

Université du Québec
Institut national de la recherche scientifique
Centre Eau Terre Environnement

**MODÉLISATION DE L'ALÉA STRUCTUREL DES FLEURS D'EAU
D'ALGUES SUR LES LACS DU QUÉBEC MÉRIDIONAL**

Par
Claudie RATTÉ-FORTIN

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

Jury d'évaluation

Examinatrice externe	M ^{me} Beatrix BEISNER, professeure Université du Québec à Montréal
Examinatrice externe	M ^{me} Sarah DORNER, professeure Polytechnique de Montréal
Examineur interne et président du jury	M. André ST-HILAIRE, professeur INRS-ETE
Codirectrice de recherche	M ^{me} Isabelle LAURION, professeure INRS-ETE
Directeur de recherche	M. Karem CHOKMANI, professeur INRS-ETE

Résumé

Au Québec, les principaux intervenants dans le dossier de la problématique d'efflorescences d'algues et/ou de cyanobactéries peinent à concilier le maintien de la santé publique, la protection de l'environnement et la conservation de la biodiversité dans une optique de durabilité économique. Les outils de surveillance et d'analyse présentement utilisés constituent les points de mire à considérer dans l'élaboration d'une procédure de gestion et de suivi des fleurs d'eau d'algues (FEA) qui soit rigoureuse et complète. D'autre part, la complexité de l'aléa spatio-temporel des FEA de même que la méconnaissance de certains processus physico-chimiques sous-tendant leur apparition restreignent la capacité des modèles orientés processus à expliquer ce phénomène avec précision. La modélisation de l'aléa spatio-temporel des FEA nécessite une approche stochastique dans un contexte de gestion du risque associé à leur émergence. Par ailleurs, une modélisation prédictive est cruciale sachant que ce phénomène augmente en fréquence et en intensité en raison du réchauffement global de la planète et de la croissance démographique, plaçant ainsi le maintien de la santé publique au premier plan d'une procédure de gestion du risque. L'impact des changements climatiques et anthropiques sur l'occurrence du phénomène demeure cependant difficile à quantifier par les méthodes standards de suivi d'échantillonnage, celles-ci étant restreintes à un cadre spatio-temporel ponctuel.

L'objectif général de cette thèse est de développer un modèle stochastique permettant d'estimer la prédisposition des lacs québécois à développer des FEA en fonction des conditions climatiques et physiographiques prévalant sur le lac et son bassin versant. L'élaboration d'une procédure semi-automatisée d'estimation de la phénologie des FEA des lacs du Québec méridional a d'abord été mise en place, celle-ci étant nécessaire d'une part dans un contexte de suivi des épisodes en temps quasi-réel et, d'autre part, en raison de l'acquisition et du traitement d'une quantité importante de données. Une étude de la dynamique spatio-temporelle des FEA a ainsi pu être élaborée, de même que leurs liens potentiels avec les descripteurs morphologiques, physiographiques et climatiques des lacs et de leur bassin versant. Un modèle stochastique non-stationnaire régional a ensuite été développé dans le but de modéliser l'aléa structurel des FEA. Ce modèle est issu d'une approche par statistique fréquentielle à probabilités conditionnelles, où la probabilité d'apparition des FEA dépend de variables environnementales spécifiques. Cette thèse a conduit au développement (i) d'un outil de suivi des FEA en temps quasi-réel de même qu'un outil de gestion des risques socio-économiques et environnementaux qu'elles engendrent, (ii) d'un outil prévisionnel d'aide à la décision quant à l'optique de proposer des solutions d'intervention en matière de protection ou de restauration d'un plan d'eau, et (iii) au développement d'un modèle fréquentiel intégrant la structure spatio-temporelle de données.

Mots-clés Cyanobactéries Algues Eau Lac Bassin Versant Climat Physiographie Morphologie Modélisation Stochastique Probabilité Conditionnelle Fréquentielle

Remerciements

Je tiens d'abord à remercier mon superviseur Karem Chokmani pour son soutien constant au cours des dernières années. Karem a toujours su se distinguer par sa créativité, sa facilité d'approche et sa capacité à établir facilement des relations de partenariat. J'ai adoré collaborer à divers projets avec lui et discuter de sujets chauds tel que le végétarisme ou à l'écouter argumenter de sujets d'actualité politique avec Jean-Daniel ou Yves. Je remercie également ma codirectrice, Isabelle Laurion. J'ai énormément apprécié travailler et discuter avec Isabelle et ses commentaires judicieux ont grandement amélioré la thèse. Je tiens à remercier les membres de l'équipe de recherche en télédétection environnementale et nordique (TENOR) dont Anas El Alem, Yves Gauthier, Jasmin Gill-Fortin, Sarah Goubet, Jimmy Poulin, Tahiana Ratsimbazafy et Sophie Roberge pour leur aide et leur support (technique et moral). Je tiens aussi à remercier les examinateurs interne et externes, Beatrix Beisner, Sarah Dorner et André St-Hilaire pour leurs commentaires pertinents.

De nombreuses personnes ont fait en sorte que l'INRS représente plus qu'un lieu de travail pour moi. Il est difficile de tous les nommer mais je tiens à mentionner Anne Carabin, Patricia Gomez, Gilles Guérin et Alain Rousseau; merci pour tout, je vous adore. Une pensée toute particulière à Sophie Duchesne et Jean-Pierre Villeneuve avec qui j'ai adoré travailler et envers qui je suis grandement reconnaissante. Je remercie également Jean-Pierre Dedieu, l'équipe GeoLearn et Erwan Gloaguen avec qui j'ai énormément apprécié travailler et discuter.

Finalement, je remercie profondément ma famille dont mes parents et Étienne, Claude et Marcel, Michèle, feu Luce, mes frères et mes soeurs qui m'ont toujours supportée et que j'adore.

Table des matières

Résumé	iii
Remerciements	v
Table des matières	vii
Liste des figures	xi
Liste des tableaux	xv
Liste des abréviations	xvii
I INTRODUCTION	1
1 Les fleurs d'eau d'algues et/ou de cyanobactéries	3
1.1 Outils de surveillance et de suivi des FEA	4
1.2 Procédure de suivi des FEA par le Ministère	6
1.3 Modélisation de l'aléa structurel des FEA	7
1.4 Objectifs	9
1.4.1 Premier objectif spécifique	9
1.4.2 Deuxième objectif spécifique	10
1.4.3 Troisième objectif spécifique	10
1.5 Hypothèses	10
1.6 Structure de la thèse	11
II BASE DE DONNÉES HISTORIQUES D'ÉPISODES DE FEA	13
2 Introduction au développement de la base de données de FEA	15
2.1 Détection du couvert nuageux en milieux lacustre et océanique	16
2.2 Définition des caractéristiques phénologiques d'une FEA	17
2.3 Sommaire	21
3 Estimation phénologique des FEA	23
3.1 Données satellitaires	23
3.2 Prétraitement des données	25
3.3 Détection du couvert nuageux	27
3.3.1 Description du modèle	28

3.3.2	Calage et validation	29
3.4	Algorithme d'estimation des concentrations en chl-a	30
3.5	Caractérisation des FEA sur les lacs du Québec méridional	33
3.5.1	Choix des lacs	33
3.5.2	Caractérisation phénologique des FEA	34
4	Résultats	37
4.1	Détection du couvert nuageux	37
4.2	Procédure d'estimation phénologique via l'imagerie satellitaire MODIS	41
4.3	Analyse descriptive des événements algaux	44
4.4	Sommaire	47
 III PORTRAIT DE LA VARIABILITÉ SPATIO-TEMPORELLE DES FEA SUR LE QUÉBEC MÉRIDIONAL		49
5	Introduction aux sources de perturbation liées au développement les FEA	51
5.1	Prédisposition d'un lac aux FEA	52
5.2	Sommaire	56
6	Étapes méthodologiques	57
6.1	Sites d'étude et base de données géospatiales	57
6.2	Portrait du milieu physique	58
6.3	Analyse canonique des corrélations	63
6.4	Sommaire	65
7	Résultats	67
7.1	Variabilité spatio-temporelle des FEA	67
7.2	Tendances phénologiques du lac Brome et de la baie Missisquoi	72
7.3	Analyse des corrélations de Pearson	74
7.4	Analyse canonique des corrélations	75
7.5	Sommaire	78
7.5.1	Tendances phénologiques	78
7.5.2	Liens avec le climat et la physiographie du milieu	80
7.5.3	Incertitude sur l'estimation de la biomasse et des tendances globales	82
 IV MODÉLISATION DE L'ALÉA STRUCTUREL DES FEA		85
8	Revue de littérature des modèles développés pour les FEA	87
8.1	Modélisation déterministe (basée processus)	87
8.2	Modélisation stochastique (basée statistique)	88
8.3	Méthodes d'analyse fréquentielle	89
8.4	Modélisation par apprentissage automatique	90
8.5	Sommaire	92
9	Approche statistique d'analyse fréquentielle régionale	95
9.1	Introduction	95
9.2	Description des données	96

9.3	Description du modèle	97
9.4	Définition de la densité de probabilité conditionnelle	100
9.5	Évaluation du modèle régional	102
9.6	Hypothèses clés	102
9.7	Sommaire	103
10	Résultats	105
10.1	Analyse de la structure spatio-temporelle	105
10.2	Calage du modèle	110
10.3	Application du modèle	117
10.4	Sommaire	126
V	CONCLUSION	129
11	Conclusion	131
11.1	Résumé des principaux résultats	131
11.1.1	Partie II	131
11.1.2	Partie III	133
11.1.3	Partie IV	135
11.2	Retombées scientifiques	136
11.2.1	Portée thématique	136
11.2.2	Portée méthodologique	138
11.3	Perspectives	138
	Références	141
VI	ANNEXES	157
A	Imputation des données manquantes	159
B	Portrait de la qualité de l'eau des plans d'eau du Québec	161
C	Types de données climatiques et physiographiques	163
D	Corrélations de Pearson	165
E	Résultats supplémentaires	167

Liste des figures

2.1	Exemples de seuils décisionnels pour l'évaluation du risque associé à la présence de cyanobactéries (tiré de Warren (2011)).	20
3.1	Zone couverte par la base de données satellitaires (source: Google Earth).	25
3.2	(a) Image MODIS en vraie couleur, (b) masque de nuage développé par le CCTS, et (c) masque de nuage développé par le groupe MODIS Atmosphere.	27
3.3	Méthode détaillée utilisée pour distinguer les classes [nuage - eau] avec l'analyse linéaire discriminante.	29
3.4	Méthode détaillée utilisée pour estimer la distribution du succès global (%) et du κ de la classification en utilisant une validation croisée par Monte Carlo.	30
3.5	Concentration en chlorophylle-a estimée à partir des quatre modèles comparée aux mesures in situ pour les données supérieures à 10 $\mu\text{g/L}$, avec les indices de performance des modèles (figure tirée de El-Alem (2015)).	31
3.6	Concentration en chlorophylle-a estimée à partir du modèle adaptatif 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. (figure tirée de El-Alem (2015)).	32
3.7	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. (figure tirée de El-Alem (2015)).	33
4.1	(a) Bandes MODIS R-PIR-B et R-V-B de la baie Missisquoi et du lac Champlain, (b) les trois masques générés, et (c) les couches de concentrations en chl-a estimées avec les pixels d'eau restants. Les cercles rouges présentent des concentrations en chl-a élevées où il n'y a <i>a priori</i> aucune FEA.	39
4.2	Bandes MODIS R-PIR-B et R-V-B du fleuve Saint-Laurent (a) et du lac Macamic (c), ainsi que les trois masques de nuages (b) et (d) correspondant.	40
4.3	Bandes MODIS R-PIR-B et R-V-B de la baie de Fundy (a) et du fleuve Saint-Laurent (c), ainsi que les trois masques de nuages (b) et (d) correspondant	41
4.4	(a) Exemple d'une scène du lac Saint-Jean et du fleuve Saint-Laurent présentée (a) avec les bandes MODIS R-PIR-B, (b) sans présence de pixels de terre ni de pixels mixtes, (c) sans présence du couvert nuageux, et (d) avec la couche des concentrations en chl-a estimées.	43
4.5	Distribution des variables phénologiques de 2000 à 2016: (a) Fréquence d'apparition annuelle, (b) intensité maximale annuelle, (c) étendue maximale annuelle, et (d) Date d'initiation annuelle.	45

4.6	Distribution cumulée complémentaire des variables phénologiques de 2000 à 2016: (a) Fréquence d'apparition annuelle, (b) intensité maximale annuelle, (c) étendue maximale annuelle, et (d) Date d'initiation annuelle.	46
6.1	Région d'étude et limites géographiques des zones de gestion intégré de l'eau (gris foncé) couvrant 580 lacs observables par le capteur MODIS.	58
6.2	Exemple de calcul de l'aire de drainage: (a) MNA et lac Saint-Joseph (bleu), (b) zones de dépressions naturelles du sol, (c) raster de la direction du flux d'écoulement et lac Saint-Joseph (bleu), (d) bassin de drainage (bleu pâle) du lac Saint-Joseph (bleu).	60
6.3	Couverture spatiale des données physiographiques et climatiques: (a) MNA et données sur le relief, (b) précipitations, températures, degrés-jours, vents (c) occupation du sol, (d) écoumène agricole et écoumène de population, (e) données d'érodabilité des sols, et (f) données d'érosivité des sols.	61
7.1	Fréquence d'apparition de FEA (relatif au nombre de lacs) en fonction de la longitude (haut-gauche), de la latitude (haut-droit) et du temps (bas).	68
7.2	Intensité des FEA en fonction de la longitude (haut-gauche), de la latitude (haut-droit) et du temps (bas).	69
7.3	Étendue des FEA (relatif à la superficie totale du lac) en fonction de la longitude (haut-gauche), de la latitude (haut-droit) et du temps (bas).	70
7.4	Début des épisodes annuels de FEA en fonction de la longitude (haut-gauche), de la latitude (haut-droit) et du temps (bas).	71
7.5	Fin des épisodes annuels de FEA en fonction de la longitude (haut-gauche), de la latitude (haut-droit) et du temps (bas).	72
7.6	Fréquence, intensité, étendue et début annuels de 2000 à 2016 sur la Baie Missisquoi.	73
7.7	Fréquence, intensité, étendue et début annuels de 2000 à 2016 sur le lac Brome.	74
7.8	Pourcentage moyen (%) entre 2000 et 2016 de données manquantes reliées à la couverture nuageuse trop importante (plus de 75% du lac) sur un lac donné ou présentant des distorsions géométriques de l'image MODIS. Le pourcentage moyen de jours manquants est présenté par mois, par superficie du lac, et selon l'emplacement géographique (latitude longitude).	84
9.1	Présentation schématique des données d'entrée (x et covariable) dans le modèle. ACC signifie l'analyse canonique des corrélations.	96
10.1	Répartition spatiale des centroides des lacs (points) en fonction de la latitude et de la longitude sur la région d'étude (contour noir). Le gradient de couleur reflète le nombre de lacs présents dans une région donnée.	106
10.2	Fonctions $L(h)$ pondérés et non-pondérés en fonction de la distance euclidienne h des lacs à l'étude.	107
10.3	Répartition spatiale de la fréquence des FEA sur la région d'étude.	108
10.4	Répartition spatiale de l'intensité des FEA sur la région d'étude.	109
10.5	Répartition spatiale de la fréquence des FEA sur la région d'étude.	110
10.6	Test d'autocorrélation spatiale sur l'intensité maximale.	110
10.7	Moyennes et écart-types de la variation des fréquences d'apparition de FEA pour chaque classe de valeurs de la covariable.	112
10.8	Moyennes et écart-types de la variation d'intensité ($\ln[mgL^{-1}]$) de FEA pour chaque classe de valeurs de la covariable.	113

10.9	Simulations bootstraps non-paramétriques des retours aux 2 ans des fréquences annuelles du lac Saint-Jean estimés par le modèle local (haut) et le modèle régional (bas).	118
10.10	Simulations bootstraps non-paramétriques des retours aux 2 ans des fréquences annuelles du lac Aylmer estimés par le modèle local (haut) et le modèle régional (bas).	119
10.11	Simulations bootstraps non-paramétriques des retours aux 2 ans des fréquences annuelles du lac Brome estimés par le modèle local (haut) et le modèle régional (bas).	120
10.12	Simulations bootstraps non-paramétriques des retours aux 2 ans des fréquences annuelles de la Baie Missisquoi estimés par le modèle local (haut) et le modèle régional (bas).	121
10.13	Temps de retour 2 ans de la fréquence d'apparition annuelle de FEA sur le lac Brome en 2000 estimés par le modèle local (haut) et le modèle régional (bas).	123
10.14	Localisation de la tuile (rouge) d'où proviennent les simulations climatiques futures par Ouranos. Le lac Brome et son bassin versant (polygone blanc) sont inclus à l'intérieur de cette région.	124
10.15	Fréquence annuelle future du lac Brome présentée en fonction du temps (haut), de la covariable environnementale (milieu) et des degrés-jour cyanobactériens.	125
E.1	Répartition spatiale de l'étendue des FEA sur la région d'étude.	167
E.2	Répartition spatiale de la date de début des FEA sur la région d'étude.	168
E.3	Test d'autocorrélation spatiale sur l'étendue maximale.	168
E.4	Test d'autocorrélation spatiale sur la date de début annuel d'épisodes.	169
E.5	Moyennes et écart-types de la variation de l'étendue de FEA pour chaque classe de valeurs de la covariable.	170
E.6	Moyennes et écart-types de la variation de date d'initiation annuelle de FEA pour chaque classe de valeurs de la covariable.	171

Liste des tableaux

3.1	Caractéristiques du satellite TERRA et du capteur MODIS.	24
3.2	Caractéristiques des bandes 1-7 du capteur MODIS.	26
3.3	Liste des images MODIS utilisées pour le calage et la validation.	30
4.1	Résultats de la matrice de confusion avec un intervalle de confiance à 95 % des valeurs moyennes de succès global et de l'indice Kappa.	38
4.2	Résultats de la classification des deux produits MODIS (1 km et 250 m) et de l'algorithme basé sur l'analyse linéaire discriminante.	38
6.1	Liste des descripteurs environnementaux. Les variables annotées en gras correspondent aux descripteurs utilisés pour l'analyse canonique des corrélations.	62
7.1	Corrélations canoniques et test de signification des corrélations canoniques.	75
7.2	Coefficients canoniques, coefficients de structure et coefficients de communalité des deux premières variables canoniques. Les coefficients de communalité plus élevés que 45 % sont présentés en gras.	77
10.1	Variables phénologiques et leur corrélation avec les variables environnementales.	111
10.2	Log-vraisemblances des modèles testés sur la fréquence de FEA en considérant une dépendance linéaire sur la covariable environnementale. Le modèle sélectionné est présenté en caractère gras.	114
10.3	Log-vraisemblances des modèles testés sur l'intensité de FEA en considérant une dépendance linéaire sur la covariable environnementale. Le modèle sélectionné est présenté en caractère gras.	114
10.4	Log-vraisemblances des modèles locaux testés sur la fréquence de FEA en considérant une dépendance linéaire dans le temps. Les modèles sélectionnés sont présentés en caractère gras.	115
10.5	Log-vraisemblances des modèles locaux testés sur l'intensité de FEA en considérant une dépendance linéaire dans le temps. Les modèles sélectionnés sont présentés en caractère gras.	116
10.6	Statistiques comparatives des temps de retour 2 ans obtenus par le modèle régional et le modèle local. Les biais absolus et les RMSE sont présentés à titre indicatif seulement.	122
B.1	Nombre de plans d'eau touchés selon la zone de gestion intégrée de l'eau, de 2007 à 2012 (MELCC, 2014a).	162
D.1	Analyse des corrélations de Pearson entre les variables phénologiques et les variables environnementales.	165

D.2	Corrélations de Pearson entre les variables environnementales (gauche) et entre les variables phénologiques (droite).	166
E.1	Variabes phénologiques et leur corrélation avec les variables environnementales. . . .	169

Liste des abréviations

ACC	Analyse canonique des corrélations
AIC	Critère d'information d'Akaike
ALD	Analyse linéaire discriminante
AVHRR	Advanced Very High Resolution Radiometer
BIC	Critère d'information Bayésien
CART	Classification and Regression Tree
CCEQ	Centre de contrôle environnemental du Québec
CCL	Projection conique conforme de Lambert
CCTS	Centre Canadien de Télédétection Spatiale
CEAEQ	Centre d'expertise en analyse environnementale du Québec
Chl-a	Chlorophylle a
CLIR	Courtes longueurs d'ondes de l'infrarouge
EOS	Earth Observing System
FAI	Floating algae index
FEA	Fleurs d'eau d'algues et/ou de cyanobactéries
GRIL	Groupe de recherche inter-universitaire en limnologie et en environnement aquatique
KW	Test de Kruskal-Wallis
MA	Modèle adaptatif
MBME	Modèle basé sur les méthodes de l'ensemble
MELCC	Ministère de l'Environnement et de la Lutte contre les Changements Climatiques
MNA	Modèle numérique d'altitude
MOD	Matière organique dissoute
MODC	Matière organique dissoute colorée
MODIS	Moderate Resolution Imaging Spectroradiometer
MODIS-D-250	Bandes 1-7 du capteur MODIS désagrégées à 250 m de résolution spatiale
MSS	Matière solide en suspension

NT	Azote total
ONA	Oscillation nord-atlantique
PIR	Proche infrarouge
PT	Phosphore total
ROBVQ	Regroupement des organismes de bassins versants du Québec
RSVL	Réseau de surveillance volontaire des lacs du Québec
SIN	Projection sous le système sinusoïdal
SMAC	Simplified Model for Atmospheric Correction
SSTTA	Structure spatio-temporelle totalement aléatoire
ZGIE	Zones de gestion intégrée de l'eau

Première partie

INTRODUCTION

Chapitre 1

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

La hausse marquée des fleurs d'eau d'algues et/ou de cyanobactéries (FEA) dans les écosystèmes d'eau douce est d'intérêt majeur pour les agences gouvernementales et les organismes de santé publique en ce qui a trait à la gestion et à la consommation de l'eau. Ces micro-organismes menacent désormais l'intégrité écologique des plus importants milieux lacustres sur la planète, dont le lac Érié, USA-Canada (Michalak *et al.*, 2013), le lac Ontario, USA-Canada (Higgins *et al.*, 2012), le lac Taihu, Chine (Duan *et al.*, 2015), le lac Okeechobee, USA (Rosen *et al.*, 2017) et le lac Victoria, Afrique de l'Est (Simiyu *et al.*, 2018). Les problèmes de santé liés aux toxines que les cyanobactéries en particulier produisent incluent des réactions allergiques, des irritations cutanées, des gastro-entérites, et parfois même des problèmes potentiellement mortels dont le développement de tumeurs et de maladies du foie (Chorus *et al.*, 2000; Carmichael, 2001; Huisman *et al.*, 2018). Leur présence plus fréquente et plus soutenue affecte la structure et le fonctionnement du réseau alimentaire aquatique (Bartram & Chorus, 1999), limite les activités récréo-touristiques (Steffensen, 2008) et menace les sources d'eau potable (Zamyadi *et al.*, 2012). Le suivi des FEA surtout dominées par les cyanobactéries demeure cependant difficile à réaliser en raison de leur distribution extrêmement variable dans le temps et dans l'espace.

1.1 Outils de surveillance et de suivi des FEA

Tout programme de protection et de gestion des lacs requière la détection rapide des FEA et de leurs cyanotoxines potentielles puisque, même à faible abondance, elles peuvent nuire à la santé humaine et animale (Huisman *et al.*, 2018). Il s'avère donc essentiel de développer des outils de détection qui puissent assurer de manière efficace la détection et le suivi spatio-temporel des FEA. Il existe actuellement plusieurs outils de surveillance des FEA adaptés aux milieux naturels qui permettent d'effectuer une estimation adéquate de leur prolifération. Le moyen le plus précis est l'échantillonnage *in situ*, et ce, aussi bien pour les contaminants que pour les micro-organismes. L'identification taxonomique et le calcul de l'abondance des cyanobactéries et du phytoplancton peuvent être obtenus par microscopie inverse en laboratoire. Cette méthode se base sur les caractéristiques morphologiques des cellules, et lors d'une analyse complète, elle peut fournir l'identification, le dénombrement et la biomasse des espèces d'algues. Des méthodes basées sur les propriétés optiques des algues permettent également d'estimer la biomasse par fluorescence *in vivo* (Bastien *et al.*, 2011; Zamyadi *et al.*, 2012; Choo *et al.*, 2018), la composition taxonomique de certains groupes par la signature spectrale (Kudela *et al.*, 2015) et l'état physiologique du phytoplancton par fluorescence à amplitude modulée (Jakob *et al.*, 2005). Notamment, la fluorescence *in vivo* est utilisée afin d'estimer *in situ* la biomasse algale à l'aide de sondes submersibles permettant un profilage vertical de la colonne d'eau, une cartographie de la répartition algale en surface (transects) ou la surveillance d'une station spécifique (bouée). Ces sondes enregistrent le signal de fluorescence puis le convertissent en biomasse (concentration en chl-a) par le biais de relations empiriques établies par le fabricant (et très souvent inconnues des utilisateurs). Cependant, sachant que la fluorescence est affectée par les caractéristiques physiologiques et morphologiques des algues et que plusieurs facteurs interfèrent avec l'estimation de la biomasse à partir du signal de fluorescence (par ex.: la matière solide en suspension (MSS), la matière organique dissoute (MOD)), il est essentiel de calibrer l'appareil *a posteriori* en établissant ses propres relations empiriques selon le plan d'eau étudié. Ces méthodes basées sur la fluorescence des pigments sont plus rapides que la microscopie conventionnelle mais nécessitent un calage et une validation régulière par microscopie et montrent une incertitude sur les résultats très variable d'un plan d'eau à l'autre.

La télédétection constitue quant à elle une alternative à ces méthodes de suivi et s'est avérée être un outil de détection et de suivi efficace, permettant une estimation rapide et automatisée de

la biomasse des FEA en milieu océanique et lacustre (Kahru *et al.*, 1993, 2004; Kutser *et al.*, 2006; Becker *et al.*, 2009; Tarrant & Neuer, 2009; Hu *et al.*, 2010; Wynne *et al.*, 2010; Li *et al.*, 2015; Lunetta *et al.*, 2015; Palmer *et al.*, 2015). Les méthodes de suivi et d'estimation de la concentration en chl-a via l'imagerie satellitaire contribuent à réduire le temps et le coût d'acquisition des données sur la qualité des eaux lacustres, mettant à la disposition des gestionnaires de l'eau un outil efficace de suivi des FEA dans le temps et dans l'espace. La télédétection permet alors une meilleure compréhension de la dynamique spatiale des FEA, autant au niveau local que régional, en offrant la possibilité de couvrir certaines régions sensibles à plus petite échelle. La dynamique temporelle peut également être étudiée grâce aux bases de données historiques d'images disponibles pour certains capteurs. Dans le cadre d'un suivi des lacs de petites superficies qui ne sont pas observables via l'imagerie satellitaire, l'utilisation de drones ou de capteurs aéroportés est une option envisageable. Notons cependant que la détection des FEA par l'imagerie satellitaire ne permet pas d'identifier les espèces, présente une plus grande incertitude lors d'une floraison à faible densité (Franks & Keafer, 1995; El-Alem *et al.*, 2014a) et est restreinte par la couverture nuageuse (International Council for the Exploration of the Sea, 2004). De plus, l'imagerie satellitaire est restreinte quant à la profondeur de détection des composantes optiquement actives dans l'eau qui peuvent renvoyer un signal détectable. La profondeur de pénétration dépend de la transparence de l'eau, où les algues et en particulier les cyanobactéries peuvent migrer dans la colonne d'eau et se retrouver en forte densité plus profondément. Notons également que les cyanobactéries (particulièrement les buoyant bloom-formers) peuvent s'accumuler à la surface des lacs de manière plus marquée que les autres groupes phytoplanctoniques (sauf dans certains cas, par ex. *Planktothrix rubescens* dont l'habitat est le métalimnion).

Bien que l'échantillonnage *in situ* soit un processus de surveillance considérablement fiable et précis à l'échelle locale, il demeure laborieux et comporte des impératifs budgétaires évidents. À cela s'ajoute le délai entre le signalement et les résultats d'analyse, et le besoin de personnel spécialisé et abondant en période d'efflorescence importante d'algues. Ainsi, ce processus de surveillance demeure limité dans le temps et dans l'espace (Allan *et al.*, 2006). Pourtant, une bonne compréhension de la variabilité spatio-temporelle des FEA repose sur la capacité à fournir des données fiables et représentatives de la réalité du plan d'eau. À cet aspect s'ajoute la mise en œuvre d'un protocole d'échantillonnage et d'analyse rigoureux et standardisé permettant la comparaison des résultats à d'autres études. Ainsi, les limites de cet outil et l'importance socio-économique du problème incitent

à développer des outils de surveillance en temps quasi-réel permettant une estimation automatisée de la prolifération algale, et ce, avec comme objectif de réduire le temps et les coûts d'acquisition de données sur la qualité des eaux lacustres.

1.2 Procédure de suivi des FEA par le Ministère

Depuis 2004, le ministère de l'Environnement et de la Lutte contre les Changements Climatiques (MELCC) et le ministère de la Santé et des Services Sociaux (MSSS) travaillent conjointement dans le dossier des FEA et ont mis en place un plan de gestion des épisodes de fleurs d'eau d'algues bleu-vert (MELCC, 2015). La procédure du plan de gestion débute d'abord par le signalement de la présence d'une FEA par un citoyen ou un organisme auprès du MELCC. Avant 2013, le Centre de contrôle environnemental du Québec (CCEQ) se déplaçait sur le terrain et évaluait l'étendue et l'intensité du phénomène, déterminait les usages touchés et prélevait des échantillons qui étaient analysés au Centre d'expertise en analyse environnementale du Québec (CEAEQ) (MELCC, 2014a). Le suivi consistait alors à informer les municipalités et le Regroupement des organismes de bassins versants du Québec (ROBVQ) de la présence d'une FEA. Cette procédure a permis de générer une base de données sur un ensemble de plans d'eau touchés par les FEA entre 2007 et 2012 au Québec, faisant principalement état du nombre de plans d'eau qui ont été touchés par une FEA et du nombre de plans d'eau qui ont été visités par le ministère. Des analyses en laboratoire ont également été faites, dénombrant les genres de cyanobactéries dans les échantillons et les concentrations en microcystine. Depuis 2013, des modifications ont été apportées à la procédure de gestion et le MDDELCC ne visite plus systématiquement tous les plans d'eau où il y a eu un signalement de FEA. En effet, seuls les plans d'eau dont la présence de FEA est nouvellement signalée de même que ceux possédant certaines particularités, telles que les réservoirs d'eau potable, sont visités (MELCC (2014b)). Ainsi, les FEA signalées ne sont plus toutes analysées et ne peuvent pas être comparées avec celles signalées des années entre 2007 et 2012, car les plans d'eau ne sont plus systématiquement visités. En date d'octobre 2019, le MELCC n'a pas établi de bilans de la gestion des épisodes de FEA depuis 2015.

Par ailleurs, notons qu'il existe un réseau de surveillance volontaire des lacs du Québec (RSVL), issu d'un partenariat entre le MELCC, les associations de riverains, les municipalités et des organismes partenaires. Ce réseau a comme objectifs d'établir le niveau trophique d'un grand nombre de lacs, d'assurer un suivi de l'évolution de la qualité de l'eau, de détecter les lacs montrant des

signes d'eutrophisation, de dresser un portrait global de l'état des lacs de villégiature au Québec et de sensibiliser et informer les participants. Le RSVL permet de manière générale d'établir un portrait sommaire de l'état trophique du milieu, incluant une mesure de la transparence de l'eau et une analyse de la qualité de l'eau 3 fois par année pendant 2 ans, suivant ensuite une pause d'échantillonnage pendant 4 ans.

Bien que les procédures de gestion du MELCC aient permis un suivi de la prolifération de FEA, les coûts alloués aux analyses d'échantillons en laboratoire sont élevés. À titre d'exemple, le MELCC a alloué près de 411 500 \$ pour l'analyse en laboratoire de 965 échantillons en 2011, et près de 542 000 \$ pour l'analyse de 1058 échantillons en 2012 (MELCC, 2011, 2013). Le RSVL relate qu'il en coûte en moyenne 476\$ annuellement par station d'échantillonnage, ce montant incluant l'analyse de 3 échantillons d'eau, les frais de transport, les taxes et les documents d'accompagnement. Ces impératifs budgétaires limitent la surveillance de la prolifération de FEA dans le temps et dans l'espace, et ce, particulièrement pour des territoires vastes et possédant un volume important de milieux lacustres. De plus, cette procédure n'assure pas de suivi complet de la phénologie des FEA (i.e. fréquence, intensité, durée) car cela nécessite un suivi beaucoup plus complet dans le temps et dans l'espace. Par ailleurs, il n'existe pas de base de données historiques d'au moins 15 ans couvrant l'ensemble du Québec méridional qui soient relativement complète dans le temps et dans l'espace, et encore moins sur les caractéristiques phénologiques des FEA. Considérant que l'échantillonnage *in situ* limite grandement le suivi d'un phénomène très variable sur le plan spatiotemporel (Rantajarvi *et al.*, 1998), un portrait régional des FEA sur le Québec méridional via des données satellitaires permettrait d'établir un bilan complet et précis de la situation au Québec, et ainsi, d'orienter les décisions subséquentes à une meilleure gestion des milieux lacustres.

1.3 Modélisation de l'aléa structurel des FEA

Les milieux lacustres ont été sévèrement touchés par les rejets diffus et ponctuels de contaminants provenant d'activités anthropiques depuis les dernières décennies. Sur une grande partie du territoire du Québec méridional, le développement économique et la croissance de la population ont altéré de manière importante la santé de ces écosystèmes. En raison de cette pression anthropique, plusieurs plans d'eau souffrent d'eutrophisation précoce causée par l'enrichissement en phosphore et en azote. Ces nutriments constituent des éléments nutritifs essentiels à la croissance des algues qui prolifèrent

lorsqu'ils sont en excès. Parmi les facteurs environnementaux influençant la prolifération de FEA en eaux douces, le phosphore est considéré comme l'élément limitant et comme le principal facteur expliquant leur variabilité (Taranu *et al.*, 2015; Przytulaska *et al.*, 2017; Huisman *et al.*, 2018). De plus, des facteurs climatiques jouent un rôle dans la prolifération des FEA; notons entre autre la température, les vents dominants et l'intensité des précipitations (Huisman *et al.*, 2018). Par ailleurs, de nombreux experts soulignent que les changements climatiques ont le potentiel d'agir comme un catalyseur sur la prolifération des cyanobactéries et ne feront qu'amplifier le problème (Winter *et al.*, 2011; Paerl & Paul, 2012; Michalak, 2016; Huisman *et al.*, 2018; Townhill *et al.*, 2018; Bartosiewicz *et al.*, 2019a). Ces facteurs climatiques et anthropiques contribuent à créer un patron spatio-temporel variable dans la distribution régionale des FEA, rendant plus difficile leur suivi (Figueredo & Giani, 2005).

Afin de comprendre les mécanismes d'initiation, de développement, de maintien et de déclin des FEA, d'étudier les liens entre leur occurrence et les facteurs environnementaux et de comprendre les mécanismes régissant les FEA, il est nécessaire de mettre en lumière les sources de perturbation et leur lien avec la phénologie des FEA, d'une part, et de développer des modèles stochastiques qui permettent une modélisation complète et précise du phénomène. La modélisation de l'aléa structurel (déf.1.1) des FEA nécessite l'utilisation de modèles stochastiques, et plus particulièrement des modèles stochastiques fréquentiels afin de réaliser une meilleure gestion du risque associé à l'émergence des FEA. Bien que plusieurs modèles se soient intéressés à la dynamique de développement de FEA en fonction des facteurs environnementaux (par ex.: Binding *et al.* (2011); Liu *et al.* (2013)), ils ne permettent pas de décrire la variabilité spatio-temporelle des facteurs structurels des bassins versants en lien avec l'occurrence des FEA.

Définition 1.1 (Aléa structurel). *Potentiel de développement d'un épisode de FEA sur un lac donné, au cours d'une période donnée, en relation avec ses caractéristiques morphologiques et les caractéristiques physiographiques et climatiques de son bassin versant.*

Les modèles statistiques fréquemment utilisés dans la prédiction de la probabilité d'occurrence d'un phénomène sont très souvent utilisés de manière locale, c'est-à-dire que la modélisation est faite à partir de données provenant d'un milieu récepteur, d'une station hydrométrique ou d'une station météorologique. L'étude des changements climatiques ou physiographique sur l'occurrence

des FEA nécessite l'acquisition d'une quantité considérable de données temporelles afin d'assurer une analyse précise de leur évolution. Une modélisation fréquentielle locale de l'occurrence des FEA nécessite minimalement l'obtention d'une trentaine d'années de données disponibles afin d'assurer une modélisation juste du phénomène. Peu de bases de données permettent l'étude adéquate de l'impact des changements climatiques ou de l'impact de facteurs environnementaux sur l'apparition des FEA en raison de la difficulté à établir une base de données complète dans le temps et dans l'espace lorsque les études sont établies à une échelle locale (par ex.: l'étude d'un milieu lacustre en particulier).

1.4 Objectifs

L'objectif général de la thèse est d'estimer la prédisposition des lacs québécois à développer des FEA en fonction des caractéristiques physiographiques et climatiques du lac et de son bassin versant par le développement d'un modèle stochastique. Ce modèle sera basé sur l'analyse de données historiques de FEA obtenues par le biais d'algorithmes satellitaires exploitant l'imagerie MODIS. La grande quantité de données fournies par cette approche permettra d'assurer un suivi efficace de leur évolution dans le temps et dans l'espace. Ainsi, la fréquence d'occurrence des FEA résultera des conditions climatologiques prédominantes, de la signature morphologique du plan d'eau, des conditions physiographiques du bassin versant et des caractéristiques d'occupation et d'utilisation du territoire.

1.4.1 Premier objectif spécifique

Le premier objectif spécifique est de mettre en place une procédure semi-automatisée d'estimation de la chl-a afin d'établir une base de données historiques de la phénologie des FEA des lacs du Québec méridional. Les principales étapes qui permettront de répondre à cet objectif spécifique sont d'abord de développer un algorithme de détection du couvert nuageux pour les eaux lacustres à l'aide des données satellitaires du capteur MODIS. Par la suite, une base de données géospatiales des concentrations en chl-a sera mise en place sur les lacs québécois de 2000 à 2016, puis finalement une base de données tabulaires de la phénologie des FEA des lacs québécois.

1.4.2 Deuxième objectif spécifique

Le deuxième objectif spécifique consiste à établir un portrait régional de l'occurrence des FEA et de caractériser leurs liens potentiels avec les caractéristiques environnementales prévalant sur le bassin versant. Les principales étapes qui permettront de répondre à cet objectif spécifique sont d'abord d'établir un portrait régional de l'état des lacs québécois, de mettre en place une base de données tabulaires des variables environnementales liées à l'aléa structurel des FEA sur l'ensemble des lacs québécois à l'étude, puis finalement d'analyser les relations potentielles entre la phénologie des FEA et les variables environnementales.

1.4.3 Troisième objectif spécifique

Le troisième objectif spécifique est de modéliser l'aléa structurel (déf. 1.1) des FEA afin d'estimer la prédisposition des lacs à développer des FEA en relation avec les variables environnementales prévalant sur le bassin versant du plan d'eau donné. La première étape sera de valider les hypothèses d'ajustement d'une densité de probabilité conditionnelle à des données spatio-temporelles, d'ajuster un modèle fréquentiel régional non-stationnaire aux variables phénologiques, puis de démontrer la pertinence du modèle à travers un exemple d'application.

1.5 Hypothèses

Deux hypothèses de recherche constituent la base des travaux de cette thèse. Les postulats menant à ces hypothèses de même que des hypothèses d'aspects plus techniques liées au développement du modèle sont présentés plus en détail dans le chapitre 9. La première hypothèse stipule que:

Hypothèse 1.1. *Il existe une variabilité spatio-temporelle dans la phénologie des FEA qui peut être expliquée par le biais de descripteurs environnementaux sous-tendant au phénomène.*

Nous supposons qu'il existe une variabilité spatio-temporelle dans la phénologie des FEA, et celle-ci porte l'empreinte de descripteurs environnementaux tels que la physiographie du milieu,

le climat et la morphologie du plan d'eau, qui eux, évoluent dans le temps et dans l'espace. La deuxième hypothèse avance que:

Hypothèse 1.2. *Une modélisation fréquentielle régionale permet d'améliorer la précision et l'exactitude des estimations comparativement à une modélisation fréquentielle locale.*

Nous supposons que le développement d'un modèle captant la structure spatio-temporelle d'un phénomène tel que les FEA (dans le cadre de cette thèse, nous l'appelons un modèle régional) permettrait une meilleure représentativité du phénomène comparativement à une modélisation locale.

1.6 Structure de la thèse

La structure de la thèse s'articule comme suit. La partie II incluant les chapitres 2, 3 et 4 porte sur la mise en place d'une procédure permettant l'estimation de la phénologie des FEA sur l'ensemble des lacs du Québec méridional et répond ainsi au premier objectif spécifique de la thèse. La partie III incluant les chapitres 5, 6 et 7 porte sur l'établissement d'un portrait régional de la phénologie des FEA sur les lacs du Québec méridional et des liens potentiels avec les descripteurs environnementaux. Cette partie répond au deuxième objectif spécifique de cette thèse. La partie IV incluant les chapitres 8, 9 et 10 porte sur le développement d'un modèle fréquentiel régional et répond ainsi au troisième objectif spécifique. Finalement la cinquième et dernière partie inclut la conclusion, suivie des références et des annexes.

Deuxième partie

**BASE DE DONNÉES
HISTORIQUES D'ÉPISODES DE
FEA**

Chapitre 2

Introduction au développement de la base de données de FEA

La couverture nuageuse s'avère être un défi majeur à la télédétection optique dans le cadre d'études sur la qualité de l'eau. Plusieurs travaux combinent des approches de détection du couvert nuageux, mais les principaux algorithmes se divisent en cinq groupes distincts (Chen & Zhang, 2015): le seuillage dans le PIR, le seuillage dans le CLIR, l'algorithme de Hu (2011), les méthodes basées sur la variabilité spatiale de la réflectance et celles basées sur la variabilité spectrale du signal. Par ailleurs, bien qu'elle intègre généralement des mesures d'intensité et de fréquence d'occurrence, la caractérisation phénologique d'un épisode de FEA ne présente pas de mesures constantes dans la littérature. La European Water Framework Directive (WFD; 2000/60/EC) fait part que la quantification d'une FEA devrait incorporer l'intensité en termes de mesure de la magnitude/abondance, et la fréquence d'occurrence sur une période de temps spécifique. Dans ce chapitre, une revue des méthodes de détection du couvert nuageux en milieux lacustre et océanique est présentée suivie d'une revue des approches de caractérisation de la phénologie des FEA et des méthodes d'imputation de données manquantes.

2.1 Détection du couvert nuageux en milieux lacustre et océanique

En milieu océanique, les masques de nuages se basent habituellement sur l'hypothèse que la réflectance de l'eau est presque nulle dans le PIR. Cette approche est cependant beaucoup moins efficace en présence d'éléments optiquement actifs dans l'eau tels que les sédiments en suspension et les organismes phytoplanctoniques. Le masque de nuage développé par la NASA se base sur un seuil dans le PIR appliqué après correction atmosphérique. Le seuil fixé est tenu assez bas afin d'assurer un traitement de bonne qualité. Cependant, des éléments optiquement actifs telle qu'une efflorescence d'algues peuvent s'avérer être faussement exclus des données de par l'effet additif qu'ils apportent à la signature spectrale de l'eau dans le PIR.

Le masque de nuage développé par Wang & Shi (2006), lors d'une efflorescence d'algues sur la mer Baltique, utilise un seuillage dans le CLIR (courtes longueurs d'ondes de l'infrarouge) et une méthode hybride combinant ce dernier avec celui développé par MODIS Atmosphere Group. Cette méthode a montré les résultats les plus concluants comparativement aux algorithmes utilisant des seuils dans le PIR (Banks & Mélin, 2015). Également, afin d'éviter l'influence de la turbidité des eaux, Wang & Shi (2006) ont proposé un seuil de réflectance dans le CLIR qui a su discriminer les eaux turbides des nuages. Cet algorithme a été utilisé dans plusieurs études (Wang *et al.*, 2009; Chen *et al.*, 2014).

Par ailleurs, Chen & Zhang (2015) fait part qu'il est nécessaire de discriminer les nuages de la réflectance spéculaire (sunglint) affectant le signal de la couleur des eaux. Plusieurs travaux portent sur le sujet (Doerffer *et al.*, 2008; Hu, 2011; Steinmetz *et al.*, 2011). L'effet de réflectance spéculaire (sunglint) a une forte influence sur la réflectance dans le PIR et le CLIR. De ce fait, les algorithmes utilisant des seuils dans le PIR et le CLIR peuvent faussement identifier ces pixels comme étant des nuages.

