Université du Québec Institut National de la Recherche Scientifique Centre Eau Terre Environnement

Développement d'une approche de suivi des fleurs d'eau d'algues à l'aide de l'imagerie désagrégée du capteur MODIS, adaptée aux lacs du Québec méridional

Par

El Alem Anas

Thèse présentée pour l'obtention du grade de Philosophiae doctor (Ph.D.) en sciences de l'eau

Jury d'évaluation

	Président du jury et examinateur externe	Yannick Huot Université de Sherbrooke
	Examinateur externe	Martin Montes-Hugo Département d'océanographie UQAR
	Examinatrice interne	Monique Bernier INRS Centre Eau-Terre-Environnement
	Directeur de recherche	Karem Chokmani INRS Centre Eau-Terre-Environnement
	Codirectrice de recherche	Isabelle Laurion INRS Centre Eau-Terre-Environnement
	Codirecteur de recherche	Salah Eddine El Adlouni Moncton University
© Dro	its réservés de <i>(Anas El Alem)</i> , 2014	

DÉDICACE

¹ Pour mes parents, pour mes frères et pour ma sœur, que dieu les protègent tous.

REMERCIEMENTS

La présente thèse est le fruit de six années passées au sein de l'équipe du laboratoire de télédétection du Centre Eau-Terre-Environnement (ETE) de l'Institut National de la Recherche Scientifique (INRS). Ce travail n'aurait pas vu le jour sans la contribution ou l'encouragement de plusieurs personnes dont la liste est très longue. Les remercier comme il se doit et réussir à exprimer toute ma gratitude envers eux avec des mots justes me semble aujourd'hui bien plus délicat que de rédiger cette thèse. Alors merci à :

Monsieur **Karem Chokmani**, pour votre confiance, patience et soutien qu'il soit scientifique, moral ou financier. Merci de m'avoir accordé la chance de travailler sur cette thématique à caractère multidisciplinaire et époustouflante. Merci également de m'avoir accordé le statut de pionnier au sein de notre équipe de recherche dans l'axe de la modélisation des «fleurs d'eau d'algues». Depuis le début, j'ai su que le défi était énorme, mais je savais aussi que je pouvais compter sur un chercheur dévoué, perspicace, rigoureux et un homme humainement exceptionnel afin de m'épauler et de m'assister à relever ce défi avec excellence.

Madame **Isabelle Laurion**, pour son enthousiasme, sa disponibilité, sa générosité de partage de son expertise en limnologie qui a amplement aidé à améliorer la qualité des travaux de cette thèse. Merci aussi au temps accordé et à la finesse des corrections qu'il soit aux articles ou aux différentes sections de cette thèse. En résumé, merci pour l'attention que vous m'avez manifestée tout au long de mes études à l'INRS.

Monsieur **Salah Eddine El Adlouni**, que puis-je dire pour vous exprimer mon plaisir de vous avoir eu comme co-directeur et ma reconnaissance envers votre personne ? Merci de m'avoir été toujours présent, malgré la distance, et d'avoir été à l'écoute et pour vos prestigieuses remarques et commentaires en matière statistique et vos conseils humains.

Aux piliers de notre équipe de recherche, Madame Monique Bernier et Messieurs Yves Gauthier, Simon Perreault, Andres Jacome et Jimmy Poulin. Toute l'équipe de télédétection de l'INRS-ETE, spécifiquement aux Mesdames Chaima Touati, Kim-Huong Hoang, Parvin Kalantari et Sophie Roberge et aux Messieurs Khalid Oubennaceur et Sena Noumonvi-Yawu. En votre compagnie, j'ai beaucoup apprécié le climat d'entre aides qui règne dans notre équipe, la notion de communication et des relations interpersonnelles exemplaires. Félicitation à Madame Kim-Huong pour avoir soutenu sa thèse en juillet 2014 et bon courage à vous toutes et à vous tous.

Aux Mesdames Sylvie Blais et Nathalie Bourbonnais du Ministère du Développement Durable, Environnement et Lutte contre les Changements Climatiques pour le temps qui ont fourni à préparer les bases de données et aussi pour leur disponibilité.

Mesdames : Bouchra Nasri, Dorra Hammami, Fatima Laanaya, Hind El Housni, Imen Ghannem, Malika Bendouz et Rym El Khourchani, et Messieurs : Ahmed Boulahcen, Ahmed Hatitou, Djeed Benbouzid, Mohamed Alaoui, Mohamed badouli, Mostafa Talbi, Jamal Fenane, Jawad Grar, Tarik et Nidal Ziane et Youssef El Khamlichi, ... Bref, À tous mes ami(e)s.

Finalement, merci aux membres du jury d'évaluation Madame Monique Bernier (INRS-ETE) et Messieurs Martin Montes-Hugo (UQAR) et Yannick Huot (US) d'avoir accepté de juger ce travail ainsi que de leur patience à lire ce long document.

RÉSUMÉ

L'occurrence et l'intensité de l'extension des fleurs d'eau d'algues et/ou de cyanobactéries (FEA) ont connu un développement important au cours des dernières années dans les plans d'eau douce partout sur la planète. Bien que ces microorganismes aient toujours fait partie intégrante des écosystèmes aquatiques, leur prolifération reflète un état avancé d'eutrophisation et causent divers dommages variant d'une simple perturbation à un débalancement plus sérieux et une réduction des services rendus par ces écosystèmes. Le moyen le plus utilisé actuellement pour détecter la présence des FEA est l'échantillonnage in situ. Toutefois, ce processus de surveillance, laborieux et coûteux, demeure limité dans le temps et dans l'espace. Par sa vision synoptique, la télédétection satellite s'avère être un outil intéressant pour fournir la couverture spatiale et la fréquence temporelle nécessaire pour suivre et mieux comprendre les facteurs de contrôle des FEA. L'activité bio-optique de la chlorophylle-a (Chl-a), pigment principal des algues et des cyanobactéries, est caractérisée par une signature spectrale typique détectable par les capteurs satellitaires. L'objectif de cette étude était de développer une approche satellitaire d'estimation de la concentration en Chl-a adaptée aux lacs du Québec méridional permettant la surveillance en temps quasi réel de la dynamique spatiale des FEA, via l'imagerie du capteur MODIS (Moderate Resolution Imaging Spectroradiometer) désagrégée à 250 mètres de résolution spatiale. L'étude a permis de mettre en évidence que : 1) les 7 premières bandes désagrégées du capteur MODIS, non-conçues initialement pour l'étude des propriétés de l'eau, ont la capacité de détecter les FEA dans les eaux douces avec un niveau de précision satisfaisant, en particulier pour les eaux fortement chargées en Chl-a (R² jusqu'à 0.93); 2) les modèles standards (Gitelson, FAI (*floating algae index*), Kahru et APPEL (APProach by ELimination)), lorsqu'ils sont appliqués sur les images MODIS désagrégées, ne permettent pas de produire des estimations précises de la concentration en Chl-a (R^2 jusqu'à 0.11) durant la phase d'initiation des FEA (< 50 mg Chl-a m⁻³); 3) la détermination *a priori* de la classe de concentration en Chl-a d'un pixel donné augmente significativement la précision des estimations, non seulement pour les fortes concentrations ($R^2 = 0.96$) mais aussi durant la phase d'initiation des FEA

 $(R^2 = 0.56)$ et 4) l'utilisation des méthodes d'ensemble permet d'améliorer la qualité et la robustesse des estimations quelle que soit la charge en Chl-a ($R^2 = 0.98$ pour les fortes floraison et $R^2 = 0.77$ durant leur d'initialisation) tout en fournissant un niveau d'incertitude sur l'estimation. En effet, tout processus de modélisation est une simplification de la réalité et est porteur d'erreurs, le contrôle de ces erreurs ainsi que leur prise en considération par les méthodes d'ensemble lors de l'estimation finale de la concentration en Chl-a ont significativement amélioré la qualité de la modélisation des FEA, particulièrement durant leur phase d'initiation. Le développement d'une telle approche se veut complémentaire aux méthodes standards de suivi puisqu'elle permet d'identifier les plans d'eau affichant des charges chlorophylliennes problématiques sur un territoire aussi étendu que le Québec méridional (les images MODIS couvrent une scène de plus de 23,300 km²). De plus, l'automatisation de cette approche, pour des fins opérationnelles, permettra aux intervenants de suivre quotidiennement la dynamique spatiale des FEA, et ce en temps quasi réel puisque les images MODIS sont disponibles en ligne environ deux heures après leur acquisition.

TABLE DES MATIÈRES

PA	RTIE 1	: SYNTHÈSE	1
1		INTRODUCTION GÉNÉRALE	3
	1.1	MISE EN CONTEXTE	3
	1.2	PROBLEMATIQUE	4
	1.3	HYPOTHESES ET OBJECTIFS	6
	1.4	STRUCTURE DE LA THESE	8
2		ÉTAT DE L'ART	11
	2.1	LES FLEURS D'EAU D'ALGUES ET/OU DE CYANOBACTERIES	11
	2.1.1	Écologie	11
	2.1.2	Toxines	12
	2.2	PROLIFERATION DES CYANOBACTERIES	13
	2.3	METHODES DE SUIVI DES PROLIFERATIONS D'ALGUES : APPROCHES STANDARDS	15
	2.3.1	Caractéristiques physico-chimiques du plan d'eau	15
	2.3.2	Suivi direct des populations phytoplanctoniques	17
	2.3.3	Surveillance par le réseau de volontaire	19
	2.4	METHODES DE SUIVI DES PROLIFERATIONS D'ALGUES : TELEDETECTION	20
	2.4.1	Cartographie qualitative	22
	2.4.2	Estimation quantitative	24
3		METHODOLOGIE GENERALE	33
	3.1	ZONE D'ETUDE	33
	3.2	DONNEES IN SITU	34
	3.3	DONNEES MODIS	38
	3.4	PRETRAITEMENTS DES IMAGES	39
	3.4.1	Désagrégation des bandes 3 à 7	40
	3.4.2	Re-projection	41
	3.4.3	Correction atmosphérique	43
	3.4.4	Masque de l'eau	44
	3.4.5	Masque de nuages	47
	3.4.6	Indices d'évaluation statistiques	48
4		ÉTUDE DU POTENTIEL DES 7 PREMIERES BANDES DU CAPTEUR MODIS	
DE	SAGRE	GEES A 250 M DE RESOLUTION SPATIALE A SUIVRE LES FEA DANS LES EAUX	

DOUCES 51

	4.1		51
	4.2	APPROCHE METHODOLOGIQUE	52
	4.3	RESULTATS ET DISCUSSION	54
	4.4	Conclusions	57
5		DEVELOPPEMENT D'UN MODELE ADAPTATIF DE SUIVI DES FEA DANS LES EAUX	
DC	OUCES	EN UTILISANT LES DONNEES DESAGREGEES DU CAPTEUR MODIS	59
	5.1	INTRODUCTION	59
	5.2	APPROCHE METHODOLOGIQUE	60
	5.3	RESULTATS ET DISCUSSION	62
	5.4	CONCLUSIONS	69
6		DEVELOPPEMENT D'UN MODELE BASE SUR LES METHODES D'ENSEMBLE POUR	ł
Ľ	ESTIMA	ATION DE LA CONCENTRATION EN CHLOROPHYLLE-A	71
	6.1	INTRODUCTION	71
	6.2	APPROCHE METHODOLOGIQUE	72
	6.3	RESULTATS ET DISCUSSION	77
	6.4	CONCLUSIONS	80
7		CONCLUSION GENERALE	83
	7.1	ÉVALUATION DES HYPOTHESES	83
	7.2	LIMITES	87
	7.3	ASPECTS INNOVATEURS, CONTRIBUTIONS ET RETOMBEES	87
	7.4	PERSPECTIVES	89
RÉ	FÉREI	VCES	91
PA	RTIE 2	? : ARTICLE-1	.101
1		INTRODUCTION	.109
2		MATERIAL AND METHODS	.112
	21	ΙΝ SITU DATA	112
	2.1		114
2	_		
3		ALGORITHMS USED TO ESTIMATED CHL-A CONCENTRATION	. 1 1 5
	3.1	Kahru's Model	. 116
	3.2	FLOATING ALGAE INDEX (FAI)	. 117
	3.3	GITELSON MODEL	. 118
	3.4	NOVEL APPROACH APPEL (APPROACH BY ELIMINATION)	. 120
4		ACCURACY ASSESSMENT	.122

5		RESULTS AND DISCUSSION	123
	5.1	EVALUATION OF DOWNSCALING THE MODIS BAND SIGNAL	123
	5.2	MODEL CALIBRATION	124
	5.3	MODEL VALIDATION	125
	5.4	QUALITATIVE EVALUATION	132
6		CONCLUSIONS	
AC	KNOW	/LEDGEMENTS	
RE	FEREN	NCES AND NOTES	141
РА	RTIE 3	3 : ARTICLE-2	147
1		INTRODUCTION	155
2		MATERIALS AND METHODS	157
	2.1	STUDY AREA AND IN SITU DATA	157
	2.2	MODIS DATA	159
	2.3	ADAPTIVE MODEL PARAMETERIZATION	162
	2.4	ACCURACY ASSESSMENT AND VALIDATION DATA	166
3		RESULTS AND DISCUSSION	
	3.1	CALIBRATION	168
	3.2	EVALUATION OF <i>ESTIMATORS</i>	170
	3.3	EVALUATION OF THE ADAPTIVE MODEL: CROSS-VALIDATION	171
	3.4	VALIDATION BY INDEPENDENT DATA	174
	3.5	QUALITATIVE VALIDATION: MODEL'S APPLICATION	177
4		CONCLUSIONS	
AC	KNOW	/LEDGMENTS	
AU	THOR	CONTRIBUTIONS	
co	NFLIC	TS OF INTEREST	
RE	FEREN	NCES	
РА	RTIE 4	: ARTICLE-3	
1		INTRODUCTION	
2		STUDY AREA AND IN SITU DATA	203
	2.1	CALIBRATION DATABASE	203
	2.2	VALIDATION DATABASE	204

3		MODIS IMAGES PRE-PROCESSING	
	3.1	SPATIAL RESOLUTION DOWNSCALING	
	3.2	ATMOSPHERIC CORRECTION	
	3.3	LAND/WATER MASK	
4		THE ADAPTIVE MODEL	211
5		METHODOLOGICAL APPROACH AND EVALUATION INDICES	213
	5.1	ENSEMBLE-BASED SYSTEM PARAMETERIZATION	213
	5.2	ACCURACY ASSESSMENT	219
6		RESULTS AND DISCUSSION	
	6.1	CALIBRATION OF THE SPECIFIC ESTIMATORS	
	6.2	ENSEMBLE-BASED SYSTEM QUANTITATIVE EVALUATION	
	6.2.1	The k-fold cross-validation	223
	6.2.2	Confusion matrix: (independent database)	225
	6.2.3	8 Ensemble-based system qualitative evaluation	226
	6.2.4	Comparison with MODIS and MERIS chlorophyll-a standard products	228
7		CONCLUSIONS	232
RE	FEREN	ICES	

LISTE DES TABLEAUX

TABLEAU 1-1 SITUATION DES PLANS D'EAU QUEBECOIS FACE AUX PROLIFERATIONS DES FLEURS D'EAU D'ALGUES
ET/OU DE CYANOBACTERIES AU COURS DES SIX DERNIERES ANNEES (MDDEP, 2002)4
TABLEAU 3-1 NOMS, SURFACES, ET DENOMBREMENT MINIMAL ET MAXIMAL DE LA DENSITE EN CYANOBACTERIES
ENTRE L'ANNEE 2007 ET 2010 DE L'ENSEMBLE DES PLANS D'EAU AYANT SERVI A LA VALIDATION DES
APPROCHES DEVELOPPEES
TABLEAU 3-2 CARACTERISTIQUES DES 16 PREMIERES BANDES DU CAPTEUR MODIS
TABLEAU 3-3 DIAGRAMME SIMPLIFIE DE LA MATRICE DE CONFUSION ET DES PARAMETRES IMPLIQUES DANS SON
CALCUL AINSI QUE DANS LE CALCUL DE L'INDICE KAPPA, OU A ET D SONT LES NOMBRES DES ESTIMATIONS
BIEN-CLASSEES, B ET C SONT LES NOMBRES DES ESTIMATIONS MAL-CLASSEES, ${f n1, n2, n3}$ ET ${f n4}$ SONT
RESPECTIVEMENT LES TOTAUX A+B, C+D, A+C ET B+D, N EST LA TAILLE DE LA BASE DE DONNEES, ${f P0}$ ET ${f Pc}$
SONT TELS QUE DEFINIS DANS LES EQUATIONS (3-7) ET (3-8)49
TABLEAU 5-1. RESULTAT DE LA MATRICE DE CONFUSION DU MODELE ADAPTATIF (MA). 65
TABLEAU 5-2. RESULTAT DE LA MATRICE DE CONFUSION D'APPEL. 65
TABLEAU 5-3. RESULTAT DE LA MATRICE DE CONFUSION DU MODELE FAI. 66
TABLEAU 5-4. RESULTAT DE LA MATRICE DE CONFUSION DU MODELE KAHRU. 66
TABLEAU 6-1 ABSCISSES ET POIDS DE LA DISTRIBUTION NORMAL STANDARD
TABLEAU 6-2 RESULTAT DE LA MATRICE DE CONFUSION DU MODELE BASE SUR LA METHODE D'ENSEMBLE
TABLEAU 2-1 CHARACTERISTICS OF THE MODERATE RESOLUTION IMAGING SPECTRORADIOMETER (MODIS)
BANDS USED IN THE PRESENT STUDY115
TABLEAU 5-1. MODEL PERFORMANCE USING CROSS-VALIDATION RESULTS FROM THE WHOLE DATA SERIES126
TABLEAU 5-2. MODEL PERFORMANCE USING CROSS-VALIDATION RESULTS FOR THE DATA SUBSET CHL-A > 50
MG·M ⁻³
TABLEAU 5-3. MODEL PERFORMANCE USING CROSS-VALIDATION RESULTS FOR THE DATA SUBSET WHEN CHL-A IS
BETWEEN 10 AND 50 MG·M ⁻³
TABLEAU 2-1. CHARACTERISTICS OF THE MODIS BANDS USED IN THE PRESENT STUDY. NIR = NEAR INFRARED;
SWIR = SHORTWAVE INFRARED; LCAB = LAND/CLOUD/AEROSOL BOUNDARIES; AND LCAP =
LAND/CLOUD/AEROSOL PROPERTIES160
TABLEAU 2-2. EQUATIONS OF THE THREE CALIBRATED MODELS (OR ESTIMATORS) USING A MULTIVARIATE
REGRESSION TO ESTIMATE CHL-A CONCENTRATION
TABLEAU 2-3. SIMPLIFIED DIAGRAM OF THE PARAMETERS USED IN THE CONFUSION MATRIX AND TO CALCULATE
THE KAPPA INDEX: A AND D ARE THE NUMBER OF WELL-CLASSIFIED VALUES, B AND C ARE THE NUMBER OF
MIS-CLASSIFIED VALUES, N1, N2 , N3, AND N4 ARE RESPECTIVELY TOTALS OF A + B, C + D, A + C, AND B + D, N
IS THE SAMPLE SIZE, $Po = a + dN$, and $Pc = n1 imes n3 + n2 imes n4N2$ 168

TABLEAU 3-1. EVALUATION OF THE THREE MODELS (ESTIMATORS) USING CROSS-VALIDATION TECHNIQUE171
TABLEAU 3-2. THE ADAPTIVE MODEL CONFUSION MATRIX RESULTS. 175
TABLEAU 3-3. THE APPEL CONFUSION MATRIX RESULTS. 176
TABLEAU 3-4. THE FAI CONFUSION MATRIX RESULTS. 176
TABLEAU 3-5. THE KAHRU CONFUSION MATRIX RESULTS
TABLEAU 1-1. MOST CURRENT OCEAN COLOR SENSORS. 201
TABLEAU 2-1. LIST OF LAKES AND RESERVOIRS USED IN THE VALIDATION OF THE ENSEMBLE-BASED SYSTEM206
TABLEAU 3-1. CHARACTERISTICS OF THE MODIS BANDS USED IN THE PRESENT STUDY207
TABLEAU 5-1. ABSCISSAS AND WEIGHTS FOR THE STANDARD NORMAL DISTRIBUTION214
TABLEAU 5-2. SIMPLIFIED DIAGRAM OF THE PARAMETERS USED IN THE CONFUSION MATRIX AND TO CALCULATE
THE KAPPA INDEX: A AND D ARE THE NUMBER OF WELL-CLASSIFIED VALUES, B AND C ARE THE NUMBER OF
MIS-CLASSIFIED VALUES, N1, N2 , N3, AND N4 ARE RESPECTIVELY TOTALS OF A + B, C + D, A + C, AND B + D, N
IS THE SAMPLE SIZE, $Po = a + dN$, and $Pc = n1 imes n3 + n2 imes n4N2$ 221
TABLEAU 6-1. THE ENSEMBLE BASED SYSTEM CONFUSION MATRIX RESULTS. (FALSE POSITIVES AND NEGATIVE
WERE CALCULATED USING RED AND BROWN NUMBERS, RESPECTIVELY)225
TABLEAU 6-2. THE ADAPTIVE MODEL CONFUSION MATRIX RESULTS. (FALSE POSITIVES AND NEGATIVE WERE
CALCULATED USING RED AND BROWN NUMBERS, RESPECTIVELY)

LISTE DES FIGURES

- FIGURE 2-1 ARBRE DECISIONNEL PUBLIE PAR L'ORGANISATION MONDIALE DE LA SANTE (CHORUS *ET AL.*, 1999). 14

- FIGURE 2-4 LES PLANS D'EAU REÇOIVENT L'ECLAIREMENT SOLAIRE (ESOLEIL) ET DU CIEL (ECIEL). LE RAYONNEMENT TOTAL RETOURNE DU PLAN D'EAU VERS LE CAPTEUR EST UNE FONCTION DU RAYONNEMENT ATMOSPHERIQUE DIFFUSE (LP), DU RAYONNEMENT DE L'EAU DE SURFACE (LS), DU RAYONNEMENT VOLUMETRIQUE DE SUB-SURFACE (LV) ET DU RAYONNEMENT DE FOND (LB) DU PLAN D'EAU (BUKATA *ET AL.*, 1995). 21

FIGURE 2-5	FLEUR D'EAU D'ALGUE QUI A ETE DETECTEE AU LAC ERIE EN OHIO, 2011	23
FIGURE 2-6	FLORAISON DETECTEE PAR LE CAPTEUR MERIS EN MER DE BARENTS (NORVE	GE), 201123

- FIGURE 2-7 HAUTEUR DE LA LIGNE DE BASE DE FLUORESCENCE (FLH), INSPIREE DE GILERSON *ET AL.* (2007). 26

FIGURE	4-2	CONCENTRATION EN CHLOROPHYLLE-A ESTIMEE A PARTIR DES QUATRE MODELES COMPAREE AUX
M	ESURE	ES IN SITU POUR LES DONNEES ENTRE 10 ET 50 MG CHL-A M ⁻³ , AVEC LES INDICES DE PERFORMANCE
DE	ES MOI	DELES
FIGURE	5-1	SIGNATURES SPECTRALES CARACTERISTIQUES DES TROIS CONDITIONS DE CONCENTRATIONS DE LA
С	HL-A:	A) CONCENTRATIONS FAIBLES ET MOYENNES ET B) CONCENTRATIONS FORTES
FIGURE	5-2	RESULTATS DE LA SUBDIVISION DE LA BASE DE CALIBRATION EN TROIS SOUS-GROUPES
SF	PECIFIC	QUES. 1) LE SOUS-GROUPE DE DONNEES QUI A ETE UTILISE POUR CALIBRER L' <i>ESTIMATEUR</i>
SF	PECIFIC	QUE DES EAUX FORTEMENT CHARGEES EN CHL-A, 2) LE SOUS-GROUPE DE DONNEES QUI A ETE
U	TILISE	POUR CALIBRER L'ESTIMATEUR SPECIFIQUE DES EAUX MOYENNEMENT CHARGEES EN CHL-A ET 3) LE
SC	OUS-G	ROUPE DE DONNEES QUI A ETE UTILISE POUR CALIBRER L'ESTIMATEUR SPECIFIQUE DES EAUX
FA	AIBLEN	IENT CHARGEES EN CHL-A61
FIGURE	5-3	STRUCTURE INTERNE DU MODELE ADAPTATIF62
FIGURE	5-4	CONCENTRATION EN CHLOROPHYLLE-A ESTIMEE A PARTIR DES QUATRE MODELES COMPAREE AUX
M	ESURE	ES IN SITU POUR L'ENSEMBLE DE LA BASE DE DONNEES, AVEC LES INDICES DE PERFORMANCE DES
M	ODELE	S63
FIGURE	5-5	CONCENTRATION EN CHLOROPHYLLE-A ESTIMEE A PARTIR DES QUATRE MODELES COMPAREE AUX
M	ESURE	ES IN SITU EN UTILISANT LES CONCENTRATIONS INFERIEURES A 50 MG CHL-A M-3, AVEC LES INDICES
DE	E PERF	ORMANCE DES MODELES
FIGURE	5-6	APPLICATION DES QUATRE MODELES (MODELE ADAPTATIF ET LES MODELES APPEL, FAI ET
K	AHRU)	A UNE SERIE D'IMAGES MODIS ACQUISES DURANT L'ETABLISSEMENT D'UNE FEA SUR LA BAIE
М	lissisc	QUOI DU LAC CHAMPLAIN (LA LEGENDE EST DANS ECHELLE LOGARITHMIQUE)
FIGURE	6-1	SCHEMA ILLUSTRATIF DE L'UTILISATION DE LA QUADRATURE GAUSSIENNE SUR DEUX DIMENSIONS
El	N UTILI	SANT DEUX VARIABLES (${f SVis} - {f PIR}$ et ${f SR} - {f CLIR})$ et son application afin de calibrer a la
FC	DIS UN	CLASSIFICATEUR BASE SUR L'ENSEMBLE (LIGNES VERTES, ROUGES ET BLEUES) ET UN ESTIMATEUR
BA	ASE SL	IR L'ENSEMBLE. PAR EXEMPLE: LE SOUS-GROUPE DE DONNEES ENCERCLE PAR L'ELLIPSE BLEUE
C	ORRES	POND AUX DONNEES UTILISEES POUR CALIBRER UN DES ESTIMATEURS SPECIFIQUES DESIGNE A
ES	STIMEF	R LES FORTES CONCENTRATIONS EN CHLOROPHYLLE-A ET LE SOUS-GROUPE DE DONNEES ENCERCLE
PA	AR L'EL	LIPSE ROUGE CORRESPOND AUX DONNEES UTILISEES POUR CALIBRER UN DES ESTIMATEURS
SF	PECIFIC	QUES DESIGNE A ESTIMER LES FAIBLES CONCENTRATIONS EN CHLOROPHYLLE-A76
FIGURE	6-2	CONCENTRATION EN CHLOROPHYLLE-A ESTIMEE A PARTIR DU MODELE BASE SUR LA METHODE DE
L'I	ENSEM	IBLE COMPAREE AUX MESURES IN SITU POUR : A) L'ENSEMBLE DE LA BASE DE DONNEES ET B) EN
U	TILISAN	NT LES CONCENTRATIONS INFERIEURES A 50 MG CHL-A M^{-3} , AVEC LES INDICES DE PERFORMANCE
DE	ES MOI	DELES
FIGURE	6-3	APPLICATION DES DEUX MODELES (MODELE BASE SUR LA METHODE DE L'ENSEMBLE ET LE MODELE
A	DAPTA	TIF (MA)) SUR LES MEMES IMAGES UTILISEES A LA (FIGURE 5-6) ; RAPPELONS QUE LA LEGENDE EST
DA	ANS UN	NE ECHELLE LOGARITHMIQUE

FIGURE 6-4 EXEMPLE DE L'APPLICATION DU MODELE BASE SUR LA METHODE DE L'ENSEMBLE (MBME) SUR UNE
FLORAISON DETECTEE LE 19 SEPTEMBRE 2001 ET DE L'EVALUATION DE SES ESTIMATIONS PAR LE
COEFFICIENT DE VARIATION
FIGURE 2-1. GEOGRAPHIC DISPOSITION OF THE STUDIED LAKES
FIGURE 2-2. WATER SAMPLING STATIONS ON THE FOUR STUDIED LAKES
FIGURE 3-1. SIMPLIFIED REFLECTANCE SPECTRA OF WATER AND PHYTOPLANKTON (OBTAINED FROM
HTTP://WWW.A-A-R-S.ORG/ACRS/PROCEEDING/ACRS1994/PAPERS/AGS94-5.HTM)121
FIGURE 5-1. CORRELATION BETWEEN THE ORIGINAL AND THE DOWNSCALED REFLECTANCE USED IN THE
ESTIMATION MODELS OF CHLOROPHYLL-A CONCENTRATION. (A) BAND 3, (B) BAND 4, AND (C) BAND 5123
FIGURE 5-2. EXPONENTIAL REGRESSION OBTAINED BETWEEN CHLOROPHYLL-A CONCENTRATION (WHOLE DATA
SET) AND THE OUTPUTS OF THE MODIS REFLECTANCE BANDS FOR THE FOUR MODELS (Y-AXIS OF INSERTS
IS ON A LOGARITHMIC SCALE)
FIGURE 5-3. EXPONENTIAL REGRESSION OBTAINED BETWEEN CHLOROPHYLL-A CONCENTRATION (> 10 $MG \cdot M^{-3}$)
AND THE OUTPUTS OF THE MODIS REFLECTANCE BANDS FOR THE FOUR MODELS (Y-AXIS OF INSERTS IS ON
A LOGARITHMIC SCALE)
FIGURE 5-4. CHLOROPHYLL-A CONCENTRATION ESTIMATED FROM THE FOUR MODELS COMPARED TO IN SITU
MEASUREMENTS (DATA SET CHL-A > 10 $\text{MG} \cdot \text{M}^{-3}$)
FIGURE 5-5. CHLOROPHYLL-A CONCENTRATION AND CYANOBACTERIAL DENSITY USED IN DECISION TREE
PUBLISHED BY THE WORLD HEALTH ORGANIZATION
FIGURE 5-6. CHLOROPHYLL-A CONCENTRATION ESTIMATED FROM THE FOUR MODELS COMPARED TO IN SITU
MEASUREMENTS UNDER HEAVY BLOOM CONDITIONS (CHL-A > 50 MG·M ⁻³)
FIGURE 5-7. CHLOROPHYLL-A CONCENTRATION ESTIMATED FROM THE FOUR MODELS COMPARED TO IN SITU
MEASUREMENTS UNDER MODERATE CONCENTRATIONS (BETWEEN 10 AND 50 $\text{MG} \cdot \text{M}^{-3}$)132
FIGURE 5-8. COMPARISON BETWEEN TRUE-COLOR IMAGE FROM MODIS SENSOR AND THE CORRESPONDING FIELD
MAP PREPARED BY THE MINISTÈRE DU DÉVELOPPEMENT DURABLE DE L'ENVIRONNEMENT ET DES PARCS
(MDDEP) TECHNICIANS AT MISSISQUOI BAY ON LAKE CHAMPLAIN.
FIGURE 5-9. MODIS IMAGES OF MISSISQUOI BAY ON LAKE CHAMPLAIN ON THREE DATES AND CORRESPONDING
ESTIMATES OF CHLOROPHYLL-A CONCENTRATION BY THE FOUR MODELS.
FIGURE 5-10. COMPARISON BETWEEN ESTIMATED CHLOROPHYLL-A CONCENTRATION CALCULATED BY THE FOUR
MODELS AND IN SITU CONCENTRATION OBTAINED BY THE MDDEP AT TWO STATIONS ON MISSISQUOI BAY ON
LAKE CHAMPLAIN
FIGURE 2-1. GEOGRAPHIC LOCATION OF THE WATER BODIES USED FOR MODEL CALIBRATION AND VALIDATION157
FIGURE 2-2. WATER SAMPLING STATIONS ON THE FOUR STUDIED LAKES USED FOR CALIBRATION
FIGURE 2-3. HISTOGRAM OF THE FREQUENCY OF CHLOROPHYLL-A VALUES OBSERVED (COMPLETE DATA SET,
N = 363)159

FIGURE 2-4. COMPARISON OF THE MODIS SIGNAL FROM LAKE CHAMPLAIN DURING AN ALGAL BLOOM ON 19
SEPTEMBER 2001 WITH DIFFERENT ATMOSPHERIC CORRECTION MODELS
FIGURE 2-5. SPECTRAL SIGNATURE BEHAVIOUR FOR THREE CHLOROPHYLL-A CONCENTRATION CLASSES: (A) LOW
(<10 Mg·M ⁻³) AND MODERATE (10-50 Mg·M ⁻³) CONCENTRATIONS COMBINED; AND (B) HIGH
CONCENTRATIONS (>50 MG·M ⁻³). DIFFERENT COLORS REPRESENT THE SIGNATURE OF INDIVIDUAL SAMPLES.
FIGURE 2-6. THE SPLITTING ALGORITHM OF THE CLASSIFICATION AND REGRESSION TREE (CART), WHERE
tParent, $tLeft - child$, and $tRight - child$ are parent, left, and right nodes, xj is the splitting
VARIABLE J, AND \mathbf{xjB} IS THE BEST SPLITTING VALUE OF THE \mathbf{xj}
FIGURE 2-7. THRESHOLDS VALUES ($\times 10^{6}$) USED TO DISTINGUISH BETWEEN THE THREE CHLOROPHYLL-A
BLOOMING CLASSES USING THE CLASSIFICATION AND REGRESSION TREE METHOD (CART)165
FIGURE 3-1. RESULTS OF MULTIVARIATE REGRESSION ADJUSTMENTS BETWEEN THE MEASURED CHLOROPHYLL-A
CONCENTRATIONS AND THE RETURN SIGNALS OF MODIS IMAGES FOR THE THREE BLOOMING CLASSES: (A)
LOW; (B) MODERATE; (C) HIGH CHL-A CONCENTRATIONS. THE SCALE OF THE Y-AXIS IN THE INSERT IS
LOGARITHMIC TO BETTER ILLUSTRATE THE CORRELATION
FIGURE 3-2. CHLOROPHYLL-A CONCENTRATION ESTIMATED FROM THE FOUR MODELS COMPARED TO IN SITU
MEASUREMENTS FOR THE COMPLETE DATABASE, WITH MODEL PERFORMANCE INDICES172
FIGURE 3-3. CHLOROPHYLL-A CONCENTRATION ESTIMATED FROM THE FOUR MODELS COMPARED TO IN SITU
MEASUREMENTS FOR THE DATABASE USING ONLY VALUES ${<}50$ MG CHL-A M $^{-3}$
FIGURE 3-4. APPLICATION OF THE FOUR MODELS (ADAPTIVE MODEL, APPEL, FAI, AND KAHRU) TO A SERIES OF
MODIS IMAGES COLLECTED DURING THE ESTABLISHMENT OF A HAB ON MISSISQUOI BAY, LAKE
CHAMPLAIN, COMPARED TO THE TRUE COLOR COMPOSITE IMAGES (RGB FOR RED, GREEN, BLUE). THE RED
POLYGON INDICATES THE SHORELINE AND SOUTHERN BOUNDARY OF THE BAY. CHLOROPHYLL-A
CONCENTRATIONS ARE IN A NAPIERIAN LOGARITHMIC SCALE
FIGURE 3-5. COMPARISON BETWEEN ESTIMATED CHLOROPHYLL-A CONCENTRATION CALCULATED BY THE
ADAPTIVE MODEL, APPEL, FAI, AND KAHRU MODELS AND IN SITU MEASUREMENTS OBTAINED BY THE
MDDEFP AT TWO STATIONS ON MISSISQUOI BAY ON LAKE CHAMPLAIN (19 SEPTEMBER 2001)180
FIGURE 1-1. RELATIONSHIP, IN A LOGARITHMIC SCALE, BETWEEN CYANOBACTERIA DENSITY AND CHLOROPHYLL-A
CONCENTRATION. THE DATABASE IS COLLECTED BY THE MINISTERE DU DEVELOPPEMENT DURABLE, DE
L'ENVIRONNEMENT ET DE LA LUTTE CONTRE LES CHANGEMENTS CLIMATIQUES OF QUEBEC OVER NINE
YEARS (2000-2008) FROM SEVEN DIFFERENT LAKES
FIGURE 2-1. WATER SAMPLING STATIONS ON THE FOUR STUDIED LAKES USED FOR CALIBRATION: 1) MISSISQUOI
BAY OF CHAMPLAIN LAKE, 2) BROME LAKE, 3) WILLIAM LAKE, AND 4) NAIRNE LAKE
FIGURE 2-2. GEOGRAPHIC LOCATION OF THE WATER BODIES USED FOR MODEL CALIBRATION AND VALIDATION205
FIGURE 3-1. TRUE COLOR COMPOSITE OF ST-JEAN LAKE USING: A) MODIS-D-250 IMAGES (LCC PROJECTION)
AND B) 500 M SPATIAL RESOLUTION MODIS LANDS BANDS (SINUSOIDAL PROJECTION)207

FIGURE 3-2. COMPARISON BETWEEN LAND/WATER MASKS COVERING THE MISSISQUOI BAY OF CHAMPLAIN LAKE PRODUCED BY: A) THE ORIGINAL 250 M SPATIAL RESOLUTION NIR MODIS BAND, B) THE LAND/WATER MASK PRODUCED BY THE NASA, C) THE MODIS-D-250 NIR BAND, AND E) THE PROPOSED LAND/WATER MASK.

240

FIGURE 3-3. THE REFLECTANCE HISTOGRAM OF THE NIR COMPOSITE IMAGE
FIGURE 4-1. OPERATING MODE OF THE ADAPTIVE MODEL (AM). THE SUBGROUPS OF DATA SURROUNDED BY THE
BLACK, GREEN, AND BLUE DASHED LINES REPRESENT THE TRAINING DATABASE USED TO CALIBRATE THE
ESTIMATORS DESIGNED TO ESTIMATE HIGH, MODERATE, AND LOW CHLOROPHYLL-A CONCENTRATIONS,
RESPECTIVELY. BOXES 1 TO 3 INDICATE POSSIBLE SCENARIOS TO ESTIMATE CHLOROPHYLL-A USING THE
AM. V1 AND V2 ARE DETAILED IN APPENDIX 1212
FIGURE 5-1. THE USE SCHEME OF THE GAUSSIAN QUADRATURE FORMULA ON 2 DIMENSIONS USING TWO
VARIABLES (V1 AND V2) AND ITS APPLICATION TO CALIBRATE THE ENSEMBLE-BASED SYSTEM (EBS). RED,
BLUE, AND GREEN LINES REPRESENT RESPECTIVELY THE OPTIMAL THRESHOLDS (NOMINAL, UPPER, AND
LOWER) FOR V1 (TV1N, TV1U, AND TV1L) AND V2 (TV2N, TV2U, AND TV2L). $k(1:3)$ and $p(1:3)$ are
THE INDICES OF THE OPTIMAL THRESHOLDS RELATED TO V1 AND V2. THE SURROUNDED DATA SUBGROUP BY
THE BLUE ELLIPSE REPRESENTS THE TRAINING DATA USED TO CALIBRATE THE SPECIFIC ESTIMATOR
DESIGNED TO ESTIMATE HIGH CHLOROPHYLL-A CONCENTRATIONS (CHL-A), AND THE SURROUNDED DATA
SUBGROUP BY THE RED ELLIPSE REPRESENTS THE TRAINING DATA USED TO CALIBRATE THE SPECIFIC
ESTIMATOR DESIGNED TO ESTIMATE LOW CHL-A CONCENTRATIONS. BOXES 1 TO 13 INDICATE THE POSSIBLE
MODELING CASES TO ESTIMATE CHL-A USING THE EBS
FIGURE 5-2. EXAMPLE OF SUBGROUPS SELECTION SAMPLES (BLACK POINTS) USED TO CALIBRATE THE SPECIFIC
ESTIMATORS FOR THE CASE OF THE BLUE STAR. ARROWS INDICATE WHETHER THE SUBGROUP SELECTION IS
MADE BEFORE (SPECTRAL INDEX OF SAMPLES IS INFERIOR) OR AFTER (SPECTRAL INDEX OF SAMPLES IS
SUPERIOR) A GIVEN OPTIMAL THRESHOLD, WHETHER FOR V1 OR V2218
FIGURE 5-3. K-FOLD CROSS-VALIDATION ADOPTED IN THIS STUDY
FIGURE 6-1. CHLOROPHYLL-A CONCENTRATION ESTIMATED BY THE TWO MODELS COMPARED TO IN SITU
MEASUREMENTS FOR THE COMPLETE DATABASE, WITH THE MODEL PERFORMANCE INDICES224
FIGURE 6-2. CHLOROPHYLL-A CONCENTRATION ESTIMATED BY THE TWO MODELS COMPARED TO IN SITU
MEASUREMENTS FOR THE DATABASE ONLY INCLUDING VALUES < 50 Mg CHL-A M ⁻³ , WITH THE MODEL
PERFORMANCE INDICES

- FIGURE 6-3. COMPARISON BETWEEN THE ADAPTIVE MODEL (AM) AND ENSEMBLE BASED SYSTEMS (EBS) APPLICATION ON MODIS IMAGES. A) TRUE COLOR, B) EBS, AND C) AM (SEPT.19, 2001).227
- FROM THE MODIS IMAGE DOWNSCALED TO 250 M SPATIAL RESOLUTION (B) AND THE NASA'S MODIS

DUCT AT 4 KM SPATIAL RESOLUTION (C). CHL-A CONCENTRATIONS ARE IN A LOG SCALE. A) IS
JE COLOR IMAGE AT 250 M SPATIAL RESOLUTION230
PHYLL-A CONCENTRATIONS (CHL-A) SPATIAL DISTRIBUTION IN LAKE ERIE ON OCT. 9, 2011
IS IMAGE DOWNSCALED TO 250 M SPATIAL RESOLUTION (B) AND THE ESA'S MERIS
DUCT AT 300 M SPATIAL RESOLUTION (C). THE (A) IMAGE IS THE MODIS TRUE COLOR IMAGE
L RESOLUTION. CHL-A CONCENTRATIONS ARE IN A LOG SCALE

LISTE DES ABRÉVIATIONS ET DES SYMBOLES

Abréviations	Définition du terme						
FEA	Fleurs d'eau d'algues et/ou de cyanobactéries						
Chl-a	Chlorophylle-a						
MODIS	Moderate Resolution Imaging Spectroradiometer						
FAI	Floating algae index						
APPEL	APProach by ELimination						
MDDELCC	Ministère du Développement Durable, Environnement et de la Lut contre les Changements Climatiques						
PIR	Proche infrarouge						
CLIR	Courtes longueurs d'ondes de l'infrarouge						
MSS	Matière solide en suspension						
MODC	Matière organique dissoute colorée						
CZCS	Coastal Zone Color Scanner						
MERIS	MEdium Resolution Imaging Spectrometer						
ССТ	Centre Canadien de Télédétection						
IR	Infrarouge						
MA	Modèle adaptatif						
MBME	Modèle basé sur les méthodes de l'ensemble						
CBME	Classificateur basé sur les méthodes de l'ensemble						
EBME	Estimateur basé sur les méthodes de l'ensemble						
OMS	Organisation mondiale de la santé						
POI	propriétés optiques inhérentes						
RVB	Rouge vert bleu						
CASI-2	Compact Airborne Spectrographic Imager-2						
SVM	Support vector machine						
FLH	Fluorescence line height						
R	Réflectance						
λ	Longueur d'onde						
b _b	Rétrodiffusion						
f	Sensibilité de la réflectance aux variations de l'angle zénithale solaire						
Q	Propriétés bidirectionnelles de la réflectance						
a _{tot}	Absorption totale						

a _φ	Absorption par phytoplancton					
a _{PNA}	Absorption par les particules non algales					
a _{MODC}	Absorption par la MODC					
a _{eau}	Absorption par les molécules d'eau					
NDVI	Normalized difference vegetation index					
EVI	Enhanced vegetation index					
SIN	Sinusoïdal					
LCC	Conique conforme de Lambert					
SMAC	Simplified Model for Atmospheric Correction					
6S	Second Simulation of Satellite Signal in the Solar Spectrum					
MODTRAN	MODerate resolution atmospheric TRANsmittance and radiance code					
ATCOR	ATmospheric CORrection					
DOS	Dark Object Subtraction					
COST	COSine Transmission for atmospheric correction					
R ²	Coefficient de détermination					
NASHr	Critère de Nash relatif					
EQMr	Erreur quadratique moyenne relative					
BIAISr	Biais relatif					
Er	Erreur relative					
CV	Coefficient de variation					
σ	Variance					
μ	Moyenne					
S	Surface					
QG	Quadrature gaussienne					
n_{bag}	Nombre d'itérations du bagging					
n _{so}	Nombre des solutions optimales					
SO	Seuils optimaux					
CEAEQ	Centre d'expertise en analyse environnementale du Québec					

PARTIE 1 : SYNTHÈSE

1 INTRODUCTION GÉNÉRALE

1.1 Mise en contexte

Face aux éventuels changements planétaires, qu'ils soient d'origine naturelle (climatique et météorologique) ou anthropique (développement socio-économique, agricultural ou industriel), plusieurs pays se retrouvent aux prises avec la prolifération des fleurs d'eau d'algues et/ou de cyanobactéries (FEA) dans leurs systèmes aquatiques. Ce phénomène, considéré par certains auteurs de même ampleur que les autres catastrophes naturelles (tremblements de terres, glissements de terrain et tsunamis (Anderson et al., 2002)), a fait la manchette de plusieurs revues et articles scientifiques et une multitude de congrès, forums et conférences internationales (Huisman et al., 2005). Bien que ces microorganismes aient toujours fait partie intégrante de la chaîne alimentaire, leur apparition massive reflète un état avancé d'eutrophisation (AFSSA & AFSSET, 2006). Une fois établies dans un plan d'eau, les FEA sont extrêmement difficiles à contrôler, surtout lorsque composées de cyanobactéries nuisibles, et spécialement en cas de fortes concentrations. L'apparition récurrente des FEA mène à d'évidentes perturbations des écosystèmes, à la production de composés malodorants et à la dégradation de l'aspect esthétique des plans d'eau. De plus, la capacité de certaines espèces à synthétiser, emmagasiner et libérer des toxines dans l'eau, notamment les cyanotoxines, constitue un problème de taille touchant la santé publique, en particulier dans le cas où les plans d'eau représentent une réserve d'eau potable. De plus, la présence de grandes quantités de toxines dans l'eau peut mener à des pertes monétaires considérables soit de façon directe (mortalité des individus dans un contexte d'aquaculture (Kahru et al., 1993)) ou indirecte (augmentation des coûts de gestion des ressources hydriques ; traitement de l'eau potable par exemple (Huisman *et al.*, 2005)).

Le grand nombre de rapports et d'articles scientifiques portant sur l'occurrence des FEA dans les milieux lacustres québécois au cours de ces dernières années souligne la vulnérabilité de ses ressources hydriques. Alors qu'en 2007, 167 plans d'eau étaient recensés comme étant « touchés » par une cyanobactérie (fleur d'eau avec densité

3

supérieure à 20,000 cellules mL⁻¹), le nombre de ces plans d'eau est demeuré relativement élevé depuis (139 ont été touchés par une FEA de cyanobactéries en 2012, dont 48 l'ont été pour la première fois (MDDEP, 2002)). Bien que les proliférations recensées se soient probablement accrues suite à une conscientisation du public par l'intermédiaire des médias (radio, télévision, internet, ... etc.), ce phénomène n'est pas nouveau au Québec et il est évident que la problématique prend de l'ampleur d'une année à l'autre Tableau 1-1).

Tableau 1-1Situation des plans d'eau québécois face aux proliférations des fleurs d'eau
d'algues et/ou de cyanobactéries au cours des six dernières années (MDDEP,
2002).

Année	Plans d'eau visités	Plans d'eau touchés	Nouveaux cas répertoriés	Plans d'eau avec restrictions d'usages	Avec fermeture de plages
2013	121	85	20 (24%)	6	-
2012	211	139	48 (35%)	7	2
2011	209	149	36 (24%)	8	2
2010	227	150	53 (35%)	3	3
2009	226	134	38 (28%)	5	2
2008	234	116	47 (41%)	7	7
2007	273	167	117 (70%)	6	14

D'après le Ministère du Développement Durable, de l'Environnement et de la Lutte contre les Changements Climatiques (MDDELCC), 94% des plans touchés en 2012 sont constitués de lacs et de réservoirs, et ils sont localisés sur une quinzaine de régions administratives dont les plus fréquemment touchées sont les régions des Laurentides, de Lanaudière, de la Montérégie, de l'Estrie et du Saguenay Lac-Saint-Jean, parmi les régions les plus peuplées du Québec. Par conséquent, la présence d'un tel nombre de plans d'eau touchés peut présenter de sérieuses menaces pour la santé publique québécoise.

1.2 **Problématique**

Habituellement, le moyen le plus sûr et le plus utilisé pour surveiller les proliférations d'algues est l'échantillonnage *in situ* via l'identification des espèces par microscopie et

par dosage de la concentration en Chl-a et des toxines en laboratoire (par exemple. par spectroscopie et HPLC). Toutefois, ce processus de surveillance est laborieux et pour des impératifs budgétaires évidents il demeure limité dans le temps et dans l'espace. En effet, la fréquence spatio-temporelle de ces méthodes standards est insuffisante pour détecter les variations des FEA (Eija *et al.*, 1998) et les coûts alloués aux analyses en laboratoire sont souvent élevés, particulièrement dans un territoire aussi large que celui de la province de Québec. En 2012, environ \$542,000 CAN ont été dépensés pour analyser 1,058 échantillons d'algues bleu-vert récoltés sur 211 plans d'eau (MDDEP, 2002). Il est alors temps de développer des outils permettant une gestion optimale des méthodes de suivi et de surveillance des FEA dans le temps et dans l'espace en exploitant l'information produite par ces programmes d'échantillonnage.

La télédétection s'avère un outil prometteur, capable à la fois de fournir la couverture spatiale et la fréquence temporelle essentielle à la détection des fleurs d'eau pour une large diversité de plans d'eau. En effet, grâce à sa vision synoptique, sa répétitivité et sa capacité à opérer sur une large gamme du spectre électromagnétique, la télédétection optique répond parfaitement aux critères nécessaires de suivi et de surveillance des proliférations d'algues et/ou de cyanobactéries. La détection des FEA est possible grâce à l'activité bio-optique de leur pigment principal, la chlorophylle-a (Chl-a). En effet, les eaux chargées en Chl-a sont caractérisées par une faible réflectance dans le bleu et dans le rouge et une forte réflectance dans le vert, le proche infrarouge (PIR) et une partie des courtes longueurs d'ondes de l'infrarouge (CLIR). Certains capteurs satellitaires ont la capacité de détecter l'énergie réémise de la Chl-a dans ces différentes bandes spectrales. Ainsi, par l'intermédiaire de modèles bio-optiques reliant les propriétés optiques inhérentes aux propriétés optiques apparentes des eaux, il est possible de détecter et d'estimer la biomasse algale (Carder *et al.*, 1999).

Si toutefois la détection des FEA en haute mer (*case-1 waters*) est une tâche relativement simple (Ahn *et al.*, 2006, Becker *et al.*, 2009, Gitelson *et al.*, 2007, Hu, 2009, Stumpf *et al.*, 2000) puisque la Chl-a est la seule composante optiquement active (Darecki *et al.*, 2004), cette opération est beaucoup plus complexe pour les zones littorales et les eaux douces (*case-2 waters*). L'information spectrale reçue de ces eaux

5

plus complexes est généralement constituée d'une combinaison diverse de plusieurs composantes optiquement actives telles que la Chl-a du phytoplancton et des plantes aquatiques, la matière solide en suspension (MSS), la matière organique dissoute colorée (MODC), et parfois l'effet de bordure et la réflexion par le fond (Mishra *et al.*, 2009). De plus, la résolution spatiale de l'ancienne génération des capteurs destinés à l'étude des propriétés de l'eau (CZCS (*Coastal Zone Color Scanner*), SeaSAT, et SeaWIFS) ne permettait pas la détection des FEA dans les lacs de moyenne à petite taille. Or, depuis le nouveau millénaire, des données satellitaires à 250 et à 300 mètres de résolution spatiale sont disponibles respectivement via les capteurs MODIS (*MODerate Resolution Imaging Spectroradiometer*) et MERIS (*MEdium Resolution Imaging Spectrometer*) les qualifiant pour de telles applications. Certes, des images à plus haute résolution spatiale (Landsat, QuickBird ou IKONOS) sont disponibles offrant un détail spatial intéressant, mais ces dernières sont couteuses et leur temps de revisite est trop long (16 jours pour le Landsat, par exemple) pour le suivi de ce phénomène très dynamique dans le temps et l'espace.

1.3 Hypothèses et objectifs

La présente thèse propose une méthodologie pratique combinant télédétection optique, mesures de terrain et outils statistiques pour le suivi de la dynamique spatio-temporelle des FEA dans les eaux douces de moyenne à petite taille du Québec méridional (> 2.5 km²). Ainsi, l'objectif principal de cette thèse était de *développer une approche d'estimation de la concentration en ChI-a pour le suivi des FEA en exploitant les méthodes d'ensemble et des données MODIS désagrégées à 250 mètres de résolution spatiale.* Cet objectif a été scindé en trois objectifs spécifiques décrits dans les paragraphes suivants.

Les bandes du capteur MODIS destinées à l'étude des propriétés de l'eau et centrées sur les pics d'absorption et de réflectance de la Chl-a sont à 1 km de résolution spatiale. Cette résolution n'est pas adéquate pour suivre les FEA dans les lacs de moyenne à petite taille. Récemment, une méthode de désagrégation (downscaling) de la résolution spatiale des bandes 3 à 7 du capteur MODIS a été développée au Centre Canadien de

6

Télédétection (CCT (Trishchenko *et al.*, 2006)). Cette approche permet l'utilisation des sept premières bandes de ce capteur à 250 mètres de résolution spatiale. Ainsi, l'hypothèse de cette première partie de la thèse se formule comme suit :

 «Quoique non-destinées initialement à l'étude de la couleur de l'eau, les bandes MODIS 1 à 7 désagrégées à 250 mètres de résolution spatiale peuvent être utilisées pour le suivi des FEA dans les lacs de moyenne à petite taille, et ce en raison de leur résolution spatiale plus fine.».