Des méthodes basées sur la variabilité spatiale de la réflectance au-dessus de l'atmosphère dans les longueurs d'ondes du visible (bande à 555 nm (Martins *et al.*, 2002)), du PIR (centré à 865 nm (Nicolas *et al.*, 2005)) et du CLIR (centré à 2130 nm (Chen & Zhang, 2015)) sont également proposées dans la littérature. L'idée est de définir les valeurs d'écart-type de chaque fenêtre de pixels (par ex. 2X2 dans les travaux de Chen & Zhang (2015)) et de les comparer à un seuil établi permettant de classer ces pixels comme étant des nuages ou non. Bien que cette technique puisse

efficacement identifier les nuages (Chen & Zhang, 2015), une efflorescence d’algues présentant une forte variabilité spatiale dans les longueurs d’onde du visible et du PIR peut être encore une fois faussement discriminée comme étant des nuages (Lubac & Loisel, 2007).

Cet algorithme a été proposé par Nordkvist *et al.* (2009) qui utilise une série de seuils dans les longueurs d’onde du visible et du PIR. La présence d’aérosols peut influencer considérablement la détection des nuages via l’utilisation de cet algorithme (Nordkvist *et al.*, 2009). Un produit de masque de nuage (MOD35) a été développé par des membres de MODIS atmosphere science à une résolution spatiale de 1 km et de 250 m (Platnick *et al.*, 2017). L’algorithme utilise un seuil de réflectance dans le PIR (threshold of Rayleigh-corrected reflectance at the near infrared band of 869nm) ce qui réduit son efficacité lorsqu’il est appliqué en eaux turbides (Robinson *et al.*, 2003). Banks & Mélin (2015) ont d’ailleurs montré que ce produit masquait incorrectement de larges zones d’efflorescence d’algues sur une image prise dans la mer Baltique, similairement à la méthode standard de seuil dans le PIR et à la méthode de seuillage dans le visible/PIR développé par Nordkvist *et al.* (2009). Banks & Mélin (2015) mentionnent également que le problème de masquage de ces 3 méthodes est essentiellement dû au fait qu’ils utilisent une approche de seuil dans le PIR.

2.2 Définition des caractéristiques phénologiques d’une FEA

Aucune définition universelle d’une FEA n’est présente (Carvalho *et al.*, 2013) puisqu’elle se réfère à une abondance massive d’une classe particulière d’algues et qu’elle se rattache au niveau trophique du milieu. Elle se caractérise généralement par une forte abondance phytoplanctonique et une composition inégale de la communauté phytoplanctonique. En effet, une FEA est souvent dominée par une ou deux espèces et est accompagnée d’une diminution des autres espèces phytoplanctoniques présentes dans le plan d’eau (Agence Française de Sécurité Sanitaire des Aliments & Agence Française de Sécurité Sanitaire de l’Environnement et du Travail, 2006). À forte concentration, les espèces dominantes sont souvent d’origine cyanobactérienne, favorisant la présence de FEA mono-spécifiques et denses durant l’été en milieux eutrophes (Carvalho *et al.*, 2011). À cet effet, plusieurs travaux se réfèrent à des fleurs d’eau d’algues bleu-vert, des fleurs d’eau de cyanobactéries ou simplement des cyanobactéries pour désigner une FEA. Au Québec, le MDDELCC définit

une FEA comme étant caractérisée par une densité très importante d'algues microscopiques ou de cyanobactéries résultant à une prolifération excessive de ces micro-organismes.

Plusieurs approches ont été utilisées afin de quantifier la phénologie des FEA dans les lacs par le biais d'images satellitaires. Brody *et al.* (2013) ont comparé trois méthodes permettant de déterminer la date d'initiation d'une FEA phytoplanctonique dans l'océan Atlantique. La première méthode consiste à calculer un ratio de changement (*rate of change method*) en interpolant les séries temporelles afin de générer les valeurs manquantes. L'idée est ensuite d'identifier tous les maxima locaux dans la série comme étant l'intensité maximale d'une FEA, et de définir la date d'initiation comme la dérivée $d \text{ chl-a} / dt$ maximale de chaque maxima. La deuxième méthode est d'établir la somme cumulative (*cumulative sum method*) de la concentration en chl-a de la série temporelle. La date d'initiation correspond au moment où la courbe de la somme cumulée atteint un certain pourcentage de la biomasse totale. La dernière méthode, largement utilisée dans la littérature, consiste à fixer un seuil d'initiation d'une FEA (*threshold method*). À titre d'exemple, les travaux de Palmer *et al.* (2015) exploitent le logiciel TIMESAT permettant de définir le temps d'initiation et de déclin d'une FEA par deux approches: i) la date à laquelle une variable (ici, la concentration en chl-a) passe au-delà ou en-deçà d'un certain seuil défini comme le pourcentage de l'intensité maximal détectée, ou ii) la date à laquelle une variable passe au-delà ou en-deçà d'un certain seuil fixé par l'utilisateur. En milieu océanique, l'initiation d'une FEA est fréquemment définie en fixant le seuil comme étant la concentration en chl-a dépassant de 5 % la concentration médiane Thomalla *et al.* (2011); Cole *et al.* (2012); Racault *et al.* (2012). Palmer *et al.* (2015) ont utilisé cette approche et ont assigné un seuil d'initiation de 10 mg m^{-3} ; le seuil de déclin d'un épisode a été fixé à 12 mg m^{-3} (à une date suivant le pic d'intensité). L'auteur rappelle cependant que l'utilisation de ces seuils sur d'autres lacs n'est pas appropriée puisque la dynamique phytoplanctonique est fortement variable. D'autres études ont établi un seuil basé sur l'indice FAI (*floating algae index*) indiquant la présence ou l'absence d'écume à la surface afin de déterminer le temps d'initiation et la durée de prolifération d'une FEA (Hu *et al.*, 2010; Duan *et al.*, 2014). Dans le cas des travaux de Duan *et al.* (2014), ils ont défini la date d'initiation comme étant la première observation dans l'année où l'étendue de la FEA (définie selon l'indice FAI) couvre une superficie plus grande que 150 km^2 . Par ailleurs, Obenour *et al.* (2014) ont défini un épisode de FEA comme étant la période de 30 jours présentant l'indice de cyanobactéries (CI) la plus élevée, où un CI correspond à environ à 1.2×10^{20} cellules de cyanobactéries.

Les états d'initiation et de déclin d'une FEA peuvent être également établis selon des seuils décisionnels liés au risque associé aux FEA, et plus particulièrement aux cyanobactéries. Warren (2011) cite plusieurs exemples de seuils décisionnels (figure 2.1). Au Québec le MELCC (2015) considère qu'un plan d'eau est touché par une FEA lorsque la densité des cellules atteint 20 000 cellules mL^{-1} . Silva (2014) se réfère au seuil de 100 000 cellules mL^{-1} au-dessus duquel l'Organisation Mondiale de la Santé préconise l'interdiction de la baignade (World Health Organization, 2003). Un seuil de 2 000 cellules mL^{-1} est défini pour l'eau de consommation selon Bartram & Chorus (1999), qui considèrent également la densité de 100 000 cellules mL^{-1} comme seuil d'alerte modéré d'usages à des fins récréatives. L'institut national de santé publique du Québec (INSPQ) a retenu également ce seuil pour les eaux récréatives. Le problème demeure entier quant à la précision de la conversion de ces seuils décisionnels en biomasse algale qui reflète mieux le risque associé à la présence des cyanobactéries, celles-ci montrant une grande diversité de taille et de morphologie. En effet, la quantité de cyanotoxines est mieux corrélée à la biomasse qu'à la densité. Finalement, certains auteurs utilisent des modèles conceptuels tels que des modèles Gaussien (Zhai *et al.*, 2011) et sinusoidal (Sapiano *et al.*, 2012) pour estimer la date d'initiation et de déclin mais ces derniers peuvent perdre de leur robustesse dans des cas où le cycle saisonnier diffère du modèle. Duan *et al.* (2014) ont établi la date d'initiation comme étant le jour défini par la première observation de l'étendue d'une FEA couvrant plus du 1/15 de la superficie totale du lac Taihu en Chine ($> 150 \text{ km}^2$).

La durée d'un épisode de FEA peut se définir comme le nombre de jours entre la date d'initiation et de fin d'une FEA. Les travaux de Duan *et al.* (2009) ont par ailleurs défini la durée d'un épisode comme le nombre de mois auxquels un épisode de FEA a été détecté pour une année spécifique.

En termes de mesure sur la magnitude ou l'abondance d'un épisode, la grande majorité des études portant sur la phénologie des FEA s'est surtout intéressée à une quantification de l'intensité maximale donnée par la valeur maximale de chl-a (Kahru *et al.*, 2011; Palmer *et al.*, 2015) ou par une tendance centrale (Edwards & Richardson, 2004). C'est le cas par exemple d'études de séries temporelles des valeurs médianes des concentrations en chl-a des pixels d'eau valides (Binding *et al.*, 2011; Matthews, 2014).

Certains autres descripteurs ont été définis dans la littérature, incluant notamment le pourcentage de superficie des niveaux trophiques et le jour de la concentration maximale en chl-a. Le premier descripteur, défini par Binding *et al.* (2011), se base sur la définition des statuts trophiques

Densité (cellules m^{-1})	Biomasse* ($\mu\text{g Chl-a L}^{-1}$)	Actions à entreprendre en fonction du risque associé	Référence
500	0.25	Augmenter la fréquence d'échantillonnage. Échantillonner à une fréquence bi-mensuelle.	Carmichael 2001 Watzin <i>et al.</i> 2006
2 000	1.00	Niveau 1 «Prudence» en Corée du Sud. Doser les cyanotoxines. Alerte de niveau 1 selon l'OMS.	Ahn <i>et al.</i> 2007 Chorus et Bartram 1999
4 000	2.00	Échantillonner à une fréquence hebdomadaire.	Chorus et Bartram 1999
5 000	2.50	Niveau 2 «Alerte» en Corée du Sud	Ahn <i>et al.</i> 2007
10 000	5.00	Protéger et suivre la qualité de l'eau brute à l'entrée de l'usine de traitement de l'eau potable. Mettre en place des traitements pour éliminer les cellules et les toxines dans l'eau brute. Seuil proposé par le Service municipal des eaux du Québec.	Ellis 2008
15 000	7.50	Émettre un avis de non-consommation en absence de traitement adéquat.	Carmichael 2001
20 000	10.00	Densité cellulaire néfaste pour la santé (limite inférieure). Éviter la baignade. Plan d'eau «touché» par une fleur d'eau. Seuil utilisé au Québec.	Oliver et Ganf 2000 OMS 1998 MDDEP 2010a
100 000	50.00	Utiliser une source d'eau potable alternative. Échantillonner à une fréquence journalière. Interdire la baignade. Alerte de niveau 2 selon l'OMS.	Chorus et Bartram 1999
1 000 000	500.00	Niveau 2 «Épidémie» en Corée du Sud	Ahn <i>et al.</i> 2007

* Valeur estimée à partir du ratio 100 000 cellules m^{-1} pour 50 $\mu\text{g Chl-a L}^{-1}$ en condition de dominance par la cyanobactérie *Microcystis*, donné dans Chorus et Bartram (1999).

Figure 2.1 – Exemples de seuils décisionnels pour l'évaluation du risque associé à la présence de cyanobactéries (tiré de Warren (2011)).

de Vollenweider (1982) pour catégoriser chaque pixel de l'image, permettant de déterminer la proportion de chaque catégorie par rapport au nombre de pixel total. Trois statuts trophiques sont utilisés:

1. Conditions hypereutrophes: $[\text{chl}_a] > 25 \text{ mg/m}^3$
2. Conditions eutrophes: $[\text{chl}_a] > 8 \text{ mg/m}^3$ et $[\text{chl}_a] < 25 \text{ mg/m}^3$
3. Conditions mésotrophes et oligotrophes: $[\text{chl}_a] < 8 \text{ mg/m}^3$

Le jour de la concentration maximale en chl-a, également défini par Binding *et al.* (2011), correspond au jour où une concentration maximale de chl-a a été observée pour chaque année.

2.3 Sommaire

La mise en place d'un algorithme semi-automatisée du suivi des concentrations en chl-a en temps quasi-réel nécessite la détection adéquate du couvert nuageux pour des capteurs optiques. De ce fait, peu de masques de nuages existants ont été développés spécifiquement dans la cadre d'études sur la qualité de l'eau. Parmi ceux existants, la plupart ont été calés et validés en milieu océanique, où l'apport en éléments optiquement actifs est beaucoup moins important qu'en milieux lacustre et fluviale. De plus, seul le produit récemment développé par MODIS Atmosphere Science propose un masque de nuage à une résolution spatiale de 250 m. Ce produit est somme toute basé sur le masque MODIS à 1 km de résolution spatiale, et ce dernier semble difficilement parvenir à performer adéquatement en eaux turbides (Robinson *et al.*, 2003; Banks & Mélin, 2015). Il s'avère donc nécessaire de développer un masque de nuage spécifiquement voué aux eaux libres qui puisse bien performer en milieux lacustres afin de permettre *a posteriori* l'étude de la qualité de l'eau via l'imagerie satellitaire.

D'autre part, les épisodes de FEA devraient être considérées comme des événements caractérisés par leur phénologie. La définition même d'une FEA dans le cadre de ce projet se réfère à une forte abondance en biomasse phytoplanctonique ou cyanobactérienne dans le milieu. Notons qu'à fortes concentrations, les espèces dominantes sont souvent d'origine cyanobactérienne. La définition du seuil d'initiation d'un épisode de FEA est importante puisqu'elle sous-tend à la définition des autres variables phénologiques. Ce seuil est habituellement fixé selon des seuils décisionnels liés au risque associé aux FEA. De manière générale, la fréquence d'apparition et l'intensité du phénomène sont les deux variables priorisées dans la caractérisation de la phénologie des FEA parce qu'il s'agit de paramètres faciles à évaluer dans le cadre d'un suivi de la qualité de l'eau par échantillonnage *in situ* ou via l'utilisation de sondes de qualité de l'eau. L'utilisation de l'imagerie satellitaire apporte une précision dans l'estimation spatio-temporelle de ces paramètres et permet l'évaluation de d'autres descripteurs phénologiques telle que l'envergure du phénomène.

Chapitre 3

Estimation phénologique des FEA

Les concentrations en chl-a sur les eaux intérieures du Québec méridional sont issues de données satellitaires du capteur MODIS (Moderate Resolution Imaging Spectroradiometer). Ce chapitre présente la procédure semi-automatisée d'estimation de la concentration en chl-a, décrit les étapes de prétraitement des images, le modèle d'estimation de la concentration en chl-a et présente l'algorithme basé sur l'analyse discriminante utilisé pour détecter le couvert nuageux. La description des variables phénologiques issues des concentrations en chl-a est ensuite présentée.

3.1 Données satellitaires

L'imagerie du capteur MODIS, un spectroradiomètre à résolution modérée, a été utilisée dans le développement d'une approche satellitaire d'estimation de la concentration en chl-a. Ce capteur est installé à bord du satellite TERRA (EOS AM-1) qui fut lancé en 1999. La mission spécifique de TERRA est de surveiller l'état de l'environnement de la Terre et les changements climatiques. Les caractéristiques spécifiques de ce capteur établissent un bon compromis entre la haute résolution temporelle qu'il possède et sa résolution spatiale, tout en possédant une très bonne résolution spectrale et radiométrique (tableau 3.1). Les images sont accessibles gratuitement (Goddard Space Flight Center, 2014) et sont généralement disponibles 24 heures après le passage du satellite. L'accessibilité et la fréquence temporelle de ces données leur permettent d'effectuer un suivi de l'état des lacs en temps quasi-réel. Un suivi d'épisodes de FEA sur une base journalière est un atout pour

tout gestionnaire de l'eau souhaitant effectuer un suivi quotidien de la qualité de l'eau et limiter les risques possibles sur la santé humaine. De plus, l'étude de la dynamique temporelle des FEA requière une plus haute fréquence temporelle afin de bien caractériser l'événement, ce qui est le cas pour ce capteur qui montre un temps de revisite de 1 jour sur le Québec méridional.

Tableau 3.1 – Caractéristiques du satellite TERRA et du capteur MODIS.

Satellite TERRA (EOS AM-1)	
Opérateur	NASA
Lancement	1999
Espérance de vie	2020
Altitude	705 km (Noeud ascendant)
Inclinaison	98,5 °
Période	99 minutes
Capteur MODIS	
Bandes spectrales	36 bandes (0,4 µm- 14,4 µm)
Résolution spatiale	250 m (bandes 1-2)
	500 m (bandes 3-7)
	1 km (bandes 8-36)
Résolution radiométrique	12 bits
Temps de revisite	1-2 jours
Largeur de fauchée	2330 km

La base de données satellitaires couvre l'ensemble du Québec méridional (44°- 50°Nord et 67°- 80°Ouest). Cette zone est présentée à la figure 3.1.

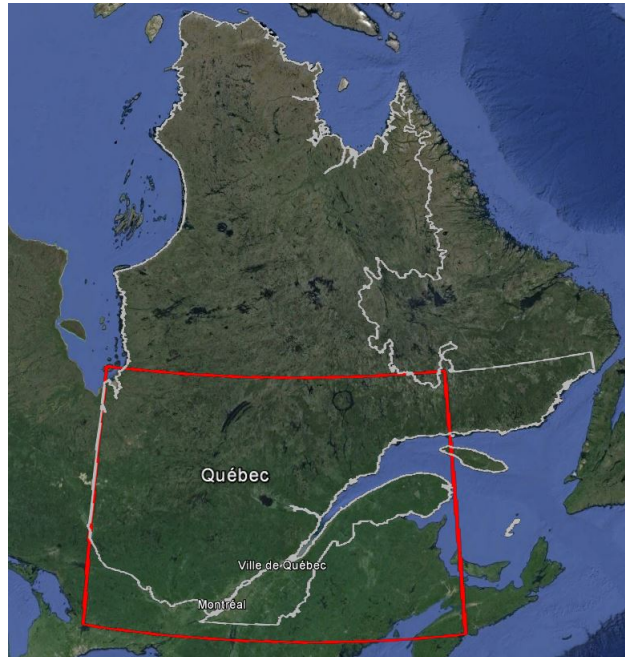


Figure 3.1 – Zone couverte par la base de données satellitaires (source: Google Earth).

3.2 Prétraitement des données

Les sept premières bandes du capteur MODIS couvrent les longueurs d'onde du visible au CLIR sur le spectre électromagnétique (tableau 3.2). Ces longueurs d'onde sont très sensibles à l'activité bio-optique de la chl-a ainsi qu'à d'autres éléments optiquement actifs dans l'eau telle que la matière en suspension (MSS) et la matière organique dissoute colorée (MODC). La résolution spatiale des bandes 1-2 est de 250 m alors que celle des bandes 3-7 est de 500 m. Afin d'améliorer la résolution spatiale de ces dernières, une méthode de désagrégation développée par l'équipe du Centre Canadien de Télédétection Spatiale (CCTS) a été utilisée afin de désagréger les bandes 3-7 d'une résolution de 500 m à une résolution de 250 m (Trishchenko *et al.*, 2007). Cette approche a été validée en utilisant des données à haute résolution spatiale (Landsat ETM+ (30 m)) et a montré que la procédure de désagrégation n'altère pas les propriétés radiométriques de la scène. Cette résolution spatiale à 250 m améliore la précision de détection des FEA et permet l'acquisition de données sur des plans d'eau de plus petites tailles comparativement aux anciennes générations de capteurs (AVHRR, SeaSAT, et SeaWIFS). Les bandes 1-7 ont ensuite été projetées sous la projection Conique Conforme de Lambert (LCC) qui est utilisée au Canada pour des besoins cartographiques à une petite échelle (échelle nationale). Ce système de projection conserve les formes et présente des distorsions minimales près des deux parallèles de référence (47°Nord et 77°Nord) dont fait partie la

province du Québec. Une correction atmosphérique a ensuite été effectuée sur les bandes MODIS au moyen de l’algorithme SMAC (*Simplified Model for Atmospheric Correction*). Ce modèle de correction atmosphérique utilise sept variables d’entrée dérivées directement des images MODIS: la réflectance à la limite de l’atmosphère, l’angle zénithal solaire, l’angle de vue zénithal, l’angle relatif azimutal, l’épaisseur optique des aérosols à 550 nm, la teneur de la vapeur d’eau atmosphérique et la teneur en ozone atmosphérique. Le modèle SMAC est habile à corriger l’effet de bruit additif et de bruit multiplicatif engendrés par des perturbations atmosphériques (nuage, brume, humidité, *etc*). Par ailleurs, plusieurs modèles déterministes tels que le 6S (*Second Simulation of Satellite Signal in the Solar Spectrum*), le MODTRAN (*MODerate resolution atmospheric TRANsmittance and radiance code* (Berk *et al.*, 1999)) et le ATCOR (*ATmospheric CORrection* (Richter, 1996)) peuvent s’avérer être moins efficaces en présence d’aérosols (Bailey & Werdell, 2006; Gordon, 1997) ou en présence d’eaux turbides (Dekker *et al.*, 1997; Lavender *et al.*, 2005), ce qui peut mener à des erreurs considérables dans l’estimation de la réflectance. De plus, les modèles empiriques tels que le DOS (*Dark Object Subtraction* (Chavez, 1988)) et le COST (COSine Transmission for atmospheric correction (Chavez, 1996)) corrigent uniquement le biais additif causé par la brume et non le biais multiplicatif causé par les aérosols, la vapeur d’eau et l’ozone dans l’atmosphère. Finalement, un masque de terre (El-Alem *et al.*, 2012) a été appliqué sur l’ensemble des images afin de retirer les pixels de terre mais également les pixels mixtes (eau-terre) qui peuvent mener à des erreurs considérables dans l’estimation de la concentration en chl-a.

Tableau 3.2 – Caractéristiques des bandes 1-7 du capteur MODIS.

Bande (résolution)	Largeur des bandes (nm)	Nom des bandes
1 (250 m)	620 - 670	Visible (rouge)
2 (250 m)	841 - 876	PIR
3 (500 m)	459 - 479	Visible (bleu)
4 (500 m)	545 - 565	Visible (vert)
5 (500 m)	1230 - 1250	PIR
6 (500 m)	1628 - 1652	CLIR
7 (500 m)	2105 -2155	CLIR

3.3 Détection du couvert nuageux

Le modèle d'estimation de la chl-a a été calé et validé en utilisant des images sans couvert nuageux et sur un certain nombre de lacs spécifiques. Une étude à l'échelle régionale et sur un ensemble considérable d'images nécessite la mise en place d'un masque de nuage performant en milieu lacustre. En effet, la détection du couvert nuageux est nécessaire dans la mise en place d'un outil de surveillance régionale semi-automatisé. La procédure de pré-traitement et de désagrégation des bandes 1-7 développée par le CCTS génère un produit de masque de nuage à une résolution spatiale de 250m, mais celui-ci ne présente généralement pas des performances optimales sur les eaux intérieures (Figure 3.2, centre). Par ailleurs, les produits de masque de nuage disponibles pour les images MODIS sont générés à des résolutions spatiales de 250m et 1km (Platnick *et al.*, 2017). Cependant, le produit à 1km de résolution limite le suivi de la qualité de l'eau sur les lacs de petite à moyenne taille, et semble être inefficace lorsque qu'il est appliqué en eaux turbides (Figure 3.2, droite). Le produit MODIS à 250m de résolution incorpore les résultats des tests à 1km de résolution afin de maintenir une consistance avec le produit à 1km de résolution spatiale. Ce produit semble donc, *a priori*, présenter des problèmes similaires au produit à 1km de résolution spatiale.

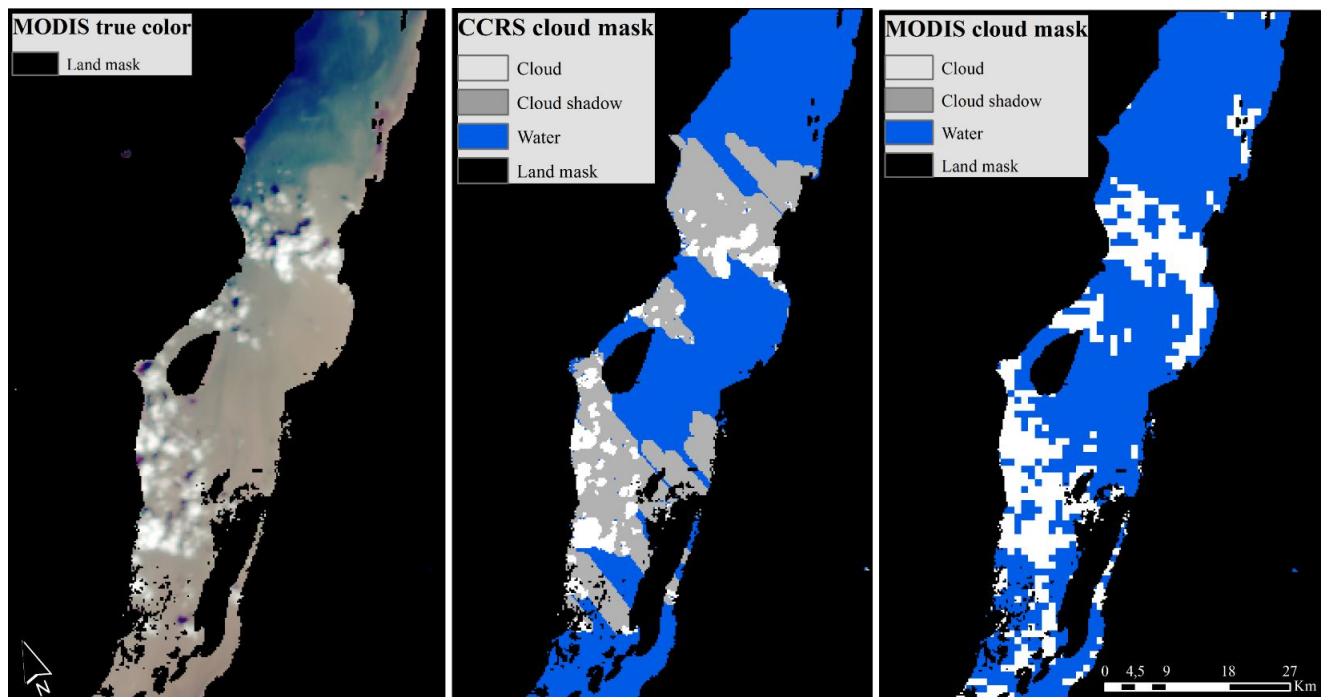


Figure 3.2 – (a) Image MODIS en vraie couleur, (b) masque de nuage développé par le CCTS, et (c) masque de nuage développé par le groupe MODIS Atmosphere.

3.3.1 Description du modèle

L'analyse linéaire discriminante (ALD) est une alternative intéressante aux algorithmes par seuil. Cette méthode est souvent utilisée dans le cadre de classifications de la couverture de sol (Friedl & Brodley, 1997; Xia *et al.*, 2014; Priedītis *et al.*, 2015) ou pour le développement d'indices d'eau liquide (Fisher & Danaher, 2013). Les techniques multivariées fournissent une information plus riche et plus complète au modèle prédictif et sont favorisées aux algorithmes par seuil pour générer un modèle discriminant optimal. L'ALD a été originellement proposée par Fisher (1936) pour des scénarios à deux groupes ($g = 2$) et ses résultats ont été étendus à plusieurs groupes ($g \geq 2$) par Rao (1948); Radhakrishna Rao (1952). L'analyse se base sur une matrice de variables explicatives quantitatives X de taille $(m \times n)$, où n objets sont décrits par m descripteurs. La variable réponse y correspond au critère de classification. L'ALD consiste à trouver une projection qui minimise la variation intra-classes tout en maximisant la variation inter-classes. Ce type de problème fait appel aux vecteurs propres et aux valeurs propres de la matrice $V^{-1}A$:

$$(V^{-1}A - \lambda_k I)u_k = 0 \quad (3.1)$$

Où V correspond à la matrice de dispersion intra-groupe et A à la matrice de dispersion inter-groupe. Nous avons au départ une variable dépendante qualitative correspondant aux classes: 1) Eau 2) Brume (*a priori*) 3) Nuage. Ainsi que des variables indépendantes correspondant aux valeurs spectrales des bandes MODIS-D-250. Les variables indépendantes sont transformées pour respecter l'hypothèse de normalité. Le modèle discrimine en premier lieu les trois classes [eau, brume (*a priori*), nuage] en assumant que les variables indépendantes sont distribuées normalement et possède la même matrice de covariance pour chaque classe (Figure 3.3). L'eau claire se distingue facilement du couvert nuageux de par sa faible réflectance spectrale dans le visible et le PIR. Cependant, l'eau contenant des éléments optiquement actifs tels que la MES, la CDOM et la chl-a est plus difficile à distinguer de la brume de par l'effet additif que cette dernière apporte à la signature spectrale de l'eau pure. Afin de réduire les erreurs de commission qui sont particulièrement nuisibles dans le cadre d'études sur la qualité de l'eau, une seconde ALD est appliquée sur la classe brume (*a priori*) afin de distinguer la brume des eaux turbides ou à moyenne-fortes concentrations en chl-a. Les résultats sont ainsi séparés en trois classes: 1) Eau (teneur en MES ou en MODC élevée) 2) Eau (teneur en chl-a élevée) 3) Brume. La classification de ces trois groupes a été effectuée à l'aide de

l’algorithme d’estimation de la concentration en chl-a développé pour les eaux intérieures (El-Alem *et al.*, 2014a), permettant de sélectionner des événements importants d’efflorescence d’algues, ainsi que des lacs connus à présenter de fortes turbidités.

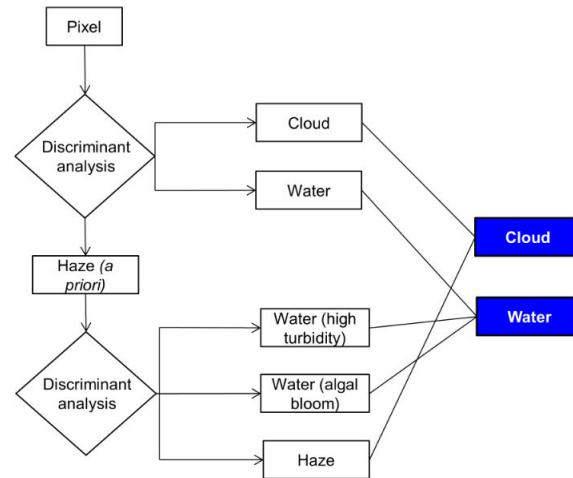


Figure 3.3 – Méthode détaillée utilisée pour distinguer les classes [nuage - eau] avec l’analyse linéaire discriminante.

3.3.2 Calage et validation

Un ensemble d’échantillons provenant de 26 images MODIS ont été sélectionnées durant la saison de mai à novembre entre 2000 et 2015 (Tableau (3.3)). Des échantillons d’eau libre (lacs, rivières, baies, estuaires) ont été sélectionnées pour chaque scène MODIS et pour qu’ils soient représentatifs des classes trophiques en eaux intérieures (classes oligotrophe, mésotrophe, eutrophe, hypereutrophe). Conjointement à l’information connue dans la littérature concernant les lacs fortement turbides, l’estimateur des concentrations en chl-a a été utilisé afin de distinguer l’eau claire, les FEA et l’eau turbide. Les échantillons couvrent ainsi les classes trophiques basées sur des faibles concentrations en chl-a ($0.1\mu gL^{-1}$) jusqu’aux fortes concentrations en chl-a ($1000\mu gL^{-1}$).

Les échantillons ont ensuite été partitionnés en deux ensembles, le premier servant au calage et contenant 70 % des données (6186 pixels) et le deuxième servant à la validation et contenant 30 % des données (2651 pixels). La performance du modèle a été évaluée en utilisant une validation croisée par Monte Carlo: la séparation aléatoire de l’échantillon de départ en données de calage et de validation a été répétée 10,000 fois afin d’obtenir une distribution du succès global et du coefficient Kappa (κ) issus de la classification (Figure (3.4)). À titre de validation qualitative, le modèle a été

Tableau 3.3 – Liste des images MODIS utilisées pour le calage et la validation.

Jour julien	Année
185-217-243-299	2000
262	2001
141-200-246-282	2002
133-195-231-267	2005
262	2007
136-189-234-293	2010
147-217-237-268	2013
170-201-234-266	2015
Nombre d'images:	26

appliquée sur différentes images MODIS afin d'évaluer la performance de l'algorithme de masque de nuage. Les images utilisées à cet effet n'ont pas été utilisées dans le calage / validation du modèle. Les scènes incluant des lacs et estuaires fortement turbides et lors de périodes avec des efflorescences d'algues ont été sélectionnées. L'algorithme d'estimation de la chl-a a également été appliqué sur les images afin de détecter la présence d'algues.

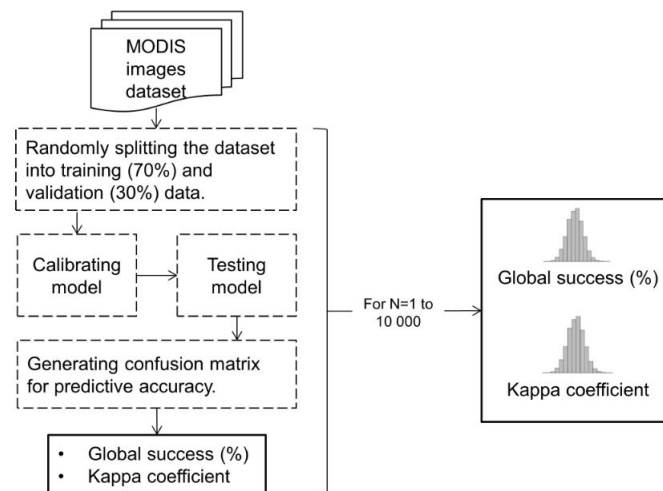


Figure 3.4 – Méthode détaillée utilisée pour estimer la distribution du succès global (%) et du κ de la classification en utilisant une validation croisée par Monte Carlo.

3.4 Algorithme d'estimation des concentrations en chl-a

L'algorithme d'estimation de la concentration en chl-a s'est développé en trois étapes. D'abord, la capacité des bandes MODIS 1-7 désagrégées à 250 m (MODIS-D-250) à effectuer le suivi des FEA sur des lacs de moyenne à petite taille a été validée (El-Alem *et al.*, 2012). Les modèles d'estimation

de la concentration en chl-a les plus utilisés dans la littérature (Karhu, Gitelson et FAI (*floating algae index*)) et le modèle APPEL (*APProach by ELimination*) ont été calés et validés à l'aide des données MODIS désagrégées et des données *in situ* de concentration en chl-a collectées par le MDDELCC sur quatre lacs au sud du Québec durant la période 2000-2008 (lac Champlain, lac Brome, lac William, Lac Nairne). Les résultats ont montré le potentiel des données MODIS à estimer la concentration en chl-a, et particulièrement les fortes concentrations (figure 3.5). Par ailleurs, nous supposons que la biomasse cyanobactérienne benthique ne contribue pas aux estimations obtenues par imagerie MODIS puisque les bordures des plans d'eau jusqu'à une largeur de 250 m ne sont pas considérées dans l'analyse.

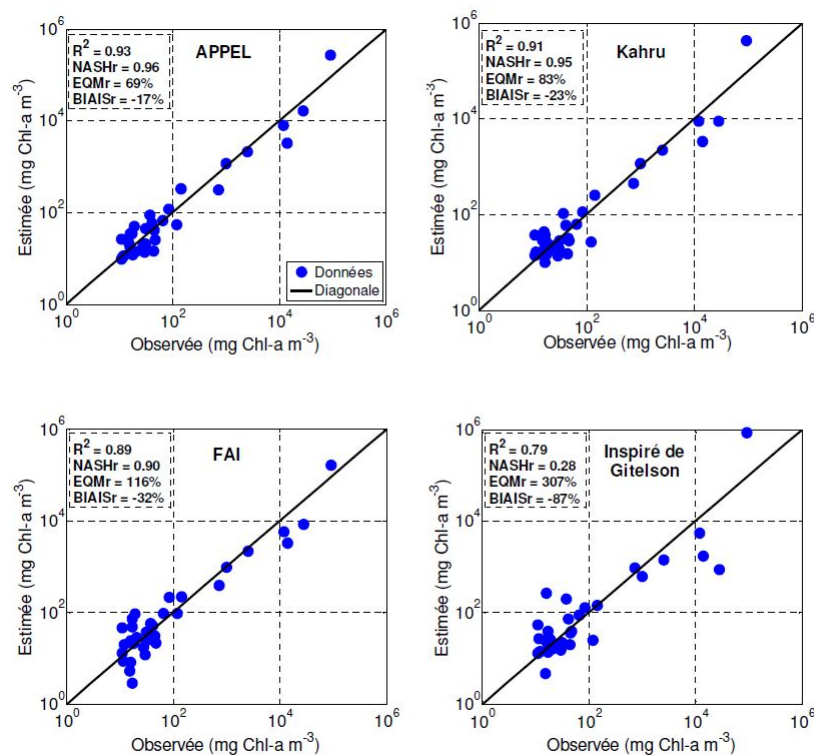


Figure 3.5 – Concentration en chlorophylle-a estimée à partir des quatre modèles comparée aux mesures *in situ* pour les données supérieures à 10 µg/L, avec les indices de performance des modèles (figure tirée de El-Alem (2015)).

Ensuite, un modèle adaptatif (MA) d'estimation de la chl-a a été développé afin de pallier à la faible performance d'estimation sous les conditions de faibles concentrations en chl-a (El-Alem *et al.*, 2014a). Ce modèle a montré que la combinaison des informations spectrales du visible aux CLIR peut faciliter la discrimination de la chl-a des autres éléments optiquement actifs présents dans les milieux lacustres, tels que la MSS et la MODC, et que le développement d'un modèle spécifique pour

trois classes de concentrations en chl-a améliore la précision des estimations. Le MA est constitué d'un classificateur, celui-ci ayant été calé à l'aide d'une méthode d'apprentissage par arbres de décision CART (*Classification and Regression Tree*), et de trois estimateurs, calés par régression linéaire / non-linéaire multiple pas-à-pas (stepwise). Ces trois estimateurs sont spécifiques à trois niveaux trophiques ou classes de biomasse: des eaux faiblement, moyennement et fortement chargées en chl-a. Les résultats ont montré que le MA était plus performant que les modèles APPEL et FAI à détecter les FEA durant leur phase d'initiation (à faible concentration) et ont également mis en évidence la reconnaissance a priori des classes de concentrations qui ont amélioré considérablement l'estimation des concentrations en chl-a (figure 3.6).

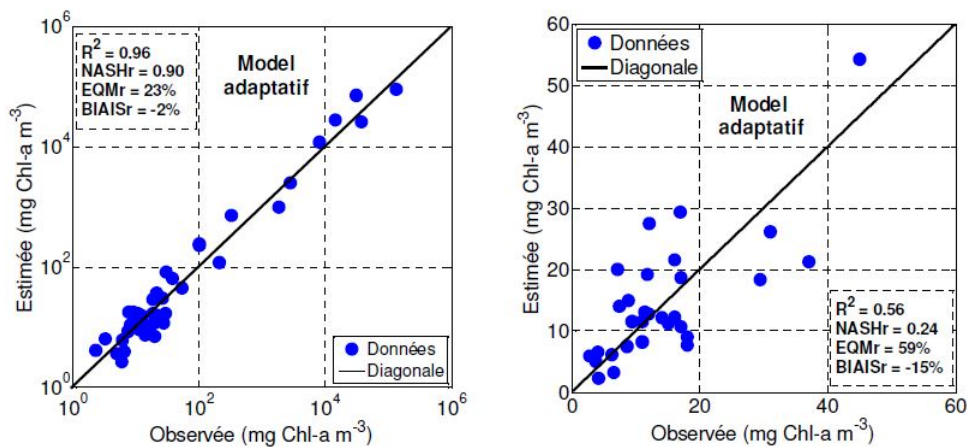


Figure 3.6 – Concentration en chlorophylle-a estimée à partir du modèle adaptatif comparée aux mesures in situ pour : A) l'ensemble de la base de données et B) en utilisant les concentrations inférieures à 50 mg Chl-a m⁻³, avec les indices de performance des modèles. (figure tirée de El-Alem (2015)).

Finalement, afin de réduire les sources d'erreur possibles provenant soit de la classification, soit de l'estimation des concentrations, un modèle basé sur la méthode de l'ensemble (MBME) a été développé et validé à l'aide des mêmes données utilisées pour le MA (El-Alem, 2015; El-Alem *et al.*, 2019). L'idée dans cette étape de l'étude était de valider si l'utilisation de méthodes statistiques avancées telles 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. En comparant la performance des deux modèles (MBME et MA), les résultats ont montré une performance assez similaire pour les fortes concentrations, mais une nette amélioration de l'estimation des faibles à moyennes concentrations avec le modèle MBME où les indices de performance ont doublé (figure 3.7).

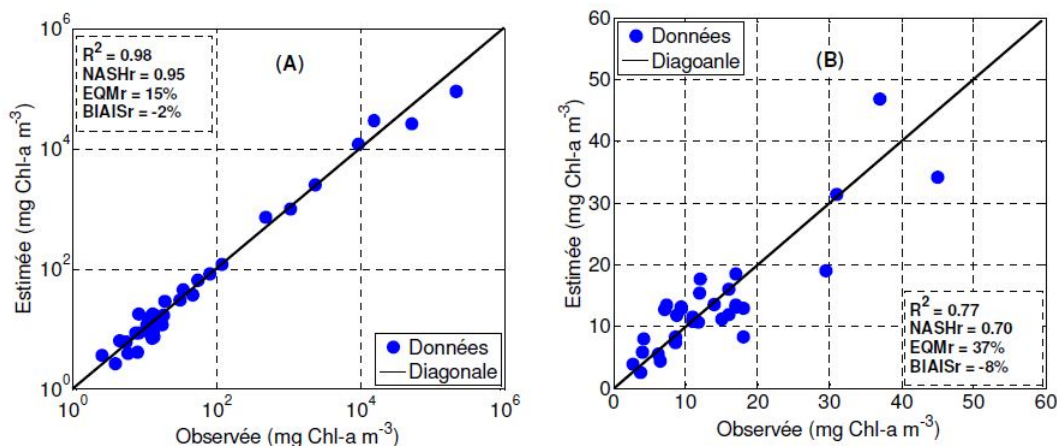


Figure 3.7 – Concentration en chlorophylle-a estimée à partir du modèle basé sur la méthode de l'ensemble comparée aux mesures in situ pour : A) l'ensemble de la base de données et B) en utilisant les concentrations inférieures à 50 mg Chl-a m^{-3} , avec les indices de performance des modèles. (figure tirée de El-Alem (2015)).

3.5 Caractérisation des FEA sur les lacs du Québec méridional

3.5.1 Choix des lacs

L'unité spatiale utilisée dans le cadre de la caractérisation phénologique est le lac. Une proportion minimale du lac visible sans couverture nuageuse a été fixée à 75 % (Duan *et al.*, 2014) afin de respecter la représentativité du phénomène sur le plan d'eau. Les pixels localisés en bordure des lacs ont été auparavant retirés jusqu'à une largeur de 250 m, ce qui réduit considérablement le nombre de valeurs aberrantes reliées aux effets de bordure (chalet, bateau, profondeur du lac) mais qui minimise également la superficie observable des FEA. Afin de disposer de lacs observables par MODIS et d'obtenir un échantillon représentatif de la réalité du plan d'eau, le nombre minimal de pixels est fixé à 30 (5×6) après le rognage des bordures en représentant le lac par son rectangle équivalent. Ce seuil minimal correspond à une superficie de 3.5 km^2 et doit couvrir un bloc de pixels de 1,75 km par 2 km. Le rectangle équivalent a été calculé via le coefficient de compacité de Gravelius K (Gravelius, 1914; Roche, 1963). Celui-ci correspond à un indice de forme défini comme le rapport du périmètre du lac à celui d'un cercle de même surface:

$$K = \frac{P}{2\sqrt{\pi A}} \quad (3.2)$$

où P correspond au périmètre du lac et A à l'aire du lac. La longueur l_1 et la largeur l_2 du rectangle équivalent sont :

$$\begin{aligned} A &= l_1 l_2 \\ P &= 2(l_1 + l_2) \\ l_1, l_2 &= \frac{K_g \sqrt{\pi} \sqrt{A}}{2} \pm \sqrt{\left(\frac{\pi K_g^2}{4} - 1\right) A} \end{aligned} \quad (3.3)$$

Où

$$K_g \geq \sqrt{\frac{4}{\pi}}$$

3.5.2 Caractérisation phénologique des FEA

Afin de quantifier les épisodes de FEA qui ont eu lieu entre 2000 et 2016, la phénologie des FEA a été définie à partir des concentrations en chl-a. L'apparition d'une FEA a été établie à partir du seuil décisionnel ϕ pour les eaux récréatives (niveau 1) tels qu'établis par le (World Health Organization, 2003) et le MELCC (2015) qui est d'une concentration en chl-a atteignant

$$\phi = 10 \mu\text{gL}^{-1} \cong 20\,000 \text{ cellules mL}^{-1} \quad (3.4)$$

L'utilisation de ce seuil, qui correspond à des effets mineurs sur la santé (effets irritatifs et allergiques), facilite une meilleure gestion du risque engendré par les FEA et facilite une meilleure compréhension de l'amplitude de l'événement puisqu'il est connu et déjà utilisé par plusieurs organismes de bassin versant, gestionnaires de l'eau, municipalités et usagers des plans d'eau. Ainsi, l'état d'initiation D (et de déclin F) d'une FEA sur un lac donné correspond à la journée à laquelle la concentration en chl-a $[chl_a]$ d'au moins un pixel passe significativement au-delà (ou en deçà) de ce seuil ($[chl_a] - 1,96s_e > 10 \mu\text{gL}^{-1}$). Pour un pixel i situé sur un lac donné, les jours de début D et de fin F d'une FEA sont définis comme suit:

$$D = \{\exists i \mid [chl_a]_i > \phi\} \quad (3.5)$$

$$F = \{\forall i, [chl_a]_i < \phi\} \quad (3.6)$$

Ainsi, pour un lac et une année donnés, les variables phénologiques établies sont:

1. La fréquence correspond au nombre de jours où une FEA est détectée.
2. L'intensité correspond à la concentration maximale de chl-a détectée en période de FEA.
3. L'étendue correspond à la superficie maximale occupée par une FEA par rapport à la superficie du lac.
4. La date d'initiation correspond au premier jour de l'année où une FEA est détectée.
5. La date de déclin correspond au dernier jour de l'année où une FEA est détectée.
6. La durée correspond au nombre de jours entre la date d'initiation et la date de déclin.

Chapitre 4

Résultats

4.1 Détection du couvert nuageux

Le tableau 4.1 présente la matrice de confusion générée à partir de l’algorithme de masque de nuage basé sur l’ALD. Les résultats montrent que la classification de l’eau et des nuages n’est problématique. Le modèle classe adéquatement les pixels d’eau avec 0 % de faux négatifs et sous-estime la détection des nuages dans 1 % des cas (faux négatifs), mais ces derniers sont classifiés comme de la brume ce qui n’est pas problématique dans le cadre d’étude sur la qualité de l’eau. En résumé, aucun pixel d’eau n’est faussement classifié ce qui s’avère être le défaut majeur aux algorithmes de masque de nuages actuels en présence d’éléments optiquement actifs dans l’eau tels que la MES, la CDOM et la chl-a (Banks & Mélin, 2015). Dans l’ensemble, la performance du modèle montre un κ moyen de 0.982 avec comme intervalle de confiance des valeurs entre 0.979 et 0.986. Le succès global moyen de la classification est de 99.0 % variant entre 98.8 % et 99.2 % (intervalle de confiance à 95%).

Afin de comparer l’algorithme de masque de nuage avec les produits MODIS à 250 m et 1 km, le succès global et le κ ont été calculés sur les deux classes [nuage, sans nuage] disponibles pour ces produits. Le tableau 4.2 présente les résultats obtenus pour les trois masques de nuage appliqués sur les mêmes ensembles de données de validation. La performance du produit MODIS à 250 m est évidemment plus élevée que celle du produit MODIS à 1 km. Cependant, le succès global de

même que le κ de l’algorithme basé sur l’ALD, qui présente la même résolution spatiale de 250 m, montrent des valeurs plus élevées que celles obtenues avec le produit MODIS à 250 m.

Tableau 4.1 – Résultats de la matrice de confusion avec un intervalle de confiance à 95 % des valeurs moyennes de succès global et de l’indice Kappa.

		Observed						
		Water	Fog	Cloud	Total	Commission error (%)	Success rate (%)	
Predicted	Water	1059	0	0	1059	0	100	
	Fog	0	236	0	236	0	100	
	Cloud	0	27	1329	1356	2	98	
	Total	1059	263	1329	2651			
	Omission error (%)	0	10.3	0				
	Success rate (%)	100	89.7	100			I.C. 95%	
	Global success (%)						98.8	99.0
κ						0.979	0.982	0.986

Tableau 4.2 – Résultats de la classification des deux produits MODIS (1 km et 250 m) et de l’algorithme basé sur l’analyse linéaire discriminante.

	MODIS 1 km	MODIS 250 m	ADL 250 m
Succès global (%)	91.3	95.3	99
κ	0.827	0.905	0.982

À titre de validation qualitative, l’algorithme de masque de nuage a été appliqué sur de nouvelles images MODIS-D-250 et comparé aux produits MODIS à 250 m et 1 km. La figure 4.1 présente les résultats obtenus sur la baie Missiquoi et le lac Champlain en période de forte efflorescence algale. Les masques de nuages MODIS semblent moins sensibles à la présence de brume ce qui peut mener à des surestimations lors de l’estimation de la concentration en chl-a à ces endroits. La figure 4.1 en présente un exemple et montre l’amélioration apportée par le nouveau masque de nuage dans la classification de la brume. Elle présente la baie Missiquoi (haut de l’image) en période de forte efflorescence algale (concentrations moyennes au-dessus de $50 \mu g L^{-1}$) et le lac Champlain (bas de l’image) partiellement couvert de nuage et de brume. Les trois masques de nuages sont ensuite présentés (produit MODIS 1 km produit MODIS 250 m, nouvel algorithme) de même que les concentrations en chl-a estimées sur les pixels d’eau restants. Les concentrations en chl-a ont été estimées via l’algorithme développé par El-Alem *et al.* (2014b). Les deux produits MODIS présentent une insuffisance à masquer la brume, ce qui mène à de fortes estimations de la concentration en chl-a sur ces pixels sans qu’il y ait, *a priori*, d’efflorescence algale à ces endroits.

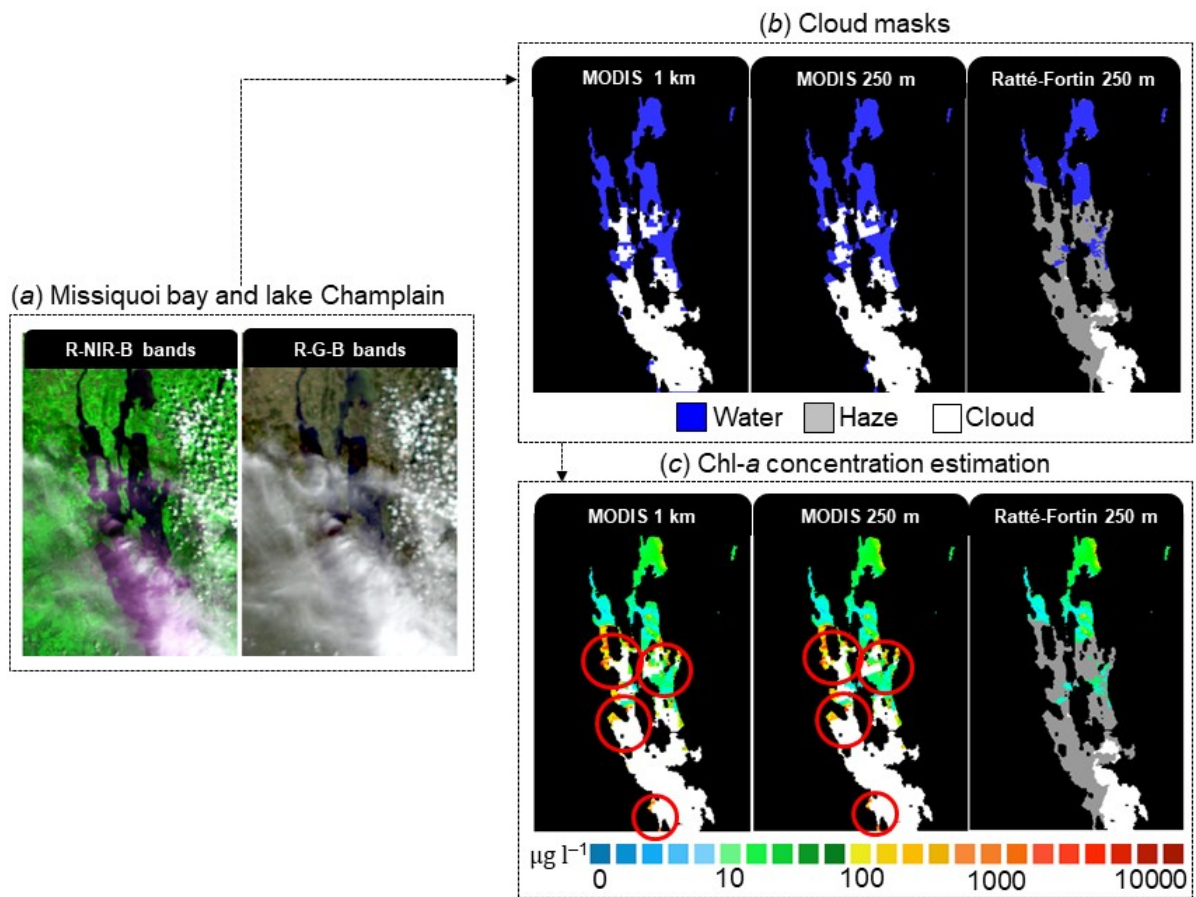


Figure 4.1 – (a) Bandes MODIS R-PIR-B et R-V-B de la baie Missisquoi et du lac Champlain, (b) les trois masques générés, et (c) les couches de concentrations en chl-a estimées avec les pixels d'eau restants. Les cercles rouges présentent des concentrations en chl-a élevées où il n'y a *a priori* aucune FEA.

La figure 4.2 présente les résultats obtenus sur le fleuve saint-Laurent et le lac Macamic. Cette portion du fleuve Saint-Laurent est caractérisé par une turbidité et une teneur en chl-a modérées, alors que le lac Macamic présente une forte turbidité. On constate que les produits MODIS montrent une performance beaucoup plus faible en présence d'eau très turbide, où les algorithmes les classifient faussement comme des nuages. La scène MODIS sur le fleuve Saint-Laurent à la figure 4.2 montre que la classification est améliorée avec le nouvel algorithme comparativement aux produits MODIS. Les eaux turbides localisées en bordure des zones urbaines sont mieux classifiées. Rappelons par ailleurs que les zones de transition de la terre à l'eau (pixels mixtes) ont été retirées jusqu'à une largeur de 250 m des bordures des lacs. De plus, sur les lacs de petite à moyenne taille et particulièrement ceux à forte teneur en turbidité, la mauvaise classification des produits de masque de nuage devient

un problème majeur en terme de données exploitables. La figure 4.2 (bas de l'image) présente une scène MODIS sur le lac Macamic d'une superficie de 45 km². Les produits de masque de nuage MODIS classifient faussement comme des nuages l'équivalent approximatif de 16 % de la superficie du lac.

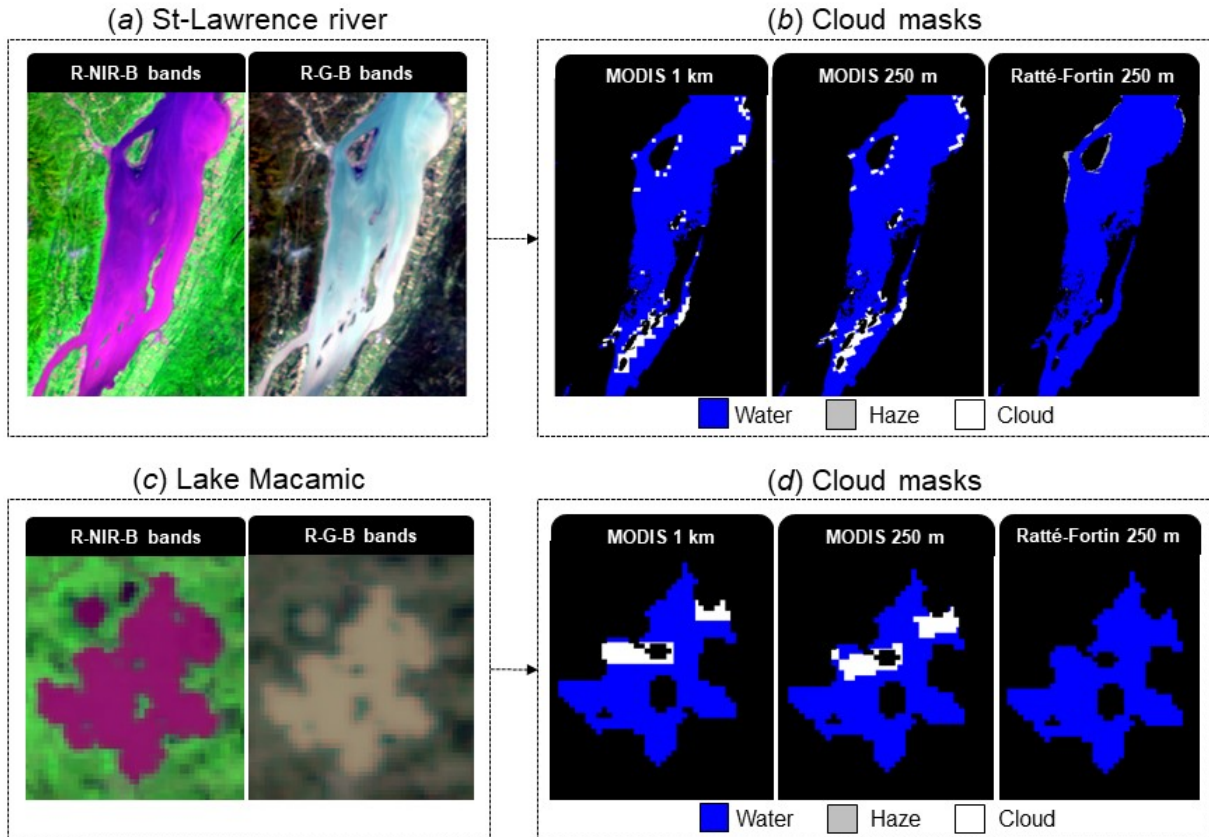


Figure 4.2 – Bandes MODIS R-PIR-B et R-V-B du fleuve Saint-Laurent (a) et du lac Macamic (c), ainsi que les trois masques de nuages (b) et (d) correspondant.

La figure 4.3 présente la performance du masque de nuage en présence de cirrus et de fine brume. La scène présentant la baie de Fundy (image du haut) montre la très bonne performance de l'algorithme à détecter la fine brume comparativement aux masques de nuage MODIS. La seconde scène sur le fleuve Saint-Laurent (image du bas) montre clairement la faible performance des produits MODIS à détecter le couvert nuageux en présence de cirrus. Comme il a été mentionné auparavant, cette faible sensibilité à la brume peut mener à de mauvaises estimations des paramètres de qualité de l'eau.

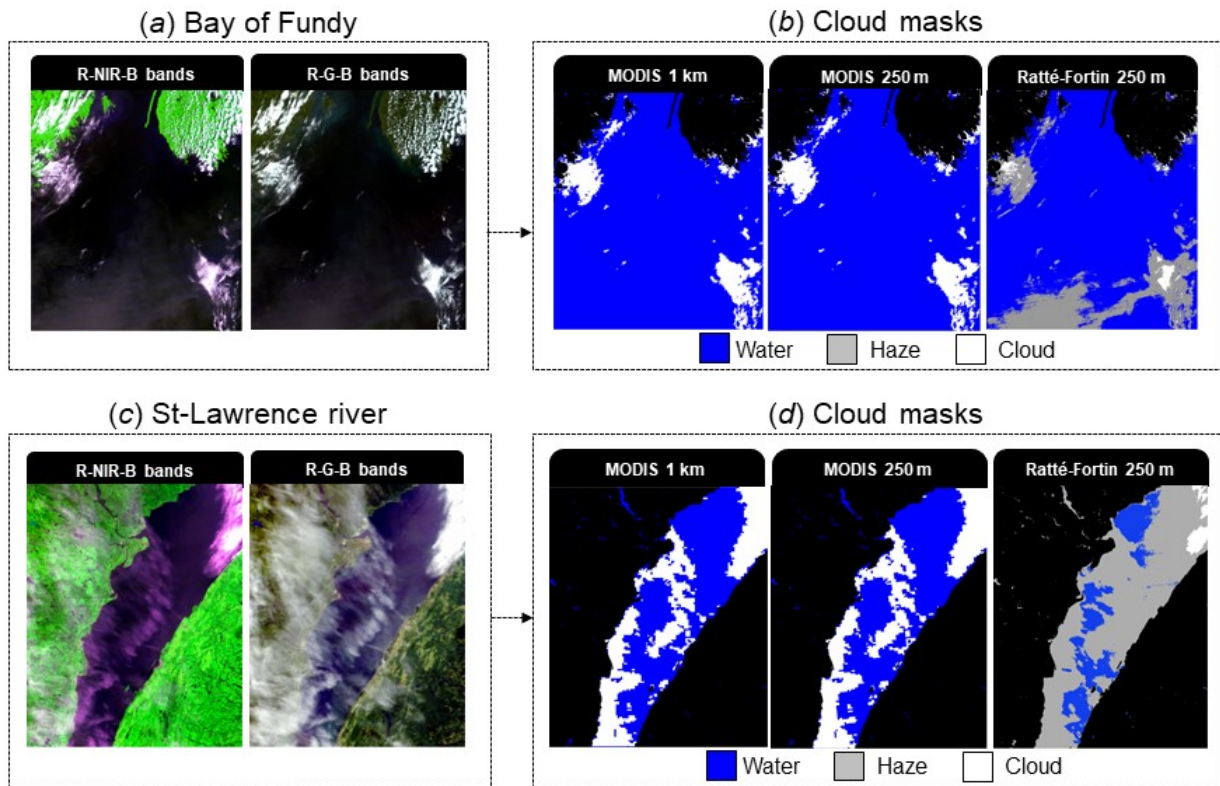


Figure 4.3 – Bandes MODIS R-PIR-B et R-V-B de la baie de Fundy (a) et du fleuve Saint-Laurent (c), ainsi que les trois masques de nuages (b) et (d) correspondant .

4.2 Procédure d'estimation phénologique via l'imagerie satellitaire MODIS

La procédure d'estimation phénologique à partir de l'imagerie satellitaire MODIS a été implantée dans le but de mettre en place un suivi journalier des épisodes sur l'ensemble des lacs du Québec et, par le fait même, permettre de générer une base de données phénologiques complètes dans le temps et dans l'espace. La mise en place de la base de données d'épisodes de FEA s'est construite en huit étapes, celles-ci pouvant être exécutées successivement par commandes MS-DOS:

1. Téléchargement des produits MODIS

De 2000 à 2016 de mai à novembre, quatre produits du satellite Terra MODIS ont été télé-

chargés: 1) MOD02HKM – Level 1B Calibrated Radiances - 500 m, 2) MOD02QKM – Level 1B Calibrated Radiances - 250 m, 3) MOD03 - Geolocation - 1 km et 4) MOD08 D3 - Level 3 Daily Joint Aerosol/Water Vapor/Cloud Product. Ces produits proviennent de la collection 5.1 – Selected Atmos and Land Products.

2. Mise à l'échelle

La mise à l'échelle s'est fait via l'exécution d'un script Matlab qui génère et initialise les fichiers et les dossiers nécessaires, puis exécute l'ensemble des commandes développées par le Centre canadien de cartographie et d'observation de la Terre (Trishchenko *et al.*, 2007). Cette mise à l'échelle inclue une désagrégation spatiale des bandes 3-7 de 500m à 250m.

3. Corrections atmosphériques et conversion des bandes 1-7 MODIS-D-250

Les corrections atmosphériques et le regroupement des images binaires en fichier d'extension .PIX en fonction de la date d'acquisition de l'image MODIS (figure 4.4, (a)) s'est fait ensuite via l'exécution d'un script EASI (C1_MOD_250_pix.EAS).

4. Masquage des pixels de terre et des pixels mixtes

L'application du masque de terre (figure 4.4, (b)) est appliqué sur les images .PIX (C2_CloudMaskLDA.EAS).

5. Détection du couvert nuageux

L'algorithme de détection du couvert nuageux (figure 4.4, (c)) est appliqué sur les images .PIX (C2_CloudMaskLDA.EAS).

6. Estimation des concentrations en chl-a

L'algorithme d'estimation de la concentration en chl-a figure 4.4, (d)) est appliqué sur les images .PIX (C3_MatConfAMOAM.EAS).

7. Extraction des données de chl-a

L'extraction des données de concentrations en chl-a a été effectuée sur l'ensemble des images MODIS avec l'exécution d'un premier script EASI (mcd1.EAS) nécessaire à l'ajustement du

nom des bandes 1-10, suivi d'un script Python (`batch_extract_pixel_value.xlsx.py`). Les valeurs de concentrations en chl-a à l'échelle du pixel sont enregistrées en format `.xlsx` et classifiées par lac et selon le bassin versant prioritaire dans lequel le lac se situe. Dans l'ensemble, une base de données des concentrations en chl-a est disponible sur un total de 580 lacs entre 2000 et 2016.

8. Estimation phénologique

La procédure de calcul des variables phénologiques a été effectuée avec le logiciel Matlab R2017a.

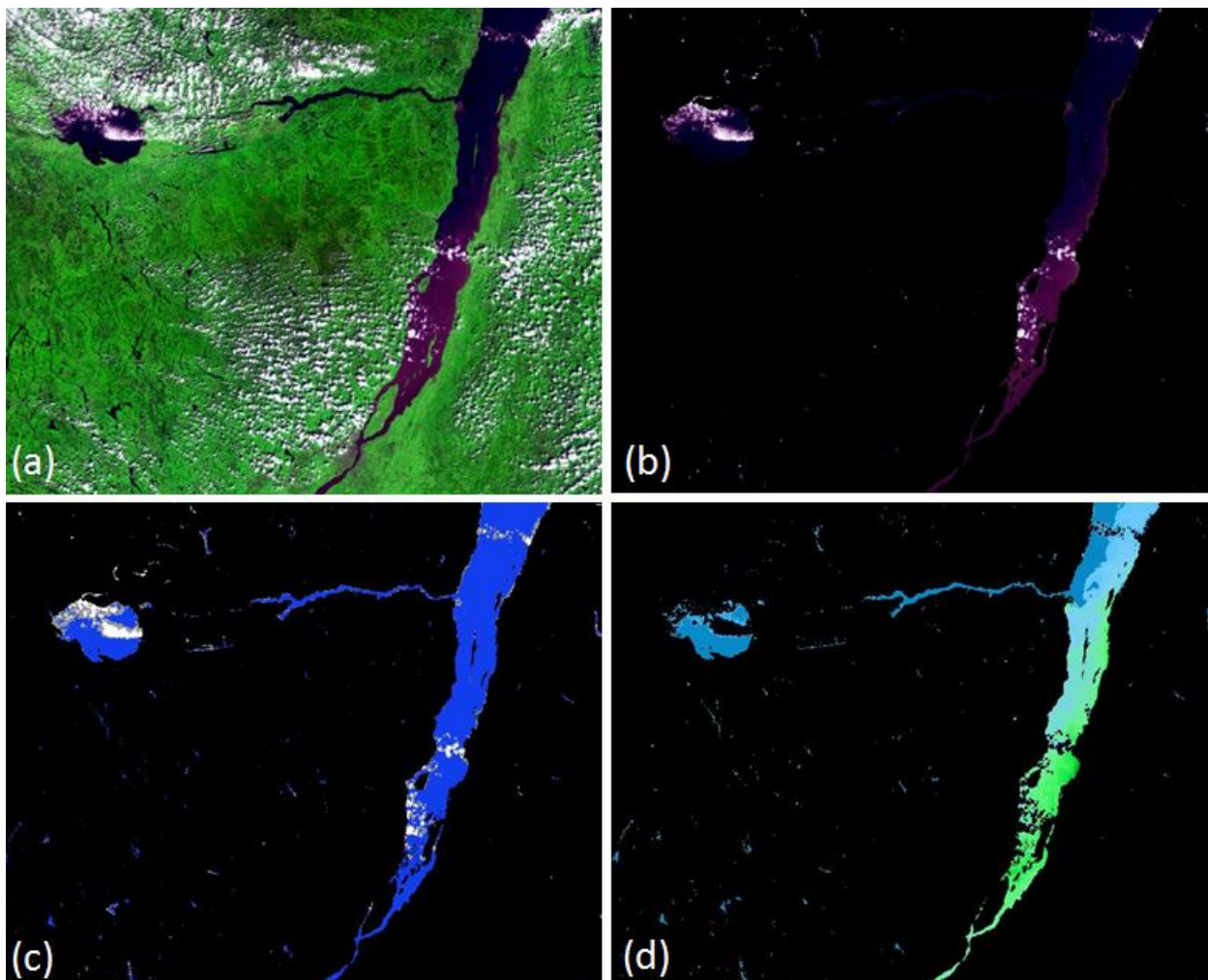


Figure 4.4 – (a) Exemple d'une scène du lac Saint-Jean et du fleuve Saint-Laurent présentée (a) avec les bandes MODIS R-PIR-B, (b) sans présence de pixels de terre ni de pixels mixtes, (c) sans présence du couvert nuageux, et (d) avec la couche des concentrations en chl-a estimées.

4.3 Analyse descriptive des événements algaux

La figure 4.5 présente l’histogramme de la distribution des variables phénologiques, et la Figure 4.6 présente l’histogramme cumulé complémentaire (similaire à la fonction de densité cumulée complémentaire (CCDF) où $\bar{F}(x) = P(X > x) = 1 - F(x)$) des variables phénologiques. La fréquence médiane d’apparition est de 15 jours par année (Figure 2.a). L’intensité (valeur maximale de concentration en chl-a) médiane est de $110 \mu g L^{-1}$ et l’étendue (valeur maximale de superficie par rapport à l’aire du lac) médiane est de 19 % (Figure 4.5.b, 4.5.c). L’allure des distributions est similaire d’une variable phénologique à l’autre, où l’on constate une asymétrie positive (Figure 4.5). L’allure de la distribution de la date de début est moins évidente que les autres variables phénologiques (Figure 4.5.d). La distribution débute le 15 mai puisqu’il s’agit de la date seuil utilisée pour considérer que les lacs localisés aux plus basses latitudes sont complètement dégelés. Bien que la portion extrême gauche de la distribution soit absente, on constate tout de même la même allure de courbe que les autres variables phénologiques, où la date médiane de la première efflorescence d’algues est le 4 juin. Notons que cette date serait théoriquement un peu plus hâtive avec une distribution complète de cette variable phénologique. Cinq lacs ont montré au moins 50 jours de FEA lors d’une année spécifique entre 2000 et 2016 (Figure 4.6.a): la baie Missisquoi (2012), le lac Saint-Louis (2005, 2012, 2013), le lac des Deux-Montagnes (2005, 2012), le lac Saint-Pierre (2013) et le lac Saint-François (2012). Treize lacs ont montré un épisode de FEA avec une étendue d’au moins 75% de la superficie totale observée entre 2000 et 2016; notons entre autre le lac Brome et la Baie Missisquoi.

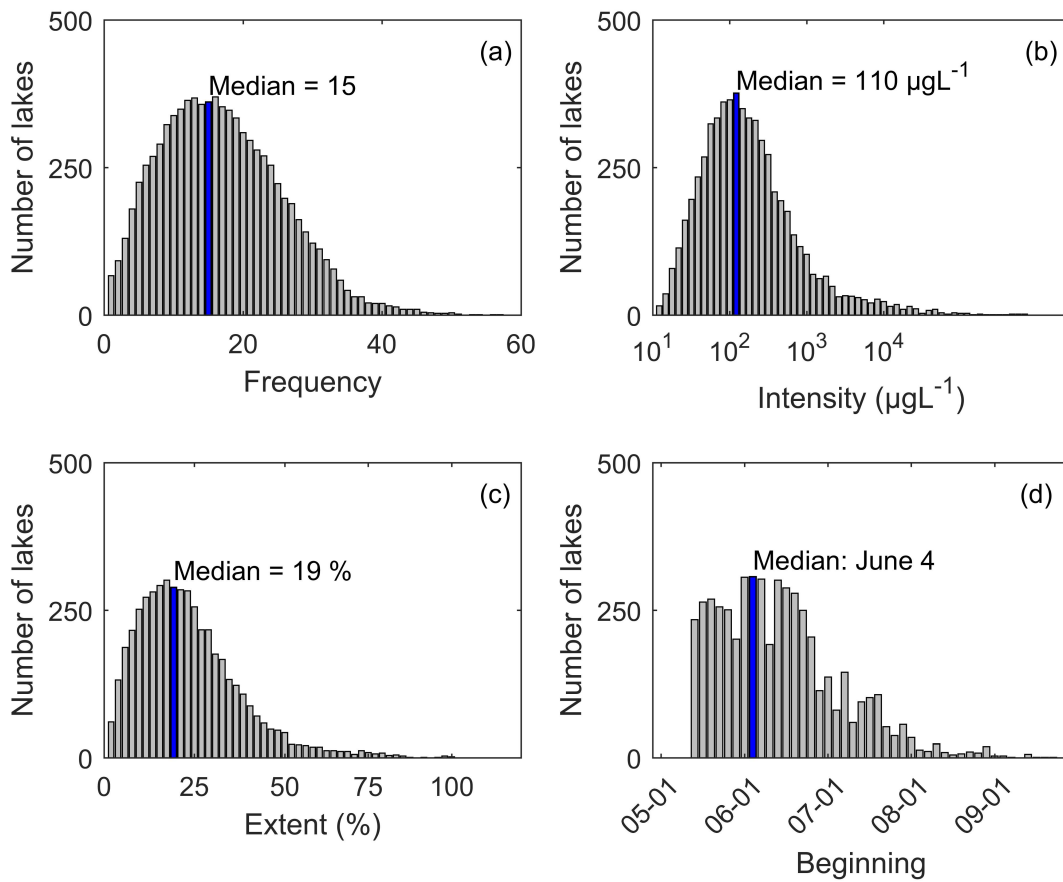


Figure 4.5 – Distribution des variables phénologiques de 2000 à 2016: (a) Fréquence d'apparition annuelle, (b) intensité maximale annuelle, (c) étendue maximale annuelle, et (d) Date d'initiation annuelle.

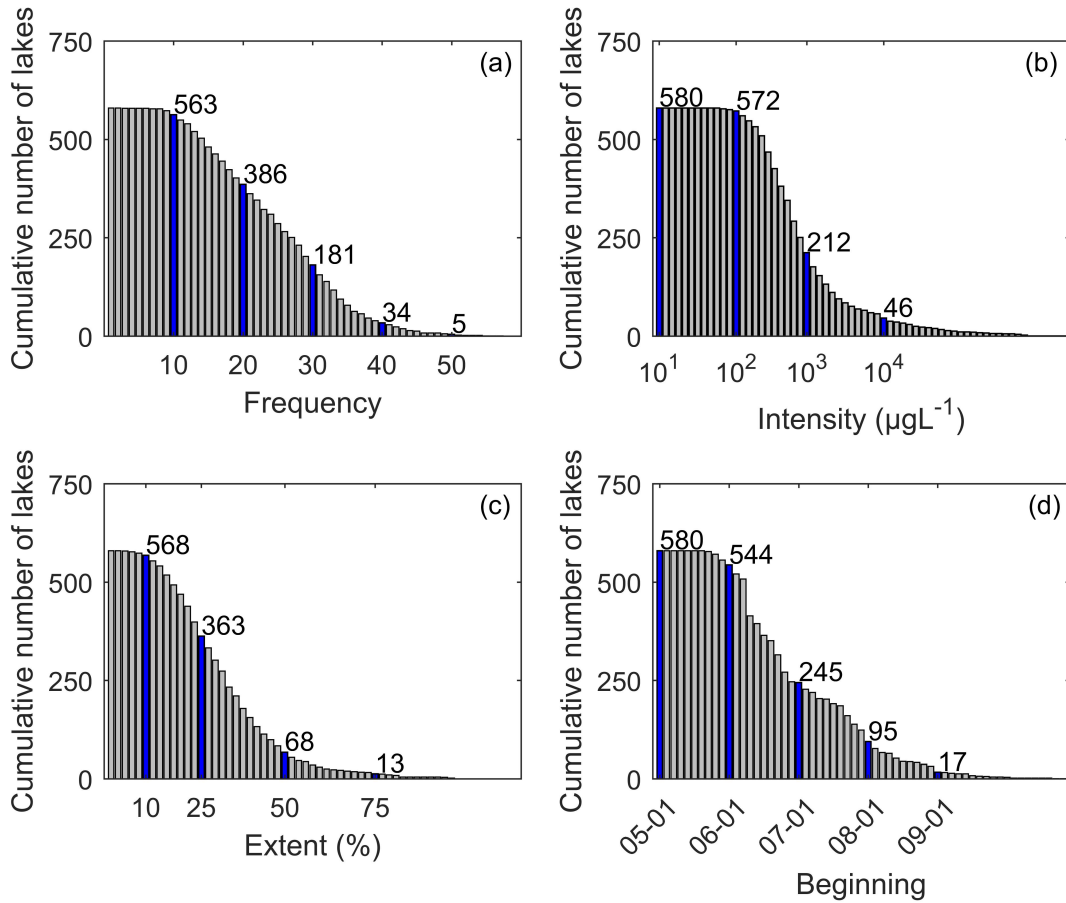


Figure 4.6 – Distribution cumulée complémentaire des variables phénologiques de 2000 à 2016: (a) Fréquence d'apparition annuelle, (b) intensité maximale annuelle, (c) étendue maximale annuelle, et (d) Date d'initiation annuelle.

4.4 Sommaire

Bien que le nouvel algorithme de masque de nuage présente la meilleure performance, le produit de masque de nuage MODIS à 250 m de résolution spatiale montre une performance dans l'ensemble assez bonne, avec un succès global de 95.3 % et un κ de 0.905. Cependant, la validation qualitative a permis de cibler les conditions où ce masque est plus restrictif. En effet, ce produit est limité sous des conditions d'eau à forte teneur en turbidité, et ces régions sont généralement localisées en bordure des lacs. Sa faible sensibilité à la brume limite son utilisation dans le cadre d'études sur la qualité de l'eau sous les régions spectrales exploitées dans le cadre de ce projet. Le nouvel algorithme de masque de nuage permet d'améliorer la précision dans la détection des nuages en milieux lacustre et fluvial. Sa résolution spatiale améliorée à 250 m améliore la précision et l'ampleur de données exploitable dans le cadre de suivi de la qualité de l'eau sur des plans d'eau de petite à moyenne taille. Les erreurs de commission potentielles sont plus faibles que les produits MODIS, moins sensibles à la brume. Cette réduction dans l'erreur de commission est essentielle dans le cadre de la détection et du suivi des FEA en raison de l'effet additif qu'apporte la brume à la signature spectrale de l'eau. Cet effet mène à des surestimations de la concentration en chl-a. L'utilisation d'une méthode multivariée comparativement aux méthodes couramment proposées dans littérature basées sur des algorithmes par seuillage permet de générer un modèle de classification beaucoup plus précis et optimal. Les résultats de la matrice de confusion mettent en évidence la très bonne concordance entre les classes observées et prédites en utilisant les bandes MODIS-D-250, avec un succès global moyen de 99.0 % variant entre 98.8 % et 99.2 % (I.C. à 95%) et un κ moyen de 0.982 variant entre 0.979 et 0.986 (I.C. à 95%). L'ajout de ce masque de nuage à l'ensemble de la procédure permet la mise en place d'un système semi-automatisé d'estimation de la concentration en chl-a en temps quasi-réel.

Les variables phénologiques (c'est-à-dire la fréquence, l'intensité, l'étendue et le début) générées à partir de la procédure semi-automatisée présentent des résultats réalistes. L'analyse descriptive des événements algaux qui se sont produits sur les 580 lacs entre 2000 et 2016 montre les variables phénologiques présente des distributions de type asymétrique vers la gauche. Entre 2000 et 2016, le nombre médian de jours par année où le seuil de qualité de l'eau de niveau 1 fixé à $10 \mu g L^{-1}$ a été dépassé est de 15 jours. Notons qu'il s'agit des jours où la couverture nuageuse pour un lac donné est inférieure à 25 % de la superficie observable, seuil tel qu'établi dans les travaux de Duan

et al. (2014). Également, la valeur médiane de la concentration maximale annuelle en chl-a est de $110 \mu gL^{-1}$ lorsqu'il y a présence de FEA, et l'étendue médiane s'élève à 19 %. Bien sûr, les données manquantes ont un impact sur l'exactitude des variables phénologiques, particulièrement sur la fréquence et la date d'initiation annuelle. Cole *et al.* (2012) a trouvé des erreurs sur la date d'initiation de FEA en milieu océanique de l'ordre de 2 à 3 jours avec 10 % de données manquantes et de 15 à 30 jours avec 80 % de données manquantes via l'utilisation d'images du capteur SeaWiFS en milieu océanique. La probabilité d'erreur sur la date d'initiation (date plus tardive qu'en réalité) est cependant de 50 % près de la région d'étude (Figure 6; Cole *et al.* (2012)), signifiant qu'il y a autant de chances que la date d'initiation soit en réalité plus hâtive que plus tardive. Le pourcentage de données manquantes relié à la couverture nuageuse est discuté plus en profondeur à la section 7.5.3. Par ailleurs, l'analyse descriptive des événements algaux à l'échelle du lac a été utilisée afin de comparer les résultats obtenus via le modèle d'estimation de la concentration en chl-a et les études disponibles dans la littérature traitant de suivis quantitatifs (concentrations médianes ou moyennes de chl-a) de la qualité de l'eau en milieux lacustres. Parmi les deux lacs (Deux-Montagnes et Saint-Louis) baignant l'île de Montréal, le lac des Deux-Montagnes présente une valeur médiane de concentration en chl-a de $4,8 \mu gL^{-1}$ en 2000, ce qui rejoint les résultats obtenus des médianes estivales de 1997-1998 qui se situent entre 2 et $5 \mu gL^{-1}$ à proximité de ce lac (MELCC, 2000).

Troisième partie

**PORTRAIT DE LA VARIABILITÉ
SPATIO-TEMPORELLE DES FEA
SUR LE QUÉBEC MÉRIDIONAL**

Chapitre 5

Introduction aux sources de perturbation liées au développement les FEA

Les FEA peuvent être composées de cyanobactéries surtout lorsque la biomasse est élevée. Ce groupe fonctionnel a le potentiel de produire des toxines, celles-ci causant potentiellement une dégradation des services écologiques que nous procurent les plans d'eau. Les cyanobactéries se sont dotées d'une multitude de stratégies adaptatives leur permettant de coloniser et dominer plusieurs types d'environnements tout en possédant des avantages compétitifs face à d'autres micro-organismes. Parmi ces stratégies, notons leur capacité à migrer dans la colonne d'eau, leur permettant ainsi d'exploiter les eaux profondes et riches en nutriments et le rayonnement solaire en surface. Leur pigmentation diversifiée (incluant les phycobiliprotéines qui leur procure une signature optique unique) leur permet de réaliser la photosynthèse à de faibles intensités lumineuses. De plus, elles peuvent faire des réserves en nutriments et entrer en dormance lorsque le milieu n'est pas favorable à leur prolifération. Certaines espèces ont également la capacité d'utiliser l'azote atmosphérique (via la fixation du diazote (Gallon, 1992)) si les formes d'azote inorganique deviennent limitées.

Ces avantages compétitifs ne permettent toutefois pas d'expliquer à eux seuls l'abondance et la persistance des FEA dans certains plans d'eau. En effet, l'ampleur du phénomène reflète également l'état avancé d'eutrophisation de plusieurs systèmes aquatiques, causé par le développement urbain,

agricole et industriel. Ces développements génèrent un transfert plus important de nutriments vers les milieux aquatiques et un déséquilibre dans leur biodiversité. Des facteurs endogènes tels que les caractéristiques physico-chimiques de l'eau, de même que des facteurs exogènes tels que les caractéristiques structurelles du bassin versant et les caractéristiques climatiques ambiantes peuvent jouer un rôle dans le développement et la prolifération des FEA. Une revue de ces facteurs est présentée dans ce chapitre, suivie d'une revue des modèles développés afin d'expliquer la présence de FEA. Une discussion conclue ce chapitre.

5.1 Prédiposition d'un lac aux FEA

Facteurs physico-chimiques:

De manière générale, plus les concentrations en nutriments augmentent, plus la biomasse phytoplanctonique présente fréquemment des intensités élevées et de manière soutenue durant la saison estivale (Reynolds, 1984). Parmi les facteurs physico-chimiques favorisant la prolifération d'algues, le phosphore est considéré comme étant le premier prédicteur de FEA et comme étant le principal facteur expliquant la variabilité des FEA (Guildford *et al.*, 2005; Davis *et al.*, 2009; Obenour *et al.*, 2014; Steffen *et al.*, 2014), mais l'enrichissement en azote est également un facteur clé à la prolifération de FEA notamment lorsque le phosphore n'est plus limitant (Davis *et al.*, 2009; Chaffin *et al.*, 2013). Plusieurs travaux se sont aussi intéressés à la relation existant entre les FEA et le ratio de NT/PT dans l'eau (Havens *et al.*, 2003; Lilover & Stips, 2008) mais la véritable influence de ce paramètre sur la prolifération de FEA demeure mitigée puisque certaines études appuient cette relation (Ferber *et al.*, 2004) et d'autres ne l'appuient pas Jayatissa *et al.* (2006). Des travaux de Duan *et al.* (2014) ont montré que le ratio TN/TP était relié au moment d'initiation de FEA sur le lac Taihu en Chine. Pour leur part, des travaux de Bergström (2010) et de Ptacnik *et al.* (2010) ont montré qu'une approche intégrant un ratio de l'azote inorganique dissous (DIN) par rapport au PT (lorsque le ratio DIN/PT < 2) s'avère être un meilleur prédicteur de la biomasse cyanobactérienne.

Bien que son effet sur la croissance soit spécifique aux différentes espèces d'algues (Mehnert *et al.*, 2010), la température de l'eau est aussi considérée comme un facteur essentiel à la prolifération des FEA (Davis *et al.*, 2009; Winder *et al.*, 2012). Le taux de croissance de la plupart des cyanobactéries est maximal à 25°C selon Seaburg *et al.* (1981), à des températures supérieures à 25°C selon Robarts

& Zohary (1987) ou à des températures entre 25°C et 35°C d'après les travaux de Reynolds & Walsby (1975). De manière générale, la température optimale de croissance des cyanobactéries est plus élevée que celle des diatomées et des algues vertes (Lavoie *et al.*, 2007b), et leur croissance est limitée en deçà de 5°C et au delà de 35°C (Tang *et al.*, 1997).

D'autres facteurs peuvent aussi influencer la dynamique des FEA et sont utilisés comme indicateurs de l'état trophique d'un lac tels que la disponibilité en lumière (Winder *et al.*, 2012), sans nécessairement prédisposer le lac à des FEA. Les indices de qualité de l'eau généralement utilisés pour évaluer l'état d'eutrophisation d'un lac ou d'un réservoir sont la température de l'eau, l'oxygène dissous, l'azote total, le phosphore total et la chl-a (Takamura *et al.*, 2003; Wang *et al.*, 2007).

Facteurs structurels:

Plusieurs facteurs structurels du bassin versant influencent directement et indirectement la quantité et la qualité de l'eau. Au niveau de la qualité de l'eau, la quantité d'éléments nutritifs d'origine anthropique pouvant s'écouler dans les lacs et les rivières est liée à la physiographie et à l'occupation du sol du bassin versant (Lavoie *et al.*, 2007a; Langlois, 2008). De plus, puisque l'apport d'eau est directement relié à l'aire de réception des précipitations sur le bassin versant, la quantité et la qualité de l'eau acheminée vers le lac est en partie reliée à la superficie du bassin. La capacité de stockage des polluants augmente avec la superficie, mais pour une même période de temps, l'apport sera plus estompé sur de grands bassins. La forme du bassin versant aura également une influence directe sur l'allure de l'hydrogramme à l'exutoire du bassin, et ainsi sur les débits de pointe qui influencent les risques d'érosion des sols. Par ailleurs, le relief peut être caractérisé d'abord par la courbe hypsométrique, qui représente l'altitude en fonction de la surface cumulée du bassin, ou par l'altitude minimale, maximale, moyenne et médiane qui mettent en évidence l'influence du relief sur l'écoulement. La pente nous renseigne sur la topographie et nous informe sur la capacité de transport des nutriments et des sédiments sur le bassin versant. La capacité de ruissellement et le taux d'infiltration sont influencés par le type de sol sur le bassin versant. La couverture végétale influence directement l'écoulement de surface et limite les flux de contaminants vers les cours d'eau. D'autre part, les sols nus montrent une faible capacité de rétention et favorisent un ruissellement de surface plus rapide. De par leur imperméabilité, ces caractéristiques sont d'autant plus importantes pour les surfaces urbanisées. De plus, les sources de rejets diffus de nutriments qui se déversent ultimement dans les cours d'eau sont directement liées aux superficies agricoles et aux types de

cultures. D'autres apports en contaminants d'origine anthropique telles que les sources de pollution ponctuelles (installation septique, rejet de produits domestiques) sont associés à la démographie sur le bassin versant. La densité de drainage (longueur totale du réseau hydrologique par rapport à la superficie totale du bassin) et la densité hydrographique (nombre de canaux d'écoulement par rapport à la superficie totale du bassin) sont reliées au relief et à la perméabilité du sol qui dépendent en partie de la géologie du milieu. Le rapport de confluence (nombre de canaux d'ordre n par rapport au nombre de canaux d'ordre $(n + 1)$) est un indicateur de l'influence de la géologie sur le bassin versant. Certaines caractéristiques du plan d'eau peuvent avoir une influence notable sur l'intensité, le début, la durée et l'occurrence de FEA, incluant le rapport de superficie entre le plan d'eau et le bassin versant (plus haut) et la profondeur du plan d'eau. En effet, certaines études ont montré une relation inversement proportionnelle entre la profondeur de la colonne d'eau et la prolifération de cyanobactéries (Berlamont *et al.*, 2013).