En effet, l'utilisation structurée de ces bandes, couvrant la partie du bleu aux CLIR peut mener à une meilleure estimation de la concentration en Chl-a dans ces milieux aquatiques à signature spectrale complexe. Pour vérifier la validité de cette hypothèse, il a été question d'*étudier le potentiel des sept premières bandes du capteur MODIS, désagrégées à 250 mètres de résolution spatiale, à détecter les FEA, et ce en développant un modèle d'estimation de la concentration en Chl-a et en comparant sa performance à trois modèles très répondus dans la littérature (floating algae index (FAI), Kahru et Gitelson).*

La deuxième partie de la thèse se base sur le principe que chaque modèle possède des forces et des faiblesses (théorème du *no free lunch*) et que l'utilisation simultanée de plusieurs régions spectrales peut améliorer la performance de la modélisation. D'autre part, il a été démontré que le comportement spectral des eaux faiblement chargées en Chl-a (Moses *et al.*, 2009, Yacobi *et al.*, 2011) diffère de celui des eaux moyennent ou fortement chargées en Chl-a (Gómez *et al.*, 2011). De ces constats, deux hypothèses ont été formulées à la base des travaux de recherche de cette deuxième partie de thèse :

- «La combinaison des informations spectrales allant du visible aux CLIR pourrait faciliter la détection de la Chl-a et sa discrimination des autres éléments optiquement actifs présents dans les milieux lacustres, tels que la MSS et la MODC.»;
- 3. «Le développement d'un modèle spécifique à chaque niveau ou classe de concentrations en Chl-a peut améliorer la précision des estimations.».

Ainsi, l'objectif était de développer un modèle adaptatif (MA) de suivi des FEA composé d'un classificateur et de trois estimateurs spécifiques à trois niveaux ou classes de concentration en Chl-a (eaux faiblement, moyennement ou fortement chargées en Chl-a), étalonnés par régression multivariée.

Les algorithmes de classification tels qu'utilisés pour étalonner le *classificateur* du MA sont généralement connus pour leur instabilité et leur application spatiale restreinte (Li *et al.*, 2002). Une mauvaise sélection de l'*estimateur* final du MA, suite à une mauvaise classification, peut ainsi induire à des erreurs d'estimation de la concentration en Chl-a et mener à un effet d'escalier lors de la modélisation. D'un autre côté, vu la complexité du phénomène à modéliser, il est souvent recommandé de se baser sur plusieurs décisions avant de prendre une décision finale (Polikar, 2006). C'est en partant de ces deux concepts que l'hypothèse de la troisième partie de la thèse a été formulée :

4. «L'utilisation des méthodes statistiques avancées telle les méthodes d'ensemble à des fins de classification et de régression peut améliorer la précision des estimations en Chl-a et réduire l'effet en escalier causé par la conception générale du MA.».

Le troisième objectif des travaux de cette thèse était ainsi *d'optimiser la performance du MA via le développement d'un modèle basé sur les méthodes de l'ensemble (MBME) pour l'estimation de la concentration en ChI-a dans les eaux douces du Québec méridional*.

1.4 Structure de la thèse

Le reste de la synthèse est organisée comme suit : Un état de l'art détaillant les propriétés écologiques des cyanobactéries ainsi que l'ensemble des méthodes utilisées pour leur suivi (standards et par télédétection) ; une présentation des données (Terrain et satellitaires), méthodes de prétraitements et indices d'évaluation statistiques utilisés, qui seront suivi par une section synthétique de l'ensemble des travaux réalisés au cours de cette recherche doctorale. La synthèse se termine par une conclusion générale. Les

trois autres parties de la thèses présentent les articles publiés et en soumission avec un bref descriptif des tâches réalisées par les co-auteurs.

2 ÉTAT DE L'ART

2.1 Les fleurs d'eau d'algues et/ou de cyanobactéries

Une fleur d'au d'algues et/ou de cyanobactéries (FEA), appelée aussi prolifération, floraison, efflorescence ou *bloom* en anglais est le résultat d'une augmentation massive en biomasse phytoplanctonique lorsque les conditions environnementales lui est favorable. Elle est souvent dominée par une ou deux espèces et accompagnée d'une diminution des autres espèces phytoplanctoniques présentes dans le plan d'eau (AFSSA & AFSSET, 2006). Malgré leur taille micrométrique, les individus composant la FEA peuvent colorer les masses d'eau qui prendront généralement la couleur de l'espèce dominante dans ce système aquatique (AFSSA & AFSSET, 2006). À forte concentration, les espèces qui dominent la FEA sont souvent d'origine cyanobactérienne. Ainsi, dans la littérature, une FEA fait souvent référence aux fleurs d'eau d'algues bleu-vert, aux fleurs d'eau de cyanobactéries (FEC) ou tout simplement aux cyanobactéries. Dans la présente thèse, les termes FEA, FEC et cyanobactéries sont utilisés de façon interchangeable.

2.1.1 Écologie

L'appellation cyanobactérie vient du nom grec «*cyano*» qui fait référence à la couleur bleu-vert qui distingue la majorité des espèces cyanobactériennes. Autrefois, ces microorganismes étaient nommés algues bleues ou (bleu-vert) puisque, comme les algues, elles utilisent la photosynthèse pour leur développement. Or, avec l'avancement de la science, les chercheurs ont ensuite constaté que les cyanobactéries possédaient des caractéristiques plus proches des bactéries que celles des algues. C'est pour cette raison que l'appellation cyanobactéries leur a été attribuée faisant référence à des bactéries qui font de la photosynthèse (GRIL, 2007). Elles sont parmi les organismes photosynthétiques les plus vieux de l'histoire de la planète. Il y a environ 3.5 milliards d'années, grâce à leur système photosynthétique, d'énormes quantités d'oxygène ont été produites au point d'enrichir assez l'atmosphère pour permettre la respiration des animaux. Leur rôle dans les écosystèmes aquatiques est comparable à celui des

plantes terrestres du fait qu'elles utilisent la lumière comme source d'énergie pour la photosynthèse, mais la grande majorité des espèces cyanobactériennes produisent des composés malodorants et des toxines (Lavoie *et al.*, 2007).

L'écologie des cyanobactéries est assez bien connue. En milieu aquatique, elles sont planctoniques ou pélagiques lorsqu'elles prolifèrent en suspension dans la colonne d'eau, ou benthiques quand elles sont fixées sur les sédiments (Bouchard Valentine, 2004). Leur mode de reproduction se fait seulement de façon asexuée. La plupart d'entre elles sont des photo-autotrophes². Elles sont capables de se développer dans des milieux aérobiques³ ou anaérobiques⁴. Elles ont besoin d'eau, de dioxyde de carbone, de nutriments et de la lumière. Comme les autres microorganismes, les cyanobactéries sont à la base de la chaîne alimentaire des écosystèmes aquatiques et représentent donc une source d'aliments pour les maillons supérieurs du réseau trophique. Cependant, elles sont des proies difficiles à ingérer puisqu'elles sont souvent sous forme de colonies dans la nature, expliquant en partie une croissance rapide des populations si la pression de broutage est réduite. Lorsque les conditions leur sont favorables, elles acquièrent une grande capacité à se développer aux dépend des autres espèces phytoplanctoniques conduisant à la détérioration de l'écosystème qui les abrite (Lavoie *et al.*, 2007).

2.1.2 Toxines

Certaines espèces des cyanobactéries ont la capacité de synthétiser des toxines nommées *cyanotoxines*. Selon plusieurs scientifiques, l'occurrence des *cyanotoxines* dans les écosystèmes aquatiques est très ancienne (Carmichael, 1992, Carmichael, 1994, Chorus *et al.*, 1999, Fawell *et al.*, 1993, Svrcek *et al.*, 2004) et leur présence a été reportée pratiquement dans toutes les régions où les cyanobactéries ont été répertoriées. Les toxines cyanobactériennes sont généralement classées selon leur mode d'action : les hépatotoxines (foie), les neurotoxines (système nerveux) et les dermatoxines (peau). Ces molécules sont synthétisées et donc localisées à l'intérieur

² Photoautotrophes : capables de générer leur propre matière organique à partir de la lumière, au contraire des hétérotrophes.

³ Aérobique : présence d'oxygène dans un milieu aquatique donné.

⁴ Anaérobique : milieu dépourvu d'oxygène

des cellules pendant la phase de croissance des cyanobactéries. Durant cette phase, les toxines libérées dans l'eau représentent de 10% à 20% de la teneur totale en toxine (Jones *et al.*, 1994). Cependant, à leur mort naturelle ou lors de la lyse par des algicides, les toxines sont libérées de façon massive dans le milieu aquatique (AFSSA & AFSSET, 2006).

La liste des problèmes causés par les cyanobactéries et leurs toxines est longue. À ce jour, plusieurs cas de perturbations gastriques liées à la présence de ces microorganismes dans l'eau ont été rapportés. En 1979, dans un réservoir d'eau de Palm Island dans le Queensland (Australie), la propagation de Cylindrospermopsis raciborskii a causé une gastroentérite à 140 enfants (Byth, 1980). En 1996, suite à l'ingestion d'une eau contaminée en cyanobactéries et en microcystines, 76 personnes ont connu la mort par une hépatotoxicité aiguë au Brésil (Pouria et al., 1998). Le premier cas de décès lié aux activités récréatives a été rapporté aux États-Unis en 2002, alors que l'anatoxine-a a intoxiqué un jeune de 17 ans, conséquence d'une ingestion accidentelle d'eau contaminée lors d'une baignade dans un étang sur un terrain de golf (Stewart, 2004). Au Canada, bien qu'aucune déclaration de mortalité humaine n'ait été déclarée suite à une intoxication par les cyanotoxines, elles ont été la cause de nombreux problèmes de santé (maux de tête, nausée, diarrhée et douleur musculaire) chez 13 personnes suite à leur baignade dans un lac contaminé par Microcystis spp. et Anabaena circinalis (Dillenberg et al., 1960). Le cas le plus impressionnant d'intoxication est survenu en 1992 le long de la rivière Darling en Australie, où environ 10,000 bêtes ont connu la mort suite à une fleur d'eau massive de l'espèce neurotoxique Anabaena circinalis (Lavoie et al., 2007).

2.2 Prolifération des cyanobactéries

Il est très difficile de définir un seuil de biomasse caractérisant une prolifération. En effet, ce seuil devrait être défini en fonction du statut trophique de l'écosystème aquatique concerné. Des concentrations de 30 à 50 mg Chl-a m⁻³ dans les milieux oligo-mésotrophes représentent des valeurs très élevées de biomasse. En revanche, ces mêmes valeurs sont considérées comme faibles ou moyennes dans les milieux

eutrophes. Le MDDELCC, quant-à-lui, subdivise les FEA en 3 catégories : 1) lorsque la densité des cellules varient grandement dans la colonne d'eau, allant de quelques individus à plusieurs cellules clairsemées, on parle de catégorie 1; 2) lorsque la transparence de l'eau est réduite en raison d'une présence anormale de cellules et l'eau est anormalement brouillée et de couleur verte à bleu-vert, mais sans présence d'écume, on parle alors de catégorie 2a et 3) lorsque la colonne d'eau est couverte d'écumes, qui sont des dépôts colorés verdâtres où la densité des algues bleu-vert est très élevée, on parle de **catégorie 2b**. Généralement, ces écumes proviennent d'une prolifération massive de cellules qui s'accumulent sur les rives d'un plan d'eau (Blais, 2008). les proliférations d'alques Néanmoins. peuvent être catégorisées quantitativement par l'intermédiaire de certains seuils préétablis par l'ensemble des Ministères et des organismes de gestion des eaux (Figure 2-1).



Figure 2-1 Arbre décisionnel publié par l'Organisation mondiale de la santé (Chorus *et al.*,

1999).

Globalement, deux seuils (10 et 50 mg Chl-a m⁻³) sont utilisés mondialement pour évaluer les risques liés à l'utilisation d'une eau contaminée (Chorus *et al.*, 1999, Monestier, 2006). Plusieurs organismes sont en effet impliqués dans la gestion des eaux de consommation (MDDELCC, Santé canada, *United States environmental protection agency*), mais seule l'Organisation mondiale de la santé (OMS) a développé
des critères pour les eaux récréatives. Ces critères reposent sur les effets irritatifs causés par les cyanobactéries et du potentiel d'exposition liée à l'ingestion accidentelle de cyanotoxines, particulièrement des microcystines qui sont les toxines le plus souvent mesurées. Ainsi, l'OMS a présenté trois niveaux de risque : 1) mineur ou faible probabilité d'effets sur la santé lorsque la densité est inférieure à 20,000 cellules de cyanobactéries mL⁻¹, équivalente à 10 mg Chl-a m⁻³ avec dominance de cyanobactéries ; 2) modéré ou probabilité modérée d'effets sur la santé lorsque la densité se situe entre 20,000 et 100,000 cellules mL⁻¹ avec dominance de cyanobactéries et 3) élevé ou forte probabilité d'effets sur la santé lorsque la densité est supérieure à 100,000 cellules mL⁻¹, équivalente à 50 mg Chl-a m⁻³ avec dominance de cyanobactéries (Figure 2-1). Au Québec, le niveau d'alerte utilisé par l'Institut national de santé publique du Québec est de 100,000 cellules mL⁻¹ et de 16 µg L⁻¹ en microcystine-LR toxicité équivalente pour annoncer une alerte dans le cas des eaux récréatives (Boissonneault *et al.*, 2007).

2.3 Méthodes de suivi des proliférations d'algues : Approches standards

2.3.1 Caractéristiques physico-chimiques du plan d'eau

Certains indicateurs physiques sont en étroite relation avec le développement d'une biomasse algale ou cyanobactérienne. La présence d'une importante biomasse chlorophyllienne dans une masse d'eau réduit la **transparence de l'eau**, ce qui peut constituer un signal d'alerte de prolifération cyanobactérienne, spécialement dans les milieux aquatiques ayant déjà connus cette problématique et pour lesquels la couleur de l'eau n'est pas dominée par la matière organique dissoute (matière humique). Ainsi, un milieu devient **turbide** lorsqu'il s'enrichit en matière en suspension qui survient de la prolifération des cellules cyanobactériennes. Ces dernières ont la capacité de se déplacer verticalement dans la colonne d'eau, à une profondeur donnée, en fonction de l'intensité lumineuse afin d'échapper au fort ensoleillement pendant les périodes de canicule, par exemple, ou afin de profiter des nutriments à la surface des sédiments au fond du lac et de se remobiliser par la suite vers la surface du plan d'eau, lorsque les

conditions leurs sont favorables, pour la photosynthèse (Lavoie *et al.*, 2007). Ce phénomène augmente la turbidité du plan d'eau qui peut être un indicateur de présence de ces microorganismes (Kahru *et al.*, 2004).

Certains indicateurs chimiques sont aussi liés au comportement métabolique propre à l'activité des algues et des cyanobactéries. En effet, durant le jour, on assiste à une forte activité photosynthétique et pendant le soir c'est la respiration qui l'emporte. Ainsi, la forte activité métabolique lorsque la densité est associée à une grande production d'**oxygène** dû à la photosynthèse pendant la journée avant que cet oxygène ne soit consommé par la respiration des microorganismes, des végétaux et des animaux pendant la nuit. Par conséquent, l'amplitude des variations circadiennes⁵ de la concentration en oxygène peut donner une indication sur la présence et l'importance d'une prolifération. Par ailleurs, en raison de sa forte relation avec la photosynthèse et la respiration, une forte activité des cyanobactéries provoque aussi un changement du **pH** du milieu. L'amplitude de la variation circadienne du pH, qui dépend de l'activité photosynthétique, du pouvoir tampon de l'eau et de son pH naturel, peut donc être un indicateur de l'importance d'une prolifération (Figure 2-2, (AFSSA & AFSSET, 2006)).



Figure 2-2 Schéma de principe illustrant une variation jour-nuit de la saturation en oxygène et du pH en présence d'une prolifération phytoplanctonique (AFSSA & AFSSET, 2006).

⁵ Variation d'un phénomène durant 24 heures.

2.3.2 Suivi direct des populations phytoplanctoniques

Dénombrement par microscopie

La composition taxonomique et l'abondance spécifique des cyanobactéries et du phytoplancton sont souvent analysées par microscopie inverse (Desai *et al.*, 1978). Cette méthode standard d'identification des espèces du phytoplancton est basée sur les caractères morphologiques des cellules. Deux méthodes de dépistage sont souvent utilisées : 1) l'**analyse complète** qui permet l'identification, le dénombrement et la quantification des algues au genre et 2) l'**analyse en mode dépistage** qui peut être réalisée plus rapidement (environ 75 minutes par échantillon) et fournit des informations suffisamment détaillées pour évaluer le risque associé à la présence des cyanobactéries et prendre les décisions appropriées, en particulier dans le cas de l'approvisionnement en eau potable (CEAEQ, 2010).

Suivi de la chlorophylle-a

La concentration en Chl-a est très bien corrélée à la biomasse des populations photosynthétiques (AFSSA & AFSSET, 2006). Ce pigment, aussi présent chez les cyanobactéries, assure la photosynthèse de toutes les espèces photosynthétiques. Il est donc moins spécifique aux cyanobactéries. Toutefois, la mesure de la Chl-a sur le terrain demeure une solution relativement simple pour évaluer sa biomasse. Une faible concentration en Chl-a indiquera l'absence d'algues, et par conséquent des cyanobactéries, alors que les plus fortes concentrations ne permettront pas d'indiquer si cette biomasse est dominée par les cyanobactéries ou par d'autres espèces d'algues (Ellis, 2009). Toutefois, la probabilité que les fortes biomasses soient composées d'espèces cyanobactériennes est plus élevée (Figure 2-3). Il demeure ainsi plus intéressant d'utiliser une mesure spécifique aux cyanobactéries pour compléter l'information fournie par la mesure de la Chl-a, soit la phycocyanine (AFSSA & AFSSET, 2006).



Figure 2-3 Relation entre les densités en cyanobactéries (log nb cellules mL⁻¹) et les concentrations en chlorophylle-a (log mg Chl-a m⁻³) collectées au cours de la période (2000 à 2008) par le Ministère du Développement Durable, de l'Environnement et de la Lutte contre les Changements Climatiques.

Suivi de la phycocyanine

Étant donné qu'il est dificile de quantifier de façon précise les cyanobactéries via la Chl-a, le suivi de la phycocyanine devient une alternative intéressante. La phycocyanine est la phycobiliprotéine principale chez de nombreuses cyanobactéries, mais elle est aussi présente, comme pigment secondaire, chez certaines algues rouges (Datla, 2011). Des études comparatives ont démontré une bonne corrélation (R² = 0.73) entre la concentration en phycocyanine et le dénombrement des cyanobactéries par microscopie (Brient *et al.*, 2008). Cependant, le dosage de la phycocyanine en laboratoire s'avère plus complexe que celui de la Chl-a. L'identification des cellules cyanobactériennes et leur comptage basée sur le dénombrement microscopique ou le dosage de la phycocyanine nécessite généralement le recours à un personnel hautement qualifié et à un laboratoire spécialisé, ce qui s'avère coûteux et laborieux (de \$85 à \$280 CAN par échantillon (Ellis, 2009)). Comme alternative, l'utilisation de sondes capables de mesurer la phycocyanine est de plus en plus favorisée. Ces sondes sont pratiques et relativement moins chères (quelques milliers de dollars), mais une analyse de fiabilité doit être réalisée avant l'utilisation de cet outil (Ellis, 2009).

Analyse de la fluorescence pigmentaire

Grâce à la diversité de la composition pigmentaire existant chez les différents microorganismes photosynthétiques, il est possible de quantifier les différentes classes phytoplanctoniques présentes dans un plan d'eau donné par analyse de la fluorescence des cellules excitées par stimulation lumineuse. Ces méthodes, connues sous le nom de fluorescence *in vivo*, sont des approches prometteuses, relativement accessibles et déjà utilisées par certains intervenants dans la gestion de l'eau pour l'estimation *in situ* de la biomasse chlorophyllienne. Deux sortes de sondes submersibles basées sur ce principe sont actuellement disponibles sur le marché : 1) celles qui sont capables de distinguer les différentes classes de phytoplancton (chlorophycées, diatomées, cyanobactéries, cryptophycées) par la mesure de l'émission fluorescente à 685 nm après excitation par diverses longueurs d'onde (par exemple. le Fluoroprobe de BBE Moldaenke (Beutler *et al.*, 2002)) et 2) celles qui sont spécifiques aux cyanobactéries puisqu'elles reposent uniquement sur l'excitation de la phycocyanine à 590 nm (par exemple, la sonde EXO de YSI).

2.3.3 Surveillance par le réseau de volontaire

Depuis 2004, le Québec a adopté un système de surveillance des lacs connu sous le nom du réseau de surveillance volontaire des lacs. Ce dernier est le résultat d'un partenariat entre le MDDELCC, les associations de propriétaires riverains et les organisations participant à la protection et à la gestion des plans d'eau. Les différentes tâches attribuées à ce réseau se résument en deux activités principales :1) **le suivi de la qualité de l'eau** soit par mesure de la transparence ou par l'échantillonnage *in situ* pour mesurer en laboratoire le phosphore, le carbone organique dissous et la Chl-a et 2) **l'évaluation de la bande riveraine et de la zone littorale autour du lac** à l'aide des protocoles mis à la disposition des participants. De cette façon, les riverains sont capables de mieux gérer et participent activement à la protection de leurs lacs, et ce par : 1) la récolte des données afin d'établir le niveau trophique d'un grand nombre de lacs et suivre leur évolution dans le temps ; 2) le dépistage des lacs montrant des signes d'eutrophisation et de dégradation ; 3) éducation, sensibilisation, soutient et

information des associations des riverains et les autres participants et 4) dressage d'un portrait général de la situation des lacs de villégiature au Québec .

2.4 Méthodes de suivi des proliférations d'algues : Télédétection

Par définition la télédétection est la discipline qui :

« regroupe l'ensemble des connaissances et techniques utilisées pour déterminer des caractéristiques physiques et biologiques d'objets par mesures effectuées à distance sans contact matériel avec ceux-ci (COMITAS, 1988). ».

Ses domaines d'applications sont multiples et divers, allant de la géographie, géologie, hydrologie, foresterie et biologie, jusqu'aux études sociales et à la médecine. Comme les eaux couvrent approximativement 74% de la surface terrestre, plusieurs chercheurs se sont ainsi intéressés au développement de méthodes basées sur la télédétection pour étudier l'eau et ses composantes (Gregg *et al.*, 2004). Toutefois, avant de détailler les différentes méthodes et approches utilisées pour détecter et suivre les FEA, il est important d'introduire les principes fondamentaux de l'étude de l'eau par télédétection.

Le rayonnement total (L_t) enregistré par les capteurs aéroportés ou satellitaires est une fonction de l'énergie électromagnétique provenant de quatre sources identifiées sur la Figure 2-4 (Bukata *et al.*, 1995) :

$$L_{t} = L_{p} + L_{s} + L_{v} + L_{b}$$
 (2-1)

où

- ✓ L_p est la portion du rayonnement enregistré par l'instrument de télédétection résultant de l'éclairement, émis du ciel (E_{ciel}) et du soleil (E_{soleil}), qui est incapable d'atteindre la surface de l'eau ;
- ✓ L_s est le rayonnement émis du ciel et du soleil qui atteint l'interface air-eau (appelé en anglais *free-surface layer* ou *boundary layer*), mais la pénètre à peine par quelques millimètres avant qu'il soit totalement réfléchie par l'eau de surface. Cette énergie réfléchie contient de précieuses informations spectrales sur les caractéristiques de sub-surface du plan d'eau ;

- L_v est la portion du rayonnement émis du ciel et du soleil qui éventuellement pénètre l'interface air-eau, interagit avec les constituants organiques et/ou inorganiques du plan d'eau et quitte la colonne d'eau sans contact avec le fond (appelé rayonnement volumétrique de sub-surface). Ce rayonnement enregistre également de précieuses informations concernant les caractéristiques internes (inhèrentes) de la masse d'eau ;
- ✓ L_b est la portion du rayonnement émis du ciel et du soleil qui atteint le fond du plan d'eau. Cette portion d'énergie est intéressante pour des études des fonds de lacs (par exemple, la cartographie bathymétrique). Cependant, ce type de rayonnement rend difficile la caractérisation de la colonne d'eau parce qu'il est difficile de le séparer du L_v (Islam *et al.*, 2004).

Rayonnement total



Figure 2-4 Les plans d'eau reçoivent l'éclairement solaire (Esoleil) et du ciel (Eciel). Le rayonnement total retourné du plan d'eau vers le capteur est une fonction du rayonnement atmosphérique diffusé (Lp), du rayonnement de l'eau de surface (Ls), du rayonnement volumétrique de sub-surface (Lv) et du rayonnement de fond (Lb) du plan d'eau (Bukata *et al.*, 1995).

21

Tel que mentionné ci-haut, les eaux sont généralement groupées en deux classes : 1) les eaux de type cas-1 (*case-1 waters*) qui incluent les eaux en haute mer caractérisées par des propriétés optiques inhérentes simples et 2) les eaux de type cas-2 (*case-2 waters*) qui incluent les lacs (eaux douces) et les eaux côtières et estuariennes, caractérisées par des propriétés optiques inhérentes beaucoup plus complexes. Les eaux de type cas-2 sont souvent chargées en MSS, MODC et en biomasse phytoplanctonique, et elles peuvent être également colonisées par des plantes aquatiques. Ainsi, l'étude de la qualité de ce type d'eau peut être effectuée via le suivi de ces composantes.

En raison de l'activité bio-optique de la Chl-a chez le phytoplancton ou des autres pigments accessoires tels que la phycocyanine (Mishra *et al.*, 2009) et la phycoérythrine (Bresciani *et al.*, 2011) chez les cyanobactéries, il est possible de suivre l'évolution spatio-temporelle de ces microorganismes par des capteurs satellitaires capables de détecter les variations en couleur induites à la colonne d'eau dues à leur présence (Dall'Olmo *et al.*, 2005, Ha *et al.*, 2013, Hu *et al.*, 2004, Lyu *et al.*, 2013, Tarrant *et al.*, 2009, Yacobi *et al.*, 2011). Les capteurs satellitaires tels SeaWIFS, MERIS, MODIS, Landsat, IKONOS, ... etc. sont ainsi capables d'enregistrer les ondes émises de la surface de la Terre et de restituer des images reflétant la biomasse algale. Des informations qualitatives et quantitatives sur les formations végétales aquatiques ou terrestres peuvent être indiquées, suite à un traitement adapté aux images. Les méthodes de détection des FEA par satellite peuvent être divisées en deux grands secteurs : 1) la **cartographie qualitative** de l'extension des floraisons de surface et 2) l'**estimation quantitative** de la biomasse algale (Kuster, 2009).

2.4.1 Cartographie qualitative

Que ce soit en haute mer, en zones côtières ou dans les lacs, la cartographie qualitative des FEA repose sur le même principe, la photo-interprétation. Cette technique ne permet pas d'évaluer l'intensité ou l'importance du risque liés aux proliférations algales, mais elle aide à les localiser et à les suivre dans le temps et dans l'espace. Elle est considérée comme une étape essentielle avant de passer à l'estimation quantitative de la biomasse algale. Plusieurs capteurs ont été utilisés pour

cartographier l'extension des FEA, qu'ils soient aéroportés (Figure 2-5) ou satellitaires (Figure 2-6), via la constitution de composés en couleur vraie (rouge vert bleu ou RVB). Les capteurs radars ont aussi contribué à la cartographie qualitative des FEA, quoique les ondes radar ne peuvent pas pénétrer la surface de l'eau, la température élevée de la surface de l'eau peut être considérée comme un indicateur de présence de fortes concentrations algales (Kahru *et al.*, 1993).



Figure 2-5 Fleur d'eau d'algue qui a été détectée au lac Érié en Ohio, 2011⁶.

Figure 2-6 Floraison détectée par le capteur MERIS en mer de Barents (Norvège), 2011⁷

La cartographie qualitative peut aussi se réaliser en se basant sur des algorithmes de classification non supervisée, c'est-à-dire par classification de l'image sans reconnaissance *a priori*. En effet, ces modèles de classification ont la capacité de regrouper un ensemble de pixels ayant des caractéristiques spectrales similaires, via des algorithmes tel *le maximum de vraisemblance,* ou encore en se basant sur le principe inverse regroupant les pixels selon des critères de dissimilarité. Cette pratique est toutefois moins utilisée par rapport au composé RVB, mais son avantage est qu'elle permet une meilleure discrimination des pixels de l'eau de ceux des FEA, particulièrement en haute mer. Toutefois, cet avantage est moins prépondérant pour les zones côtières et les lacs en raison de la complexité de la signature spectrale de ces

⁶http://photography.nationalgeographic.com/photography/photos/best-pod-october-2013/#/algae-bloom-lake-erieessick_72092_600x450.jpg.

⁷ http://image-cnes.fr/1-loeil-du-satellite/le-plancton-vegetal-sepanouit-en-mer-de-barents/

systèmes aquatiques. Une sur- ou sous-classification des pixels des FEA peut résulter de la présence simultanée d'autres composantes (MSS et MODC) optiquement actives dans ces plans d'eau.

D'autres méthodes de détection des FEA par classification sont aussi répertoriées dans la littérature. Ces méthodes se basent sur des algorithmes de classification plus poussés, qui nécessitent une reconnaissance d'un historique de données avec des classes préassignées menant au développement de classificateurs, qui par la suite sont généralisés sur l'ensemble des images. L'application de cette technique est spécifique aux eaux douces en raison de leur diversité en composantes optiquement actives (Chl-a, MODC, et TSS). Les réseaux de neurones sur les données Landsat ont par exemple été utilisés sur des lacs au Japon afin de discriminer entre les différentes composantes optiquement actives (Baruah et al., 2002). Un modèle basé sur les arbres décisionnels en utilisant les données du Compact Airborne Spectrographic Imager-2 a également été développé pour l'étude de la variation saisonnière de la Chl-a et de la phycocyanine à des fins préventives aux risques liés à la présence des cyanobactéries dans les lacs (Hunter et al., 2009). D'autre part, la technique de la machine à vecteurs de support, connue sous l'acronyme SVM (Support vector machine), a été appliquée aux données du capteur MERIS pour étudier la qualité de l'eau de l'estuaire Pearl River à Hong Kong (Yuanzhi Zhang et al., 2009b). Finalement, des travaux plus récents ont montré qu'il était possible de différencier avec succès entre plusieurs niveaux de concentration en Chl-a d'un ensemble de lacs américains en utilisant un algorithme d'arbres décisionnels sur les données Landsat et IKONOS (Mackie, 2010).

2.4.2 Estimation quantitative

Télé-estimation de la Chl-a par approches empiriques et semi-empiriques

Afin de télé-estimer la concentration en Chl-a par des capteurs satellitaires, les scientifiques cherchent à relier de façon empirique ou semi-empirique les propriétés inhérentes d'un plan l'eau donné (concentration *in situ* de la Chl-a ou autres pigments) à ses propriétés optiques apparentes (l'énergie réfléchie par les eaux et captée par un senseur, spécialement la portion L_v (Figure 2-4)) afin de développer des modèles

bio-optiques capables de modéliser la biomasse des FEA. Gordon *et al.* (1983) ont proposé la première fonction (équation (2-2)) d'estimation de la Chl-a en haute mer :

$$L_{t} = x \left[\frac{L(\lambda_{1})}{L(\lambda_{2})} \right]^{\nu}$$
(2-2)

où $L(\lambda_1)$ et $L(\lambda_2)$ sont les signaux de retour à des longueurs d'onde spécifiques enregistrées par le système de télédétection, et *x* et *y* sont des constantes dérivées empiriquement. L'emploi de ce principe simple dans les algorithmes standards d'estimation de la Chl-a, utilisant des bandes spectrales localisées généralement dans le bleu et le vert, a connu un grand succès en haute mer (Jensen, 2006). Toutefois, le transfert des réalisations de ces algorithmes vers les environnements continentaux n'a pas été tâche facile. En effet, l'utilisation de la bande bleue doit se faire avec précaution dans les eaux de type cas-2, étant donné sa corrélation avec la MODC qui peut induire une sur- ou sous-estimation de la concentration en Chl-a lorsqu'elle est assez concentrée pour interférer et qu'elle varie en concentration, ou encore dû à la correction atmosphérique, toujours problématique dans la partie bleue du spectre et où beaucoup d'erreurs subsistent (Dekker *et al.*, 1991, Gons, 1999, Gons *et al.*, 2008, Gower, 1980, Kuster, 2009, Mittenzwey *et al.*, 1991, Neville *et al.*, 1977).

Une autre génération d'algorithmes a aussi été répertoriée dans la littérature, se basant davantage sur la partie rouge et PIR du spectre électromagnétique (Moses *et al.*, 2009, Reinart *et al.*, 2006). L'avantage d'utiliser cette partie du spectre est que : 1) l'effet du bruit apporté par les particules atmosphériques est moins important ; 2) l'effet du signal de retour des eaux et du fond est faible puisque l'eau a une grande capacité à absorber ces longueurs d'onde et 3) la séparation de la Chl-a des autres matières organiques et inorganiques est possible grâce à la forte réflectance de ce pigment dans le PIR. La liste des capteurs utilisés est par contre moins longue et se limite généralement aux capteurs à moyenne (MERIS (Matthews *et al.*, 2010) et MODIS (Moses *et al.*, 2009)), haute (Landsat (Vincent *et al.*, 2004)) et très haute résolution spatiale (QuickBird (Wheeler *et al.*, 2012) et IKONOS (Ormeci *et al.*, 2009)).

Par ailleurs, certains algorithmes se sont basés sur la quantification du pic de réflectance à 685 nm causé par la fluorescence de la Chl-a induite par l'énergie solaire

(Gower, 1980, Gower *et al.*, 1990, Neville *et al.*, 1977). En effet, lorsque le phytoplancton absorbe la lumière, une partie de l'énergie absorbée est utilisée pour la photosynthèse, une autre partie est dissipée en chaleur, et le reste est réémis comme un faible liseré rouge fluorescent. Ce signal est observable soit activement par modification de l'éclairement incident avec une source de lumière artificielle, ou encore passivement par mesure de la quantité de la lumière fluorescée sous des conditions de luminance naturelle (Butler, 1978, Govindje, 1995, Joliot *et al.*, 2005). Cette relation, fluorescence/énergie-solaire, a rendue possible l'utilisation de la magnitude du pic à 685 nm au-dessus d'une ligne de base à partir de 650 à 730 nm (Figure 2-7), connue en anglais sous l'acronyme FLH (*fluorescence line height*), pour l'estimation de la concentration en Chl-a (Gower, 1980). La FLH est généralement utilisée dans les eaux littorales et en haute mer (Gower *et al.*, 2007, Ryan *et al.*, 2009), mais récemment elle a aussi été utilisée pour la détection des FEA dans les eaux internes (Binding *et al.*, 2011).



Figure 2-7 Hauteur de la ligne de base de fluorescence (FLH), inspirée de Gilerson *et al.* (2007).

Plusieurs algorithmes de télé-estimation de la concentration en Chl-a dans les lacs se sont basés sur le pic de réflectance aux alentours de 700 nm. étant donné le minimum d'absorption à cette longueur d'onde par les pigments phytoplanctoniques et l'eau (Gitelson, 1992, Han, 1997, Yacobi et al., 2011), et le pic de réflectance à 675 nm fortement lié au maximum d'absorption de la Chl-a dans le rouge. D'autres algorithmes ont exploré le rapport de réflectances aux pics près de 700 et 560 nm. Ce dernier est minimum d'absorption causé par la combinaison des relié au pigments phytoplanctoniques, des particules en suspension et de la MODC. De fortes relations ont été trouvées, avec des coefficients de corrélation variant de 0.91 à 0.99, entre $\frac{R(\lambda_{700})}{R(\lambda_{770})}$ et la concentration en Chl-a dans les rivières de Don et Donec en Russie, la mer Azov et le lac Balaton en Hongrie (Gitelson et al., 1986). L'utilisation de ce rapport ne s'était pas limitée à ces deux pics de réflectance, les rapports $\frac{R(\lambda_{705})}{R(\lambda_{675})}$ et $\frac{R(\lambda_{705})}{R(\lambda_{670})}$ ont été aussi répertoriés dans la littérature (Gitelson et al., 1985, Mittenzwey et al., 1991).

La seule approche de télé-estimation de la concentration en Chl-a dans les eaux douces qui se base sur la quantification du contraste entre le maximum de réflectance des eaux chargées en Chl-a dans le PIR et leur maximum d'absorption dans le rouge est celle de Kahru *et al.* (2004). Dans un premier temps, un composé en couleur vrai est utilisé pour localiser la distribution des fleurs d'eau. Ensuite, comme mesure semi-quantitative de la quantité de particules dans l'eau près de la surface, l'indice de turbidité est calculé en soustrayant la réflectance du PIR du rouge. Même si la turbidité n'est pas spécifique à la prolifération d'algues, il s'agit d'une estimation quantitative de l'intensité de la turbidité créer par la FEA, une fois que son existence est détectée par les images en couleurs vraies. Cette approche a été appliquée avec succès pour détecter une FEA dans la baie Paracas au Pérou en utilisant les données MODIS centrées sur les régions spectrales à 645 et 859 nm (Kahru *et al.*, 2004).

Plusieurs modèles d'estimation de la concentration en Chl-a sont basés sur la relation fondamentale entre la réflectance de télédétection (R) et les propriétés optiques inhérentes de l'eau par l'intermédiaire de l'équation 3.

$$R(\lambda) = \frac{f(\lambda)}{Q(\lambda)} \frac{b_{b}(\lambda)}{a_{tot}(\lambda) + b_{b}(\lambda)}$$
(2-3)

où $a(\lambda)$ est le coefficient d'absorption, $b_b(\lambda)$ le coefficient de rétrodiffusion, $f(\lambda)$ décris la sensibilité de la réflectance aux variations de l'angle zénithale solaire (Morel et al., 1991) et $Q(\lambda)$ exprime les propriétés bidirectionnelles de la réflectance (Morel *et al.*, 1993).

Ainsi, Gons (1999) avait reformulé le rapport $\frac{R(\lambda_{705})}{R(\lambda_{675})}$ en termes de $a(\lambda)$ et $b_b(\lambda)$ et avait partitionné $a_{tot}(\lambda)$ en quatre coefficients d'absorption : 1) par le phytoplancton $(a_{\varphi}(\lambda))$; 2) par les particules non algales $(a_{PNA}(\lambda))$; 3) par la MODC $(a_{MODC}(\lambda))$ et 4) par les molécules d'eau $(a_{eau}(\lambda))$ afin de développer un modèle d'estimation de la concentration en Chl-a pour les eaux internes et les eaux côtières (Gons, 1999, Gons *et al.*, 2000) sous la forme :

$$[Chl - a] = \frac{\frac{R(\lambda_2)}{R(\lambda_1)}(a_{eau}(\lambda_2) + b_b) - (a_{eau}(\lambda_1) + b_b^p)}{a_{\varphi}^*(\lambda_1)}$$
(2-4)

où R(λ_1) est la réflectance à 672 nm, une longueur d'onde proche du maximum d'absorption de la Chl-a dans le rouge, R(λ_2) est la réflectance à 704 nm, $a_{eau}(\lambda_1)$ et $a_{eau}(\lambda_2)$ sont les coefficients d'absorption de l'eau à 672 et à 704 nm, b_b est le coefficient de rétrodiffusion, $a_{\phi}^*(\lambda_1)$ est le coefficient d'absorption spécifique à la Chl-a à 672 nm et *p* est un coefficient introduit afin d'améliorer l'ajustement du modèle.

D'autre part, Dall'Olmo *et al.* (2003) ont appliqué un modèle conceptuel aux eaux turbides développé originalement pour l'estimation de la concentration en Chl-a dans la végétation terrestre :

$$[Chl - a] \propto [R^{-1}(\lambda_1) - R^{-1}(\lambda_2)] \times R(\lambda_3)$$
(2-5)

où le terme $R^{-1}(\lambda_1)$ doit être fortement sensible à l'absorption de la Chl-a (dans la gamme de 660 à 690 nm (Dall'Olmo *et al.*, 2005)). Toutefois, $R^{-1}(\lambda_1)$ est toujours affectée par l'absorption de la MSS, MODC et la rétrodiffusion de l'eau. Cet effet

(dénominateur de l'équation (2-3)) peut être réduit en utilisant une deuxième bande, où $R^{-1}(\lambda_2)$ est le moins sensible à l'absorption de la Chl-a et la plus proche de λ_1 ; il a été démontré que λ_2 devrait appartenir à la gamme de 710 à 730 nm (Dall'Olmo *et al.*, 2005). La différence ente $R^{-1}(\lambda_1)$ et $R^{-1}(\lambda_2)$ élimine par conséquent l'effet de la MSS et de la MODC, mais la Chl-a est toujours affectée par la rétrodiffusion de l'eau. Une troisième bande est par conséquent utilisée afin de réduire cet effet. Cette bande doit être le moins affectée par la MSS, et la MODC, et absorbe fortement les molécules de l'eau. Le PIR du spectre électromagnétique répond parfaitement à cette condition. Ainsi, l'équation (2-5) dépend uniquement de l'activité bio-optique de la Chl-a.

Télé-estimation de Chl-a par indices de végétation

Les indices de végétation figurent aussi parmi les méthodes d'estimation de la concentration en Chl-a pour le suivi des FEA. Le concept des indices de végétation les plus populaires tels le NDVI (*Normalized difference vegetation index*, (Kiage *et al.*, 2009)) et le EVI (*Enhanced vegetation index*, (Jiang *et al.*, 2007)) a été utilisé sur les données MODIS (Hu *et al.*, 2008). Toutefois, le succès de tels indices en milieu marin n'a pas été aussi probant qu'en milieu terrestre. En fait, la délimitation des FEA et leur séparation des pixels de l'eau a été possible, mais l'application de ces indices était sensible aux conditions environnementales variables et aux conditions d'observation. Ces dernières créent certains problèmes de visualisation et de quantification puisque non seulement ils affectent le contraste visuel entre les FEA et l'eau, mais ils changent aussi leurs valeurs absolues (Hu, 2009).

Récemment, un nouvel indice de végétation au nom du FAI (*floating algae index*, (Hu, 2009)) a été développé spécifiquement pour les algues flottantes et a montré un grand potentiel à détecter les FEA que ce soit en haute mer (Hu, 2009) ou en milieux lacustres (Hu *et al.*, 2010). Cet indice utilise trois régions spectrales allant du rouge au CLIR. Cette dernière est utilisée en raison de la forte absorption de l'eau dans cette région spectrale. Ainsi, l'eau apparait opaque (noir) même dans les environnements les plus turbides. Les algues flottantes ont une plus forte réflectance dans le PIR que dans d'autres longueurs d'ondes, et par conséquent elles peuvent être facilement distinguées de l'eau pure (Hu, 2009). Basé sur ce principe, le FAI a surmonté les problèmes des

autres indices de végétation (NDVI et EVI) lors de son application pour la détection des FEA. Le FAI est défini comme suit :

$$FAI = R_{PIR} - R'_{PIR}$$
(2-6)

où :

$$R'_{PIR} = R_R - [R_{CLIR} - R_R] \times \left[\frac{\lambda_{PIR} - \lambda_R}{\lambda_{CLIR} - \lambda_R}\right]$$
(2-7)

où R'_{PIR} est la réflectance de référence dans la bande du PIR, dérivée à partir de l'interpolation entre le rouge et le CLIR. Pour MODIS, $\lambda_R = 645$ nm, $\lambda_{PIR} = 859$ nm et $\lambda_{CLIR} = 1,240$ nm. Le FAI prend par conséquent la forme :

$$FAI = R(\lambda_{b2}) - [(R(\lambda_{b5}) - R(\lambda_{b1}) + R(\lambda_{b1})] \times 0.359$$
(2-8)

Estimation quantitative de la biomasse via les pigments accessoires

Certains groupes phytoplanctoniques contiennent des pigments accessoires qui leurs sont spécifiques. La détection de ces pigments par télédétection s'avère donc possible. Puisque les cyanobactéries contiennent de la phycocyanine et/ou de la phycoérythrine, il est possible de les distinguer des autres regroupements phytoplanctoniques par la détection de ces deux pigments. Certains algorithmes appliqués en eaux douces ont utilisé avec succès les rapports de réflectance $\frac{R(\lambda_{650})}{R(\lambda_{620})}$ (Dekker, 1993, Dekker *et al.*, 1992, Simis *et al.*, 2005, Simis *et al.*, 2007) et $\frac{R(\lambda_{650})}{R(\lambda_{620})}$ pour estimer la concentration en phycocyanine au lac Carter aux États-Unis (Schalles *et al.*, 2000). Bresciani *et al.* (2011) ont proposé un rapport de bande qui explore la partie du 560 à 620 nm du spectre pour détecter et estimer la biomasse des cyanobactéries en utilisant la phycoérythrine avec les données MERIS (Bresciani *et al.*, 2011). Cependant, bien que ces deux pigments peuvent être utilisés pour distinguer les cyanobactéries des autres espèces phytoplanctoniques, ils ne sont détectables qu'à fortes concentrations (Reinart *et al.*, 2006), limitant ainsi leur utilisation pour la cartographie quantitative des cyanobactéries via la télédétection.

Pour la présente thèse, malgré l'intérêt porté au suivi des cyanobactéries et leur floraison, c'est la concentration en Chl-a qui a été utilisée pour la calibration de l'ensemble des modèles. Deux points majeurs ont conduit à ce choix : 1) en raison de la faible taille de la base de donnée de la phycocyanine (N = 74) fournit par le MDDELCC et du grand nombre de journées nuageuses par année sur le territoire du Québec méridional, il a malheureusement été impossible de bâtir une base de données exploitable avec ces mesures et 2) la détection de la phycocyanine est généralement possible en utilisant l'information spectrale de deux à trois bandes localisées dans la partie du 600 à 650 nm du spectre (Ogashawara et al., 2013). Or, seule la bande 1 des sept premières du capteur MODIS est localisée dans cette partie du spectre (620 nm). L'utilisation unique de cette bande, également sensible à la présence des MSS et MODC, rendrait la télé-estimation de la phycocyanine difficile, voire impossible. Ainsi nous avons misé sur le fait que la Chl-a est très corrélée avec la présence des cyanobactéries spécialement en cas de fortes concentrations (Figure 2-3), spécialement dans un ensemble de lacs où la problématique des fleurs d'eau de cyanobactéries a déjà été identifiée, d'où l'adoption de la nomination fleur d'eau d'algues et/ou de cyanobactéries le long des travaux de cette thèse.

3 MÉTHODOLOGIE GÉNÉRALE

3.1 Zone d'étude

Située entre les latitudes 44° à 50° Nord, et les longitudes 67° à 80° Ouest, la zone d'étude est composée de 22 plans d'eau douce faisant partie du Québec méridional (Figure 3-1) dont les noms et les surfaces sont résumées dans le Tableau 3-1. Ces plans d'eau ont été suivis par le MDDELCC en raison de la récurrence des fleurs d'eau observée d'une année à l'autre suite aux réclamations des riverains ou du réseau de surveillance volontaire des lacs. Ces plans d'eau représentent différents statuts trophiques allant de l'oligotrophie à l'hyper-eutrophie. Ils sont localisés sur plusieurs régions administratives et appartient à des bassins versants possédant différentes caractéristiques d'occupations du sol allant de milieux forestiers à des milieux à forte occupation urbaine ou agricole.



Figure 3-1 Localisation des lacs sur la zone d'étude (les lacs marqués par leurs noms, ont servi à la calibration des approches développées).

Tableau 3-1Noms, surfaces, et dénombrement minimal et maximal de la densité en
cyanobactéries entre l'année 2007 et 2010 de l'ensemble des plans d'eau ayant
servi à la validation des approches développées.

Nom des lacs	Surface (Km ²)	Dénombrement min/max annuel de la densité cyanobactérienne (cellules mL ⁻¹)				
		2007	2008	2009	2010	
Adélard	5	******	******	> 20,000	> 20,000	
Aylmer	32		1-2,000,000	> 20,000	> 20,000	
Missisquoi Bay	47	20,000-500,000	1-2,000,000	> 20,000	> 20,000	
Bouchette	7	******	*******	> 20,000	> 20,000	
Brome	15	1-2,000,000	50,000-500,000	> 20,000		
Choinière	5	20,000-100,000	1-2,000,000	> 20,000	> 20,000	
Des-Commissaires	28	500,000-2,000,000	*******	*******	> 20,000	
Etchemin	2	******	10,000-500,000	> 20,000	********	
Fréchette	5	******	******	*******	> 20,000	
Labrecque	4	******	1,000-2,000,000	*******	> 20,000	
Lovering	5	2,000-2,000,000	******	> 20,000	********	
Mandeville	2	100,000-500,000	100,000-2,000,000	> 20,000	> 20,000	
Maskinongé	10	1-2,000,000	1-100,000	> 20,000	> 20,000	
Massawippi	19	1-2,000,000	1,000-100,000	> 20,000	********	
Nairne	2	10,000-100,000	20,000-2,000,000	> 20,000	********	
Ouareau	13	******	2,000-2,000,000	> 20,000	> 20,000	
Perchaude	2	******	20,000-2,000,000	********	*******	
Pohénégamook	9	10,000-2,000,000	500,000-2,000,000	> 20,000	> 20,000	
Roxton	2	2,000-2,000,000	2,000-2,000,000	> 20,000	> 20,000	
Taureau	111	2,000-2,000,000	1-2,000,000	> 20,000	*******	
Tortue	3	******	100,000-500,000	> 20,000	*******	
William	5	2,000-2,000,000	2,000-500,000	*******	> 20,000	

3.2 Données in situ

Les données de calibration ont été collectées sur une période de neuf ans (2000-2008) dans le cadre d'un projet réalisé par les différentes directions du MDDELCC afin d'identifier les principales causes des fleurs d'eau de cyanobactéries ayant eu lieu dans les lacs au Québec, particulièrement en 2007. Ces données incluent des variables : de toxicité (microcystine LR total équivalent, anatoxine-a intracellulaire, anatoxine-a libre et anatoxine-a totale), photosynthétiques (phycocyanine, Chl-a, phéophytine-a et Chl-a totale), et physico-chimiques (phosphore total, azote total, ratio N/P molaire, pH, turbidité, profondeur de Secchi et température de l'eau de surface). Seules les concentrations en Chl-a ont été utilisées pour la calibration et la validation des modèles étudiés au cours de la présente thèse. La base de données fournit par le MDDELCC comportait des mesures collectées sur sept lacs (la baie Missisquoi du lac Champlain et les lacs Brome, Forgeron, Gauvreau, Gros-Ruisseau, Nairne et William) totalisant ainsi

363 échantillons. Cependant, en raison de la présence des nuages sur certains sites d'échantillonnage, ou à cause de la mauvaise qualité de certaines images satellitaires (présence d'artefacts), seuls les données des lacs Brome, Nairne et William et la baie Missisquoi du lac Champlain (Figure 3-2) ont pu être utilisées pour bâtir la base de calibration, réduisant ainsi la taille de la base de calibration à 46 échantillons.



Figure 3-2 Disposition des sites échantillonnés par le MDDELCC sur les quatre lacs qui ont servi à la calibration.

Concernant le prélèvement et la conservation des échantillons, c'est le Centre d'expertise en analyse environnementale du Québec (CEAEQ) qui est chargé de ce processus et procède de la façon suivante. Un volume d'environ 225 ml d'échantillon est impérativement prélevé dans un contenant de 250 ml en polypropylène opaque. L'échantillon est conservé au froid (près de 4 °C) et est acheminé au laboratoire le plus tôt possible. L'extraction est effectuée dans un délai n'excédant pas 72 heures après le prélèvement et elle est effectuée immédiatement après la filtration. L'extrait dans l'acétone est conservé au congélateur et à l'obscurité avant de procéder au dosage. Dans cette méthode, le filtre utilisé se dissout instantanément dans l'acétone.

Le dosage de la Chl-a est effectuée par la suite au laboratoire du CEAEQ. La méthode employée est le dosage par fluorométrie, élaborée originalement par Yentsch et al. (1963). Son avantage est d'être très sensible et ne nécessite qu'un faible volume d'échantillon. Dans un premier temps, l'échantillon est filtré et les algues sont recueillis sur un filtre et extraits dans l'acétone à 90%. La concentration de Chl-a est déterminée en mesurant la fluorescence émise à une longueur d'onde supérieure à 665 nm, à la suite d'une excitation à une longueur d'onde de 340-500 nm. La méthode apporte une correction pour la présence de phéophytine-a, forme de dégradation des pigments chlorophylliens (chlorophylles a, b, c1, c2 et d), laquelle absorbe à une longueur d'onde voisine de la Chl-a. La mesure fluorométrique est effectuée avant acidification (mesure de la Chl-a) et après acidification qui transforme toute Chl-a en phéophytine-a. Cette opération permet de corriger l'interférence de la phéophytine-a sur la Chl-a et, en même temps, de doser la phéophytine-a. L'ensemble de la procédure (extraction et dosage) est répliquée (10 essais avec des volumes de 100 ml de culture d'algues) afin de calculer la précision du dosage. Ainsi, la réplicabilité de la méthode pour un échantillon de densité cellulaire de 80,000 cell ml⁻¹ est de 4.4 μ g/l ± 0.19 (coefficient de variation (CV) = 5.4%). La réplicabilité de la méthode pour un échantillon de densité cellulaire de 500,000 cell ml⁻¹ est de 27 μ g/l ± 1.4 (CV = 6.4%).

D'autre part et en raison des chevauchements des spectres d'absorption ainsi que d'émission de fluorescence par les différentes formes de chlorophylles et de leurs intermédiaires de dégradation, une sur- ou sous-estimation de la Chl-a peut subsister pendant les analyses. Lors d'une étude réalisée au laboratoire du CEAEQ et utilisant des ajouts dosés pour déterminer l'interférence des autres pigments sur le dosage de la Chl-a dans 101 échantillons d'eau de surface au cours de l'été 1999, il a été démontré que le pourcentage de sous estimation médian est de 0.3% pour un CV de 18.7%. Il est également intéressant de noter que l'interférence présente des variations saisonnières, étant en sur-estimation de 7% en mai et en sous-estimation de 5% en juillet et août (CEAEQ, 2012a).