Facteurs climatiques:

Les variables climatiques jouent un rôle indéniable dans le développement de FEA (Elliott, 2012; Reichwaldt & Ghadouani, 2012; Bartosiewicz *et al.*, 2019a). Plusieurs travaux se sont intéressés à la relation existant entre les FEA et les variables météorologiques et/ou climatiques (Lilover & Stips, 2008; Liu *et al.*, 2013; Ndong *et al.*, 2014). La température de l'air (Hu *et al.*, 2009; Elliott, 2012; Reichwaldt & Ghadouani, 2012; Zhang *et al.*, 2012), les précipitations (Hu *et al.*, 2009; Reichwaldt & Ghadouani, 2012), le vent (Zhang *et al.*, 2012; Liu *et al.*, 2013) ainsi que les heures d'ensoleillement (Zhang *et al.*, 2012) sont des variables ayant montré une influence notable sur la prolifération de FEA. Notons par ailleurs que la probabilité d'apparition d'une FEA diminue avec de forts vents (> 5 m/s) (Zhang, 2007) et avec toute autre processus naturel provoquant une turbulence de l'eau de surface (Ahn *et al.*, 2002). Des travaux de Hu *et al.* (2009) sur le lac Monduran (Australie) ont montré que la température mensuelle minimale et les précipitations totales mensuelles avaient un effet significatif sur l'abondance de cyanobactéries. Zhang *et al.* (2012) ont étudié l'effet des variables climatiques sur la prolifération des FEA au lac Taihu en Chine depuis les 23 dernières années. Leurs résultats ont montré que l'augmentation de la température, des heures d'ensoleillement et de la radiation globale de même que la diminution de la vitesse des vents ont comme effet un développement plus hâtif et une persistance plus soutenue des FEA. La vitesse du vent et les heures d'ensoleillement ont permis d'expliquer près de 85 % de la variabilité du modèle de régression multiple. Des travaux de Binding *et al.* (2011) ont montré que la précipitation totale entre avril

et juillet et la température cumulée entre janvier et août sur le lac des Bois (Canada) expliquait 77 % de la variabilité des pics annuels d'intensité de FEA. Plusieurs études se sont également intéressées à l'impact de l'oscillation nord-atlantique (ONA) sur la variabilité et la prolifération de FEA. Blenckner *et al.* (2007) ont montré qu'il existait des réponses directes et indirectes sur les écosystèmes lacustres, incluant une augmentation de la biomasse cyanobactérienne lors d'indices annuels positifs de l'ONA. Les résultats de Janssen *et al.* (2004) suggèrent également un impact de l'ONA sur l'occurrence des FEA dans la mer Baltique.

La prolifération de FEA est sujette à augmenter en fréquence et en intensité en réponse aux changements climatiques (Mooij *et al.*, 2007; Wagner & Adrian, 2009; Moss, 2012; Zhang *et al.*, 2012). Par changements climatiques, on entend par exemple la hausse de la température de l'air et par conséquent de la température à la surface de l'eau, la réduction de la période avec un couvert de glace (étés plus longs), la hausse en intensité et en fréquence des tempêtes, les changements dans le régime hydrologique et la persistance accrue des périodes de sécheresse. D'autres études montrent que la phase d'initiation des fleurs d'eau semble être plus hâtive au printemps (Edwards & Richardson, 2004; Winder & Sommer, 2012). Les organismes phytoplanctoniques peuvent être affectés par les changements climatiques directement à travers leur physiologie, et indirectement par des changements dans la stratification thermique et la disponibilité en nutriments et en lumière. En effet, l'augmentation de la température des eaux de surface intensifie la stratification verticale des écosystèmes d'eau douce, prenant place plus tôt au printemps, se maintenant de manière plus stable tout au long de l'été et prenant fin plus tard à l'automne (Elliott, 2010; Winder & Sommer, 2012). Sachant que la température joue un rôle majeur dans les réponses phénologiques des FEA (Elliott *et al.*, 2006; Thackeray *et al.*, 2008), son augmentation graduelle affectera les mécanismes d'initiation, de développement, de maintien et de déclin des FEA (Winder *et al.*, 2012; Winder & Sommer, 2012), de même que l'intensité moyenne de la biomasse présente lors d'une floraison (Winder & Schindler, 2004). Le réchauffement planétaire touche particulièrement la croissance des cyanobactéries puisqu'en tant que procaryotes, leur croissance est optimale à des températures relativement plus élevées que les autres groupes phytoplanctoniques (Watkinson *et al.*, 2005), leur procurant un avantage compétitif face aux producteurs primaires eucaryotes (diatomés, chlorophytes, dinoflagellés, *etc*) (Paerl *et al.*, 2011).

5.2 Sommaire

Plusieurs travaux ont permis de cibler les éléments perturbateurs et caractériser leurs effets sur le développement des FEA. Les principaux facteurs sont bien entendu le P (Guildford *et al.*, 2005; Davis *et al.*, 2009; Steffen *et al.*, 2014) et le NT lorsque le PT n'est plus limitant (Davis *et al.*, 2009; Chaffin *et al.*, 2013). Le ratio N : P est également utilisé comme indicateur à l'apparition de biomasse cyanobactérienne (Havens *et al.*, 2003; Lilover & Stips, 2008; Duan *et al.*, 2014) ou de biomasse phytoplanctonique (Bergström, 2010; Ptacnik *et al.*, 2010). Bien que spécifique à chaque espèce d'algue (Mehner *et al.*, 2010), la température de l'eau et la disponibilité en lumière jouent également un rôle prépondérant sur le développement des FEA cyanobactériennes (Davis *et al.*, 2009) et phytoplanctoniques (Winder & Sommer, 2012). Il demeure cependant difficile de circonscrire tous les éléments perturbateurs à l'échelle du plan d'eau de manière efficace et économique. Dans l'optique de présenter des solutions d'intervention en matière de protection ou de restauration d'un plan d'eau, il est nécessaire de mettre en lumière les sources de perturbation et leur lien avec la fréquence, l'intensité et la durée des épisodes de FEA. L'augmentation en fréquence et en intensité des FEA dans les écosystèmes d'eau douce est en partie causée par les multiples perturbations anthropiques sur ces milieux. L'apport en éléments nutritifs d'origine diffuse ou ponctuelle sur un plan d'eau est lié à la physiographie de son bassin versant. Par exemple, la capacité de stockage d'éléments nutritifs qui augmente avec la superficie du bassin, le type de sol qui influence la capacité de ruissellement et le taux d'infiltration, la forme et le relief du bassin qui ont une influence directe sur les débits de pointe et donc sur l'érosion des sols. Par ailleurs, plusieurs variables climatiques jouent un rôle indéniable dans le développement des FEA, dont la température de l'air (Hu *et al.*, 2009; Paerl & Huisman, 2009; Reichwaldt & Ghadouani, 2012; Zhang *et al.*, 2012), les précipitations (Hu *et al.*, 2009; Reichwaldt & Ghadouani, 2012), le vent (Zhang *et al.*, 2012; Liu *et al.*, 2013) et les heures d'ensoleillement (Zhang *et al.*, 2012). En réponse aux changements climatiques, les FEA sont sujettes à augmenter en intensité et en fréquence (Wagner & Adrian, 2009; Moss, 2012; Zhang *et al.*, 2012) et à montrer une phase d'initiation plus hâtive au printemps (Edwards & Richardson, 2004; Winder *et al.*, 2012).

Chapitre 6

Étapes méthodologiques

6.1 Sites d'étude et base de données géospatiales

Les lacs à l'étude sont situés dans la province du Québec, Canada (44° - 50° Nord, 67° - 80° Ouest) et se regroupent plus précisément dans les 40 zones de gestion intégrée des ressources en eau (Figure 6.1), ces zones découlant du plan d'intervention sur les algues bleu-vert mis en place par le gouvernement du Québec (MELCC, 2014c). L'ensemble du territoire est constitué majoritairement de sols podzoliques (of Canada Working Group *et al.*, 2010) et est caractérisé par un climat continental humide au sud et un climat subarctique au nord (Peel *et al.*, 2007). Le développement industriel et agricole est majoritairement répandu au sud-ouest du territoire, alors que le nord est très peu développé.

Les concentrations en chl-a à l'échelle du pixel d'une résolution spatiale de 250 m ont été estimées par le biais des bandes MODIS-D-250. Rappelons que les images font fies des pixels mixtes (eau-terre) et de la couverture nuageuse par l'application d'un masque de terre (El-Alem *et al.*, 2014b) et d'un masque de nuage (Ratté-Fortin *et al.*, 2018) développé pour les eaux intérieures (lacs, rivières, estuaires). L'incertitude du signal spectral en bordure des plans d'eau causé par des éléments hétérogènes (fond du lac, quais, embarcations de plaisance) est réduite par la suppression des bordures des plan d'eau d'une largeur maximale de 250 m. Afin de disposer de lacs observables par le capteur MODIS et d'obtenir un ensemble de pixels qui soit représentatif de la réalité du plan d'eau, les lacs d'une superficie minimale de 3.5 km² ont été gardés pour l'étude (se référer à

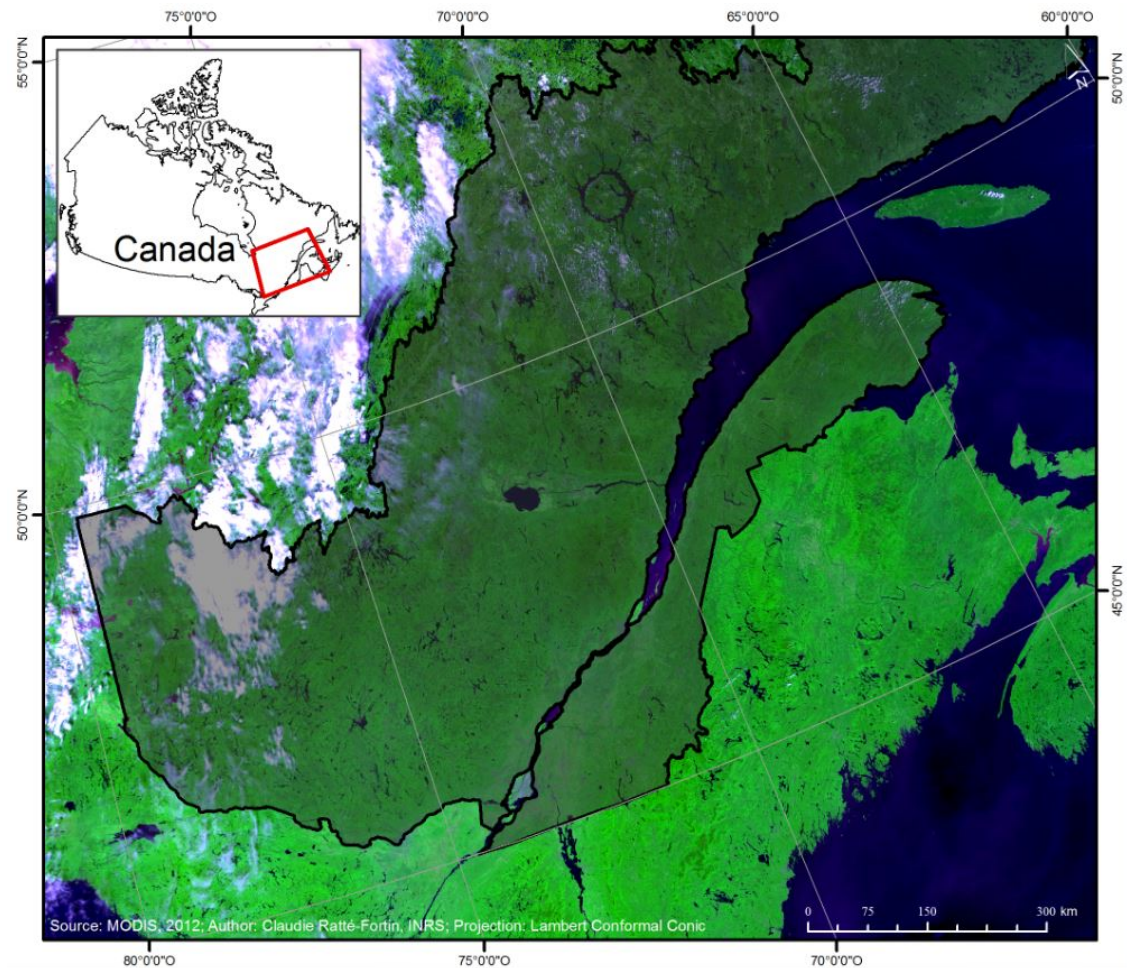


Figure 6.1 – Région d'étude et limites géographiques des zones de gestion intégré de l'eau (gris foncé) couvrant 580 lacs observables par le capteur MODIS.

la section 3.5.1). Au total, les concentrations en chl-a de 580 lacs ont été extraites sur la période de mai à novembre de 2000 à 2016. À partir des concentrations disponibles à l'échelle du pixel pour l'ensemble des lacs, les six variables phénologiques sélectionnées pour l'étude ont été générées sur l'ensemble des lacs entre 2000 et 2016: 1) la fréquence, 2) l'intensité, 3) l'étendue, 4) la date d'initiation, 5) la date de déclin et 6) la durée (se référer à la section 3.5.2 pour plus d'information).

6.2 Portrait du milieu physique

Une base de données géo-référencées des caractéristiques morphologiques, physiographiques et climatiques des bassins versants de chacun des lacs à l'étude a été mise en place. Le choix dans les différents types de données climatiques et physiographiques sélectionnés est basé sur une revue de

littérature des données accessibles dans le cadre de cette thèse et est présentée à l'annexe C. Les limites des bassins de drainage de chaque lac et, par extension, les descripteurs morphologiques des bassins versants (par ex. superficie, pente), ont été calculées à partir du modèle numérique d'altitude (MNA) du Canada d'une résolution spatiale de ~ 30 m (Gouvernement du Canada, 2017). Le MNA est disponible sur l'ensemble du territoire à l'étude c'est-à-dire le Québec méridional (figure 6.3, (a)). Le MNA (figure 6.2, (a)) inclut les dépressions naturelles du sol qui se retrouvent habituellement à quelques mètres d'abaissement de la surface environnante. Certains milieux lacustres en sont des exemples, où l'eau qui y pénètre s'écoule soit par voies souterraines, s'évapore ou s'accumule dans la zone jusqu'à écoulement de surface. Afin de calculer la direction d'écoulement via les outils d'analyse spatiale dans ArcToolbox, ces dépressions sont retirées en générant un MNA sans dépressions (figure 6.2, (b)). Notons que les dépressions naturelles doivent être maintenues dans le cas de bassins endoréiques, ce que l'on ne retrouve pas au Québec (Ricketts *et al.*, 1999). La direction d'écoulement au niveau du pixel peut ensuite être calculée (figure 6.2, (c)). Cette étape permet ultimement de délimiter les bassins de drainage selon un point donné à l'exutoire. Dans notre cas, le point à l'exutoire correspond au format raster des lacs à l'étude (figure 6.2, (d)).

Les données climatiques sont issues de réanalyses régionales de l'Amérique du Nord (Mesinger *et al.*, 2006) et possèdent une résolution spatiale de ~ 32 km. Les produits téléchargés doivent dans un premier temps être convertis en multi-bandes d'extension .tif avant d'effectuer une analyse statistique focale qui permet de extraire les valeurs climatiques moyennes ou totales sur chaque bassin versant (ou lac). Les données de réanalyse sont disponibles sur l'ensemble du territoire à l'étude (figure 6.3, (b)). Les données d'occupation du sol (figure 6.3, (c)) et d'économie agricole et de population (figure 6.3, (d)) sont fournies par le Government of Canada (2015, 2016) via leur portail de données ouvertes. Dans l'ensemble, vingt-sept variables environnementales (descripteurs physiographiques, climatiques et morphologiques) ont été extraites pour chacun des lacs et leur bassin versant (tableau 6.1).

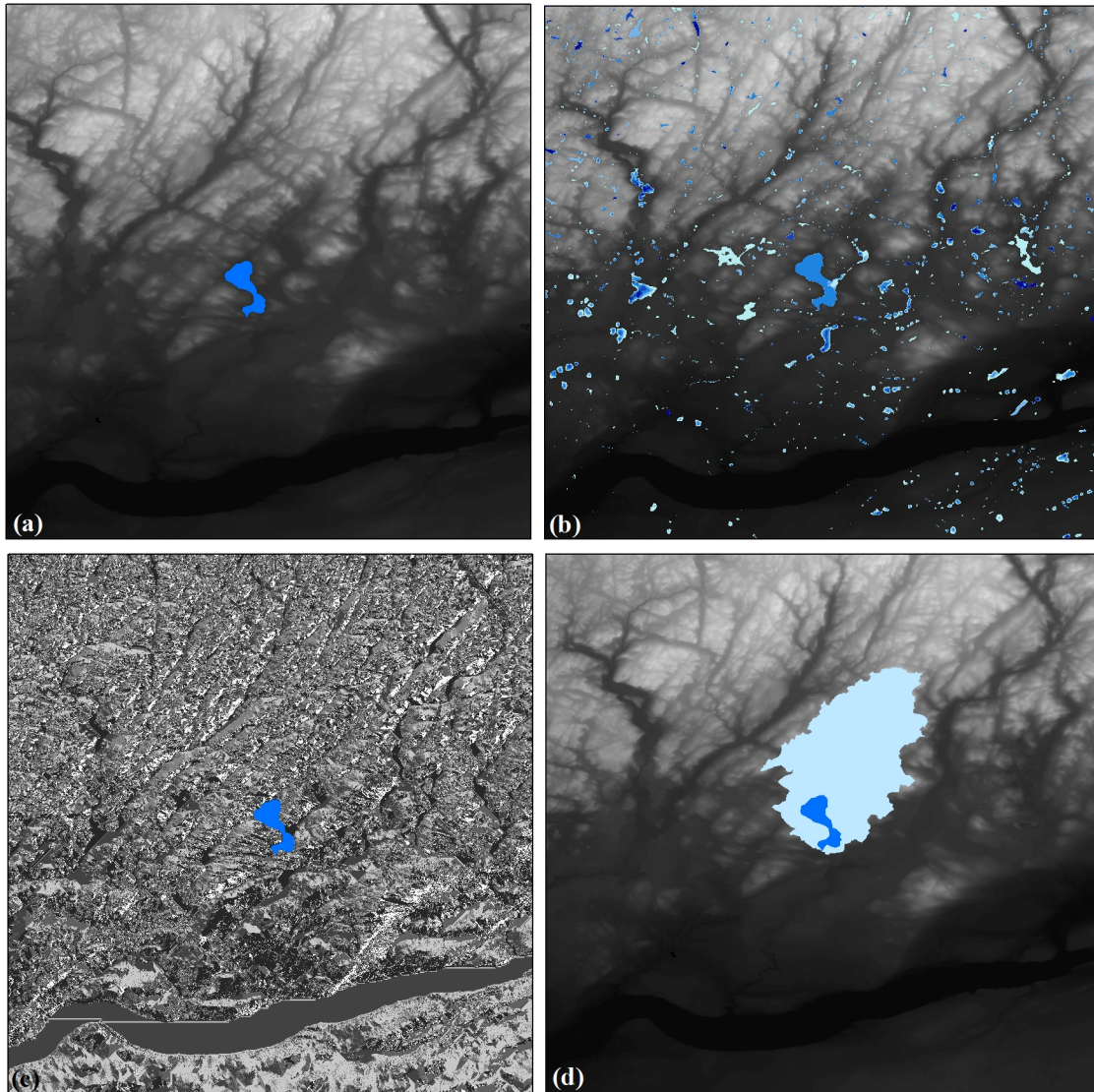


Figure 6.2 – Exemple de calcul de l'aire de drainage: (a) MNA et lac Saint-Joseph (bleu), (b) zones de dépressions naturelles du sol, (c) raster de la direction du flux d'écoulement et lac Saint-Joseph (bleu), (d) bassin de drainage (bleu pâle) du lac Saint-Joseph (bleu).

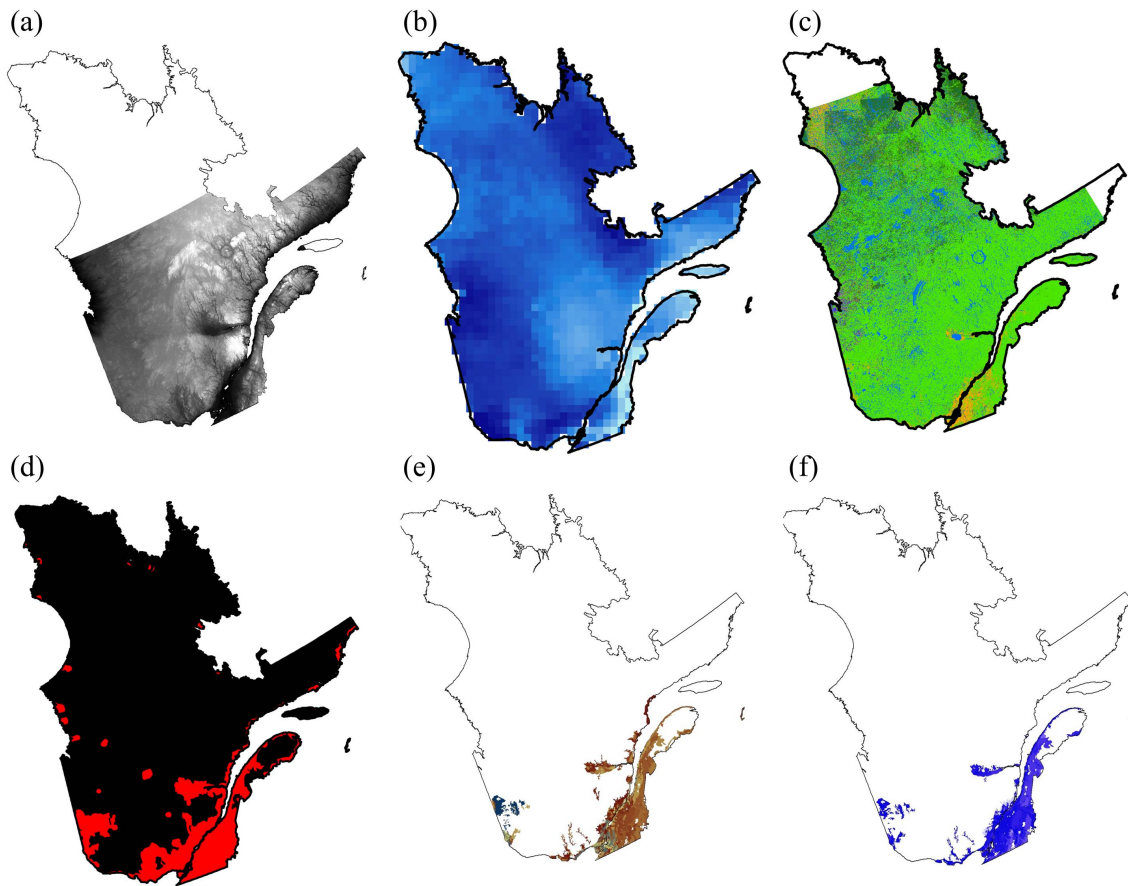


Figure 6.3 – Couverture spatiale des données physiographiques et climatiques: (a) MNA et données sur le relief, (b) précipitations, températures, degrés-jours, vents (c) occupation du sol, (d) écoumène agricole et écoumène de population, (e) données d'érodabilité des sols, et (f) données d'érosivité des sols.

Tableau 6.1 – Liste des descripteurs environnementaux. Les variables annotées en gras correspondent aux descripteurs utilisés pour l’analyse canonique des corrélations.

Descripteurs environnementaux	Caractéristiques
1 – Superficie du lac	
2 – Périmètre du lac	
3 – Indice de forme du lac	Coefficient de Gravelius
4 – Longueur du lac	Rectangle de Gravelius
5 – Largeur du lac	Rectangle de Gravelius
6 – Superficie du bassin versant	
7 – Périmètre du bassin versant	
8 – Indice de forme du lac	Coefficient de Gravelius
9 – Longueur du lac	Rectangle de Gravelius
10 – Largeur du lac	Rectangle de Gravelius
11 – Pente sur le bassin versant	Moyenne
12 – Pente sur le bassin versant	Écart-type
13 – Occupation du sol	Forêt
14 – Occupation du sol	Zones développées et routes
15 – Occupation du sol	Agriculture, terres cultivées
16 – Occupation du sol (rel)	Forêt
17 – Occupation du sol (rel)	Zones développées et routes
18 – Occupation du sol (rel)	Agriculture, terres cultivées
19 – Écoumène de population	Aires de diffusion habitée
20 – Écoumène agricole	Aires de diffusion où règne des activités agricoles
21 – Précipitations	Totales annuelles
22 – Précipitations	Totales (mai-nov)
23 – Température	Moyenne annuelle
24 – Température	Moyenne (mai-nov)
25 – Degrés-jour	FEA; seuil de croissance de 20°C (Ralston <i>et al.</i> , 2014; Pick, 2016; Larson <i>et al.</i> , 2018)
26 – Vents	Moyenne annuelle
27 – Vents	Moyenne (mai-nov)

6.3 Analyse canonique des corrélations

Dans un premier temps, la distribution spatiotemporelle des événements de FEA sur le Québec méridional est étudiée. Si une variabilité spatiale et/ou temporelle est notable pour une variable phénologique donnée, celle-ci est sous-tendante aux conditions climatiques prédominantes, à la physiographie du bassin versant et à la morphologie du plan d'eau. Pour les variables phénologiques présentant une variabilité spatiotemporelle sur le territoire, une analyse statistique multivariée entre celles-ci et les variables environnementales est effectuée. Une analyse des corrélations de Pearson intra et inter variables phénologiques et variables environnementales est d'abord calculée. Cette analyse permet de réduire le nombre de variables environnementales afin de réduire la redondance d'information tout en sélectionnant les variables les plus corrélées à la phénologie des FEA (pour plus d'information, voir la section 6.4). Notons que des régressions linéaires multiples entre les variables phénologiques et les variables environnementales, de même qu'une analyse des composantes principales (ACP) appliquée sur les variables environnementales ont également permis de résumer la variabilité des descripteurs environnementaux et de réduire le nombre de variables environnementales utilisées pour l'analyse multivariée.

À partir des variables environnementales sélectionnées et des variables phénologiques, une analyse canonique des corrélations (ACC) est calculée afin d'analyser simultanément les deux groupes de variables et en identifiant et quantifiant leurs associations. Les objets (lacs) à l'étude sont décrits par deux ensembles de descripteurs quantitatifs : le premier ensemble X_1 de p descripteurs phénologiques, et le deuxième ensemble X_2 de q descripteurs environnementaux. La matrice de dispersion S de ces $p + q$ descripteurs est construite à partir de 4 sous-matrices :

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \quad (6.1)$$

Les sous-matrices S_{11} (ordre $p \times p$) et S_{22} (ordre $q \times q$) se réfèrent aux matrices de covariance du premier et du deuxième ensemble de descripteurs respectivement, alors que S_{12} (ordre $p \times q$) et sa transposée $S'_{12} = S_{21}$ (ordre $q \times p$) se correspondent à la matrice de covariance entre les deux ensembles de descripteurs. Des combinaisons linéaires de chaque ensemble de descripteurs

sont calculées :

$$\begin{aligned} U_1 &= a' X_1 \\ V_1 &= b' X_2 \end{aligned} \quad (6.2)$$

Où les vecteurs de coefficients a et b sont générés de telle sorte que la corrélation canonique

$$R_1^* = \text{Corr}(U_1, V_1) = \frac{a' S_{12} b}{\sqrt{a' S_{11} a} \sqrt{b' S_{22} b}} \quad (6.3)$$

est la plus élevée possible. La première paire de variables canoniques est la paire de combinaisons linéaires U_1 et V_1 qui maximise la corrélation (6.3). La deuxième paire de variables canoniques est la paire de combinaisons linéaires U_2 et V_2 qui maximise la corrélation (6.3) et qui n'est pas corrélée avec la première paire de variables canoniques, et ainsi de suite jusqu'à atteindre k couples de variables canoniques tel que $k = \min(p, q)$. La signification des corrélations canoniques a été testée avec la Bartlett's approximate chi-squared statistic (Bartlett, 1938, 1939; Johnson & Wichern, 2007):

$$H_0: \text{All } R_i^* = 0$$

$$H_1: R_i^* \neq 0 \text{ pour un } i$$

Le rejet de H_0 au seuil α est donné si

$$-(n-1 - \frac{1}{2}(p+q+1)) \ln \prod_{i=1}^k (1 - R_i^{*2}) > \chi_{pq}^2(\alpha) \quad (6.4)$$

Si l'hypothèse nulle est rejetée, les corrélations canoniques après la première peuvent être examinées individuellement afin de vérifier leur signification :

$$H_0: R_1^* \neq 0, R_2^* \neq 0, \dots, R_s^* \neq 0, R_{s+1}^* = 0, \dots, R_k^* = 0$$

$$H_1: R_i^* \neq 0 \text{ pour un } i \geq s+1$$

$$-(n-1 - \frac{1}{2}(p+q+1)) \ln \prod_{i=s+1}^k (1 - R_i^{*2}) > \chi_{(p-s)(q-s)}^2(\alpha) \quad (6.5)$$

L'interprétation des variables canoniques s'effectue via l'identification (1) des coefficients canoniques standardisés a_i et b_j , (2) des coefficients de structure $R(U_i, Z_j)$ et $R(V_i, Z_j)$, et (3) des coefficients canoniques de communalité h_2 . Les contributions des variables originales standardisées

$Z_1 = (z_{11}, z_{12}, \dots, z_{1p})$ et $Z_2 = (z_{21}, z_{22}, \dots, z_{2q})$ à la formation des variables canoniques ont été évaluées à partir des coefficients canoniques standardisés a_i et b_j . Ces poids sont générés dans le but de maximiser la corrélation canonique R_1^* et sont ainsi analogues aux poids en régression. Les coefficients standardisés permettent d'évaluer l'importance d'une variable donnée par rapport aux autres et reflètent ainsi leur contribution à la corrélation canonique. Il peut être également utile de considérer les coefficients de structure afin d'évaluer l'importance d'une variable donnée sans rapport aux autres. Ces coefficients correspondent aux corrélations entre les variables originales et les variables canoniques R :

$$\begin{aligned} R_{(U_i, Z_{1i})} &= \text{corr}(U_i, Z_{1i}) \\ R_{(V_i, Z_{2j})} &= \text{corr}(V_i, Z_{2j}) \quad i = 1, 2, 3, \dots, k \quad j = 1, 2, 3, \dots, \max(p, q) \end{aligned} \quad (6.6)$$

Ces corrélations indiquent la proportion de variance (lorsqu'elles sont élevées au carré) une variable observée partage linéairement avec la variable canonique. Notons que les corrélations R ne sont pas affectées par la standardisation des variables originales. Enfin, les coefficients canoniques de communalité h_2 , issus des R , correspondent à la somme du carré des R de toutes les variables canoniques interprétées dans l'analyse. Ils nous informent sur la proportion de variance d'une variable qui est expliquée par l'ensemble des variables canoniques utilisées dans l'analyse. Les variables présentant des coefficients canoniques de communalité faibles (i.e. en-deçà de 45 %) sont généralement omises lors de l'analyse (Thompson, 1991).

6.4 Sommaire

Les données relatives à l'érodabilité et à l'érosivité des sols présentent une couverture spatiale largement insuffisante pour qu'elles puissent être intégrées à l'étude. Il s'agit des seules données mises de côté en raison de l'ampleur de données manquantes sur l'ensemble des lacs à l'étude. De plus, étant donné l'ampleur et la diversité de données générées et la complexité des interactions régissant la phénologie des FEA, le choix d'analyse s'est porté vers les méthodes d'analyse canonique. L'objectif est de décrire les corrélations possibles entre les variables phénologiques et les variables environnementales sans, à terme, expliquer un groupe de variables par l'autre. La CCA a été choisie à cet effet car elle permet l'analyse simultanée de deux groupes de variables X_1 et X_2 en identifiant

et en quantifiant leurs associations. Notons que bien que la multicolinéarité entre les variables ne présente pas de difficulté analytique en ACC, elles peuvent compliquer l'interprétation des résultats en brouillant l'origine des effets observés (Reichwein Zientek & Thompson, 2006). L'utilisation jointe des coefficients standardisés et des coefficients de structure est donc recommandée puisque ces derniers ne sont pas affectés par la multicolinéarité et nous informe sur la contribution potentielle des variables observées sur le développement des variables canoniques (Henson, 2002). En plus de l'utilisation jointe de ces coefficients, la base de données environnementales utilisée pour l'ACC a tout de même été réduite à 14 variables (tableau 6.1) par le biais d'une analyse des corrélations de Pearson, de régressions linéaires multiples entre celles-ci et les variables phénologiques et d'une analyse des composantes principales appliquée sur les variables environnementales.

Chapitre 7

Résultats

7.1 Variabilité spatiotemporelle des FEA

La fréquence d'apparition des FEA est variable sur l'ensemble du territoire à l'étude (figure 7.1). Les apparitions sont plus fréquentes à l'ouest de la zone d'étude, et cette fréquence tend à diminuer systématiquement de l'ouest vers l'est (test de Kruskal-Wallis (KW): $\chi^2 = 934$; $p < 0.001$). Une augmentation systématique est également apparente du nord au sud de la zone d'étude (KW: $\chi^2 = 433$; $p < 0.001$). Les régions situées au sud-ouest du Québec sont donc beaucoup plus marquées par les événements de FEA. L'ensemble des lacs à l'étude présente une hausse significative (KW: valeur médiane, $p = 0.091$) de la fréquence d'apparition des FEA de 2000 à 2016 (figure 7.1). Afin d'évaluer s'il y a une tendance temporelle significative, un test paramétrique (t-test) de même qu'un test non-paramétrique (Mann-Kendall) ont été testés en parallèle sur les variables phénologiques. Notons que puisque les résultats obtenus étaient similaires, seuls les résultats issus des t-tests sont présentés. La médiane du nombre de jours d'épisodes de FEA a augmenté de 23 % entre 2000 et 2016. Une hausse significative (percentile 95, $p = 0.043$) est également présente et plus marquée pour les fréquences d'apparition élevées (non présenté sur la figure). Le 95^e percentile du nombre de jours d'épisodes a également augmenté de 24 % entre 2000 et 2016. Les fréquences d'apparition plus élevées sont généralement présentes sur les secteurs fortement urbanisés et agricoles (sud-ouest de la zone d'étude). À l'opposé, les fréquences d'apparition moins prononcées (percentile 5) ne montrent pas de tendances significatives dans le temps. Il s'agit généralement des fréquences d'apparition de lacs situés en régions très peu développées (nord-est de la zone d'étude).

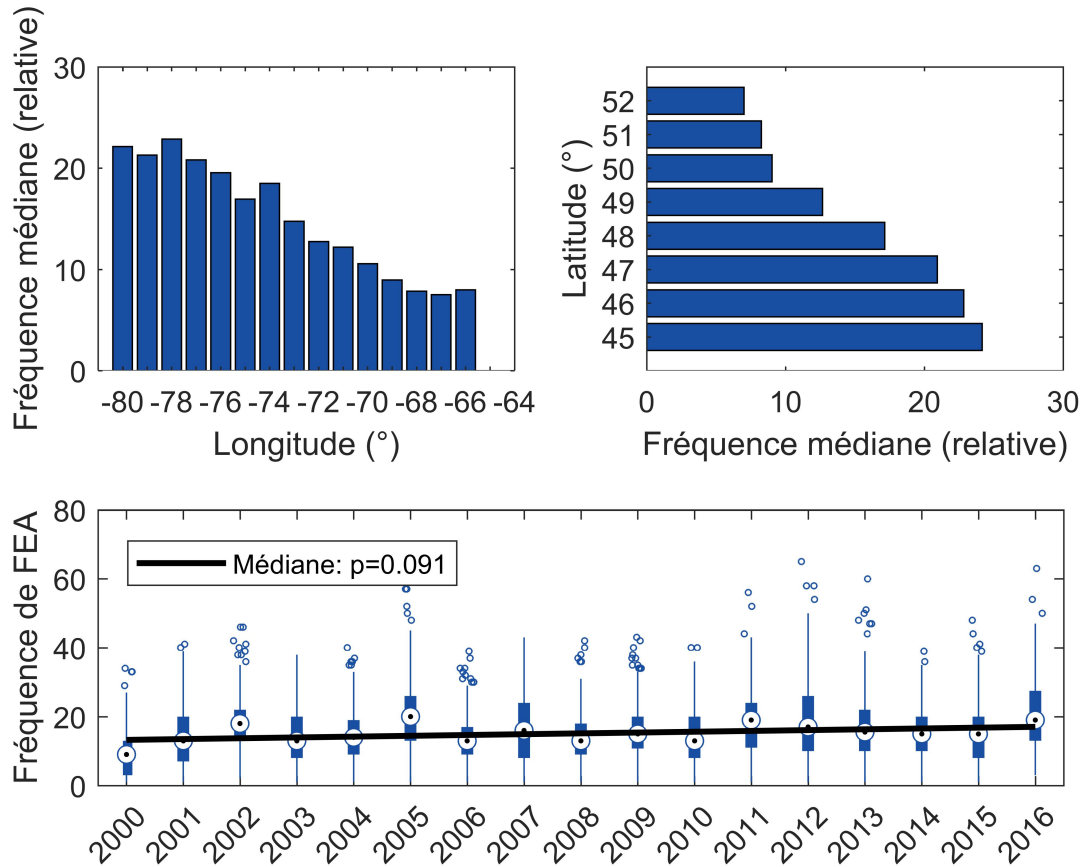


Figure 7.1 – Fréquence d’apparition de FEA (relatif au nombre de lacs) en fonction de la longitude (haut-gauche), de la latitude (haut-droit) et du temps (bas).

Bien qu’elle soit moins marquée que la fréquence d’apparition, l’intensité des FEA présente une variabilité spatiale sur l’ensemble du territoire (figure 7.2). Cette variabilité s’apparente à celle de la fréquence, soit une augmentation de l’intensité de l’est vers l’ouest (KW: $\chi^2 = 199$; $p < 0.001$) et du nord au sud du territoire (KW: $\chi^2 = 78$; $p < 0.001$). Aucune tendance temporelle de l’intensité des FEA n’est apparente de 2000 à 2016 (médiane, $p = 0.220$).

L’étendue des FEA (relatif à la superficie du lac) montre une variabilité spatiale très prononcée et différente de la variabilité de la fréquence d’apparition et de l’intensité des FEA (figure 7.3). L’étendue relative lors des épisodes de FEA est plus importante sur les lacs situés le long du 74°W au 77°W (KW: $\chi^2 = 512$; $p < 0.01$). La hausse marquée de l’étendue à 74°-75°W (N=117 lacs) correspond à la région qui débute au sud-ouest de l’île de Montréal, qui remonte vers Mont-Tremblant et qui se termine près de la réserve Obedjiwan. Une hausse est également présente du nord au sud de la zone d’étude, où l’étendue des FEA sur les lacs entre le 45°N et le 49°N est

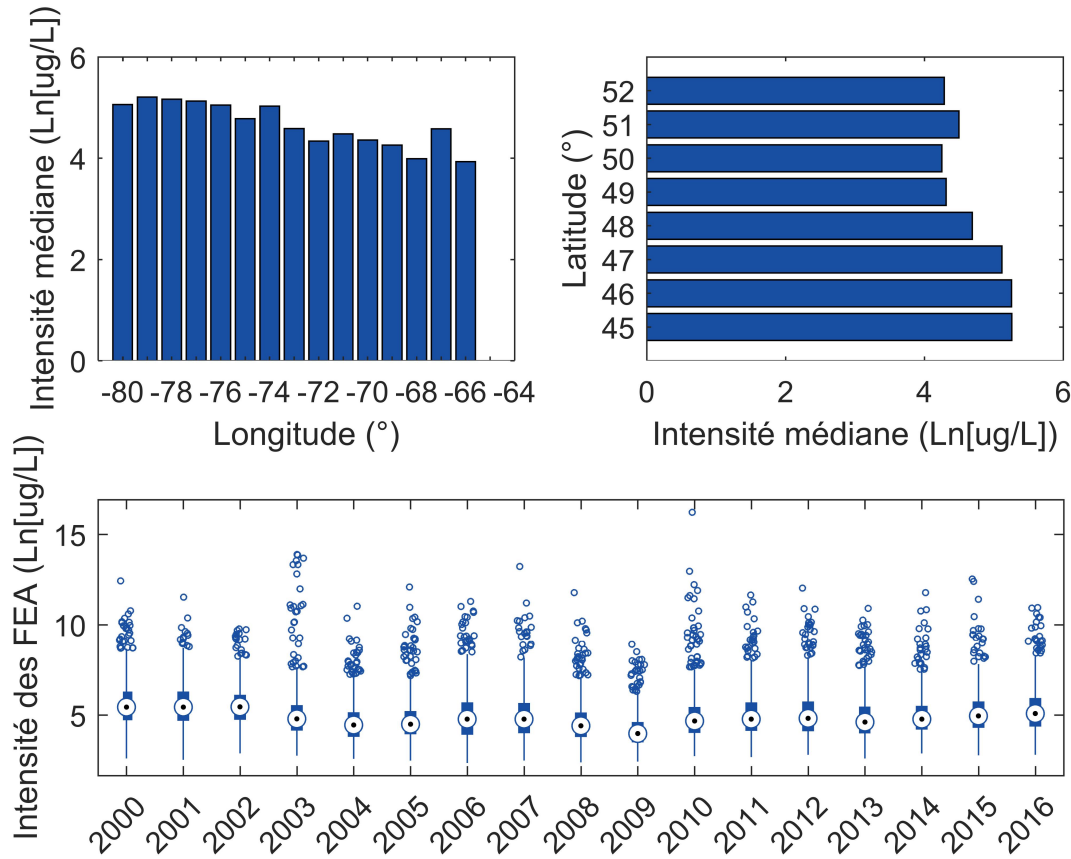


Figure 7.2 – Intensité des FEA en fonction de la longitude (haut-gauche), de la latitude (haut-droit) et du temps (bas).

généralement deux fois plus élevée que sur les lacs situés entre le 50°N et le 52°N (KW: $\chi^2 = 198$; $p < 0.01$). Aucune tendance temporelle significative de l'étendue n'est présente entre 2000 et 2016 (médiane, $p = 0.1836$). Cependant, les événements de FEA caractérisés par de très vastes étendues présentent une augmentation significative (percentile 95, $p = 0.025$), où l'étendue relative passe de 50 % à 55 % de 2000 à 2016.

La date d'apparition du premier épisode de FEA est plus hâtive sur les lacs localisés à l'ouest de la zone d'étude (figure 7.4). La date médiane de la première apparition est le 21 mai à 80°W alors que cette date est repoussée au 25 juin à 66°W (KW: $\chi^2 = 828$; $p < 0.01$). L'apparition est également de plus en plus hâtive sur les lacs situés plus au sud du territoire (KW: $\chi^2 = 508$; $p < 0.01$). La date médiane de la première apparition est le 21 mai à 45°N et le 27 juin à 52°N. Par ailleurs, une diminution significative (médiane, $p = 0.089$) de la première date d'apparition des FEA est détectée entre 2000 et 2016 sur l'ensemble des lacs du territoire (Figure 7.4); la date médiane est

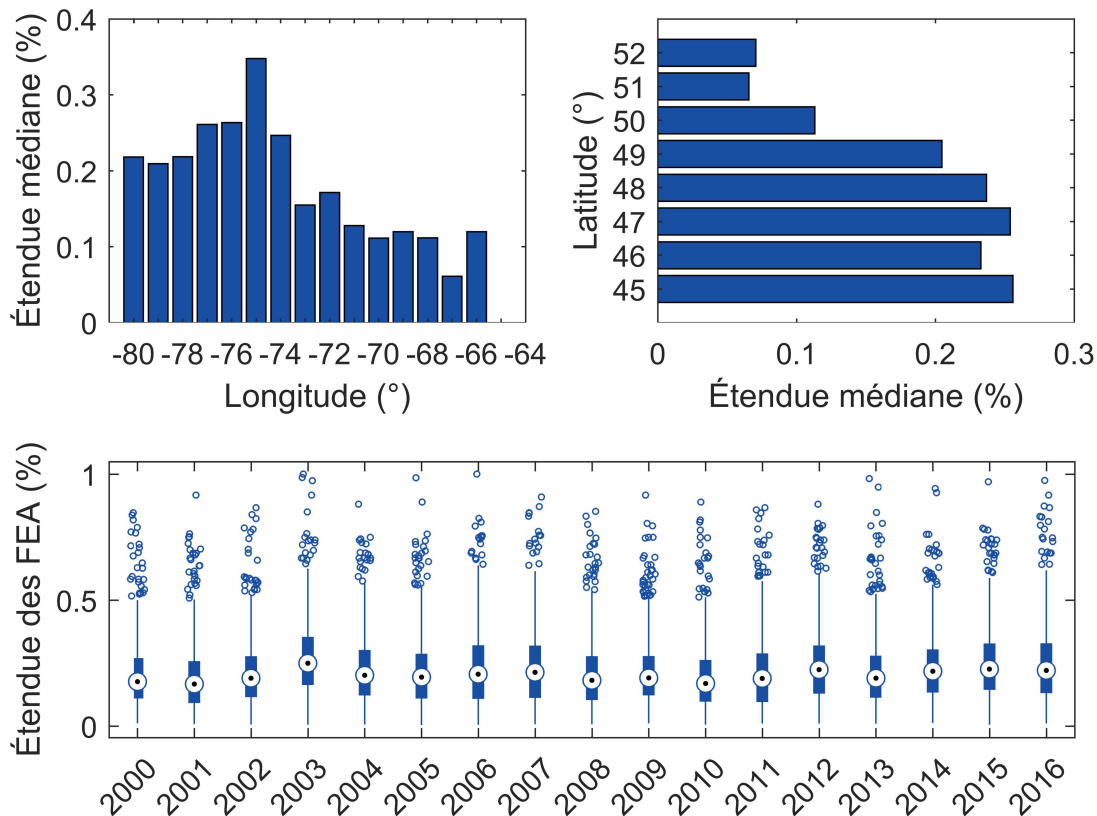


Figure 7.3 – Étendue des FEA (relatif à la superficie totale du lac) en fonction de la longitude (haut-gauche), de la latitude (haut-droit) et du temps (bas).

passée du 6 au 3 juin. Le percentile 95 de la première date d'apparition des FEA présente également une diminution mais non significative (percentile 95, $p=0.1065$), passant du 14 juillet au 1 juillet de 2000 à 2016. Ces lacs sont généralement localisés au nord de la zone d'étude, où très peu de développements agricole et urbain ne sont présents. Les lacs présentant généralement une première apparition de FEA très hâtive au printemps ne montrent également pas de tendance significative. La date d'apparition se maintient annuellement au 18 ou 19 mai de 2000 à 2016 (percentile 5, $p=0.826$). Il s'agit de lacs généralement localisés au sud-ouest de la zone d'étude, des secteurs fortement urbanisés et agricoles.

La date du dernier épisode de FEA annuel montrent des distributions annuelles très asymétriques (figure 7.5). Cela est dû à la couverture nuageuse très fréquente au Québec durant l'automne et qui réduit considérablement la base de données d'images MODIS disponibles pour cette période. Le mois d'octobre, plus précisément, présente en moyenne deux fois moins d'images MODIS sans

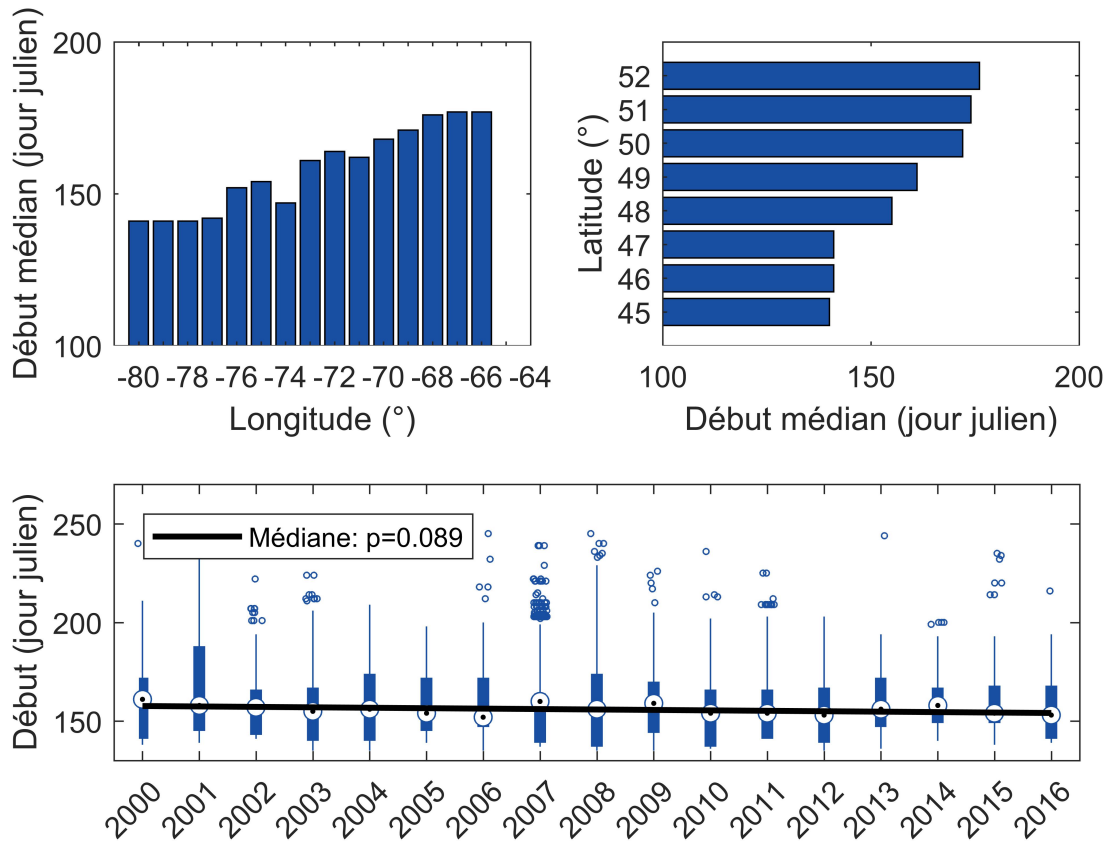


Figure 7.4 – Début des épisodes annuels de FEA en fonction de la longitude (haut-gauche), de la latitude (haut-droit) et du temps (bas).

couvert nuageux complet entre 2000 et 2016 que les mois de mai à septembre. Pour cette raison, la date de fin, et par extension la durée, n'ont pas été intégrées à l'étude.

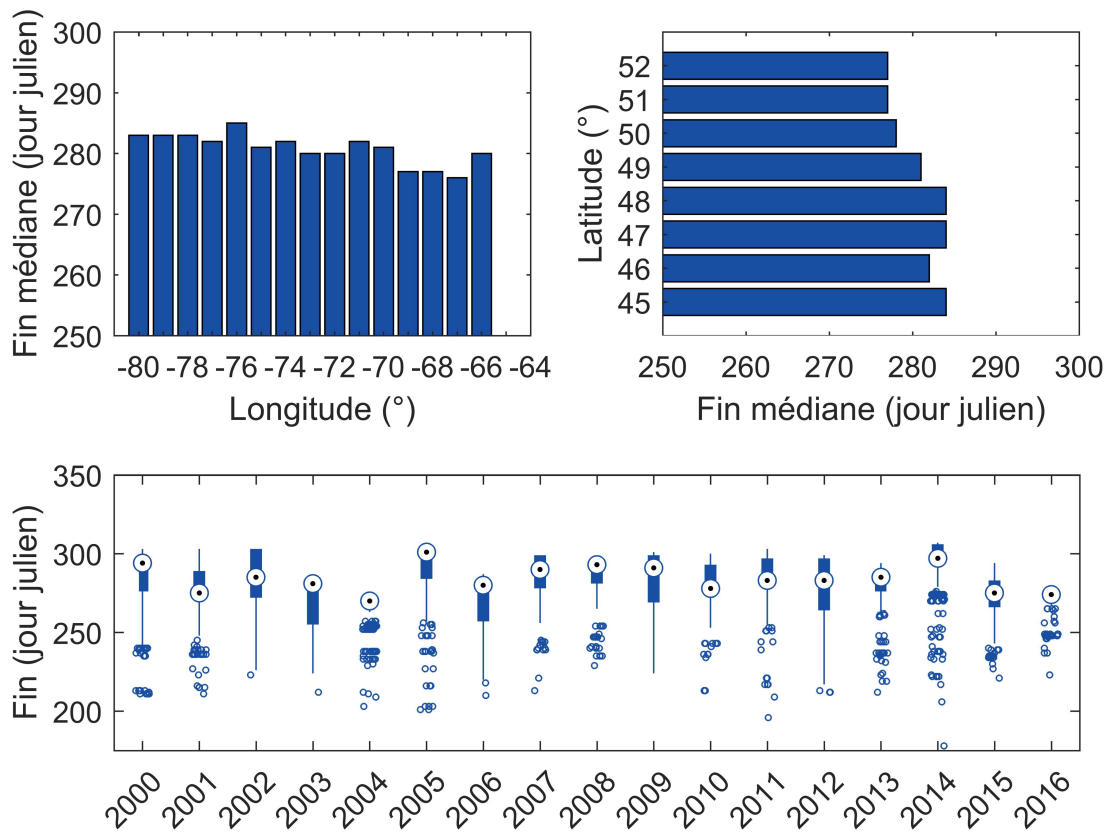


Figure 7.5 – Fin des épisodes annuels de FEA en fonction de la longitude (haut-gauche), de la latitude (haut-droit) et du temps (bas).

7.2 Tendances phénologiques du lac Brome et de la baie Missisquoi

La baie Missisquoi et le lac Brome se retrouvent parmi les milieux lacustres les plus étudiés dans la région d'étude (Rashidan & Bird, 2001; Rolland *et al.*, 2005; Fortin *et al.*, 2010; Wheeler *et al.*, 2012) en raison de l'ampleur des événements algaux et des problèmes socio-économiques qu'ils engendrent à chaque année. Les tendances phénologiques de ces deux lacs sont présentés aux figures 7.6 et 7.7. La fréquence d'apparition annuelle a augmenté significativement sur la Baie Missisquoi, où le nombre de jours de FEA est passé de 25 en 2000 à 47 en 2016. L'intensité et l'étendue de ces efflorescences sont importantes, en raison de concentrations maximales en chl-a $1315\mu gL^{-1}$ en moyenne et d'étendues maximales de 62% en moyenne, mais ne présentent pas de hausses ou baisses significatives tout comme pour la date de début. En ce qui a trait au lac Brome, la fréquence annuelle présente également une hausse significative passant de 23 jours en 2000 à 30 jours en 2016. Bien que

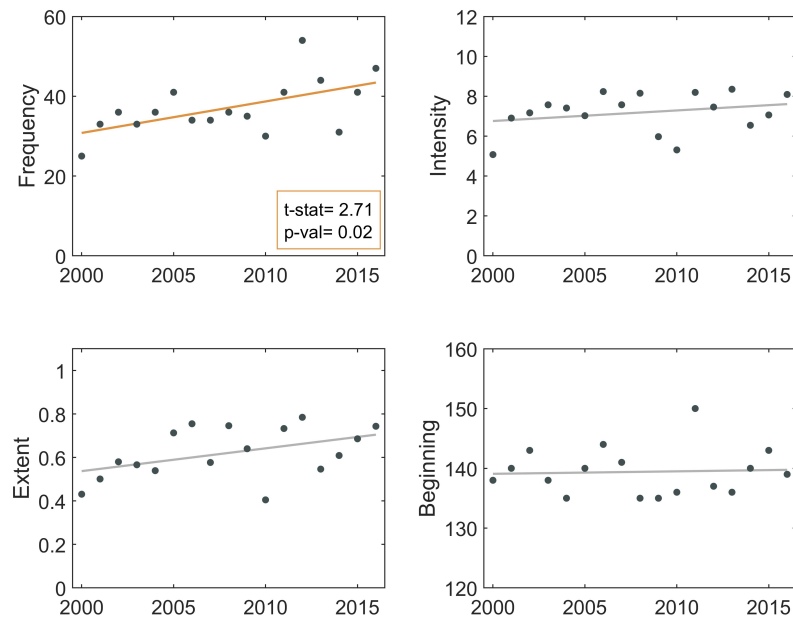


Figure 7.6 – Fréquence, intensité, étendue et début annuels de 2000 à 2016 sur la Baie Missisquoi.

l'intensité maximale ($219\mu gL^{-1}$ en moyenne) soit moins importante que pour la Baie Missisquoi, l'étendue maximale annuelle atteint entièrement (sauf en 2010) la superficie du lac. Quoique non significative, la première FEA annuelle apparaît de plus en plus hâtivement au printemps.

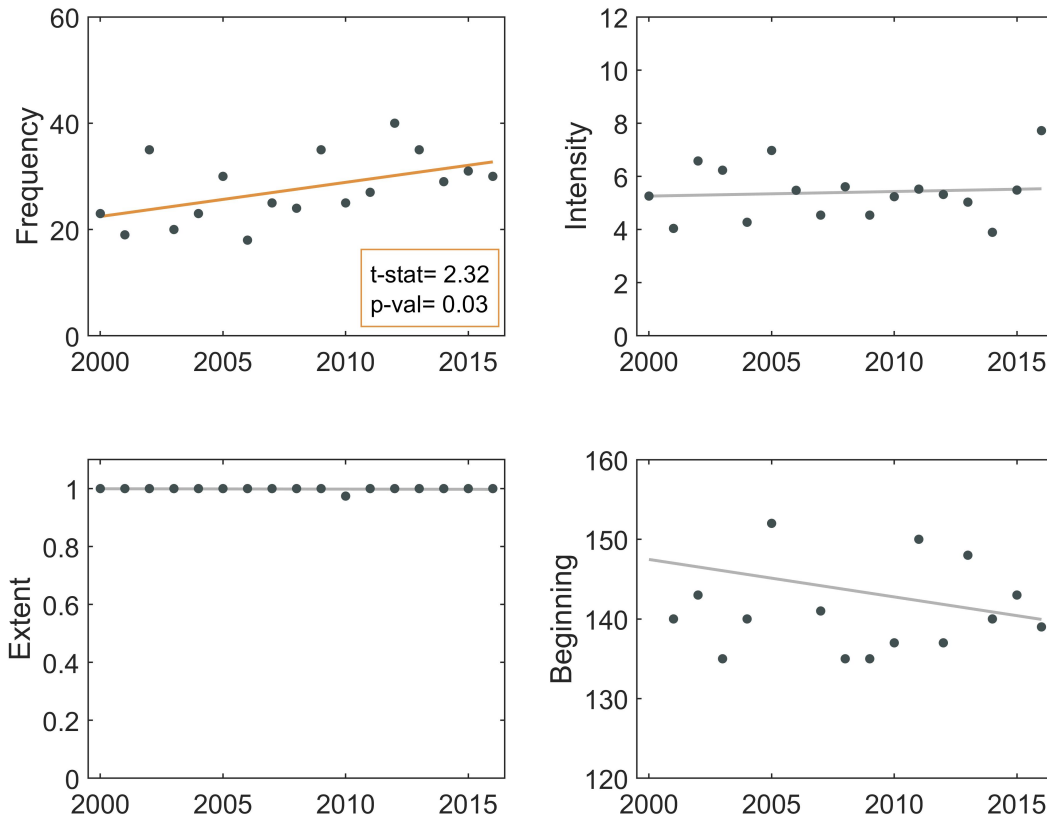


Figure 7.7 – Fréquence, intensité, étendue et début annuels de 2000 à 2016 sur le lac Brome.

7.3 Analyse des corrélations de Pearson

Les analyses des corrélations de Pearson intra et inter variables phénologiques et environnementales sont présentées à l'annexe D. Cette analyse a permis de réduire le nombre de variables environnementales afin d'éliminer la redondance d'information, tout en sélectionnant les variables les plus corrélées à la phénologie des FEA. Notons que la fréquence d'apparition et la date d'apparition de la première FEA annuelle montrent une forte corrélation avec les degré-jours, alors que l'intensité est plutôt corrélée à la morphologie du plan d'eau et à la température de l'eau. L'étendue ne présente aucune corrélation notable avec les variables environnementales.

7.4 Analyse canonique des corrélations

Les résultats de la CCA appliquée entre les facteurs environnementaux et la phénologie des FEA montrent des R_i^* élevés pour les deux premières paires de variables canoniques (0.78 et 0.72) et qui diminuent considérablement pour les dernières corrélations canoniques (0.27, 0.22, et 0.14). Le test de Bartlett montre que ces corrélations canoniques sont significatives au seuil $\alpha = 0.01$ (tableau 7.1). En se basant sur les valeurs de R_i^* obtenues, l'interprétation des résultats s'est, par conséquent, conduite sur les deux premières paires de variables canoniques.

Le tableau 7.2 présente les coefficients canoniques standardisés, les coefficients de structure et les communality coefficients. Les coefficients de structure où $|R| \geq 0.45$ et les communality coefficients où $h^2 \geq 45\%$ sont soulignés afin de mettre l'accent sur les variables les plus utiles au développement du modèle.

Tableau 7.1 – Corrélations canoniques et test de signification des corrélations canoniques.

i	R_i^*	Test statistique	Degrés de liberté	χ^2 distribution ($\alpha = 0.01$)	p -value
1	0.78	14992.1	70	100.43	<0.01
2	0.72	7355.07	52	78.62	<0.01
3	0.27	1198.8	36	58.62	<0.01
4	0.22	575.72	22	40.29	<0.01
5	0.14	169.71	10	23.21	<0.01

En observant d'abord les résultats obtenus de la fonction 1, on remarque que la variable phénologique contribuant le plus à la corrélation entre U_1 et V_1 est la fréquence d'apparition ($b_1 = -0.58$), suivie de l'intensité ($b_2 = -0.47$), la date de début ($b_4 = 0.30$) et finalement l'étendue ($b_4 = 0.28$). Les variables avec le plus haut niveau d'utilité dans le développement du modèle apparaissent dans le même ordre, c'est-à-dire avec d'abord la fréquence d'apparition ($R = -0.86$), suivie de l'intensité ($R = -0.70$), la date de début ($R = 0.65$) et finalement l'étendue ($R = -0.09$). La phénologie des FEA est donc caractérisée principalement par la fréquence et l'intensité, suivie du début qui apporte une contribution notable. Chacune de ces variables apporte une information complémentaire puisqu'elles ne présentent pas de forte colinéarité entre elles ($r < 0.7$, voir annexe D). Les coefficients de structure de la fréquence, de l'intensité et de l'étendue sont inversement reliés à celui de la date de début, en raison de leur signe inverse. Ainsi, les FEA présentant des fréquences, intensités et étendues élevées ont eu tendance à apparaître plus tôt dans la saison. En ce qui a trait à la fonction

2, elle présente une forte contribution de l'étendue ($b_3 = -0.72$), suivie de l'intensité ($b_2 = 0.53$), de la date de début ($b_4 = 0.29$) et finalement de la fréquence ($b_1 = -0.14$).

Les variables environnementales contribuant le plus à la corrélation canonique entre U_1 et V_1 sont l'aire du lac ($a_1 = -0.75$), l'aire du bassin versant du lac ($a_3 = 0.62$), les degré-jours ($a_{13} = -0.42$), la couverture forestière ($a_6 = -0.41$), la superficie urbaine ($a_7 = -0.24$) et les précipitations annuelles ($a_{11} = 0.22$). Les variables les plus utiles au développement des variables canoniques ($|R| \geq 0.45$) sont généralement les mêmes que celles présentant des coefficients standardisés élevés, à l'exception de la couverture forestière de la fonction 2, qui présente un coefficient standardisé modeste ($a_6 = 0.19$) et un coefficient de structure élevé ($R = 0.51$). Tel que mentionné précédemment, ce résultat est lié à la multicollinéarité que cette variable présente par rapport aux autres. Le coefficient de structure de la couverture forestière ($R = -0.49$) étant de même signe que ceux de la fréquence ($R = -0.86$), de l'intensité ($R = -0.70$) et de l'étendue ($R = -0.09$), cela indique que cette variable reflèterait plutôt sa contribution à la superficie totale du bassin versant plutôt que la superficie forestière. L'apport en superficie forestière est plutôt reflété dans la fonction 2, où le signe du coefficient de structure diffère ($R = 0.51$). De manière générale, la superficie forestière contribue fortement à expliquer la variabilité spatio-temporelle des variables phénologiques ($h^2 = 49.8\%$), mais celle-ci porte en partie l'empreinte de l'aire du bassin dans la fonction 1. Les superficies en zones urbanisées correspondent au caractère physiographique le plus relié à la phénologie des FEA après la couverture forestière ($b_7 = -0.24$), et montre une corrélation directe avec la hausse en fréquence, en intensité et en étendue (R de signe inverse). Quant à la fonction 2, les variables environnementales contribuant le plus à cette corrélation canonique est l'aire du lac ($a_1 = 0.71$), les degré-jours ($a_{13} = -0.43$) et l'écoumène de population ($a_9 = -0.27$). Toutes les variables phénologiques contribuent au développement des deux premières paires de variables canoniques ($h^2 \geq 45\%$), tandis que l'aire du lac, l'aire du bassin versant, l'occupation du sol (forêts et zones urbanisées) et les degré-jours annuels sont les variables environnementales qui contribuent le plus.

Tableau 7.2 – Coefficients canoniques, coefficients de structure et coefficients de communalité des deux premières variables canoniques. Les coefficients de communalité plus élevés que 45 % sont présentés en gras.

		Morphologie						Physiographie				Climat							
Phénologie des FEA		Superficie lac	Forme lac	Superficie bassin	Forme bassin	Pente	Forêt	Zones développées	Agriculture	Écume pop	Écume agr	Précipitations	Température	Degrés-jour	Vents				
a_i, b_j :		-0.58	-0.47	0.28	0.3	-0.75	0.04	0.62	-0.12	-0.12	-0.41	-0.24	0.13	-0.07	0.01	0.22	-0.05	-0.42	-0.01
R :		-0.86	-0.7	-0.09	0.65	-0.75	-0.53	-0.53	-0.47	0.23	-0.49	-0.68	-0.2	-0.39	-0.33	0.47	0.33	-0.65	-0.11
a_i, b_j :		-0.14	0.53	-0.72	0.29	0.71	-0.19	-0.08	-0.07	0.15	0.19	-0.22	0.19	-0.27	-0.17	-0.15	0.04	-0.43	0.04
R :		-0.4	0.36	-0.82	0.56	0.59	0.38	0.51	0.27	0.21	0.51	-0.14	0.4	-0.35	-0.17	-0.24	-0.39	-0.6	-0.09
h_2 :		89.2%	62.3%	67.5%	73.3%	90.7%	43.0%	53.9%	29.4%	9.5%	49.8%	48.8%	19.9%	27.3%	13.8%	27.4%	26.6%	77.9%	2.1%

7.5 Sommaire

Cette étude présente la dynamique spatio-temporelle de FEA sur 580 lacs du Québec méridional entre 2000 et 2016, et présente la caractérisation de leurs liens potentiels avec des descripteurs physiographiques, morphologiques et climatiques prévalant sur les lacs et leur bassin versant. Les résultats montrent des tendances réalistes et attendues, qui valident l'utilité de cette approche pour analyser l'évolution temporelle de l'état trophique des lacs en réponse aux changements historiques et futurs du bassin versant et du climat. À titre d'exemple, les données montrent l'augmentation attendue de l'importance des FEA (exprimée par la fréquence, l'étendue ou l'intensité) du nord au sud et de l'est vers l'ouest, ainsi qu'une apparition plus hâtive au sud et à l'ouest de la région d'étude. Ces tendances spatiales avaient été remarquées durant les dernières décennies par le MELCC et sont généralement localisées sur les secteurs fortement urbanisés et agricoles. On remarque également une plus grande étendue du phénomène pour les lacs situés dans le corridor entre 74 et 76°W s'étalant du sud-ouest de Montréal jusqu'au Mont-Tremblant, une région de forte villégiature.

Un résultat particulièrement intéressant issu de cette approche est qu'elle permet de mettre en évidence l'augmentation temporelle du phénomène entre 2000 et 2016 tel qu'obtenu par Winter *et al.* (2011) de 1994 à 2009 en Ontario, et infirmant l'hypothèse que cette tendance serait simplement liée à une augmentation de l'attention portée au phénomène. De manière générale, cette étude permet de quantifier les tendances phénologiques dans le temps et en fonction des caractéristiques du paysage, alimentant ainsi les modèles prédictifs.

7.5.1 Tendances phénologiques

Durant la période d'étude, la biomasse algale a dépassé le seuil de 10 $\mu\text{g chl a L}^{-1}$ 15 jours par année (mai à novembre) en moyenne sur l'ensemble des lacs. Notons qu'il s'agit des jours où la couverture nuageuse pour un lac donné est inférieure à 25% de la superficie observable (seuil établi suivant les travaux de Duan *et al.* (2014)), ce résultat est donc conservateur. Il est intéressant de constater que seuls 17 lacs ont vu leur biomasse augmenter au-delà du seuil seulement après le 1er septembre, alors que ce nombre augmente à 95 pour les efflorescences apparaissant après le 1er août. D'ailleurs, la plupart des cas de dépassement surviennent dès le mois de juin. Ainsi, lorsqu'un lac est touché par des FEA, la croissance algale a tendance à apparaître relativement tôt durant

l'été et à se maintenir. Par ailleurs, lorsqu'un lac est touché par les FEA, il est plus probable qu'il le soit de nouveau les années subséquentes. Ce résultat pourrait suggérer qu'une fois les conditions favorisant une forte croissance algale dans un lac atteintes, cet état est plus propice à se maintenir dans le temps (alternative stable state; Carpenter *et al.* (1999))

En moyenne sur les lacs étudiés (valeurs médianes), la biomasse atteignait au maximum $110 \mu\text{gL}^{-1}$, et l'étendue de cette biomasse occupait au maximum 19% de la superficie totale du lac. Ainsi, la plupart des cas de FEA était plutôt restreinte spatialement, si l'on compare cette étendue avec les critères du MELCC qui qualifie une FEA de très limitée lorsque l'étendue demeure inférieure à 25% de la superficie (S. Blais, personal communication). En effet, 572 lacs ont montré une FEA d'intensité maximale à $100 \mu\text{gL}^{-1}$ alors que seulement 363 lacs ont montré une FEA d'étendue supérieure à 25% de la superficie du lac, ce nombre diminuant à 68 lacs pour une étendue supérieure à 50%. L'atteinte de fortes biomasses est donc fréquente mais le problème est plus rarement généralisé à l'ensemble du lac.

La phénologie des FEA montre un patron spatiotemporel clair. Une hausse significative de la médiane des fréquences d'apparition des FEA sur les lacs du Québec méridional est notable, où le nombre de jours d'épisodes a augmenté de 23% entre 2000 et 2016. On retrouve des fréquences annuelles élevées de FEA particulièrement sur les lacs localisés dans des secteurs très développés (forte urbanisation et/ou occupation agricole). De plus, l'augmentation en fréquence a affecté principalement les lacs moyennement touchés (fréquence moyenne annuelle de 15 jours) à fortement touchés (28 jours).

L'ampleur de l'étendue des FEA varie également spatialement, avec une hausse marquée entre les longitudes 74 et 75° W. Cette hausse régionale (moyennée sur la période d'étude 2000-2016) est alimentée en partie par le lac des Deux-Montagnes et le lac Saint-François montrant des efflorescences d'algues étendues sur l'ensemble du lac mais à faible teneur en chl a (valeurs médianes lorsque le seuil est dépassé: $14,4 \pm 1,1 \mu\text{gL}^{-1}$ et $12,6 \pm 1,1 \mu\text{gL}^{-1}$ respectivement pour les deux lacs). D'autre part, la médiane de l'étendue n'a pas augmenté significativement entre 2000 et 2016, alors que le percentile 95 de l'étendue, qui correspond à des événements de très vastes étendues (proportionnellement), montre une augmentation significative. Ainsi, les lacs présentant des efflorescences très fréquentes et très étendues (incluant par exemple le lac Brome et la Baie Missisquoi) ont montré

une augmentation significative entre 2000 et 2016. Ce résultat suggère que les lacs montrant un stade plus avancé d'eutrophisation sont en augmentation.

L'apparition de la première FEA annuelle est plus hâtive en 2016 qu'en 2000, (recul de 3 jours). À titre indicatif, ce recul en terme absolu est passé du 6 au 3 juin. Rappelons toutefois que la précision de l'estimation sur la date d'initiation est influencée par l'ampleur de données manquantes. Cette tendance temporelle s'accorde avec les quelque études de tendances des dates d'initiation, telle que sur le lac Taihu, Chine, où la date d'initiation de FEA a reculé de 11 jours par année entre 1998 et 2007 (Duan *et al.*, 2009) et de 0.9 jour par année entre 1987 et 2009 (Zhang *et al.*, 2012). La date d'initiation de FEA est également de plus en plus hâtive sur les lacs situés plus au sud du territoire. Il est possible que ce résultat soit lié de manière directe à une hausse des températures de l'air, provoquant ainsi une diminution de la période d'englacement et une prolongation de la saison estivale ce qui a pour effet de hausser la température de surface et intensifier la stratification thermique des milieux lacustres (Dibike *et al.*, 2011). Ces conditions stimulent la croissance des communautés phytoplanctoniques en milieux eutrophes (Kraemer *et al.*, 2017) et tout particulièrement celle des cyanobactéries.

7.5.2 Liens avec le climat et la physiographie du milieu

Ces tendances spatiotemporelles dans l'apparition, le maintien, le développement et le déclin des FEA laissent supposer qu'elles dépendent des caractéristiques physiographiques du bassin versant, des caractéristiques morphologiques du plan d'eau et des conditions climatiques prédominantes sur le bassin versant. On suppose évidemment que l'apport et la quantité d'éléments nutritifs d'origine anthropique sera plus importante sur les secteurs urbanisés et agricoles. Ainsi, les milieux lacustres du sud sont plus sujets à la pollution diffuse (liées généralement à l'agriculture) et ponctuelle (e.g., les installations sceptiques, les rejets de produits domestiques ou industriels), tandis que ces sources de pollution sont moins abondantes au nord de la zone d'étude qui est peu développée. Par ailleurs, puisque le territoire à l'étude s'étend sur près de 1520 km d'est en ouest et 885 km du nord au sud, le climat joue sur cette zone un rôle indéniable dans la phénologie des FEA (Zhang *et al.*, 2012). Entre 2000 et 2016, on constate une augmentation de la fréquence d'apparition des FEA et une apparition de plus en plus hâtive au début de l'été. Il est probable que ces changements surviennent en réponse aux changements climatiques, mais aussi potentiellement aux changements dans l'occupation du

territoire à travers la période d'étude. Les résultats indiquent également que sur le territoire à l'étude, il existe une différence de 37 jours dans la date de première apparition des FEA, ce délai pouvant être exploité dans le cadre d'un space-for-time substitution afin d'évaluer la réponse des lacs au climat futur à ces latitudes plus nordiques (Blois *et al.*, 2013; Pickett, 1989). Les tendances temporelles observées sur une période de 17 ans démontrent l'importance de cette problématique dans un contexte d'expansion des populations survenant en parallèle au réchauffement climatique, une question particulièrement préoccupante dans un pays qui compte plus d'un million de lacs sur son territoire (Pick, 2016).

L'analyse canonique a permis de faire ressortir les caractéristiques qui sous-tendent cette variabilité spatiotemporelle, et d'expliquer la dynamique régionale des FEA au Québec. La fréquence et l'intensité des épisodes de FEA sont les variables phénologiques les plus fortement reliées aux caractéristiques environnementales prévalant sur les lacs et leur bassin versant. Ces deux variables apportent chacune une information spécifique au phénomène et devraient être étudiées de manière simultanée. Une apparition plus hâtive du premier épisode annuel est notable sur les lacs présentant des épisodes de FEA plus fréquents et plus intenses. La morphologie du plan d'eau et celle de son bassin versant jouent un rôle prépondérant dans la variabilité spatiotemporelle de la fréquence, de l'intensité et du début des FEA, particulièrement la superficie totale du lac et celle du bassin versant, alors que cette dernière se réfère directement à la capacité de stockage d'éléments nutritifs.

Plus particulièrement, la superficie forestière (fortement reliées à la taille du bassin versant) et la superficie en zones urbanisées sont les variables physiographiques expliquant le plus la variabilité spatiotemporelle de la phénologie des FEA. L'ampleur des milieux urbanisés est directement proportionnelle à la fréquence et l'intensité des épisodes de FEA sur le bassin versant. Bien qu'il soit connu que l'enrichissement en P et en N dans les milieux lacustres engendre une prolifération des FEA, les sources responsables de leur augmentation en fréquence, en intensité et en étendue sont encore très suggestives. Les résultats de l'ACC montrent que les sources de contamination diffuse provenant des milieux agricoles ne sont pas les seuls responsables, puisque les variables reflétant l'urbanisation (settlement et population ecumene) expliquent beaucoup plus la variabilité spatiotemporelle des FEA que les variables reflétant l'agriculture (cropland and agriculture ecumene).

Les degrés-jours correspondent de loin à la variable climatique expliquant le mieux les variables phénologiques, suivis par la précipitation totale annuelle. Comparativement aux degrés-jour, les

températures moyennes annuelle et saisonnière de l'eau ne présentent aucune relation évidente avec les variables phénologiques à l'étude. Bien que dans la littérature plusieurs descripteurs de température aient montré une influence notable sur la prolifération de FEA, incluant la température de l'air (Hu *et al.*, 2009; Zhang *et al.*, 2012), la température de l'eau (Elliott, 2012), les heures d'ensoleillement (Zhang *et al.*, 2012) et les degrés-jour (Ralston *et al.*, 2014; Larson *et al.*, 2018), il semble que les degrés-jour soient plus directement reliés au taux de croissance des organismes ectothermes (Trudgill *et al.*, 2005) et devraient être privilégiés dans le cadre d'une étude sur le suivi ou la prédiction des FEA.

7.5.3 Incertitude sur l'estimation de la biomasse et des tendances globales

Il existe évidemment une incertitude sur l'estimation des concentrations en chl a obtenues avec le modèle basé sur les méthodes d'ensemble développé par El-Alem *et al.* (2019). Ce modèle présentait un coefficient de détermination de 0.85 lorsqu'il a été testé sur une banque de données indépendantes récoltées sur un ensemble de lacs du Québec méridional, et où la concentration en chl a avait été mesurée en parallèle par des méthodes d'échantillonnage conventionnelles. Cependant, le modèle présentait une performance globale beaucoup plus élevée sur l'estimation des fortes concentrations où l'erreur d'estimation à des concentrations en deçà de $50 \mu\text{gL}^{-1}$ était de 37% alors que l'erreur d'estimation de l'ensemble des concentrations était 15%.

D'autre part, notons que le calage et la validation du modèle n'a pas intégré de lacs en milieux fluviaux, où le régime hydrologique et les composantes optiquement actives diffèrent des autres milieux lacustres, cela menant notamment à une plus forte teneur en particules inorganiques en suspension. Cela pourrait causer un biais sur les estimations de chl a lorsque les particules non algales sont proportionnellement élevées, quoi qu'une évaluation qualitative effectuée par El-Alem *et al.* (2019) sur ce genre d'interférence semble indiquer que le modèle d'estimation se comporte bien en présence d'une plus forte teneur en particules inorganiques au lac Huron. Ainsi, les lacs des Deux-Montagnes, Saint-Louis, Saint-François et Saint-Pierre, lacs localisés le long du fleuve Saint-Laurent, ont été présentés à titre indicatif seulement. L'ajout de ces lacs n'influence pas les résultats globaux présentés dans le cadre de cette étude, puisque l'analyse de tendance spatiotemporelle et l'ACC ont aussi été effectués sans ces lacs et les résultats ne changeaient pas. Dans le cas du lac des Deux-Montagnes en bordure de la Ville de Montréal, la concentration médiane estivale en chl

a était de $4.8 \mu\text{gL}^{-1}$ en 2000 selon nos résultats, alors que les mesures in situ obtenues pour les années 1997-1998 variaient entre 2 et $5 \mu\text{gL}^{-1}$ à proximité de ce lac (MELCC, 2000).

Finalement, le pourcentage de données manquantes relié à une couverture nuageuse importante (plus de 75% du lac) ou à des distorsions géométriques des images MODIS est d'environ 68% à 45°N pour la présente étude, et augmente graduellement avec la latitude pour atteindre 80% de données manquantes à 52°N (figure 7.8). Le nombre d'heures d'ensoleillement au Québec (de Jong *et al.*, 2017; Canadian Geographic, 2018) explique bien cette variabilité spatiale du pourcentage de données manquantes. Certaines études ont démontré que les données manquantes ont un impact sur l'exactitude des variables phénologiques, particulièrement sur la fréquence et la date d'initiation d'une FEA. À titre d'exemple, Cole *et al.* (2012) ont trouvé des erreurs sur la date d'initiation de FEA en milieu océanique de l'ordre de 2 à 3 jours avec 10% de données manquantes, et de 15 à 30 jours avec 80% de données manquantes via l'utilisation d'images du capteur SeaWiFS en milieu océanique. Toutefois, Cole *et al.* (2012) ont obtenu une probabilité d'erreur (date plus hâtive que celle observée) de 50% (Figure 6 dans Cole *et al.* (2012)), signifiant qu'il y avait autant de chances que la date d'initiation soit en réalité plus hâtive que plus tardive. C'est pourquoi les résultats obtenus dans la présente étude sont potentiellement empreints d'une incertitude reliée à la couverture nuageuse qui, certes, demeure difficilement quantifiable mais l'apport en données satellitaires permet d'améliorer considérablement la précision et l'exactude de la modélisation spatio-temporelle du phénomène comparativement à l'échantillonnage *in situ*.

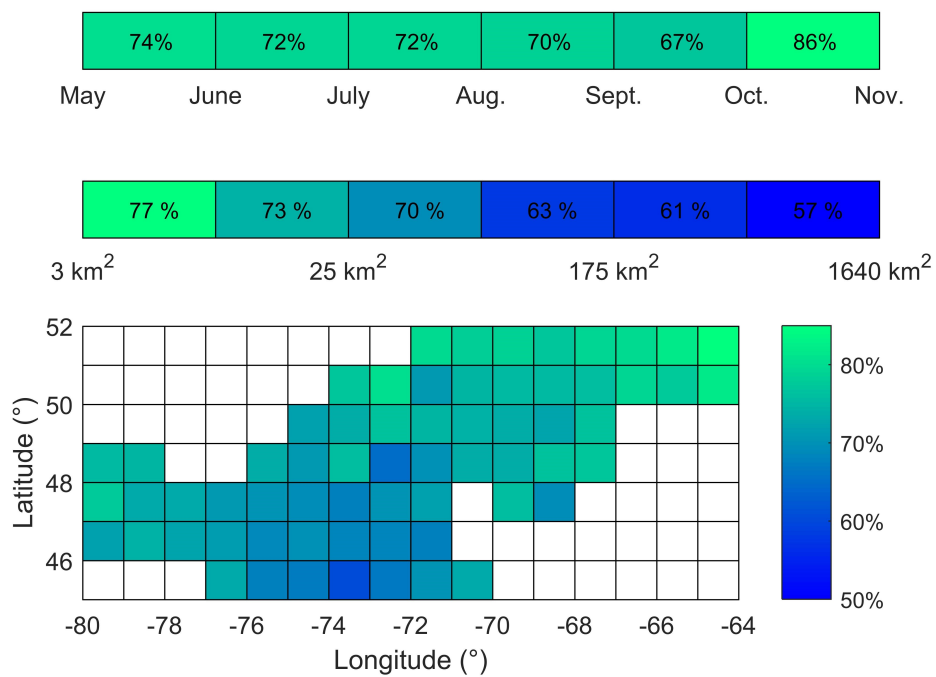


Figure 7.8 – Pourcentage moyen (%) entre 2000 et 2016 de données manquantes reliées à la couverture nuageuse trop importante (plus de 75% du lac) sur un lac donné ou présentant des distorsions géométriques de l'image MODIS. Le pourcentage moyen de jours manquants est présenté par mois, par superficie du lac, et selon l'emplacement géographique (latitude longitude).

Quatrième partie

**MODÉLISATION DE L'ALÉA
STRUCTUREL DES FEA**

Chapitre 8

Revue de littérature des modèles développés pour les FEA

La modélisation des FEA d'eau douce se divise en deux principales catégories: les modèles basés processus et les modèles stochastiques (basés statistiques). Une revue des études employant ces techniques est présentée dans ce chapitre.

8.1 Modélisation déterministe (basée processus)

Les modèles déterministes s'efforcent de modéliser les processus physiques régissant, par exemple, la taille des populations de phytoplancton en utilisant le maximum de connaissances sur le sujet. Plusieurs travaux se sont intéressés à l'utilisation de modèles déterministes afin de décrire les mécanismes d'initiation, de développement, de maintien et de déclin des FEA, incluant des modèles écologiques traitant des mécanismes de prolifération des FEA (Capblancq & Décamps, 2002; Bonnet & Poulin, 2004; Janssen *et al.*, 2004; Laanemets *et al.*, 2006; Onderka, 2007) et des modèles hydrologiques traitant surtout des mécanismes de production, de transport et d'accumulation des nutriments (Garnier *et al.*, 2005; Silva *et al.*, 2011). Entres autres, Bonnet & Poulin (2004) ont développé le modèle DyLEM-1D qui décrit la structure thermique et biogéochimique (incluant le cycle phytoplanctonique et la variabilité des nutriments et de l'oxygène dissous) d'un réservoir sur une période de 10 ans. Garnier *et al.* (2005) ont exploré les liens entre l'eutrophisation, les charges

en nutriments et les conditions hydrologiques sur le bassin de la rivière Marne (France) à l'aide du modèle écologique Riverstrahler. Les travaux de Janssen *et al.* (2004) ont porté sur l'étude de la variabilité inter-annuelle de données simulées de FEA à l'aide du modèle biogéochimique ERGOM.

8.2 Modélisation stochastique (basée statistique)

Les modèles stochastiques peuvent s'appuyer sur des bases de données historiques afin d'obtenir la meilleure relation possible entre, par exemple, la biomasse phytoplanctonique et des variables environnementales. La modélisation des FEA peut s'orienter d'abord vers l'analyse de tendances temporelles des concentrations en chl-a en lien ou non avec des facteurs environnementaux prévalant dans le milieu (par ex. : température de l'eau, P, N, ratio N:P). Les modèles les plus simples pour ces types de relations proposent une relation directe entre la chl-a et le P (par ex., Jones & Bachmann (1976)). D'autres modèles de régression multiple issus de données *in situ* ont été développés afin de mettre en évidence ces relations (Onderka, 2007; Liu *et al.*, 2013; Laanemets *et al.*, 2006; Hu *et al.*, 2009; Wang *et al.*, 2007, 2008, 2012) de même que des modèles issus de données satellitaires (par ex. : MERIS (Matthews, 2014; Binding *et al.*, 2011), Landsat et MODIS (Duan *et al.*, 2014)), permettant une modélisation spatio-temporelle plus complète du phénomène. Par exemple, les travaux de Hu *et al.* (2009) ont porté sur l'utilisation d'un modèle de régression linéaire de Poisson afin d'étudier la relation entre le nombre de cellules cyanobactériennes et des variables environnementales sur le lac Monduran, Australie. Matthews (2014) s'est intéressé aux séries temporelles des concentrations médianes en chl-a sur 50 lacs d'Afrique du Sud via des images MERIS de 2002 à 2012. À l'échelle du pixel, ils ont déterminé la valeur médiane de chl-a sur l'ensemble du lac pour chaque image donnée. Duan *et al.* (2014) se sont intéressés aux efflorescences de FEA sur le lac Taihu en Chine et ont exploré les tendances temporelles des dates d'initiation en comparant celles-ci avec des descripteurs environnementaux (température, phosphore total (PT), azote total (NT), et ratio NT:PT) via une régression linéaire multiple. Binding *et al.* (2011) se sont intéressés aux concentrations médianes de chl-a sur le lac des Bois au Canada/États-Unis via des images MERIS de 2003 à 2011. D'autres travaux portant sur le développement de modèles plus complexes d'analyse multivariée se sont avérés être très utiles dans la compréhension de la relation entre les facteurs environnementaux et l'influence qu'ils ont sur la concentration en chl-a ou la biomasse phytoplanctonique. Par exemple, une analyse en composantes principales (ACP) couplée à une régression orthogonale ont été utilisées sur les

concentrations en chl-a et des variables environnementales sur le lac Dianshan, Chine (Wang *et al.*, 2008) et sur le lac Taihu, Chine (Wang *et al.*, 2007). Van Tongeren *et al.* (1992) ont effectué une ACP et une analyse de redondance afin d'établir la relation entre les communautés phytoplanctoniques, les communautés zooplanctoniques, la température de l'eau et la profondeur sur les lacs de Loosdrecht, Pays-Bas.

Par ailleurs, les modèles basés-statistiques peuvent utiliser des bases de données historiques afin d'ajuster une loi de probabilité sur une variable aléatoire donnée en lien ou non avec des variables environnementales. L'application de la théorie des valeurs extrêmes pour modéliser l'émergence de FEA est d'ailleurs de plus en plus utilisée dans un contexte où les FEA sont considérées comme un événement extrême pouvant engendrer un risque. Ce risque concerne notamment la santé des usagers des plans d'eau, la santé des écosystèmes et les activités récréotouristiques et socioéconomique. L'analyse fréquentielle permet d'établir la relation existant entre des événements extrêmes qui peuvent engendrer un risque et leur probabilité au dépassement ou au non-dépassement. Wang *et al.* (2012) ont d'ailleurs utilisé la technique des copules afin d'analyser les liens entre la concentration en chl-a et des variables environnementales (température de l'eau, PT, NT, oxygène dissous) sur le lac Wulihu, Chine.