Une deuxième base de données ordinales constituée de trois classes de densité en cyanobactéries : 1) dépassant 100,000 cellules mL⁻¹ (équivalent à 50 mg Chl-a m⁻³) ; 2) entre 100,000 et 20,000 cellules mL⁻¹ et 3) en bas de 20,000 cellules mL⁻¹ (équivalent

à 10 mg Chl-a m⁻³), a servi à valider les résultats des modèles calibrés. Ces mesures ont été collectées sur l'ensemble des plans d'eau illustrés à la Figure 3-1 et au Tableau 3-1. Pour les mêmes raisons citées ci-haut, seuls 103 des 677 échantillons collectés par le MDDELCC ont été utilisés.

La procédure employée dans le dosage des cyanobactéries de cette deuxième base de données est le **dépistage** qui consiste à identifier les trois ou quatre genres dominants de cyanobactéries (selon la densité cellulaire) et à leur attribuer une classe d'abondance en cell ml⁻¹. De façon occasionnelle, jusqu'à 6 genres peuvent être rapportés au certificat d'analyse lorsque l'échantillon contient une communauté fortement diversifiée dans laquelle aucun genre n'est vraiment dominant. L'analyse en dépistage exclut l'identification et le dénombrement des picocyanobactéries (diamètre inférieur à 2 μ m), lesquelles sont peu problématiques en termes de production de toxines. Des classes d'abondance sont attribuées pour le total des cyanobactéries à potentiel toxique, pour le total des cyanobactéries sans potentiel toxique et pour le total des cyanobactéries tous genres confondus.

Les échantillons d'eau destinés au dépistage des cyanobactéries sont préservés directement sur le terrain à l'aide d'une solution de lugol à 1% V/V. Le lugol préserve relativement bien les fines structures des cyanobactéries et il améliore leur sédimentation dans la chambre d'Utermöhl. Par contre, il colore les cellules, ce qui peut nuire à l'identification. La concentration de lugol ne doit pas dépasser 1% V/V pour éviter une déformation des cellules et une coloration trop intense. Il est également primordial de s'assurer que les bouchons des bouteilles soient hermétiques aux gaz et que les échantillons sont à l'abri de la lumière. Le temps de contact de l'échantillon avec le lugol doit être d'au moins 24 heures avant l'analyse pour favoriser une bonne sédimentation des cyanobactéries. Les échantillons lugolés sont conservés à la température ambiante, à l'obscurité et dans un endroit bien aéré s'ils sont analysés dans les 30 jours suivant leur prélèvement. Si les échantillons doivent être conservés sur une plus longue période avant d'être analysés ou s'ils sont conservés à des fins de connaissance, ils devraient être gardés au froid (près de 4 ºC) et à l'obscurité. Finalement, les bouteilles doivent être remplies à environ 80% pour que le jeu d'air permette un brassage adéquat de l'échantillon.

Comme toute procédure de dosage, le dépistage des cyanobactéries est associé à une incertitude de calcul. Les principaux facteurs liés à cette l'incertitude sont : la représentativité de l'échantillon prélevé, l'erreur liée au sous-échantillonnage effectué au laboratoire, la procédure de comptage, l'estimation du nombre de cellules par colonie ou par filament, la conversion du nombre de colonies ou de filaments en densité cellulaire et le nombre total de cellules comptées. Toutefois, cette incertitude est quantifiable via l'équation (3-1) qui permet de calculer la précision du dénombrement pour un seuil de confiance de 95% (CEAEQ, 2012b):

% précision =
$$\frac{2}{\sqrt{nonbre \ de \ cellules \ comptées}} \times 100$$
 (3-1)

3.3 Données MODIS

En raison de ses résolutions (spatiale, spectrale et temporelle), MODIS figure parmi les capteurs les plus utilisés pour une multitude d'applications en télédétection terrestre, atmosphérique ou marine. Le capteur MODIS embarqué à bord de la plate-forme *TERRA* du programme *the Earth observation system* de la NASA a été lancé en 2000. Ses données sont acquises avec une fréquence quotidienne et sont disponibles via leur serveur ftp⁸, en format HDF, 2 heures après leur acquisition. Le niveau 1B du capteur MODIS contient un ensemble de données calibrées et géo-localisées. Le capteur couvre la région spectrale de 0,4 à 14,4 microns sur 36 bandes, avec des résolutions spatiales variant de 250 mètres à 1 km. L'heure du passage du satellite sur la zone d'étude varie entre 2:00 et 4:30 GMT (équivalent en heure locale) sur une orbite héliosynchrone.

Un ensemble d'images du capteur MODIS, couvrant la période de 2000 à 2010, ont été téléchargées et prétraitées à des fins de calibration et de validation. En raison de leur résolution spatiale, seules les sept premières bandes du capteur ont été prises en considération dans ce travail (Tableau 3-2). Afin d'avoir une résolution spatiale

⁸ http://ladsweb.nascom.nasa.gov/data/search.html

commune entre l'ensemble des bandes utilisées, les bandes 3 à 7 ont été désagrégées à 250 mètres de résolution spatiale par l'intermédiaire d'une approche développée au sein du Centre canadien de télédétection (CCT, (Trishchenko *et al.*, 2006)).

Utilisation primaire	Bande	Longueur d'onde (nm)	Résolution spatiale (m)	
Limitas : Tarra/Nuagas/Aárasala	1	620–670	250	
Limites . Terre/Nuages/Aerosois	2	841–876	250	
	3	459–479	500	
Propriété :	4	545–565	500	
	5	1,230–1,250	500	
Terre/Nuages/Aerosois	6	1,628–1,652	500	
	7 2,105–2,155	500		
	8	405-420	1000	
	9	438-448	1000	
	10	483-493	1000	
Coulour d'agu/phytoplanetons	11	526-536	1000	
/biochimio	12	546-556	1000	
/biocilime	13	662-672	1000	
	14	673-683	1000	
	15	743-753	1000	
	16	862-877	1000	

 Tableau 3-2
 Caractéristiques des 16 premières bandes du capteur MODIS⁹.

3.4 Prétraitements des images

Les bandes du capteur MODIS destinées à l'étude des propriétés de l'eau sont à 1 km de résolution spatiale (bandes 8 à 16 du Tableau 3-2). Une telle résolution limite leur utilisation à détecter les FEA aux grands lacs et en haute mer. Or, l'objectif de cette thèse était de développer une approche capable d'étudier l'évolution des floraisons dans les lacs de moyenne à petite taille. Certes, les deux premières bandes (rouge au PIR) du capteur MODIS sont à 250 mètres de résolution spatiale mais vu la difficulté de modéliser les FEA dans ces systèmes aquatiques à signature spectrale complexe, il était nécessaire d'utiliser plusieurs régions spectrales pour atteindre une modélisation optimale tout en ayant une résolution spatiale la plus fine possible.

⁹ http://modis.gsfc.nasa.gov/about/specifications.php

3.4.1 Désagrégation des bandes 3 à 7

L'équipe du CCT a développé une méthode de désagrégation des bandes 3 à 7 du capteur MODIS permettant de passer de 500 à 250 mètres de résolution spatiale (Trishchenko et al., 2006). En effet, un certain degré de corrélation existe parmi les différentes bandes spectrales pour les classes similaires d'occupation du sol ; c'est sur cette propriété que se base la classification d'occupation du sol issue des données de télédétection. D'un autre côté, il a été démontré qu'il existe peu de formes spectrales de l'albédo de base (Bowker et al., 1985). Rappelons que l'albédo est une mesure de la fraction du rayonnement incident qui est réfléchie par une surface, alors que la réflectance est une mesure de la même fraction pour un angle d'incidence unique. Ainsi, l'équipe du CCT a pu trouver un certain nombres de corrélations (positives et négatives, linéaires et non-linéaires) entre l'albédo des bandes 3, 4, 5, 6 et 7, et l'albédo des bandes 1 et 2 et le NDVI. De plus, les auteurs ont constaté que les bandes ayant une bonne corrélation avec la bande 1 sont moins corrélées avec la bande 2 et vice-versa. Ainsi, les fonctions des deux bandes (1 et 2) peuvent être considérées comme orthogonales et contribuent par conséquent aux comportements basiques dans la reconstruction des autres bandes. En tenant compte de ces propos, il était alors possible de bâtir une fonction générique (équation (3-2)) entre les bandes 3 à 7 avec les bandes 1 et 2 et le NDVI, en utilisant les étapes suivantes :

- ✓ Développement d'un modèle de régression adaptatif basé sur les données du produit MOD02HKM qui contiennent l'information spectrale des sept premières bandes à 500 mètres de résolution spatiale. Les paramètres de régression sont obtenus pour différents types de scènes génériques tels la végétation, les terres désertes, la neige, l'eau et les nuages ;
- ✓ Utilisation du modèle de régression développé, en prenant comme données d'entrées le NDVI et la réflectance des bandes 1 et 2 (à 250 mètres de résolution spatiale), afin de générer cinq néo-bandes à 250 mètres de résolution spatiale correspondante aux bandes 3 à 7 ;
- ✓ Normalisation des néo-images aux valeurs numériques des bandes à 500 mètres afin de préserver les propriétés radiométriques des bandes originales.

Le modèle de régression non-linéaire suivant a été appliqué aux bandes 3 à 7 pour générer les cinq néo-bandes :

$$R(\lambda_{i}) = a_{0,i} + (a_{1,i} \times R(\lambda_{1}) + a_{2,i} \times R(\lambda_{2}))(1 + a_{3,i} \times NDVI + a_{4,i} \times NDVI^{2})$$
(3-2)

où : NDVI = $\frac{(R(\lambda_2)-R(\lambda_1))}{(R(\lambda_2)+R(\lambda_1))}$ et $R(\lambda_i)$ est la réflectance observée pour la bande i =1 : 7. Les détails d'implémentation telle la paramétrisation du modèle adaptatif et la normalisation des bandes sont présentées dans le papier de Trishchenko *et al.* (2006). Cette approche a été évaluée en utilisant les données à haute résolution spatiale du capteur Landsat. Les résultats ont démontré que, généralement, les coefficients de corrélation des bandes désagrégées sont meilleurs que ceux des bandes originales (amélioration qui variait entre 0.02 à 0.07). Cela démontre que l'approche de désagrégation n'altère pas les propriétés radiométriques des bandes tout en améliorant la résolution spatiale.

3.4.2 Re-projection

Plusieurs produits prétraités du capteur MODIS sont disponibles sur leur site. Ces produits sont généralement divisés en tuiles dans le système sinusoïdal (SIN). Ce système dessine les lignes de latitudes comme des lignes droites horizontalement parallèles, avec des longueurs égales à leurs longueurs sur le globe, et les places de façon systématique et équitable par rapport au méridiens centraux (vertical et horizontal). Ceci crée des distorsions géométriques (de superficie, de forme ou d'orientation) pour les régions éloignées des méridiens centraux, spécifiquement à partir des latitudes moyennes (Bottomley, 2003). Or, contrairement à la projection SIN, la projection conique conforme de Lambert (LCC) permet de représenter de façon précise, sur le plan de l'orientation et de la forme, les régions des latitudes moyennes. L'échelle est exacte le long des parallèles de références (49° N et au 77° N) uniquement¹⁰, dont fait partie la région du Québec.

¹⁰ http://www.statcan.gc.ca/pub/92-195-x/2011001/other-autre/mapproj-projcarte/m-c-fra.htm

De ces faits, il a été nécessaire de re-projeter les images MODIS de leur système de projection initial (SIN) au système LCC. La Figure 3-3 montre un composé en couleur vrai d'une image prise par le capteur MODIS (19-09-2001) projetée dans les deux projections (SIN et LCC). Par comparaison, il est clair que l'image de la Figure 3-3.B, projetée dans le système LCC, est moins affectée par les effets de distorsions notamment ceux de la forme et de l'orientation. Afin de mieux percevoir l'impact de ces distorsions géométriques sur la qualité de l'image, un zoom sur la région du lac Saint-Jean a été effectué sur les deux images (Figure 3-3.C et D), mais cette fois avec la même résolution spatiale (250 mètres). Il est clair que l'image-D montre une meilleure qualité spatiale que l'image-C (lacs entourés par des cercles rouges).



Figure 3-3 A) et B) image MODIS projetée respectivement dans le système sinusoïdal (500 mètres) et dans le système conique conforme de Lambert (250 mètres); C) et D) Zoom sur la région du lac Saint-Jean (250 mètres).

3.4.3 Correction atmosphérique

Pour plusieurs applications, spécialement pour des analyses multi-temporelles, les valeurs des pixels brutes ou les nombres digitaux des images de chaque bande doivent être corrigés pour les effets atmosphériques avant le traitement des images (Moran et al., 1997). Une mauvaise correction atmosphérique pourrait mener à des erreurs significatives dans le calcul de la réflectance et affecter la précision des estimations (Rahman, 2001). En raison de sa simplicité et de son habilité à corriger les deux effets atmosphériques (additif et multiplicatif), les données de MODIS ont été corrigées atmosphériquement au moyen de l'algorithme de SMAC (Simplified Model for Atmospheric Correction). L'avantage de ce modèle de correction atmosphérique et sa position entre les modèles déterministes plus sophistiqués tels le 6S (Second Simulation of Satellite Signal in the Solar Spectrum (Vermote et al., 1997)), le MODTRAN (MODerate resolution atmospheric TRANsmittance and radiance code (Berk et al., 1999)) et le ATCOR (ATmospheric CORrection (Richter, 1996)) et les modèles à base de l'image (appelés en anglais *image-based*) y compris le DOS (*Dark* Object Subtraction (Chavez, 1988)) et le COST (COSine Transmission for atmospheric correction (Chavez, 1996)). En effet, à cause de leur hypothèse supposant que l'influence des aérosols dans les régions marines au PIR est nulle (Gordon, 1978, Gordon et al., 1994, Viollier et al., 1980), les modèles déterministes présentent beaucoup de problèmes à estimer la réflectance en présence d'aérosols absorbants (Bailey et al., 2006, Gordon, 1997) ou en cas d'eaux turbides (Dekker et al., 1997, Lavender et al., 2005, Morel et al., 2006), menant à des erreurs considérables. D'un autre côté, les modèles à base de l'image corrigent seulement l'effet additif causé par la brume. Certes, l'influence atmosphérique est réduite, mais la réflectance calculée est toujours affectée par l'effet multiplicatif causé par les aérosols, la vapeur de l'eau et l'ozone. Par contre, le modèle SMAC utilise un ensemble d'équations simples qui nécessitent seulement sept variables : la réflectance à la limite de l'atmosphère, l'angle zénithal solaire, l'angle de vue zénithale et l'angle relative azimutale pour la correction de l'effet additif et l'épaisseur optique des aérosols à 550 nm, la teneur de la vapeur d'eau atmosphérique et la teneur en ozone atmosphérique pour la correction de l'effet multiplicatif. Ces variables d'entrées sont dérivées directement du produit MOD08 du

niveau 3 du capteur MODIS. L'ensemble des prétraitements (désagrégation, re-projection et correction atmosphérique par le modèle SMAC) ont été réalisés en utilisant un outil gratuit disponible sur le serveur ftp¹¹.

3.4.4 Masque de l'eau

Parmi les étapes importantes de cette thèse est la création du masque d'eau (c-à-d. détermination de la limite entre la terre (sol, végétation, terrain agricol, ...etc.) et l'eau). En effet, pour bâtir une bonne base de calibration pour le suivi des FEA, le signal capté des sept bandes utilisées doit être une mesure pure de d'eau. Ainsi, l'objectif était d'identifier les pixels de l'eau pure et de les différencier des pixels de la terre et des pixels mixtes. Ces derniers ont tendance à hausser la réflectance influencée par le signal de la végétation plus que celui de l'eau. Au fait, la différenciation entre les pixels de la terre des pixels de l'eau n'est pas vraiment une problématique puisque le signal de l'eau et de la végétation qui l'entoure forme généralement un grand contraste spectrale dans le PIR permettant leur séparation facilement. Basé sur ce principe, la NASA avait mise au point depuis 2009 un masque d'eau (MOD44W005) à 250 mètres de résolution spatiale (Figure 3-4.B) disponible gratuitement sur leur serveur ftp¹². Le masque standard de la NASA est capable de discriminer la terre de l'eau, mais il présente certains problèmes spécifiquement en cas des pixels mixtes qui les considèrent comme pixel d'eau pure (Figure 3-4.A). Ceci présentait un handicap pour développer une base de calibration exploitable.

¹¹ ftp.ccrs.nrcan.gc.ca/ad/CCRS_CANADA/Software/

¹² http://e4ftl01.cr.usgs.gov/MOLT



Figure 3-4 Différents masques eaux, couvrant la baie Missisquoi du lac Champlain, produits par trois différentes méthodes : A) image originale du capteur MODIS à 250 mètres de résolution spatiale (projection SIN); B) masque eau produit par la NASA; C) et E) composé en couleur vrai des bandes désagrégées à 250 mètresde résolution spatiale (projection LCC); D) masque eau fournit par le CCT et F) masque généré par l'INRS.

La Figure 3-4.D montre un produit de classification de l'occupation du sol qui est généré en même temps que les images désagrégées par l'outil développé au sein du CCT. Les pixels noirs représentent la terre, les blancs la brume et les gris l'eau. L'algorithme de l'occupation du sol (terre, eau, neige, brume, nuage et ombre de nuage) généré par cette approche est détaillé dans Luo *et al.* (2008). Par comparaison au composé en couleur vrai sur la Figure 3-4.C, il est évident que l'algorithme de classification de l'occupation du sol du CCT ne distingue pas les pixels d'eau en présence de fortes floraisons d'algues et aussi classifie les pixels mixtes comme eau pure.

Comme les deux masques d'eau présentaient des faiblesses de discrimination entre les pixels de la Terre et de l'eau pure, il a été nécessaire de générer un masque d'eau spécifique à la présente étude. L'utilisation de la bande du PIR pourrait être la solution, mais la génération d'un masque basé sur une seule bande serait toujours entaché d'erreurs, spécifiquement en cas de présence de fortes floraisons algales et des plantes aquatiques. Ainsi, l'idée était de créer une image composite formée de valeurs minimales de la bande du PIR d'un ensemble de plus de 1600 images. De cette façon, la réflectance des pixels d'eau et de la végétation temporaire (floraisons d'algues et plantes aquatiques) présentes dans les plans d'eau serait faible, alors que, la réflectance des pixels de végétation permanente aux alentours des lacs serait élevée. Les images utilisées pour la génération de l'image composite ont été choisies de telle sorte qu'elles présentent une forte abondance végétative (images prises entre mai et octobre).

La Figure 3-5 montre l'histogramme de la réflectance de l'image composite, où les faibles valeurs correspondent aux pixels de l'eau pure (ellipse rouge), les fortes valeurs font référence aux pixels de la terre (ellipse bleue), alors que la zone intermédiaire entourée par l'ellipse brune représente les pixels mixtes. Il est clair que les réflectances en bas de 10% correspondent aux pixels de l'eau pure. De ces constats, il a ainsi été possible de générer un masque d'eau en utilisant 10% de réflectance comme seuil de discrimination. La Figure 3-4.F montre le résultat du nouveau masque qui couvre parfaitement les pixels d'eau pure de la baie Missisquoi, tout en évitant de couvrir les pixels mixtes.



Figure 3-5 Histogramme de la réflectance de l'image composite.

3.4.5 Masque de nuages

Pour la présente étude, il n'a pas été nécessaire de créer un masque de nuages. L'objectif de ce travail était de développer et d'évaluer la performance d'un modèle d'estimation de la concentration en Chl-a dans les eaux douces. Dans cette perspective, seules des images sans couverture nuageuse ont ainsi été utilisées pour construire la base de calibration et de validation. Également, lors de l'application du modèle sur les images pour la validation avec les données ordinales, encore une fois, seules les images sans nuage ont été considérées. En effet, le masque de nuages disponibles via le site web de la NASA est à 1 km de résolution spatiale. L'utilisation de ce produit masquera une grande majorité des eaux douces de moyenne à petite taille. Alors, il aurait fallu soit développer une méthode de détection des nuages à partir des bandes désagrégées, soit développer une méthode de désagrégation de la résolution spatiale des bandes de MODIS conçues initialement à détecter les nuages et de les utiliser ensuite pour produire un masque de nuages à 250 mètres de résolution spatiale. Le développement de telles approches colossales, laborieuses et complexes et qui sortes du contexte de l'objectif initial de cette thèse, a par conséquent été jugé inutile.

3.4.6 Indices d'évaluation statistiques

L'ensemble des modèles ont été validés par la technique de la *validation-croisée*. Cette technique consiste à enlever temporairement une mesure de la concentration en Chl-a de la base de données et d'utiliser le reste de la base comme données de calibration, pour ensuite estimer la concentration de la mesure enlevée. Cette opération se répète pour l'ensemble de la base de données pour aboutir à une estimation de toutes les mesures de concentration en Chl-a permettant une comparaison entre valeurs estimées/mesurées via des indices d'évaluation. Pour la présente thèse, quatre indices d'évaluation statistique (le coefficient de détermination (R^2), le critère de Nash relatif (NASHr), l'erreur quadratique moyenne relative (EQMr) et le Biais relatif (BIAISr)) ont été utilisés dont les formules sont représentées par les équations (3-3) à (3-6). Le critère du NASH évalue la performance des modèles en comparant les valeurs estimées avec la moyenne de celles mesurées. Pour un NASH négatif, il serait préférable d'utiliser la moyenne des valeurs mesurées que celles estimées par le modèle, ce qui est très peu performant. Le modèle commence à être satisfaisant à partir d'un NASH \geq à 0.8 ; le modèle est parfait pour un NASH=1 (Nash *et al.*, 1970).

$$R^{2} = \left[\frac{\sum_{i=1}^{n} (M_{i} - \overline{M})(Es - \overline{Es})}{\sqrt{\sum_{i=1}^{n} (M_{i} - \overline{M})^{2}} \sqrt{\sum_{i=1}^{n} (Es_{i} - \overline{Es})^{2}}}\right]^{2}$$
(3-3)

$$BIAISr = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{Es_i - M_i}{M_i} \right)$$
(3-4)

EQMr =
$$\sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(\frac{Es_i - M_i}{M_i}\right)^2}$$
 (3-5)

NASHr =
$$1 - \frac{\sum_{i=1}^{n} \left(\frac{M_{i} - Es_{i}}{M_{i}}\right)^{2}}{\sum_{i=1}^{n} \left(\frac{M_{i} - \overline{M}}{\overline{M}}\right)^{2}}$$
 (3-6)

où : n est la taille de l'échantillon, M et Es sont les valeurs mesurées et estimées, et $\overline{\text{M}}$ et $\overline{\text{Es}}$ sont les moyennes des valeurs mesurées et estimées.

Les indices utilisés sont sous forme relative dans le but de réduire le contraste entre les valeurs mesurées et celles estimées tant pour les fortes valeurs que les plus faibles. Les formes absolues des indices quantifient les différences absolues entre les valeurs observées et estimées. Une sur- ou sous-estimation des fortes valeurs a une influence plus grande que celles des plus faibles. En utilisant les formes relatives des indices, on réduit l'influence de la différence absolue calculée durant les périodes de fortes biomasses. En contrepartie, la différence absolue calculée durant les périodes de faibles de faibles biomasses sont pareillement haussées (Krause *et al.*, 2005).

Tableau 3-3Diagramme simplifié de la matrice de confusion et des paramètres impliqués dans
son calcul ainsi que dans le calcul de l'indice Kappa, où a et d sont les nombres
des estimations bien-classées, b et c sont les nombres des estimations
mal-classées, n_1 , n_2 , n_3 et n_4 sont respectivement les totaux a+b, c+d, a+c et b+d,
N est la taille de la base de données, P_0 et P_c sont tels que définis dans les
équations (3-7) et (3-8).

			Me	surées		
		VRAI	FAUX	Total	Taux de succès (%)	Erreur de commission (%)
	VRAI	а	b	n ₁	$\frac{\mathbf{a}}{\mathbf{n}_1}$	$\frac{\mathbf{b}}{\mathbf{n_1}}$
G	FAUX	С	d	n ₂	$\frac{\mathbf{d}}{\mathbf{n}_2}$	$\frac{c}{n_2}$
Jée	Total	n ₃	n ₄	Ν	_	
Estin	Taux de succès (%)	$\frac{a}{n_3}$	$\frac{d}{n_4}$			
	Erreur d'omission (%)	C n ₃	$\frac{\mathbf{b}}{\mathbf{n_4}}$			
	Succès global (%)					P ₀
	Indice Kappa					$\frac{P_0-P_c}{1-P_c}$
$=\frac{a}{a}$	+ d					()

$$P_{c} = \frac{n_{1} \times n_{3} + n_{2} \times n_{4}}{N^{2}}$$
(3-8)

La performance des modèles a aussi été évaluée avec des données ordinales qui ont servi aux calculs des matrices de confusion. Le principe de la matrice de confusion consiste à comparer les classes mesurées par rapport aux classes estimées et de

quantifier la précision de l'approche utilisée via le calcul des taux de succès, des erreurs (commission et omission) et d'indices de précision, notamment l'indice *Kappa*. Ce dernier permet de quantifier le niveau de concordance de la classification des concentrations en Chl-a mesurées avec celles estimées. Pour un *Kappa* négatif, la concordance est exécrable ; la concordance est bonne pour un *Kappa* supérieur à 0.6 et elle est excellente à partir de 0.8. Le Tableau 3-3 résume tous les paramètres impliqués dans le calcul de la matrice de confusion et de l'indice *Kappa (Cohen, 1960)*.

Un dernier indice statistique a aussi été utilisé dans la présente thèse, ayant uniquement servi à l'évalution du modèle basé sur les méthodes de l'ensemble (MBME). Tel que décrit à la section 3.4, l'estimation de la concentration en Chl-a via le MBME est une mesure de la moyenne pondérée d'un ensemble d'*estimateurs*. Ceci permet également de mesurer la variance d'estimation de la concentration en Chl-a de cet ensemble d'*estimateurs*. À partir de ces moments statistiques (moyenne et variance), il est alors possible de calculer le coefficient de variation (CV) qui est une mesure d'homogénéité (équation (3-9)). En d'autres termes, il permet de quantifier la dispersion des individus d'une population par rapport à leur moyenne. Ainsi, pour un CV inférieur à 16%, la moyenne est considérée comme fiable ; pour un CV supérieur à 33%, la dispersion est importante et la moyenne n'est plus fiable. Le calcul du CV serait d'une importance considérable en matière de prise de décision puisqu'il permet de quantifier la fiabilité de l'estimation de la concentration en Chl-a de cuantifier la fiabilité de l'estimation de la concentration puisqu'il permet de quantifier la fiabilité de l'estimation de la concentration en Chl-a de chaque pixel traité dans l'image.

$$CV(\%) = \frac{\sqrt{\sigma}}{\mu} \times 100 \tag{3-9}$$

où σ est la varaiance et μ est la moyenne.
4 ÉTUDE DU POTENTIEL DES 7 PREMIÈRES BANDES DU CAPTEUR MODIS DÉSAGRÉGÉES À 250 M DE RÉSOLUTION SPATIALE À SUIVRE LES FEA DANS LES EAUX DOUCES

4.1 Introduction

En raison de la résolution spatiale généralement supérieure à 1 km de l'ancienne génération des capteurs satellitaires, le suivi des FEA par télédétection a pour longtemps été limité aux océans (O'Reilly et al., 1998). Or, à partir de l'année 2000, des données à meilleure résolution spatiale, telles que produites par les capteurs MODIS et MERIS sont gratuitement disponibles, et ont été largement utilisées avec succès pour suivre les FEA dans les eaux douces (Moses et al., 2009). Le capteur MERIS possède cinq bandes centrées sur les régions d'absorption et de réflectance de la Chl-a¹³ à 300 mètres de résolution spatiale, soient les bandes 2, 3, 5, 7 et 8. Avec une telle résolution, il est possible d'utiliser ses données pour suivre les FEA dans les lacs de moyenne à petite taille (Wheeler et al., 2012). Les bandes équivalentes de MODIS (8 à 16 du Tableau 3-2) sont pour leur part à 1 km de résolution spatiale, limitant leur application aux plans d'eau de plus grande taille tels que l'océan et les grands lacs (Becker et al., 2009). Or, en avril 2012, la liaison avec le satellite ENVISAT a été brusquement interrompue et l'Agence spatiale européenne a déclaré la fin de la mission de cette plate forme porteuse du capteur MERIS un mois plus tard (9 mai 2012). D'autre part, dans un but d'implémentation opérationnelle de l'imagerie satellitaire pour un suivi quotidien des FEA, l'utilisation des images MODIS serait plus avantageuse pour les gestionnaires de l'eau en raison de sa fréquence temporelle (1 journée versus 3 jours pour MERIS).

Récemment, il a été démontré que les bandes du capteur MODIS à résolution spatiale plus fine (500 et 250 mètres), quoiqu'originalement développées pour d'autres applications (bandes 1 à 7 du Tableau 3-2), ont la capacité d'étudier les propriétés de l'eau (Zhiqiang. Chen *et* al., 2007). Avec l'approche de désagrégation (*downscaling*) de

¹³ http://www.crisp.nus.edu.sg/~research/tutorial/meris.htm.

la résolution spatiale des bandes 500 à 250 mètres développée au CCT, il devenait alors possible d'exploiter les sept premières bandes de MODIS avec une meilleure résolution spatiale, couvrant du visible aux CLIR du spectre électromagnétique. Rappelons que ces régions spectrales sont sensibles à la fois à l'activité bio-optique de la Chl-a et à la présence de la MODC et de la MSS. Par conséquent, l'exploitation de l'information spectrale issue de ces bandes désagrégées serait très utile pour discriminer la Chl-a des autres éléments optiquement actifs.

Le premier objectif de cette partie de la thèse était de tester le potentiel des sept premières bandes du capteur MODIS désagrégées à 250 mètres de résolution spatiale à suivre les FEA dans les eaux douces de moyenne à petite taille du Québec méridional. Pour ce faire, il a été question de développer un modèle d'estimation de la concentration en Chl-a nommé APPEL (*APProach by ELimination*) et de tester sa performance par rapport aux modèles les plus prometteurs reportés dans la littérature (inspiré de Gitelson (Gitelson *et al.*, 2009), Kahru (Kahru *et al.*, 2004) et FAI (Hu, 2009)). Il est à préciser qu'en raison de l'absence de bandes couvrant certaines régions spectrales utilisées dans la version originale du modèle Gitelson parmi les bandes désagrégées, ces dernières ont été substituées par une combinaison de bandes 1, 2 et 4 d'où l'appellation «modèle inspiré de Gitelson». Les raisons du choix de ces bandes sont détaillées dans le premier article (El-Alem *et al.*, 2012) joint à cette thèse.

4.2 Approche méthodologique

Les quatre modèles (APPEL, Kahru, FAI et inspiré de Gitelson) étudiés ont fait l'objet d'une analyse comparative en utilisant les images désagrégées du capteur MODIS. Dans un premier temps, les modèles ont été adaptés aux conditions environnementales des milieux lacustres du Québec via leur calibration par les données *in situ* de la concentration en Chl-a collectées sur quatre lacs différents (baie Missisquoi du lac Champlain, et lacs Brome, Nairne et William). La calibration a été effectuée par <u>régression simple</u> entre les indices spectraux calculés à partir des modèles et les mesures *in situ* de la concentration en Chl-a. Une fois calibrés, le potentiel des bandes désagrégées du capteur MODIS à détecter les FEA a été ensuite testé via l'évaluation

52

de la performance des modèles par l'intermédiaire des indices d'évaluation statistiques prédécrit ci-haut (R², NASHr, BIAISr et EQMr), sur la base des résultats de la technique de la *validation-croisée*. Comme les principes fondamentaux des modèles Kahru, Gitelson et FAI sont détaillés dans la section 2.4.2, seul le principe du modèle APPEL, développé au cours des travaux de la présente thèse, sera détaillé dans la section ci-dessous.

La réflectance spectrale captée des eaux douces varie en fonction de quatre éléments (la Chl-a du phytoplancton et/ou des plantes aquatiques, la MSS, la MODC et la b_b de l'eau), et peut être exprimée par l'équation (4-1). Afin que cette réflectance soit uniquement une fonction de la Chl-a, l'effet des autres éléments doit être réduit au maximum. Pour ce faire, trois étapes sont nécessaires :

- ✓ Réduire l'effet de la MODC en utilisant la bande-3 de MODIS pour sa forte sensibilité à la présence de ces matières (Luciani *et al.*, 2007). Cependant, la réflectance de la bande-3 est aussi influencée par l'information spectrale de la Chl-a et de la b_b de l'eau. Afin que la réflectance soit uniquement en fonction de la MODC, l'effet de ces deux éléments (Chl-a et b_b) doit également être réduit. En raison de la forte réflectance de la Chl-a dans le PIR ainsi que la forte absorption de l'eau dans cette partie du spectre, l'effet de ces deux éléments peut être réduit de la bande-3 en utilisant la bande-2 du capteur MODIS. Ceci est possible en soustrayant la bande-2 de la bande-3 ce qui réduira l'effet de la Chl-a, et en multipliant le résultat de la soustraction par la même bande qui permettra de réduire la b_b de l'eau. Cette bande a aussi été utilisée par Dall'Olmo *et al.* (2005) pour le même objectif. Ainsi l'équation (4-1) prend la forme de l'équation (4-2) ;
- Réduire l'effet de la MSS en utilisant la bande-1 pour sa grande corrélation avec ces particules (Hu *et al.*, 2004). Pareillement, il a été nécessaire de réduire l'effet de la Chl-a de la bande-1 afin qu'elle soit uniquement une fonction de la MSS. Étant donné que b_b de l'eau est déjà faible dans la bande-1 en raison de la forte absorption de l'eau dans le rouge, il n'a pas été nécessaire de la réduire. La

soustraction de la bande-2 de la bande-1 a permis de réduire l'effet de la Chl-a de cette bande. Ainsi, l'équation (4-2) prend la forme de l'équation (4-3) ;

Réduire l'effet de la b_b de l'eau par le choix de la région spectrale utilisée pour le développement du modèle APPEL. La bande utilisée doit en effet être fortement sensible à la réflectance de la Chl-a et moins sensible à la présence de la MSS, MODC et la b_b de l'eau. Le PIR répond parfaitement à ces conditions (forte réflectance de la Chl-a versus forte absorption de l'eau), d'où l'utilisation de la bande-2 a été favorisée. Ainsi APPEL prend la forme finale décrite par l'équation (4-4).

$$R(\lambda_i) \alpha [Chl - a] + [MODC] + [MSS] + b_b$$
(4-1)

où λ_i est la réflectance corrigée des bandes MODIS (i = 1 à 7).

$$R(\lambda_i) \alpha [Chl - a] + [(R(\lambda_{b3}) - R(\lambda_{b2})) \times R(\lambda_{b2}) + [MSS]]$$
(4-2)

$$R(\lambda_i) \alpha \left[Chl - a \right] + \left[\left(R(\lambda_{b3}) - R(\lambda_{b2}) \right) \times R(\lambda_{b2}) + \left(R(\lambda_{b1}) - R(\lambda_{b2}) \right) \right]$$
(4-3)

$$APPEL = R(\lambda_{b2}) - [(R(\lambda_{b3}) - R(\lambda_{b2})) \times R(\lambda_{b2}) + (R(\lambda_{b1}) - R(\lambda_{b2}))]$$
(4-4)

4.3 Résultats et discussion

Les premiers résultats obtenus, par *validation-croisée*, ont démontré la capacité des bandes désagrégées du capteur MODIS à estimer la concentration en Chl-a, tel qu'illustré à la Figure 4-1. Les valeurs du NASHr pour APPEL, FAI et Kahru ont été assez similaires (NASHr jusqu'à 0.96) avec un léger avantage pour les modèles APPEL et Kahru. Le NASHr pour le modèle inspiré de Gitelson, quant-à-lui, était beaucoup plus faible (NASHr = 0.28). Les résultats de l'EQMr ont montré que les estimations de la concentration en Chl-a calculées par le modèle APPEL étaient les meilleures (EQMr d'APPEL < Kahru < FAI < Gitelson). Cependant, l'erreur des estimations demeure forte pour l'ensemble des modèles (EQMr du meilleur modèle était de 69%). De plus, les BIAISr négatifs de l'ensemble des modèles indiquent que ces derniers ont tendance à sous-estimer la concentration en Chl-a. Les faibles performances enregistrées pour le

modèle inspiré de Gitelson peuvent s'expliquer par les substitutions apportées aux bandes initialement utilisées par le calcul de néo-bandes. Certes, ces dernières réduisent l'effet de la MODC et de la MSS mais aussi interagissent avec la Chl-a en raison de l'utilisation de la bande-4 (sensible à la Chl-a) dans le calcul des néo-bandes.



Figure 4-1 Concentration en chlorophylle-a estimée à partir des quatre modèles comparée aux mesures *in situ* pour les données supérieures à 10 mg Chl-a m⁻³, avec les indices de performance des modèles.

Dans un contexte d'intervention rapide à des fins de gestion du risque lié à la présence d'une FEA et de protection de la santé publique, les eaux déjà contaminées par une FEA sont moins intéressantes à surveiller que celles relativement saines, mais qui présentent un risque de développement de floraisons. Ainsi, il a été nécessaire

d'évaluer la performance des modèles pour ces conditions de floraison (concentrations en bas de 50 mg Chl-a m⁻³ ; seuil établit par l'OMS pour déclarer une FEA (Figure 2-1)). Les résultats de la *validation-croisée* (Figure 4-2) démontrent la limite des quatre modèles à produire des estimations de qualité pour les moyennes à faibles concentrations en Chl-a (NASHr < -0.65 pour le meilleur modèle) avec les données désagrégées du capteur MODIS. Rappelons que pour un NASHr négatif, la moyenne des mesures *in situ* est considérée plus précise que les résultats de la modélisation.



Figure 4-2 Concentration en chlorophylle-a estimée à partir des quatre modèles comparée aux mesures in situ pour les données entre 10 et 50 mg Chl-a m⁻³, avec les indices de performance des modèles.

4.4 Conclusions

Le premier objectif de la thèse était de tester le potentiel des bandes désagrégées du capteur MODIS à suivre les FEA dans les lacs de moyenne à petite taille du Québec méridional. Quoiqu'originalement conçues pour d'autres applications (terre, nuage et aérosols), les résultats de la *validation-croisée* des modèles APPEL, Kahru et FAI ont démontré le potentiel des bandes étudiées à produire des estimations acceptables, spécialement en cas de fortes concentrations (R² > 0.89 et NASHr > 0.90). En raison des substitutions apportées aux bandes utilisées pour le modèle Gitelson, ses résultats n'ont pas atteint le même degré de qualité d'estimation que ceux des autres modèles. Cependant, la performance de l'ensemble des modèles étudiés a significativement chuté pour les faibles à moyennes concentrations (< 50 mg Chl-a m⁻³).

À la lumière de ces résultats, il est clair qu'aucun modèle n'est parfait, et que chaque modèle possède ses forces et ses faiblesses. Si la performance des modèles a été assez similaire pour les fortes concentrations, leur performance a toutefois été médiocre pour les moyennes à faibles concentrations (NASHr < 0). Du coup, des efforts de recherche devront être dirigés pour comprendre le comportement spectral des eaux moyennement à faiblement chargées en Chl-a afin d'améliorer la qualité de l'estimation dans ces conditions de floraisons. Ceci permettra un suivi optimal des FEA via leur identification dès leur phase d'éclosion jusqu'à leur extinction et contribuera par conséquent, d'une part, à mieux gérer et contrôler les plans d'eau touchés par les cyanobactéries, et d'autre part, à comprendre les mécanismes de stimulation, de développement, de maintien et de déclin de ces microorganismes.

NB :Il est important à préciser qu'une erreur de calcul du NASHr s'était glissée au moment de la présentation des résultats du premier article. Ceci n'a pas influencé sur les conclusions générales des travaux de cette partie de la thèse. Cette erreur a été corrigée pour la présente synthèse. De ce fait, les valeurs du NASHr présentées dans cette section diffèrent de celles présentées dans l'article #1 (El-Alem et al., 2012).

5 DÉVELOPPEMENT D'UN MODÈLE ADAPTATIF DE SUIVI DES FEA DANS LES EAUX DOUCES EN UTILISANT LES DONNÉES DÉSAGRÉGÉES DU CAPTEUR MODIS

5.1 Introduction

La plupart des études considèrent que les différents niveaux de concentration d'un élément donné (MSS, MODC ou Chl-a) d'un plan d'eau pourraient être modélisés en utilisant une même fonction (Jun. Chen et al., 2013, Gilerson et al., 2010, Yu et al., 2014). Cependant, comme la réponse spectrale est fortement contrôlée par les concentrations de tous ces éléments optiquement actifs, leur fonction de calibration peut différer d'un niveau de concentration à un autre. Par conséquent, ces fonctions peuvent être sous forment linéaire, exponentielle, ou polynomiale ou autres dépendamment des concentrations relatives de chacun de ces éléments. L'utilisation d'une fonction de calibration inappropriée pour une certaine gamme de concentrations peut ainsi mener à sur- ou à sous-estimer une composante donnée. Ceci a été clairement démontré dans l'étude précédente, où la performance des quatre modèles comparés (calibrés par des fonctions exponentielles) avait grandement chuté pour les moyennes à faibles concentrations. D'ailleurs, d'autres études ont montré que l'estimation des faibles concentrations en Chl-a en utilisant des fonctions linaires (Moses et al., 2009) ou polynomiales (Yacobi et al., 2011), était plus précise, alors que l'estimation des concentrations moyennes ou fortes étaient plus précises en utilisant des fonctions exponentielles (Yunlin Zhang et al., 2009a).

Le second objectif de cette partie de la thèse était alors de développer un modèle adaptatif (MA) d'estimation de la concentration en Chl-a, composé d'un *classificateur* et de trois *estimateurs* spécifiques à trois classes de floraisons (eaux faiblement, moyennement et fortement chargées en Chl-a), et d'évaluer sa performance sur une série de plans d'eau (Tableau 3-1) dans le Québec méridional. La performance de ce dernier a ensuite été comparée à celle des modèles APPEL, FAI et Kahru. En raison de sa faible performance, le modèle inspiré de *Gitelson* n'a pas fait partie de la présente comparaison. Le MA ainsi que les trois autres modèles ont été évalués par

59

validation-croisée et par matrice de confusion en utilisant des données ordinales indépendantes des données de calibration.

5.2 Approche méthodologique

De l'analyse du comportement spectrale des eaux fortement chargées en Chl-a (> à 50 mg m⁻³), il s'était avéré que ces dernières possèdent une allure spectrale typique dans la partie du rouge aux CLIR du spectre électromagnétique. Alors que les eaux faiblement à moyennement chargées en Chl-a (< à 50 mg m⁻³), présentent une certaine corrélation avec les bandes du visible (Figure 5-1). Ainsi, il a été décidé d'utiliser cette information afin de développer trois *estimateurs* spécifiques à trois différentes classes de floraisons (eaux faiblement, moyennement et fortement chargées en Chl-a). D'autre part, il a été démontré que les modèles calibrés par <u>régression multiple</u>, améliore significativement la qualité des estimations (Hodjaty *et al.*, 2010). Ainsi, l'utilisation structurée d'*estimateurs* spécifiques a été effectuée par la régression multiple devait permettre d'atteindre une meilleure modélisation de la concentration en Chl-a. La calibration des *estimateurs* spécifiques a été effectuée par la régression pas-à-pas (*stepwise*) en utilisant les différents rapports et différences de bandes (susceptibles de rehausser l'activité bio-optique de la Chl-a) des bandes désagrégées de MODIS et les différents modèles pré-étudiés (Appendix 1 de l'article #3).



Figure 5-1 Signatures spectrales caractéristiques des trois conditions de concentrations de la Chl-a: A) concentrations faibles et moyennes et B) concentrations fortes.

Si la séparation entre la signature spectrale des fortes concentrations de celles des moyennes ou faibles concentrations est évidente (Figure 5-1.B), la séparation des faibles concentrations des moyennes n'est, toutefois, pas tâche facile (Figure 5-1.A). La méthode CART (*Classification and Regression Tree* (Leo. Breiman, 1984)), largement utilisée dans le domaine de classification, a été ainsi appliquée à quatre variables, calculées à partir des bandes désagrégées de MODIS, pour aboutir à cette fin. Les variables d'entrainement utilisées pour la présente étude sont des surfaces (S) calculées au-dessous des courbes de réflectance pour :1) les bandes du visibles (S_{Vis}) ; 2) les bandes du vert au PIR (S_{V-PIR}) ; 3) les bandes du visible au PIR (S_{Vis-PIR}) et 4) les bandes du rouge au CLIR (S_{R-CLIR}) en utilisant l'équation (5-1).

$$S = \frac{1}{2} \sum_{i=1}^{l} |R(\lambda_i) \times \lambda_{i+1} - R(\lambda_{i+1}) \times \lambda_i|$$
(5-1)

où I est le nombre des bandes de MODIS et $R(\lambda_i)$ et λ_i sont respectivement la réflectance et l'absorption de la longueur d'onde de la i^{ième} bande.



Figure 5-2 Résultats de la subdivision de la base de calibration en trois sous-groupes spécifiques. 1) le sous-groupe de données qui a été utilisé pour calibrer l'*estimateur* spécifique des eaux fortement chargées en Chl-a, 2) le sous-groupe de données qui a été utilisé pour calibrer l'*estimateur* spécifique des eaux moyennement chargées en Chl-a et 3) le sous-groupe de données qui a été utilisé pour calibrer l'*estimateur* spécifique des eaux fortement chargées en Chl-a.

Le *classificateur* développé par CART a permis ensuite de subdiviser spectralement la base de calibration en trois sous-groupes spécifiques, permettant alors la calibration des trois *estimateurs* spécifiques (Figure 5-2). Ainsi l'estimation de la concentration en Chl-a pour un pixel donné via cette approche s'effectue sur deux étapes :1) la détermination *a priori* de la classe de floraison à modéliser via le *classificateur* et 2) l'estimation de la concentration en Chl-a par l'intermédiaire de l'*estimateur* spécifique de la classe de floraison prédéterminée (Figure 5-3). D'où l'appellation modèle adaptatif. Les expressions mathématiques des trois *estimateurs* spécifiques du MA sont présentées au Tableau 2 de l'article 2 (El-Alem *et al.*, 2014).



Figure 5-3 Structure interne du modèle adaptatif.

5.3 Résultats et discussion

L'évaluation des quatre modèles par *validation-croisée* a démontré que le MA performe le mieux ($R^2 = 0.96$, NASHr = 0.90 et EQMr = 23% (Figure 5-4)), alors que la performance des autres modèles a été moins bonne et presque similaire entre eux. La Figure 5-4 montre également la bonne dispersion du nuage de points par rapport à la diagonale pour l'ensemble des modèles, avec un léger avantage pour le MA pour les

faibles concentrations, mettant en évidence la robustesse des modèles même à leurs extrémités. Toutefois, malgré la bonne performance de l'ensemble des modèles, il est clair une fois de plus qu'ils détiennent leur force des fortes concentrations. Ainsi, comme pour le premier axe de la thèse et en raison de l'importance du suivi des eaux faiblement à moyennement chargées en Chl-a, il a été nécessaire d'évaluer la performance des modèles dans ces conditions de floraisons. Cette évaluation (Figure 5-5) a démontré que seule la performance du MA peut être considérée acceptable ($R^2 = 0.56$ et NASHr = 0.24); la performance des autres modèles était significativement plus faible, particulièrement pour le NASHr (< à 0).



Figure 5-4 Concentration en chlorophylle-a estimée à partir des quatre modèles comparée aux mesures in situ pour l'ensemble de la base de données, avec les indices de performance des modèles.



Figure 5-5 Concentration en chlorophylle-a estimée à partir des quatre modèles comparée aux mesures in situ en utilisant les concentrations inférieures à 50 mg Chl-a m-3, avec les indices de performance des modèles.

L'évaluation par matrice de confusion (Tableau 5-1 au Tableau 5-4), quant-à-elle, a confirmé les résultats de la *validation-croisée* avec un *Kappa* de 0.51 pour le MA versus 0.21, 0.15 et 0.10 pour les modèles APPEL, FAI et Kahru, respectivement. La classification des estimées en ChI-a pour les fortes concentrations n'a été problématique pour aucun des modèles (erreur de commission < 18%). Cependant, la performance de l'ensemble des modèles a été très modeste pour les eaux moyennement chargées en ChI-a (erreur de commission > 61%). La performance du

MA pour les eaux faiblement chargées en Chl-a a été très acceptable (erreur de commission =19 %) par comparaison aux autres modèles (> à 43%). Les faux positifs pour les eaux faiblement chargées en Chl-a¹⁴ pour le MA étaient aussi acceptables (11%) et les faux positifs pour les eaux moyennement chargées en Chl-a¹⁵ étaient de l'ordre de 22%. Comptabilisant ainsi un risque total de déclaration de fausses floraisons de 33%. Ce risque a été de l'ordre de 49%, 54% et 54% pour APPEL, FAI et Kahru, respectivement. D'autre part, les faux négatifs pour les eaux fortement chargées en Chl-a¹⁶, représentant le risque d'un *«oubli»* de déclaration (ou non-déclaration) d'une FEA, pour le MA étaient de l'ordre de 20%. Ce risque était de l'ordre de 18%, 18% et 12% pour les modèles APPEL, FAI et Kahru, respectivement.

				Observée			
		[Chl-a]<10	10<[Chl-a]<50	[Chl-a]>50	Total	Erreur commission	Taux de succès
	[Chl-a]<10	30	3	4	37	19%	81%
a)	10<[Chl-a]<50	14	14	8	36	61%	39%
Estimé	[Chl-a]>50	2	3	25	30	17%	83%
	Total	46	20	37	103	****	****
_	Erreur omission	35%	30%	32%	*****	****	****
	Taux de succès	65%	70%	68%	*****	****	****
	Succès global	****	****	****	*****	****	67%
	Indice Kappa	****	****	****	****	****	0.51

Tableau 5-1. Résultat de la matrice de confusion du modèle Adaptatif (MA).

Tableau 5-2. Résultat de la matrice de confusion d'APPEL.

				Observée			
		[Chl-a]<10	10<[Chl-a]<50	[Chl-a]>50	Total	Erreur commission	Taux de succès
	[Chl-a]<10	17	5	8	30	43%	57%
a)	10<[Chl-a]<50	29	11	11	51	78%	22%
mé	[Chl-a]>50	0	4	18	22	18%	82%
Esti	Total	46	20	37	103	****	****
_	Erreur omission	63%	45%	51%	****	****	****
	Taux de succès	37%	55%	49%	****	****	****
	Succès global	****	****	****	****	****	45%
	Indice Kappa	****	****	****	****	****	0.21

¹⁴Failes concentrations (04) classées comme fortes sur le total de commission (37).

¹⁵moyennes concentrations (08) classées comme fortes sur le total de commission (36).

¹⁶Somme des fortes concentrations classées comme moyennes (02) ou faibles (03) sur le total de commission (30).

				Observée			
		[Chl-a]<10	10<[Chl-a]<50	[Chl-a]>50	Total	Erreur commission	Taux de succès
	[Chl-a]<10	24	10	13	47	49%	51%
đ	10<[Chl-a]<50	22	7	10	39	82%	18%
mé	[Chl-a]>50	0	3	14	17	18%	82%
Esti	Total	46	20	37	103	****	****
_	Erreur omission	48%	65%	62%	*****	****	****
	Taux de succès	52%	35%	38%	****	****	****
	Succès global	****	****	****	****	****	44%
	Indice Kappa	****	****	****	****	****	0.15

Tableau 5-3. Résultat de la matrice de confusion du modèle FAI.

Tableau 5-4. Résultat de la matrice de confusion du modèle Kahru.

				Observée			
		[Chl-a]<10	10<[Chl-a]<50	[Chl-a]>50	Total	Erreur commission	Taux de succès
	[Chl-a]<10	8	6	6	20	60%	40%
6	10<[Chl-a]<50	38	12	16	66	82%	18%
Estimée	[Chl-a]>50	0	2	15	17	12%	88%
	Total	46	20	37	103	****	****
	Erreur omission	83%	40%	59%	****	****	****
	Taux de succès	17%	60%	41%	****	****	****
	Succès global	****	****	****	****	****	34%
	Indice Kappa	****	****	****	****	****	0.10

La Figure 5-6 illustre le résultat de l'application des quatre modèles sur trois images MODIS. La première (17 septembre. 2001) montre le début de l'établissement d'une floraison sur la baie Missisquoi du lac Champlain, la deuxième (30 septembre. 2001) montre l'expansion totale de la floraison sur la baie et la troisième montre une floraison centrale (16 octobre 2001). Il est clair que l'ensemble des modèles ont été capable, de façon équitable, à détecter la FEA bien établie (30 septembre) et que la détection de la fleur d'eau durant sa phase d'initiation (17 septembre) était problématique pour les modèles standards (APPEL, FAI et Kahru). Cependant, des résultats de l'image du 16 octobre, il est évident que le MA montre un effet d'escalier (à la limite de la fleur d'eau) et qu'il a tendance à surestimer la concentration en Chl-a par rapport aux autres modèles.



Figure 5-6 Application des quatre modèles (modèle adaptatif et les modèles APPEL, FAI et Kahru) à une série d'images MODIS acquises durant l'établissement d'une FEA sur la baie Missisquoi du lac Champlain (la légende est dans échelle logarithmique).

5.4 Conclusions

Par comparaison aux modèles standards, l'évaluation du MA par *validation-croisée* a montré les meilleurs résultats aussi bien pour les fortes que pour les faibles concentrations (seul le NASHr du MA était positif). Les résultats de la matrice de confusion confirment la bonne performance du MA avec un indice de *Kappa* deux fois supérieur à celui du meilleur des modèles standards. Certes la performance pour les eaux fortement chargées en Chl-a était quasi-similaire pour l'ensemble des modèles, mais la performance du MA était significativement distingue pour les eaux faiblement chargées en Chl-a (erreur de commission de 19% versus 43%, 49% et 60% pour APPEL, FAI et Kahru, respectivement), et a montré le plus faible risque de fausses déclaration de floraisons (33%). De plus, lors de l'application des modèles sur les images MODIS, seul le MA a été capable de détecter les FEA durant leur phase d'initiation.