8.3 Méthodes d'analyse fréquentielle

L'application de la théorie des valeurs extrêmes pour la modélisation d'événements extrêmes est devenue un champ de recherche très actif de nos jours, notamment en raison des impacts socioéconomiques que ces événements engendrent. La théorie des valeurs extrêmes cherche à caractériser les distributions asymptotiques des *extrema* (ou *minima*) lorsque la taille de l'échantillon tend vers l'infini. L'analyse fréquentielle locale, largement utilisée en hydrologie, vise à ajuster une densité de probabilité à une série de données qui, généralement, varie temporellement. Cette base de données peut provenir d'une station d'échantillonnage, d'une station hydrométrique ou d'un milieu récepteur (lac, rivière). Lorsque la probabilité d'occurrence d'un événement extrême dépend de plus d'une variable ou que le risque engendré est attribué à plus d'un facteur, l'analyse fréquentielle multivariée peut être utilisée afin de limiter une surestimation potentielle du risque associé à l'événement (Yue & Rasmussen, 2002; De Michele *et al.*, 2005) et par conséquent entrainer des coûts supplémentaires liés à la gestion et au suivi de l'événement. Dans plusieurs domaines d'application tels qu'en hy-

drologie, l'analyse d'événements multivariés est grandement utilisée (Voir par ex. Adamson *et al.* (1999); Yue (2000)). Il peut arriver cependant que ces modèles ne soient pas bien adaptés à des événements extrêmes en raison, entre autre, des distributions multivariées qui imposent aux marges des distributions d'être identiques et qui sont limitées dans les types de distributions exportables. En effet, un nombre limité de distributions usuelles peut être utilisé (par ex: Normal, Gamma) et cet aspect s'avère plus problématique dans des domaines d'application impliquant des événements extrêmes qui peuvent montrer des lois marginales dissymétriques et des queues de distribution lourdes. Les distributions multivariées sont également moins flexibles sur la structure de dépendance entre les variables. Le concept de copule (Sklar, 1959) permet la construction de lois multivariées qui ne souffrent pas des inconvénients mentionnés plus haut. La copule, utilisée fréquemment en modélisation de la dépendance, permet de contenir toute l'information à propos de la structure de dépendance de variables dans une fonction appelée la copule. Il s'agit d'un outil puissant et flexible qui n'impose pas de marges identiques et qui ne tient pas compte de l'effet du comportement des marges (c'est-à-dire des fonctions de répartitions des variables). Cette technique a été développée initialement dans le domaine des finances et a récemment été introduite en hydrologie (Genest *et al.*, 2007; Chebana & Ouarda, 2011; Gräler *et al.*, 2013).

Par ailleurs, l'analyse fréquentielle régionale consiste à regrouper les données d'une même région considérée homogène afin d'améliorer la précision de l'estimation des quantiles. Cette méthode suppose donc que les distributions de fréquence de la variable étudiée sont identiques pour chaque station à l'intérieur d'une région. La première étape consiste ainsi à définir les régions homogènes de la zone d'étude. Diverses techniques sont présentées dans la littérature afin d'identifier les régions homogènes, dont la région d'influence (Burn, 1990), l'analyse canonique des corrélations (Cavadias, 1990) et les L-moments (Guttman *et al.*, 1993). La deuxième étape consiste à ajuster une distribution régionale sur chaque site, de calculer les quantiles d'intérêt pour finalement estimer ceux-ci sur des sites non-jaugés.

8.4 Modélisation par apprentissage automatique

Les algorithmes d'apprentissage automatique peuvent être définis comme des algorithmes qui ont la capacité d'apprendre d'une expérience E à partir d'une classe de tâches T et d'une mesure de performance P , et qui peuvent améliorer automatiquement cette performance P avec l'expérience

E (Mitchell *et al.*, 1997). Plusieurs types de tâches T peuvent être résolues avec l'apprentissage automatique tels que les classifications, les régressions, la détection d'anomalies, l'imputation de données manquantes, le débruitement ou l'estimation de densités de probabilité. Évaluer l'habileté de l'algorithme d'apprentissage se traduit par l'estimation d'une mesure quantitative reflétant sa performance P, tels qu'une mesure de la précision du modèle lors de la prédiction de valeurs numériques ou du ratio d'erreurs lors d'une classification par exemple. De manière générale, ces algorithmes se catégorisent en algorithmes d'apprentissage supervisés et non-supervisés. L'apprentissage non-supervisé permet l'interprétation de données (par ex.: partitionnement de données) en se basant uniquement sur la structure des données d'entrées. L'apprentissage supervisé permet l'interprétation de données (par ex.: régressions, classifications) en se basant sur des données d'entrées et des données de sorties.

Les algorithmes d'apprentissage automatique traditionnels répondent à une vaste variété de problèmes mais deviennent limités lorsque la complexité du problème est élevée. Notons entre autres l'augmentation de la dimensionnalité des données (spatial ou temporel) ou la connectivité de points voisins. Les algorithmes d'apprentissage profond ont été développés à cet effet. Les approches souvent utilisées en industries sont entre autres les réseaux de neurones à propagation avant et les réseaux de neurones convolutifs. L'utilisation d'autres approches plus avancées se poursuit présentement dans la communauté scientifique, où la problématique centrale est encore une fois la dimensionnalité élevée des variables aléatoires qui sont modélisées.

Les thèmes de recherche portant sur l'apprentissage automatique sont de plus en plus présents en sciences environnementales. Recknagel *et al.* (2002) ont utilisés des algorithmes génétiques d'apprentissage automatique sur des séries temporelles de caractéristiques physico-chimiques de l'eau afin de prédire l'abondance algale (chl-a $\mu g L^{-1}$ et cellules ml^{-1}) sur le lac Kasumigaura, Japon. Similairement, Cho *et al.* (2014) ont développé un modèle prédictif de l'abondance algale (chl-a $\mu g L^{-1}$) avec un réseau de neurones artificiel à partir de données physico-chimiques, climatiques et hydrologiques sur le lac Juam, Corée du Sud. Maier *et al.* (1998) ont évalué l'effet de huit variables physico-chimiques sur la durée et l'intensité d'événements algaux sur la rivière Murray, Australie à l'aide de réseaux de neurones artificiels. Ces modèles intègrent comme données d'entrées des séries temporelles de données de qualité de l'eau dans un objectif de prédiction à court terme (jusqu'à 7 jours) et moyen terme (4 semaines). Plusieurs autres études portent également sur le sujet (French & Recknagel, 1970; Karul *et al.*, 2000; Lee *et al.*, 2003; Muttill & Chau, 2006; Oh *et al.*, 2007;

Guallar *et al.*, 2016). Entre autre, French & Recknagel (1970) ont utilisé un réseaux de neurones avec propagation arrière afin de prédire l'apparition de sept types d'algues connaissant 9 variables environnementales dans le réservoir Saldenbach, Allemagne. Recknagel *et al.* (1997) ont également développé des réseaux de neurones reliant des espèces d'algues à des variables causales sur divers lacs au Japon, en Finlande et en Australie. Maier *et al.* (1998) ont utilisé des réseaux de neurones avec propagation arrière afin de prédire l'incidence et l'amplitude des cyanobactéries *Anabaena* dans la rivière Murray, Australie.

8.5 Sommaire

Les modèles-basés processus ont l'avantage de s'adapter à des changements majeurs dans le système à modéliser contrairement aux méthodes statistiques qui ne peuvent être utilisées à l'extérieur des marges de données empiriques qui ont été utilisées dans le développement du modèle (processus d'extrapolation). Par contre, puisqu'ils sont basés sur des relations physiques fondamentales, les processus physiques ou physico-chimiques sous-tendant une variable étudiée doivent être entièrement compris et de manière précise afin d'être décrits mathématiquement, ce qui n'est souvent pas le cas en raison de la complexité du phénomène naturel à l'étude. De ce fait, les modèles-basés processus représentent souvent une approximation grossière de la réalité d'un phénomène naturel (Hipel, 1985), considérant la quantité parfois importante de paramètres et où l'incertitude qui y est associée n'est souvent pas considérée lors de prédictions. De manière générale, les différentes études ont porté sur des analyses de régression de concentrations en chl-a (ou de la biomasse algale) d'organismes phytoplanctoniques et/ou de cyanobactéries afin de limiter la contribution de certains éléments perturbateurs tels que le P, le N, la température ou la turbidité. À l'exception des approches d'ajustement de densités de probabilité, les méthodes-basées processus et statistiques ne permettent pas de prédire la fréquence, l'intensité ou la date d'initiation des FEA avec précision, et encore moins la probabilité d'occurrence du phénomène basée sur les variables environnementales prévalant sur le lac et le bassin versant. Les méthodes fréquentielles sont utilisées de manière locale, c'est-à-dire que l'ajustement d'une loi de probabilité est effectuée sur des données issues d'un milieu (lac, rivière, etc.), d'une station météorologique ou d'une station hydrométrique (dimensionnalité temporelle).

Par ailleurs, l'utilisation de l'apprentissage automatique et de l'apprentissage profond est de plus en plus présente dans le domaine des sciences environnementales. Cependant, la complexité des processus spatio-temporels à modéliser dans le cadre d'une étude de l'aléa structurel des FEA est difficilement explicable avec l'utilisation des modèles présentement disponibles. D'une part, le développement de modèles à données qualitatives a progressé plus rapidement que la modélisation de données quantitatives par le biais de modèles d'intelligence artificielle. La lourdeur de temps de calcul et du volume de données à acquérir expliquent en partie la raison pour laquelle ces modèles tirent de l'arrière. Également, peu de modèles d'intelligence artificielle ne permettent l'estimation d'une densité de probabilité, et encore moins d'une densité de probabilité conditionnelle. Au meilleur de notre connaissance, un modèle d'estimation d'une densité de probabilité conditionnelle est disponible (Pospisil & Lee, 2018), mais il n'était pas possible de le tester sous le système d'exploitation Windows.

Chapitre 9

Approche statistique d'analyse fréquentielle régionale

Ce chapitre présente les étapes méthodologiques au développement et l'application d'un modèle fréquentiel régional non-stationnaire permettant l'ajustement d'une densité de probabilité conditionnelle à une base de données spatio-temporelles.

9.1 Introduction

La complexité de l'aléa structurel des FEA de même que la méconnaissance de certains processus physico-chimiques sous-tendant leur apparition restreignent les modèles basés-processus à modéliser adéquatement et avec précision ce phénomène. La modélisation de l'aléa structurel des FEA nécessite l'utilisation de modèles-basés statistiques, et plus particulièrement des modèles basés-statistiques fréquentielles dans un contexte de meilleure gestion du risque associé à l'émergence des FEA. Par ailleurs, la hausse marquée en fréquence et en intensité des FEA dans les écosystèmes d'eau douce porte l'empreinte du réchauffement global de la planète et des changements dans le régime hydrologique. L'impact des changements climatiques sur l'occurrence de ce phénomène demeure cependant difficile à quantifier par les méthodes de suivi d'échantillonnage standards, restreintes à un cadre spatio-temporel ponctuel. En effet, l'étude des changements climatiques sur l'occurrence des FEA nécessite l'acquisition d'une quantité considérable de données temporelles afin d'assurer

une analyse précise de leur évolution dans le temps. Une analyse fréquentielle locale (c.-à-d. sur un lac donné) de l'occurrence des FEA nécessite minimalement l'obtention d'une trentaine d'années de données disponibles afin d'assurer une modélisation juste du phénomène. Peu de bases de données permettent l'étude adéquate de l'impact des changements climatiques ou de l'impact de facteurs environnementaux sur l'apparition des FEA en raison de la difficulté à établir une base de données complète dans le temps et dans l'espace, d'où l'intérêt de développer un modèle régional permettant de décrire l'aléa structurel des FEA.

9.2 Description des données

La base de données comprend, comme variable dépendante, la phénologie des FEA dont la fréquence d'apparition et l'intensité. À titre d'exemple, des observations des fréquences d'apparition de FEA d'algues sur un lac donné et pour une année donnée sont disponibles pour un ensemble de 580 lacs et 17 années. De même, les descripteurs environnementaux sont disponibles pour l'ensemble de ces lacs et de ces années. En tout, 9860 observations (580 lacs pour 17 ans d'étude) sont répertoriées en entrée au modèle (figure 9.1).

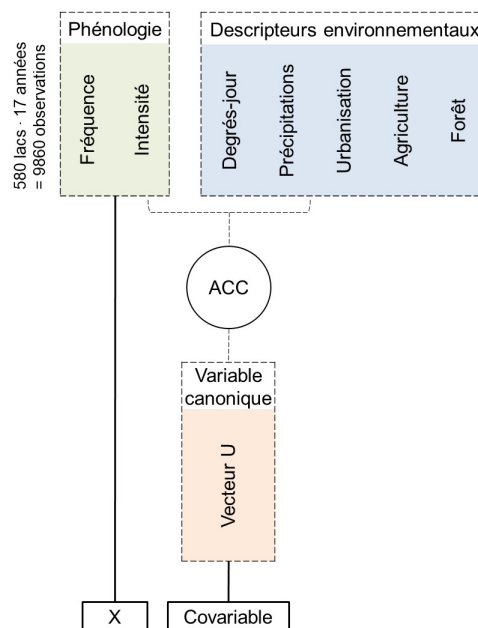


Figure 9.1 – Présentation schématique des données d'entrée (x et covariable) dans le modèle. ACC signifie l'analyse canonique des corrélations.

La base de données utilisée dans le cadre de cette étude est spatio-temporellement discrète (Diggle, 2013). Les FEA se produisent à des endroits spécifiques (lacs) sur la zone d'étude, et les variables phénologiques sont calculées annuellement et ainsi générées à des moments spécifiques dans le temps. Une hypothèse sur les données est énoncée:

Hypothèse 9.1. *L'intensité (densité de points par unité de surface) du processus est spatio-temporellement séparable.*

Puisque les données temporelles s'échelonnent de 2000 à 2016 pour chaque lac à l'étude, la composante temporelle est constante spatialement et il y a ainsi *a priori* une indépendance de premier ordre entre les composantes temporelles et spatiales. Un processus spatio-temporel est séparable de premier ordre si son intensité $\lambda(s, t)$ peut être factorisé par

$$\lambda(s, t) = m(s)\mu(t) \text{ pour tout } (s, t) \in SXT \quad (9.1)$$

On suppose également une séparabilité de deuxième ordre. Pour une région spatiale S prédéfinie et une région temporelle T . Un processus spatio-temporel stationnaire est séparable de deuxième ordre si la densité de covariance $\gamma(u, v) = \lambda_2(u, v) - \lambda^2$ peut être factorisé par

$$\gamma(u, v) = \gamma_s(u)\gamma_t(v) \quad (9.2)$$

Nous assumons ainsi une séparabilité spatio-temporelle de $\lambda(s, t)$. Dans ce qui suit, l'analyse de l'intensité est traitée séparément de l'espace et du temps.

9.3 Description du modèle

Cette approche est basée sur l'idée que la fréquence d'apparition des FEA x (ou tout autre variable phénologique) est une variable aléatoire pouvant être représentée par une distribution statistique $f(x|\varphi)$, où φ correspond aux paramètres de la distribution. Un premier postulat est d'abord émis:

Postulat 9.1. *L'aléa spatio-temporel de l'occurrence des FEA porte l'empreinte de descripteurs environnementaux, qui eux, varient dans le temps et l'espace.*

L'étude liée à la section précédente de cette thèse a d'ailleurs permis de caractériser les liens potentiels entre la phénologie des FEA et un ensemble de 25 variables environnementales (descripteurs physiographiques, climatiques et morphologiques). Des variables environnementales sont ressorties de cette étude, démontrant ainsi la présence de liens forts entre l'aléa structurel des FEA et des descripteurs environnementaux. Les descripteurs ressortant de cette étude étaient les degrés-jour, l'occupation du territoire, la superficie du lac et, à moindre mesure, la précipitation. Nous postulons ensuite que :

Postulat 9.2. *L'ajustement d'une densité de probabilité conditionnelle à des données de dimensionnalité élevée est possible si celles-ci respectent les hypothèses d'ajustement.*

L'ajustement d'une loi de probabilité dans un cadre stationnaire à un échantillon nécessite la vérification d'hypothèses d'ajustement. Le caractère aléatoire des données doit d'abord être considéré. L'ajustement de données à une densité de probabilité donnée suppose que les données sont issues de la même population et qu'elles ont la même probabilité d'être prélevées. De plus, l'indépendance des données nécessite que l'occurrence d'une observation ne soit pas affectée par les autres observations. Enfin, l'ajustement d'une distribution dans un cadre stationnaire nécessite que les données soient homogènes. Rappelons que dans un cadre d'analyse spatio-temporelle, un processus dit homogène requiert que les observations proviennent d'une structure spatio-temporelle totalement aléatoire (SSTTA):

1. Le nombre d'événements $Y(S \times T)$ dans une région spatiale donnée S et une région temporelle donnée T suit une distribution de Poisson avec une moyenne $\lambda|S||T|$, où λ est l'intensité ou le nombre moyen d'événements par unité de surface et $|\cdot|$ fait référence à l'aire (2D) ou à la longueur (1D).
2. Pour $Y(S \times T) = n$ donné, le nombre n d'événements dans une région $S \times T$ donnée forment un échantillon indépendants et identiquement distribués.

La première propriété implique que le nombre moyen d'événements est constant à l'intérieur de la région. La deuxième propriété implique que l'occurrence d'un événement n'affecte pas la probabilité d'occurrence d'un autre événement, et qu'il n'y a donc pas d'interaction entre les événements d'un même voisinage. La deuxième propriété implique également que les données sont issues de la même population et qu'elles ont la même probabilité d'être prélevées. L'indépendance et le caractère aléatoire des données sont présentées plus bas. Ces deux propriétés vérifiées impliquent que le processus est **homogène**. Lorsque seule la deuxième propriété est vérifiée, le processus est dit **inhomogène**. Plusieurs approches sont utilisées afin de gérer l'inhomogénéité (Illian *et al.*, 2008) de données, ce qui est souvent le cas dans le cadre d'études épidémiologiques. Une série de données est réputée hétérogène lorsqu'elle provient d'un phénomène dont les caractéristiques évoluent dans le temps ou dans l'espace (variations climatiques, variations physiographiques, changement à l'équipement de mesure). Ainsi, le cadre stationnaire nécessite que les propriétés statistiques de la loi de probabilité qui régit le phénomène (moyenne, variance ou autres moments) soient invariantes dans le temps et dans l'espace. Ajoutons que l'échantillon d'observations doit être suffisamment grand afin d'assurer une bonne représentativité du processus et que les valeurs aberrantes soient retirées, particulièrement lors d'études d'événements rares.

De ces postulats, trois hypothèses constituent la base du développement de cette méthode. Les deux premières hypothèses ont été présentées dans le chapitre 1 (hypothèses 1.1 et 1.2), où la première énonce qu'*il existe une variabilité spatio-temporelle dans la phénologie des FEA qui peut être expliquée par le biais de descripteurs environnementaux sous-tendant au phénomène*. Le modèle fréquentiel régional proposé sera non-stationnaire puisque les propriétés statistiques de la loi de probabilité varient dans le temps et dans l'espace. La variabilité spatio-temporelle de la phénologie des FEA a été vérifiée dans le chapitre précédent. Cette non-stationnarité sera modélisée en reliant les propriétés statistiques (espérance, variance) de la densité de probabilité aux descripteurs environnementaux (climat, physiographie) qui évoluent dans le temps et dans l'espace.

L'hypothèse 1.2 énonce qu'*une modélisation fréquentielle régionale permettrait d'améliorer la précision et l'exactitude des estimations comparativement à une modélisation fréquentielle locale*. Cette hypothèse s'appuie majoritairement sur l'apport en volume de données qui permettrait une meilleure représentativité du phénomène comparativement à une modélisation locale. Bien que la précision des modèles (régional et local) puisse être évaluée et comparée, l'exactitude demeure cependant plus difficile à évaluer quantitativement. Elle sera cependant évaluée à titre indicatif seulement en

comparant les médianes des fréquences d'apparition obtenues pour chaque modèle (régional et local) avec la fréquence d'apparition moyenne observée. Nous supposons finalement qu'une potentielle non-stationnarité des événements algaux soit linéaire. Le choix de cette forme fonctionnelle sera appuyé par une visualisation des moyennes et des coefficients de variation des fréquences d'apparition par rapport à la covariable.

Hypothèse 9.2. *Il existe une corrélation linéaire entre la moyenne et le coefficient de variation des distributions des variables phénologiques et les covariables environnementales.*

9.4 Définition de la densité de probabilité conditionnelle

Nous supposons que les variables à l'étude peuvent être représentées par les distributions Lognormale et Gamma. Les fonctions de densité de probabilité de ces distributions sont respectivement:

$$f_{LN}(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} \quad (9.3)$$

$$f_G(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \quad (9.4)$$

Où x correspond à la variable phénologique. Le choix de ces fonctions est appuyé par les quelque études portant sur la modélisation de la chl-a qui se sont basées sur ces distributions (par ex.: Wang *et al.* (2012)) et par leur simplicité mathématique qui facilite la convergence numérique lorsque les degrés de liberté sont élevés. Connaissant la covariable C , le modèle définit la relation entre l'espérance $E(x|C)$ et la variance $Var(x|C)$ de la fréquence d'apparition comme suit:

$$E(x|c) = g(C) \quad (9.5)$$

$$Var(x|c) = h(C) \quad (9.6)$$

Où $g(C)$ et $h(C)$ sont des fonctions à définir. Deux formes fonctionnelles sont testées pour chacune de ces fonctions : le cas constant et le cas linéaire. L'inspection visuelle de la moyenne et la variance des fréquences d'apparition en fonction de C appuie ce choix méthodologique dans les formes

fonctionnelles testées.

$$g(C) = a + bC \quad (9.7)$$

$$h(C) = c + dC \quad (9.8)$$

Où $p = [a, b, c, d]$ sont les paramètres des formes fonctionnelles. Le cas constant (stationnarité de la moyenne) correspond à $b = 0$ et l'autre cas constant (stationnarité de la variance) correspond à $d = 0$. Les formes fonctionnelles sélectionnées dépendent de l'analyse exploratoire des données qui permet d'explorer le type de relation existant entre une variable donnée et la covariable. Les relations entre les paramètres de la distribution et la moyenne et variance de la distribution sont utilisées afin d'exprimer les paramètres des distributions testées. À titre d'exemple, la moyenne $E(x)$ et la variance $Var(x)$ de la distribution lognormale sont :

$$E(x) = e^{\mu + \sigma^2/2} \quad (9.9)$$

$$Var(x) = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2} \quad (9.10)$$

En isolant numériquement μ et σ , l'expression des paramètres de la distribution lognormale est :

$$\sigma^2 = \ln\left(1 + \frac{Var(x)}{E(x)^2}\right) \quad (9.11)$$

$$\mu = \ln(E(x)) - \frac{\sigma^2}{2} \quad (9.12)$$

$$L(\theta) = \prod_{i=1}^n f(x_i, \theta) \quad (9.13)$$

Le maximum de la fonction de vraisemblance s'obtient en résolvant le système de k équations à k inconnues suivant:

$$\frac{\partial L(\theta)}{\partial \theta_j} = 0, \quad j = 1, \dots, k \quad (9.14)$$

Notons par ailleurs que la qualité de l'estimation repose davantage sur le choix du modèle que sur le choix de la méthode d'ajustement utilisée. Le modèle fréquentiel le mieux approprié pour chaque variable sera déterminé par le critère d'information bayésienne (BIC) (Schwarz *et al.*, 1978), qui

recherche un compromis entre le nombre de paramètres et l'ajustement:

$$BIC = -2 \log(L) + 2k \log(N) \quad (9.15)$$

Où L est la vraisemblance, k le nombre de paramètres et N la taille de l'échantillon. À titre d'exemple, en exprimant les paramètres de la distribution en fonction de $E(x)$ et $Var(x)$ qui sont exprimés par des formes fonctionnelles (voir plus haut), la fonction de log-vraisemblance est définie comme suit :

$$\ln L(p|x_i, C_i) = \sum_{i=1}^n \ln f(x_i|p, C_i) \quad (9.16)$$

9.5 Évaluation du modèle régional

Le modèle régional sera comparé à un modèle local, où la sélection et l'ajustement d'une densité de probabilité seront également effectuées. La modélisation locale signifie que l'ajustement d'une densité de probabilité (conditionnelle ou non) sera appliquée sur un milieu récepteur donné, par exemple le lac Brome. Les périodes de retour 2 ans, 20 ans et 100 ans seront estimées pour chacun des modèles par le biais de deux méthodes de rééchantillonnage: le bootstrap non-paramétrique et le bootstrap paramétrique. La première méthode est non-paramétrique et permettra de générer l'incertitude des temps de retour issus de la variabilité des données d'entrée. La deuxième méthode est paramétrique et mettra plutôt l'accent sur l'incertitude dans l'estimation des paramètres de la densité de probabilité conditionnelle. La précision de chacun des modèles sera présentée à titre indicatif seulement, en comparant les périodes de retour 2 ans aux fréquences (ou intensités) moyennes observées sur un lac donné pour une année donnée.

9.6 Hypothèses clés

Un tel modèle contient quelques hypothèses clés incluant: 1) L'ensemble de la structure spatio-temporelle de la variable phénologique est capturée par la fonction de densité de probabilité conditionnelle. 2) Le champ résiduel du modèle est homoscédastique et indépendant dans le temps et dans l'espace. Cette hypothèse signifie qu'il faut une répartition spatio-temporelle équidistante avec

des valeurs manquantes occasionnelles qui seront traitées comme des données manquantes. La répartition de premier ordre et de deuxième ordre sera vérifiée. 3) Il y a absence d'auto-corrélation spatiale et temporelle. Cette hypothèse sera également vérifiée.

9.7 Sommaire

L'ajustement d'une densité de probabilité à un échantillon nécessite avant tout une analyse exploratoire des données, puis par une vérification (et validation) des hypothèses d'ajustement. Nous pouvons par la suite considérer un échantillon d'observations indépendantes et identiquement distribuées. Puisque l'analyse fréquentielle permet de produire la distribution des probabilités d'une variable donnée (par ex: fréquence, intensité), elle permet de connaître la probabilité d'apparition d'une FEA à une période donnée avec une fréquence ou une intensité donnée, permettant de mieux planifier les activités de surveillance et de gestion. Notons qu'une analyse fréquentielle multivariée ou par copule a été considérée. Cependant, l'analyse descriptive des données phénologiques présentée à la section précédente n'a pas permis de montrer une corrélation évidente entre les variables phénologiques. La fréquence d'apparition et la date de début de l'événement montrent la plus importante corrélation avec un coefficient de détermination à 0.436. Dans une moindre mesure, une analyse multivariée pourrait être effectuée entre ces deux variables.

Les différentes variables physiographiques, climatiques et anthropiques sont décrites via l'utilisation d'une covariable représentant la variable canonique U , issue d'une analyse canonique des corrélations. Les modèle final sélectionné intègrera cette covariable qui elle, décrit la non-stationnarité spatio-temporelle des variables phénologiques. La covariable ainsi intégrée au modèle définira les relations entre celle-ci, qui représente les descripteurs environnementaux, et les moments théoriques (ordre 1 (espérance), et centré d'ordre 2 (variance)) des variables phénologiques. Notons qu'étant donné la complexité du modèle, il est important d'établir un bon compromis entre le nombre de covariables utilisées et le nombre de cofonctions testées car celles-ci sont liées au nombre total de paramètres du modèle. En tout, 2 densités X 2 moments X 1 covariable X 2 formes fonctionnelles = 8 modèles seront ajustés pour chaque variable phénologique étudiée (fréquence et intensité).

Chapitre 10

Résultats

10.1 Analyse de la structure spatio-temporelle

Dans cette section, la répartition spatio-temporelle des événements (propriétés de premier ordre) et leur niveau d'interaction entre eux (propriétés de deuxième ordre) sont évalués. Rappelons que la base de données utilisée dans le cadre de cette étude est spatio-temporellement discrète (Diggle, 2013). De plus, la dimension temporelle est constante dans l'espace, c'est-à-dire que le nombre d'années d'étude pour chaque lac est le même (17 ans de données) à quelques exceptions près en raison de valeurs aberrantes qui ont été retirées de la base de données. Ainsi, nous supposons *a priori* que les dimensions spatiale et temporelle seront séparables de premier ordre. Notons que lorsqu'elles ne sont pas séparables, une analyse de la structure spatio-temporelle est complémentaire à une analyse de la structure spatiale et de la structure temporelle.

D'abord, une analyse de la répartition spatiale et temporelle de premier ordre est nécessaire à la validation de l'homoscédasticité des résidus qui nécessite que la distribution des données soit équidistante spatialement. Notons que puisque la dimension temporelle représente des valeurs annuelles de fréquences ou d'intensités, la répartition des données est équidistante temporellement et n'est donc pas présentée. Une visualisation de la localisation des lacs permet d'établir une première validation qualitative de la répartition des lacs à l'étude qui doit être issue d'un processus aléatoire. La figure 10.1 présente la répartition des centroides des lacs sur la région d'étude. La région d'étude représente l'ensemble des zones de gestion intégrée de l'eau par bassin versant. À l'exception de

la région gaspésienne, la répartition des centroides semble être issue d'un processus aléatoire. La

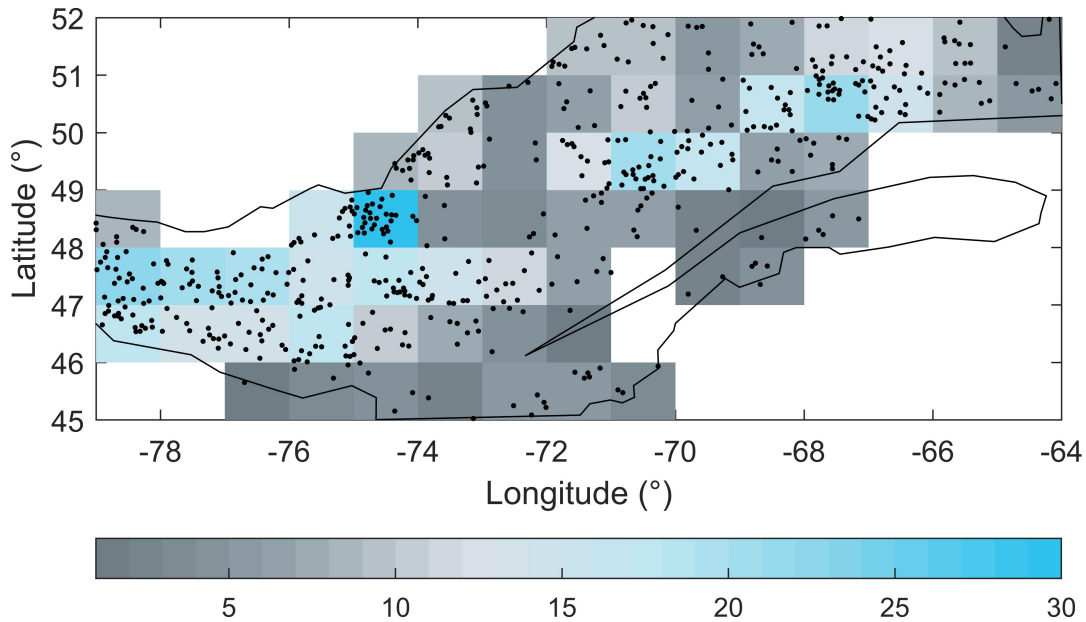


Figure 10.1 – Répartition spatiale des centroides des lacs (points) en fonction de la latitude et de la longitude sur la région d'étude (contour noir). Le gradient de couleur reflète le nombre de lacs présents dans une région donnée.

figure 10.2 présente la fonction K de Ripley standardisée, $L(h) = \sqrt{\frac{K(h)}{\pi}} - r$, où $L(h)$ est positif lorsqu'il y a une agrégation et négatif lorsqu'il y a une dispersion à une distance euclidienne de h . La fonction calculée sur les points sans pondération indique l'intensité d'agrégation associée aux seuls emplacements des lacs. Les résultats pondérés indiquent l'intensité d'agrégation ou de dispersion supplémentaire ajoutée par la prise en compte de la fréquence d'apparition des FEA. Les valeurs L observées sont plus grandes que 0 jusqu'à une distance d'environ 75 km. Cette distance est en partie expliquée par le polygone de la région d'étude utilisé dans le calcul de la fonction de Ripley. En observant la répartition des lacs à la figure 10.1, l'inclusion de la région gaspésienne induit une hausse dans l'estimation de la distance euclidienne entre des lacs localisés dans cette région et ceux localisés sur la Côte-Nord, la fonction utilisée pour estimer la fonction L ne prend pas en charge la non-convexité du polygone représentant la région d'étude. Afin de répondre en totalité à l'hypothèse

d'homoscédasticité des résidus, nous pourrions dans une telle situation sélectionner aléatoirement un sous-échantillon de lacs répondant au critère de répartition complètement aléatoire, mais dans le cadre de cette étude, nous considérons que la répartition des lacs répond au critère de répartition aléatoire en considérant la concavité de la région d'étude. Les répartitions spatiales d'ordre 2 de la

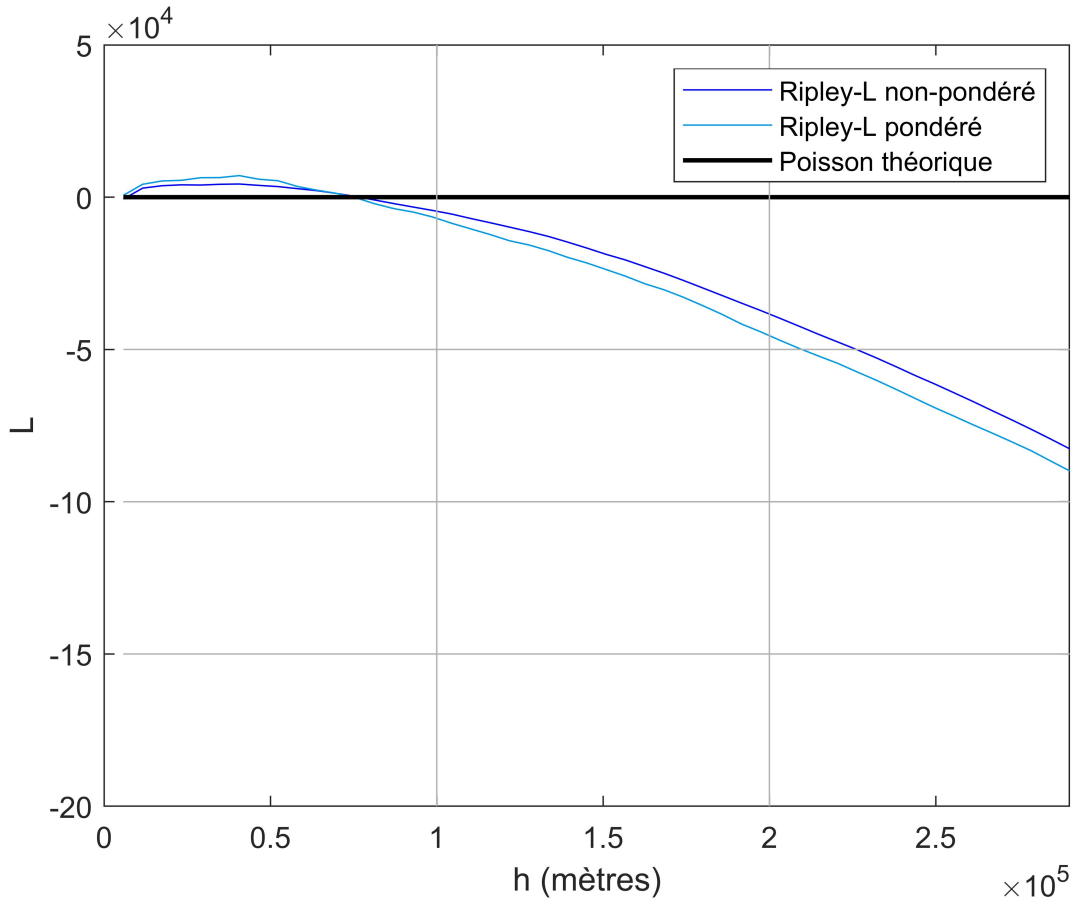


Figure 10.2 – Fonctions $L(h)$ pondérés et non-pondérés en fonction de la distance euclidienne h des lacs à l'étude.

fréquence et de l'intensité des FEA sont présentées aux figures 10.3 et 10.4 respectivement (étendue et début présentées à l'annexe E). On remarque une hétérogénéité évidente dans la fréquence d'apparition et l'intensité des événements sur la région d'étude. Cette hétérogénéité sera prise en compte dans le modèle en intégrant une non-stationnarité spatio-temporelle issue des descripteurs environnementaux qui varient dans le temps et dans l'espace.

L'auto-corrélation spatiale est vérifiée par un test d'hypothèse paramétrique issu de la statistique d'indice global de Moran qui mesure simultanément l'auto-corrélation selon la localisation des lacs et selon les valeurs d'entité (fréquence, intensité, etc.). Le test d'hypothèse est le suivant:

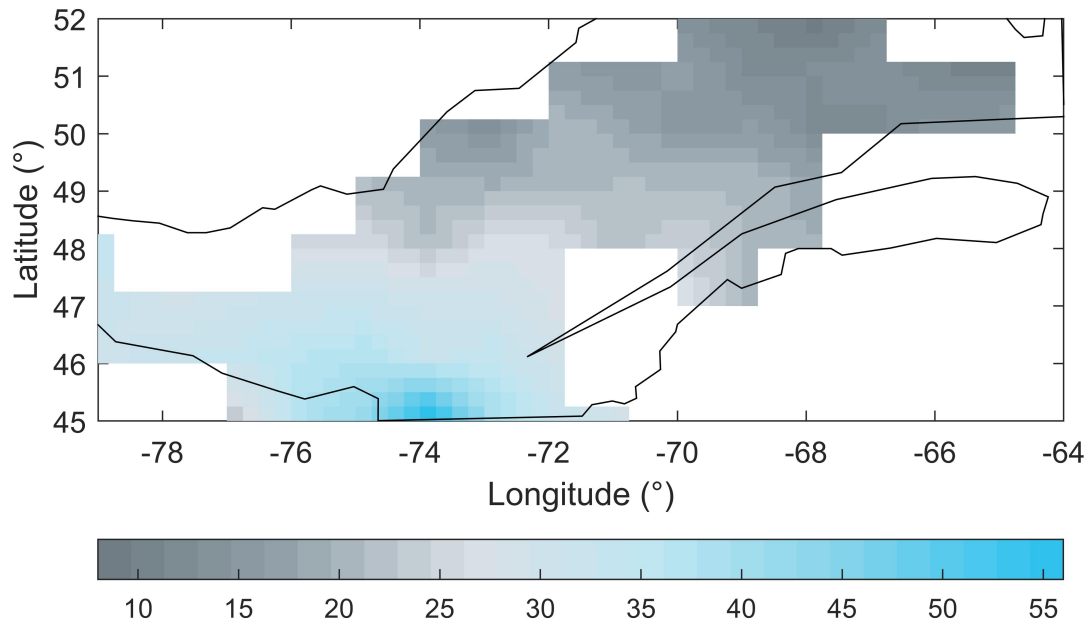


Figure 10.3 – Répartition spatiale de la fréquence des FEA sur la région d'étude.

H_0 : La distribution spatiale des valeurs d'entité est le résultat d'un processus aléatoire.

H_1 : (Score positif) La distribution spatiale fait l'objet d'une dispersion spatiale plus importante que si les processus sous-jacents étaient aléatoires. (Score négatif) La distribution spatiale des valeurs d'entité fait l'objet d'une agrégation spatiale plus importante que si les processus spatiaux sous-jacents étaient aléatoires.

Les résultats du test sur la fréquence d'apparition (figure 10.5) montrent que le rejet de l'hypothèse nulle n'est pas possible sous 95% de certitude. La fréquence d'apparition sur le territoire est ainsi issue d'un processus aléatoire à 95% de certitude ($p\text{-value}=0.055590$). Ce même test a été effectué sur les autres variables phénologiques (l'intensité à la figure 10.6, l'étendue et le début à l'annexe E), où les résultats obtenus montrent qu'il y a présence d'auto-corrélation pour les variables d'intensité et de date de début d'épisodes. Ainsi, un rééchantillonnage aléatoire des lacs a

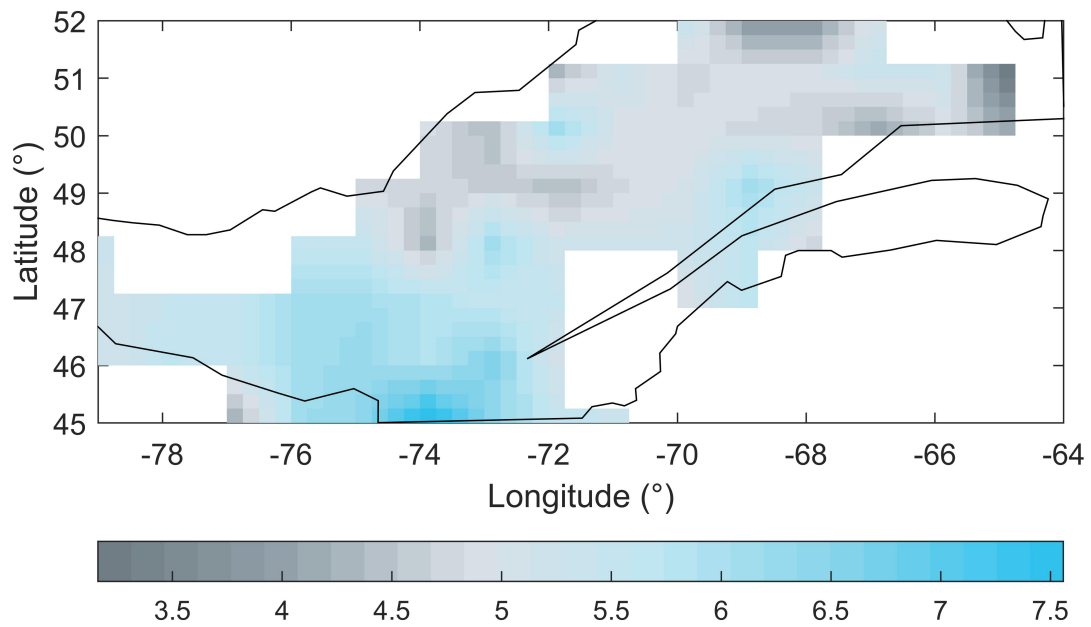


Figure 10.4 – Répartition spatiale de l'intensité des FEA sur la région d'étude.

été effectué pour ces deux variables afin d'assurer l'indépendance des données d'entrée au modèle régional fréquentiel.

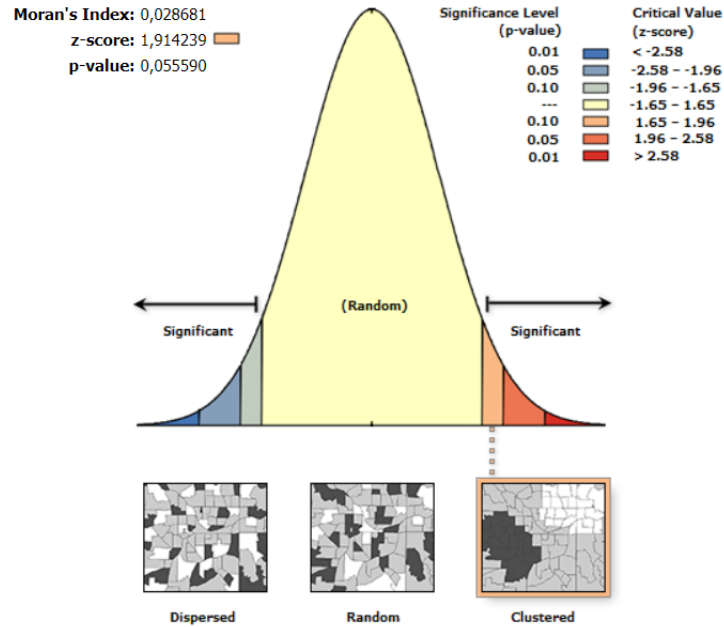


Figure 10.5 – Répartition spatiale de la fréquence des FEA sur la région d'étude.

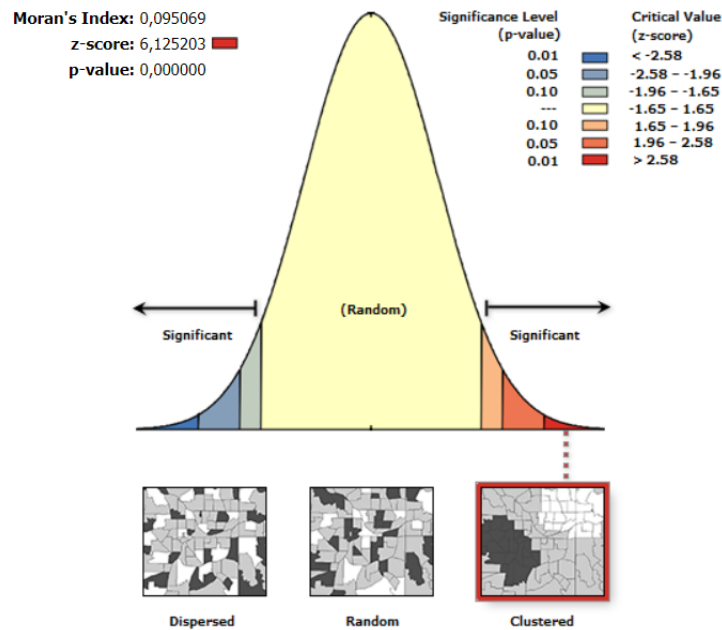


Figure 10.6 – Test d'autocorrélation spatiale sur l'intensité maximale.

10.2 Calage du modèle

Les descripteurs environnementaux en lien avec la phénologie des FEA ont été soulevés à la section précédente. L'ACC a d'ailleurs permis d'identifier les variables phénologiques les plus cor-

réelées, celles-ci étant la fréquence et l'intensité. Également, afin de réduire la dimensionnalité des données d'entrée et de permettre de réduire le nombre de paramètres d'ajustement liés aux descripteurs environnementaux, une analyse canonique des corrélations a été appliquée entre les variables phénologiques à l'étude et les variables environnementales liées à l'aléa structurel des FEA. Les variables environnementales sélectionnées ont montré une corrélation avec les variables phénologiques de $R > |0.3|$ et ne sont pas colinéaires entre elles:

Tableau 10.1 – Variables phénologiques et leur corrélation avec les variables environnementales.

Fréquence	Intensité
Degrés-jour (R=0.61)	Température moyenne (R=-0.47)
Précipitation annuelle (R=-0.30)	Précipitation annuelle (R=-0.43)
Urbanisation (R=0.30)	Périmètre du lac (R=0.54)
Forêt (R=-0.33)	Périmètre du bassin (R=0.46)
Agriculture (R=0.30)	

Le lecteur peut se référer à la section précédente pour une description détaillée de ces descripteurs. Notons que les résultats qui suivront issus du calage, des exemples d'application et des simulations sont présentés pour la fréquence d'apparition uniquement. Le vecteur U résultant de l'analyse canonique des corrélations sur ces données correspond à la covariable reflétant les caractéristiques physiographiques et climatiques sur le lac et son bassin versant. Afin de mieux visualiser la dépendance de cette covariable par rapport à la fréquence d'apparition, les valeurs de covariable ont été divisées en classes de valeurs. Les figures 10.7 et 10.8 présentent les moyennes et écart-types de la fréquence et l'intensité en fonction des valeurs de covariable pour ces différentes classes (étendue et début présentées à l'annexe E). La figure montre que la fréquence d'apparition diminue linéairement en fonction de la variable canonique, et que l'écart-type ne montre pas de tendance particulière. Quoique moins prononcée, la moyenne d'intensité semble montrer une tendance décroissante linéaire en fonction de la variable canonique, alors que l'écart-type ne montre pas de tendance particulière. Ces considérations supportent le choix des fonctions testées qui décrivent la relation entre la moyenne, la variance et la covariable.

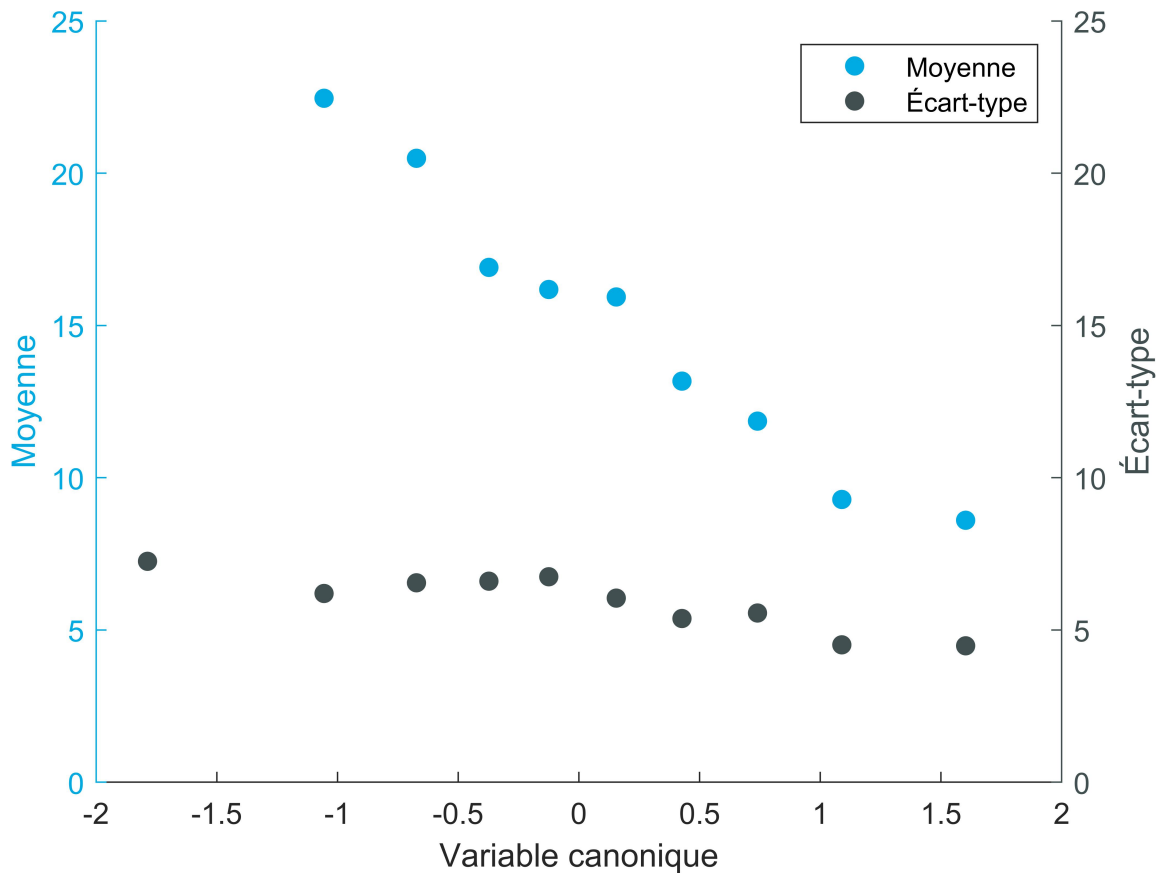


Figure 10.7 – Moyennes et écart-types de la variation des fréquences d’apparition de FEA pour chaque classe de valeurs de la covariable.

Au total, 8 modèles ont été ajustés sur les variables phénologiques de la fréquence d’apparition, de même pour l’intensité. Les tableaux 10.2 et 10.3 présentent les valeurs de log-vraisemblance et les critères d’information obtenus sur l’ajustement des fréquences d’apparition et de l’intensité à ces 8 modèles. Le critère d’information d’Akaike (AIC) et le critère d’information Bayésien (BIC) sont des mesures de la qualité d’un modèle statistique à ajuster des données. Les modèles présentant une fonction de la moyenne linéaire et une fonction de la variance constante correspondent aux modèles avec les critères d’information les plus faibles.

Le lac Brome, la Baie Missiquoi, le lac Aylmer et le lac Saint-Jean ont été sélectionnés à titre d’exemple comparatif des résultats obtenus avec un modèle calé localement sur chacun de ces milieux récepteurs. Les 8 modèles ont été ajustés pour chacun des lacs, en considérant une potentielle non-stationnarité temporelle. Notons que ces résultats sont présentés à titre indicatif seulement, puisque

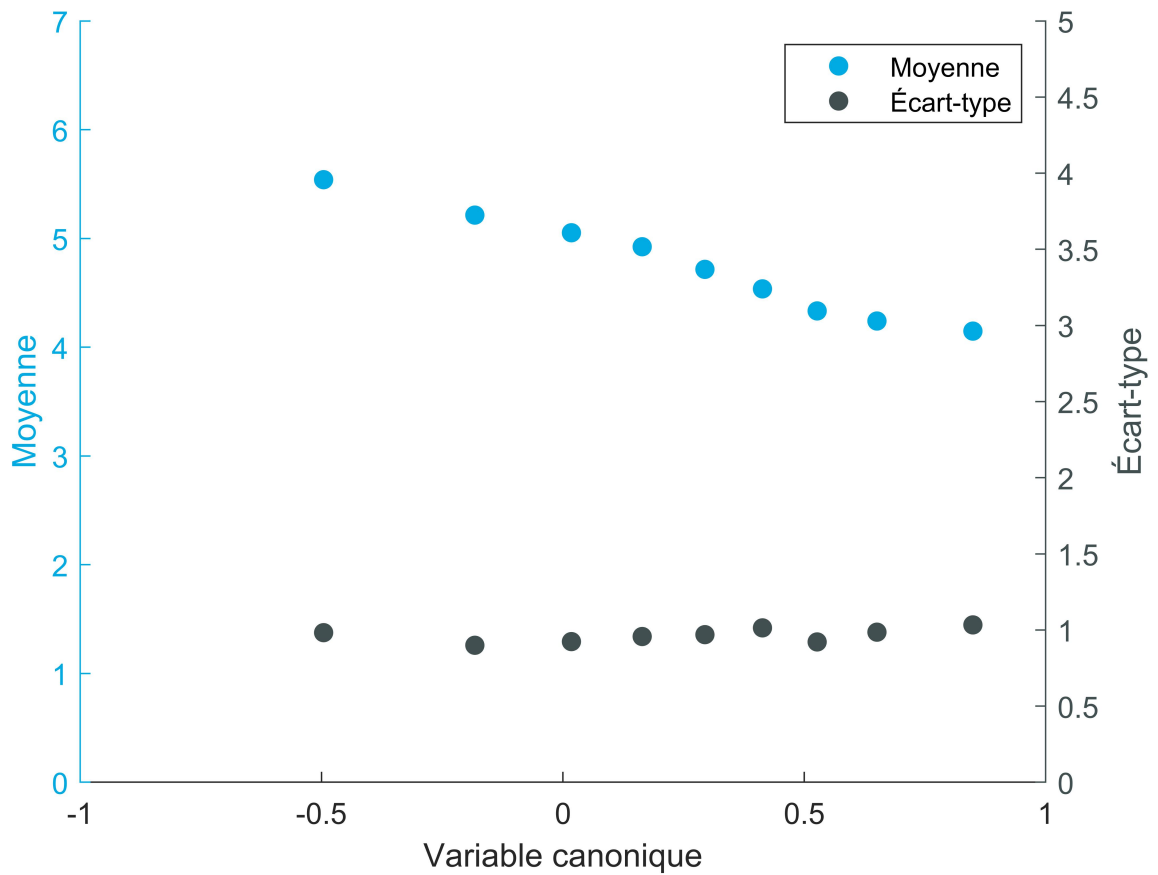


Figure 10.8 – Moyennes et écart-types de la variation d'intensité ($\ln[\text{mgL}^{-1}]$) de FEA pour chaque classe de valeurs de la covariable.

la taille d'échantillons utilisée est faible. Les résultats des ajustements sont présentés aux tableaux ci-dessous, où chaque modèle sélectionné est présenté en caractère gras.

Tableau 10.2 – Log-vraisemblances des modèles testés sur la fréquence de FEA en considérant une dépendance linéaire sur la covariable environnementale. Le modèle sélectionné est présenté en caractère gras.

Distribution	Fonction moyenne	Fonction variance	Paramètres	Log V	AIC	BIC
Log-Normal	Éq. 9.7, b=0	Éq. 9.8, d=0	2	-4.85E+04	9.70E+04	9.71E+04
		Éq. 9.8	3	-3.07E+04	6.15E+04	6.15E+04
	Éq. 9.7	Éq. 9.8, d=0	3	-2.80E+04	5.60E+04	5.61E+04
		Éq. 9.8	4	-2.80E+04	5.59E+04	5.60E+04
Gamma	Éq. 9.7, b=0	Éq. 9.8, d=0	2	-5.00E+04	1.00E+05	1.00E+05
		Éq. 9.8, d=0	3	-3.00E+04	6.00E+04	6.00E+04
	Éq. 9.7	Éq. 9.8, d=0	3	-2.81E+04	5.63E+04	5.63E+04
		Éq. 9.8	4	-2.81E+04	5.63E+04	5.63E+04

Tableau 10.3 – Log-vraisemblances des modèles testés sur l'intensité de FEA en considérant une dépendance linéaire sur la covariable environnementale. Le modèle sélectionné est présenté en caractère gras.

Distribution	Fonction moyenne	Fonction variance	Paramètres	Log V	AIC	BIC
Log-Normal	Constante	Constante	2	-2.69E+04	5.38E+04	5.38E+04
		Linéaire	3	-1.35E+04	2.69E+04	2.70E+04
	Linéaire	Constante	3	-1.19E+04	2.38E+04	2.39E+04
		Linéaire	4	-9.90E+03	1.98E+04	1.99E+04
Gamma	Constante	Constante	2	-2.52E+04	5.04E+04	5.04E+04
		Linéaire	3	-1.34E+04	2.68E+04	2.69E+04
	Linéaire	Constante	3	-1.20E+04	2.40E+04	2.40E+04
		Linéaire	4	-9.86E+03	1.97E+04	1.98E+04

Tableau 10.4 – Log-vraisemblances des modèles locaux testés sur la fréquence de FEA en considérant une dépendance linéaire dans le temps. Les modèles sélectionnés sont présentés en caractère gras.

Modèle local - Lac Brome						
Distribution	Fonction moyenne	Fonction variance	Paramètres	Log V	AIC	BIC
Log-Normal	Constante	Constante	2	-5.47E+01	1.13E+02	1.21E+02
		Linéaire	3	-5.49E+01	1.18E+02	1.33E+02
	Linéaire	Constante	3	-5.18E+01	1.10E+02	1.21E+02
		Linéaire	4	-5.20E+01	1.10E+02	1.21E+02
Gamma	Constante	Constante	2	-5.47E+01	1.13E+02	1.21E+02
		Linéaire	3	-5.66E+01	1.19E+02	1.30E+02
	Linéaire	Constante	3	-5.45E+01	1.15E+02	1.26E+02
		Linéaire	4	-5.65E+01	1.19E+02	1.30E+02

Modèle local - Baie Missiquoi						
Distribution	Fonction moyenne	Fonction variance	Paramètres	Log V	AIC	BIC
Log-Normal	Constante	Constante	2	-5.59E+01	1.16E+02	1.23E+02
		Linéaire	3	-5.59E+01	1.20E+02	1.34E+02
	Linéaire	Constante	3	-5.24E+01	1.11E+02	1.22E+02
		Linéaire	4	-5.22E+01	1.12E+02	1.27E+02
Gamma	Constante	Constante	2	-5.60E+01	1.16E+02	1.23E+02
		Linéaire	3	-5.59E+01	1.18E+02	1.29E+02
	Linéaire	Constante	3	-5.58E+01	1.18E+02	1.29E+02
		Linéaire	4	-5.55E+01	1.17E+02	1.28E+02

Modèle local - Lac Aylmer						
Distribution	Fonction moyenne	Fonction variance	Paramètres	Log V	AIC	BIC
Log-Normal	Constante	Constante	2	-5.20E+01	1.08E+02	1.15E+02
		Linéaire	3	-5.18E+01	1.12E+02	1.26E+02
	Linéaire	Constante	3	-5.19E+01	1.10E+02	1.21E+02
		Linéaire	4	-5.19E+01	1.12E+02	1.26E+02
Gamma	Constante	Constante	2	-5.25E+01	1.09E+02	1.16E+02
		Linéaire	3	-1.02E+02	2.10E+02	2.21E+02
	Linéaire	Constante	3	-5.07E+01	1.07E+02	1.18E+02
		Linéaire	4	-9.99E+01	2.06E+02	2.17E+02

Tableau 10.5 – Log-vraisemblances des modèles locaux testés sur l'intensité de FEA en considérant une dépendance linéaire dans le temps. Les modèles sélectionnés sont présentés en caractère gras.

Modèle local - Lac Brome						
Distribution	Fonction M	Fonction C	Paramètres	Log V	AIC	BIC
Log-Normal	Constante	Constante	2	-1.16E+02	2.36E+02	2.43E+02
		Linéaire	3	-1.16E+02	2.40E+02	2.54E+02
	Linéaire	Constante	3	-1.16E+02	2.38E+02	2.49E+02
		Linéaire	4	-1.16E+02	2.40E+02	2.54E+02
Gamma	Constante	Constante	2	-1.18E+02	2.41E+02	2.48E+02
		Linéaire	3	-1.18E+02	2.43E+02	2.54E+02
	Linéaire	Constante	3	-1.18E+02	2.43E+02	2.54E+02
		Linéaire	4	-1.18E+02	2.43E+02	2.54E+02

Modèle local - Baie Missiquoi						
Distribution	Fonction M	Fonction C	Paramètres	Log V	AIC	BIC
Log-Normal	Constante	Constante	2	-1.46E+02	2.95E+02	3.02E+02
		Linéaire	3	-1.46E+02	2.99E+02	3.14E+02
	Linéaire	Constante	3	-1.45E+02	2.95E+02	3.06E+02
		Linéaire	4	-1.45E+02	2.97E+02	3.12E+02
Gamma	Constante	Constante	2	-1.44E+02	2.92E+02	2.99E+02
		Linéaire	3	-1.44E+02	2.94E+02	3.05E+02
	Linéaire	Constante	3	-1.43E+02	2.92E+02	3.03E+02
		Linéaire	4	-1.43E+02	2.91E+02	3.02E+02

Modèle local - Lac Aylmer						
Distribution	Fonction M	Fonction C	Paramètres	Log V	AIC	BIC
Log-Normal	Constante	Constante	2	-1.21E+02	2.45E+02	2.53E+02
		Linéaire	3	-1.21E+02	2.49E+02	2.64E+02
	Linéaire	Constante	3	-1.21E+02	2.47E+02	2.58E+02
		Linéaire	4	-1.21E+02	2.49E+02	2.64E+02
Gamma	Constante	Constante	2	-1.21E+02	2.46E+02	2.53E+02
		Linéaire	3	-1.21E+02	2.48E+02	2.59E+02
	Linéaire	Constante	3	-1.21E+02	2.48E+02	2.59E+02
		Linéaire	4	-1.21E+02	2.48E+02	2.59E+02

10.3 Application du modèle

Le modèle régional ajusté sur la fréquence d'apparition des FEA a été comparé à des modèles locaux qui ont été ajustés par lacs. Quatre lacs ont été sélectionnés pour cette comparaison: le lac Brome, la baie Missiquoi, le lac Aylmer et le lac Saint-Jean. Ces lacs présentent des classes de niveaux trophiques différentes et des caractéristiques physiographiques et climatiques variées. Les figures 10.9, 10.10, 10.11 et 10.12 présentent les temps de retours aux 2 ans estimés par les 2 modèles. 1000 itérations de bootstrap non-paramétrique ont été appliquées afin de capter l'incertitude sur les données d'entrée aux modèles. Notons que les simulations via la méthode bootstrap paramétrique donnent des résultats similaires et ont ainsi été mises de côté par souci de concision. Les meilleurs modèles locaux ajustés sur les données présentent tous une non-stationnarité temporelle, d'où la tendance positive du temps de retour sur les figures. Le modèle régional capte beaucoup plus la variabilité spatio-temporelle dans les temps de retour. On constate une nette amélioration dans la précision de l'estimation par le modèle régional pour l'ensemble des lacs comparés. L'exactitude d'estimation est difficilement quantifiable, mais les temps de retours aux 2 ans ont été comparés aux fréquences moyennes qui ont été observés sur lac donné à une année donnée. Le tableau 10.6 présente des statistiques comparatives entre les deux modèles. Une définition détaillée des statistiques calculées est présentée ci-dessous:

1. Étendue: Moyenne annuelle des écarts entre le 95e percentile et le 5e percentile des simulations bootstrap de la fréquence d'apparition.
2. Écart-type: Moyenne annuelle de la racine carré de la variance des simulations bootstrap de la fréquence d'apparition.
3. Biais absolu moyen ou mesure d'asymétrie (à titre indicatif seulement): Moyenne annuelle des biais absolus entre la médiane des temps de retours estimée et la moyenne observée.
4. Biais absolu maximal ou mesure d'asymétrie (à titre indicatif seulement): Moyenne annuelle de la valeur maximale des biais absolus entre les simulations bootstrap des temps de retour estimée et la moyenne observée.
5. RMSE (à titre indicatif seulement): Écart quadratique moyen entre la médiane des temps de retour estimée et la moyenne observée.
6. RMSE maximal (à titre indicatif seulement): Écart quadratique moyen entre la médiane des temps de retour ayant le biais absolu le plus élevé et la moyenne observée.

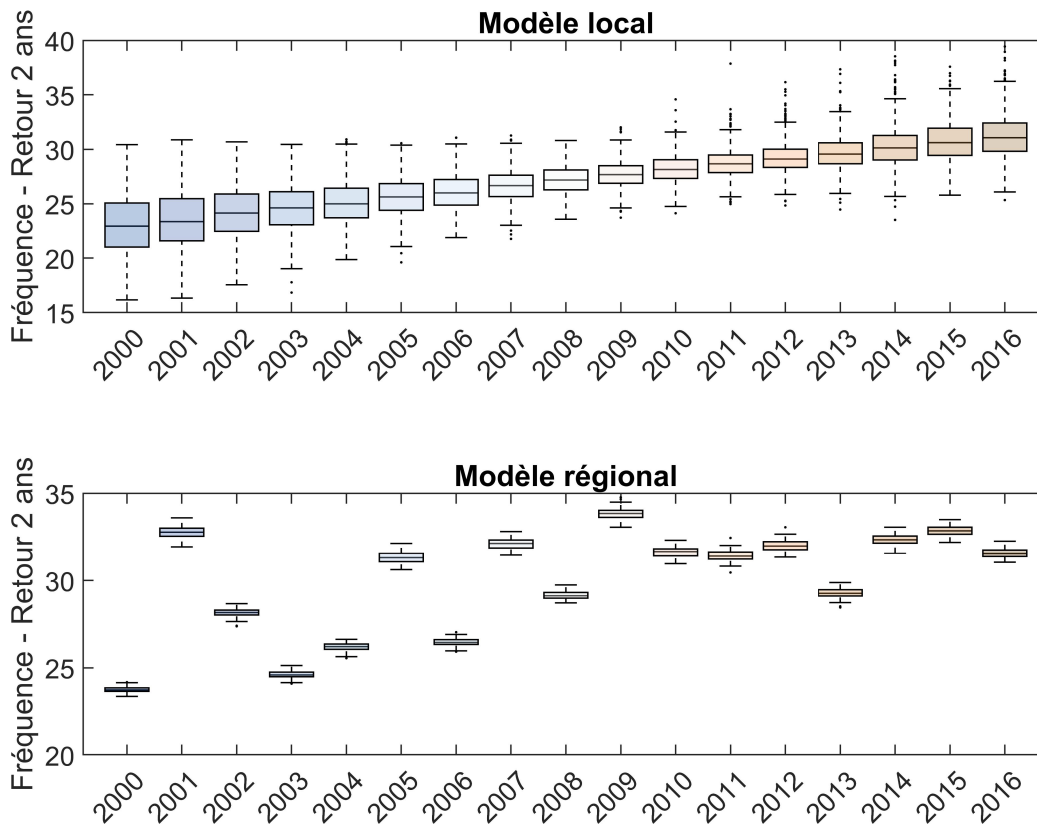


Figure 10.9 – Simulations bootstraps non-paramétriques des retours aux 2 ans des fréquences annuelles du lac Saint-Jean estimés par le modèle local (haut) et le modèle régional (bas).

Le modèle régional permet une estimation plus précise dans l'estimation, où l'étendue et l'écart-type sont moins élevés que les modèles locaux. Le modèle régional présente également des biais et RMSE un peu plus faibles que les modèles locaux, supposant ainsi que l'exactitude dans l'estimation semble bonne, mais une quantification de cette exactitude n'est pas possible. Un exemple plus détaillée de simulations des temps de retours 2 ans obtenus sur le lac Brome en 2000 est présenté à la figure 10.13. On constate encore un fois une meilleure exactitude dans l'estimation. Bien que la figure compare la valeur médiane à la valeur moyenne observée, on constate tout de même que l'estimation est potentiellement assez précise.

Des simulations de la fréquence d'apparition à des horizons futurs ont été calculées par le biais de scénarios climatiques futures fournies par l'organisme Ouranos (agr, version 4.2.3 du modèle climatique régional canadien). Cet organisme propose des projections climatiques régionales à haute résolution telles que la précipitation totale et la température moyenne sur l'ensemble de l'Amé-

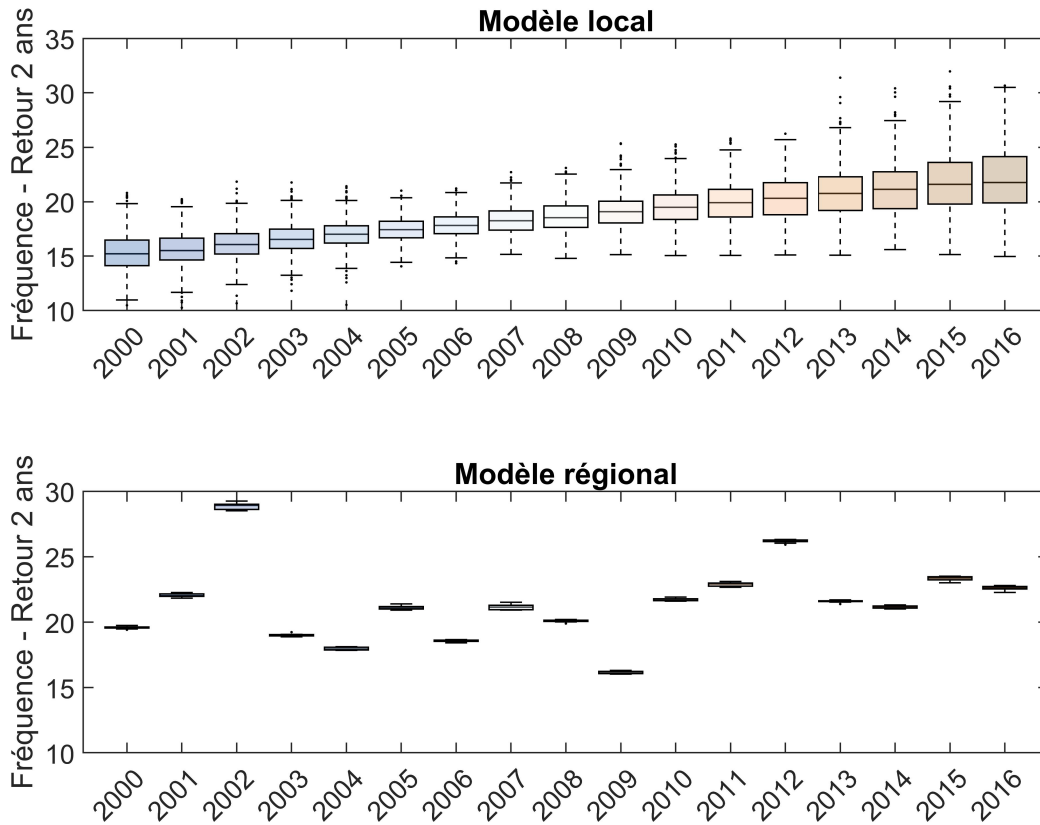


Figure 10.10 – Simulations bootstraps non-paramétriques des retours aux 2 ans des fréquences annuelles du lac Aylmer estimés par le modèle local (haut) et le modèle régional (bas).

rique du Nord en continu jusqu'en 2100. Le lac Brome a été sélectionné pour cette analyse. La tuile utilisée, qui représente une région de longueur approximative de 45km , couvre le lac Brome et son bassin versant (figure 10.14). Les degrés-jour cyanobactériens ont été calculés via les températures moyennes journalières simulées jusqu'à 2100. Notons que les degrés-jour cyanobactériens sont calculées à partir d'un seuil de croissance de la biomasse algale égal à 20°C . Les simulations considèrent ainsi une non-stationnarité temporelle des degrés-jour et une stationnarité temporelle des autres descripteurs environnementaux comme l'occupation du territoire. Notons cependant que des simulations intégrant à la fois une non-stationnarité du climat et de la physiographie seraient possible. La figure 10.15 présente les simulations des fréquences annuelles estimées jusqu'à 2100. On constate évidemment une hausse des fréquences annuelles à travers le temps. En période de référence, le lac Brome présente en moyenne 29 jours annuels d'événements algaux, qui sera de 40 jours annuels sur l'horizon 2050 (2041-2060) et de 48 jours annuels sur l'horizon 2090 (2081-2010). Ainsi,

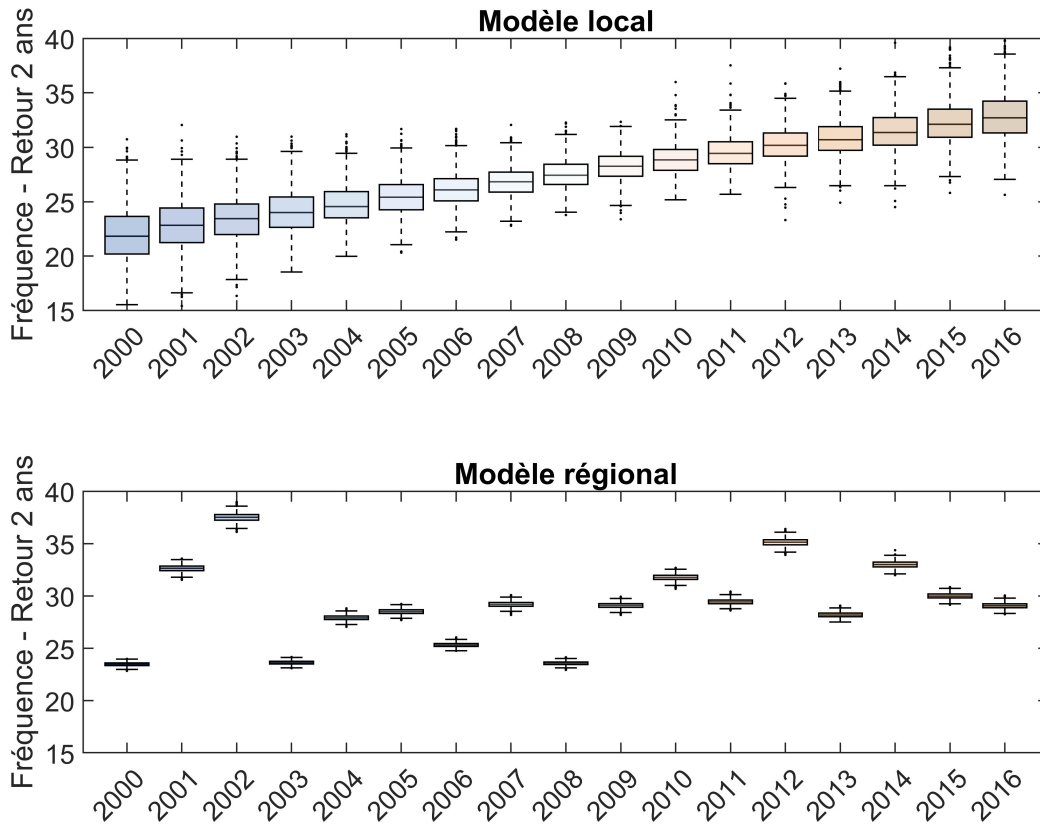


Figure 10.11 – Simulations bootstraps non-paramétriques des retours aux 2 ans des fréquences annuelles du lac Brome estimés par le modèle local (haut) et le modèle régional (bas).

le modèle prévoit une hausse de 65% des fréquences d'apparition de FEA sur le lac Brome d'ici l'horizon 2090. La relation linéaire entre les événements moyens et les degrés-jours (autres descripteurs environnementaux fixés à stationnaires) est présentée dans le graphique du bas de la figure 10.15. Une hausse annuelle d'environ 104 degrés-jour augmente de un jour le nombre d'événements algaux annuels. Cet exemple d'application du modèle est possible grâce à l'obtention des simulations du climat où leur localisation coïncide avec celle du lac Brome, un des lacs à l'étude. D'autres analyses seraient possibles avec l'obtention de simulations climatiques localisées sur d'autres régions de la zone d'étude.

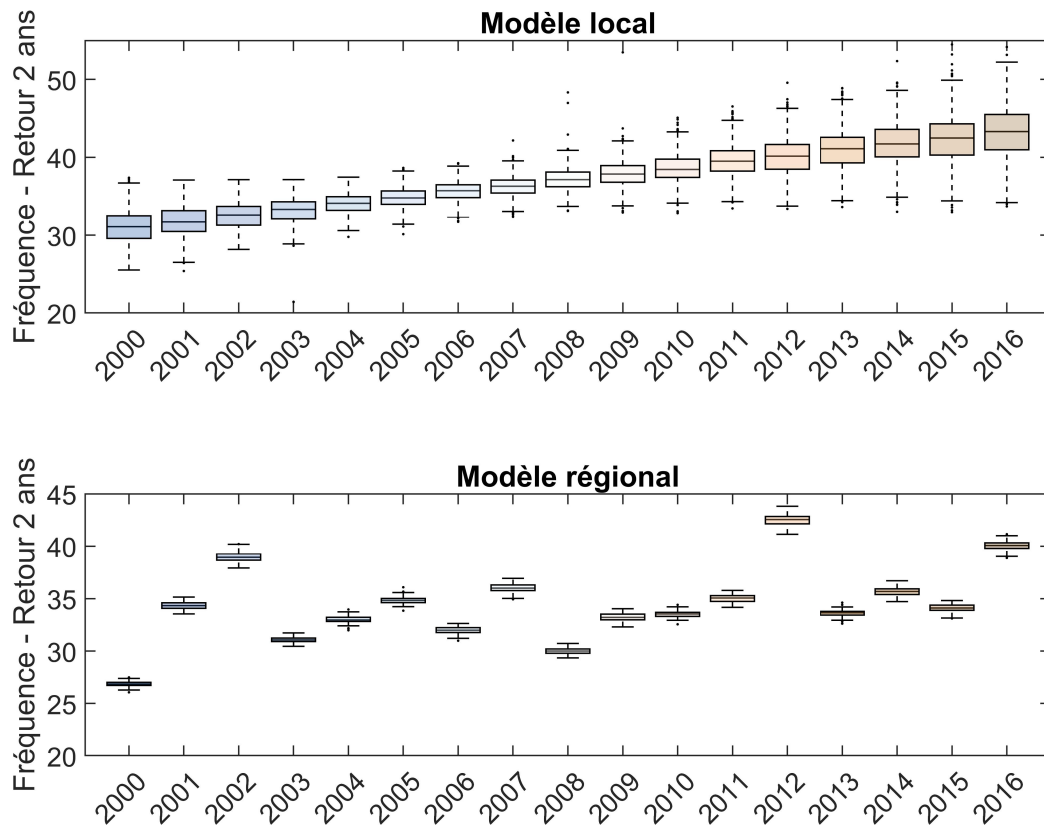


Figure 10.12 – Simulations bootstraps non-paramétriques des retours aux 2 ans des fréquences annuelles de la Baie Missisquoi estimés par le modèle local (haut) et le modèle régional (bas).

Tableau 10.6 – Statistiques comparatives des temps de retour 2 ans obtenus par le modèle régional et le modèle local. Les biais absolus et les RMSE sont présentés à titre indicatif seulement.

Lacs	Statistiques		Régional	Local
Missisquoi	Étendue		1.21	5.74
	Écart-type		0.36	1.80
	Biais absolu	Moyen	4.78	4.45
		Maximal	5.73	7.36
	RMSE	Moyen	5.59	5.74
		Maximal	6.47	8.54
Brome	Étendue		0.75	5.39
	Écart-type		0.24	1.68
	Biais	Moyen	3.14	4.21
		Maximal	3.54	6.80
	RMSE	Moyen	3.67	5.21
		Maximal	4.01	7.36
Aylmer	Étendue		0.40	5.48
	Écart-type		0.13	1.68
	Biais absolu	Moyen	2.97	4.06
		Maximal	3.17	6.93
	RMSE	Moyen	3.47	5.38
		Maximal	3.68	7.75
Saint-Jean	Étendue		0.87	5.35
	Écart-type		0.27	1.74
	Biais absolu	Moyen	2.98	3.78
		Maximal	3.66	6.43
	RMSE	Moyen	3.64	4.38
		Maximal	4.19	6.94

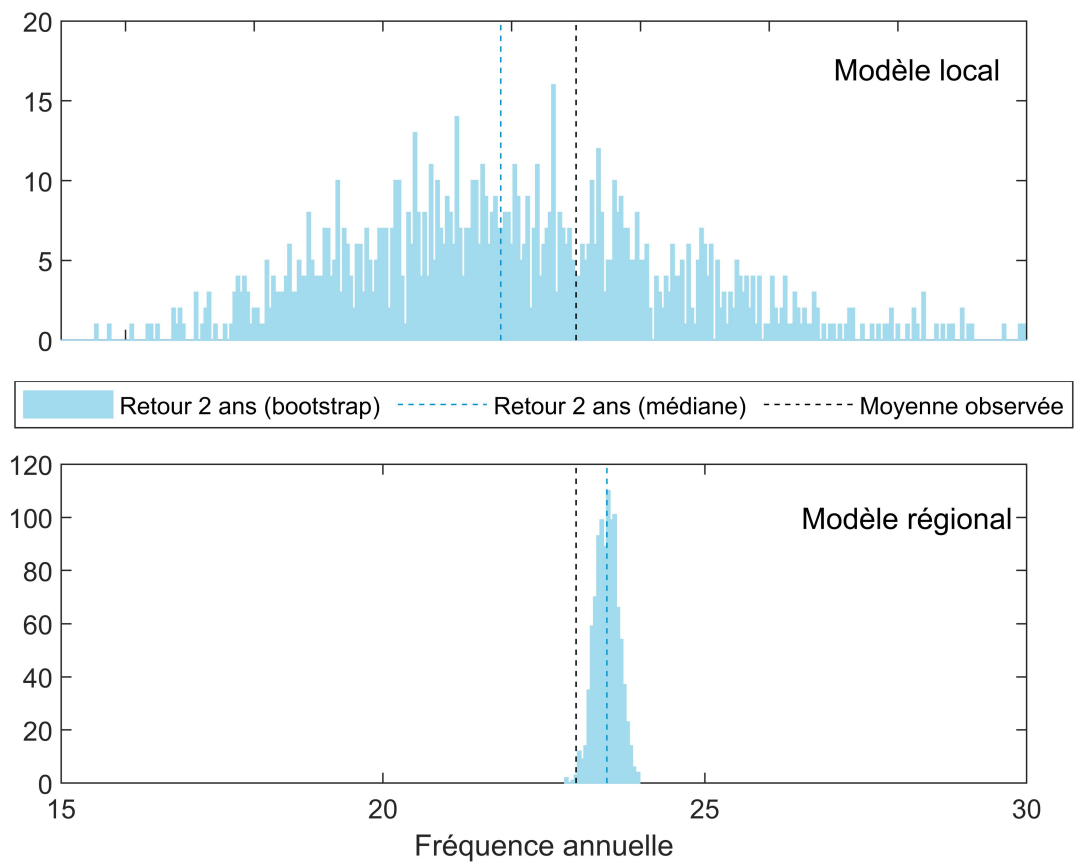


Figure 10.13 – Temps de retour 2 ans de la fréquence d’apparition annuelle de FEA sur le lac Brome en 2000 estimés par le modèle local (haut) et le modèle régional (bas).

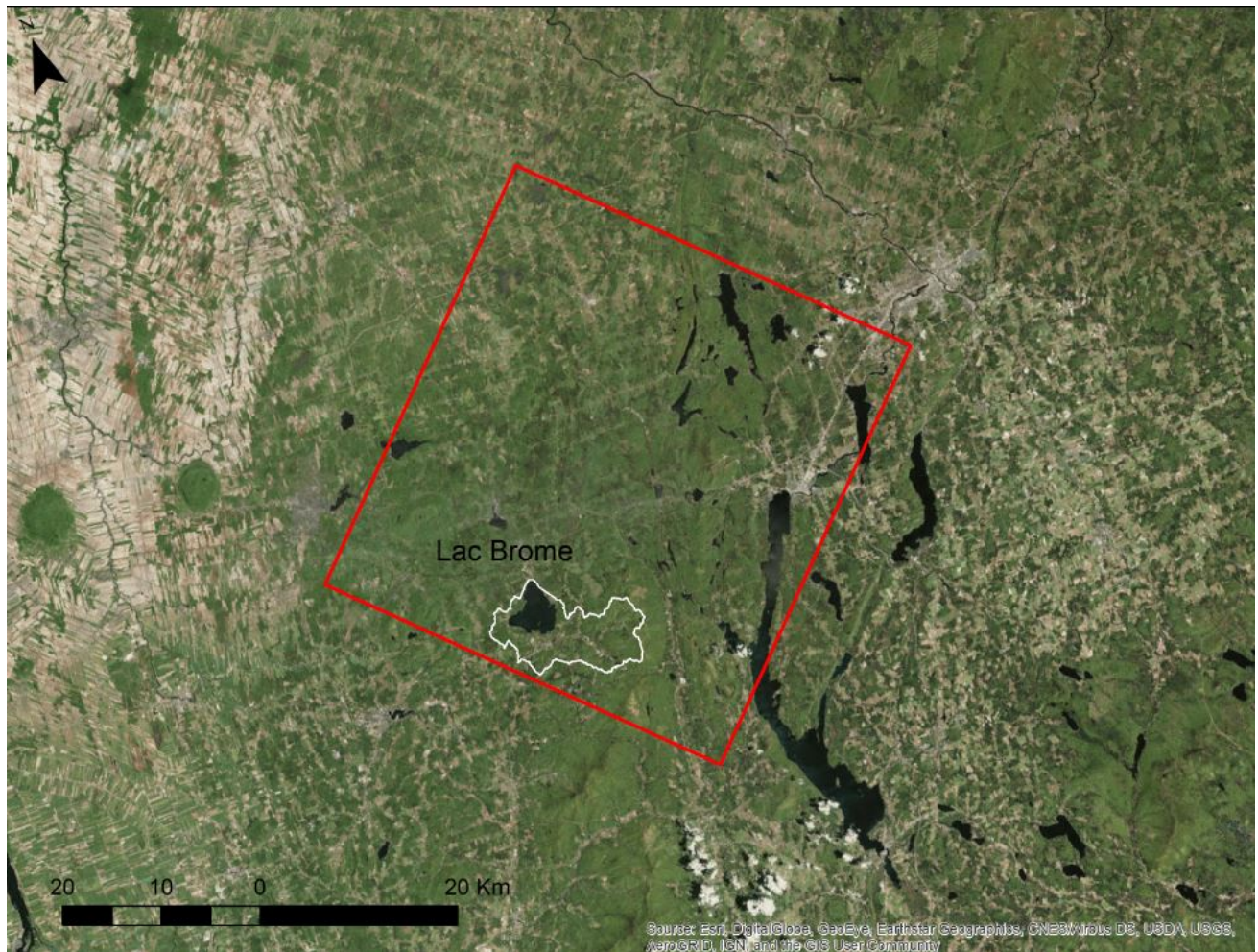


Figure 10.14 – Localisation de la tuile (rouge) d'où proviennent les simulations climatiques futures par Ouranos. Le lac Brome et son bassin versant (polygone blanc) sont inclus à l'intérieur de cette région.

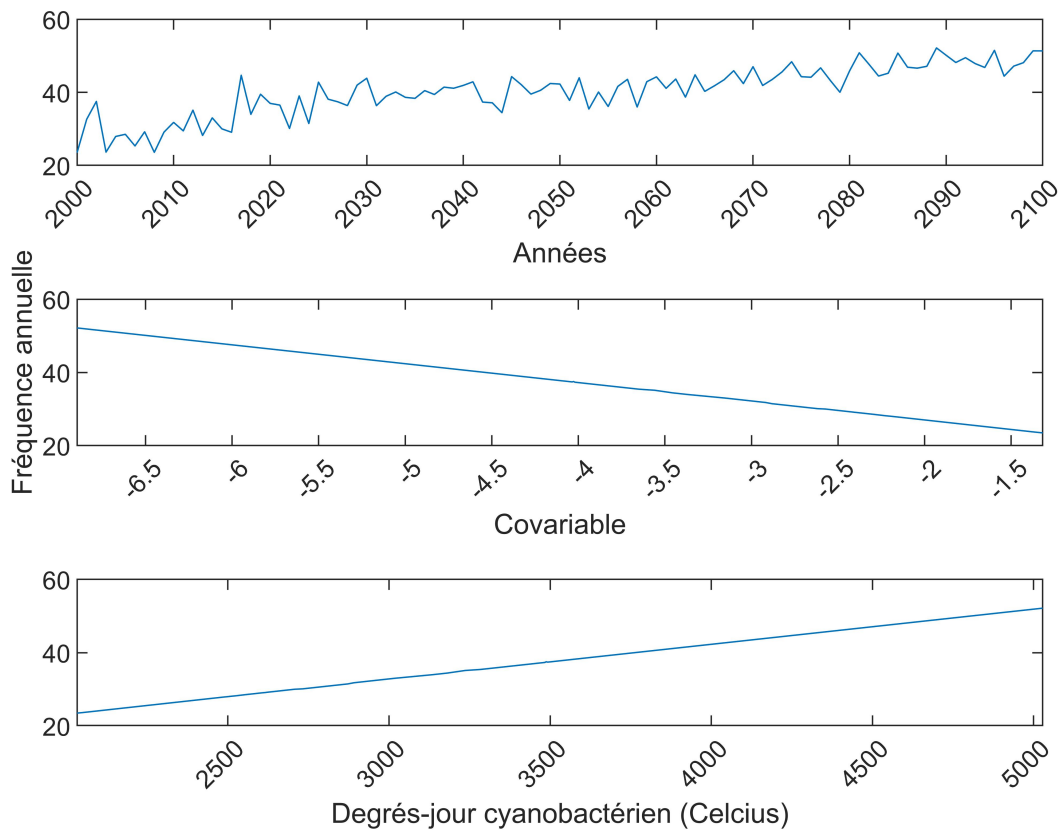


Figure 10.15 – Fréquence annuelle future du lac Brome présentée en fonction du temps (haut), de la covariable environnementale (milieu) et des degrés-jour cyanobactériens.