Certes, le MA a montré de bons résultats sur plusieurs plans et son application sur les images MODIS a permis la détection des FEA durant leur phase d'initiation, mais ses performances demeurent modestes pour les eaux moyennement à faiblement chargées en Chl-a (NASHr = 0.24, risque total de déclaration de fausses floraisons de 33% et erreur de commission de 61% pour les eaux moyennement chargées en Chl-a). De plus, son risque de non-déclaration d'une FEA était le plus grand (20% versus 18%, 18% et 12% pour APPEL, FAI et Kahru, respectivement), risque d'un grand intérêt en matière de gestion de l'eau pour des fins de protection de la santé publique. D'autre part, un effet d'escalier lors de la modélisation a été perçu spécifiquement pour les concentrations moyennes à faibles. Tenant compte de toutes ces considérations, il devenait important de concentrer nos travaux dans la prochaine étape sur les faiblesses du MA afin d'optimiser les estimations de la concentration en Chl-a, et ce pour l'ensemble des classes de floraisons et, par conséquent, réduire les risques de fausse et de non-déclaration des floraisons lors de la modélisation des FEA.

6 DÉVELOPPEMENT D'UN MODÈLE BASÉ SUR LES MÉTHODES D'ENSEMBLE POUR L'ESTIMATION DE LA CONCENTRATION EN CHLOROPHYLLE-A

6.1 Introduction

L'étape critique lors de l'estimation de la concentration en Chl-a en utilisant le MA est la sélection de l'*estimateur* final. Une mauvaise sélection de ce dernier, suite à une mauvaise classification d'un pixel donné, pourrait conduire à une sur- ou sous-estimation de la concentration en Chl-a et induire un effet d'escalier au cours de son application sur les images satellitaires. En effet, plusieurs études ont démontré l'instabilité et la localité des techniques de classification basées sur les décisions par arbre (Timofeev, 2004), tel qu'utilisé pour étalonner le *classificateur* du MA. D'autre part, l'estimation finale de la concentration en Chl-a avec le MA est basée sur la décision d'un seul *estimateur* des trois. Or, dans de telles conditions de modélisation complexes, il est souvent recommandé de se baser sur plusieurs décisions avant de prendre la dernière. La combinaison de l'ensemble de ces décisions serait alors la plus exacte. Cette technique connue sous le nom des méthodes d'ensemble est utilisable dans les deux contextes, classification et/ou estimation, et peut être considérée comme la «*révolution»* de la modélisation moderne par apprentissage automatique, connu également sous le nom anglais «*machine learning»*.

Une des deux principales clés pour développer un modèle basé sur les méthodes d'ensemble (MBME) est la diversité de ses éléments (*estimateurs* et/ou *classificateurs* (Polikar, 2006)). Une des plus simples et intuitives méthodes souvent utilisée pour développer un MBME est le *bagging* (Leo Breiman, 1996). La diversité dans le *bagging* est obtenue en utilisant un ré-échantillonnage avec remplacement sur les données de calibration. Toutefois, la principale faiblesse du *bagging* est le temps de calcul dû au nombre des éléments du MBME souvent trop élevé, particulièrement en imagerie. L'utilisation de la formule de la Quadrature gaussienne (QG) serait une bonne solution pour résoudre cette problématique. Cette méthode a le potentiel de convertir le problème de ré-échantillonnage illustré mathématiquement sous forme d'intégration,

71

demandant des calculs ardus et complexes, à un problème de sommation de probabilités pondérées où les techniques de résolutions numériques sont plus simples et sont éprouvées et précises (Tørvi *et al.*, 1997).

Ainsi, notre troisième objectif de la thèse était d'optimiser la performance du MA par le développement d'un MBME pour l'estimation de la concentration en Chl-a dans les eaux douces du Québec méridional et de tester sa performance sur l'ensemble des lacs énumérés dans le Tableau 3-1. L'évaluation du MBME a été effectuée par *validation-croisée*, par matrice de confusion et par coefficient de variation (CV). Basé sur les résultats de la *validation-croisée* et de la matrice de confusion, issue des mêmes jeux de données (continus et ordinals) qui ont servi à l'évaluation des modèles MA, APPEL, Kahru et FAI, une analyse de performance entre le MBME et le MA a été également réalisée.

6.2 Approche méthodologique

Afin d'atteindre une modélisation optimale de la concentration en Chl-a en utilisant le MA, deux sources d'erreurs doivent impérativement être contrôlées. Cela était possible par le développement d'un MBME, composé par un classificateur basé sur les méthodes d'ensemble (CBME) et un estimateur basé sur les méthodes d'ensemble (EBME), par les étapes suivantes:

Erreurs de classification

Cette source d'erreur est reliée aux nœuds du *classificateur* du MA (lignes rouges à la Figure 5-2). En effet, les seuils proposés par la méthode CART sont optimaux pour la base d'entrainement utilisée, mais pas uniques. Une simple modification dans cette base d'entrainement peut mener à de considérables changements dans l'arbre de décisions. Une façon intéressante pour contrôler cette erreur est de la <u>quantifier</u> et de la <u>prendre en considération</u> au moment de l'estimation finale de la concentration en Chl-a. La **quantification** de l'erreur du *classificateur* du MA était possible en utilisant l'algorithme du *bagging* (ré-échantillonnage par remplacement; n_{bag} était fixé à 25,000 pour cette étude). Ceci, avait permis de générer deux *vecteurs-aléatoires* v_1 et v_2

composés de milliers de seuils formant ce qu'on appel un classificateur basé sur les méthodes d'ensemble (CBME), composé de $\frac{n_{bag}}{N_{occur}V1} \times \frac{n_{bag}}{N_{occur}V2}$ classificateurs (avec $N_{occur}V1$ et $N_{occur}V2$ sont respectivement le nombre d'occurrence des seuils obtenus dans v_1 et v_2). L'occurrence et la variance des seuils avaient permis par la suite de déterminer une distribution de probabilité de seuils, sur chacun des nœuds du *classificateurs* du MA (Figure 6-1), qui est caractérisée par une moyenne (μ) et une variance (σ). Basé sur ces moments statistiques, il a été ainsi possible de quantifier l'erreur de classification par les équations suivantes :

$$\mu = \int_{v \in V} f(v) \cdot P(v) dv \tag{6-1}$$

$$\sigma = \int_{v \in V} (f(v) - \mu)^2 P(v) dv$$
(6-2)

ou v est le vecteur-aléatoire appartenant à V qui représente l'espace des variables d'entrée du modèle, f(v) est la sortie du modèle et P(v) et la distribution conditionnelle des variables d'entrée.

Toutefois, **la prise en considération** de ce grand nombre de classificateurs, composant le CBME, pour prendre une décision demandera sûrement un temps de calcul énorme. Tørvi *et al.* (1997) ont proposé une approche basé sur la formule de la quadrature Gaussienne (QG), qui a le potentielle de convertir ces intégrales probabilistes (équations (6-1) et (6-2)) à des somations pondérées qui sont en fonction de n-nœuds (n-seuils optimaux (n_{SO}), qui a était fixé à 3 pour la présente étude) formant au final n_{SO}^2 classificateurs, tout en approximant la performance de l'utilisaition de $\frac{n_{bag}}{N_{occur}V1} \times \frac{n_{bag}}{N_{occur}V2}$ classificateurs. Ainsi, les équations (6-1) et (6-2) prennent les formules suivantes:

$$\mu = \sum_{i=0}^{n_{so}} \omega_i \times f(z_i)$$

$$\sigma = \sum_{i=0}^{n_{so}} \omega_i \times f(z_i - \mu)^2$$
(6-3)
(6-4)

où $f(z_i)$ fonction du *vecteur-aléatoire* standardisé, z_i et ω_i sont respectivement les abscisses et les coefficients de pondération calculés par la formule de la QG pour un SO donné (Tableau 5-4), tel que proposés par Tørvi *et al.* (1997).

$$SO_i = \mu + \sqrt{\sigma \times z_i} \tag{6-5}$$

où: μ et σ sont respectivement la moyenne et la variance de f(z_i).

n _{so}	$\boldsymbol{z}_{\mathrm{i}}$	ω
1	0	1
2	-1,+1	$\frac{1}{2}, \frac{1}{2}$
3	$-\sqrt{3}$, 0 , $+\sqrt{3}$	$\frac{1}{6}, \frac{2}{3}, \frac{1}{6}$

 Tableau 6-1
 Abscisses et poids de la distribution normal standard.

Erreurs d'estimation

En raison de la complexité du phénomène à modéliser, moyennant plusieurs estimations de concentrations en Chl-a réduira le risque d'utiliser une seule estimation. qui pourrait être erronée. Partant de ce principe, et en se basant sur les résultats du CBME, il a été possible de développer un estimateur basé sur les méthodes d'ensemble (EBME). En effet, le CBME avait permis de subdiviser la base de calibration en 21 sous-groupes : 3 (n_{so}) sous-groupes des fortes concentrations en Chl-a, qui sont composés d'échantillons où l'indice spectrale est supérieur aux SO (supérieur, nominal ou inférieur) de la variable S_{R-CLIR} ; 9 (n_{SO}^2) sous-groupes des moyennes concentrations en Chl-a, qui sont composés d'échantillons où les indices spectraux sont à la fois supérieurs aux SO de la variable S_{Vis-PIR} et aux SO de la variable S_{R-CLIR} et 9 sous-groupes des faibles concentrations en Chl-a, qui sont composés d'échantillons où les indices spectraux sont à la fois inférieurs aux SO des deux variables (S_{R-CLIR} et S_{Vis-PIR}) (Figure 6-1). Au moyen de la régression pas-à-pas (*stepwise*) et en utilisant l'ensemble des variables indépendantes utilisées pour calibrer le MA, il a ainsi été possible de calibrer 21 estimateurs spécifiques formant ce qu'on appel un EBME. En utilisant ce schéma, les équations (6-3) et (6-4) prennent les formes des formules génériques suivantes :

$$\mu = \sum_{i=0}^{k} \omega_{i} \times \text{Est}_{\text{SOV1}_{i}^{\uparrow}} + \sum_{i=0}^{n_{so}-k} \omega_{i} \times \left[\sum_{j=0}^{p} \omega_{j} \times \left[\text{Est}_{\text{SOV2}_{j}^{\uparrow};\text{TV1}_{i}^{\downarrow}} \right] + \sum_{j=0}^{n_{so}-p} \omega_{j} \times \left[\text{Est}_{\text{SOV2}_{j}^{\downarrow};\text{TV1}_{i}^{\downarrow}} \right] \right]; \text{ où } \omega_{0} = 0$$
(6-6)

$$\sigma = \sum_{i=0}^{k} \omega_{i} \times \left[\text{Est}_{\text{SOV1}_{i}^{\dagger}} - \mu \right]^{2} + \sum_{i=0}^{n_{so}-k} \omega_{i} \times \left[\sum_{j=0}^{p} \omega_{j} \times \left[\text{Est}_{\text{SOV2}_{j}^{\dagger};\text{SOV1}_{i}^{\dagger}} - \mu \right]^{2} + \sum_{j=0}^{n_{so}-p} \omega_{j} \times \left[\text{Est}_{\text{SOV2}_{j}^{\dagger};\text{SOV1}_{i}^{\dagger}} - \mu \right]^{2} \right]$$
(6-7)

où : n_{so} est le nombre de seuils optimaux ; i et j sont les indices correspondant à chaque seuil optimal (SO) particulier pour V1 et V2 (i = 1: n_{so} et j = 1: n_{so} et où 1 réfère au seuil supérieur (S), 2 réfère au seuil nominal (N) et 3 réfère au seuil inférieur (I)) ; ω_i et ω_j sont les poids associés à chaque SO pour V1 et V2 ; $Est_{SOV1_i}^{\dagger}$ est l'ensemble d'*estimateurs* spécifiques calibrés pour les fortes concentrations en Chl-a ; $Est_{SOV2_j}^{\dagger}$; SOV1_i^{\dagger} est l'ensemble d'*estimateurs* spécifiques calibrés pour les moyennes concentrations en Chl-a ; $Est_{SOV2_j}^{\dagger}$; SOV1_i^{\dagger} est l'ensemble d'*estimateurs* spécifiques calibrés pour les moyennes concentrations en Chl-a ; $Est_{SOV2_j}^{\dagger}$; SOV1_i^{\dagger} est l'ensemble d'*estimateurs* spécifiques calibrés pour les faibles concentrations en Chl-a ; SOV1_i et SOV2_j sont respectivement les seuils optimaux déterminés par la formule de la QG pour V1 et V2 ; V1 et V2 sont respectivement S_{R-CLIR} et $S_{Vis-PIR}$ et k et p sont les indices des SO reliés respectivement à SOV1_i et à SOV2_j (k $\leq n_{so}$ et p $\leq n_{so}$).

Au final, il a ainsi été possible de développer un modèle basé sur les méthodes d'ensemble (MBME) mixte, composé d'un CBME et d'un EBME. L'utilisation de la formule de la QG était bénéfique sur deux plans : 1) réduire le temps de calcul et 2) développer un modèle souple (fonction générique (équation (6-6)), capable de s'adapter à la classe de floraison à modéliser, et où l'estimation finale est une moyenne pondérée d'un ensemble d'*estimateurs* spécifiques (de 3 à 9). Concrètement, prenons le cas de l'étoile bleue dans la classe de floraison numéro 1 à la Figure 6-1 (cas des très fortes concentrations). Malgré que la classification de ce pixel ne soit pas problématique, l'estimation finale de la concentration en Chl-a, dans ce cas, est la moyenne pondérée de la sortie des trois *estimateurs* spécifiques aux fortes concentrations. Pour cette exemple, k = p = 3, ainsi les équations (6-6) et (6-7) prennent les formes suivantes :

$$[Chl - a]; \ \mu = \frac{1}{6} \times Est_{SOV1_{I}^{\uparrow}} + \frac{2}{3} \times Est_{SOV1_{N}^{\uparrow}} + \frac{1}{6} \times Est_{SOV1_{S}^{\uparrow}}$$

$$\sigma_{[Chl-a]}; \ \sigma = \frac{1}{6} \times \left(Est_{SOV1_{I}^{\uparrow}} - \mu \right)^{2} + \frac{2}{3} \times \left(Est_{SOV1_{N}^{\uparrow}} - \mu \right)^{2} + \frac{1}{6} \times \left(Est_{SOV1_{S}^{\uparrow}} - \mu \right)^{2}$$

où, [Chl – a] est l'estimation de la concentration en Chl-a, $\sigma_{[Chl-a]}$ est la variance calculée sur cette estimation et $Est_{SOV1_1^{\uparrow}}$, $Est_{SOV1_N^{\uparrow}}$ et $Est_{SOV1_S^{\uparrow}}$ sont respectivement les *estimateurs* spécifiques calibrés à partir des sous-groupes d'échantillons où l'indice spectrale est supérieur soit au SO inférieur, nominal ou supérieur de S_{R-CLIR} .





6.3 Résultats et discussion

Les résultats de la *validation-croisée* ont été satisfaisants avec un $R^2 = 0.98$ démontrant que le MBME est capable d'expliquer 98% de la variance de la concentration en Chl-a. Le NASHr indique que la modélisation par le MBME est robuste avec un taux de succès de 95%. La robustesse du modèle a aussi été confirmée par le graphique des estimations mesurées en fonction de celles estimées (Figure 6-2.A) montrant une bonne distribution du nuage de points par rapport à la diagonale 1:1. Par comparaison à la performance du MA et des modèles standards (Figure 5-4), tous les indices statistiques se sont améliorés, spécialement pour le NASHr et le EQMr. Toutefois, il était important d'évaluer la performance du MBME pour les moyennes à faibles concentrations en Chl-a. Cette analyse a mis en évidence la bonne qualité d'estimation de la concentration en Chl-a du MBME dans ces conditions de floraisons, particulièrement pour le NASHr qui est passé de 0.24 (très modéré) à 0.70 (trés acceptable). Quant aux erreurs du BIAISr et de l'EQM, elles ont respectivement baissé d'environ 50% et 38% (Figure 6-2.B et Figure 5-5).



Figure 6-2 Concentration en chlorophylle-a estimée à partir du modèle basé sur la méthode de l'ensemble comparée aux mesures *in situ* pour : A) l'ensemble de la base de données et B) en utilisant les concentrations inférieures à 50 mg Chl-a m⁻³, avec les indices de performance des modèles.

Par comparaison aux résultats de la matrice de confusion du MA, l'indice de Kappa ainsi que le succès global ont progressé, en faveur du MBME, de 0.51 à 0.71 et de 67% à 82%, respectivement (Tableau 6-2). Ces résultats démontrent la bonne concordance de la classification entre les estimations et les observations (Kappa > 0.70). Tel que prévu, l'amélioration de la classification des estimations concernait les classes des eaux faiblement à moyennement chargées en Chl-a, plus particulièrement des concentrations moyennes où l'erreur de commission a baissé de 61% à 35%. Cette classe est particulièrement sensible aux mauvaises classifications parce qu'elles peuvent se produire aux extrémités (El-Alem et al., 2014). De l'analyse des deux matrices (AM et MBME), la classe des concentrations moyennes a gagnée plus de précision des concentrations faibles que des fortes. Les faux positifs des eaux faiblement et moyennement chargées en Chl-a ont baissé de 11% à 9% et de 22% à 12%, respectivement, comptabilisant ainsi un risque total de 21% pour la déclaration de fausses floraisons, erreur qui est acceptable. De plus, les faux négatifs pour les eaux fortement chargées en Chl-a ont été de l'ordre de 10% enregistrant ainsi la meilleure performance par rapport aux modèles préétudiés (12% pour Kahru, (Tableau 5-4)).

				Observée			
		[Chl-a]<10	10<[Chl-a]<50	[Chl-a]>50	Total	Erreur commission	Taux de succès
	[Chl-a]<10	39	3	4	46	15%	85%
a)	10<[Chl-a]<50	6	17	3	26	35%	65%
mé	[Chl-a]>50	1	2	28	31	10%	90%
Esti	Total	46	20	37	103	****	****
-	Erreur omission	15%	15%	24%	****	****	****
	Taux de succès	85%	85%	76%	****	****	****
	Succès global	****	****	****	****	****	82%
	Indice Kappa	****	****	****	*****	****	0.71

Tableau 6-2Résultat de la matrice de confusion du modèle basé sur la méthode d'ensemble.

La Figure 6-3 montre l'application du MBME et du MA sur la même série d'images utilisées auparavant (Figure 5-6) ; comme la détection de la FEA durant sa phase d'expansion totale ne présentait aucune problématique pour l'ensemble des modèles, cette image n'a pas été présentée dans cet exemple. L'analyse qualitative des sorties des deux modèles montre clairement que le MBME arrive à : 1) mieux copier les floraisons aperçues sur le composé RVB ; 2) détecter les floraisons durant leur phase

d'initiation (17 septembre. 2001) et 3) atténuer l'effet d'escalier aperçu sur l'image du 16 octobre. 2001 (Figure 5-6) rendant ainsi la modélisation plus réaliste. D'un autre côté, comme l'estimation de la concentration en Chl-a est effectuée sur la base d'un ensemble d'*estimateurs* spécifiques, ceci permet de calculer une variance sur cette estimation (équation (6-7)). Via ces moments statistiques, il est également possible de calculer le coefficient de variation (équation (3-9)) qui est une mesure d'homogénéité des individus d'une population donnée par rapport à la moyenne. Basé sur cet indice, dont les résultats sont présentés à la Figure 6-4, l'estimation de la concentration en Chl-a issue du MBME peut être considérée fiable (teinte blanche à bleue claire (< 10%) sur la presque totalité de la baie Missisquoi) pour l'ensemble des classes de floraison, exception faite pour quelques pixels au sud de la baie où le CV est plus important ; voir interprétation du CV dans la section «3.1.5. *indices d'évaluation statistiques*».



Figure 6-3 Application des deux modèles (modèle basé sur la méthode de l'ensemble et le Modèle adaptatif (MA)) sur les mêmes images utilisées à la (Figure 5-6) ; rappelons que la légende est dans une échelle logarithmique.

19 Sept. 2001



Figure 6-4 Exemple de l'application du modèle basé sur la méthode de l'ensemble (MBME) sur une floraison détectée le 19 septembre 2001 et de l'évaluation de ses estimations par le coefficient de variation.

6.4 Conclusions

Par comparaison aux modèles standards et au MA, l'évaluation du MBME par *validation-croisée* a montré les meilleurs résultats pour l'ensemble des conditions de floraison, spécifiquement pour les eaux moyennement à faiblement chargées en Chl-a où la performance du NASHr s'était significativement améliorée. Les résultats de la matrice de confusion étaient satisfaisants avec un succès global de 82% et un *Kappa* de 0.71 démontrant la bonne concordance de classification des concentrations en Chl-a estimées par le MBME versus leurs correspondantes mesurées. L'erreur de commission pour les eaux moyennement chargées en Chl-a, qui était problématique pour l'ensemble des modèles préétudiés (AM, APPEL, FAI et Kahru), a connu une nette amélioration (35% versus 61% pour la meilleure performance (AM)). De plus, le risque de non-déclaration de FEA, qui était aussi problématique pour le MA, a été réduit de 20% à 10% et est maintenant le plus faible de tous les modèles tout en gardant le plus faible risque (21%) de fausses déclarations de fleurs d'eau. D'autre part, l'évaluation par le CV a montré que les estimations de la concentration en Chl-a par le MBME sont qualifiées de fiables à très fiables (CV < 10% pour la majorité de la modélisation).

Les résultats ci-haut démontrent clairement la valeur de l'utilisation des méthodes d'ensemble qui a conduit au développement d'un modèle d'estimation de la

concentration en Chl-a robuste (meilleures performances en utilisant les données continues et ordinales) et innovateur (production de cartes d'auto-évaluation via le calcul du CV). En dépit de l'utilisation des bandes du capteur MODIS non centrées sur les pics d'absorption et de réflectance de la Chl-a, le MBME a démontré une grande capacité à suivre l'évolution des FEA dans les lacs de moyenne à petite taille du Québec méridional, même durant leur phase d'initiation avec des indices de performances satisfaisants (NASHr = 0.70, EQMr = 37%, erreur de commission pour les eaux faiblement chargées en Chl-a de 15%).

7 CONCLUSION GÉNÉRALE

7.1 Évaluation des hypothèses

Le suivi de la fréquence temporelle et de la distribution spatiale des fleurs d'eau d'algues et/ou de cyanobactéries (FEA) revêt un grand intérêt pour la communauté scientifique et les gestionnaires en eau. Cet intérêt est justifié par les risques environnementaux, économiques ou de santé publique, associés à la présence de ces floraisons. Le suivi des plans d'eau susceptibles à développer une FEA en utilisant les méthodes d'échantillonnage standards, avec une fréquence temporelle élevée est une tâche laborieuse et coûteuse et est limitée dans le temps et dans l'espace. Ainsi, l'objectif de cette recherche était de développer une approche de suivi des FEA à l'aide de l'imagerie du capteur MODIS désagrégée à 250 mètres de résolution spatiale, adaptée aux lacs du Québec méridional.

Les travaux de la présente recherche doctorale se sont basés sur un ensemble d'hypothèses dont la première était :

 «Quoique non-destinées initialement à l'étude de la couleur de l'eau, les bandes MODIS 1 à 7 désagrégées à 250 mètres de résolution spatiale peuvent être utilisées pour le suivi des FEA dans les lacs de moyenne à petite taille, et ce en raison de leur résolution spatiale plus fine».

Ainsi, il a été question dans un premier lieu de tester le potentiel des bandes désagrégées du capteur MODIS à détecter les FEA dans les eaux douces de moyenne à petite taille. Pour ce faire, les modèles d'estimation de la concentration en Chl-a pour les eaux douces, les plus utilisés dans la littérature (Kahru, Gitelson et FAI (*floating algae index*)) ainsi que le modèle APPEL (*APProach By ELimination*), développé au cours de la présente thèse, ont été calibrés et validés à l'aide des données MODIS désagrégées et des données *in situ* de la concentration en Chl-a collectées sur quatre lacs différents du sud du Québec (durant la période 2000-2008). Les résultats de validation des quatre modèles ont été satisfaisants, spécifiquement pour les fortes concentrations, démontrant ainsi **le potentiel des bandes désagrégées du capteur MODIS à estimer la concentration en Chl-a, en particulier lorsque les FEA sont**

bien établies. Toutefois, les modèles étudiés ont eu de la difficulté à produire une bonne qualité d'estimation pour les eaux faiblement chargées en Chl-a. Ceci correspond à la phase d'initiation des FEA, phase importante puisque c'est à ce moment où les efforts de gestion doivent être mis en place. Deux facteurs peuvent expliquer la faible performance des modèles dans ces conditions de floraisons : 1) la région spectrale utilisée (rouge/PIR) est plus adaptée à la détection des mécanismes de diffusion causés par la turbidité du phytoplancton à forte concentration et 2) la structure mathématique des modèles consiste en une et une seule fonction, de type exponentiel, qui n'est adapté qu'aux eaux moyennement à fortement chargées en Chl-a. Ces deux constats ont conduit à émettre les deux hypothèses suivantes :

- «La combinaison des informations spectrales allant du visible aux courtes longueurs d'onde de l'infrarouge pourrait faciliter la détection de la Chl-a et sa discrimination des autres éléments optiquement actifs présents dans les milieux lacustres, tels que la MSS et la MODC.» ;
- 3. «Le développement d'un modèle spécifique à chaque niveau ou classe de concentrations en Chl-a peut améliorer la précision des estimations.».

Afin de tester ces deux hypothèses un modèle adaptatif (MA) d'estimation de la concentration en Chl-a a été développé. Ce modèle est constitué d'un *classificateur*, calibré à l'aide de la méthode d'apprentissage par arbres de décision (appelée en anglais CART (*Classification and Regression Tree*)), et de trois *estimateurs* spécifiques pour trois niveaux trophiques ou classes de floraisons (eaux faiblement, moyennement et fortement chargées en Chl-a). Ces trois *estimateurs* ont été calibrés par la technique de régression multiple pas-à-pas (*stepwise*) linéaires et non-linéaires en utilisant divers rapports et différences des bandes désagrégées (susceptible de rehausser l'activité bio-optique de la Chl-a). Ainsi le MA procède sur deux étapes : 1) identification de la calisse de floraisons à modéliser via son *classificateur* et 2) estimation de la concentration en Chl-a par l'intermédiaire de l'*estimateur* correspondant à la classe pré-identifiée. Les résultats de la calibration ont mis en évidence la valeur de l'utilisation de la régression pas-à-pas, qui a permis la sélection des régions spectrales adéquates pour chaque *estimateur*, spécifiquement pour la classe des moyennes concentrations où la détection de la Chl-a est contrôlée à la fois par les mécanismes d'absorption et de

84

diffusion. Par comparaison aux modèles standards (APPEL, Kahru et FAI), la performance du MA était la meilleure aussi bien sur la base des résultats des indices d'évaluation statistiques (R², EQMr, BIAISr et NASHr) que des résultats des matrices de confusion (*Kappa* et succès global). De plus, lors de l'analyse qualitative des quatre modèles, seul le MA était capable de détecter les FEA durant leur phase d'initiation.

Cette étude a ainsi mis en évidence l'importance de l'utilisation des régions spectrales appropriées pour modéliser une classe de floraison donnée et que la reconnaissance *a priori* des classes à modéliser peut grandement améliorer les résultats de l'estimation de la concentration en Chl-a, spécifiquement durant la phase d'initiation des FEA. Toutefois, la performance du MA demeure modérée pour la classe des concentrations moyennes, classe de transition où les erreurs de classification peuvent être commises aux deux extrémités (concentrations moyennes-faibles et moyennes-fortes). De plus, outre sa faible performance de non-déclaration de floraisons, un certain effet d'escalier était observé lors de l'application du MA sur certaines images MODIS.

En développant le MA, les limites technologiques liées aux bandes désagrégées du capteur MODIS pour la détection de la Chl-a ont été atteintes. Certes, des améliorations ont été apportées, mais la modélisation des concentrations moyennes et faibles demeure peu acceptable. Pour atteindre une modélisation optimale de la concentration en Chl-a en utilisant le MA, deux sources d'erreurs devaient être minimisées : 1) erreur de **classification** en raison de l'instabilité et de la localité des méthodes d'apprentissage par arbres de décision (tel utilisé pour étalonner le *classificateur* du AM) et 2) erreur **d'estimation** de la concentration en Chl-a en raison de la complexité du phénomène à modéliser aux plus faibles concentrations où les facteurs d'interférence prennent une plus grande importance relative. En effet, tous processus de modélisation (qu'il soit utilisé à des fins de classification ou de régression) est une simplification de la réalité et est porteur d'incertitude. Le contrôle de ces sources d'erreurs pourrait réduire les erreurs de modélisation du MA, spécifiquement au niveau des zones de transition d'une classe de floraison à une autre. Ainsi, la dernière hypothèse de recherche de la thèse était :

85

4. «L'utilisation de méthodes statistiques avancées telles que les méthodes d'ensemble à des fins de classification et de régression peut améliorer la précision des estimations en Chl-a et réduire l'effet en escalier causé par la conception générale du MA.».

Pour tester cette dernière hypothèse un modèle basé sur la méthode de l'ensemble (MBME) a été développé et validé sur le même jeu de données et ses performances ont été comparées au MA. En plus de la modélisation de la concentration en Chl-a, la conception générale du MBME permet aussi la production de supports matriciels annexes d'aide à la décision tels que les cartes de la variance et du coefficient de variation (CV). La majeure limite d'application des méthodes de l'ensemble, spécifiquement en télédétection, est le temps de calcul en raison du grand nombre des éléments constituant l'ensemble (qu'ils soient classificateurs ou estimateurs). Cette problématique a toutefois été évitée en utilisant la formule de la quadrature Gaussienne qui a le potentiel de transformer les problèmes de ré-échantillonnages en problèmes numériques permettant ainsi de réduire significativement de temps de traitement tout en gardant la même performance.

L'utilisation des méthodes de l'ensemble a en effet permis d'améliorer la qualité de l'estimation de la concentration en Chl-a. Certes, pour les eaux fortement chargées en Chl-a, la performance des deux modèles (MBME et AM) a été assez similaire, avec un léger avantage pour le MBME. Cependant, pour les eaux moyennement à faiblement chargées en Chl-a, l'amélioration était nettement significative où les indices de performance du MBME ont doublé et où le CV n'a pas dépassé le 10% démontrant la fiabilité des estimations dans ces conditions de floraisons. De plus, le MBME était capable d'atteindre des taux de succès élevés pour la classe des concentrations moyennes malgré son exposition au risque de mauvaises classifications aux deux extrémités. Les risques de déclaration de fausses floraisons ou de non-déclaration de FEA, qui sont considérables en matière de gestion des ressources hydriques et prises de décisions, étaient les plus minimes pour le MBME. De plus, le MBME a aussi été capable de réduire l'effet d'escalier en permettant un passage graduel des fortes aux faibles concentrations en Chl-a, tout en gardant la capacité à détecter les FEA durant leur phase d'initiation. Donc, **le contrôle des sources d'erreurs (classification et**
estimation) par une approche statistique ainsi que leur prise en considération lors de l'estimation finale de la concentration en Chl-a ont significativement amélioré la qualité de la modélisation des FEA, particulièrement durant leur phase d'initiation, et ont également mis en évidence la valeur de l'utilisation des méthodes d'ensemble pour de telles applications.

7.2 Limites

L'approche proposée a permis le développement d'un modèle d'estimation de la concentration en Chl-a dans les eaux douces. Cependant, comme pour l'ensemble des capteurs optiques, la présence de **nuages** limite l'application du modèle sur les images. D'autre part, à l'heure actuelle, le masque de nuages disponible sur le site de la NASA est à 1 km. L'utilisation de ce produit mène ainsi à une perte de résolution spatiale et réduit, par conséquent, l'avantage d'utiliser les images désagrégées du capteur MODIS pour suivre les FEA dans les lacs de moyenne à petite taille. La deuxième limitation, est en relation avec les pixels mixtes. En fait, ces derniers présentent une grande problématique lors de la modélisation des FEA puisqu'ils ont tendance à surestimer la concentration en Chl-a due à la présence de plantes aquatiques près des berges, alors que cette zone demeure aussi un endroit où les fleurs d'eau peuvent s'accumuler par l'action du vent. Une méthode d'identification et d'élimination de ces pixels a été développée durant les travaux de cette thèse. Or, leur élimination conduit à un manque d'information pour les cent premiers mètres près des berges des lacs, zone d'intérêt en termes de santé publique, limitant ainsi l'implémentation opérationnelle de cet outil pour ces zones.

7.3 Aspects innovateurs, contributions et retombées

Un des aspects innovateurs de la présente thèse réside dans le fait d'utiliser **les sept premières bandes du capteur MODIS désagrégées à 250 mètres de résolution spatiale, bandes initialement** <u>non-conçues</u> **pour le suivi de la qualité de l'eau**. Ceci a permis : 1) de bénéficier d'une résolution optimale pour le suivi des FEA dans les eaux douces de moyenne à petite taille ; 2) d'exploiter une large gamme spectrale,

allant du bleu aux CLIR, permettant d'optimiser la détection de la Chl-a et sa discrimination des autres éléments optiquement actifs (TSS et MODC) et 3) la possibilité de couvrir un immense territoire en temps-quasi-réel, avec une fréquence quotidienne et pour un coût réduit.

Le deuxième aspect innovateur de la présente thèse réside dans le développement méthodologique, inédit, combinant à la fois régression multiple, classification par arbres de décision, quadrature Gaussienne et méthodes d'ensemble. Cet effet conjugué a permis d'exploiter les forces de chacune de ces méthodes et a abouti au **développement d'un modèle robuste et fiable pour l'estimation de la concentration en Chl-a dans les eaux douces**, capable non seulement de produire des cartes de la qualité de l'eau mais aussi de calculer la variance de l'estimation pour chaque pixel traité.

Il est à préciser que l'approche développée peut être entièrement automatisée et la carte de distribution des concentrations en Chl-a peut être diffusée en temps-quasi-réel sur Google Earth, par exemple, aux différentes entités décisionnelles en gestion de l'eau. L'implémentation de l'approche en mode opérationnel mènera à d'importantes retombées économiques, et ce en permettant de : 1) profiter de la disponibilité quotidienne et de la gratuité des images afin de suivre, en temps-quasi-réel, les plans d'eaux susceptibles à développer une FEA à moindre coût, laissant aux gestionnaires une marge de temps pour prendre les mesures nécessaires ; 2) planifier a priori les missions de terrain en fonction du niveau de détérioration et de l'importance des plans d'eau touchés, et par conséquent minimiser les coûts alloués aux déplacements inutiles et 3) suivre simultanément la qualité des eaux d'un ensemble de lacs sur un territoire aussi vaste que le Québec méridional, et ce en un minimum de temps. En fait, le développement de cette approche n'avait pas pour but de remplacer les méthodes standards de suivi mais de développer un outil d'aide à la décision fiable venant en appui aux programmes de suivi des FEA pour une meilleure gestion logistique. Les méthodes standards sont d'ailleurs nécessaires pour la calibration et la validation de ces approches, et pour mesurer d'autres variables associées à la présence des cyanobactéries, qui sont fondamentales dans un contexte de gestion telles que les cyanotoxines.

88

7.4 **Perspectives**

L'application de l'approche développée dans cette thèse pourra contribuer à la connaissance et la compréhension des facteurs menant au développement, au maintien et au déclin des FEA. En effet, en plus de produire des images quotidiennes de la distribution de la concentration en Chl-a dans les eaux douces du Québec méridional, l'approche développée peut être utilisée pour **construire une base historique de l'occurrence des épisodes des FEA durant les quinze dernières années (2000 à 2014)**. Une telle base de données pourrait par la suite être sujette à diverses études statistiques allant d'une simple analyse descriptive à des statistiques plus avancées (modélisation du risque conjoncturel ou structurel).

L'analyse descriptive permettrait de dresser un portrait historique de l'occurrence des FEA dans les eaux douces du Québec méridional. Cette étude inédite, fournira alors des informations de base pour étudier le lien entre la fréquence et l'intensité des épisodes de FEA en fonction de la taille, la répartition géographique ou encore le type d'occupation du sol avoisinant les plans d'eau, pour une meilleure compréhension de cette problématique au Québec. De plus, cette base historique pourrait constituer le point de départ pour développer un modèle de l'aléa structurel des FEA, et ce en combinant les conditions climatologiques prédominantes des bassins versants où se développent une FEA, avec les facteurs physiographiques écologiques ou liées à l'action anthropique et la fréquence de l'occurrence des FEA issue de la base historique de télédétection. Un tel modèle permettrait, d'une part, de décrire les relations entre le développement des FEA et les variables géo-environnementales d'un bassin versant donné et, d'autre part, d'estimer la probabilité d'apparition des FEA dans un plan d'eau donné pour lequel on ne dispose que des données caractéristiques de son bassin versant. Il permettrait également : 1) d'améliorer notre compréhension sur l'apparition des FEA; 2) d'étudier l'impact des changements climatiques ou d'un accident écologique sur ces écosystèmes et 3) d'évaluer des scénarii d'atténuation visant à améliorer la qualité de l'eau avant leur mise en œuvre (pratiques de gestion bénéfiques). Le développement d'un modèle de l'aléa conjoncturel des FEA serait aussi possible par combinaison des différents facteurs hydrométéorologiques menant à

la stimulation des FEA aux données synchrones en biomasse algale issues de la base historique de télédétection. Un tel modèle permettrait de prédire l'évolution spatio-temporelle des FEA en tenant compte de l'effet conjugué de l'ensemble des facteurs hydrométéorologiques. Malgré qu'un tel modèle demeure spécifique au lac étudié, il permettrait toutefois de fournir le cadre conceptuel pour la modélisation de l'éminence d'une FEA à l'aide uniquement des données hydrométéorologiques, plus facilement accessibles que les données de biomasse algale.

D'autre part, afin de suivre les FEA dans les zones à proximité des berges (1^{er} cent mètres) pour des fins de sécurité publique, il est possible d'adapter la méthodologie développée aux données d'autres capteurs à haute résolution spatiale tels Spot, RapidEye ou encore FORMOSAT, ou à très haute résolution spatiale tels KOMPSAT-3, QuickBird ou GeoEye-1 dont la fréquence temporelle est équivalente à celle du capteur MODIS (de 1 à 3 jours). Cependant, certains coûts seront associés à l'emploi de ces images limitant ainsi leur utilisation quotidienne par les gestionnaires des ressources en eau. D'autre part, il est également possible d'adopter les méthodes d'ensemble pour le suivi des FEA en haute mer où l'utilisation d'une résolution spatiale grossière ne constitue pas une problématique, en utilisant l'information spectrale des bandes 8 à 16 du capteur MODIS (1 km de résolution spatiale) qui sont centrées sur les régions d'absorption et de réflectance de la Chl-a.

RÉFÉRENCES

- AFSSA & AFSSET (2006) Rapport sur l'évaluation des risques liés à la présence de cyanobactéries et de leurs toxines dans les eaux destinées à l'alimentation, à la baignade et autres activités récréatives. (Agence Française de Sécurité Sanitaire des Aliments et l'Agence Française de Sécurité Sanitaire de l'Environnement et du Travail, France), p 231.
- Ahn YH & Shanmugam P (2006) Detecting the red tide algal blooms from satellite ocean colour observations in optically complex Nourtheast-Asia coastal waters. *Remote Sensing of Environment* 103:419-437.
- Anderson DM, Glibert PM & Burkholder JM (2002) Harmful algal blooms and eutrophication: nutrient sources, composition, and consequences. Estuarine Research Federation, Lawrence, KS, ETATS-UNIS. 23 p
- Bailey SW & Werdell PJ (2006) A multi-sensor approach for the on-orbit validation of ocean color satellite data products. *Remote Sensing of Environment* 102(1–2):12-23.
- Baruah PJ, Tamura M, Oki K & Nishimura H (2002) Neural network modeling of surface chlorophyll and sediment content in inland water from Landsat Thematic Mapper imagery using multidate spectrometer data. p 205-212.
- Becker RH, Sultan MI, Boyer GL, Twiss MR & Konopko E (2009) Mapping cyanobacterial blooms in the Great Lakes using MODIS. *Journal of Great Lakes Research* 35(3):447-453.
- Berk A, Anderson GP, Bernstein LS, Acharya PK, Dothe H, Matthew MW, Adler-Golden SM, Chetwynd JH, Richtsmeier SC, Pukall B & Others (1999) MODTRAN 4 radiative transfer modeling for atmospheric correction. *Proceedings of SPIE- The International Society for Optical Engineering.* p 348-353.
- Beutler M, Wiltshire K, Meyer B, Moldaenke C, Lüring C, Meyerhöfer M, Hansen UP & Dau H (2002) A fluorometric method for the differentiation of algal populations in vivo and *in situ. Photosynthesis Research* 72(1):39-53.
- Binding CE, Greenberg TA, Jerome JH, Bukata RP & Letourneau G (2011) An assessment of MERIS algal products during an intense bloom in Lake of the Woods. *Journal of Plankton Research* 33(5):793-806.
- Blais S (2008) Guide d'identification des fleurs d'eau de cyanobactéries. Comment les distinguer des végétaux observés dans nos lacs et nos rivières. 3e édition, Direction du suivi de l'état de l'environnement, Ministère du Développement durable, de l'Environnement et des Parcs, ISBN : 978-2-550-52408-3 (version imprimée):p 54.
- Boissonneault Y, Laurion I & Vincent W (2007) Les floraisons de cyanobactéries : analyse des données du MDDEP sur sept lacs québécois. Édit R-918 (Institut national de la recherche scientifique - Centre Eau, Terre et Environnement Québec).

- Bottomley H (2003) Between the Sinusoidal projection and the Werner: an alternative to the Bonne. *European Journal of Geography* 10.4000/cybergeo.3977.
- Bouchard Valentine M (2004) *Floraisons de cyanobactéries au lac Saint-Augustin; Dynamique à court terme et stratification*. Maître ès Sciences (M.Sc.) (Laval, Québec). 130 p
- Bowker DE, Davis RE, Myrick DL, Stacy K & Jones WT (1985) Spectral Reflectances of Natural Targets for Use in Remote Sensing Studies.181.
- Breiman L (1984) Classification and regression trees. Chapman & Hall.
- Breiman L (1996) Bagging Predictors. *Machine Learning* 24(2):123-140.
- Bresciani M, Giardino C, Bartoli M, Tavernini S, Bolpagni R & Nizzoli D (2011) Recognizing harmful algal bloom based on remote sensing reflectance band ratio. *APPRES* 5(1):053556-053556.
- Brient L, Lengronne M, Bertrand E, Rolland D, Sipel A, Steinmann D, Baudin I, Legeas M, Le Rouzic B & Bormans M (2008) A phycocyanin probe as a tool for monitoring cyanobacteria in freshwater bodies. *Journal of Environmental Monitoring* 10(2):248-255.
- Bukata RP, John H J, Alexander SK & Dimitry VP (1995) Optical Properties and Remote Sensing of Inland and Coastal Waters. NY:CRC
- Butler WL (1978) Energy Distribution in the Photochemical Apparatus of Photosynthesis. *Annual Review of Plant Physiology* 29(1):345-378.
- Byth S (1980) Palm Island mystery disease. *Medical Journal of Australia* 2(1):40-42.
- Carder KL, Chen FR, Lee ZP, Hawes SK & Kamykowski D (1999) emianalytic Moderate-Resolution Imaging Spectrometer Algorithms for Chlorophyll A and Absorption with Bio-Optical Domains Based on Nitrate-Depletion Temperatures. *Marine Science Faculty Publications* <u>http://scholarcommons.usf.edu/msc facpub/8:8p</u>.
- Carmichael WW (1992) Cyanobacteria secondary metabolites-the Cyanotoxins (a review). *Journal of Applied Bacteriology* 72 (6):445-459.
- Carmichael WW (1994) An overview of toxic cyanobacteria research in the United States: Toxic cyanobacteria: current status of research and management. (Australian Centre for Water Quality Research, Salisbury, Australie).
- CEAEQ (2010) Centre d'expertise en analyse environnementale du Québec. <u>http://www.ceaeq.gouv.qc.ca/index.asp</u>.
- CEAEQ (2012a) (CENTRE D'EXPERTISE EN ANALYSE ENVIRONNEMENTALE DU QUÉBEC). (Détermination de la chlorophylle a: méthode par fluorométrie, MA. 800 Chlor. 1.0, Rév. 2, Mnistère du Développement durable, de l'Environnement et des Parcs du Québec), p 16.
- CEAEQ (2012b) (CENTRE D'EXPERTISE EN ANALYSE ENVIRONNEMENTALE DU QUÉBEC). (Dépistage des cyanobactéries, MA. 800 Cya.dep 1.0, M, Mnistère du Développement durable, de l'Environnement et des Parcs du Québec), p 16.

- Chavez PS (1988) An Improved Dark-Object Substraction Technique for Atmospheric Scattering Correction of Multispectral Data. *Remote Sensing of Environment* 24:459-479.
- Chavez PS (1996) Image-based atmospheric correction-Revisited and improved. *Photogrammetric Engineering and Remote Sensing* 62:1025-1036.
- Chen J, Zhang X & Quan W (2013) Retrieval chlorophyll-a concentration from coastal waters: three-band semi-analytical algorithms comparison and development. *Opt. Express* 21(7):9024-9042.
- Chen Z, Hu C & Muller-Karger F (2007) Monitoring turbidity in Tampa Bay using MODIS/Aqua 250-m imagery. *Remote sensing of Environment* 109(2):207-220.
- Chorus I & Bartram J (1999) Toxic cyanobacteria in water. A guide to their public health consequences, monitoring and management. *WHO. E & FN Spon, London*:416 pp.
- Cohen J (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20:37-46.
- COMITAS (1988) Glossaire des termes officiels de la télédétection aéroporté. *Bulletin de la Société française de la photogrammétrie de télédétection* 112:1-63.
- Dall'Olmo G & Gitelson AA (2005) Effect of bio-optical parameter variability on the remote estimation of chlorophyll-a concentration in turbid productive waters: experimental results. *Appl. Opt.* 44(3):412-422.
- Darecki M & Stramski D (2004) An evaluation of MODIS and SeaWiFS bio-optical algorithms in the Baltic Sea. *Remote Sensing of Environment* 89(3):326-350.
- Datla P (2011) The wonder molecule called phycocyanin. http://www.valensa.com/pdf/Phycocyanin The%20Wonder%20Molecule.pdf.
- Dekker AG (1993) Detection of optical water quality parameters for eutrophic waters by high resolution remote sensing. PhD (Vrijie Amsterdam, The Netherlands).
- Dekker AG, Hoogenboom HJ, Goddijn LM & Malthus TJM (1997) The relation between inherent optical properties and reflectance spectra in turbid inland waters. *Remote Sensing Reviews* 15(1-4):59-74.
- Dekker AG, Malthus TJ & Seyhan E (1991) Quantitative modeling of inland water quality for high-resolution MSS systems. *Geoscience and Remote Sensing, IEEE Transactions on* 29(1):89-95.
- Dekker AG, Malthus TJM & Goddijn LM (1992) Monitoring cyanobacteria in eutrophic waters using airborne imaging spectroscopy and multispectral remote sensing systems. *6th Australian Remote Sensing Conference.* (Wellington, New Zealand, 2-6 November 1992).
- Desai S, Dhargalkar V, Ingole B & Kavlekar D (1978) Phytoplankton Identification Manual. <u>http://nio.org/userfiles/file/biology/Phytoplankton-manual.pdf</u>.
- Dillenberg HO & Dehnel MK (1960) Toxic waterbloom in Saskatchewan, 1959. *Can Med Assoc J* 83:1151-1154.

- Eija R, Rvi RO, Seija HI, Juha-Markku Ln & Mika R (1998) Effect of sampling frequency on detection of natural variability in phytoplankton: unattended high-frequency measurements on board ferries in the Baltic Sea. *Journal of Marine Science* 55:8p.
- El-Alem A, Chokmani K, Laurion I & El-Adlouni SE (2012) Comparative Analysis of Four Models to Estimate Chlorophyll-a Concentration in Case-2 Waters Using MODerate Resolution Imaging Spectroradiometer (MODIS) Imagery. *Remote Sensing* 4(8):2373-2400.
- El-Alem A, Chokmani K, Laurion I & El-Adlouni SE (2014) An Adaptive Model to Monitor Chlorophyll-a in Inland Waters in Southern Quebec Using Downscaled MODIS Imagery. *Remote Sensing* 6(7):6446-6471.
- Ellis D (2009) Guide d'intervention pour les propriétaires, les exploitants ou les concepteurs de stations de production d'eau potable municipales aux prises avec une problématique de fleurs d'eau de cyanobactéries. in *ISBN 978-2-550-53297-2* (Ministère du Développement durable, de l'Environnement et des Parcs, Direction des politiques de l'eau, Québec), p 46 p.
- Fawell JK, Hart JJ, Ames HA & Parr W (1993) Blue-green algae and their toxinsanalysis, toxicity, treatment and environmental control. *Water Supply* 11 : 3-4:109-121.
- Gilerson AA, Gitelson AA, Zhou J, Gurlin D, Moses W, Ioannou I & Ahmed SA (2010) Algorithms for remote estimation of chlorophyll-a in coastal and inland waters using red and near infrared bands. *Opt. Express* 18(23):24109-24125.
- Gilerson AA, Zhou J, Hlaing S, Ioannou I, Schalles J, Gross B, Moshary F & Ahmed S (2007) Fluorescence component in the reflectance spectra from coastal waters. Dependence on water composition. *Opt. Express* 15(24):15702-15721.
- Gitelson AA (1992) The peak near 700 nm on radiance spectra of algae and water: relationships of its magnitude and position with chlorophyll concentration. *International Journal of Remote Sensing* 13(17):3367-3373.
- Gitelson AA, Gurlin D, Moses WJ & Barrow T (2009) A bio-optical algorithm for the remote estimation of the chlorophyll-a concentration in case 2 waters. *Environmental Research Letters* 4(4):045003.
- Gitelson AA, Keydan G & Shishkin V (1985) Inland waters quality assessment from satellite data in visible range of the spectrum. *Sovjet Journal of Remote Sensing* 6:28-36.
- Gitelson AA, Nikanorov AM, Szabo GY & Szilagyi. F (1986) Etude de la qualité des eaux de surface par télédétéction. Monitoring to Detect Changes in Water Quality Series. (*Proceedings of the Budapest Symposium, July 1986*), IAHS Publ 157:111-121.
- Gitelson AA, Schalles JF & Hladik CM (2007) Remote chlorophyll-a retrieval in turbid, productive estuaries: Chesapeake Bay case study. *Remote Sensing of Environment* 109(4):464-472.

- Gómez JA, Alonso CA & García AA (2011) Remote sensing as a tool for monitoring water quality parameters for Mediterranean Lakes of European Union water framework directive (WFD) and as a system of surveillance of cyanobacterial harmful algae blooms (SCyanoHABs). *Environmental monitoring and assessment* 181(1-4):317-334.
- Gons HJ (1999) Optical Teledetection of Chlorophyll a in Turbid Inland Waters. *Environmental Science & Technology* 33(7):1127-1132.
- Gons HJ, Auer MT & Effler SW (2008) MERIS satellite chlorophyll mapping of oligotrophic and eutrophic waters in the Laurentian Great Lakes. *Remote Sensing of Environment* 112(11):4098-4106.
- Gons HJ, Rijkeboer M, Bagheri S & Ruddick KG (2000) Optical Teledetection of Chlorophyll a in Estuarine and Coastal Waters[†]. *Environmental Science & Technology* 34(24):5189-5192.
- Gordon HR (1978) Removal of atmospheric effects from satellite imagery of the oceans. *Appl. Opt.* 17(10):1631-1636.
- Gordon HR (1997) Atmospheric correction of ocean color imagery in the Earth Observing System era. *Journal of Geophysical Research: Atmospheres* 102(D14):17081-17106.
- Gordon HR & Wang M (1994) Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with SeaWiFS: a preliminary algorithm. *Appl. Opt.* 33(3):443-452.
- Govindje e (1995) Sixty-three Years Since Kautsky: Chlorophyll-a Fluorescence. *Functional Plant Biology* 22(2):131-160.
- Gower JFR (1980) Observations of in situ fluorescence of chlorophyll-a in Saanich Inlet. *Boundary-Layer Meteorol* 18(3):235-245.
- Gower JFR & Borstad GA (1990) Mapping of phytoplankton by solar-stimulated fluorescence using an imaging spectrometer. *International Journal of Remote Sensing* 11(2):313-320.
- Gower JFR & King S (2007) Validation of chlorophyll fluorescence derived from MERIS on the west coast of Canada. *International Journal of Remote Sensing* 28(3-4):625-635.
- Gregg WW & Casey NW (2004) *Global and regional evaluation of the SeaWiFS chlorophyll data set.* Elsevier, New York, NY, United States. 17 p
- GRIL (2007) Les cyanobactéries dans les lacs québécois : Un portrait de la situation selon les chercheurs du GRIL. (Groupe de recherche interuniversitaire en limnologie et en environnement aquatique), p 10.
- Ha N, Koike K & Nhuan M (2013) Improved Accuracy of Chlorophyll-a Concentration Estimates from MODIS Imagery Using a Two-Band Ratio Algorithm and Geostatistics: As Applied to the Monitoring of Eutrophication Processes over Tien Yen Bay (Northern Vietnam). *Remote Sensing* 6(1):421-442.

- Han L (1997) Spectral reflectance with varying suspended sediment concentrations in clear and algae-laden waters. *Photogrammetric Engineering and Remote Sensing* 63:701-705.
- Hodjaty MM, Saradjian MR & Jamshidi S (2010) Chlorophyll a Estimation in Sefidrud Estuarine Using MODIS Images. in *FIG Congress 2010*Sydney, Australia), p 11.
- Hu C (2009) A novel ocean color index to detect floating algae in the global oceans. *Remote Sensing of Environment* 113(10):2118-2129.
- Hu C, Chen Z, Clayton TD, Swarzenski P, Brock JC & MullerKarger FE (2004) Assessment of estuarine water-quality indicators using MODIS mediumresolution bands: Initial results from Tampa Bay. *Remote Sensing of Environment* 93(3):423-441.
- Hu C & He MX (2008) Origin and offshore extent of floating algae in Olympic sailing area. *Eos, Transactions American Geophysical Union* 89(33):302-303.
- Hu C, Lee Z, Ma R, Yu K, Li D & Shang S (2010) Moderate Resolution Imaging Spectroradiometer (MODIS) observations of cyanobacteria blooms in Taihu Lake, China. *Journal of Geophysical Research: Oceans* 115(C4):C04002.
- Huisman J, Matthijs HCP & Visser PM (Édit) (2005) *Harmful cyanobacteria.* Springer, Dordrecht, Netherlands, Vol 3.
- Hunter PD, Tyler AN, Gilvear DJ & Willby NJ (2009) Using Remote Sensing to Aid the Assessment of Human Health Risks from Blooms of Potentially Toxic Cyanobacteria. *Environmental Science & Technology* 43(7):2627-2633.
- Islam A, Gao J, Ahmad W, Neil D & Bell P (2004) A composite DOP approach to excluding bottom reflectance in mapping water parameters of shallow coastal zones from TM imagery. *Remote Sensing of Environment* 92(1):40-51.
- Jensen JR (2006) *Remote Sensing of the Environment: An Earth Resource Perspective (2nd Edition).* Prentice Hall
- Jiang Z, Huete AR, Youngwook K & Didan K (2007) 2-band enhanced vegetation index without a blue band and its application to AVHRR data. p 9.
- Joliot P & Joliot A (2005) Excitation transfer between photosynthetic units: the 1964 experiment. *Discoveries in Photosynthesis,* (Advances in Photosynthesis and Respiration, Govindjee, Beatty JT, Gest H & Allen J (Édit.) Springer Netherlands, Vol 20. p 187-191.
- Jones GJ & Orr PT (1994) Release and degradation of microcystin following algicide treatment of a Microcystis aeruginosa bloom in a recreational lake, as determined by HPLC and protein phosphatase inhibition assay. *Wat. Res.* 28 (4), :871-876.
- Kahru M, Leppanen JM & Rud O (1993) Cyanobacterial Blooms Cause Heating of the Sea-Surface. *Marine Ecology Progress Series* 101 1-7.

- Kahru M, Michell BG, Diaz A & Miura M (2004) MODIS detects a devastating algal bloom in Paracas Bay, Peru. *Eos, Transactions American Geophysical Union* 85 (45):465-472.
- Kiage L & Walker N (2009) Using NDVI from MODIS to Monitor Duckweed Bloom in Lake Maracaibo, Venezuela. *Water Resour Manage* 23(6):1125-1135.
- Krause P, Boyle D & Bäse F (2005) Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences* 5:89-97.
- Kuster T (2009) Passive optical remote sensing of cyanobacteria and other intense phytoplancton blooms in coastal and inland waters. *Journal of Remote Sensing* 30:4401-4425.
- Lavender SJ, Pinkerton MH, Moore GF, Aiken J & Blondeau-Patissier D (2005) Modification to the atmospheric correction of SeaWiFS ocean colour images over turbid waters. *Continental Shelf Research* 25(4):539-555.
- Lavoie I, Laurion I, Warren A & Vincent W (2007) Les fleurs d'eau de cyanobactéries: Revue de Littérature. (Institut National de le Recherche Scientifique - Centre Eau, Terre et Environnement, Québec), p 120 p
- Li R-H & Belford GG (2002) Instability of decision tree classification algorithms. in *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining* (ACM, Edmonton, Alberta, Canada), p 570-575.
- Luciani X, Pierre C, Frédéric T, David B, André B, Stéphane M & Claude J (2007) *Analyse numérique des spectres de 3D issus de mélanges non linéaires.* doctorate (Sud Toulon Var). 165 p(Consulté le 14 Decembre)
- Luo Y, Trishchenko AP & Khlopenkov KV (2008) Developing clear-sky, cloud and cloud shadow mask for producing clear-sky composites at 250-meter spatial resolution for the seven MODIS land bands over Canada and North America. *Remote Sensing of Environment* 112(12):4167-4185.
- Lyu H, Wang Q, Wu C, Zhu L, Yin B, Li Y & Huang J (2013) Retrieval of phycocyanin concentration from remote-sensing reflectance using a semi-analytic model in eutrophic lakes. *Ecological Informatics* 18(0):178-187.
- Mackie TN (2010) Public Health Surveillance of Toxic Cyanobacteria in Freshwater Systems Using Remote Detection Methods. <u>http://www.escholarship.org/uc/item/1w3686nh</u>.
- Matthews MW, Bernard S & Winter K (2010) Remote sensing of cyanobacteriadominant algal blooms and water quality parameters in Zeekoevlei, a small hypertrophic lake, using MERIS. *Remote Sensing of Environment* 114(9):2070-2087.
- MDDEP (2002) Bilan des lacs et cours d'eau touchés par les fleurs d'eau d'algues bleuvert au Québec. Ministère du Développement durable, de l'Environnement et des Parcs), <u>http://www.mddep.gouv.qc.ca/eau/algues-bv/gestion/index.htm#resultats</u>

- Mishra S, Mishra D & Schluchter W (2009) A Novel Algorithm for Predicting Phycocyanin Concentrations in Cyanobacteria: A Proximal Hyperspectral Remote Sensing Approach. *Remote Sensing* 1(4):758-775.
- Mittenzwey KH, Breitwieser S, Penig J, Gitelson AA, Dubovitzkii G, Garbusov G, Ullrich S, Vobach V & Müller A (1991) Fluorescence and Reflectance for the in-situ Determination of Some Quality Parameters of Surface Waters. *Acta hydrochimica et hydrobiologica* 19(1):1-15.
- Monestier T (2006) La problématique cyanobactéries sur le bassin versant de la Dordogne étude de cas : retenue de bort les orgues. (Université Blaise Pascal).
- Moran MS, Inoue Y & Barnes EM (1997) Opportunities and limitations for image-based remote sensing in precision crop management. *Remote Sensing of Environment* 61(3):319-346.
- Morel A & Bélanger S (2006) Improved detection of turbid waters from ocean color sensors information. *Remote Sensing of Environment* 102(3–4):237-249.
- Morel A & Gentili B (1991) Diffuse reflectance of oceanic waters: its dependence on Sun angle as influenced by the molecular scattering contribution. *Appl. Opt.* 30(30):4427-4438.
- Morel A & Gentili B (1993) Diffuse reflectance of oceanic waters. II. Bidirectional aspects. *Appl. Opt.* 32(33):6864-6879.
- Moses WJ, Gitelson AA, Berdnikov S & Povazhnyy V (2009) Estimation of chlorophyll- a concentration in case II waters using MODIS and MERIS data—successes and challenges. *Environmental Research Letters* 4(4):045005.
- Nash JE & Sutcliffe JV (1970) River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology* 10(3):282-290.
- Neville RA & Gower JFR (1977) Passive remote sensing of phytoplankton via chlorophyll α fluorescence. *Journal of Geophysical Research* 82(24):3487-3493.
- O'Reilly JE, Maritorena S, Mitchell BG, Siegel DA, Carder KL, Garver SA, Kahru M & McClain C (1998) Ocean color chlorophyll algorithms for SeaWiFS. *Journal of Geophysical Research: Oceans (1978–2012)* 103(C11):24937-24953.
- Ogashawara I, Mishra D, Mishra S, Curtarelli M & Stech J (2013) A Performance Review of Reflectance Based Algorithms for Predicting Phycocyanin Concentrations in Inland Waters. *Remote Sensing* 5(10):4774-4798.
- Ormeci C, Sertel E & Sarikaya O (2009) Determination of chlorophyll-a amount in Golden Horn, Istanbul, Turkey using IKONOS and in situ data. *Environmental Monitoring and Assessment* 155(1-4):83-90.
- Polikar R (2006) Ensemble based systems in decision making. *Circuits and Systems Magazine, IEEE* 6(3):21-45.
- Pouria S, de Andrade A, Barbosa J, Cavalcanti RL, Barreto VTS, Ward CJ, Preiser W, Poon GK, Neild GH & Codd GA (1998) Fatal microcystin intoxication in haemodialysis unit in Caruaru, Brazil. *The Lancet* 352(9121):21-26.

- Rahman H (2001) Influence of atmospheric correction on the estimation of biophysical parameters of crop canopy using satellite remote sensing. *International Journal of Remote Sensing* 22(7):1245-1268.
- Reinart A & Kuster T (2006) Comparison of different satellite sensors in detecting cyanobacterial bloom events in the Baltic Sea. *Remote Sensing of Environment* 102(1):74-85.
- Richter R (1996) A spatially adaptive fast atmospheric correction algorithm. *International Journal of Remote Sensing* 17(6):1201-1214.
- Ryan JP, Fischer AM, Kudela RM, Gower JFR, King SA, Marin lii R & Chavez FP (2009) Influences of upwelling and downwelling winds on red tide bloom dynamics in Monterey Bay, California. *Continental Shelf Research* 29(5–6):785-795.
- Schalles JF & Yacobi YZ (2000) Remote detection and seasonal patterns of phycocyanin, carotenoid and chlorophyll pigments in eutrophic waters. *Limnology and Lake Management* 55:153-168.
- Simis SGH, Peters SWM & Gons HJ (2005) Remote sensing of the cyanobacterial pigments phycocyanin in turbid inland water. *Limnology and Oceanography* 50:237-245.
- Simis SGH, Ruiz-Verdu A, Dominguez-Gomez JA, Pena-Martinez R, Peters SWM & Gons HJ (2007) Influence of Phyoplankton pigment composition on remote sensing of cyanobacterial biomass. *Remote Sensing of Environment* 106:414-427.
- Stewart I (2004) Recreational exposure to freshwater cyanobacteria: epidemiology, dermal toxicity and biological activity of cyanobacterial lipopolysaccharides. Thèse de Doctorat de Philosophie (Université de Queensland). 418 p
- Stumpf RP, Arnone RA, Gould RW, Martinolich V, Tester PA, Steward RG, Subramaniam A, Culver ME & Pennock JR (2000) SeaWiFS ocean colour data for US Southeast coastal waters. In *proceeding of 6th Internaional Conference on Remote Sensing for Marine and Coastal Environment.* in *Environmental Research Institute of Michigan*), p 25-27.
- Svrcek C & Smith DW (2004) Cyanobacteria toxins and the current state of knowledge on water treatment options: a review. *Environmental Engineering and Science* 3 : 3:155-185.
- Tarrant P & Neuer S (2009) Monitoring Algal Blooms in a Southwestern U.S. Reservoir System. *Eos, Transactions American Geophysical Union* 90(5):38-39.
- Timofeev R (2004) *Classification and regression trees (CART) theory and applications.* (Humboldt University, Berlin).
- Tørvi H & Hertzberg T (1997) Estimation of uncertainty in dynamic simulation results. *Computers & Chemical Engineering* 21, Supplement(0):S181-S185.

- Trishchenko AP, Luo Y & Khlopenkov KV (2006) A method for downscaling MODIS land channels to 250 m spatial resolution using adaptive regression and normalization. *Remote Sensing for Environmental Monitoring* 6366:36607-36607.
- Vermote EF, Tanre D, Deuze JL, Herman M & Morcette JJ (1997) Second Simulation of the Satellite Signal in the Solar Spectrum, 6S: an overview. *Geoscience and Remote Sensing, IEEE Transactions on* 35(3):675-686.
- Vincent RK, Qin X, McKay RML, Miner J, Czajkowski K, Savino J & Bridgeman T (2004) Phycocyanin detection from LANDSAT TM data for mapping cyanobacterial blooms in Lake Erie. *Remote Sensing of Environment* 89(3):381-392.
- Viollier M, Tanré D & Deschamps PY (1980) An algorithm for remote sensing of water color from space. *Boundary-Layer Meteorol* 18(3):247-267.
- Wheeler SM, Morrissey LA, Levine SN, Livingston GP & Vincent WF (2012) Mapping cyanobacterial blooms in Lake Champlain's Missisquoi Bay using QuickBird and MERIS satellite data. *Journal of Great Lakes Research* 38, Supplement 1(0):68-75.
- Yacobi YZ, Moses WJ, Kaganovsky S, Sulimani B, Leavitt BC & Gitelson AA (2011) NIR-red reflectance-based algorithms for chlorophyll-a estimation in mesotrophic inland and coastal waters: Lake Kinneret case study. Water research 45(7):2428-2436.
- Yentsch CS & Menzel DW (1963) A method for the determination of phytoplankton chlorophyll and phaeophytin by fluorescence. *Deep Sea Research and Oceanographic Abstracts.* Elsevier, p 221-231.
- Yu G, Yang W, Matsushita B, Li R, Oyama Y & Fukushima T (2014) Remote Estimation of Chlorophyll-a in Inland Waters by a NIR-Red-Based Algorithm: Validation in Asian Lakes. *Remote Sensing* 6(4):3492-3510.
- Zhang Y, Liu M, Qin B, van der Woerd HJ, Li J & Li Y (2009a) Modeling remote-sensing reflectance and retrieving chlorophyll-a concentration in extremely turbid case-2 waters (Lake Taihu, China). *Geoscience and Remote Sensing, IEEE Transactions on* 47(7):1937-1948.
- Zhang Y, Wang Y, Wang Y & Xi H (2009b) Investigating the Impacts of Landuselandcover (LULC) Change in the Pearl River Delta Region on Water Quality in the Pearl River Estuary and Hong Kong's Coast. *Remote Sensing* 1(4):1055-1064.

PARTIE 2 : ARTICLE-1

Avant-propos

Cet article a été publié dans le journal *Remote sensing* du *MDPI*. Les co-auteurs qui ont participés à la rédaction de ce manuscrit sont :

1. Anas El Alem :

Responsable de la conception de recherches, prétraitements et traitements des images, le développement des modèles mathématiques, l'analyse et l'interprétation des résultats et à la rédaction du manuscrit.

2. Karem Chokmani :

Responsable également de la conception de recherches et a contribué à l'édition, l'analyse, et à l'examen du manuscrit.

3. Laurion Isabelle :

Personne intermédiaire avec le MDDELCC pour l'octroi des données *in situ* et a contribuée à l'édition, l'analyse et à la révision du manuscrit.

4. Salah Eddine El Adlouni :

Responsable des interprétations mathématiques et statistiques et à la révision du manuscrit.

Abstract

The occurrence and extent of intense harmful algal blooms (HABs) have increased in inland waters during recent decades. Standard monitor networks, based on infrequent sampling from a few fixed observation stations, are not providing enough information on the extent and intensity of the blooms. Remote sensing has great potential to provide the spatial and temporal coverage needed. Several sensors have been designed to study water properties (AVHRR, SeaBAM, and SeaWIFS), but most lack adequate spatial resolution for monitoring algal blooms in small and medium-sized lakes. Over the last decade, satellite data with 250 m spatial resolution have become available with MODIS. In the present study, three models inspired by published approaches (Kahru, Gitelson, and Floating Algae Index (FAI)) and a new approach named APPEL (APProach by ELimination) were adapted to the specific conditions of southern Quebec and used to estimate chlorophyll-a concentration (Chl-a) using MODIS data. Calibration and validation were provided from in situ Chl-a measured in four lakes over 9 years (2000-2008) and concurrent MODIS imagery. MODIS bands 3 to 7, originally at 500 m spatial resolution, were downscaled to 250 m. The APPEL, FAI, and Kahru models yielded satisfactory results and enabled estimation of Chl-a for heavy blooming conditions (Chl-a > 50 mg·m⁻³), with coefficients of determination reaching 0.95, 0.94, and 0.93, respectively. The model inspired from Gitelson did not provide good estimations compared to the others ($R^2 = 0.77$). However, the performance of all models decreased when Chl-a was below 50 mg·m⁻³

Keywords: Remote sensing; MODIS; Chl-a; Lake; HABs; Quebec; Champlain.

Remote Sens. **2012**, *4*, 1-x manuscripts; doi:10.3390/rs40x000x



www.mdpi.com/journal/remotesensing

Article

Comparative Analysis of Four Models to Estimate Chlorophyll-a Concentration in Case-2 Waters Using MODerate Resolution Imaging Spectroradiometer (MODIS) Imagery

Anas El-Alem^{1,*}, Karem Chokmani¹, Isabelle Laurion¹ and Sallah E. El-Adlouni²

- ¹ Centre Eau Terre Environnement, INRS, 490 de la Couronne, Québec, QC G1K 9A9, Canada;
 - E-Mail: karem.chokmani@ete.inrs.ca (K.C.); isabelle.laurion@ete.inrs.ca (I.L.)
- ² Département de Mathématiques et de Statistique, Université de Moncton, 18 avenue Antonine-Maillet, Moncton, NB E1A 3E9, Canada; E-Mail: salaheddine.el.adlouni@umoncton.ca
- * Author to whom correspondence should be addressed; E-Mail: anas.el_alem @ete.inrs.ca; Tel.: +1-418-654-2570; Fax: +1-418-654-2600.

Received: 12 June 2012; in revised form: 27 July 2012 / Accepted: 2 August 2012 / Published: 10 August 2012

1 INTRODUCTION

Sustainable management of water resources became a real interest for the international community a few decades ago. Although Canada holds 9% of the renewable supply of freshwater for 0.5% of the worldwide population [1], this precious resource needs to be managed responsibly. Evidence that a lot of countries currently suffer from water penury cannot be denied [2]. Globally, the origin of the penury is attributed in part to the deterioration of water quality caused by the development of harmful algal blooms (HABs), such as cyanobacterial blooms. The increasing development of such HABs reflects the advanced state of aquatic ecosystem eutrophication caused by urban, agricultural, and industrial developments. Once established in lakes, cyanobacterial populations are extremely difficult to control except through the long-term reduction of nutrient inputs from the watershed and from internal sources.

Currently, the most common way to detect the presence of blooms is *in situ* sampling, generally through riverine claims [3]. However, neither the spatial distribution nor the temporal frequency of such monitoring programs is sufficient to detect the spatial extent, intensity, and duration of HABs in water bodies, especially during periods of heavy blooms [4], and for a country like Canada that covers a very large area with a great number of lakes. *In situ* sampling and sample analyses are laborious, and the results obtained are often difficult to interpret because of HAB heterogeneous spatial distribution and sudden rise and fall [5]. In addition, standard *in situ* sampling and sample analyses are expensive. More than \$650,000 CAN was spent to collect and analyze *in situ* samples from 150 water bodies in the province of Quebec alone during 2009 [6].

Remote sensing, with its synoptic viewing, consistent recurrence and capacity to provide information over a wide range of wavelengths represents an attractive alternative method to monitor algal blooms. Satellite sensors that provide data in visible and near infrared (NIR) wavelengths can be used to estimate chlorophyll-a concentration (Chl-a) based on its high absorption of the blue and red part of the electromagnetic spectrum, and its high reflectance of the green and NIR wavebands.

109

Thus, bio-optical models that relate the apparent optical properties of water bodies to their inherent optical properties can be used to estimate Chl-a [7].

To date, however, the use of satellite data has been limited to open ocean waters classified as Case-1 waters, in which phytoplankton and co-varying material of biological origin are the principal constituents responsible for variations in ocean inherent optical properties [8] and algal pigments are often the only component optically active in the water [9]. Remote sensing data have been efficiently used to detect marine algal blooms in recent decades [10-14]. The basic approach has been to model the apparent optical properties of oceans from remote sensing data as a function of Chl-a alone [15]. However, while the estimation of Chl-a is a relatively easy task for Case-1 water, monitoring algal blooms in lakes and coastal water classified as Case-2 waters is a challenging one. The remote sensing reflectance (R) spectra of these water bodies is generally complex because of the presence of many optically active components, such as chlorophylls and carotenoids from phytoplankton and aquatic plants, total suspended solids (TSS), colored dissolved organic matter (CDOM), and sometimes bottom reflectance [16]. Generally, the algorithms developed to detect blooms based on the absorption properties of a single variable absorbing near 700 nm (typical of Chl-a) are inadequate for this type of waters [15]. Furthermore, the older generation of sensors dedicated to the study of water properties lack sufficient spatial resolution to assess Chl-a in Case-2 waters. Nevertheless, recently, sensors such as the MODerate Resolution Imaging Spectroradiometer (MODIS) and the MEdium Resolution Imaging Spectrometer (MERIS), with their full spatial resolution of 250 m and 300 m respectively, are increasingly used to achieve this goal. This makes the two sensors very concurrent in such applications. Also, many other sensors were used successfully, such as Landsat TM to retrieve Chl-a and suspended solid concentrations in lake Kasumigaura using neural network technique [9] and to develop an algorithm for assessing phycocyanin concentrations in lake Erie in order to improve the understanding of temporal and spatial dynamics of cyanobacterial blooms. SeaWiFS data were also used to establish the seasonal distribution pattern and intensity of phytoplankton and terrigenous input [17], and AHVRR to point out the behavior of main taxonomic groups of Baikal lake phytoplankton in relation to ice conditions [18], or to

110

assess other water quality parameters [19,20]. Recently, QuickBird and MERIS data were also used to assess cyanobacterial blooms based on their specific pigment (phycocyanin) in Lake Champlain [21]. Thus, many semi-analytical algorithms [12,22–26] and indexes [13] are now available in the literature to retrieve the Chl-a concentration of inland water bodies.

For this study, because of the large number of days with cloud cover over the selected lakes and the relatively low number of *in situ* measurements, we chose to use MODIS (revisit time of 1 day) rather than MERIS (revisit time of 3 days) in order to collect enough data for model calibration and validation. The first two of the 36 MODIS bands in the red/NIR have 250 m spatial resolution. The rest of the visible and infrared (IR) bands which are the most appropriate for the detection of Chl-a, CDOM and suspended particles have resolutions larger than 250 m. However, the spatial resolution of MODIS bands 3 to 7 can be downscaled from 500 to 250 m spatial resolution using the approach of Trishchenko et al. [27]. In this manner, the first seven bands of MODIS sensor become available at 250 m spatial resolution, covering the visible, NIR, and IR parts of the spectrum. It is important to notice that the bands originally at 250 m spatial resolution and those downscaled were originally designed for aerosol, cloud and land applications. The MODIS bands designed for surface water studies are at 1 km spatial resolution, which is not optimal for monitoring algal bloom in lakes, especially in small and medium-sized ones. Thus, the challenge is to demonstrate that the first seven bands of the MODIS sensor can be used to estimate Chl-a in water bodies after spatial resolution enhancement.

The purpose of this research is to evaluate the performance of the three bio-optical models most commonly used in the literature (Kahru, Gitelson, and FAI (floating algae index)) and of a novel approach named APPEL (APProach by ELimination) to estimate ChI-a in a series of lakes in the southern part of Quebec province. The assumption that MODIS bands, not originally developed for water applications, can be used to detect algal blooms in small and medium-sized inland water bodies was also tested.

2 MATERIAL AND METHODS

2.1 In situ Data

Four lakes were considered in this study: Missisquoi Bay of Lake Champlain, Lake Brome, Lake William, and Lake Nairne, with surfaces of 45, 41, 4, and 2 km², respectively (Figure 2-1). These lakes are part of a program to monitor water bodies with recurrent cyanobacterial blooms carried out by the Ministère du Développement Durable de l'Environnement et des Parcs (MDDEP). The data were collected over a period of 9 years (2000 to 2008) at several sites on the four lakes (Figure 2-2) and included cyanotoxins (equivalent total microcystin-LR and anatoxin-a), algal pigments (Chl-a, pheophytine-a), phytoplankton abundance, and physicochemical characteristics of lake water (total phosphorous, total nitrogen, pH, turbidity, Secchi depth, surface temperature).



Figure 2-1. Geographic disposition of the studied lakes.

Chl-a was used as a proxy of the total algal biomass and potential indicator of HABs. A total of 168 Chl-a *in situ* measurements and a corresponding 62 MODIS images were obtained. Of these, 51 values were used for the model calibration/validation, corresponding to a sub-group of 22 images. The rest of the data set was unexploited due to the presence of clouds and/or cloud shadows on the measurement sites or to the bad quality of the images. The largest sample size was in Missisquoi Bay (N = 37) and the smallest was in Lake Brome (N = 6). The concentration of Chl-a varied between 2.5 mg·m⁻³ (Lake Nairne, 2007) and 91,000 mg·m⁻³ (Missisquoi Bay of Champlain Lake, 2003).

A second data set collected by the MDDEP over 5 years (2004 to 2008) was used for field truth validation. These data contained two kinds of observations: (1) field maps where the spatial distribution of algal blooms is delimited, and (2) records of observations made by technicians from regional directorates of the MDDEP as to whether blooms were present on a lake or not. It is important to note that no laboratory analysis was made for this data set. We used this latter to evaluate the ability of the models to detect the presence of algal blooms by comparing *in situ* observations to the corresponding dates of MODIS images.



Figure 2-2. Water sampling stations on the four studied lakes.

2.2 MODIS Data

The remote sensing data used were acquired by the MODIS sensor aboard the TERRA platform of NASA's Earth Observation System (EOS). The MODIS Level 1B product is a set of geo-located and calibrated data. The MODIS sensor operates across a very wide spectrum, with 36 bands that cover the region from 0.4 to 14.4 μ m, observing with a spatial resolution varying from 250 m to 1 km. Sensor data are available from the website (http://ladsweb.nascom.nasa.gov/data/search.html) in HDF format. The satellite passes over the study site between 2:00 and 4:30 pm GMT on an ascending orbit. For this study, images from 2000 to 2008 were downloaded and pre-processed. Only the first seven MODIS bands were utilized, the first two being available at 250 m spatial resolution, and bands 3 to 7 available at 500 m but downscaled to 250 m. Tableau 2-1 summarizes the characteristics of the bands used in this study.

Two pre-processing steps were performed: spatial resolution downscaling and atmospheric correction. Downscaling entails three phases. <u>Phase 1</u>: Downscaling of the MODIS/Terra bands 3-7 at 250 m spatial resolution using adaptive regression and radiometric normalization, as described in Trishchenko et al. [27]. Phase 2: Improvement of the spatial resolution by re-projecting the images from the Sinusoidal to the Lambert conformal conic (LCC) projection. Originally, the L1B products of MODIS data are projected on a sinusoidal system, introducing a substantial distortion of areas distant from the central meridian. However, the LCC superimposes a cone over the Earth sphere with two reference parallels secant to the globe and intersecting it, which minimizes distortion when projecting a three-dimensional surface to a two-dimensional surface, and enhances the quality of the spatial resolution especially in North-American and polar zones. *Phase 3:* Production of clear sky, cloud, and cloud shadow masks by applying a scene identification algorithm. The algorithm uses an exclusive criteria sequence. Bare grounds are identified and excluded, based on a high contrast of reflectance between the band-1 and band-6. Thereafter, snow and ice pixels are excluded, followed by water pixels and cloud pixels. After excluding all classes above, the remaining pixels likely belong to cloud free land class.

Recent works clearly show that atmospheric correction found to be significant in water applications, especially in NIR part of the spectrum [28–30]. Several atmospheric correction models are available and can be divided into two categories: deterministic such as MODTRAN [31], EXACT [32], 6S [33] or SMAC [34], and empirical like empirical line method [35] and the dark pixel or histogram method [36]. However, it has been shown that the most effective atmospheric correction algorithms are the simplest [30]. Thereby in our work, the MODIS data were corrected using the SMAC model (Simplified Method for Atmospheric Correction), a simplified and operational version of the 5S codes (Simulation of the Satellite Signal in the Solar Spectrum), because it can perform an atmospheric correction within a short time [34]. All pre-processing steps (downscaling, re-projection, production of a clear-sky mask and atmospheric correction) were performed using an automatic tool developed by the Canadian Centre for Remote Sensing [37].

Tableau 2	2-1	Characteristics	of	the	MODerate	Resolution	Imaging	Spectroradiometer	(MODIS)
		bands used in	the	pre	sent study.				

Primary Use	Ban	Bandwidth	Spatial Resolution
-	d	(nm)	(m)
Land/Cloud/Aeroso	1	620–670	250
Is Boundaries	2	841–876	250
	3	459–479	500
Land/Cloud/Aarooa	4	545–565	500
Lanu/Cloud/Aeroso	5	1,230–1,250	500
is properties	6	1,628–1,652	500
	7	2,105–2,155	500

3 ALGORITHMS USED TO ESTIMATED CHL-A CONCENTRATION

Three semi-empirical algorithms (Kahru, Gitelson, and FAI (floating algae index)) and a novel approach named APPEL (APProach by ELimination) developed at the remote sensing laboratory of the Institut National de la Recherche Scientifique (INRS), Quebec, were selected for this study. Many Case-2 water models designed to estimate Chl-a using remote sensing data are already in use [11,38,39], but most rely on the typical signal of Chl-a near 700 nm, which is not covered by the first seven bands of the

MODIS sensor (Tableau 2-1). The MODIS bands originally designed for ocean color studies, which do cover 700 nm, are at 1 km spatial resolution, which is not suitable for small and medium-sized water bodies. The purpose of the present study is to assess the potential of the first seven MODIS bands to estimate Chl-a at 250 m spatial resolution. These bands do not cover the frequency of maximal Chl-a absorption, but their better spatial resolution might compensate for their lack of spectral resolution. This assumption is supported by several studies [40–42]. Although originally designed for land studies and cloud detection, by comparison to sensors such as Landsat-7 ETM+, CZCS (Coastal Zone Color Scanner) and SeaWiFS, the MODIS 250 m and 500 m bands have potential for water applications as they are 4–5 times more sensitive than L7/ETM+ bands, nearly twice as sensitive as the corresponding CZCS blue-green bands, and 3–4 times more sensitive than the red and NIR ones, while they are only 1-2 times (500 m, blue-green) less sensitive than the corresponding SeaWiFS bands [19].

3.1 Kahru's Model

Most Chl-a estimation models have been based on the ultraviolet (UV) and the visible part of the spectrum [10,43–45]. This was not problematic for Case-1 waters, but there are many problems associated with the use of algorithms based on this part of the spectrum for Case-2 waters, especially in the UV portion, where several complications arise in relation to atmospheric correction [46] and the overestimation of Chl-a due to the presence of CDOM and TSS [47].

The semi-analytical approach of Kahru uses the red/NIR part of the spectrum [13]. MODIS band-1 has been shown to be sensitive to changes in concentrations of cyanobacteria grown in cultures and has therefore been suggested for the qualitative mapping of cyanobacterial blooms [48], while MODIS band-2 has been successfully used to quantify algal blooms in the Baltic Sea [49]. Kahru *et al.* [26] suggested using MODIS band 1, 3, and 4 as true-color images to detect the presence of blooms, and as a quantitative quantification, they proposed using the difference between band-2 and band-1 as a water surface turbidity index. Turbidity is not specific to algal blooms (e.g., suspended solids can result from erosion within the drainage basin), but when mainly

caused by phytoplankton biomass, turbid features appear in blue-green in the true-color image. Thus, in this case the turbidity index can be used as an estimator of Chl-a. This approach has been successfully used to detect an algal bloom in Paracas Bay in Peru [26] and to monitor water quality in the southwestern US reservoir system [14].

3.2 Floating algae index (FAI)

Chuanmin and Ming-Xia developed an ocean color index to detect floating algae based on red, NIR, and shortwave infrared (SWIR [50]). This part of the spectrum was used because water strongly absorbs these wavelengths. Thus, water appears opaque (black) even in the most turbid environments. Floating algae on the water surface, on the other hand, have higher reflectance in the NIR than other wavelengths and thus can be easily distinguished from the surrounding clear waters [13].

The FAI avoids the problems related to ocean indexes such as NDVI (normalized difference vegetation index) and EVI (enhanced vegetation index) when applied to the detection of floating algal blooms. The latter two indexes have been useful for delineating floating algae from adjacent water, but the pixel values of floating algae and adjacent waters are sensitive to variable environmental and observing conditions such as aerosols and solar/viewing geometry. These variable conditions create difficulties in visualization and quantification since they affect the visual contrast between floating algae and adjacent waters in addition to change their absolute values [13]. Thus, the FAI is defined as the difference between reflectance in the NIR (R_{NIR}) and a corrected reflectance of this band by the red and the SWIR bands (R'_{NIR}):

$$FAI = R_{PIR} - R'_{PIR}$$
(3-1)

where

$$R'_{PIR} = R_{R} - [R_{CLIR} - R_{R}] \times \left[\frac{\lambda_{PIR} - \lambda_{R}}{\lambda_{CLIR} - \lambda_{R}}\right]$$
(3-2)

 R'_{PIR} is the baseline reflectance in the NIR band derived from linear interpolation between the red and SWIR bands. For MODIS, $\lambda_R = 645 \text{ nm}$, $\lambda_{NIR} = 859 \text{ nm}$, and $\lambda_{SWIR} = 1,240 \text{ nm}$ [13]. The FAI is then calculated as follows:

$$FAI = R(\lambda_{b2}) - [R(\lambda_{b1}) + (R(\lambda_{b5}) - R(\lambda_{b1})) \times (\lambda_{b2} - \lambda_{b1}) / (\lambda_{b5} - \lambda_{b1})]$$
(3-3)

The FAI has been successfully used to detect an extensive bloom of a floating green macroalgae, *Enteromorpha prolifera*, in open ocean near Qingdao in China under a range of atmospheric environments (clear, hazy, and sunlight conditions [50]) and to reconstruct the temporal frequency and spatial distribution of cyanobacterial blooms over 9 years (2000-2008) in Taihu Lake in China [51] using MODIS data.

3.3 Gitelson Model

The principle of the Gitelson model is based on the relation between the detected reflectance and the total absorption and backscattering coefficients [52].

$$R_{rs}(\lambda) \propto \gamma \, \frac{b_b(\lambda)}{a_{tot}(\lambda) + b_b(\lambda)} \tag{3-4}$$

where γ depends of the radiation emerging from water, b_b is the backscattering coefficient, and a_{tot} (λ) is the sum of the absorption coefficients of phytoplankton cells (a_{pigm}), CDOM (a_{CDOM}), non-algal particles or TSS ($a_{tripton}$), and pure water (a_w) as shown in the following equation:

$$a_{tot} (\lambda) = a_{pigm}(\lambda) + a_{CDOM}(\lambda) + a_{tripton}(\lambda) + a_{w}(\lambda)$$
(3-5)

This semi-analytical model uses three bands to estimate Chl-a that relates pigment concentration to reflectance, $R(\lambda_i)$ [53], defined as:

$$[Chl - a] \propto [R^{-1}(\lambda_1) - R^{-1}(\lambda_2)] \times R(\lambda_3)$$
(3-6)

The reflectance reciprocal of the first band, $R^{-1}(\lambda_1) \propto \frac{a_{tot}(\lambda_1) + b_b}{b_b}$, must be strongly sensitive to Chl-a absorption (λ_1 falls within the 660-690 nm range) [54]. However,

 $R^{-1}(\lambda_1)$ is also affected by TSS, CDOM and pure water absorption, and backscatter from particulate matter. This effect (denominator of Equation (3-6)) can be minimized using a second spectral band, where $R^{-1}(\lambda_2)$ is the least sensitive to Chl-a absorption and the nearest to λ_1 . It has been demonstrated that λ_2 must belong to the 710-730 nm range [12,54]. The difference between $R^{-1}(\lambda_1)$ and $R^{-1}(\lambda_2)$ thus eliminates the effects of TSS and CDOM, but equation (3-6) is still affected by backscattering. A third band, λ_3 , is therefore used to reduce the influence of b_b . This band must be the least affected by pigments, TSS and CDOM, and the total absorption is a measure of pure water. The NIR range corresponds to this condition $(a(\lambda_3) \sim a_w)$ [55,56]. The output of these bands combination is mostly dependent on Chl-a. Thus, the model takes the following form:

$$[R^{-1}(\lambda_1) - R^{-1}(\lambda_2)] \times R(\lambda_3) \propto a_{chla}(\lambda_1)$$
(3-7)

The two-band combination $R^{-1}(\lambda_1) \times R(\lambda_3) \propto a_{chla}(\lambda_1)$ is a special case of this conceptual model, where $a_{pigm}(\lambda_1) \gg b_b$ and $a_{pigm}(\lambda_1) \gg a_{tripton}(\lambda_1) + a_{CDOM}(\lambda_1)$ [22]. To test the performance of this approach, the Gitelson model was used to estimate Chl-a of Chesapeake estuary [12].

However, the λ_1 and λ_2 wavelengths used in this approach do not fall within the coverage of the first seven MODIS bands. Since λ_1 must strongly respond to Chl-a, it should belong to the blue or red parts of the spectrum. Due to the aforementioned problems when using the blue region of the spectrum, we chose MODIS band-1 as the band of maximum Chl-a absorption. Regarding λ_2 , used to minimize the effect of TSS and CDOM, Wei et al. [57] demonstrated that MODIS band-3 could explain up to 88% of the variance in turbidity of Poyang Lake in China [57] (turbidity being an index of TSS). In addition, Chuanmin *et al.* [19] found that bands 1, 3, and 4 have the potential to detect CDOM in turbid water, as they were able to explain between 59 and 80% of the variance of CDOM concentration. In the present study we use a combination of MODIS bands 1, 3, and 4 as a substitute for the wavelength of λ_2 used in Gitelson's model:

$$\lambda_2$$
(substuted) $\propto (R_{b3} + \frac{\Delta R_{b3}}{\Delta R_{b4}})$ (3-8)

where:

$$\Delta R_{b3} = R_{b3} - R_{b1} \tag{3-9}$$

$$\Delta R_{b4} = R_{b4} - R_{b1} \tag{3-10}$$

In this equation, R_{b3} is used to minimize the effect of TSS, and the ratio $\left(\frac{\Delta R_{b3}}{\Delta R_{b4}}\right)$ is used to minimize the effect of CDOM. The third band, λ_3 should fall in the NIR according to Gitelson. MODIS band-2 was chosen for this purpose. Therefore, the final new formulation of the model inspired from the Gitelson approach was as follows:

$$[Chl - a] \propto \left[R^{-1}(\lambda_{b1}) - R^{-1} \left(\lambda_{b3} + \frac{\lambda_{\Delta b3}}{\lambda_{\Delta b4}} \right) \right] \times R(\lambda_{b2})$$
(3-11)

3.4 Novel Approach APPEL (APProach by ELimination)

The principle of APPEL is similar to that of the Gitelson approach. All of the MODIS bands capture signals responding to Chl-a, CDOM, TSS, and particle backscattering (b_b) from water bodies, but they do so in varying proportions according to the wavelength range of each band. The maximum reflectance of phytoplankton cells is in the NIR, where water has its minimum reflectance (Figure 3-1). Therefore, the maximum information regarding Chl-a can be derived from MODIS band-2. However, this band is affected by CDOM, TSS, and particle backscattering, as expressed by the following relation:

$$R(\lambda_{b2}) \alpha [Chl - a] + [CDOM] + [TSS] + b_b$$
(3-12)

To minimize the effect of CDOM, we have to use a band that is highly sensitive to the presence of organic matter and captures a maximum of reflection. Luciani et al. [58] have shown that organic matter has maximal reflection in the region of 400-470 nm [58]. This range corresponds to MODIS band-3. However, since band-3 is also affected by Chl-a and b_b , a second band that is highly sensitive to the presence of Chl-a is also required. MODIS band-2 achieves this purpose, as the maximum reflectance of Chl-a is

in the NIR. The same band is also used to reduce the effect of b_b from the water, since water strongly absorbs in the NIR; Gitelson *et al.* [54] also used this region to minimize the effect of b_b [55,56].

To minimize the effect of TSS, we proceed the same way as above. For this purpose, a third band is used, where the maximum of reflectance captured by the band is due to the presence of TSS. Hu *et al.* [19] have demonstrated that MODIS band-1 is highly correlated to the presence of TSS [19]. As above, we use MODIS band-2 to minimize the effect of Chl-a on MODIS band-1; no minimization of particle backscattering is needed because MODIS band-1 is in the red part of the spectrum, where water absorbs strongly. Thereby, particle backscattering is already feeble.

Equation (3-12) is then a function of Chl-a only, and the APPEL Model can be written as follows:

$$APPEL = R(\lambda_{b2}) - \left[\left(R(\lambda_{b3}) - R(\lambda_{b2}) \right) \times R(\lambda_{b2}) + \left(R(\lambda_{b1}) - R(\lambda_{b2}) \right) \right]$$
(3-13)



Figure 3-1. Simplified reflectance spectra of water and phytoplankton (obtained from http://www.aa-r-s.org/acrs/proceeding/ACRS1994/Papers/AGS94-5.htm).

4 ACCURACY ASSESSMENT

Four indexes are used to evaluate the performance of the models: coefficient of determination (R²), relative bias (BIASr), relative root mean square error (RMSEr), and relative Nash-Sutcliffe efficiency (Nr). The latter evaluates model performance by comparing the accuracy of estimated values to the observed mean value of the entire data set. If the result is negative, the estimate is worse than using the mean value; the criterion is equal to one for a perfect estimation. We chose to use the relative form of the indexes in order to reduce the contrast between the observed and estimated high values and the lower ones. Because the absolute forms of the indexes quantify absolute differences, over- or under-estimates of high values have, in general, greater influence than those of lower values. By using the relative forms of the indexes, we reduce, for example, the influence of the absolute differences recorded during periods of no blooming are enhanced [59]. The equations of the indexes are as follows:

$$R^{2} = \left[\frac{\sum_{i=1}^{n} (0_{i} - \overline{0})(Es - \overline{Es})}{\sqrt{\sum_{i=1}^{n} (0_{i} - \overline{0})^{2}} \sqrt{\sum_{i=1}^{n} (Es_{i} - \overline{Es})^{2}}}\right]^{2}$$
(4-1)

$$BIASr = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{Es_i - O_i}{Es_i} \right)$$
(4-2)

RMSEr =
$$\sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(\frac{Es_i - O_i}{Es_i}\right)^2}$$
 (4-3)

$$Nr = 1 - \frac{\sum_{i=1}^{n} \left(\frac{O_i - Es_i}{O_i}\right)^2}{\sum_{i=1}^{n} \left(\frac{O_i - \overline{O}}{\overline{O}}\right)^2}$$
(4-4)

where n is the sample size, O and Es are the observed and estimated values, and \overline{O} and \overline{Es} are the means of the observed and estimated values.
5 RESULTS AND DISCUSSION

5.1 Evaluation of Downscaling the MODIS Band Signal

MODIS bands 3 to 7, originally at 500 m spatial resolution, were downscaled to 250 m. To ensure that the downscaling process did not affect the signal of the original bands, we compared the spectral index of the downscaled images to that of the originals. Figure 5-1 shows the correlation between the MODIS signal sampled from areas on lakes captured on three images from three different years (2001-2003). The y-axis corresponds to the spectral index of the downscaled image, and the x-axis to the spectral index of the original image (500 m spatial resolution) re-sampled to 250 m using a method that does not enhance the resolution of the image but simply divides the surface of each pixel into four (from 500×500 m to 250×250 m).



Figure 5-1. Correlation between the original and the downscaled reflectance used in the estimation models of chlorophyll-a concentration. (A) Band 3, (B) Band 4, and (C) Band 5.

The results show a significant correlation between the original and downscaled signals (R^2 varies from 0.94 to 0.98). However, for a proper model assessment, the gradient "a" (the slope of the regression equation Y=a*X+b) should be used as a weighting index:

$$wr^{2} = \begin{cases} |a| * r^{2} \text{ for } a \leq 1, \\ |a|^{-1} * r^{2} \text{ for } a > 1 \end{cases}$$
(5-1)

In this manner, under- or over-estimations are also taken into account [59]. According to Equation (5-1), the weighted correlation coefficients are respectively 0.88, 0.77, and 0.82 for bands 3, 4, and 5. This means that the downscaled images lose from 12 to 23% of their original signal quality, which is acceptable since the MODIS band signal is 4 to 5 times higher than the corresponding bands of other sensors such as Land-Sat7/ETM or CZCS [19].

5.2 Model Calibration

As depicted in Figure 5-2, all models show a good correlation between Chl-a and the spectral index derived from MODIS data by each model, except for the Gitelson-inspired model, which does not perform as well. However, the models produce only a small error at high Chl-a values, with most of the error coming from low concentrations.

The World Health Organization (WHO) has established an alert threshold for water utilization of 10 mg Chl-a m⁻³ [60]. Water bodies with Chl-a lower than this threshold are not considered dangerous for recreational activities. Thus, to minimize a potential error related to high *in situ* data dispersion at low Chl-a, samples with *in situ* concentrations below 10 mg·m⁻³ were removed from the data set to test the accuracy of each model. As anticipated, the accuracy of the recalibrated models was improved (Figure 5-3). Therefore, most of the error was produced from low Chl-a values. This is not surprising, since most of the studied models (APPEL, Kahru, and FAI) are based on the red/NIR part of the spectrum where water has very low reflection. This means that water appears transparent when biomass is low, and the reflected signal is very weak (Figure 5-1), which increases the noise/signal ratio of the bands precluding a good correlation between Chl-a and the spectral index. Also, at low Chl-a the algal community is more

likely to be composed of a larger diversity of species [61], generating a larger diversity of optical signatures obtained by the satellite. Other considerations may also help to explain the high dispersion observed at low Chl-a. (1) The MODIS sensor bands cover a relatively broad range of the electromagnetic spectrum, possibly lowering their sensitivity at low Chl-a. (2) The atmospheric correction (SMAC) can produce additional noise; this would be especially apparent at low Chl-a. (3) The downscaling procedure may have caused a loss of signal quality. An additional source of noise can also be associated to *in situ* sampling of Chl-a, as positioning the sampling sites on the satellite images was difficult. The MDDEP technicians did not have GPS to locate their positions, and thus the sampling locations recorded were approximate. To overcome this problem, we averaged the satellite spectral index from four pixels around each *in situ* sampling site during the calibration step.

5.3 Model Validation

To evaluate the accuracy of the four models studied, a *cross-validation* technique was applied. This technique involves temporarily removing a given value of Chl-a from the sampling data set and using the remaining observations as a calibration group to estimate the withdrawn value. This operation is repeated for the whole data set. The performance of each model was assessed using the performance indexes presented above.

All models except Gitelson-inspired model produced satisfactory results (Tableau 5-1). The Nash criterion values of the APPEL, Kahru, and FAI models were quite similar and indicated very good correspondence between observed and estimated ChI-a (i.e., close to 1), with the APPEL and Kahru models performing slightly better than the FAI. The Nash criterion value of the Gitelson-inspired model was much lower (Nr = 0.34). RMSEr results showed that the ChI-a estimates calculated with the APPEL model were more precise than those of the other models (RMSEr of APPEL < Kahru < FAI < Gitelson). However, the error in ChI-a estimates was relatively high for all models, since the best performance produced a RMSEr of 69%. Moreover, relative BIAS results were negative for all of the models, indicating that all tended to under-estimate ChI-a. Again, the

APPEL model showed the best performance (-17%), while the other models ranged from -23% to -87%.

Indexes for Model Evaluation	APPEL	Kahru	FAI	Gitelson
R ²	0.93	0.91	0.89	0.79
Nr	0.97	0.95	0.91	0.34
RMESr (%)	69	83	116	307
BIASr (%)	–17	-23	-32	-87

Tableau 5-1. Model performance using cross-validation results from the whole data series.

The poor performance of the Gitelson-inspired model can be explained by the inability of the available MODIS bands to optimally respond to Chl-a absorption. The Chl-a pigment has two large absorption peaks (absorption maxima around 430 and 680 nm), with an optimized range of absorption between 660 and 690 nm. The MODIS band-1 used in the Gitelson-inspired model is indeed sensitive to a region where this pigment absorbs, but it is not centered on its maximal absorption. In addition, λ_2 , used to minimize the effect of TSS and CDOM, utilizes bands 1, 3, and 4 in this model, which are also sensitive to the presence of Chl-a. Consequently, the λ_1 wavelength shift and the λ_2 band substitution may be the cause of the lower performance of this model. APPEL, Kahru, and FAI, on the other hand, based on the contrast between the water and Chl-a reflectance in the red/NIR part of the spectrum. In this spectral region, pixels of Chl-a are easily distinguished from water pixels. Thus, for this study, the models based on reflectance (APPEL, Kahru, and FAI) to estimate Chl-a were more accurate than the Gitelson-inspired model based on absorption. Figure 5-4 shows the Chl-a estimated by the four models as a function of the observed Chl-a. This figure shows the decrease in the dispersion of estimated values as Chl-a increases, except for the Gitelson-inspired model, where the dispersion is large for most estimated concentrations.



Figure 5-2. Exponential regression obtained between chlorophyll-a concentration (whole data set) and the outputs of the MODIS reflectance bands for the four models (y-axis of inserts is on a logarithmic scale).



Figure 5-3. Exponential regression obtained between chlorophyll-a concentration (> 10 mg·m⁻³) and the outputs of the MODIS reflectance bands for the four models (y-axis of inserts is on a logarithmic scale).

Article-1



Figure 5-4. Chlorophyll-a concentration estimated from the four models compared to *in situ* measurements (data set Chl-a > 10 mg·m⁻³).



Figure 5-5. Chlorophyll-a concentration and cyanobacterial density used in decision tree published by the World Health Organization.

WHO has fixed a second threshold (Chl-a > 50 mg·m⁻³), with stricter recommendations than the earlier-mentioned threshold for recreational water use (10 mg·m⁻³), at which water utilization should be avoided (Figure 5-5). The performance of the models was specifically assessed for values at or above these concentrations. Thereby, the *cross-validation* results were divided into two groups: the first contains Chl-a between 10 and 50 mg·m⁻³ (N = 24), and the second contains the remaining values (above 50 mg·m⁻³; N = 12). The resulting evaluation indexes can be used to determine the behavior of the models in both heavy and low-moderate blooming conditions.

Tableau 5-2. Model performance using *cross-validation* results for the data subset Chl-a > 50 $\text{mg}\cdot\text{m}^{-3}$.

Indexes for Model Evaluation	APPEL	Kahru	FAI	Gitelson
R ²	0.95	0.93	0.94	0.77
Nr	0.927	0.88	0.934	0.30
RMESr (%)	61	80	58	189
BIASr (%)	1	-2	4	-28

Tableau 5-3. Model performance using *cross-validation* results for the data subset when Chl-a is between 10 and 50 mg⋅m⁻³.

Indexes for Model Evaluation	APPEL	Kahru	FAI	Gitelson
R ²	0.30	0.18	0.15	0.11
Nr	0.62	0.49	0.34	-7.87
RMSEr (%)	71	82	132	341
BIASr (%)	- 23	-30	-47	-110

Tables (5-2 and 5-3) summarize the performance of the models at Chl-a values above 50 mg·m⁻³ and between 10 and 50 mg·m⁻³, respectively. In general, the performance indexes calculated for the high concentrations (Chl-a > 50 mg·m⁻³) show better results than those calculated for the models calibrated with all Chl-a > 10 mg·m⁻³, except for the Nash criterion, which shows a slight decrease in the precision of models at the high concentrations. This may be due the smaller size of sample used in this step (N = 12).

Figure 5-6 and 5-7 show the scatter plots of the observed and estimated (*cross-validation*) Chl-a values greater than 50 mg·m⁻³ and between 10 and 50 mg·m⁻³, respectively. It is clear that the correlation is much more significant in heavy bloom conditions (Figure 5-6). The improvement in performance is not large for APPEL, Kahru, and the Gitelson-inspired model (Tableau 5-2), but for the FAI model, performance improves a lot when using only high Chl-a. The RMSEr changes from 119% to 58%, the BIASr from -32% to 4%, the R² from 0.89 to 0.94, and the Nr from 0.91 to 0.93. This is not surprising since the accumulated cells form scums at the surface of the water, which will improve the accuracy of this model originally developed to monitor floating algae. At these concentrations, FAI performs as well as APPEL, slightly overestimating Chl-a by 1% and 4%, respectively, while the Kahru model under-estimates Chl-a by 2%. In contrast, the BIASr of the Gitelson-inspired model remains high and does not reach the performance of the other models even for heavy algal blooms.



Figure 5-6. Chlorophyll-a concentration estimated from the four models compared to *in situ* measurements under heavy bloom conditions (Chl-a > 50 mg·m⁻³).

Article-1



Figure 5-7. Chlorophyll-a concentration estimated from the four models compared to *in situ* measurements under moderate concentrations (between 10 and 50 mg·m⁻³).

5.4 Qualitative Evaluation

Chuanmin *et al.* [51] emphasized the difficulty of determining the distribution of blooms and identified three methods to evaluate its accuracy: (1) comparison of field data with concurrent high-resolution sensor observations; (2) examination of the reflectance spectra of the identified bloom; and (3) evaluation of the observed patterns against existing knowledge from historical field surveys [51]. In order to evaluate the accuracy of the four models used in the present study, we opted for the third method proposed above. The MDDEP provided historical *in situ* field surveys on the presence/absence of cyanobacterial blooms and a set of sketch maps showing the extent of the blooms (as described in Section 2.1). Thus, it is possible to compare the sketches drawn on field maps to the true-color images of the corresponding dates obtained from the MDDEP had observed one, we applied each of the four models to the images in order to visually evaluate their ability to monitor the spatial distribution of blooms.

Figure 5-8 shows an example of the field map made by MDDEP technicians and the corresponding true-color image from the MODIS sensor for a bloom that took place on 19 September 2001. A clear correspondence between the sketch and the bloom extension visible on the image (green hue) can be seen both for heavy bloom conditions (the northeast to east part of the lake) and for low-moderate blooming conditions (central part of the lake). The same exercise was done for a series of bloom cases documented by the MDDEP (three dates are shown in Figure 5-9). In this figure, the three left-hand panels show the MODIS true-color images, and the remaining blocks of three panels show the results when the APPEL, Kahru, FAI, and Gitelson models were applied, respectively. The three consecutive dates show the progress of the algal bloom on Missisquoi Bay. The first image corresponds to the beginning of the bloom, the second depicts its expansion on the bay, and the last shows the relative extinction of the bloom episode. From Figure 5-9, it can be seen that all of the models detected a westward movement of the phytoplankton biomass. Since currents and swells are known to be low in bays, we conclude that the movement of this algal biomass was mainly caused by wind. In order to verify this fact, the mean wind direction of this day (15-07-2005 to 16-07-2005) was extracted from the two nearest meteorological stations (within a radius of 10 km of the bay). The average wind direction at "Sainte-Sabine" station (45°13'24.000"N; 73°01'24.000"W) was 90° (i.e., wind blowing from the east) and the average wind direction at "Frelighsburg" station (45°03'01.000"N; 72°51'42.000"W) was 140° (i.e., wind blowing from east-south-east). Therefore, the direction of the displacement of the algal bloom on the MODIS images coincides with the average wind direction at the two stations on that day. This demonstrates the capacity of remote sensing data to monitor algal blooms in inland water bodies on a daily basis.

Comparison of the bloom extension obtained from the model outputs and the true-color images (Figure 5-10) shows that the FAI model tends to amplify the bloom expansion, while the Gitelson-inspired model tends to reduce it compared to the results of APPEL and Kahru. The latter two models seem to better follow the spatial distribution, and this is true for all blooming conditions (low, moderate, and high). Also, in terms of bloom quantification, the APPEL and Kahru models seem to yield the same range of Chl-a,

133

while FAI tends to give higher and Gitelson lower values compared to APPEL and Kahru. In order to spatially compare the model results to *in situ* measured ChI-a, Figure 5-10 shows two stations on Missisquoi Bay (D1 and D2) sampled by the MDDEP (data from these stations were part of the calibration data set). In this example, APPEL and Kahru are the best estimators of ChI-a, while FAI tends to overestimate the concentration in moderate bloom conditions and underestimate it under low bloom conditions, and while the Gitelson-inspired model underestimates the concentrations in all blooming conditions.



Figure 5-8. Comparison between true-color image from MODIS sensor and the corresponding field map prepared by the Ministère du Développement Durable de l'Environnement et des Parcs (MDDEP) technicians at Missisquoi Bay on Lake Champlain.



Figure 5-9. MODIS images of Missisquoi Bay on Lake Champlain on three dates and corresponding estimates of chlorophyll-a concentration by the four models.





Figure 5-10. Comparison between estimated chlorophyll-a concentration calculated by the four models and *in situ* concentration obtained by the MDDEP at two stations on Missisquoi Bay on Lake Champlain.

It is important to note that the largest limitation with the use of satellite-based models for operational monitoring of algal blooms is the presence of clouds that prevent the detection of the blooms and consequently their quantification. In Quebec, cloud cover (overcast) is present more than 50% of the year. However, solar radiation is known to

be among the conditions that favour the development of algal blooms. Thus, blooms are more likely to occur under cloud-free conditions, a fact favoring the use of satellite data as an operational tool to monitor algal blooms and estimate their intensity. Another problem is that pixels along the edge of water bodies tend to overestimate Chl-a (see the red pixels on the shore of Champlain Lake in Figures 5-9 and 5-10). Generally, these pixels depict the combined signal from water and riparian vegetation surrounding the lakes. As most the models use the red/NIR wavelengths, most of the information detected by the sensor comes from Chl-a reflectance. Thus, the models confuse these pixels with heavy bloom conditions. In order to better distinguish water pixels from mixed (land-water) pixels, a land mask is being developed using a data set of more than 1600 MODIS images covering a period of 11 years (2000-2010).

6 CONCLUSIONS

Our aim in this study was to assess the performance of three existing models: Kahru, FAI and a model inspired from Gitelson, as well as a novel approach named APPEL for the estimation of ChI-a in southern Quebec inland waters as an indicator of water quality. One innovative element of this project is its use of the highest resolution of the first seven MODIS bands, enabling the exploitation of a wider spectral range at 250 m spatial resolution. Our results clearly demonstrate that this method can be efficiently used to monitor lake water quality as models can explain up to 95% of the variance. The analysis suggests that APPEL model yields the best overall estimates of ChI-a concentrations. However, Kahru remains a good alternative for ease of use since the model requires only the first two MODIS bands that are already at 250 m spatial resolution (no downscaling procedure needed), and it achieves performances comparable to APPEL, particularly for moderate to high ChI-a concentrations. In heavy blooming conditions, the FAI model can be considered, since it produces the lowest error under these conditions. On the other hand, all models showed poor accuracy to estimate ChI-a at low concentrations, with a slight advantage for APPEL.

The objective of this work was not to replace standard sampling methods but to provide a tool to help reducing the efforts and costs associated with *in situ* sampling. Firstly, the

use of remotely sensed data has limitations related to the presence of clouds, and to border effects where Chl-a values can be highly overestimated. Secondly, the lower performance of all models under low Chl-a concentrations is largely explained by the low signal received from water in the NIR part of the spectrum, making the signal/noise ratio much smaller. It may also be related to the atmospheric correction process, and to the reduction in signal quality caused by the downscaling process in which up to 23% of the original signal can be lost. These factors limit the application of developed algorithms during the initiation phase of blooms corresponding to low Chl-a concentrations.

In short, the model analysis has demonstrated that (1) the first seven bands of MODIS sensor can be used to estimate Chl-a in small lakes with high accuracy during heavy bloom, and consequently this can be used as an affordable and accessible tool to manage water quality; (2) better spatial resolution can compensate for the lack of spectral resolution, since the wavebands used to estimate Chl-a yielded satisfactory results in the majority of cases albeit they are not optimal for the detection of this pigment; (3) estimating low Chl-a concentration was an issue. Since the errors related to downscaling process could not be further improved, focus should be on improvement of the atmospheric correction by testing others approaches. Dark pixel has been shown to be a promising approach and should be considered in such applications [30]; (4) none of the models was perfect, and each had its strengths and weaknesses. Development of an 'ensemble system' thus seems to be a favorable approach for more accurate estimates.

ACKNOWLEDGEMENTS

We thank S. Blais from the Ministère du Développement Durable, de l'Environnement et des Parcs for providing in situ data, the NASA/GSFC for making the MODIS data available, and R. Latifovic and A.P. Trishchenko from the Canadian Centre of Remote Sensing for providing the tool to downscale the resolution of MODIS bands.

REFERENCES AND NOTES

- 1. Gleick, P.H. *The World's Water*, In Island Press: Chicago 2000; Vol. 7, p. 315.
- 2. Bouchard,V.M. *Floraisons de Cyanobactéries au lac Saint-Augustin: Dynamique à Court Terme et Stratification*; Université Laval: Québec, QC, Canada, 2004.
- Laurion, I.; Rousseau, A.; Chokmani, K.; Drogui, P.; Bourget, S.; Warren, A.; Drevnick, P. *Mémoire sur la Situation des lacs au Québec en regard des Cyanobactéries*; INRS-Centre Eau Terre Environnement: Québec, QC, Canada, 2009; p. 25.
- 4. Eija, R.; Rvi, R.O.; Seija, H.I.; 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.
- Agence Française de Sécurité Sanitaire des Aliments, A. Rapport sur l'Évaluation des Risques liés à la Présence de Cyanobactéries et de leurs Toxines dans les eaux Destinées à l'Alimentation, à la Baignade et autres Activités Récréatives; AFSSET: Maisons-Alfort, France, 2006; p. 231.
- 6. MDDEP Bilan des lacs et cours d'eau touchés par une fleur d'eau d'algues bleuvert au québec. Available online: http://www.mddep.gouv.qc.ca/eau/algues-bv/bilan/saison2009/bilan2009.pdf (accessed on 2009).
- 7. Carder, K.L.; Chen, F.R.; Lee, Z.P.; Hawes, S.K.; Kamykowski, D. Semianalytic moderate-resolution imaging spectrometer algorithms for chlorophyll a and absorption with bio-optical domains based on nitrate-depletion temperatures. *J. Geophys. Res.* **1999**, *104*, 5403–5421.
- 8. Morel, A.; Prieur, L. Analysis of variations in ocean color. *Limnol. Oceanogr.* **1977**, *22*, 709–722.
- Baruah, P.J.; Tumura, M.; Oki, K.; Nishimura, H. Neural Network Modeling of Lake Surface Chlorophyll and Sediment Content from Landsat TM Imagery. In *Proceedings of 22nd Asian Conference on Remote Sensing*, Singapore, 5–9 November2001; p. 6.
- 10. Ahn, Y.H.; Shanmugam, P. Detecting the red tide algal blooms from satellite ocean colour observations in optically complex Nourtheast-Asia coastal waters. *Remote Sens. Environ.* **2006**, *103*, 419-437.

- 11. Becker, R.H.; Sultan, M.I.; Boyer, G.L.; Twiss, M.R.; Konopko, E. Mapping cyanobacterial blooms in the great lakes using modis. *J. Great Lakes Res.* **2009**, *35*, 447-453.
- 12. Gitelson, A.A.; Schalles, J.F.; Hladik, C.M. Remote chlorophyll-a retrieval in turbid, productive estuaries: Chesapeake bay case study. *Remote Sens. Environ.* **2007**, *109*, 464-472.
- 13. Hu, C. A novel ocean color index to detect floating algae in the global oceans. *Remote Sens. Environ.* **2009**, *113*, 2118-2129.
- 14. Tarrant, P.; Neuer, S. Monitoring algal blooms in a southwestern U.S. Reservoir system. *EOS* **2009**, *90*, 38.
- 15. Darecki, M.; Stramski, D. An evaluation of modis and seawifs bio-optical algorithms in the baltic sea. *Remote Sens. Environ.* **2004**, *89*, 326-350.
- 16. Mishra, S.; Mishra, D.; Schluchter, W. A novel algorithm for predicting phycocyanin concentrations in cyanobacteria: A proximal hyperspectral remote sensing approach. *Remote Sens.* **2009**, *1*, 758–775.
- 17. Heim, B.; Oberhaensli, H.; Fietz, S.; Kaufmann, H. Variation in Lake Baikal's phytoplankton distribution and fluvial input assessed by seawifs satellite data. *Glob. Planet. Change* **2005**, *46*, 9–27.
- Semovski, S.V.; Mogilev, N.Y.; Sherstyankin, P.P. Lake Baikal ice: Analysis of AVHRR imagery and simulation of under-ice phytoplankton bloom. *J. Marine Syst.* 2000, *27*, 117–130.
- 19. Hu, C.; Chen, Z.; Clayton, T.D.; Swarzenski, P.; Brock, J.C.; Muller-Karger, F.E. Assessment of estuarine water-quality indicators using MODIS medium-resolution bands: Initial results from tampa bay. *Remote Sens. Environ.* **2004**, *93*, 423–441.
- 20. Matthews, M.W.; Bernard, S.; Winter, K. Remote sensing of cyanobacteriadominant algal blooms and water quality parameters in Zeekoevlei, A small hypertrophic lake, using MERIS. *Remote Sens. Environ.* **2010**, *114*, 2070–2087.
- Wheeler, S.M.; Morrissey, L.A.; Levine, S.N.; Livingston, G.P.; Vincent, W.F. Mapping cyanobacterial blooms in Lake Champlain's missisquoi bay using Quickbird and MERIS satellite data. *J. Great Lakes Res.* 2012, *38*, *Supplement 1*, 68–75.
- 22. Dall'Olmo, G.; Gitelson, A.A.; Rundquist, D.C.; Leavitt, B.; Barrow, T.; Holz, J.C. Assessing the potential of seawifs and modis for estimating chlorophyll concentration in turbid productive waters using red and near-infrared bands. *Remote Sens. Environ.* **2005**, *96*, 176–187.

- Gitelson, A.A.; Dall'Olmo, G.; Moses, W.; Rundquist, D.C.; Barrow, T.; Fisher, T.R.; Gurlin, D.; Holz, J. A simple semi-analytical model for remote estimation of chlorophyll-a in turbid waters: Validation. *Remote Sens. Environ.* 2008, *112*, 3582– 3593.
- 24. Gitelson, A.A.; Kondratyev, K.Y. Optical models of mesotrophic and eutrophic water bodies. *Int. J. Remote Sens.* **1991**, *12*, 373–385.
- Gons, H.J.; Rijkeboer, M.; Ruddick, K.G. A chlorophyll-retrieval algorithm for satellite imagery (medium resolution imaging spectrometer) of inland and coastal waters. *J. Plankton Res.* 2002, *24*, 947–951.
- 26. Kahru, M.; Mitchell, B.G.; Diaz, A.; Miura, M. Modis detects a devastating algal bloom in Paracas Bay, Peru. *EOS* **2004**, *85*, 465–472.
- 27. 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.
- 28. Kallio, K.; Attila, J.; Härmä, P.; Koponen, S.; Pulliainen, J.; Hyytiäinen, U.-M.; Pyhälahti, T. Landsat ETM+ images in the estimation of seasonal lake water quality in boreal river basins. *Environ. Manag.* **2008**, *42*, 511–522.
- 29. Moses, W.J.; Gitelson, A.A.; Berdnikov, S.; Povazhnyy, V. Estimation of chlorophylla concentration in Case II waters using modis and meris data—Successes and challenges. *Environ. Res. Lett.* **2009**, *4*, 045005.
- Siegel, D.A.; Wang, M.; Maritorena, S.; Robinson, W. Atmospheric correction of satellite ocean color imagery: The black pixel assumption. *Appl. Opt.* 2000, *39*, 3582–3591.
- 31. Ferrier, G., *Evaluation of Apparent Surface Reflectance Estimation Methodologies*; Taylor & Francis: Abingdon, UK, 1995; Vol. 16.
- 32. Popp, T. Correcting atmospheric masking to retrieve the spectral albedo of land surfaces from satellite measurements. *Int. J. Remote Sens.* **1995**, *16*, 3483–3508.
- Vermote, E.F.; Tanre, D.; Deuze, J.L.; Herman, M.; Morcette, J.J. Second simulation of the satellite signal in the solar spectrum, 6s: An overview. *IEEE Trans. Geosci. Remote Sens.* 1997, 35, 675–686.
- Proud, S.R.; Rasmussen, M.O.; Fensholt, R.; Sandholt, I.; Shisanya, C.; Mutero, W.; Mbow, C.; Anyamba, A. Improving the smac atmospheric correction code by analysis of meteosat second generation ndvi and surface reflectance data. *Remote Sens. Environ.* **2010**, *114*, 1687–1698.
- 35. Smith, G.M.; Milton, E.J. The use of the empirical line method to calibrate remotely sensed data to reflectance. *Int. J. Remote Sens.* **1999**, *20*, 2653–2662.

- 36. Chavez, P.S. *Image-Based Atmospheric Corrections : Revisited and Improved*; American Society for Photogrammetry and Remote Sensing: Bethesda, MD, USA, 1996; Vol. 62.
- Trishchenko, A.P.; Luo, Y.; Khlopenkov, K.V.; Park, W.M. Multi-Spectral Clear-Sky Composites of MODIS/Terra Land Channels (B1–B7) over Canada at 250 m Spatial Resolution and 10-Day Intervals since March, 2000: Top of the Atmosphere (TOA) Data. GeoGratis Client Services: Sherbrooke, QC, Canada, **2010**.
- 38. Giardino, C.; Bresciani, M.; Pilkaityte, R.; Bartoli, M.; Razinkovas, A. *In situ* measurements and satellite remote sensing of Case 2 waters: First results from the curonian lagoon. *Oceanologia* **2010**, *52*, 197–210.
- 39. González Vilas, L.; Spyrakos, E.; Torres Palenzuela, J.M. Neural network estimation of chlorophyll a from meris full resolution data for the coastal waters of galician rias (NW Spain). *Remote Sens. Environ.* **2011**, *115*, 524–535.
- 40. Gordon, H.R.; Wang, M. Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with seawifs: A preliminary algorithm. *Appl. Opt.* **1994**, *33*, 443–452.
- Hu, C.; Muller-Karger, F.E.; Andrefouet, S.; Carder, K.L. Atmospheric correction and cross-calibration of Landsat-7/ETM+ imagery over aquatic environments: A multiplatform approach using SeaWifs/MODIS. *Remote Sens. Environ.* 2001, *78*, 99–107.
- Markham, B.L.; Barker, J.L.; Kaita, E.; Seiferth, J.; Morfitt, R. On-orbit performance of the Landsat-7 ETM+ radiometric calibrators. *Int. J. Remote Sens.* 2003, *24*, 265– 285.
- 43. Kahru, M.; Mitchell, B.G. Spectral reflectance and absorption of massive red tide off southern california. *J. Geophys. Res.* **1998**, *103*, 21 601–621 609.
- 44. Miller, P.I.; Shutler, J.D.; Moore, G.F.; Groom, S.B. SeaWifs discrimination of harmful algal bloom evolution. *Int. J. Remote Sens.* **2006**, *27*, 2287–2301.
- 45. Aiken, J.; Fishwick, J.R.; Lavender, S.J.; Barlow, R.; Moore, G.F.; Sessions, H.; Bernard, S.; Ras, J.; Hardman-Mountford, N.J. Validation of meris reflectance and chlorophyll during the bencal cruise october 2002: Preliminary validation of new demonstration products for phytoplankton functional types and photosynthetic parameters. *Int. J. Remote Sens.* **2007**, *30*, 497–516.
- 46. Kuster, T. Passive optical remote sensing of cyanobacteria and other intense phytoplancton blooms in coastal and inland waters. *Int. J. Remote Sens.* **2009**, *30*, 4401–4425.

- Stumpf, R.P.; Arnone, R.A.; Gould, R.W.; Martinolich, V.; Tester, P.A.; Steward, R.G.; Subramaniam, A.; Culver, M.E.; Pennock, J.R. Seawifs Ocean Colour Data for Us Southeast Coastal Waters. In *Proceeding of 6th Internaional Conference on Remote Sensing for Marine and Coastal Environment*, Charleston, SC, USA, 1–3 May 2000; pp. 25–27.
- 48. Kuster, T.; Metsamaa, L.; Vahtmae, E.; Strombeck, N. On suitablility of MODIS 250m resolution band data for quntitative mapping cyanobacterial blooms. *Proc. Estonian Acad. Sci. Biol. Ecol.* **2006**, *55*, 318–529.
- 49. Reinart, A.; Kuster, T. Comparison of different satellite sensors in detecting cyanobacterial bloom events in the Baltic Sea. *Remote Sens. Environ.* **2006**, *102*, 74–85.
- 50. Hu, C.; He, M. Origin and offshore extent of floating algae in olympic sailing area. *EOS Trans. AGU* **2008**, *89*, 302.
- 51. Hu, C.; Lee, Z.; Ma, R.; Yu, K.; Li, D.; Shang, S. Moderate resolution imaging spectroradiometer (MODIS) observations of cyanobacteria blooms in Taihu Lake, China. *J. Geophys. Res.* **2010**, *115*, 20.
- 52. Gordon, H.R.; Brown, O.B.; Evans, R.H.; Brown, J.W.; Smith, R.C.; Baker, K.S.; Clark, D.K., A semianalytic radiance model of ocean color. *J. Geophys. Res.* **1988**, *93*, 10909.
- 53. Gitelson, A.A.; Gritz, Y.; Merzlyak, M.N., Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *J. Plant Physiol.* **2003**, *160*, 271–282.
- 54. Gitelson, A.A.; Dall'Olmo, G. Effect of bio-optical parameter variability on the remote estimation of chlorophyll-a concentration in turbid productive waters: Experimental results. *Appl. Opt.* **2005**, *44*, 412–422.
- 55. Babin, M.; Stramski, D. Light absorption by aquatic particles in the near-infrared spectral region. *Limnol. Oceanogr.* **2002**, *47*, 911–915.
- 56. Gons, H.J. Optical teledetection of chlorophyll a in turbid inland waters. *Environ. Sci. Technol.* **1999**, *33*, 1127–1132.
- 57. Liu, W.; Liu, Y.L.; Chris, M.; Wu, G. Monitoring variation of water turbidity and related environmental factors in Poyang Lake National Nature Reserve, China. *Proc. SPIE* **2007**, *6754*, 67541H.
- 58. Luciani, X.; Pierre, C.; Frédéric, T.; David, B.; André, B.; Stéphane, M.; Claude, J. Analyse numérique des spectres de 3d issus de mélanges non linéaires. Sud Toulon Var, 2007.

- 59. Krause, P.; Boyle, D.P.; Bäse, F., Comparison of different efficiency criteria for hydrological model assessment. *Adv. Geosci.* **2005**, *5*, 89–97.
- 60. Chorus, I.; Bartram, J. *Toxic Cyanobacteria in Water. A Guide to Their Public Health Consequences, Monitoring and Management*; WHO. E & FN Spon: London, UK, 1999; p. 416.
- 61. Watson, S.B.; Mc-Cauley, E.; Downing, J.A. *Patterns in Phytoplankton Taxonomic Composition across Temperate Lakes of Differing Nutrient Status*; American Society of Limnology and Oceanography: Waco, TX, USA, 1997; Vol. 42.

© 2012 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).

PARTIE 3 : ARTICLE-2

Avant-propos

Cet article a été publié dans le journal *Remote sensing* du *MDPI*. Les co-auteurs qui ont participés à la rédaction de ce manuscrit sont :

1. Anas El Alem :

Responsable de la conception de recherches, prétraitements et traitements des images, le développement des modèles mathématiques, l'analyse et l'interprétation des résultats et à la rédaction du manuscrit.

2. Karem Chokmani :

Responsable également de la conception de recherches et a contribué à l'édition, l'analyse, et à l'examen du manuscrit.

3. Laurion Isabelle :

Personne intermédiaire avec le MDDELCC pour l'octroi des données *in situ* et a contribuée à l'édition, l'analyse et à la révision du manuscrit.

4. Salah Eddine El Adlouni :

Responsable des interprétations mathématiques et statistiques et à la révision du manuscrit.

Abstract

The purpose of this study is to assess the performance of an adaptive model (AM) in estimating chlorophyll-a concentration (Chl-a) in optically complex inland waters. Chl-a modeling using remote sensing data is usually based on a single model that generally follows an exponential function. The estimates produced by such models are relatively accurate at high Chl-a concentrations, but accuracy drops at low concentrations. Our objective was to develop an approach combining spectral response classification and three semi-empirical algorithms. The AM discriminates between three blooming classes (waters poorly, moderately, and highly loaded in Chl-a), with discrimination thresholds set using the classification and regression tree (CART) technique. The calibration of three specific estimators for each class was achieved using a multivariate stepwise regression. Compared to published models (Floating Algae Index, Kahru model, and APProach by ELimination) using the same data set, the AM provided better Chl-a concentration estimates (R² of 0.96, relative RMSE of 23%, relative Bias of -2%, and a relative NASH criterion of 0.9). Moreover, the AM achieved an overall success rate of 67% in the estimation of blooming classes (corresponding to low, moderate, and high Chl-a concentration classes). This was done using an independent data set collected from 22 inland water bodies for the period 2007-2010 and for which the only information available was the blooming class.

Keywords: Remote sensing; MODIS; Inland waters; HABs; Chl-a; Classification; CART; Multivariate regression; *Stepwise.*

Remote Sens. **2014**, *6*, 1-x manuscripts; doi:10.3390/rs60x000x



ISSN 2072-4292 www.mdpi.com/journal/remotesensing

Article

An Adaptive Model to Monitor Chlorophyll-a in Inland Waters in Southern Quebec Using Downscaled MODIS Imagery

Anas El-Alem^{1,*}, Karem Chokmani¹, Isabelle Laurion¹ and Sallah E. El-Adlouni²

- ¹ Centre Eau Terre Environnement, INRS, 490 De la Couronne Street, Québec, QC G1K 9A9, Canada; E-Mails: karem.chokmani@ete.inrs.ca (K.C.); isabelle.laurion@ete.inrs.ca (I.L.)
- ² Mathematics and Statistics Department, Moncton University, 18 Antonine-Maillet Avenue, Moncton, NB E1A 3E9, Canada; E-Mail: salah-eddine.el.adlouni@umoncton.ca
- * Author to whom correspondence should be addressed; E-Mail: anas.elalem @ete.inrs.ca; Tel.: +1-418-654-2570; Fax: +1-418-6542-6002.

Received: 15 January 2014; in revised form: 19 June 2014 / Accepted: 24 June 2014 / Published: 15 July 2014

1 INTRODUCTION

Given its synoptic view, consistent recurrence, and capacity to provide information over a wide range of wavelengths, remote sensing has good potential to provide the data necessary to monitor harmful algal blooms (HAB). Bloom detection is possible through the bio-optical characteristics of the principal pigment in algae and cyanobacteria, chlorophyll-a (Chl-a), which is characterized by low reflectance in the red wavelengths and high reflectance in the near-infrared (NIR). This contrast makes it possible to estimate Chl-a concentration using bio-optical models that link inherent and apparent optical properties of water bodies [1]. For example, Landsat TM data have been used to retrieve Chl-a and total suspended solid (TSS) concentrations in Lake Kasumigaura using the neural network technique [2] and to assess phycocyanin concentrations in Lake Erie to study the temporal and spatial dynamics of cyanobacterial blooms [3]. In addition, data from the Advanced Very High Resolution Radiometer (AVHRR) have been used to assess the behavior of the main taxonomic groups of Lake Baikal phytoplankton as a function of ice conditions [4] and other water quality parameters [5,6]. Recently, QuickBird and MEdium Resolution Imaging Spectrometer (MERIS) data were used successfully to study cyanobacterial blooms in Lake Champlain [7]. Thus, many semi-analytical algorithms [2-4,8-21] and derived indices [22] are now available in the literature to retrieve Chl-a and phycocyanin concentrations in inland water bodies.

Although several models and approaches designed to model Chl-a in inland water bodies are now available in the literature, most assume that the concentrations of the whole range of optically active components in a water body can be modeled using the same function. However, this calibration function may be linear, exponential, or polynomial, depending on the relative concentrations of each component. Use of an inappropriate calibration function to estimate a given component may lead to over- or under-estimates. This was clearly demonstrated by El-Alem *et al.* [23], who compared four exponential models used to estimate Chl-a at low-to-moderate and high concentrations. The accuracy of all four models significantly decreased at low concentrations [24]. Conversely, when low Chl-a concentrations were estimated using a

155

linear [25,26] or polynomial [27] function, the results were more accurate. Moreover, it has recently been demonstrated that *prior* identification of the spectral type of inland waters significantly enhances the accuracy of Chl-a concentration estimates [28,29]. In addition, Yu *et al.* [29] have shown that estimation error can be reduced about 15 fold by using the appropriate spectral region to model low Chl-a concentrations.

Due to their one-day revisit time, MODIS images were privileged in this study. A short revisit time was considered desirable not only to enable us to collect enough data for model calibration and validation, but also to support eventual use of the technique in the context of risk management. The first two of the 36 MODIS bands in the red/NIR region are recorded at 250 m spatial resolution. The rest of the visible and shortwave infrared (SWIR) bands, which are more appropriate for the detection of Chl-a, colored dissolved organic matter (CDOM), and TSS, are recorded either at 500 m or at 1 km spatial resolution, making them unsuitable for monitoring algal blooms in small to medium-sized inland waters. However, the spatial resolution of MODIS bands 3-7 can be downscaled from 500 to 250 m using an approach developed at the Canadian Center for Remote Sensing (CCRS) by Trishchenko *et al.* [30]. In this manner, it is possible to acquire data from the first seven MODIS bands, originally designed for aerosol, cloud, and land applications, at 250 m spatial resolution, covering the visible, NIR, and SWIR parts of the spectrum.

The objective of the present study was to develop an adaptive model (AM) to estimate Chl-a concentration using MODIS data downscaled at 250 m spatial resolution and to evaluate its performance on a series of water bodies in southern Quebec, Canada. Performance of the AM was evaluated utilizing *cross-validation* using several statistical evaluation indices and a confusion matrix using an independent semi-quantitative database. We also used the same databases (continuous and ordinal) to compare performance of the AM to that of three other models originally developed to estimate Chl-a in inland waters (the floating algae index (FAI [13]), Kahru model [22], and APProach by ELimination (APPEL [24])).

156

2 MATERIALS AND METHODS

2.1 Study Area and in situ Data

The study area, located in southern Quebec between latitudes 44° and 50° north and longitudes 67° and 80° west, contains 22 inland water bodies that are large enough to be investigated at 250 m spatial resolution and that have been monitored by the Ministère du Développement Durable, Environnement et Lutte contre les Changements Climatiques (MDDELCC) because of cyanobacterial bloom occurrence and recurrence as observed by the local volunteer monitoring network. These water bodies include the Choinière and Taureau reservoirs and Aylmer, Bouchette, Brome, Champlain, des Commissaires, Etchemin, Fréchette, Labrecque, Lovering, Mandeville, Maskinongé, Massawippi, Nairne, Ouareau, Perchaude, Pohénégamook, Roxton, Tortue, William, and Adélard lakes. These water bodies were used to calibrate and validate performance of the models (Figure 2-1).



Figure 2-1. Geographic location of the water bodies used for model calibration and validation.

The calibration database of Chl-a concentrations was collected by the MDDELCC over nine years (2000-2008) at several sites located on four lakes (Brome, Champlain, Nairne, and William (Figure 2-2)). A total of 363 samples were collected, with a minimum value of 0.52 mg Chl-a·m⁻³ (Nairne Lake, 2005), a maximum value of 450,000 mg·m⁻³ (Missisquoi Bay of Lake Champlain, 2001), an average of 3700 mg·m⁻³, and a median of 14 mg·m⁻³ (Figure 2-3). Chl-a concentration was quantified following the protocol of the Centre d'expertise en analyse environnementale du Québec (CEAEQ; 2012a [31]). Given the presence of clouds over some sampled sites and the poor quality of some images (fuzzy images or presence of artifacts), only 46 of these 363 samples could be used to calibrate the models. In this final data set, the Chl-a concentration varied from 2.7 (Nairne Lake, 2008) to 91,000 mg·m⁻³ (Missisquoi Bay of Lake Champlain, 2003).



Figure 2-2. Water sampling stations on the four studied lakes used for calibration.

Three blooming classes were defined according to the thresholds used by the World Health Organization (WHO [32]) to characterize the quality of water bodies in relation to the hazard associated with water usage in the presence of algal blooms: the low Chl-a class corresponds to concentrations below 10 mg·m⁻³; the moderate class to concentrations between 10 and 50 mg·m⁻³; and the high class to concentrations above
50 mg·m⁻³. A second data set, used to validate the AM, was based on three cyanobacterial abundance classes: water bodies with cell densities lower than 20,000 cyanobacteria cells·mL⁻¹ (assumed to be equivalent to 10 mg Chl-a·m⁻³ [32]); water bodies with densities between 20,000 and 100,000; and water bodies with densities higher than 100,000 cells·mL⁻¹ (assumed to be equivalent to 50 mg·m⁻³). This data set was collected in the 22 water bodies listed above between 2007 and 2010 following the protocol of the CEAEQ (CEAEQ; 2012 [33]). For the same reasons as stated above, only 103 of the 677 samples collected were used to evaluate the performance of the AM.



Figure 2-3. Histogram of the frequency of chlorophyll-a values observed (complete data set, N = 363).

2.2 MODIS Data

The remotely sensed data was obtained from the MODIS Level 1B product, available in HDF format on the NASA website (http://ladsweb.nascom.nasa.gov/data/search.html). The MODIS sensor is located on the TERRA platform of the NASA earth observation system. It operates across a wide spectrum, with 36 bands covering the region from 0.4 to 14.4 µm. Spatial resolution of the images varies from 250 m to 1 km. For this study, MODIS images collected on the same dates as the *in situ* samples were downloaded and pre-processed to calibrate the AM *classifier* and *estimators*. Given their higher spatial resolution, only the first seven MODIS bands were used in this study, enabling

use of data from small lakes (but >2.25 km²). The first two bands were already at 250 m spatial resolution, while bands 3-7 were originally at 500 m (Table 1). The spatial resolution of the latter bands was downscaled to 250 m using an approach developed at the CCRS [30]. Two pre-processing steps were used in the downscaling process: (1) Translation of the data from 500 to 250 m spatial resolution using adaptive regression and radiometric normalization as described by Trishchenko *et al.* [34]; and (2) re-projection of the images from the Sinusoidal to the Lambert Conformal Conic projection.

Tableau 2-1. Characteristics of the MODIS bands used in the present study. NIR = Near infrared; SWIR = Shortwave infrared; LCAB = Land/Cloud/Aerosol Boundaries; and LCAP = Land/Cloud/Aerosol Properties.

Primary Use	Band	Band Width (nm)	Spatial Resolution(m)	Spectral Regions
	1	620–670	250	Red
LCAB	2	841–876	250	NIR
	3	459–479	500	Blue
	4	545–565	500	Green
LCAP	5	1,230–1,250	500	NIR
	6	1,628–1,652	500	SWIR
	7	2,105–2,155	500	SWIR

Level 1B of the MODIS sensor contains a set of geo-located and calibrated data. For many applications, especially multi-temporal analyses, raw relative pixel values or digital image numbers have to be corrected for atmospheric effects and converted to spectral reflectance at the surface before the images are processed [35]. Improper atmospheric correction can lead to significant errors in the retrieved reflectance and affect the accuracy of the estimates [36]. Several atmospheric correction models have been developed, including the Simplified Model for Atmospheric Correction (SMAC [37]), Second Simulation of the Satellite Signal in the Solar Spectrum (6S [38]), Moderate-Resolution Atmospheric Transmittance and Radiance Code (MODTRAN [39]), ATmospheric CORrection (ATCOR [40]), Dark Object Subtraction (DOS [41]), and COSine Transmission for atmospheric correction (COST [42]). The DOS and COST models are widely used, as they rely entirely on image-based atmospheric corrections and provide reasonably accurate reflectance estimates, but accuracy is improved by the use of more sophisticated models that exploit *in situ* optical depth measurements and

radiative transfer codes [42] to correct for both additive and multiplicative effects. A comparison study of three absolute atmospheric correction models, DOS, COST, and DOS4, produced similar results [41]. Reflectance was slightly improved, but the overall appearance was similar to the original image. The COST model was more effective in visible bands but less accurate in the NIR, particularly in humid conditions. On the other hand, a comparison analysis made by Norjamäki and Tokola [42] demonstrated that the RMSEr values for multi-temporal images decreased by an average of 6% using DOS, by 14% using SMAC, and by 15% using 6S, when compared to uncorrected images [43].

Figure 2-4 shows the signal recorded by the first seven MODIS bands during an algal bloom in Lake Champlain (September 19, 2001) as adjusted using two different atmospheric correction models (SMAC and DOS) and the apparent reflectance (AR) model. The AR model performs a simple conversion from the digital numbers in the images to spectral reflectance at the surface, while the DOS model corrects the additive effects caused by haze, and the SMAC model additionally corrects multiplicative effects caused by ozone, water vapor, and aerosols. The comparison clearly shows that the reflectance estimated by the AR model is higher than that estimated by the SMAC and DOS models, especially for the first four bands (blue to NIR), and that the behavior of the return signal from the SMAC and DOS models is almost the same for all bands except the first one (blue). Since the AR model does not correct for atmospheric effects and shorter wavelengths are easily scattered by atmospheric particles, the visible bands, particularly the blue one, were the most affected. The higher blue reflectance of the DOS model compared to SMAC was due to the huge sensitivity of this band to Rayleigh diffusion, which is mainly caused by water vapor and aerosols. The reflectance corrections performed by the SMAC model were closest to the spectral response of Chl-a, which is characterized by high absorption in the blue and red bands and high reflectance in the green and NIR bands. For these reasons, the SMAC model was chosen to correct the MODIS images. All pre-processing of images (downscaling, re-projection, and atmospheric correction) was performed using an automatic tool developed by the CCRS [30].

161

Figure 2-4. Comparison of the MODIS signal from Lake Champlain during an algal bloom on 19 September 2001 with different atmospheric correction models



2.3 Adaptive Model Parameterization

Parameterization of the adaptive model (AM) was performed to exploit as much as possible of the spectral information captured by the MODIS sensors for a given water body. Analysis of the return signal for the low-to-moderate blooming class in the calibration data set shows reasonable correlation ($R^2 = 0.48$; p-value < 0.0001) with the visible bands (triangle shape made by bands 1-3 on Figure 2-5.A), whereas the return signal for high Chl-a shows good correlation ($R^2 = 0.95$; p-value < 0.0001) with bands ranging from red to SWIR (polygon shapes made by bands 3-7 on Figure 2-5.B). However, while the distinction between the spectral signature of high Chl-a and that of moderate-to-low Chl-a is obvious, the distinction between low and moderate Chl-a is more complex (Figure 2-5.A).



Figure 2-5. Spectral signature behaviour for three chlorophyll-a concentration classes: (A) low (<10 mg·m⁻³) and moderate (10-50 mg·m⁻³) concentrations combined; and (B) high concentrations (>50 mg·m⁻³). Different colors represent the signature of individual samples.

Created in the 1980s by Breiman [44], the Classification and Regression Tree (CART) method is widely used for classification and regression purposes. To build decision trees, CART uses a so-called *learning* sample, composed of a set of historical data with pre-assigned classes for all observations [45] and a set of *spliting* varaibles. These decision trees are then used to classify new data. Classification trees are built in accordance with a splitting rule, which splits *learning* samples into smaller groups of maximum homogeneity (Figure 2-6). The maximum homogeneity of child nodes is determined by the impurity function (i(t)), which can be calculated by either the Gini or the Towing splitting rule, and is equivalent to maximization of the change of impurity function $\Delta i(t)$:

$$\Delta i(t) = i(t_{Parent}) - P_{Left}i(t_{Left-child}) - P_{Right}i(t_{Right-child})$$
(2-1)

where P_{Left} and P_{Right} are respectively the probabilities of right and left nodes. Consequently, at each node CART solves the following maximization problem:

$$\underset{x_{j} \leq x_{j}^{B}, j=1,...,M}{\operatorname{argmax}} \left[i(t_{Parent}) - P_{Left}i(t_{Left-child}) - P_{Right}i(t_{Right-child}) \right]$$
(2-2)

This enables CART to find the best value (x_j^B) to split the t_{Parent} (parent node) into $t_{Left-child}$ (left node) and $t_{Right-child}$ (right node) and maximize the change of impurity function $\Delta j(t)$.



Figure 2-6. The splitting algorithm of the Classification and Regression Tree (CART), where t_{Parent} , $t_{Left-child}$, and $t_{Right-child}$ are parent, left, and right nodes, x_j is the splitting variable j, and x_j^B is the best splitting value of the x_j .

The CART method was applied to four variables derived from the spectral response of the first seven MODIS bands to classify the calibration data set into three blooming classes (water poorly, moderately, and highly loaded in Chl-a). These variables were calculated from the surfaces (S) underneath the reflectance curves for (1) the visible bands (S_{Vis}); (2) the green to NIR bands (S_{G-NIR}); (3) the visible to NIR bands ($S_{Vis-NIR}$); and (4) the red to SWIR bands (S_{R-SWIR}) using Equation (2-3).

$$S = \frac{1}{2} \sum_{i=1}^{l} |R(\lambda_i) \times \lambda_{i+1} - R(\lambda_{i+1}) \times \lambda_i|$$
(2-3)

where I is the number of MODIS bands and $R(\lambda_i)$ and λ_i are respectively the reflectance and wavelength absorption of the ith band.

Before applying CART to the calibration database, we converted the continuous database into an ordinal one; i.e. the *in situ* measurements were classified into the three blooming classes defined above. The CART results (Figure 2-7) showed that (S_{Vis-NIR})

and (S_{R-SWIR}) were the two best *splitting* variables, enabling us to calibrate the AM *classifier*. This *classifier* was thereafter used to split the calibration database into the three blooming classes. We then calibrated three different *estimators*, each specific to a given blooming class. The AM *estimators* were calibrated using a multivariate *stepwise* regression in which explanatory variables were chosen through an automatic procedure that usually followed a sequence of F-tests. We used the *stepwise* regression in the forward selection mode, starting with none of the variables in the model, testing them one by one, and including only the statistically significant ones [46]. All of the ratios and band subtractions possibly related to the bio-optical activity of Chl-a, and a range of algorithms widely used in the literature for inland water (as evaluated in [24]), were used as explanatory variables and the three *estimators* are summarized in Tableau 2-2. The AM was thus structured in two steps: 1) determination of the blooming class of a given pixel based on the AM *classifier*, and 2) Chl-a concentration estimation of this pixel using the corresponding *estimator* for each predefined blooming class (Figure 2-7).



Figure 2-7. Thresholds values (×10⁶) used to distinguish between the three chlorophyll-a blooming classes using the classification and regression tree method (CART).

Tableau	2-2.	Equations	of	the	three	calibrated	models	(or	estimators)	using	а	multivariate
regression to estimate Chl-a concentration.												

Blooming Classe	s Regression Estimators (Issued from Stepwise)
Low Chl-a	$[Chl - a] = 6.49 \times 10^{-5} \times S_{Vis-NIR} - 0.323)$
Moderate Chl-a	$[Chl - a] = e^{(0.0031 \times R(\lambda_{b1}) + 1.132 \times 10^{-5} \times S_{Vis-NIR} + 0.49)}$
High Chl-a	$[Chl - a] = e^{(23.97 \times APPEL + 3.28)}$
whore S	is the coloulated area undernanth the reflectance signal surves from visible to NID banday

where $S_{Vis-NIR}$ is the calculated area underneath the reflectance signal curves from visible-to-NIR bands:

$$S_{Vis-NIR} = \frac{|\lambda_{b3}R(\lambda_{b4}) - \lambda_{b4}R(\lambda_{b3})| + |\lambda_{b4}R(\lambda_{b1}) - \lambda_{b1}R(\lambda_{b4})| + |\lambda_{b1}R(\lambda_{b2}) - \lambda_{b2}R(\lambda_{b1})| + |\lambda_{b2}R(\lambda_{b3}) - \lambda_{b3}R(\lambda_{b2})|}{2}$$

For MODIS bands: λ_{b1} =645, λ_{b2} =859, λ_{b3} =469, λ_{b4} =555, λ_{b5} =1240, λ_{b6} =1640 and λ_{b7} =2130 (nm)

and where APPEL is the APProach by Elimination model [24]:

 $APPEL = R(\lambda_{b2}) - \left[\left[R(\lambda_{b3}) - R(\lambda_{b2}) \right] \times R(\lambda_{b2}) + \left[R(\lambda_{b1}) - R(\lambda_{b2}) \right] \right]$

2.4 Accuracy Assessment and Validation Data

The performance of the AM and three other models (FAI [22], Kahru [13], and APPEL [24]) originally developed to estimate the Chl-a concentration of inland waters was evaluated using the *cross-validation* technique, in which a sample is temporarily removed from the calibration database and the remaining samples are then used as training data to estimate the value of the removed sample using the pre-calibrated model. This operation is then repeated for the whole database. Once all Chl-a measurements are estimated, the model's performance can be evaluated using statistical indices such as the coefficient of determination (R^2), relative root mean square error (RMSEr), relative bias (BIASr), and relative NASH criterion (NASHr). The NASHr criterion evaluates the performance by comparing the estimated values to the *in situ* measurement average, producing a result that ranges between $-\infty$ and 1.0 (inclusive). A negative NASH result means that it would be better to use the *in situ* measurement average than the model estimates, whereas values between 0.0 and 1.0 are generally viewed as acceptable levels of performance, and model performance is

satisfactory for values higher than 0.8; the model is perfect for a NASHr = 1.0. The mathematical equations of the indices are as follows:

$$R^{2} = \left[\frac{\sum_{i=1}^{n} (0_{i} - \overline{0})(Es - \overline{Es})}{\sqrt{\sum_{i=1}^{n} (0_{i} - \overline{0})^{2}} \sqrt{\sum_{i=1}^{n} (Es_{i} - \overline{Es})^{2}}}\right]^{2}$$
(2-4)

$$BIASr = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{Es_i - O_i}{Es_i} \right)$$
(2-5)

RMSEr =
$$\sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(\frac{Es_i - O_i}{Es_i}\right)^2}$$
 (2-6)

$$Nr = 1 - \frac{\sum_{i=1}^{n} \left(\frac{O_{i} - Es_{i}}{O_{i}}\right)^{2}}{\sum_{i=1}^{n} \left(\frac{O_{i} - \overline{O}}{\overline{O}}\right)^{2}}$$
(2-7)

where n is the sample size, O and Es are the observed and estimated values, and \overline{O} and \overline{Es} are the means of the observed and estimated values.

A second, independent, semi-qualitative database was also used to validate the performance of all models. This database, containing data on 22 water bodies monitored by the MDDELCC between 2007 and 2010, was composed of ordinal data that indicated only whether cell densities were lower than 20,000 cyanobacteria cells·mL⁻¹ (assumed to be equivalent to 10 mg Chl-a·m⁻³ [32]), between 20,000 and 100,000 cells·mL⁻¹, or higher than 100,000 cells·mL⁻¹ (assumed to be equivalent to 50 mg·m⁻³). A confusion matrix was used to test the accuracy of the AM and the FAI, Kahru, and APPEL models. Omission and commission errors, success rates, and Kappa index (K) were calculated. The Kappa index was used to quantify concordance between the measured and the estimated Chl-a classes. Concordance is weak when the Kappa is negative, good for a Kappa that is positive and higher than 0.6, and

excellent when the Kappa is above 0.8. Tableau 2-3 summarizes the parameters used in the confusion matrix as well as the Kappa index computation.

Tableau 2-3. Simplified diagram of the parameters used in the confusion matrix and to calculate the Kappa index: a and d are the number of well-classified values, b and c are the number of mis-classified values, n₁, n₂, n₃, and n₄ are respectively totals of a + b, c + d, a + c, and b + d, N is the sample size, $P_o = \frac{a+d}{N}$, and $P_c = \frac{n_1 \times n_3 + n_2 \times n_4}{N^2}$.

				Measu	ured	
pe		TRUE	FALSE	Total	Success Rate (%)	Commission Error (%)
	TRUE	а	b	n ₁	(a/n ₁)	(b/n ₁)
	FALSE	С	d	n ₂	(b/n ₂)	(c/n ₂)
timat	Total	n ₃	n ₄	Ν		
Ш	Success Rate (%)	(a/n ₃)	(d/n ₄)			
	Omission Error (%)	(c/n ₃)	(b/n ₄)			
	Global Success (%)					(a + b)/N

3 RESULTS AND DISCUSSION

3.1 Calibration

Eighteen samples served as the training set to calibrate the *estimator* for waters classified in the low Chl-a concentration class. The same number of samples was used to calibrate the moderate Chl-a *estimator*, and 10 samples were used to calibrate the high Chl-a *estimator*. The calculated spectral indices (existing approaches, subtractions, and band ratios) showed good correlation with *in situ* measurements for all three *estimators*. Low Chl-a concentrations were linearly correlated, while moderate and high Chl-a concentrations followed an exponential function. The determination coefficients (R^2) were respectively 0.91, 0.92, and 0.98 for the three blooming classes (Figure 3-1).

The high correlation between *in situ* measurements and spectral indices ($R^2 > 0.91$) for all three *estimators* illustrates the value of splitting the modeling workspace and using multivariate regressions. For all of the *estimators*, Chl-a estimation was limited to

MODIS bands 1-4 or combinations of these bands, which contain the spectral regions most sensitive to the optical activity of the Chl-a pigment; i.e. high absorption in the blue (band-3) and red (band-1) and high reflectance in the green (band-4) and NIR (band-2). The *stepwise* regression maximized the information on the presence of Chl-a and minimized the mis-modeling of this pigment at the expense of other optically active components, such as TSS or dissolved organic matter (DOM), which exhibit different spectral absorption and reflectance signatures.



Figure 3-1. Results of multivariate regression adjustments between the measured chlorophyll-a concentrations and the return signals of MODIS images for the three blooming classes: (A) Low; (B) Moderate; (C) High Chl-a concentrations. The scale of the y-axis in the insert is logarithmic to better illustrate the correlation.

The explanatory variable selected by the *stepwise* regression for low blooming conditions, characterized by clear, non-turbid water, was $S_{Vis-NIR}$, which is mostly controlled by the visible part of the spectrum, consistent with the assumption that waters with low-to-moderate Chl-a are highly influenced by the bio-optical activity of the Chl-a pigment. $S_{Vis-NIR}$ was also the variable selected by the CART method to discriminate

between waters poorly and moderately loaded in Chl-a (Figure 2-7). High phytoplankton biomass, on the other hand, is known to generate turbid waters, which significantly reduce water molecule absorption in the red-NIR part of the spectrum [13]. This explains why the APPEL model, which depends entirely on NIR band reflectance to estimate Chl-a concentrations, was the best predictor for waters highly loaded in Chl-a. The transition class of waters moderately loaded in Chl-a would logically be influenced by both Chl-a bio-optical activity and turbidity. The explanatory variables selected by the *stepwise* regression for these waters supported this supposition. The Chl-a variance was explained by two orthogonal variables, $S_{Vis-NIR}$ (sensitive to Chl-a activity) and $R(\lambda_{b1})$ (sensitive to turbidity), with most of the variance explained by $S_{Vis-NIR}$ (p-value < 0.0001). The two variables did not show any collinearity (for both variables the variance inflation factor was equal to 1.5, which is <<<10, the threshold usually used to report collinearity issues).

3.2 Evaluation of *Estimators*

The performance of the three *estimators* was satisfactory and increased with higher Chl-a concentrations (Tableau 3-1). The results for the NASHr criterion demonstrate that the calibrated *estimators* are robust, in particular for moderate (NASHr = 0.8) and high (NASHr = 0.89) Chl-a concentrations. This is not surprising, as the mathematical expressions of the *estimators* for waters moderately and highly loaded in Chl-a are mainly based on the red-to-NIR bands, for which Chl-a reflectance is at its maximum, as shown in Figures 3–20 of Mackie (2010; [47]). Chl-a and water pixels are easily distinguished since water reaches its maximum absorption in this part of spectrum. The RMSEr results support the above findings, with a clear decrease in error for moderate and high Chl-a concentrations. The *estimator* for waters poorly loaded in Chl-a seems to over-estimate the concentrations by 5%, while the *estimators* for moderate to high Chl-a concentrations are almost unbiased (BIASr = -1%).

The relatively lower performance of the *estimator* for low Chl-a can be explained by the fact that its explanatory variable is mostly composed of bands located in the visible part of spectrum. Since the MODIS bands of higher resolution (at 250 m and at 500 m

downscaled to 250 m spatial resolution) were designed for land, cloud, and atmosphere applications, they are not centered on Chl-a absorption and reflection peaks. Moreover, due to the low reflectance of water with low Chl-a concentrations, the return signal is more likely to be disturbed by noise caused by atmospheric particles, the reflectance of other optically active components present in the water (e.g., TSS and DOM), and the downscaling process, which can lead to the loss of up to 23% of the original MODIS signal [24].

Evoluction Indiaco	Concentration Levels							
Evaluation indices	Low	Moderate	High					
R ²	0.83	0.86	0.92					
NASHr criterion	0.44	0.8	0.89					
RMSEr (%)	38	9	8					
BIASr (%)	-5	-1	-1					

Tableau 3-1. Evaluation of the three models (estimators) using cross-validation technique.

3.3 Evaluation of the Adaptive Model: Cross-Validation

The performance of the AM, FAI, Kahru, and APPEL models was evaluated using a *cross-validation* based on the same database. Figure 3-2 shows the ChI-a estimated by the four models as a function of the observed ChI-a (measured *in situ* by the MDDELCC), along with the model performance indices. The figure clearly shows that the AM performed the best ($R^2 = 0.96$, NASHr = 0.9, and RMSEr = 23%), followed by the APPEL and FAI models with almost identical performance, while the Kahru model performed the least well. However, the overall performance of the three comparison models (FAI, Kahru, and APPEL) was quite similar. The figure also demonstrates that the dispersion of values is well distributed with respect to the 1:1 line, highlighting the robustness of the models even at their extremities.

However, although the performance of all the models is satisfactory, it is clear that they are strongest when addressing high concentration values. In the context of timely intervention to manage risk and protect human and animal health, water bodies already affected by HAB are less interesting to monitor than relatively healthy waters with the potential to develop algal blooms and be exposed to eutrophication. When the

performance of the models at moderate-to-low concentrations (<50 mg Chl-a mg⁻³, established by WHO as the threshold for declaring a HAB situation) is examined (Figure 3-3), only the AM provides acceptable estimates ($R^2 = 0.56$ and NASHr = 0.24); the performance of the other three models was significantly lower, with negative NASHr values indicating that the measured Chl-a average were better predictors than the model estimates.



Figure 3-2. Chlorophyll-a concentration estimated from the four models compared to *in situ* measurements for the complete database, with model performance indices.

Article-2



Figure 3-3. Chlorophyll-a concentration estimated from the four models compared to *in situ* measurements for the database using only values <50 mg Chl-a m⁻³.

The above results indicate that partitioning the solution space into blooming classes increased the accuracy of Chl-a estimates, with *estimators* able to explain from 91% to 98% of the Chl-a variance (Figure 3-1.C). Using an adequate calibration function for each blooming class (linear for low Chl-a and exponential for moderate-to-high Chl-a) specifically helped to improve the estimation of moderate-to-low Chl-a concentrations, which were problematic when applying the FAI, Kahru, and APPEL models. The AM error (RMSEr) was 200%-250% lower compared to the other models, and its systematic error (BIASr) was 300%-450% lower.

3.4 Validation by Independent Data

Tables 5-8 show the confusion matrix results for the AM, APPEL, FAI, and Kahru models. The AM performed the best (global success = 67% and Kappa index = 0.51), followed by the APPEL and FAI models, while the Kahru model performed the worst. From this analysis, we can see that estimating Chl-a at high concentrations was not problematic for any of the models, with commission errors of 17%, 18%, 18%, and 12% for the four models, respectively. On the other hand, the commission and omission errors were clearly higher for the APPEL, FAI, and Kahru models than for the AM for waters with low concentrations of Chl-a (commission error = 19% for AM versus 43%, 49%, and 60% for the APPEL, FAI, and Kahru models, respectively). Interestingly, the errors were relatively high for all four models in the moderate blooming class. The modest performance of the AM is explained by the fact that misclassification of Chl-a estimates can be done at both ends of this class (in the low-to-moderate and in moderate-to-high regions). Nevertheless, it is important to note that for all models, misclassification generally occurred with respect to low Chl-a concentrations. In other words, models more often underestimated the moderate concentrations (false negatives). For example, the AM generated false negatives in 39% of cases (14 of 36 moderate concentrations classified as low Chl-a (Tableau 3-2)) and false positives in 22% of cases (eight of 36 moderate concentrations classified as high Chl-a). The AM thus had a 22% chance of declaring a false high blooming condition, an error that is acceptable. By comparison, false positives were generated in about 22%, 26%, and 24% of cases by the APPEL, FAI, and Kahru models, respectively. Thus, the AM and APPEL models showed the same level of classification performance for waters moderately-to-highly loaded in Chl-a, but the AM still achieved the best performance for the overall moderate class, with a commission error of 61% versus 78% for the APPEL and 82% for the FAI and Kahru models.

It is important to note that most *in situ* sample points had to be moved from their original sampling sites by at least one pixel (equivalent to 250 m) as many samples were taken near the lake shoreline. For sensors such as MODIS, with fairly coarse resolution, this represents a significant handicap as these regions are transition zones from land to

174

water (mixed pixels). These zones are influenced by the reflectance of many different components, which can lead to biased estimates. Moreover, the validation data were provided in cyanobacteria density units, which had to be converted to Chl-a units using the WHO conversion factor of 1 µg Chl-a to 2 million cells given in Chorus and Bartram [32]. Of course, the actual conversion factor can vary extensively with cell size and light history, and depends on the dominant species of a bloom [48,49]. This conversion thus introduced varying levels of uncertainty into the validation database. Laboratory error, corresponding respectively to 1.4% and 0.6% for cyanobacteria densities of 20,000 and 100,000 cells mL⁻¹ according to the CEAEQ [33], equivalent to uncertainties of ±0.14 and ± 0.30 mg Chl-a m⁻³ for the two thresholds, could also affect the accuracy of models. In addition, the uncertainties associated with sampling methods (depth in water column, time of day, location on the lake, preservation conditions) and the presence of phytoplankton cells other than cyanobacteria in the water, especially at low Chl-a concentrations when cyanobacteria are less likely to dominate, are both likely to affect the accuracy of values in the measured database and thus the ability of the models to correctly classify Chl-a estimates. Despite the aforementioned limitations, the performance of the AM was acceptable (Kappa index = 0.51 and global success = 67%) and higher than the performance of the other three models.

	Measured								
		[Chl-a] < 10	10 < [Chl-a] < 50	[Chl-a] > 10	Total	Commission Error	Success Rate		
	[Chl-a] < 10	30	3	4	37	19%	81%		
σ	10< [Chl-a] < 50	14	14	8	36	61%	39%		
nate	[Chl-a] > 10	2	3	25	30	17%	83%		
Estir	Total	46	20	37	103	****	****		
ш	Omission Error	35%	30%	32%	*****	****	****		
	Success Rate	65%	70%	68%	****	****	****		
	Global Success	****	****	****	****	****	67%		
	Карра	****	****	****	*****	****	0.51		

Tableau 3-2. The adaptive model confusion matrix results.

			M	easured			
		[Chl-a] < 10	10 < [Chl-a] < 50	[Chl-a] > 10	Total	Commission Error	Success Rate
	[Chl-a] < 10	17	5	8	30	43%	57%
stimated	10 < [Chl-a]< 50	29	11	11	51	78%	22%
	[Chl-a] > 10	0	4	18	22	18%	82%
	Total	46	20	37	103	****	****
ш	Omission Error	63%	45%	51%	****	****	****
	Success Rate	37%	55%	49%	****	****	****
	Global Success	****	****	****	****	****	45%
	Карра	****	****	****	****	****	0.21

Tableau 3-3. The APPEL confusion matrix results.

Tableau 3-4. The *FAI* confusion matrix results.

	Measured								
		[Chl-a] < 10	- 10 < [Chl-a] < 50	[Chl-a] > 10	Total	Commission Error	Success Rate		
σ	[Chl-a] < 10	24	10	13	47	49%	51%		
	10 < [Chl-a] < 50	22	7	10	39	82%	18%		
nate	[Chl-a] > 10	0	3	14	17	18%	82%		
Estir	Total	46	20	37	103	****	****		
	Omission Error	48%	65%	62%	****	****	****		
	Success Rate	52%	35%	38%	****	****	****		
	Global Success	****	****	****	****	****	44%		
	Карра	****	****	****	****	****	0.15		

Tableau 3-5. The *Kahru* confusion matrix results.

			M	easured			
		[Chl-a] < 10	10 < [Chl-a] < 50	[Chl-a] > 10	Total	Commission Error	Success Rate
Estimated	[Chl-a] < 10	8	6	6	20	60%	40%
	10 < [Chl-a] < 50	38	12	16	66	82%	18%
	[Chl-a] > 10	0	2	15	17	12%	88%
	Total	46	20	37	103	****	****
	Omission Error	83%	40%	59%	*****	****	****
	Success Rate	17%	60%	41%	*****	****	****
	Global Success	****	****	****	*****	****	34%
	Карра	****	****	****	****	****	0.10

3.5 Qualitative Validation: Model's Application

The four models were applied to a series of MODIS images during a period when an important expansion of HAB (composed mainly of Aphanizomenon flos-aquae) was underway, as seen on the true color composite images of Missisquoi Bay, Lake Champlain (Figure 3-4). The three upper panels of this figure show the MODIS true color images on three consecutive dates (12, 17, and 30 September 2001), followed by the same images after application of the AM, APPEL, FAI, and Kahru models. A clear correspondence between the bloom shapes on the true color images and the AM outputs can be seen for all dates. This concordance is absent for the FAI, Kahru, and APPEL models on 12 and 17 September, during the early stages of the bloom. On the other hand, all of the models proved equally able to detect the well-established bloom on 30 September. Given the high negative BIASr of the FAI, Kahru, and APPEL models with respect to waters poorly-to-moderately loaded in Chl-a (Figure 3-3), these results were expected. False negatives (underestimated Chl-a) were produced by the AM, APPEL, FAI, and Kahru models in about 39%, 57%, 56%, and 58%, respectively, of moderate blooming class cases. The relative performance of the models is well illustrated by the results produced when they are applied to the MODIS images for 12 and 17 September, when bloom conditions were moderate (central part of the bay): while the AM estimates Chl-a concentrations to be between 12 and 33 mg·m⁻³ (indicating moderate blooming conditions), the other models all estimate Chl-a to be less than 10 mg m⁻³. This is a good demonstration of how the FAI, Kahru, and APPEL models can fail to detect blooms during their initial phase (higher errors, negatively biased, and high false negatives), when Chl-a concentrations are below 50 mg·m⁻³. In this case study, the AM-modified MODIS images show no apparent gaps in Chl-a estimates from one blooming class to the other. This is probably due to the fact that the AM *estimators* were trained using overlapping data. In other words, when spectrally splitting the calibration database using the CART method, the calibration sub-database of low Chl-a concentrations ranged from 2.7 to 19 mg·m⁻³, while it ranged from 9 to 84 mg \cdot m⁻³ in the moderate class, enabling a smooth transition and avoiding gaps.

Article-2





Figure 3-4. Application of the four models (Adaptive model, APPEL, FAI, and Kahru) to a series of MODIS images collected during the establishment of a HAB on Missisquoi Bay, Lake Champlain, compared to the true color composite images (RGB for red, green, blue). The red polygon indicates the shoreline and southern boundary of the bay. Chlorophyll-a concentrations are in a Napierian logarithmic scale.

In order to spatially compare the model results to *in situ* measured Chl-a concentrations, Figure 3-5 shows the position of two stations on Missisquoi Bay (A and B) sampled on 19 September 2001 by the MDDELCC during an important algal bloom event (data from these stations were part of the calibration data set). In this example, the APPEL and Kahru models under-estimate both Chl-a concentrations, the AM under-estimates the moderate and over-estimates the high concentration, and the FAI over-estimates the moderate and under-estimates the high concentration. Computation of the relative error (Re) of the two samples using Equation (3-1) demonstrates that the best performers are the AM and FAI. The AM is the most accurate with respect to the moderate concentration and is the second best choice for the high concentration, trailing the FAI model by only 3%.

$$\operatorname{Re} = \frac{\operatorname{Es}_{[\operatorname{Chl}-a]} - M_{[\operatorname{Chl}-a]}}{\operatorname{Es}_{[\operatorname{Chl}-a]}} \times 100$$
(3-1)

where $Es_{[Chl-a]}$ is the estimated Chl-a concentration and $M_{[Chl-a]}$ the measured one.

As these results demonstrate, the AM performs as well as or better than the models most commonly used to estimate Chl-a concentration in inland water bodies, and it provides the most stable results, with errors remaining below 20%.

Article-2



Figure 3-5. Comparison between estimated chlorophyll-a concentration calculated by the adaptive model, *APPEL*, *FAI*, and *Kahru* models and *in situ* measurements obtained by the MDDEFP at two stations on Missisquoi Bay on Lake Champlain (19 September 2001).

4 CONCLUSIONS

This study was designed to test the performance of an adaptive model (AM) developed to estimate Chl-a concentrations using MODIS images downscaled at 250 m spatial resolution in southern Quebec inland water bodies. Several innovative elements were tested with this approach: use of a classification method (CART) to spectrally pre-identify the blooming class of a sample (waters poorly, moderately, and highly loaded in Chl-a) and to apply the corresponding *estimator* for a final estimation; optimization of satellite information by means of a multivariate *stepwise* regression; and use of the first seven MODIS bands, originally designed for land, atmosphere, and cloud applications, downscaled at 250 m spatial resolution using an approach developed at the Canadian center for remote sensing.

Several validation techniques were used to assess the performance of the proposed approach: *cross-validation*, validation by independent ordinal data using a confusion matrix, and qualitative validation by applying the models to a series of MODIS images. The FAI, Kahru, and APPEL models were subjected to similar procedures of calibration and validation using the same databases (continuous and ordinal). The determination coefficients of the three AM *estimators* were high (>0.91), and the AM yielded the best overall estimates of ChI-a concentrations, especially for the low-to-moderate blooming classes (<50 mg ChI-a·m⁻³; negative NASHr values for the FAI, Kahru, and APPEL models). However, confusion matrix analysis revealed a decrease in performance for all four models in the case of waters moderately loaded in ChI-a. Estimates for this blooming class are highly sensitive to misclassification of the data at both extremities, but the AM remained the most efficient predictor, with the lowest false negatives (39%). In addition, qualitative validation highlighted the potential of the proposed method to detect algal blooms at their initialization stage, which is problematic for the other models.

Our goal in developing this approach was not to replace conventional monitoring methods, but to provide a tool to improve the management of fieldwork, which is expensive and complex for regions with a high density of lakes such as southern Quebec. The well-known limitations of remote sensing, including the loss of signal in the

181

presence of clouds and the low performance of most standard models at low-to-moderate Chl-a concentrations or when other optically active components are abundant, may have discouraged planners from integrating such data into their intervention plans. However, while organizations such as the World Health Organization and the Institut national de santé publique du Québec do not consider waters with less than 10 mg Chl-a·m⁻³ to pose a threat to human or animal health, monitoring of water at the initialization stage of an algal bloom remains crucial for lakes threatened by eutrophication. The AM provides an improved tool to monitor harmful algal blooms in medium-sized lakes, with a satisfactory level of performance even at low-to-moderate Chl-a concentrations.

ACKNOWLEDGMENTS

We thank Sylvie Blais and Nathalie Bourbonnais from the Ministère du Développement durable, Environnement et Lutte contre les Changements Climatiques (MDDELCC) for providing the in situ database, le Fonds québécois de la recherche sur la nature et les technologies (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.

AUTHOR CONTRIBUTIONS

Anas El-Alem wrote the manuscript and was responsible for the research design, preprocessing of images, development of the mathematical models, and analysis and interpretation of results. Karem Chokmani was responsible for the research design and contributed to editing, analysis, and review of the manuscript. Isabelle Laurion arranged for the MDDELCC to provide in situ data and contributed to editing, analysis, and review of the manuscript. Sallah E. El-Adlouni was responsible for mathematical and statistical interpretations and review of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- Carder, K.L.; Chen, F.R.; Lee, Z.P.; Hawes, S.K.; Kamykowski, D. Semianalytic moderateresolution imaging spectrometer algorithms for chlorophyll A and absorption with bio-optical domains based on nitrate-depletion temperatures. *J. Geophys. Res.* **1999**, *104*, 5403–5421.
- 2. Baruah, P.J.; Tumura, M.; Oki, K.; Nishimura, H. Neural network modeling of lake surface chlorophyll and sediment content from Landsat TM imagery. In Proceedings of the 22nd Asian Conference on Remote Sensing, Singapore, 5–9 November 2001; p. 6.
- 3. Vincent, R.K.; Qin, X.; McKay, R.M.L.; Miner, J.; Czajkowski, K.; Savino, J.; Bridgeman, T. Phycocyanin detection from LANDSAT TM data for mapping cyanobacterial blooms in Lake Erie. *Remote Sens. Environ.* **2004**, *89*, 381–392.
- Semovski, S.V.; Mogilev, N.Y.; Sherstyankin, P.P. Lake Baikal ice: Analysis of AVHRR imagery and simulation of under-ice phytoplankton bloom. *J. Mar. Syst.* 2000, *27*, 117–130.
- 5. Hu, C.; Chen, Z.; Clayton, T.D.; Swarzenski, P.; Brock, J.C.; Muller, K.F.E. Assessment of estuarine water-quality indicators using MODIS medium-resolution bands: Initial results from Tampa Bay. *Remote Sens. Environ.* **2004**, *93*, 423–441.
- 6. Matthews, M.W.; Bernard, S.; Winter, K. Remote sensing of cyanobacteriadominant algal blooms and water quality parameters in Zeekoevlei, a small hypertrophic lake, using MERIS. *Remote Sens. Environ.* **2010**, *114*, 2070–2087.
- 7. Wheeler, S.M.; Morrissey, L.A.; Levine, S.N.; Livingston, G.P.; Vincent, W.F. Mapping cyanobacterial blooms in Lake Champlain's Missisquoi Bay using QuickBird and MERIS satellite data. *J. Gt. Lakes Res.* **2012**, *38*, 68–75.
- 8. Dall'Olmo, G.; Gitelson, A.A.; Rundquist, D.C.; Leavitt, B.; Barrow, T.; Holz, J.C. Assessing the potential of SeaWiFS and MODIS for estimating chlorophyll concentration in turbid productive waters using red and near-infrared bands. *Remote Sens. Environ.* **2005**, *96*, 176–187.
- Gitelson, A.A.; Dall'Olmo, G.; Moses, W.; Rundquist, D.C.; Barrow, T.; Fisher, T.R.; Gurlin, D.; Holz, J. A simple semi-analytical model for remote estimation of chlorophyll-a in turbid waters: Validation. *Remote Sens. Environ.* 2008, *112*, 3582– 3593.
- 10. Gitelson, A.A.; Kondratyev, K.Y. Optical models of mesotrophic and eutrophic water bodies. **1991**, *12*, 373–385.
- 11. Gitelson, A.A.; Schalles, J.F.; Hladik, C.M. Remote chlorophyll-a retrieval in turbid, productive estuaries: Chesapeake Bay case study. *Remote Sens. Environ.* **2007**, *109*, 464–472.

- Gons, H.J.; Rijkeboer, M.; Ruddick, K.G. A chlorophyll-retrieval algorithm for satellite imagery (Medium Resolution Imaging Spectrometer) of inland and coastal waters. *J. Plankton Res.* 2002, *24*, 947–951.
- 13. Kahru, M.; Mitchell, B.G.; Diaz, A.; Miura, M. MODIS detects a devastating algal bloom in Paracas Bay, Peru. *Earth Observ. Satell.* **2004**, *85*, 465–472.
- Aiken, J.; Fishwick, J.R.; Lavender, S.J.; Barlow, R.; Moore, G.F.; Sessions, H.; Bernard, S.; Ras, J.; Hardman-Mountford, N.J. Validation of MERIS reflectance and chlorophyll during the BENCAL cruise October 2002: Preliminary validation of new demonstration products for phytoplankton functional types and photosynthetic parameters. *Int. J. Remote Sens.* 2007, *30*, 497–516.
- 15. Becker, R.H.; Sultan, M.I.; Boyer, G.L.; Twiss, M.R.; Konopko, E. Mapping cyanobacterial blooms in the Great Lakes using MODIS. *J. Great Lakes Res.* **2009**, *35*, 447–453.
- 16. Darecki, M.; Stramski, D. An evaluation of MODIS and SeaWiFS bio-optical algorithms in the Baltic Sea. *Remote Sens. Environ.* **2004**, *89*, 326–350.
- 17. Giardino, C.; Bresciani, M.; Pilkaityte, R.; Bartoli, M.; Razinkovas, A. *In situ* measurements and satellite remote sensing of case 2 waters: First results from the Curonian Lagoon. *Oceanologia* **2010**, *52*, 197–210.
- 18. Gons, H.J. Optical teledetection of chlorophyll a in turbid inland waters. *Environ. Sci. Technol.* **1999**, *33*, 1127–1132.
- 19. González Vilas, L.; Spyrakos, E.; Torres Palenzuela, J.M. Neural network estimation of chlorophyll a from MERIS full resolution data for the coastal waters of Galician Rias (NW Spain). *Remote Sens. Environ.* **2011**, *115*, 524–535.
- Pinkerton, M.H.; Richardson, K.M.; Boyd, P.W.; Gall, M.P.; Zeldis, J.; Oliver, M.D.; Murphy, R.J. Intercomparison of ocean colour band-ratio algorithms for chlorophyll concentration in the subtropical front east of New Zealand. *Remote Sens. Environ.* 2005, *97*, 382–402.
- 21. Wei, L.; Yanfang, L.; Chris, M.; Guofeng, W. Monitoring variation of water turbidity and related environmental factors in poyang lake national nature reserve, China. *Proc. SPIE* **2007**,*6754*, doi:10.1117/12.764879.
- 22. Hu, C. A novel ocean color index to detect floating algae in the global oceans. *Remote Sens. Environ.* **2009**, *113*, 2118–2129.
- 23. Huisman, J.; Matthijs, H.C.P.; Visser, P.M. *Harmful Cyanobacteria*; Springer: Dordrecht, The Netherlands, 2005; Volume 3.

- El-Alem, A.; Chokmani, K.; Laurion, I.; El-Adlouni, S.E. Comparative analysis of four models to estimate chlorophyll-a concentration in Case II waters using MOderate Resolution Imaging Spectroradiometer (MODIS) Imagery. *Remote Sens.* 2012, 4, 2373–2400.
- 25. Moses, W.J.; Gitelson, A.A.; Berdnikov, S.; Povazhnyy, V. Estimation of chlorophylla concentration in Case II waters using MODIS and MERIS data: Successes and challenges. *Environ. Res. Lett.* **2009**, *4*, 045005.
- Moses, W.J.; Gitelson, A.A.; Perk, R.L.; Gurlin, D.; Rundquist, D.C.; Leavitt, B.C.; Barrow, T.M.; Brakhage, P. Estimation of chlorophyll-a concentration in turbid productive waters using airborne hyperspectral data. *Water Res.* 2012, *46*, 993– 1004.
- Yacobi, Y.Z.; Moses, W.J.; Kaganovsky, S.; Sulimani, B.; Leavitt, B.C.; Gitelson, A.A. NIR-red reflectance-based algorithms for chlorophyll-a estimation in mesotrophic inland and coastal waters: Lake Kinneret case study. *Water Res.* 2011, 45, 2428–2436.
- 28. Shi, K.; Li, Y.; Li, L.; Lu, H.; Song, K.; Liu, Z.; Xu, Y.; Li, Z. Remote chlorophyll-a estimates for inland waters based on a cluster-based classification. *Sci. Total Environ.* **2013**, *444*, 1–15.
- 29. Yu, G.; Yang, W.; Matsushita, B.; Li, R.; Oyama, Y.; Fukushima, T. Remote estimation of Chlorophyll-a in inland waters by a NIR-Red-based algorithm: Validation in Asian lakes. *Remote Sens.* **2014**, *6*, 3492–3510.
- Trishchenko, A.P.; Luo, Y.; Khlopenkov, K.V.; Park, W.M. Multi-Spectral Clear-Sky Composites of MODIS/Terra Land Channels (B1-B7) Over Canada at 250m Spatial Resolution and 10-Day Intervals Since March, 2000: Top of the Atmosphere (TOA) Data. Available online: http://geogratis.gc.ca/api/beta/en/nrcan-rncan/esssst/c133a159–76d7–5ad3-a0dc-6f68c547e519.html (accessed on 1 December 2001).
- Centre d'Expertise en Analyse Environnementale du Québec (CEAEQ). Détermination de la Chlorophylle a: Méthode par Fluorométrie, MA. 800—Chlor. 1.0, Rév. 2, Mnistère du Développement Durable, de l'Environnement et des Parcs du Québec: Québec, QC, Canada, 2012; p. 16.
- 32. Chorus, I.; Bartram, J. Toxic cyanobacteria in water. In *A Guide to Their Public Health Consequences, Monitoring and Management*; WHO. E & FN Spon: London, UK, **1999**; p. 416.
- Centre d'Expertise en Analyse Environnementale du Québec (CEAEQ). Dépistage des Cyanobactéries, MA. 800–Cya.dep 1.0, Mnistère du Développement Durable, de l'Environnement et des Parcs du Québec: Québec, QC, Canada, 2012; p. 16.
- 34. 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. *Remote Sens. Environ. Monit.* **2006**, *6366*, 36607–36607.

- 35. Moran, M.S.; Inoue, Y.; Barnes, E.M. Opportunities and limitations for image-based remote sensing in precision crop management. *Remote Sens. Environ.* **1997**, *61*, 319–346.
- Rahman, H. Influence of atmospheric correction on the estimation of biophysical parameters of crop canopy using satellite remote sensing. *Int. J. Remote Sens.* 2001, 22, 1245–1268.
- Rahman, H.; Dedieu, G. SMAC: A simplified method for the atmospheric correction of satellite measurements in the solar spectrum. *Int. J. Remote Sens.* 1994, *15*, 123–143.
- 38. Vermote, E.F.; Tanre, D.; Deuze, J.L.; Herman, M.; Morcette, J.J. Second simulation of the satellite signal in the solar spectrum, 6S: An overview. *IEEE Trans. Geosci. Remote Sens.* **1997**, *35*, 675–686.
- Berk, A.; Anderson, G.P.; Bernstein, L.S.; Acharya, P.K.; Dothe, H.; Matthew, M.W.; Adler-Golden, S.M.; Chetwynd, J.H.; Richtsmeier, S.C.; Pukall, B. MODTRAN 4 radiative transfer modeling for atmospheric correction. *Proc. SPIE* **1999**, *3756*, 348– 353.
- 40. Richter, R. A spatially adaptive fast atmospheric correction algorithm. *Int. J. Remote Sens.* **1996**, *17*, 1201–1214.
- Chavez, P.S. An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. *Remote Sens. Environ.* **1988**, *24*, 459– 479.
- 42. Chavez, P.S. Image-based atmospheric correction—Revisited and improved. *Photogramm. Eng. Remote Sens.* **1996**, *62*, 1025–1036.
- 43. Norjamäki, I.; Tokola, T. Comparison of atmospheric correction methods in mapping timber volume with multitemporal Landsat images in Kainuu, Finland. *Photogramm. Eng. Remote Sens.* **2007**, *73*, 155–163.
- 44. Breiman, L. *Classification and Regression Trees*; Chapman & Hall/CRC: Monterey, Calif., U.S.A.: Wadsworth, Inc., **1984**.
- 45. Breiman, L.; Friedman, J.; Stone, C.J.; Olshen, R.A. *Classification and Regression Trees*; Monterey, Calif., U.S.A.: Wadsworth, Inc., **1984**; p. 368.
- 46. Draper, N.R.; Smith, H. *Applied Regression Analysis*; Wiley-Interscience: New York, NY, USA, **1998**; pp. 307–312.
- 47. 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**.
- 48. Bastien, C.; Cardin, R.; Veilleux, E.; Deblois, C.; Warren, A.; Laurion, I. Performance evaluation of phycocyanin probes for the monitoring of cyanobacteria. *J. Environ. Monit.* **2011**, *13*, 110–118.

 MacIntyre, H.; Lawrenz, E.; Richardson, T. Taxonomic discrimination of phytoplankton by spectral fluorescence. In *Chlorophyll a Fluorescence in Aquatic Sciences: Methods and Applications*; Suggett, D.J., Prášil, O., Borowitzka, M.A., Eds.; Springer Netherlands, **2010**; Volume 4, pp. 129–169.

© 2014 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).

PARTIE 4 : ARTICLE-3

Avant-propos

Cet article a été soumis pour une publication dans le journal *Remote sensing of environment*. Les co-auteurs qui ont participés à la rédaction de ce manuscrit sont :

1. Anas El Alem :

Responsable de la conception de recherches, prétraitements et traitements des images, le développement des modèles mathématiques, l'analyse et l'interprétation des résultats et à la rédaction du manuscrit.

2. Karem Chokmani :

Responsable également de la conception de recherches et a contribué à l'édition, l'analyse, et à l'examen du manuscrit.

3. Laurion Isabelle :

Personne intermédiaire avec le MDDELCC pour l'octroi des données *in situ* et a contribuée à l'édition, l'analyse et à la révision du manuscrit.

4. Salah Eddine El Adlouni :

Responsable des interprétations mathématiques et statistiques et à la révision du manuscrit.
Abstract

The increase in nutrient and pollutant laden of runoff, caused by urban population growth and agricultural production, leads to an intensification in algal productivity and changes in the species composition of the phytoplankton with possible harmful algal blooms (HAB) in inland water systems. Several environmental health scientists outlined the potential human health impacts associated with HAB extension. Rapid detection and daily surveillance of these microorganisms are thus the main keys for an effective prevention against their damages. Water sampling at the time of the algal bloom is the most accurate and used way to detect their presence, but it is expensive and limited in time and space. MODIS land bands (1-7) downscaled to 250 m spatial resolution had demonstrated their ability to monitor phytoplankton chlorophyll-a concentration (Chl-a), in near-real-time (NRT), as a proxy of HAB presence in medium to small-sized lakes. However, detecting algal blooms at their initialization was challenging using these bands. In this study, we propose an approach for Chl-a estimation in medium to small-sized lakes based on ensemble based system (EBS) using MODIS land bands downscaled to 250 m spatial resolution. The EBS was evaluated by cross-validation and by confusion matrix using an independent ordinal database. When compared to other models, the EBS provided the best performances either for high blooming conditions $(R^2 = 0.98)$, a relative RMSE (RMSEr) = 15%, a relative Bias (BIASr) = -2%, and a relative Nash (NASHr) = 0.95) or at their initialization ($R^2 = 0.77$, RMSEr = 37%, BIASr = -83%, and a NASHr = 0.70). Moreover, the EBS reached a global success rate of 82% and a Kappa index of 0.71 using the independent database

Keywords: Remote sensing; MODIS; Inland waters; HABs; Chl-a; Gaussian quadrature; Ensemble based system; CART; *Stepwise*, Near real time.

Near-real-time phytoplankton chlorophyll-a monitoring in medium to small-sized lakes using ensemble-based systems with MODIS imagery downscaled to 250 m spatial resolution

Anas El-Alem^{1,*}, Karem Chokmani¹, Isabelle Laurion¹ and Sallah E. El-Adlouni²

- ¹ Centre Eau Terre Environnement, INRS, 490 De la Couronne Street, Québec, QC G1K 9A9, Canada; E-Mails: karem.chokmani@ete.inrs.ca (K.C.); isabelle.laurion@ete.inrs.ca (I.L.)
- ² Mathematics and Statistics Department, Moncton University, 18 Antonine-Maillet Avenue, Moncton, NB E1A 3E9, Canada; E-Mail: salah-eddine.el.adlouni@umoncton.ca
- * Author to whom correspondence should be addressed; E-Mail: anas.elalem @ete.inrs.ca; Tel.: +1-418-654-2570; Fax: +1-418-6542-6002.

1 INTRODUCTION

Traditionally, water quality is monitored by standard sampling methods of some environment indicators such as the chlorophyll-a concentration (Chl-a) of phytoplanktons, total suspended solids, colored dissolved organic, or other physicochemical parameters (Chapman 1992). With the nutrient and pollutant enrichment of runoff caused by anthropogenic or natural sources, hydrographic changes, or climate change impacts, phytoplankton biomass have been reported to increase spatially and temporally in many offshore, coastal, and inland waters around the world (Allison *et al.* 2009; Hu *et al.* 2010; Sellner *et al.* 2003; Yentsch *et al.* 2008). When sized between less than 10 μ m to over than 100 μ m, phytoplanktons are commonly called 'algae' or 'algal' cell (Blondeau-Patissier *et al.* 2014) and in high concentration they often refer to harmful algal blooms (HAB) with probable cyanobacteria (Figure 1-1).

The occurrence, persistence, and the toxic character of some species of these blooms have been reported to cause fish mortality, shellfish poisoning, physiological impairment, and numerous ecological and health impacts (Allison *et al.* 2009; Sellner *et al.* 2003; Yentsch *et al.* 2008). Their daily monitoring and the early warning of their development are then the main keys to control and to mitigate the harmful effect of these microorganisms for public safety. This supposes investigating massive efforts in measuring and collecting samples, systematically, from all over inland waters of areas of interest. In addition, it has also been proved that standard sampling methods, based on some sample points, are not sufficient to monitor the spatio temporal evolution of algal blooms, especially in high concentrations (Eija *et al.* 1998).

Phytoplankton Chl-a bio-optical activity can cause distinct changes in the color of water by absorbing and scattering the light incident on water (Gordon *et al.* 1988). Given their capacity to provide information over a wide range of wavelengths, satellite data present an interesting tool to estimate Chl-a concentration from remotely sensed spectral reflectance water. This is possible by relating optical changes observed in the reflected light at specific wavelengths to the Chl-a (Moses *et al.* 2009). Based on this principal,

ocean color remote sensing data has provided more than two decades of near-real-time (NRT) synoptic and recurrent distribution of Chl-a estimate for all aquatic systems (offshore, coastal, and inland waters). However, because of the relatively coarse spatial resolution of ocean color sensors (Tableau 1-1), remote estimation of Chl-a was only reserved to offshore and high sized inland waters, where the spatial resolution is not a problematic.



Figure 1-1. Relationship, in a logarithmic scale, between cyanobacteria density and chlorophyll-a concentration. The database is collected by the Ministère du Développement Durable, de l'Environnement et de la Lutte contre les Changements Climatiques of Québec over nine years (2000-2008) from seven different lakes.

Insofar most of inland waters feeding cities are often medium to small of size, it would be appreciable developing approaches capable of monitoring the Chl-a spatial distribution, in NRT, for such water bodies. To do this, three essential criteria are needed: 1) low temporal resolution; 2) low spatial resolution; and 3) high spectral resolution. For those raisons, MERIS data (Tableau 1-1) were been widely used during the last decade (Binding *et al.* 2010; Feng *et al.* 2014; Wheeler *et al.* 2012). Sensors with higher spatial resolution were also been tested for this purpose (Allan *et al.* 2011;

Torbick *et al.* 2013; Wheeler *et al.* 2012). However, in both cases, the NRT aspect was missing. Indeed, even MERIS temporal frequency is relatively short (3 days), but it is not sufficient to monitor the variation of phytoplankton blooms, specifically in high concentrations as they can vary rapidly (Davis *et al.* 2007), and consequently many events can be missed with such temporal frequency.

Sensor	Agency	Satellite	Operating Dates	Swath (km)	Resolution (m)	Number of bands	Spectral coverage (nm)	Temporal frequency (days)
SeaWiFS	NASA (USA)	OrView-2 (USA)	08/1997	2806	1100	8	402-885	1
OCI	NSPO (Taiwan)	FORMOSAT-1 (Taiwan)	01/1999	690	825	6	433-12500	1
ОСМ	ISRO (India)	IRS-P4 (India)	05/1999	1420	350	8	402-885	2
OSMI	KARI (Korea)	12/1999 (Korea)	12/1999	800	850	6	400-900	28
Terra	NASA (USA)	MODIS	12/1999	2330	1000	36	405-14385	1
MERIS	ESA (Europe)	Evsisat-1 (Europe)	03/2002	1150	300/1200	15	412-1050	3
Aqua	NASA (USA)	MODIS	05/2002	2330	1000	36	405-14385	1
COCTS	NSOAS (China)	HaiYang-1 (China)	05/2002	1400	1100	10	402-12500	3

Tableau 1-1. Most current ocean color sensors.

The Canadian centre for remote sensing (CCRS) has developed an approach allowing to downscale MODIS land bands 3 7 to 250 m spatial resolution (Trishchenko *et al.* 2006). With this approach, it is possible to use the MODIS first seven bands at 250 m spatial resolution. Moreover, it has been proven that these bands are also sensitive to detect optical properties variation of waters (Hu *et al.* 2004). With such characteristics, MODIS land bands downscaled to 250 m spatial resolution (MODIS-D-250) would present the best remote sensing data for a NRT monitoring of phytoplankton spatial distribution in medium to small-sized lakes. Effectively, when tested with the most popular models designed to estimate Chl-a in inland waters (floating algae index (Hu 2009), Gitelson (Gitelson *et al.* 2008), and Kahru (Kahru *et al.* 2004)), the MODIS-D-250 bands demonstrated their ability for detecting and retrieving the Chl-a in such aquatic systems, specifically during HAB (R^2 up to 0.94). However, Chl-a estimates were much less accurate at the initialization of algal blooms (El-Alem *et al.* 2012).

To address this issue, an adaptive model (AM) that combines the results of a *classifier* and three specific *estimators* (explicitly developed for waters poorly, moderately, and highly loaded in Chl-a) was developed and was able to detect algal blooms during their initialization, but a staircase effect was sighted when modeling the Chl-a using MODIS-D-250 images. This staircase effect was related to the AM *classifier* uncertainty (El-Alem *et al.* 2014). Indeed, classification algorithms are known the be local and unstable (Alpak *et al.* 2007). This instability can significantly affect Chl-a estimation accuracy using the AM. On the other hand, due to the spectral signature complexity of inland waters (Dall'Olmo et al. 2005), Chl-a modeling in these waters is very challenging because it could co-vary with other optical elements as total suspended solids and the colored dissolved organic matters (Moses *et al.* 2009). This could also lead to over- or under-estimate the Chl-a concentration and to amplify the staircase effet.

In such complicated modeling contexts, many authors suggest the use of ensemble-based system (EBS), which seems to be more stable and accurate (Chan and Paelinckx 2008: Dasarathy and Sheela 1979: Hansen and Salamon 1990: Ismail and Mutanga 2010; Jacobs et al. 1991; Jordan and Jacobs 1994; Oza and Tumer 2008; Schapire 1990; Wang et al. 2009). Two main keys are needed to build a robust EBS: 1) to reach the highest diversity between the EBS elements (*classifiers* and/or *estimators*); and 2) to find the best combination rule of individual elements in such a way that the correct decisions are amplified and incorrect ones are cancelled out (Polikar 2006). However, for reaching the desired diversity, EBSs are usually composed of thousands of elements. This represented a big issue using this technique with remote sensing data. Therefore, researchers often find themselves in uncomfortable compromise between reducing the number of EBS elements, which could also reduce the space of solutions, or developing a robust EBS with a very long computing time. The Gaussian quadrature formula (GQF) could provide an interesting solution to this dilemma. This method, frequently used in uncertainty propagation analysis (Kelly and Krzysztofowicz 1997), has the potential to reduce the number of EBS elements, which require arduous integral computations, into a weighted probabilities summation where the numerical resolution techniques are simpler, and are accurate and approved (Tørvi and Hertzberg 1997).

The objective of this paper was then to optimize the AM performance by developing an EBS based GQF to estimate Chl-a for medium to small-sized inland waters in southern Quebec using MODIS-D-250 images. The EBS was evaluated in three ways: 1) by an adjusted k-fold *cross-validation* using a continuous database; 2) by confusion matrix using an independent ordinal database; and 3) by the coefficient of variation (CV). This latter index is an evaluation of Chl-a modeling reliability for every processed pixel (even outside the *in situ* sampling network). Based on the *cross-validation* and confusion matrix results, a performance comparison analysis between the EBS and the original model (AM) was also done using both databases (continuous and ordinal).

2 STUDY AREA AND IN SITU DATA

2.1 Calibration database

The calibration database is composed by Chl-a concentration measurements collected over 4 water bodies (Missisquoi Bay of Champlain Lake and Brome, Nairne, and William Lakes (Figure 2-1)) by the Quebec's environment ministry called the MDDELCC¹⁷ between 2000 and 2008. These water bodies were chosen because they are subject to cyanobacterial bloom occurrence and recurrence, as observed by the local volunteer monitoring network. Among the objectives of this monitoring network is to identify the main causes leading to the development of HAB in Quebec inland waters for both environmental and public health protection purposes.

Chl-a concentration was quantified following the protocol of the CEAEQ¹⁸ (2012a). The range of the Chl-a concentrations extends from ultra oligotrophic (2.7 mg m⁻³ for Nairne Lake, 2005) to hyper trophic (91,000 mg m⁻³ for Missisquoi Bay, 2003), with an average of 3700 mg Chl-a m⁻³ and a median of 14 mg Chl-a m⁻³. Because of the presence of clouds over some sampled sites or the poor quality of some images (fuzzy images and/or artifacts presence), only 46 of the 363 samples could be used to calibrate the EBS.

¹⁷ Ministère du Développement Durable, Environnement et Lutte contre les Changements Climatiques

¹⁸ Centre d'expertise en analyse environnementale du Québec.



Figure 2-1. Water sampling stations on the four studied lakes used for calibration: 1) Missisquoi bay of Champlain Lake, 2) Brome Lake, 3) William Lake, and 4) Nairne Lake.

2.2 Validation database

22 inland water bodies located between 44° and 50° north and 67° and 80° west and covering a large part of southern Quebec, Canada (Figure 2-2), were used for the validation step. These lakes and reservoirs were also been chosen because of their occurrence and recurrence to cyanobacterial blooms as stated by the MDDELCC's local volunteer monitoring network. This database is independent from the one used to calibrate the EBS, and it is composed by three classes of cyanobacteria density: 1) densities lower than 20,000 cells mL⁻¹ (assumed to be equivalent to 10 mg Chl-a m⁻³ (Chorus and Bartram 1999)), 2) densities between 20,000 and 100,000 cells mL⁻¹, and 3) densities higher than 100,000 cells mL⁻¹ (assumed to be equivalent to 50 mg Chl-a m⁻³). Cyanobacteria density was quantified following the protocol of the CEAEQ (2012b). For the same aforementioned reasons, only 103 of the 677 sampled measures by the MDDELCC between 2007 and 2010 were used to evaluate the EBS

performance. Tableau 2-1 summarises the retained water bodies, which represent different trophic statues ranging from oligo (< 20,000 cell mL⁻¹) to hypertrophic (> 2,000,000 cell mL⁻¹) as well as different lake sizes ranging from medium to small (starting from 2 km²).



Figure 2-2. Geographic location of the water bodies used for model calibration and validation.

Laba nama	Surface	Cyanobacteria	density min/max ann	ual counting	(cell mL ⁻¹)
Lаке пате	(km²)	2007	2008	2009	2010
Adélard	5	******	******	> 20,000	> 20,000
Aylmer	32		1-2,000,000	> 20,000	> 20,000
Missisquoi Bay	47	20,000-500,000	1-2,000,000	> 20,000	> 20,000
Bouchette	7	******	******	> 20,000	> 20,000
Brome	15	1-2,000,000	50,000-500,000	> 20,000	
Choinière	5	20,000-100,000	1-2,000,000	> 20,000	> 20,000
Des-Commissaires	28	500,000-2,000,000	******	*******	> 20,000
Etchemin	2	******	10,000-500,000	> 20,000	*******
Fréchette	5	******	******	*******	> 20,000
Labrecque	4	*******	1,000-2,000,000	*******	> 20,000
Lovering	5	2,000-2,000,000	******	> 20,000	*******
Mandeville	2	100,000-500,000	100,000-2,000,000	> 20,000	> 20,000
Maskinongé	10	1-2,000,000	1-100,000	> 20,000	> 20,000
Massawippi	19	1-2,000,000	1,000-100,000	> 20,000	******
Nairne	2	10,000-100,000	20,000-2,000,000	> 20,000	*******
Ouareau	13	******	2,000-2,000,000	> 20,000	> 20,000
Perchaude	2	******	20,000-2,000,000	*******	******
Pohénégamook	9	10,000-2,000,000	500,000-2,000,000	> 20,000	> 20,000
Roxton	2	2,000-2,000,000	2,000-2,000,000	> 20,000	> 20,000
Taureau	111	2,000-2,000,000	1-2,000,000	> 20,000	******
Tortue	3	******	100,000-500,000	> 20,000	******
William	5	2,000-2,000,000	2,000-500,000	******	> 20,000

Tableau 2-1. List of lakes and reservoirs used in the validation of the ensemble-based system.

3 MODIS IMAGES PRE-PROCESSING

3.1 Spatial resolution downscaling

The MODIS image Level 1B products are available in HDF format through the NASA's website (http://ladsweb.nascom.nasa.gov/data/search.html). The TERRA MODIS sensor was launched in 2000 on the NASA's Terra earth observation satellite. It operates across a large spectrum, with 36 bands that cover the region from 0.4 to 14.4 µm, with a spatial resolution varying from 250 m to 1 km. Given their higher spatial resolutions, only the first seven MODIS land bands were considered in this study. The first two are already at 250 m, while bands 3 to 7, at 500 m (Tableau 3-1), were downscaled to 250 m spatial resolution using an approach developed at the Canadian center for remote sensing (CCRS). First an adaptive regression and a radiometric normalization are made between bands 3 to 7 and bands 1, 2, and the NDVI using the NASA's MOD02HKM product at 500 m spatial resolution (Equation (3-1)). Thereafter, the NASA's MOD02QKM product at 250 m spatial resolution is used as Equation (3-1) inputs to estimate bands 3 to 7 radiance (Trishchenko *et al.* 2007). To further improve

the visual quality of the downscaled images, they are re-projected from the Sinusoidal to Lambert conical conform projection. This step reduces distortions of the orientation and form for subarctic areas. After those two transformations, both the spatial resolution and the visual quality of MODIS images are enhanced (Figure 3-1).

$$Rad(\lambda)_{i} = a_{0,i} + (a_{1,i}Rad(\lambda)_{1} + a_{2,i}Rad(\lambda)_{2})(1 + a_{3,i}NDVI + a_{4,i}NDVI^{2})$$
(3-1)

where a_0 , a_1 , a_3 , and a_4 are coefficients derived from the adaptive regression, and i = 3:7 represent the index of the estimated radiance of bands 3 to 7 at 250 m spatial resolution.



Figure 3-1. True color composite of St-Jean Lake using: A) MODIS-D-250 images (LCC projection) and B) 500 m spatial resolution MODIS lands bands (Sinusoidal projection).

Tableau 3-1	. Characteristics	of the MODIS	bands used in	the present study.
-------------	-------------------	--------------	---------------	--------------------

Primary use	Band	Bandwidth (nm)	Spatial resolution (m)
Land/Cloud/Aerosol Boundaries	1	620-670	250
	2	841-876	250
Land/Cloud/Aerosol properties	3	459-479	500
	4	545-565	500
	5	1,230-1,250	500
	6	1,628-1,652	500
	7	2,105-2,155	500

3.2 Atmospheric correction

For many applications, especially for multi-temporal analysis, the raw relative pixel (top-of-atmosphere) values of each band have to be corrected for atmospheric effects before the images are processed (Moran et al. 1997). Improper atmospheric correction could lead to significant errors in retrieved reflectance and affect the accuracy of estimations (Rahman 2001). Because of its simplicity and its ability to correct both additive and multiplicative atmospheric effects, the SMAC (Simplified Model for Atmospheric Correction) algorithm was used to correct the MODIS-D-250 bands (Rahman and Dedieu 1994). The advantage of this atmospheric correction model is its position between the more sophisticated deterministic models such as 6S (Second Simulation of the Satellite Signal in the Solar Spectrum (Vermote et al. 1997)), MODTRAN (Moderate Resolution Atmospheric Transmittance and Radiance Code (Berk et al. 1999)), and ATCOR (ATmospheric CORrection (Richter 1996)) and the image-based ones like DOS (Dark Object Subtraction (Chavez 1988)) and COST (COSine Transmission for atmospheric correction (Chavez 1996)). Because the more sophisticated deterministic models assume that aerosol retrieval over water bodies in the NIR region is equal to zero of reflectance (Gordon 1978; Gordon and Wang 1994; Viollier et al. 1980)), their estimate of reflectance are incorrect in the presence of absorbing aerosols (Bailey and Werdell 2006; Gordon 1997) and in inland turbid waters (Dekker et al. 1997; Lavender et al. 2005; Morel and Bélanger 2006), leading to considerable errors. Image-based models, on the other hand, only correct for the additive effect caused by haze. Certainly, they reduce atmospheric influences, but the estimated reflectance is still affected by the so-called multiplicative effects caused by ozone, water vapor, and aerosols, specifically in the blue region of the spectrum. While for the SMAC algorithm, it only uses a set of relatively simple equations that requires 7 input parameters: top of atmosphere reflectance, solar zenith angle, view zenith angle, and relative azimuth angle for additive effect correction and aerosol optical depth at 550 nm, water vapor content of the atmosphere, and the ozone content of the atmosphere for multiplicative effect correction. These input parameters are directly derived from MODIS data itself. All images pre-processing (downscaling, re-projection,

and atmospheric correction) was performed using an automatic tool developed by the CCRS.

3.3 Land/water mask

Among the most important steps for building a correct calibration database for remote sensing studies is the use of a good land/water mask, which has the ability to discriminate both the land and mixed pixels of pure water pixels. The 250 m spatial resolution NASA's land/water mask (MOD44W005), freely available on their ftp server (http://e4ftl01.cr.usgs.gov/MOLT), presents some issues in detecting mixed pixels (Figure 3-2.A and B). This represented a serious problematic for building the EBS calibration database due to the erroneous reflectance, picked up over those pixels (usually very high), mainly controlled by the terrestrial vegetation surrounding water bodies. Thus, it has been necessary to develop a specific land/water mask for the MODIS-D-250.

The MODIS-D-250 near infrared (NIR) band could be an interesting solution due to the existing contrast between water absorption and the surrounding vegetation reflectance in this spectral region. However, generating a land/water mask using a single band would always be tainted by errors, especially in algal blooms conditions and at the presence of aquatic plants. Thus, the idea was creating a NIR composite image composed by the minima of more than 1600 MODIS-D-250 NIR bands. In this way, the reflectance of the pure water pixels and temporary vegetation (algal blooms and aquatic plants) within water bodies would be low, while the reflectance of the permanent vegetation surrounding water bodies would be high. The Figure 3-3 shows the NIR composite image reflectance histogram that clearly shows three different reflectance groups. The lowest values correspond to pure water pixels, the highest values correspond to land pixels, and pixels between both groups correspond to the mixed pixels. It is clear that reflectances below 10% correspond to pure water pixels. By using this threshold, it has been possible to create a land/water mask, with no mixed pixels issues, where the result is presented on the Figure 3-2.D.

Chapitre-4



Figure 3-2. Comparison between land/water masks covering the Missisquoi Bay of Champlain Lake produced by: A) the original 250 m spatial resolution NIR MODIS band, B) the land/water mask produced by the NASA, C) the MODIS-D-250 NIR band, and E) the proposed land/water mask.



Figure 3-3. The reflectance histogram of the NIR composite image.

4 THE ADAPTIVE MODEL

With the characteristics advancement of the new generation of sensors (spatial and spectral resolutions), Chl-a modeling has known a significant improvement either for offshore, coastal, or inland waters during the past two decades. Since, several empirical and semi-empirical models were been developed and most of them assumed that all Chl-a concentration levels in a water bodies can be modeled using the same function (Ahn and Shanmugam 2006; Becker *et al.* 2009; Gons 1999). However, the spectral response of a given pixel is controlled by the concentration of all optically active components in the water body (Chl-a of algae, total suspended solids, and colored dissolved organic matters). Consequently, the calibration function of a given component may differ from one concentration level to another (El-Alem *et al.* 2012; Moses *et al.* 2009; Moses *et al.* 2012; Yacobi *et al.* 2011). Indeed, Recent research has demonstrated that, unlike estimation of moderate and high Chl-a concentrations, which are often based on exponential functions (El-Alem *et al.* 2012), estimation of low Chl-a concentrations are more accurate when estimated using linear (Moses *et al.* 2009) or polynomial (Yacobi *et al.* 2011) functions.

Based on these findings, an adaptive model (AM) has been developed, which is composed of a *classifier* and three specific *estimators* (explicitly calibrated for waters poorly, moderately, or highly loaded in Chl-a). The AM *classifier* was trained using the CART (Classification and Regression Tree) method. Once trained, it was possible to split the calibration database into the three aforementioned specific classes. These latter were thereafter used to calibrate the three specific *estimators* of the AM, using a *stepwise* regression (Figure 4-1). As expected, calibration functions for waters moderately and highly loaded in Chl-a were exponential, while the calibration function for waters poorly loaded in Chl-a was linear (El-Alem *et al.* 2014). Thus, the Chl-a modelization using the AM is carried out in two steps: 1) classification of the pixel to be processed into one of the three blooming classes using the AM *classifier* and 2) estimation of the Chl-a concentration of the classified pixel using the corresponding specific *estimator* (Figure 7 of El-Alem *et al.* (2014)).



Figure 4-1. Operating mode of the adaptive model (AM). The subgroups of data surrounded by the black, green, and blue dashed lines represent the training database used to calibrate the *estimators* designed to estimate high, moderate, and low chlorophyll-a concentrations, respectively. Boxes 1 to 3 indicate possible scenarios to estimate chlorophyll-a using the AM. V1 and V2 are detailed in Appendix 1

When compared to classical models (floating algae index (Hu 2009), Kahru approach (Kahru *et al.* 2004), and APProach by ELimination (El-Alem *et al.* 2012)), the AM presented the best results and a significant improvement of Chl-a estimates, particularly for low-to-moderate concentrations, which were a real problematic for classical models. On the other side, based on a comparison analysis made by an ordinal database, the AM has shown some issues in classifying Chl-a estimates for the moderate class (commission error about 69%) also a staircase effect was sighted when applying the AM on MODIS-D-250 images. The low performance of the AM at those blooming conditions was related to the instability of the AM *classifier*, which has a significant influence on Chl-a estimation accuracy (El-Alem *et al.* 2014).

5 METHODOLOGICAL APPROACH AND EVALUATION INDICES

5.1 Ensemble-based system parameterization

To reach an optimal Chl-a estimate using the AM, tow uncertainty sources must be controlled (classification and estimation). This was possible by developing an ensemble-based system (EBS), composed by an ensemble-based classifier (EBC) and an ensemble-based estimator (EBE), by the following steps:

Classification uncertainty

The source of this error is linked to the AM *classifier* nodes (red lines computed from V_1 and V_2 (Figure 4-1)). The proposed thresholds by the CART method are optimal for the used training database, but not unique. A simple modification into this training database can lead to radical changes in the decisions tree. An interesting way to control this uncertainty is to quantify it and to take it into consideration when estimating Chl-a concentrations. The quantification of the AM classifier uncertainty was possible by using the *bagging* algorithm (n-resampling with replacement; n_{bagg} was set to 25,000 for this study), which is one the most popular algorithms used for building an ensemble system (Polikar 2006). This allowed generating random-vectors (v_1 and v_2) composed by thousands $\left(\frac{n_{bagg}}{N_{occur}V1} \times \frac{n_{bagg}}{N_{occur}V2}\right)$ of classifiers that form the so-called EBC (with $N_{occur}V1$ and $N_{occur}V2$ are respectively the thresholds occurrence number within v_1 and v_2). The occurrence and variance of those thresholds have allowed thereafter determining a thresholds probability distribution on both AM classifier nodes (Figure 5-1), that are characterized by a mean (μ) and a variance (σ). Based on those statistical moments, it was possible to quantify classification uncertainty by the following equations:

$$\mu = \int_{v \in V} f(v) \cdot P(v) dv$$

$$\sigma = \int_{v \in V} (f(v) - \mu)^2 \cdot P(v) dv$$
(5-1)
(5-2)

where v is the *random-vector* belonging to V that represents the space of the model input variables, f(v) is the output of the model, and P(v) is the conditional distribution of input variables.

However, the **take into consideration** of this large number of thresholds composing the EBC for making a decision will certainly require a huge computing time. Tørvi and Hertzberg (1997) have proposed an approach based on the Gaussian quadrature formula (GQF), which has the potential to convert those probabilistic integrals (Equations (5-1) and (5-2)) into weighted summations that are in function of n-nodes (n-optimal thresholds (n_{OT}), which was set to 3 in our study) of the original distribution. Thus, Equations (5-1) and (5-2) take the following forms:

$$\mu = \sum_{i=0}^{n_{OT}} \omega_i \times f(\mathbf{z}_i)$$
(5-3)

$$\sigma = \sum_{i=0}^{\infty} \omega_i \times f(z_i - \mu)^2$$
(5-4)

and,

$$OT_i = \mu + \sqrt{\sigma} \times z_i \tag{5-5}$$

where: μ and σ are respectively the mean and variance of the standardized *random-vector* f(z), z_i and ω_i are respectively the abscissas and weights related to each optimal threshold (OT_i; (i = 1: n_{OT})). Tableau 5-1 summarizes the abscissas and weights related to each OT as proposed in the paper of Tørvi and Hertzberg (1997), where mathematical details of the GQF demonstration and its validation can also be found.

Tableau 5-1. Abscissas and weights for the standard normal distribution.

Nodes (OT)	z _i	ω _i
1	0	1
2	-1, +1	1 1
3	$-\sqrt{3}, 0, +\sqrt{3}$	$\frac{\overline{2}, \overline{2}}{1, 2, 1}$ $\frac{1}{6}, \frac{2}{3}, \frac{1}{6}$

• Estimation uncertainty

Due to the complexity of the phenomenon to be modeled, averaging several Chl-a estimates will reduce the risk of using a single estimation function, which could produce biased estimates. Starting from this principle, and based on the results of the EBC, it was possible to build an EBE. In fact, the EBC has permitted to divide the calibration database into 21 subgroups: $3 (n_{OT})$ subgroups of high Chl-a concentrations, which are composed of samples where the spectral index is higher than the V_1 OTs (lower, nominal, and upper); $9 n_{OT}^2$ subgroups of moderate Chl-a concentrations, which are composed of samples where spectral indices are both higher than the V_2 OTs and lower than the V_1 OTs; and 9 subgroups of low Chl-a concentrations, which are composed of samples where spectral indices are both lower to the V_1 and V_2 OTs (Figure 5-1). By means of a *stepwise* regression using all the ratios and band subtractions possibly related to the bio-optical activity of Chl-a (Appendix 1), and a set of algorithms widely used in the literature for inland water (as evaluated by El-Alem *et al.* (2012)), it was possible to calibrate 21 specific *estimators* composing the EBE.

By using this scheme, Equations (5-1) and (5-2) took the forms of the following generic equations:

$$\mu = \sum_{i=0}^{p} \omega_{i} \times \operatorname{Est}_{OTV1_{i}^{\dagger}} + \sum_{i=0}^{n_{OT}-p} \omega_{i} \times \left[\sum_{j=0}^{k} \omega_{j} \times \left[\operatorname{Est}_{OTV2_{j}^{\dagger};OTV1_{i}^{\dagger}} \right] + \sum_{j=0}^{OT-k} \omega_{j} \times \left[\operatorname{Est}_{OTV2_{j}^{\dagger};OTV1_{i}^{\dagger}} \right] \right]; \text{ where } \omega_{0} = 0$$

$$\sigma = \sum_{i=0}^{p} \omega_{i} \times \left[\operatorname{Est}_{OTV1_{i}^{\dagger}} - \mu \right]^{2} + \sum_{i=0}^{n_{OT}-p} \omega_{i} \times \left[\sum_{j=0}^{k} \omega_{j} \times \left[\operatorname{Est}_{OTV2_{j}^{\dagger};OTV1_{i}^{\dagger}} - \mu \right]^{2} + \sum_{j=0}^{n_{OT}-k} \omega_{j} \times \left[\operatorname{Est}_{OTV2_{j}^{\dagger};OTV1_{i}^{\dagger}} - \mu \right]^{2} \right]$$

$$(5-7)$$

where i and j are the corresponding indices to a particular OT (i = 1: n_{OT} and j = 1: n_{OT} and where 1 refers to lower (L), 2 refers to nominal (N), and 3 refers to upper (U) threshold); ω_i and ω_j are the associated weights to each particular OT for V1 and V2 ($\omega_1 = \omega_3 = \frac{1}{6}$ and $\omega_2 = \frac{2}{3}$); V1 and V2 are the discrimination variables, as computed in Appendix 1; OTV1_i and OTV2_j are the optimal thresholds computed by the GQF for V1 and V2 (Equation (5-5)); Est_{OTV1_i} is the set of *estimators* (n_{OT}) trained by the subgroups of high Chl-a concentrations; Est_{OTV2_i;OTV1_i¹ is the set of *estimators* (n_{OT}^2) trained by the}

subgroups of moderate Chl-a concentrations; $Est_{OTV2_{j}^{\downarrow};OTV1_{1}^{\downarrow}}$ is the set of *estimators* (n_{OT}^{2}) trained by the subgroups of low Chl-a concentrations; and k and p are the indices of the OTs related to V1 and V2, respectively ($k \le n_{OT}$ and $p \le n_{OT}$).



Figure 5-1. The use Scheme of the Gaussian quadrature formula on 2 dimensions using two variables (V1 and V2) and its application to calibrate the ensemble-based system (EBS). Red, blue, and green lines represent respectively the optimal thresholds (nominal, upper, and lower) for V1 (TV1_N, TV1_U, and TV1_L) and V2 (TV2_N, TV2_U, and TV2_L). k_(1:3) and p_(1:3) are the indices of the optimal thresholds related to V1 and V2. The surrounded data subgroup by the blue ellipse represents the training data used to calibrate the specific *estimator* designed to estimate high chlorophyll-a concentrations (Chl-a), and the surrounded data subgroup by the red ellipse represents the training data used to calibrate the specific estimator designed to estimate high chlorophyll-a concentrations. Boxes 1 to 13 indicate the possible modeling cases to estimate Chl-a using the EBS.

Thus, the developed EBS in this study is a combination of an EBC and an EBE. By combining the results of these two ensembles, it was possible to reach a good level of diversity for both *classifiers* and *estimators*, since the final ChI-a estimate is made by the decision of <u>at least</u> 3 *classifiers* and 3 *estimators*, depending on the blooming class to be modeled (Equation (5-6)). Also, the modeling space is refined. Instead of being based on 3 modeling cases (Figure 4-1), 13 modeling cases were identified (from very high to very low ChI-a concentrations (Figure 5-1)) based on the GQF results. This would allow a gradual transition between the modelled blooming classes when using the EBS on MODIS-D-250 images and could also reduce the staircase effect sighted when using the AM.

For example, if we take the case of the blue star on the Figure 5-2, p = 2 and k = 3 in this case. Thus, Equations (5-6) and (5-7) will take the forms:

$$\mu = \frac{1}{6} \times \text{Est}_{\text{TV1}_{L}^{\uparrow}} + \frac{2}{3} \times \text{Est}_{\text{TV1}_{N}^{\uparrow}} + \frac{1}{6} \times \left[\frac{1}{6} \times \text{Est}_{\text{TV2}_{U}^{\uparrow};\text{TV1}_{U}^{\downarrow}} + \frac{2}{3} \times \text{Est}_{\text{TV2}_{N}^{\uparrow};\text{TV1}_{U}^{\downarrow}} + \frac{1}{6} \times \text{Est}_{\text{TV2}_{L}^{\uparrow};\text{TV1}_{U}^{\downarrow}}\right]$$

$$\sigma = \frac{1}{6} \times \left(\text{Est}_{\text{TV1}_{L}^{\uparrow}} - \mu_{1}\right)^{2} + \frac{2}{3} \dots + \frac{1}{6} \times \left[\frac{1}{6} \times \left(\text{Est}_{\text{TV2}_{U}^{\uparrow};\text{TV1}_{U}^{\downarrow}} - \mu_{1}\right)^{2} + \frac{2}{3} \dots + \frac{1}{6} \times \left(\text{Est}_{\text{TV2}_{L}^{\uparrow};\text{TV1}_{U}^{\downarrow}} - \mu_{1}\right)^{2}\right]$$

where μ is the weighted mean of the Chl-a estimate, and σ is its variance, $\operatorname{Est}_{TV1_L^{\uparrow}}$ and $\operatorname{Est}_{TV1_N^{\uparrow}}$ are respectively specific *estimators* calibrated with training database subgroups where the spectral index is superior to lower and nominal OTs of V1 and $\operatorname{Est}_{TV2_U^{\uparrow};TV1_U^{\downarrow}}$, $\operatorname{Est}_{TV2_N^{\uparrow};TV1_U^{\downarrow}}$, and $\operatorname{Est}_{TV2_L^{\uparrow};TV1_U^{\downarrow}}$ are specific *estimators* calibrated with training database subgroups where spectral indices are both inferior to the upper OT of V1 and superior to upper, nominal, and lower OTs of V2 (Figure 5-2). This combination of specific *estimators* is not unique, because it varies with the blooming class to be modelled (depending on p and k values).



Figure 5-2. Example of subgroups selection samples (black points) used to calibrate the specific *estimators* for the case of the blue star. Arrows indicate whether the subgroup selection is made before (spectral index of samples is inferior) or after (spectral index of samples is superior) a given optimal threshold, whether for V1 or V2.

5.2 Accuracy assessment

The EBS was evaluated by a k-fold *cross-validation*, in which a block of samples is temporarily removed from the database and the remaining samples are then used as training data to estimate the removed samples with the pre-calibrated sub-EBS. This exercise is remade for the whole of the calibration database. In order to ensure a representative selection of all ChI-a concentration classes within each removed block, to induce a maximum of instability in the EBS training parameters, and to leave enough samples to calibrate the sub-EBS, we ensured to select at least one sample from each blooming class in the removed bloc, as described at the Figure 5-3. Four statistical indices (Equations (5-8) to (5-11)) were thereafter used to assess the EBS: the coefficient of determination (R²), relative bias (BIASr), relative root mean square error (RMSEr), and relative NASH-Sutcliffe efficiency (NASHr). The NASH criterion evaluates performance based on the estimated values and the *in situ* measurement average. A negative NASH value means that it would be better to use the measurements average rather than the model's *estimations*; a NASH value equal to 1.0 means that the model is perfect. The mathematical equations of the used statistical indices are as follows:

$$R^{2} = \left[\frac{\sum_{i=1}^{n} (O_{i} - \overline{O})(Es - \overline{Es})}{\sqrt{\sum_{i=1}^{n} (O_{i} - \overline{O})^{2}} \sqrt{\sum_{i=1}^{n} (Es_{i} - \overline{Es})^{2}}}\right]^{2}$$
(5-8)

$$BIASr = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{Es_i - O_i}{Es_i} \right)$$
(5-9)

RMSEr =
$$\sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(\frac{Es_i - O_i}{Es_i}\right)^2}$$
 (5-10)

$$Nr = 1 - \frac{\sum_{i=1}^{n} \left(\frac{O_{i} - E_{i}}{O_{i}}\right)^{2}}{\sum_{i=1}^{n} \left(\frac{O_{i} - \overline{O}}{\overline{O}}\right)^{2}}$$
(5-11)

where n is the sample size, O and Es are the observed and estimated values, and $\overline{0}$ and \overline{Es} are the means of the observed and estimated values.



Figure 5-3. K-fold cross-validation adopted in this study.

The EBS performance was also tested by an independent validation database. The performance assessment using this database was made by a confusion matrix, enabling the computation of omission and commission errors, success rates, and the Kappa index. The Kappa index allows quantifying the concordance level between

measured and estimated Chl-a classes. The concordance is weak when the Kappa value is negative, whereas it is good when the Kappa is higher than 0.6 and excellent when it is above 0.8. Tableau 5-2 summarizes the parameters involved in the confusion matrix and Kappa index computation.

Tableau 5-2. Simplified diagram of the parameters used in the confusion matrix and to calculate the Kappa index: a and d are the number of well-classified values, b and c are the number of mis-classified values, n₁, n₂, n₃, and n₄ are respectively totals of a + b, c + d, a + c, and b + d, N is the sample size, $P_o = \frac{a+d}{N}$, and $P_c = \frac{n_1 \times n_3 + n_2 \times n_4}{N^2}$.

	Measured									
Estimated		TRUE	FALSE	Total	Success Rate (%)	Commission Error (%)				
	TRUE	а	b	n ₁	(a/n ₁)	(b/n ₁)				
	FALSE	С	d	n ₂	(b/n ₂)	(c/n ₂)				
	Total	n ₃	n ₄	Ν						
	Success Rate (%)	(a/n ₃)	(d/n ₄)							
	Omission Error (%)	(c/n ₃)	(b/n ₄)							
	Global Success (%)					(a + b)/N				

Based on Equations (5-6) and (5-7), it is possible to calculate the coefficient of variation (CV) as a qualitative evaluation of Chl-a estimate using the following formula:

$$CV(\%) = \frac{\sqrt{\mu}}{\sigma} \times 100 \tag{5-12}$$

The CV is a measure of the homogeneity. In other words, it measures the dispersion of individuals of a population with respect to their average. According to Martin and Gendron (2004), when the CV is lower than 16% the mean can be considered a reliable predictor; when it is between 16% and 33.3% the mean may include some errors; and when it is higher than 33.3% the dispersion is very important and the mean is no longer reliable.

6 RESULTS AND DISCUSSION

6.1 Calibration of the specific estimators

It was recently demonstrated that use of a multivariate regression for estimating Chl-a concentration significantly enhanced the accuracy of estimates (Hodjaty *et al.* 2010). Thus, the 21 EBS specific *estimators* were calibrated using the 25 explanatory variables presented in the Appendix 1 table, by means of a *stepwise* regression. The Appendix 2 table shows the calibration characteristics of the specific *estimators* (coefficients of calibration and intercepts, explanatory variables used to train each *estimator*, coefficients of determination (R^2), size of the sub-training database, and the explanatory variable with the best p-value (BpV)). The correlation between spectral indices and the *in situ* Chl-a measurements ranged from modest ($R^2 = 0.40$ for specific *estimator-5*) to high ($R^2 = 0.96$ for specific *estimator-1*). It is also important to underline that the correlation decreases with low concentrations and specific *estimators* designed to estimate low Chl-a of concentrations presented best correlations with a linear functions, while specific *estimators* designed to estimate moderate and high Chl-a were trained using exponential functions (Appendix 2).

According to the explanatory variables with the B*p*V, we can conclude that Chl-a estimate for high blooming classes is mainly controlled by the NIR₁ (859 nm), for moderate blooming classes Chl-a estimate is mostly controlled by visible to NIR₁, and for low blooming classes Chl-a estimate is shred between the visible and NIR (NIR₂-NIR₁). These results are quite logical. In fact, it is known that the presence of algal blooms generates turbid waters, but at very high blooming conditions, where the blooms usually form scums at the surface of the water, it is the high reflectance of the Chl-a at the NIR₁ that is captured by sensors more strongly than the turbidity created below this layer of algae. For moderate Chl-a concentrations, the two most used variables are V7 ($R(\lambda_{b4}) - R(\lambda_{b3})$) and V9 ($R(\lambda_{b2}) - R(\lambda_{b3})$). Those latter would be directly related to the bio-optical activity of phytoplankton blooms. Indeed, both computed spectral contrasts increase with phytoplankton density in the water body due to the high absorption of phytoplankton Chl-a in the blue ($R(\lambda_{b3})$) and its high

reflectance in the green $(R(\lambda_{b4}))$ and NIR $(R(\lambda_{b2}))$. However, for the specific estimator-4, which represents the transition class from high to moderate concentrations, it was trained by V17 $\binom{R(\lambda_{b2})}{R(\lambda_{b1})}$. This variable would be more related to phytoplankton turbidity than to their bio-optical activity since reflectances from 620 to 705 nm show a good correlation with turbid waters (Ouillon et al. 2008). For low Chl-a concentrations, two variables controlled phytoplankton detection, which are V11 $(R(\lambda_{b2}) - R(\lambda_{b5}))$ and V8 $(R(\lambda_{b4}) - R(\lambda_{b1}))$. The V8 was selected by the *stepwise* regression as the most correlated spectral index with phytoplankton for very low concentrations, which is logical as at these blooming conditions no turbidity is made by phytoplankton and consequently Chl-a detection is controlled more by the bio-optical activity of this pigment. By contrast, the use of V11 to train low class specific *estimators* may appear incoherent, but its presence maybe explained by some high values of Chl-a concentrations (up to 18 mg m⁻³) in subgroups used to train those specific *estimators*. At these concentrations, phytoplankton Chl-a still highly reflect at the NIR₁ ($R(\lambda_{b2})$), and, on the other hand, even the very turbid waters strongly absorb at the NIR₂ ($R(\lambda_{b5})$). This contrast makes it easier to distinguish between Chl-a and water pixels in this part of the spectrum.

6.2 Ensemble-based system quantitative evaluation

6.2.1 The k-fold cross-validation

The cross-validation results were interesting with a coefficient of determination of 0.98 indicating that the EBS can explain up to 98% of the Chl-a variance. The NASHr, which is a severe evaluation index, indicates that the model is robust, with a success rate of 95%. The robustness of the EBS was also confirmed by the scatter plot of *in situ* measurements versus their estimates (Figure 6-1), which shows that all points are well distributed with respect to the 1:1 line. The EBS performed better than the AM based on all statistical indices, particularly for the NASHr criterion and the RMSEr (a decrease by 35%). Nevertheless, testing the performance of the models at lower blooming conditions (< 50 mg Chl-a m⁻³) is of greater interest, since detecting water bodies that present a potential of developing algal blooms can be more useful for water managers than

detecting waters where HAB are already pre-established. The analysis shows a significant improvement of Chl-a estimates for the EBS, especially with respect to the NASHr criterion, which jumps from 0.24 to 0.70 (an enhancement of 66%). Moreover, the BIASr is decreased by almost 50% and the RMSEr by 38% (Figure 6-1). These results demonstrate clearly the value of estimating Chl-a by several specific *estimators* trained from different spectral regions. Averaging *estimators* decisions have thus lead to reach a good quality of Chl-a estimates, specifically for moderate-to-low concentrations where the modelization is done by at least five specific *estimators*' decisions.



Figure 6-1. Chlorophyll-a concentration estimated by the two models compared to *in situ* measurements for the complete database, with the model performance indices.



Figure 6-2. Chlorophyll-a concentration estimated by the two models compared to *in situ* measurements for the database only including values < 50 mg Chl-a m⁻³, with the model performance indices.

6.2.2 Confusion matrix: (independent database)

The Kappa index and the global success of both confusion matrices (Tableau 6-1 and 6-2) have respectively jumped, in favour of the EBS, from 0.51 to 0.71 and from 67% to 82%. These results underline the good concordance between observed Chl-a classes (low, moderate, and high) and the corresponding Chl-a classification estimated by the EBS (Kappa > 0.7). As expected, the improvement in the Chl-a estimate concerned the moderate blooming class (a decreased of the commission error from 61% to 35%), specifically for waters poorly-to-moderately loaded in Chl-a. The maximum gain of the accuracy is original of waters poorly loaded in Chl-a (i.e. a decrease from 14 (Tableau 6-2) to 6 (Tableau 6-1) mis-classified Chl-a estimates for the moderate class and an increase from 30 to 39 well-classified Chl-a estimates for the low class). This is a good demonstration of how EBSs can strongly control and mitigate classification uncertainties, even for the moderate class that is very sensitive to mis-classifications since it can occur at both ends of the class (El-Alem *et al.* 2014).

Tableau 6-1. The ensemble based system confusion matrix results. (False positives and negative were calculated using red and brown numbers, respectively).

	Measured								
		[Chl-a]<10	10<[Chl-a]<50	[Chl-a]>10	Total	Commission Error	Success Rate		
	[Chl-a]<10	39	3	4	46	15%	85%		
ğ	10<[Chl-a]<50	6	17	3	26	35%	65%		
stimate	[Chl-a]>10	1	2	28	31	10%	90%		
	Total	46	20	37	103	****	****		
ш	Omission Error	15%	15%	24%	****	****	****		
	Success Rate	85%	85%	76%	*****	****	****		
	Global Success	****	****	****	****	****	82%		
	Kappa index	****	****	****	*****	****	0.71		

False positives for waters poorly loaded in Chl-a (4 of 46 low concentrations classified as high (Tableau 6-1)) decreased from 11% to 9% and for waters moderately charged in Chl-a (3 of 26 moderate concentrations classified as high; Tableau 6-1) also decreased from 22% to 12%, at the benefit of the EBS, counting then a total hazard of 21% for a declaration of false blooms versus 33% for the AM. Moreover, false negative for waters highly charged in Chl-a were of 10% (1+2 of 31 high concentration classified as

moderate and low, respectively (Tableau 6-1)) versus 17% for the AM, recording thus the best performance in terms of non-declaration of an HAB.

		[Chl-a]<10	10<[Chl-a]<50	[Chl-a]>10	Total	Commission Error	Success Rate
stimated	[Chl-a]<10	30	3	4	37	19%	81%
	10<[Chl-a]<50	14	14	8	36	61%	39%
	[Chl-a]>10	2	3	25	30	17%	83%
	Total	46	20	37	103	****	****
ш	Omission Error	35%	30%	32%	*****	****	****
	Success Rate	65%	70%	68%	*****	****	****
	Global Success	****	****	****	*****	****	67%
	Kappa index	****	****	****	*****	****	0.51

 Tableau 6-2. The adaptive model confusion matrix results. (False positives and negative were calculated using red and brown numbers, respectively).

6.2.3 Ensemble-based system qualitative evaluation

In order to qualitatively compare the performance of the EBS to the original model, both were applied to a bloom detected at the Missisquoi Bay of Lake Champlain (Figure 6-3). A clear correspondence between the bloom shape drawn by the EBS (Figure 6-3.B) and the HAB extension detected on the MODIS true color image (green hue (Figure 6-3.A)) can be seen for heavy bloom conditions (the northeast to east part of the bay) and for low-to-moderate blooming conditions (central part of the Bay). This concordance is much less clear for the AM, especially for low-to-moderate concentrations (Figure 6-3.C), where a staircase effect can be sighted. This result was somewhat expected because of the general conceptualization of the AM based on a local and an unstable *classifier* and the use of a single specific *estimator* for making Chl-a estimates. By contrast, the EBS is based on a more general and stable *classifier* (ensemble based classifier) and it is based not only on a single adequate specific *estimators*, but on a combination (at least 3; Equation (5-6)) of many specific *estimators* trained by different spectral regions (from visible to short waves of infrared). This was well illustrated in the central part of Missisquoi Bay, where the EBS has shown a great potential in smoothing

the enormous contrast between high $(max_{est} = 13.629 \ mg \ m^{-3})$ and low $(min_{est} = 0.52 \ mg \ m^{-3})$ Chl-a concentrations modelled, making the modelization more realistic.



Figure 6-3. Comparison between the adaptive model (AM) and ensemble based systems (EBS) application on MODIS images. A) True color, B) EBS, and C) AM (Sept.19, 2001).

Since the Chl-a estimate made by the EBS, for each pixel, is a weighted mean of several specific *estimators*, it is also possible to compute a Chl-a concentration variance, by means of the Equation (5-7). Based on these two statistical moments (mean and variance), it is then possible to compute a coefficient of variance (CV), which is an interesting evaluation index capable of measuring the modelization reliability, for every processed pixel, as described at «<u>5.2. accuracy assessment</u>» section. Thus, results show that Chl-a variance increases with higher concentrations, with a maximum

of about 2.45 mg m⁻³. For the low-to-moderate Chl-a estimate, the variance was almost constant varying between 1 and 1.2 mg m⁻³ (Figure 6-4.A). These results may appear inconsistent with the above findings since the performance of the EBS is supposed to be higher for high Chl-a concentrations. However, it is important to highlight that the statistical indices that were used (RMSEr, BIASr, and NASHr) provide relative results, while the calculated variance (σ) is an absolute measure. Since this latter only quantifies the absolute difference, over- or under-estimations of high Chl-a concentrations have greater influence than those of lower ones. Nevertheless, the accuracy of the EBS in the high blooming conditions was higher as the CV was almost null. For the rest of Missisquoi Bay, the CV was generally lower than 10% except for some southern areas where it reached 90% on average (Figure 6-4.B). These results lead to conclude that the EBS Chl-a estimate can be considered to be reliable most of the time for all blooming classes.



Figure 6-4. Evaluation indices of Chl-a estimate using the ensemble based system on the algal bloom detected on Sept.19, 2001. A) Variance and B) Coefficient of variation (CV).

6.2.4 Comparison with MODIS and MERIS chlorophyll-a standard products

In order to compare the EBS Chl-a product to the NASA's and ESA's Chl-a standard products, a qualitative evaluation was done. Thus, the Figure 6-5.B shows the Chl-a EBS-derived MODIS-D-250 product and the Figure 6-5.C shows the NASA's MODIS Chl-a product on a bloom detected in Lake Erie on Oct. 8, 2011. This latter is obtained by a fourth-order polynomial relationship between ratio of MODIS reflectance and *in situ*

Chl-a measurements (Equation (6-1)). This product is available on the NASA's website: http://oceancolor.gsfc.nasa.gov/. The Figure 6-5.A, is the MODIS-D-250 true color image. Two important points have led to choose this example: 1) to test the performance of the EBS on Lakes that represent physiographic characteristics completely different from those used for its calibration, and 2) the presence of two different kinds of panache on the same image (sediments in the south of Lake Huron and algal bloom in Lake Erie, as stated by the NASA's Earth Observation: http://earthobservatory.nasa.gov/IOTD/view.php?id=76115).

$$log_{10}(Chl - a) = a_0 + \sum_{i=1}^{4} a_i \log_{10} \left(\frac{R(\lambda_{blue})}{R(\lambda_{green})}\right)^i$$
(6-1)

where the numerator, $R(\lambda_{blue})$, is the greatest of several input reflectance values and the coefficients, $a_0 - a_4$, are empirically computed.

First, it is clear that visual quality of the EBS is by far better than of the MODIS standards product. When compared to the true color image, the Chl-a EBS-derived MODIS-D-250 seems to well copy the bloom in terms of spatial distribution and intensity. By contrast, the MODIS standard product seems to over-estimate Chl-a concentrations at a point to outdo its computation field, when the bloom was at its maximal intensity. Consequently, no Chl-a estimates were done over this area of the algal bloom. On the other hand, Chl-a estimates for Lake Huron, where a huge panache of sediment can be seen, are at the same level than those of Lake Ontario and the not affected part by the algal bloom of Lake Erie. This is a good demonstration of the EBS ability for modeling solely phytoplankton Chl-a concentrations, even at the presence of other bio-optical elements. This is not surprising indeed, it is known that waters loaded in Chl-a strongly absorb in blue and red, and waters loaded in sediments strongly reflect in those spectral regions, especially in the blue. Thus, the green/red or /blue and NIR/red or /blue contrasts, used to train the EBS specific estimators (Appendix 1), increase when phytoplankton is dominating water body and vice versa. On the hand, NIR bands are not sensitive to sediments, but they are for phytoplankton Chl-a in high concentrations. Thereby, use of this set of band combinations has led to develop

specific *estimators* that are insensitive to none of the other inland water bio-optical elements, but the Chl-a.



Figure 6-5. Chlorophyll-a concentrations (Chl-a) spatial distribution in Lake Erie on Oct. 8, 2011 from the MODIS image downscaled to 250 m spatial resolution (B) and the NASA's MODIS standard product at 4 km spatial resolution (C). Chl-a concentrations are in a log scale. A) is the MODIS true color image at 250 m spatial resolution.

The example at the Figure 6-6 is for a comparison purpose with the full spatial resolution of MERIS product, considered as the best ocean color sensor for the last decade. The image used here (Figure 6-6.C) is published in the Fostering Earth Observation & Global Awareness e-review. Unfortunately, no information about the modeling process was available, except that it was produced by the NOAA's National Centers for Coastal Ocean Science (Richard 2014). When comparing both Chl-a
derived images to the true color image, it is clear that the two products well copy the developed bloom in Lake Erie on Oct. 9, 2011, with a similar visual quality. However, a slight under-estimation of Chl-a can be seen for the Chl-a EBS-derived MODIS-D-250 product. Since no ground truth data are available to statistically evaluate both results, it is impossible to affirm which Chl-a estimate is more accurate.



Figure 6-6. Chlorophyll-a concentrations (Chl-a) spatial distribution in Lake Erie on Oct. 9, 2011 from the MODIS image downscaled to 250 m spatial resolution (B) and the ESA's MERIS standard product at 300 m spatial resolution (C). The (A) image is the MODIS true color image at 250 m spatial resolution. Chl-a concentrations are in a log scale.

The Pymatuning reservoir (red boxes on the Figure 6-6) can be an interesting ground truth data for a qualitative evaluation. In the zoomed box (Figure 6-6.A), it is clear that no algal bloom is developed in this reservoir. However, for the MERIS derived image, Chl-a concentrations exceed the medium (green to red) level, which is stated by the author as higher than the cyanobacteria moderate risk as established by the Word Heath Organization (about 100,000 cells mL⁻¹). In those blooming conditions, waters of the Pymatuning reservoir should also appear green in the true color image, since the

inferior limit of cyanobacterial visual identification is about 20,000 cells mL^{-1} (Lavoie et al. 2007). By contrast, the modeled Chl-a concentrations for the same reservoir using the EBS were below 10 mg m⁻³ (equivalent to 20,000 cells mL^{-1} according to the conversion system made by the Word Heath Organization (Chorus and Bartram 1999)). This results lead to conclude that the MERIS derived Chl-a could be over-estimated.

7 CONCLUSIONS

An ensemble-based system (EBS) for Chl-a concentration estimation using MODIS bands downscaled to 250 m spatial resolution (MODIS-D-250) was developed and tested for inland water bodies in southern Quebec and the great lakes. The EBS general conceptualization was developed in such way to control and to mitigate the adaptive model (AM) uncertainties, specifically those related to its *classifier*. Several innovative points were tested in this work: use of MODIS-D-250 which are originally developed for other studies (land, aerosols, and cloud); use of an ensemble system for a regression purpose in a remote sensing application; and use of the Gaussian quadrature formula for optimizing the computing time. Combination of these points has led to develop a mixed EBS capable of monitoring medium to small-sized lakes, in a near-real-time, and where the Chl-a estimate is done by a set of *classifiers* and *estimators*. In addition, for the first time, it has been possible to generate evaluative raster supports (variance and coefficient of variation (CV) images), annexed to Chl-a estimates.

Several validation techniques were used to assess the performance of the proposed approach: k-fold *cross-validation*; validation with independent ordinal database using a confusion matrix; and qualitative validation by applying the model on several study cases. The *cross-validation* results were satisfactory, especially for low-to-moderate concentrations since evaluation indices have shown a significant improvement ($R^2 = 0.77$ and NASHr = 0.70) compared to those of the original model ($R^2 = 0.56$ and NASHr = 0.24). Validation with the independent database has highlighted the good concordance (Kappa index = 0.71) between observed and estimated blooming classes. This improvement was more tangible for the moderate blooming class, where

232

commission errors decreased by 26% versus 4% and 7% for the low and high blooming classes, respectively. False declaration or non-declaration of a harmful bloom hazards were the best for the EBS with errors lower than 21%. Furthermore, the evaluation by the CV (< 10% for most pixels) underlines the reliability of the Chl-a estimates produced by the EBS for all blooming classes.

The above results demonstrate the robustness of the developed approach, when tested over inland water bodies with similar physiographic characters as Lakes used for its calibration. This robustness was also maintained when the EBS was tested on Lakes with different physiographic characteristics (Erie, Huron, and Ontario), and compared to the NASA's and ESA's Chl-a standard products. This comparative analysis has emphasized the good quality of the Chl-a modeling either visually or in terms of intensity, when it was compared to the Chl-a MODIS's standard product. By comparison to the full resolution of MERIS's Chl-a standard product, both the visual quality and spatial distribution of Chl-a concentration seemed to be similar, with a slight advantage for the EBS when modeling waters poorly-to-moderately loaded in Chl-a. Additional, the EBS has demonstrated a great potential to solely model phytoplankton Chl-a even at the presence of high concentrations of sediments.

By the present paper, we proposed a robust approach for estimating Chl-a as a proxy of phytoplankton growth, in near-real-time, using the fine spatial resolution of MODIS sensor. With the developed water/land mask, we ensured to model pixels of pure water that are off coasts by at least 250 m spatial resolution. At this distance, none of aquatic plants can reach the surface of waters, and consequently the Chl-a modeled by the developed approach using MODIS-D-250 images is of phytoplankton origin. Development of such approaches could be of great assistance to water managers looking to achieve optimal management of water body quality in large areas like the province of Quebec. On the other hand, it could bring a great help to understand phytoplankton behaviour by linking watershed physiographic factors and meteorological conditions that have ruled and have led to the stimulation, development, maintain, and decline of a given algal bloom episode.

233

REFERENCES

- Ahn, Y.H., & Shanmugam, P. (2006). Detecting the red tide algal blooms from satellite ocean colour observations in optically complex Nourtheast-Asia coastal waters. Remote Sensing of Environment, 103, 419-437
- Allan, M.G., Hamilton, D.P., Hicks, B.J., & Brabyn, L. (2011). Landsat remote sensing of chlorophyll a concentrations in central North Island lakes of New Zealand. International Journal of Remote Sensing, 32, 2037-2055
- Allison, E.H., Perry, A.L., Badjeck, M.C., Neil Adger, W., Brown, K., Conway, D., Halls, A.S., Pilling, G.M., Reynolds, J.D., & Andrew, N.L. (2009). Vulnerability of national economies to the impacts of climate change on fisheries. Fish and fisheries, 10, 173-196
- Alpak, F.O., Vink, J.C., Gao, G., & Mo, W. (2007). Techniques for effective simulation, optimization, and uncertainty quantification of the in-situ upgrading process. Journal of Unconventional Oil and Gas Resources
- Bailey, S.W., & Werdell, P.J. (2006). A multi-sensor approach for the on-orbit validation of ocean color satellite data products. Remote Sensing of Environment, 102, 12-23
- Becker, R.H., Sultan, M.I., Boyer, G.L., Twiss, M.R., & Konopko, E. (2009). Mapping cyanobacterial blooms in the Great Lakes using MODIS. Journal of Great Lakes Research, 35, 447-453
- Berk, A., Anderson, G.P., Bernstein, L.S., Acharya, P.K., Dothe, H., Matthew, M.W., Adler-Golden, S.M., Chetwynd, J.H., Richtsmeier, S.C., Pukall, B., & Others (1999). MODTRAN 4 radiative transfer modeling for atmospheric correction. In, Proceedings of SPIE- The International Society for Optical Engineering (pp. 348-353)
- Binding, C., Greenberg, T., Jerome, J., Bukata, R., & Letourneau, G. (2010). An assessment of MERIS algal products during an intense bloom in Lake of the Woods. Journal of Plankton Research, fbq133
- Blondeau-Patissier, D., Gower, J.F.R., Dekker, A.G., Phinn, S.R., & Brando, V.E. (2014). A review of ocean color remote sensing methods and statistical techniques for the detection, mapping and analysis of phytoplankton blooms in coastal and open oceans. Progress in Oceanography, 123, 123-144
- CEAEQ (2012a). (CENTRE D'EXPERTISE EN ANALYSE ENVIRONNEMENTALE DU QUÉBEC). In (p. 16). Mnistère du Développement durable, de l'Environnement et des Parcs du Québec Détermination de la chlorophylle a: méthode par fluorométrie, MA. 800 Chlor. 1.0, Rév. 2
- CEAEQ (2012b). (CENTRE D'EXPERTISE EN ANALYSE ENVIRONNEMENTALE DU QUÉBEC). In (p. 16). Mnistère du Développement durable, de l'Environnement

et des Parcs du Québec Dépistage des cyanobactéries, MA. 800 – Cya.dep 1.0, M

- Chan, J.C.-W., & Paelinckx, D. (2008). Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. Remote Sensing of Environment, 112, 2999-3011
- Chapman, D. (1992). Water Quality Assessments-A Guide to Use of Biota, Sediments and Water in Environmental Monitoring
- Chavez, P.S. (1988). An Improved Dark-Object Substraction Technique for Atmospheric Scattering Correction of Multispectral Data. Remote Sensing of Environment, 24, 459-479
- Chavez, P.S. (1996). Image-based atmospheric correction-Revisited and improved. Photogrammetric Engineering and Remote Sensing, 62, 1025-1036
- Chorus, I., & Bartram, J. (1999). Toxic cyanobacteria in water. A guide to their public health consequences, monitoring and management. WHO. E & FN Spon, London, 416 pp
- Dall'Olmo, G., Gitelson, A.A., Rundquist, D.C., Leavitt, B., Barrow, T., & Holz, J.C. (2005). Assessing the potential of SeaWiFS and MODIS for estimating chlorophyll concentration in turbid productive waters using red and near-infrared bands. Remote Sensing of Environment, 96, 176-187
- Dasarathy, B.V., & Sheela, B.V. (1979). A composite classifier system design: Concepts and methodology. Proceedings of the IEEE, 67, 708-713
- Davis, C.O., Kavanaugh, M., Letelier, R., Bissett, W.P., & Kohler, D. (2007). Spatial and spectral resolution considerations for imaging coastal waters. In, Optical Engineering+ Applications (pp. 66800P-66800P-66812): International Society for Optics and Photonics
- Dekker, A.G., Hoogenboom, H.J., Goddijn, L.M., & Malthus, T.J.M. (1997). The relation between inherent optical properties and reflectance spectra in turbid inland waters. Remote Sensing Reviews, 15, 59-74
- Eija, R., Rvi, R.O., Seija, H.I., Juha-Markku, L.n., & Mika, R. (1998). EVect of sampling frequency on detection of natural variability in phytoplankton: unattended high-frequency measurements on board ferries in the Baltic Sea. Journal of Marine Science, 55, 8
- El-Alem, A., Chokmani, K., Laurion, I., & El-Adlouni, S. (2014). An Adaptive Model to Monitor Chlorophyll-a in Inland Waters in Southern Quebec Using Downscaled MODIS Imagery. Remote Sensing, 6, 6446-6471
- El-Alem, A., Chokmani, K., Laurion, I., & El-Adlouni, S.E. (2012). Comparative Analysis of Four Models to Estimate Chlorophyll-a Concentration in Case-2 Waters Using MODerate Resolution Imaging Spectroradiometer (MODIS) Imagery. Remote Sensing, 4, 2373-2400

- Feng, L., Hu, C., Han, X., Chen, X., & Qi, L. (2014). Long-Term Distribution Patterns of Chlorophyll-a Concentration in China's Largest Freshwater Lake: MERIS Full-Resolution Observations with a Practical Approach. Remote Sensing, 7, 275-299
- Gitelson, A.A., Dall'Olmo, G., Moses, W., Rundquist, D.C., Barrow, T., Fisher, T.R., Gurlin, D., & Holz, J. (2008). A simple semi-analytical model for remote estimation of chlorophyll-a in turbid waters: Validation. Remote Sensing of Environment, 112, 3582-3593
- Gons, H.J. (1999). Optical Teledetection of Chlorophyll a in Turbid Inland Waters. Environmental Science & Technology, 33, 1127-1132
- Gordon, H.R. (1978). Removal of atmospheric effects from satellite imagery of the oceans. Applied Optics, 17, 1631-1636
- Gordon, H.R. (1997). Atmospheric correction of ocean color imagery in the Earth Observing System era. Journal of Geophysical Research: Atmospheres, 102, 17081-17106
- Gordon, H.R., Brown, O.B., Evans, R.H., Brown, J.W., Smith, R.C., Baker, K.S., & Clark, D.K. (1988). A semianalytic radiance model of ocean color. Journal of Geophysical Research: Atmospheres (1984–2012), 93, 10909-10924
- Gordon, H.R., & Wang, M. (1994). Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with SeaWiFS: a preliminary algorithm. Applied Optics, 33, 443-452
- Hansen, L.K., & Salamon, P. (1990). Neural network ensembles. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 12, 993-1001
- Hodjaty, M.M., Saradjian, M.R., & Jamshidi, S. (2010). Chlorophyll a Estimation in Sefidrud Estuarine Using MODIS Images. In, FIG Congress 2010 (p. 11). Sydney, Australia
- Hu, C. (2009). A novel ocean color index to detect floating algae in the global oceans. Remote Sensing of Environment, 113, 2118-2129
- Hu, C., Chen, Z., Clayton, T.D., Swarzenski, P., Brock, J.C., & Muller–Karger, F.E. (2004). Assessment of estuarine water-quality indicators using MODIS mediumresolution bands: Initial results from Tampa Bay, FL. Remote Sensing of Environment, 93, 423-441
- Hu, C., Lee, Z., Ma, R., Yu, K., Li, D., & Shang, S. (2010). Moderate resolution imaging spectroradiometer (MODIS) observations of cyanobacteria blooms in Taihu Lake, China. Journal of Geophysical Research: Oceans (1978–2012), 115
- Ismail, R., & Mutanga, O. (2010). A comparison of regression tree ensembles: Predicting Sirex noctilio induced water stress in Pinus patula forests of KwaZulu-Natal, South Africa. International Journal of Applied Earth Observation and Geoinformation, 12, Supplement 1, S45-S51
- Jacobs, R.A., Jordan, M.I., Nowlan, S.J., & Hinton, G.E. (1991). Adaptive mixtures of local experts. Neural Comput., 3, 79-87

- Jordan, M.I., & Jacobs, R.A. (1994). Hierarchical mixtures of experts and the EM algorithm. Neural Comput., 6, 181-214
- Kahru, M., Michell, B.G., Diaz, A., & Miura, M. (2004a). MODIS detects a devastating algal bloom in Paracas Bay, Peru. Eos, Transactions American Geophysical Union, 85, 465-472
- Kahru, M., Mitchell, B.G., Diaz, A., & Miura, M. (2004b). MODIS detects a devastating algal bloom in Paracas Bay, Peru. EOS, 85, 465-472
- Kelly, K.S., & Krzysztofowicz, R. (1997). A bivariate meta-Gaussian density for use in hydrology. Stochastic Hydrology and Hydraulics, 11, 17-31
- Lavender, S.J., Pinkerton, M.H., Moore, G.F., Aiken, J., & Blondeau-Patissier, D. (2005). Modification to the atmospheric correction of SeaWiFS ocean colour images over turbid waters. Continental Shelf Research, 25, 539-555
- Lavoie, I., Laurion, I., Warren, A., & Vincent, W.F. (2007). Les fleurs d'eau de cyanobactéries: Revue de littérature. INRS rapport no 916
- Martin, L., & Gendron, A. (2004). Méthodes statistiques appliquées à la psychologie: traitement de données avec Excel.
- Moran, M.S., Inoue, Y., & Barnes, E.M. (1997). Opportunities and limitations for imagebased remote sensing in precision crop management. Remote Sensing of Environment, 61, 319-346
- Morel, A., & Bélanger, S. (2006). Improved detection of turbid waters from ocean color sensors information. Remote Sensing of Environment, 102, 237-249
- Moses, W.J., Gitelson, A.A., Berdnikov, S., & Povazhnyy, V. (2009a). Estimation of chlorophyll- a concentration in case II waters using MODIS and MERIS dataâ€"successes and challenges. Environmental Research Letters, 4, 045005
- Moses, W.J., Gitelson, A.A., Berdnikov, S., & Povazhnyy, V. (2009b). Estimation of chlorophyll-a concentration in case II waters using MODIS and MERIS data successes and challenges. Environmental Research Letters, 4, 045005
- Moses, W.J., Gitelson, A.A., Perk, R.L., Gurlin, D., Rundquist, D.C., Leavitt, B.C., Barrow, T.M., & Brakhage, P. (2012). Estimation of chlorophyll-a concentration in turbid productive waters using airborne hyperspectral data. Water Research, 46, 993-1004
- Ouillon, S., Douillet, P., Petrenko, A., Neveux, J., Dupouy, C., Froidefond, J.-M., Andréfouët, S., & Muñoz-Caravaca, A. (2008). Optical algorithms at satellite wavelengths for total suspended matter in tropical coastal waters. Sensors, 8, 4165-4185
- Oza, N.C., & Tumer, K. (2008). Classifier ensembles: Select real-world applications. Information Fusion, 9, 4-20
- Polikar, R. (2006). Ensemble based systems in decision making. Circuits and Systems Magazine, IEEE, 6, 21-45

- Rahman, H. (2001). Influence of atmospheric correction on the estimation of biophysical parameters of crop canopy using satellite remote sensing. International Journal of Remote Sensing, 22, 1245-1268
- Rahman, H., & Dedieu, G. (1994). SMAC: a simplified method for the atmospheric correction of satellite measurements in the solar spectrum. Remote Sensing, 15, 123-143
- Richard, P.S. (2014). Satellite Monitoring of Toxic Cyanobacteria for Public Health. In. Earthzine: http://earthzine.org/2014/03/26/satellite-monitoring-of-toxiccyanobacteria-for-public-health/
- Richter, R. (1996). A spatially adaptive fast atmospheric correction algorithm. International Journal of Remote Sensing, 17, 1201-1214
- Schapire, R.E. (1990). The Strength of Weak Learnability. Mach. Learn., 5, 197-227
- Sellner, K.G., Doucette, G.J., & Kirkpatrick, G.J. (2003). Harmful algal blooms: causes, impacts and detection. Journal of Industrial Microbiology and Biotechnology, 30, 383-406
- Torbick, N., Hession, S., Hagen, S., Wiangwang, N., Becker, B., & Qi, J. (2013). Mapping inland lake water quality across the Lower Peninsula of Michigan using Landsat TM imagery. International Journal of Remote Sensing, 34, 7607-7624
- Tørvi, H., & Hertzberg, T. (1997). Estimation of uncertainty in dynamic simulation results. Computers & Chemical Engineering, 21, Supplement, S181-S185
- Trishchenko, A.P., Luo, Y., & Khlopenkov, K.V. (2006). A method for downscaling MODIS land channels to 250-m spatial resolution using adaptive regression and normalization. In (pp. 636607-636607-636608)
- Trishchenko, A.P., Luo, Y., Khlopenkov, K.V., & Park, W.M. (2007). Multi-Spectral Clear-Sky Composites of MODIS/Terra Land Channels (B1-B7) Over Canada at 250m Spatial Resolution and 10-Day Intervals Since March, 2000: Top of the Atmosphere (TOA) Data. In
- Vermote, E.F., Tanre, D., Deuze, J.L., Herman, M., & Morcette, J.J. (1997). Second Simulation of the Satellite Signal in the Solar Spectrum, 6S: an overview. Geoscience and Remote Sensing, IEEE Transactions on, 35, 675-686
- Viollier, M., Tanré, D., & Deschamps, P.Y. (1980). An algorithm for remote sensing of water color from space. Boundary-Layer Meteorology, 18, 247-267
- Wang, S.-j., Mathew, A., Chen, Y., Xi, L.-f., Ma, L., & Lee, J. (2009). Empirical analysis of support vector machine ensemble classifiers. Expert Systems with Applications, 36, 6466-6476
- Wheeler, S.M., Morrissey, L.A., Levine, S.N., Livingston, G.P., & Vincent, W.F. (2012). Mapping cyanobacterial blooms in Lake Champlain's Missisquoi Bay using QuickBird and MERIS satellite data. Journal of Great Lakes Research, 38, 68-75

- Yacobi, Y., Moses, W., Kaganovsky, S., Sulimani, B., Leavitt, B., & Gitelson, A. (2011). NIR-red reflectance-based algorithms for chlorophyll-a estimation in mesotrophic inland and coastal waters: Lake Kinneret case study. Water Research, 45, 2428
- Yentsch, C., Lapointe, B., Poulton, N., & Phinney, D. (2008). Anatomy of a red tide bloom off the southwest coast of Florida. Harmful Algae, 7, 817-826

Appendix 1. The explanatory variables used to train the 21 specific estimators. For V1 the implied λ_i are i = 1, 2, 5, 6, and 7, while for V2, V3, and V4 they are i = 3, 4, 1, and 2, i = 4, 1, and 2, and i = 3, 4, and 1, respectively. For example:

$$\begin{split} \mathbf{V1} &= \frac{1}{2} \times [|\lambda_{b1} \mathbf{R}(\lambda_{b2}) - \lambda_{b2} \mathbf{R}(\lambda_{b1})| + |\lambda_{b2} \mathbf{R}(\lambda_{b5}) - \lambda_{b5} \mathbf{R}(\lambda_{b2})| + |\lambda_{b5} \mathbf{R}(\lambda_{b6}) - \lambda_{b6} \mathbf{R}(\lambda_{b5})| + \\ & |\lambda_{b6} \mathbf{R}(\lambda_{b7}) - \lambda_{b7} \mathbf{R}(\lambda_{b6})| + |\lambda_{b7} \mathbf{R}(\lambda_{b1}) - \lambda_{b1} \mathbf{R}(\lambda_{b7})|]. \end{split}$$

V1 and V2 are the two variables used by the AM classifier. For MODIS bands λ_{b1} =645, λ_{b2} =859, λ_{b3} =469, λ_{b4} =555, λ_{b5} =1,240, λ_{b6} =1,640 and λ_{b7} =2,130 (nm).

Explanatory variables	Names	Equations		
V1	The surface			
V2	The surface underneath Vis-NIR ₁	$\frac{1}{2}\sum \mathbf{R}(\boldsymbol{\lambda}) \times \boldsymbol{\lambda} = \mathbf{R}(\boldsymbol{\lambda} - \boldsymbol{\lambda} \times \boldsymbol{\lambda})$		
V3	The surface underneath G-NIR₁	$2 \sum_{i} \frac{ \mathcal{N}(\mathcal{N}_{i}) \wedge \mathcal{N}_{bi+1}}{ \mathcal{N}(\mathcal{N}_{i+1}) \wedge \mathcal{N}_{bi} }$		
V4	The surface underneath Vis			
V5	APPEL	$R(\lambda_{b2}) - \left[[R(\lambda_{b1}) - R(\lambda_{b2})] + [R(\lambda_{b3}) - R(\lambda_{b2})] \times R(\lambda_{b2}) \right]$		
V6	FAI	$R(\lambda_{b2}) - [R(\lambda_{b1}) + [R(\lambda_{b5}) - R(\lambda_{b1})] \times (\lambda_{b2} - \lambda_{b1})/(\lambda_{b5} - \lambda_{b1})$		
V7	Green – Blue bands	$R(\lambda_{h4}) - R(\lambda_{h3})$		
V8	Green – Red bands	$R(\lambda_{h4}) - R(\lambda_{h1})$		
V9	NIR1 – Blue bands	$R(\lambda_{b2}) - R(\lambda_{b3})$		
V10	NIR₁ – Red bands	$R(\lambda_{h2}) - R(\lambda_{h1})$		
V11	$NIR_1 - NIR_2$ bands	$R(\lambda_{b2}) - R(\lambda_{b5})$		
V12	NIR ₂ – Blue bands	$R(\lambda_{b5}) - R(\lambda_{b3})$		
V13	NIR₂ – Red bands	$R(\lambda_{b5}) - R(\lambda_{b1})$		
V14	Green/Blue bands	$R(\lambda_{b4})/R(\lambda_{b3})$		
V15	Green/Red bands	$R(\lambda_{b4})/R(\lambda_{b1})$		
V16	NIR ₁ /Blue bands	$R(\lambda_{b2})/R(\lambda_{b3})$		
V17	NIR ₁ /Red bands	$R(\lambda_{b2})/R(\lambda_{b1})$		
V18	NIR ₁ /NIR ₂ bands	$R(\lambda_{b2})/R(\lambda_{b5})$		
V19	NIR ₂ /Blue bands	$R(\lambda_{b5})/R(\lambda_{b3})$		
V20	NIR ₂ /Red bands	$R(\lambda_{b5})/R(\lambda_{b1})$		
V21	Red band	$R(\lambda_{b1})$		
V22	NIR₁ band	$R(\lambda_{b2})$		
V23	Blue band	$R(\lambda_{b3})$		
V24	Green band	$R(\lambda_{b4})$		
V25	NIR₂ band	$R(\lambda_{h5})$		

Appendix 2. The calibration equations of estimators (R² is the coefficient of determination of the multivariate regressions, N is the size of the training data set of each *estimator*, and BpV is the best p-value¹⁹. U, N, and L are the upper, nominal, and lower thresholds.
↑: *Estimator* trained by data higher than a given optimal threshold (U,N, or L), and ↓: *Estimator* trained by data lower than a given optimal threshold.

Specific estimators	Equations of the calibration	R ²	Ν	BpV
$Est_{TV1_{II}}$	$e^{(0.0054 \times V22 + 1.73)}$	0.96	06	<i>V</i> ₂₂
$Est_{TV1_N}^{\uparrow}$	$e^{(0.0049 \times V22 + 2.54)}$	0.95	07	V_{22}
$Est_{TV1_{L}^{\uparrow}}$	$e^{(29.76 \times V5 + 1.73)}$	0.95	17	V_{05}
$Est_{TV2_{U};TV1_{U}}$	$e^{(1.63 \times V17 + 1.39)}$	0.73	13	<i>V</i> ₁₇
$Est_{TV2_{N};TV1_{U}^{\downarrow}}$	$e^{(0.0025 \times V9 + 2.97)}$	0.40	13	V_{09}
$Est_{TV2_{I}^{\uparrow}:TV1_{II}^{\downarrow}}$	$e^{(0.006 \times V7 + 1.93)}$	0.88	05	V_{07}
$Est_{TV2_{U};TV1_{N}^{\downarrow}}$	$e^{(0.005 \times V9 + 2.61)}$	0.71	23	V_{09}
$Est_{TV2_{N}:TV1_{N}^{\downarrow}}$	$e^{(0.0057 \times V7 + 0.001 \times V11 + 1.81)}$	0.76	22	V_{07}
$Est_{TV2_{L}^{\uparrow}:TV1_{M}^{\downarrow}}$	$e^{(0.0063 \times V7 + 1.68)}$	0.70	12	V_{07}
$Est_{TV2_{U}^{\uparrow}TV1_{U}^{\downarrow}}$	$e^{0.006 \times V7 + 1.76}$	0.68	22	V_{07}
$Est_{TV2^{\uparrow}_{V}:TV1^{\downarrow}_{V}}$	$e^{(0.0056 \times V7 + 1.79)}$	0.73	29	V_{07}
$Est_{TV2_{L}^{\uparrow}TV1_{L}^{\downarrow}}$	$e^{(0.0024 \times V21 + 0.0044 \times V7 + 1.58)}$	0.84	16	V_{07}
$Est_{TV2_{U}^{\downarrow}:TV1_{U}^{\downarrow}}$	$0.05 \times V11 + 10.47$	0.65	21	V_{11}
$Est_{TV2_{N}^{\downarrow}:TV1_{U}^{\downarrow}}$	$0.05 \times V11 + 10.47$	0.65	21	V_{11}
$Est_{TV2_{T}^{\downarrow}:TV1_{U}^{\downarrow}}$	$0.049 \times V11 + 10.53$	0.65	20	V_{11}
$Est_{TV2_{U}^{\downarrow}:TV1_{W}^{\downarrow}}$	$0.051 \times V11 + 10.41$	0.71	13	V_{11}
$Est_{TV2_{N}^{\downarrow}:TV1_{N}^{\downarrow}}$	$0.051 \times V11 + 10.41$	0.71	13	V_{11}
$Est_{TV2_{T}^{\downarrow}:TV1_{N}^{\downarrow}}$	$0.051 \times V11 + 10.41$	0.71	13	V_{11}
$Est_{TV2_{U}^{\downarrow}TV1_{U}^{\downarrow}}$	$-0.08 \times V8 + 12.96$	0.81	06	V_{08}
$Est_{TV2_{V}^{\downarrow}:TV1_{V}^{\downarrow}}$	$-0.08 \times V8 + 12.96$	0.81	06	V_{08}
$Est_{TV2_L^{\downarrow};TV1_L^{\downarrow}}$	$-0.08 \times V8 + 12.96$	0.81	06	V_{08}

¹⁹ The explanatory variable that explains the maximum of the variance.