10.4 Sommaire

Ce chapitre présente un modèle fréquentiel régional permettant d'estimer la probabilité d'occurrence de la phénologie des FEA en fonction des descripteurs environnementaux prévalant sur le lac et son bassin versant. Un tel modèle permet d'estimer explicitement la prédisposition d'un plan d'eau donné au développement d'une FEA, et ce, en fonction des descripteurs environnementaux prévalant sur le bassin versant du plan d'eau. Il permet également d'estimer la probabilité d'apparition des FEA dans un plan d'eau pour lequel on ne dispose que des données sur les caractéristiques de son bassin versant et peu ou pas de données sur l'historique d'occurrence des FEA. En plus de permettre l'estimation de la probabilité d'occurrence des FEA par le biais de descripteurs environnementaux, il permet ainsi d'estimer la prédisposition d'un lac à développer des fleurs d'eau d'algues sur un lac non jaugé ou qui ne soit pas observable par le capteur MODIS. À titre d'exemple, il serait fort intéressant de projeter vers le passé la phénologie des FEA lors de la crise des algues bleu-vert en 2007 au lac Saint-Charles afin d'évaluer si le modèle simule la hausse observée durant cette année. Le modèle présente une meilleure précision que lors de l'utilisation de modèles calés localement sur un lac. Quoiqu'il s'avère difficile de quantifier l'exactitude du modèle, une comparaison effectuée à titre indicatif entre le temps de retour 2 ans de la fréquence d'apparition et la fréquence moyenne observée montrent des biais et des RMSE comparables aux modèles locaux. Notons que pour un lac donné, par exemple le lac Brome, l'estimation des quantiles de temps de retour était calculée en excluant ce lac de la base de données qui a servi au calage du modèle. Des simulations de la fréquence d'apparition future de FEA sur le lac Brome ont été estimées par le biais de ce modèle et de simulations du climat futur fournies par Ouranos. La fréquence annuelle augmenterait ainsi de 29 jours (2000-2016) à 48 jours sur l'horizon 2090, où la hausse d'un événement algal annuel est lié à une hausse d'environ 104 degrés-jour cyanobactériens.

Le modèle fréquentiel régional a été mis en application dans le cadre de l'étude de l'occurrence des FEA sur les lacs du Québec méridional, mais peut être utilisé pour d'autres études où la problématique majeure réside dans le manque de données qui limite la modélisation rigoureuse du phénomène, et où l'on souhaite estimer la probabilité d'occurrence d'un phénomène. L'utilisation d'un tel modèle inclut évidemment des hypothèses-clés qui sont importantes à rappeler. Puisqu'aucun modèle n'est parfait, il existe évidemment une part de la structure spatio-temporelle qui n'est pas modélisée par la fonction de densité de probabilité conditionnelle, mais nous supposons que

la majeure part de l'aléa structurel des FEA est capturée par le modèle fréquentiel régional non-stationnaire. Ensuite, une hypothèse à émettre mais qui peut également être validée statistiquement stipule que le champ résiduel du modèle est homoscédastique et indépendant. Ajoutons à cela l'absence d'auto-corrélation des données, où celle-ci peut être également vérifiée et, dans des cas où il y a auto-corrélation spatiale ou temporelle, peut être modélisée.

Cinquième partie

CONCLUSION

Chapitre 11

Conclusion

L'objectif principal de la thèse était de développer un modèle statistique permettant d'estimer la prédisposition des lacs québécois à développer des FEA en fonction des caractéristiques physiographiques et climatiques du lac et de son bassin versant. Trois objectifs spécifiques ont été définis: (*i*) le développement d'une procédure semi-automatisée d'estimation de la phénologie des FEA; (*ii*) l'établissement d'un portrait régional de l'occurrence des FEA au Québec et de leur lien avec des descripteurs environnementaux; et (*iii*) la modélisation de l'aléa structurel (déf. 1.1) des FEA. Les prochaines sections présentent un retour sur les principaux résultats obtenus (sections 11.1.1, 11.1.2 et 11.1.3), un sommaire des retombées scientifiques apportées (section 11.2) et une proposition d'axes de recherche à investiguer (section 11.3).

11.1 Résumé des principaux résultats

11.1.1 Partie II

Une procédure semi-automatisée d'estimation de la phénologie des FEA à l'échelle du lac a été développée et est opérationnelle sur les images satellitaires issues du capteur MODIS. Cette procédure, spécifiquement développée pour les milieux lacustres, intègre d'abord (*i*) le téléchargement des produits satellitaires MODIS. Ensuite, (*ii*) le prétraitement des images est effectué, ce dernier incluant une mise à l'échelle, une correction atmosphériques, une désagrégation spatiale des bandes 3-7 de 500 m à 250 m puis une conversion des images binaires en fichiers d'extension .PIX. (*iii*) Les

pixels de terre et les pixels mixtes sont ensuite masqués, d'une part afin de réduire l'incertitude du signal spectral en bordure des plans d'eau causée par des éléments hétérogènes tels que le fond du lac, les quais et les embarcations de plaisance, et d'autre part afin de cibler uniquement les pixels d'eau dans l'estimation de la chl-a, ce qui optimise considérablement le temps de calcul. (iv) Le couvert nuageux est masqué afin de permettre une estimation de la concentration en chl-a adéquate. (v) La concentration en chl-a est ensuite estimée à l'échelle du pixel puis (vi) les résultats sont extraits en format tabulaires. (vii) Les variables phénologiques sont finalement calculées à l'échelle du lac.

L'algorithme de détection du couvert nuageux était essentiel dans la mise en place d'un système semi-automatisé d'estimation de la concentration en chl-a en temps quasi-réel. Ce masque performe beaucoup mieux en milieux lacustre et fluvial que les produits déjà disponibles pour le capteur MODIS à la même résolution spatiale, entre autre de par ses erreurs de commission plus faibles, celles-ci ayant tendances à générer une surestimation de la concentration en chl-a. De plus, sa résolution spatiale à 250 m améliore la précision et l'ampleur de données exploitables dans le cadre d'un suivi de la qualité de l'eau sur des plans d'eau de petite à moyenne taille. Les variables phénologiques générées de la procédure semi-automatisée montrent que pour l'ensemble des lacs à l'étude, la fréquence médiane d'apparition est de 15 jours par année, l'intensité médiane est de $110 \mu\text{g chl-a } L^{-1}$ et l'étendue médiane est de 19 % de la superficie du lac.

La procédure d'estimation phénologique présente des résultats réalistes et dont les résultats de quelques lacs ont pu être comparés et validés à des résultats issus de la littérature. L'ensemble de la procédure de calcul après le téléchargement des produits MODIS prend en moyenne 5-15 minutes sur un processeur Intel Core i5 à 8 Gig de mémoire vive pour une image journalière. Les produits MODIS sont accessibles journalièrement avec un délai moyen de quelques heures après le passage du satellite. Cette procédure peut évidemment être ajustée à d'autres types de produits satellitaires. L'ajustement devra alors inclure alors un recalage de l'algorithme d'estimation de la chl-a et de l'algorithme de masque de nuage avec les bandes spectrales du produit satellitaire utilisé. La procédure d'estimation phénologique pourrait être sujette à une amélioration. D'abord, la région d'étude est très vaste et est localisée à des latitudes élevées (44° - 50° Nord) où règnent deux types de climat (continental humide et subarctique). Les milieux lacustres seront ainsi partiellement ou totalement gelés au courant de l'hiver, et la date de dégel total du couvert de glace dépend, entre autre, de la latitude où le lac est situé. Il est nécessaire de connaître cette date de même que la date du premier gel d'un lac afin que l'algorithme d'estimation de la concentration en chl-a soit appliqué

en période de dégel complet uniquement. La procédure pourrait ainsi être améliorée en intégrant, par exemple, un algorithme de détection du couvert de glace. De plus, notons que la procédure de désagrégation nécessite une mise à jour afin qu'elle puisse exploiter les produits MODIS datant de 2017 et plus.

11.1.2 Partie III

Cette partie porte sur la dynamique spatio-temporelle des FEA sur 580 lacs du Québec méridional entre 2000 et 2016, et présente la caractérisation de leurs liens potentiels avec des descripteurs physiographiques, morphologiques et climatiques prévalant sur les lacs et leur bassin versant. Les résultats montrent des tendances réalistes et logiques, qui valident l'utilité de cette approche permettant d'analyser l'état trophique des lacs et d'établir des liens avec l'évolution historique du bassin versant et du climat. À titre d'exemple, les données montrent l'augmentation attendue de l'importance des FEA (exprimée par la fréquence, l'étendue ou l'intensité) du nord au sud et de l'est vers l'ouest, ainsi qu'une apparition plus hâtive au sud et à l'ouest de la région d'étude. Ces tendances spatiales avaient été remarquées durant les dernières décennies par le MELCC (2000; 2002; 2014c) et sont généralement localisées sur les secteurs fortement urbanisés ou agricoles. On remarque également une plus grande étendue du phénomène pour les lacs situés dans le corridor entre 74 et 76°W s'étalant du sud-ouest de Montréal jusqu'au Mont-Tremblant, une région de forte villégiature. De plus, cette approche permet de mettre en évidence l'augmentation temporelle du phénomène entre 2000 et 2016 tel qu'obtenu par Winter *et al.* (2011) de 1994 à 2009 en Ontario, et infirmant l'hypothèse que cette tendance serait attribuée à une augmentation de l'attention portée au phénomène. De manière générale, cette étude permet de quantifier les tendances phénologiques dans le temps et en fonction des caractéristiques environnementales, alimentant ainsi les modèles prédictifs (par ex.: Cremona *et al.* (2017)).

Les résultats indiquent qu'il existe une différence de 37 jours dans la date de première apparition des FEA sur le territoire à l'étude, ce délai pouvant être exploité dans le cadre d'un "space-for-time substitution" afin d'évaluer la réponse des lacs au climat futur à ces latitudes (Pickett, 1989; Blois *et al.*, 2013). Les tendances temporelles observées sur une période de 17 ans démontrent l'importance de cette problématique dans un contexte d'expansion des populations survenant en parallèle au réchauffement climatique, une question particulièrement préoccupante dans un pays

qui compte plus d'un million de milieux lacustres sur son territoire (Pick, 2016). Entre 2000 et 2016, on constate une augmentation de la fréquence d'apparition des FEA et une apparition de plus en plus hâtive au début de l'été. Il est probable que ces changements surviennent en réponses aux changements climatiques ou encore aux changements dans l'occupation du territoire à travers la période d'étude. L'effet direct d'une hausse des températures de l'air est une diminution du temps d'englacement. Une saison estivale plus longue et chaude augmente la température de l'eau en surface et la stratification thermique des milieux lacustres (Dibike *et al.*, 2011; Bartosiewicz *et al.*, 2019b). En zone tempérée, la fonte plus hâtive du couvert de glace et le devancement de la saison de croissance pourraient expliquer le devancement de l'apparition des FEA, ces conditions (température de l'eau, stabilité verticale) favorisant la croissance des communautés phytoplanctoniques en milieux eutrophes (Kraemer *et al.*, 2017; Bartosiewicz *et al.*, 2019a).

La morphologie du plan d'eau et celle de son bassin versant jouent un rôle prépondérant dans la variabilité spatiotemporelle de la fréquence, de l'intensité et du début des FEA, particulièrement la superficie totale du lac et celle du bassin versant, où cette dernière se réfère directement à la capacité de stockage d'éléments nutritifs. La superficie forestière explique fortement la variabilité spatiotemporelle des variables phénologiques, mais cette variable porte en partie l'empreinte de la taille du bassin versant. En deuxième lieu, la superficie en zones urbanisées est le paramètre physiographique expliquant le plus la variabilité spatiotemporelle des variables phénologiques. L'ampleur des milieux urbanisés est directement proportionnelle à la fréquence et l'intensité des événements sur le bassin versant. Bien qu'il soit connu que l'enrichissement en P et en N dans les milieux lacustres soient des facteurs clés à la prolifération des FEA, les sources responsables de leur augmentation en fréquence, en intensité et en étendue sont encore très suggestives. Les résultats de l'ACC montrent que les sources de contamination diffuse provenant des milieux agricoles ne sont pas les seuls responsables. Or, les variables reflétant plutôt l'urbanisation expliquent beaucoup plus la variabilité spatiotemporelle des FEA que les variables reflétant l'agriculture. Les degrés-jours correspondent à la variable climatique expliquant le mieux les variables phénologiques, suivis par la précipitation totale annuelle. Bien que dans la littérature plusieurs descripteurs de température aient montré une influence notable sur la prolifération de FEA, incluant la température de l'air, la température de l'eau, les heures d'ensoleillement et les degrés-jour, il semble que les degrés-jour soient plus directement reliés au taux de croissance des organismes ectothermes (Trudgill *et al.*, 2005) et devraient être privilégiés dans le cadre d'une étude sur le suivi ou la prédiction des FEA. Ralston *et al.* (2014)

ont notamment démontré les liens de dépendance entre les degrés-jour et la variabilité interannuelle des FEA près de la péninsule du Cap Cod au Massachusetts, de même pour Rolland *et al.* (2013) qui ont montré une corrélation entre les degrés-jour et le biovolume de cyanobactéries (*Microcystis aeruginosa*) sur le lac Saint-Charles près de Québec.

Le portrait régional de la qualité de l'eau des milieux lacustres du Québec méridional établi dans la présente étude s'avère être représentatif, précis et complet en terme de qualité d'information et du nombre de milieux récepteurs étudiés. Ces résultats permettront au MELCC de cibler, d'une part, les milieux problématiques qui exigent une surveillance accrue, et également les zones problématiques sur un lac donné et les périodes optimales pour déployer des gens sur le terrain pour le suivi *in situ*. La procédure de gestion sera améliorée de par la qualité et la quantité d'information disponible et permettra d'orienter les décisions quant au déploiement du personnel sur le terrain, réduisant ainsi les coûts alloués aux analyses en laboratoire.

11.1.3 Partie IV

Un modèle fréquentiel régional non-stationnaire a été développé afin d'estimer les probabilités d'apparition future des FEA des lacs québécois et d'étudier l'impact des changements climatiques sur la phénologie des FEA sur notre période de référence et à des horizons futurs. Ce modèle permet d'identifier les mécanismes sous-tendant l'apparition des FEA et d'estimer explicitement la prédisposition d'un plan d'eau donné aux développements d'une FEA en se basant sur les conditions physiographiques et climatiques du bassin versant. Ce modèle permet également d'estimer la probabilité d'apparition des FEA sur un plan d'eau pour lequel on ne dispose que peu ou pas de données sur le phénomène, mais pour lequel on connaît les caractéristiques physiographiques et climatiques de son bassin versant. De plus, le modèle fréquentiel régional permet d'une part d'améliorer notre compréhension sur l'apparition des FEA sur un lac donné, mais également d'étudier *a posteriori* l'impact de certains facteurs climatiques et anthropiques sur les écosystèmes d'eau douce. Des scénarios d'atténuation visant à améliorer la qualité de l'eau avant leur mise en oeuvre (comme des pratiques de gestion bénéfiques par exemple) pourraient donc être élaborés et mis en place.

L'ajustement de la fréquence d'apparition et de l'intensité des FEA en fonction des descripteurs environnementaux a été fait sur un ensemble de 580 milieux lacustres du Québec méridional. Le modèle régional (intégrant l'ensemble des lacs) a été comparé à des modèles fréquents locaux, cha-

cun étant ajusté sur un lac en particulier. Les lacs sélectionnés dans le cadre de cette comparaison étaient la Baie Missisquoi, le lac Brome, le lac Aylmer et le lac Saint-Jean, ces derniers présentant des caractéristiques climatiques et anthropiques ainsi que des niveaux trophiques variables. Les résultats issus de ces comparaisons ont montré que le modèle régional capte beaucoup plus la variabilité spatio-temporelle des quantiles estimés et améliore la précision d'estimation de la probabilité d'occurrence comparativement aux modèles locaux (étendues et écart-types plus faibles pour l'ensemble des comparaisons). De plus, bien que l'exactitude dans les estimations soit difficilement quantifiable, elle a tout de même été évaluée à titre indicatif en comparant les temps de retours aux 2 ans estimés par le modèle régional (correspondant à une estimation de la valeur médiane) avec les valeurs moyennes observées pour une année donnée et un lac donnée. Le modèle régional présente des valeurs de biais et de RMSE plus faibles pour l'ensemble des comparaisons, ce qui laisse supposer que l'exactitude est raisonnablement acceptable. Par ailleurs, le modèle régional a été appliqué dans le but d'estimer des probabilités d'apparition futures de FEA grâce à l'utilisation de simulations du climat futur fournies par l'organisme Ouranos. Des simulations de la fréquence d'apparition futures sur le lac Brome estimées jusqu'à 2100 montrent qu'il y aura une potentielle hausse de 65% des fréquences d'apparition annuelles de FEA jusqu'à l'horizon 2090. En d'autres mots, il requiert une hausse annuelle de 104 degrés-jour pour incrémenter de un la fréquence d'apparition annuelle de FEA sur ce lac.

11.2 Retombées scientifiques

Ce projet conduit à un ensemble d'outils d'aide à la décision utilisables par les gestionnaires de l'eau, les organismes de bassins versants ou les municipalités. Il répond ainsi concrètement aux besoins identifiés par le gouvernement du Québec dans son plan d'intervention sur les algues bleu-vert 2007-2017 (MELCC, 2014c).

11.2.1 Portée thématique

1. Les procédures d'estimation des concentrations en chl-a sur le Québec méridional par des images satellitaires et de caractérisation des épisodes de FEA permet de développer des outils de surveillance en temps quasi réel afin d'aider à l'optimisation du réseau de surveillance. Il s'agit d'une amélioration considérable au mode de suivi actuel employé par les intervenants

travaillant dans la gestion des épisodes de FEA. Cette procédure de suivi permettrait de diminuer drastiquement les coûts comparativement à un suivi via l'échantillonnage *in situ* tout en améliorant la précision et le temps de réponse à l'efflorescence de FEA, favorisant ainsi une meilleure gestion du risque lié à la santé publique, à l'économie locale et à la biodiversité du milieu.

2. La base de données historiques d'épisodes de FEA permet de tracer un portrait précis de la qualité de l'eau au Québec en lien avec les efflorescences de FEA, et ce, via une caractérisation quantitative des épisodes ce qui aidera à une meilleure compréhension de leur apparition. De plus, cette base de données permet d'appuyer les municipalités dans des actions de prévention et d'appuyer le programme de suivi des plans d'eau par le MELCC. Il s'agit d'un premier portrait complet de l'état de lacs dont la majorité ne sont pas (ou peu) suivis régulièrement. Dans une optique de suivi et de restauration d'un plan d'eau, une bonne compréhension du passé est cruciale pour prévenir et appréhender l'avenir.
3. Le modèle d'aléa structurel a permis d'abord d'établir les liens possibles entre la phénologie des FEA et les facteurs environnementaux, et également d'étudier les impacts des changements climatiques et autres perturbations anthropiques sur ces écosystèmes. Peu de travaux se sont intéressés explicitement à la dynamique régionale des FEA et encore moins à la variabilité spatiale des facteurs structurels des bassins versants où se développent les FEA. Concrètement, ce modèle permet de:
 - (a) Estimer la probabilité d'occurrence de la phénologie des FEA sur l'ensemble des lacs à l'étude.
 - (b) Estimer la probabilité d'occurrence de la phénologie des FEA sur un plan d'eau pour lequel on ne dispose que des données sur les caractéristiques de son bassin versant et peu ou pas de données sur l'historique d'occurrence des FEA.
 - (c) Évaluer des scénarios d'aménagement ou de restauration d'un plan d'eau visant à améliorer la qualité de l'eau avant leur implantation en prévision de changements climatiques ou de changements dans l'occupation du territoire du bassin versant du plan d'eau.
 - (d) Évaluer des scénarios climatiques en vue d'étudier l'impact des changements climatiques sur l'occurrence de la phénologie des FEA.

11.2.2 Portée méthodologique

1. Le développement technique et méthodologique menant au modèle fréquentiel régional permet la modélisation de données spatio-temporelles, un type de modélisation qui est généralement difficile à mettre en place en raison des contraintes d'indépendance, d'homogénéité et d'homoscédasticité des données.
2. Faisant suite au propos ci-haut, le modèle régional fréquentiel permet de répondre au manque de données nécessaires à la modélisation d'un phénomène, une problématique importante en modélisation d'événements rares. L'utilisation de données provenant non-seulement d'un cadre temporel mais bien spatio-temporel augmente le nombre total de données et permet ainsi de modéliser un phénomène de manière plus précise.
3. Pour toute variable aléatoire pouvant être représentée par une distribution statistique, il est possible d'utiliser une telle procédure de modélisation afin d'ajuster une fonction de densité à probabilité conditionnelle. Le modèle est ainsi adaptable pour toute autre modélisation d'événements naturels extrêmes.
4. La complexité des événements naturels extrêmes de même que les risques imminents qu'ils causent à la population nécessitent le développement de modèles qui puissent assurer une modélisation adéquate et précise de ces phénomènes. Un autre aspect innovateur au développement technique et méthodologique menant au modèle fréquentiel régional est qu'il permet de répondre à des problèmes complexes, où cette complexité devient élevée avec la dimensionnalité des variables à l'étude. La procédure de modélisation utilisée dans le cadre de cette thèse assure une réduction de la dimensionnalité des variables par l'utilisation de l'analyse canonique des corrélations.

11.3 Perspectives

Plusieurs éléments de cette thèse peuvent être approfondis dans des travaux futurs.

Faisant suite à ce qui a été dit dans la section précédente, le modèle fréquentiel régional peut être utilisé dans d'autres domaines d'application où la problématique majeure réside dans le manque ou le peu de données, ce qui limite la modélisation complète d'un phénomène, ou dans des études de gestion du risque lié à un phénomène où l'on souhaite estimer la probabilité d'occurrence d'un

événement. La portée du modèle est donc vaste et pourrait d'une part être appliquée sur d'autres caractéristiques décrivant la qualité de l'eau (MED, MOD), mais également sur d'autres phénomènes extrêmes tels qu'en hydrologie fréquentielle où la dimension temporelle et/ou spatiale des données est souvent très limitée. Ce modèle permettrait également d'estimer des quantiles sur des régions où l'on ne dispose que de peu ou pas d'information sur l'historique d'occurrence d'un phénomène.

Les simulations en climat futur ont été estimées pour un seul lac (lac Brome) et pour une variable phénologique (fréquence d'apparition). Il serait intéressant d'étendre ces analyses à plusieurs lacs et pour d'autres variables phénologiques à partir de différentes simulations climatiques fournies par l'organisme Ouranos. Également, une étude de l'impact des changements anthropiques sur le bassin versant d'un lac donné serait également intéressante dans le contexte où l'on souhaiterait évaluer des scénarios de restauration de la qualité de l'eau. Par exemple, il serait intéressant de tester l'effet de meilleures pratiques de gestion des sources de pollution pouvant affecter l'apparition de FEA dans un contexte de changements climatiques. De plus, il serait intéressant d'évaluer le potentiel du modèle à estimer l'occurrence de FEA sur des lacs non observables par le capteur MODIS en raison de leur taille ou de leur forme. À titre d'exemple, le lac Saint-Charles, où une partie de l'eau potable de la ville de Québec est puisée, n'est pas observable par le capteur MODIS en raison de sa taille. Il serait fort intéressant de projeter vers le passé la phénologie des FEA lors de la crise des algues bleu-vert en 2007 au lac Saint-Charles afin d'évaluer si le modèle simule la hausse observée durant cette année.

L'ajustement de d'autres types de densités de probabilité conditionnelles pourrait être mis en place sur la base de données historiques d'événements algaux. Puisque très peu de travaux ont permis l'étude phénologique des FEA à une telle échelle, les choix de densités et des formes fonctionnelles utilisées peuvent difficilement être basés sur des travaux antérieurs. Les distributions Lognormale et Gamma ont été choisies dans la présente thèse puisqu'une distribution classique à deux paramètres qui puisse être exprimée en fonction de l'espérance et du coefficient de détermination était désirée. D'autres distributions à deux paramètres pourraient être utilisées, ou à trois paramètres, où le troisième paramètre ajusterait le coefficient d'aplatissement de la distribution.

De plus, puisque la corrélation entre les variables phénologiques était peu significative, un modèle multivarié ou par copule n'a pas été considéré dans le cadre de cette étude, mais leur utilisation pourrait être considérée lorsque la corrélation entre les variables est positive ou négative. Les va-

riables phénologiques de la fréquence d'apparition et de la date de début de l'événement montrent le plus haut coefficient de détermination à 0.436. Bien que l'utilisation de copules soit de plus en plus utilisée en hydrologie (Zhang & Singh, 2007; Chebana & Ouarda, 2011), son utilisation dans l'évaluation de la qualité de l'eau demeure encore peu documentée jusqu'à présent et le choix des copules à tester peut, encore une fois, difficilement être basé sur des travaux antérieurs. Notons cependant que la famille de copules Archimédiennes pourrait être une option sachant qu'elles peuvent être construites facilement et possèdent une large variété de copules. Zhang & Singh (2007) et Wang *et al.* (2012) ont d'ailleurs utilisé ces types de copules pour étudier l'interaction entre la chl-a et des variables environnementales.

Finalement, l'incertitude liée à l'estimation de la concentration en chl-a obtenues avec le modèle basé sur les méthodes d'ensemble développé par El-Alem *et al.* (2019) pourrait être incluse dans l'estimation des quantiles de retour du modèle régional. Pour un pixel donné, l'algorithme génère maintenant une estimation de la valeur moyenne en concentrations en chl-a, mais également une valeur d'incertitude (écart-type) autour de cette estimation.

Références

- Adamson P, Metcalfe A & Parmentier B (1999). Bivariate extreme value distributions: an application of the gibbs sampler to the analysis of floods. *Water Resources Research*, 35(9):2825–2832.
- Agence Française de Sécurité Sanitaire des Aliments & Agence Française de Sécurité Sanitaire de l'Environnement et du Travail (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.
- Ahn CY, Chung AS & Oh HM (2002). Rainfall, phycocyanin, and n: P ratios related to cyanobacterial blooms in a korean large reservoir. *Hydrobiologia*, 474(1-3):117–124.
- Allan IJ, Vrana B, Greenwood R, Mills GA, Roig B & Gonzalez C (2006). A “toolbox” for biological and chemical monitoring requirements for the european union’s water framework directive. *Talanta*, 69(2):302–322.
- 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):12–23.
- Banks AC & Mélin F (2015). An assessment of cloud masking schemes for satellite ocean colour data of marine optical extremes. *International Journal of Remote Sensing*, 36(3):797–821.
- Bartlett MS (1938). Further aspects of the theory of multiple regression. *Mathematical Proceedings of the Cambridge Philosophical Society*, Cambridge University Press, volume 34, pages 33–40.
- Bartlett MS (1939). A note on tests of significance in multivariate analysis. *Mathematical Proceedings of the Cambridge Philosophical Society*, Cambridge University Press, volume 35, pages 180–185.
- Bartosiewicz M, Przytulska A, Deshpande BN, Antoniadou D, Cortes A, MacIntyre S, Lehmann MF & Laurion I (2019a). Effects of climate change and episodic heat events on cyanobacteria in a eutrophic polymictic lake. *Science of The Total Environment*, 693:133414.
- Bartosiewicz M, Przytulska A, Lapierre JF, Laurion I, Lehmann MF & Maranger R (2019b). Hot tops, cold bottoms: Synergistic climate warming and shielding effects increase carbon burial in lakes. *Limnology and Oceanography Letters*.
- Bartram J & Chorus I (1999). Toxic cyanobacteria in water. a guide to their public health consequences, monitoring and management. 416 pages.
- Bastien C, Cardin R, Veilleux É, Deblois C, Warren A & Laurion I (2011). Performance evaluation of phycocyanin probes for the monitoring of cyanobacteria. *Journal of Environmental Monitoring*, 13(1):110–118.

- 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. DOI:10.1016/j.jglr.2009.05.007.
- Ben Aissia MA (2014). *Étude des variables hydrologiques dans un cadre multivarié et dans un contexte de changement*. Thèse de doctorat, Université du Québec, Institut national de la recherche scientifique.
- Bergström AK (2010). The use of tn: Tp and din: Tp ratios as indicators for phytoplankton nutrient limitation in oligotrophic lakes affected by n deposition. *Aquatic Sciences*, 72(3):277–281.
- Berk A, Anderson GP, Bernstein LS, Acharya PK, Dothe H, Matthew MW, Adler-Golden SM, Chetwynd Jr JH, Richtsmeier SC, Pukall B *et al.* (1999). Modtran4 radiative transfer modeling for atmospheric correction. *SPIE's International Symposium on Optical Science, Engineering, and Instrumentation*, International Society for Optics and Photonics, pages 348–353.
- Berlamont X, Valeria O & Harvey B (2013). Détection de cyanobactéries dans les lacs abitibi, témiscamingue et malartic par le biais de l'imagerie satellite.
- Binding CE, Greenberg TA & Bukata RP (2011). Time series analysis of algal blooms in lake of the woods using the meris maximum chlorophyll index. *Journal of plankton research*, 33(12):1847–1852.
- Blenckner T, Adrian R, Livingstone DM, Jennings E, Weyhenmeyer GA, George D, Jankowski T, Järvinen M, Aonghusa CN, Noges T *et al.* (2007). Large-scale climatic signatures in lakes across europe: a meta-analysis. *Global Change Biology*, 13(7):1314–1326.
- Blois JL, Williams JW, Fitzpatrick MC, Jackson ST & Ferrier S (2013). Space can substitute for time in predicting climate-change effects on biodiversity. *Proceedings of the National Academy of Sciences*, 110(23):9374–9379.
- Bonnet M & Poulin M (2004). Dylem-1d: a 1d physical and biochemical model for planktonic succession, nutrients and dissolved oxygen cycling: Application to a hyper-eutrophic reservoir. *Ecological modelling*, 180(2):317–344.
- Brody SR, Lozier MS & Dunne JP (2013). A comparison of methods to determine phytoplankton bloom initiation. *Journal of Geophysical Research: Oceans*, 118(5):2345–2357.
- Burn DH (1990). Evaluation of regional flood frequency analysis with a region of influence approach. *Water Resources Research*, 26(10):2257–2265.
- Canadian Geographic (2018). *Extremes of weather - The Canadian Atlas Online*. http://www.canadiangeographic.com/atlas/themes.aspx?id=weathersub=weather_power_solarpowerlang=En.
- Capblancq J & Décamps H (2002). L'eutrophisation des eaux continentales: questions à propos d'un processus complexe. *Nature Sciences Sociétés*, 10(2):6–17.
- Carmichael WW (2001). Health effects of toxin-producing cyanobacteria: "the cyanohabs". *Human and ecological risk assessment: An International Journal*, 7(5):1393–1407.
- Carpenter SR, Ludwig D & Brock WA (1999). Management of eutrophication for lakes subject to potentially irreversible change. *Ecological applications*, 9(3):751–771.

- Carvalho L, Miller CA, Scott EM, Codd GA, Davies PS & Tyler AN (2011). Cyanobacterial blooms: statistical models describing risk factors for national-scale lake assessment and lake management. *Science of the Total Environment*, 409(24):5353–5358.
- Carvalho L, Poikane S, Solheim AL, Phillips G, Borics G, Catalan J, De Hoyos C, Drakare S, Dudley B, Järvinen M *et al.* (2013). Strength and uncertainty of phytoplankton metrics for assessing eutrophication impacts in lakes. *Hydrobiologia*, 704(1):127–140.
- Cavadias G (1990). The canonical correlation approach to regional flood estimation. *Regionalization in hydrology*, 191:171–178.
- Chaffin JD, Bridgeman TB & Bade DL (2013). Nitrogen constrains the growth of late summer cyanobacterial blooms in lake erie. *Advances in Microbiology*, 2013.
- Chebana F & Ouarda T (2011). Multivariate quantiles in hydrological frequency analysis. *Environmetrics*, 22(1):63–78.
- Chen S & Zhang T (2015). An improved cloud masking algorithm for modis ocean colour data processing. *Remote Sensing Letters*, 6(3):218–227.
- Chen S, Zhang T & Hu L (2014). Evaluation of the nir-swir atmospheric correction algorithm for modis-aqua over the eastern china seas. *International Journal of Remote Sensing*, 35(11-12):4239–4251.
- Cho S, Lim B, Jung J, Kim S, Chae H, Park J, Park S & Park JK (2014). Factors affecting algal blooms in a man-made lake and prediction using an artificial neural network. *Measurement*, 53:224–233.
- Choo F, Zamyadi A, Newton K, Newcombe G, Bowling L, Stuetz R & Henderson RK (2018). Performance evaluation of in situ fluorometers for real-time cyanobacterial monitoring. *H2Open Journal*, 1(1):26–46.
- Chorus I, Falconer IR, Salas HJ & Bartram J (2000). Health risks caused by freshwater cyanobacteria in recreational waters. *Journal of Toxicology and Environmental Health Part B: Critical Reviews*, 3(4):323–347.
- Cole H, Henson S, Martin A & Yool A (2012). Mind the gap: The impact of missing data on the calculation of phytoplankton phenology metrics. *Journal of Geophysical Research: Oceans (1978-2012)*, 117(C8).
- Cremona F, Vilbaste S, Couture RM, Nõges P & Nõges T (2017). Is the future of large shallow lakes blue-green? comparing the response of a catchment-lake model chain to climate predictions. *Climatic change*, 141(2):347–361.
- Davis TW, Berry DL, Boyer GL & Gobler CJ (2009). The effects of temperature and nutrients on the growth and dynamics of toxic and non-toxic strains of microcystis during cyanobacteria blooms. *Harmful algae*, 8(5):715–725.
- de Jong S, Auping WL, Oosterveld WT, Usanov A, Abdalla M, Van de Bovenkamp A & della Frattina CF (2017). *The geopolitical impact of climate mitigation policies: How hydrocarbon exporting rentier states and developing nations can prepare for a more sustainable future*. The Hague Centre for Strategic Studies.

- De Michele C, Salvadori G, Canossi M, Petaccia A & Rosso R (2005). Bivariate statistical approach to check adequacy of dam spillway. *Journal of Hydrologic Engineering*, 10(1):50–57.
- Dekker A, Hoogenboom H, Goddijn L & Malthus T (1997). The relation between inherent optical properties and reflectance spectra in turbid inland waters. *Remote Sensing Reviews*, 15(1-4):59–74.
- Dibike Y, Prowse T, Saloranta T & Ahmed R (2011). Response of northern hemisphere lake-ice cover and lake-water thermal structure patterns to a changing climate. *Hydrological Processes*, 25(19):2942–2953.
- Diggle PJ (2013). *Statistical analysis of spatial and spatio-temporal point patterns*. Chapman and Hall/CRC.
- Doerffer R, Schiller H, Fischer J, Preusker R & Bouvet M (2008). The impact of sun glint on the retrieval of water parameters and possibilities for the correction of meris scenes. *Proceedings of the 2nd MERIS/(A) ATSR User Workshop, ESA SP-666,(ESA, Frascati, Italy, 2008)*, Citeseer.
- Duan H, Loisel SA, Zhu L, Feng L, Zhang Y & Ma R (2015). Distribution and incidence of algal blooms in lake taihu. *Aquatic Sciences*, 77(1):9–16.
- Duan H, Ma R, Xu X, Kong F, Zhang S, Kong W, Hao J & Shang L (2009). Two-decade reconstruction of algal blooms in china’s lake taihu. *Environmental Science & Technology*, 43(10):3522–3528.
- Duan H, Ma R, Zhang Y & Loisel SA (2014). Are algal blooms occurring later in lake taihu? climate local effects outcompete mitigation prevention. *Journal of Plankton Research*, 36(3):866–871.
- Edwards M & Richardson AJ (2004). Impact of climate change on marine pelagic phenology and trophic mismatch. *Nature*, 430(7002):881–884.
- El-Alem A (2015). *Développement d’une approche de suivi des FEA à l’aide de l’imagerie du capteur MODIS désagrégée à 250 mètres de résolution spatiale, adaptées aux lacs du Québec méridional*.
- 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 (2014a). An adaptive model to monitor chlorophyll-a in inland waters in southern quebec using downscaled modis imagery. *Remote Sensing*, 6(7):6446–6471.
- El-Alem A, Chokmani K, Laurion I & El-Adlouni SE (2014b). An ensemble based system for chlorophyll-a estimation using modis imagery over southern quebec inland waters. *Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International*, IEEE, pages 3878–3881.
- El-Alem A, Chokmani K, Laurion I, El-Adlouni SE, Raymond S & Ratte-Fortin C (2019). Ensemble based systems to monitor algal bloom with remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*.
- Elliott J (2010). The seasonal sensitivity of cyanobacteria and other phytoplankton to changes in flushing rate and water temperature. *Global Change Biology*, 16(2):864–876.
- Elliott J, Jones I & Thackeray S (2006). Testing the sensitivity of phytoplankton communities to changes in water temperature and nutrient load, in a temperate lake. *Hydrobiologia*, 559(1):401–411.

- Elliott JA (2012). Is the future blue-green? a review of the current model predictions of how climate change could affect pelagic freshwater cyanobacteria. *water research*, 46(5):1364–1371.
- Ferber L, Levine S, Lini A & Livingston G (2004). Do cyanobacteria dominate in eutrophic lakes because they fix atmospheric nitrogen? *Freshwater Biology*, 49(6):690–708.
- Figueredo CC & Giani A (2005). Ecological interactions between Nile tilapia (*Oreochromis niloticus*, L.) and the phytoplanktonic community of the Furnas Reservoir (Brazil). *Freshwater Biology*, 50(8):1391–1403.
- Fisher A & Danaher T (2013). A water index for Spot-5 HRG satellite imagery, New South Wales, Australia, determined by linear discriminant analysis. *Remote Sensing*, 5(11):5907–5925.
- Fisher RA (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2):179–188.
- Fortin N, Aranda-Rodriguez R, Jing H, Pick F, Bird D & Greer CW (2010). Detection of microcystin-producing cyanobacteria in Missisquoi Bay, Quebec, Canada, using quantitative PCR. *Appl. Environ. Microbiol.*, 76(15):5105–5112.
- Franks P & Keafer B (1995). Sampling techniques and strategies for coastal phytoplankton blooms. *Manual on Harmful Marine Microalgae*. Paris: Unesco.
- French M & Recknagel F (1970). Modeling of algal blooms in freshwaters using artificial neural networks. *WIT Transactions on Ecology and the Environment*, 6.
- Friedl MA & Brodley CE (1997). Decision tree classification of land cover from remotely sensed data. *Remote sensing of environment*, 61(3):399–409.
- Gallon J (1992). Tansley review no. 44. Reconciling the incompatible: N₂ fixation and O₂. *New Phytologist*, pages 571–609.
- Garnier J, Nemery J, Billen G & Théry S (2005). Nutrient dynamics and control of eutrophication in the Marne river system: modelling the role of exchangeable phosphorus. *Journal of Hydrology*, 304(1):397–412.
- Genest C, Favre AC, Béliveau J & Jacques C (2007). Metaelliptical copulas and their use in frequency analysis of multivariate hydrological data. *Water Resources Research*, 43(9).
- Goddard Space Flight Center (2014). *LAADS Web: Level 1 and Atmosphere Archive and Distribution System*. <http://ladsweb.nascom.nasa.gov/index.html>.
- Gordon HR (1997). Atmospheric correction of ocean color imagery in the Earth observing system era. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 102(D14):17081–17106.
- Gouvernement du Canada (2017). *Canadian Digital Elevation Model*. Online; accessed 23 July 2018.
- Government of Canada (2015). *Land Use 1990, 2000, 2010*. <https://open.canada.ca/data/en/dataset/18e3ef1a-497c-40c6-8326-aac1a34a0dec>.
- Government of Canada (2016). *Population Ecumene Census Division, Cartographic Boundary files*. <https://open.canada.ca/data/en/dataset/8498f9b4-4914-456c-9223-4260ea3bea4d>. Online; accessed 23 July 2018.

- Gräler B, van den Berg M, Vandenberghe S, Petroselli A, Grimaldi S, Baets BD & Verhoest N (2013). Multivariate return periods in hydrology: a critical and practical review focusing on synthetic design hydrograph estimation. *Hydrology and Earth System Sciences*, 17(4):1281–1296.
- Gravelius H (1914). Grundrifi der gesamten gewässerkunde, band 1: h'lufkunde (compendium of hydrology?, vol. 1: Rivers, in german). *Goschen, Berlin, Germany*.
- Guallar C, Delgado M, Diogene J & Fernandez-Tejedor M (2016). Artificial neural network approach to population dynamics of harmful algal blooms in alfacs bay (nw mediterranean): Case studies of karlodinium and pseudo-nitzschia. *Ecological modelling*, 338:37–50.
- Guildford SJ, Hecky RE, Smith RE, Taylor WD, Charlton MN, Barlow-Busch L & North RL (2005). Phytoplankton nutrient status in lake erie in 1997. *Journal of Great Lakes Research*, 31:72–88.
- Guttman NB, Hosking J & Wallis JR (1993). Regional precipitation quantile values for the continental united states computed from l-moments. *Journal of Climate*, 6(12):2326–2340.
- Havens KE, James RT, East TL & Smith VH (2003). N: P ratios, light limitation, and cyanobacterial dominance in a subtropical lake impacted by non-point source nutrient pollution. *Environmental Pollution*, 122(3):379–390.
- Henson RK (2002). The logic and interpretation of structure coefficients in multivariate general linear model analyses.
- Higgins SN, Pennuto CM, Howell ET, Lewis TW & Makarewicz JC (2012). Urban influences on cladophora blooms in lake ontario. *Journal of Great Lakes Research*, 38:116–123.
- Hipel KW (1985). Time series analysis in perspective 1. *JAWRA Journal of the American Water Resources Association*, 21(4):609–623.
- Hu C (2011). An empirical approach to derive modis ocean color patterns under severe sun glint. *Geophysical Research Letters*, 38(1).
- 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. DOI:10.1029/2009jc005511.
- Hu W, Connell D, Mengersen K & Tong S (2009). Weather variability, sunspots, and the blooms of cyanobacteria. *EcoHealth*, 6(1):71–78.
- Huisman J, Codd GA, Paerl HW, Ibelings BW, Verspagen JM & Visser PM (2018). Cyanobacterial blooms. *Nature Reviews Microbiology*, 16(8):471.
- Illian J, Penttinen A, Stoyan H & Stoyan D (2008). *Statistical analysis and modelling of spatial point patterns*. volume 70. John Wiley & Sons.
- International Council for the Exploration of the Sea (2004). Report of the ices-ioc working group on harmful algal bloom dynamics. 42 pages.
- Jakob T, Schreiber U, Kirchesch V, Langner U & Wilhelm C (2005). Estimation of chlorophyll content and daily primary production of the major algal groups by means of multiwavelength-excitation pam chlorophyll fluorometry: performance and methodological limits. *Photosynthesis Research*, 83(3):343–361.

- Janssen F, Neumann T & Schmidt M (2004). Inter-annual variability in cyanobacteria blooms in the baltic sea controlled by wintertime hydrographic conditions. *Marine ecology. Progress series*, 275:59–68.
- Jayatissa L, Silva E, McElhiney J & Lawton L (2006). Occurrence of toxigenic cyanobacterial blooms in freshwaters of sri lanka. *Systematic and applied microbiology*, 29(2):156–164.
- Johnson R & Wichern D (2007). *Applied multivariate Statistical Analysis - 6th Edition*. Pearson Prentice Hall, New Jersey, pages 539–574.
- Jones JR & Bachmann RW (1976). Prediction of phosphorus and chlorophyll levels in lakes. *Journal (Water Pollution Control Federation)*, pages 2176–2182.
- Kahru M, Brotas V, MANZANO-SARABIA M & Mitchell B (2011). Are phytoplankton blooms occurring earlier in the arctic? *Global Change Biology*, 17(4):1733–1739.
- Kahru M, Leppanen J & Rud O (1993). Cyanobacterial blooms cause using heating of the sea surface. *Marine Ecology Progress Series*, 101:1–7.
- Kahru, M. 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. DOI:10.1029/2004EO450002.
- Karul C, Soyupak S, Çilesiz AF, Akbay N & Germen E (2000). Case studies on the use of neural networks in eutrophication modeling. *Ecological modelling*, 134(2-3):145–152.
- Kraemer BM, Mehner T & Adrian R (2017). Reconciling the opposing effects of warming on phytoplankton biomass in 188 large lakes. *Scientific reports*, 7(1):10762.
- Kudela RM, Palacios SL, Austerberry DC, Accorsi EK, Guild LS & Torres-Perez J (2015). Application of hyperspectral remote sensing to cyanobacterial blooms in inland waters. *Remote Sensing of Environment*, 167:196–205.
- Kutser T, Metsamaa L, Strömbeck N & Vahtmäe E (2006). Monitoring cyanobacterial blooms by satellite remote sensing. *Estuarine, Coastal and Shelf Science*, 67(1-2):303–312.
- Laanemets J, Lilover MJ, Raudsepp U, Autio R, Vahtera E, Lips I & Lips U (2006). A fuzzy logic model to describe the cyanobacteria nodularia spumigena blooms in the gulf of finland, baltic sea. *Hydrobiologia*, 554(1):31–45.
- Langlois G (2008). Plan d’actions visant à limiter et diminuer la prolifération des cyanobactéries dans le lac memphrémagog.
- Larson JH, Evans MA, Kennedy RJ, Bailey SW, Loftin KA, Laughrey ZR, Femmer RA, Schaeffer JS, Richardson WB, Wynne TT *et al.* (2018). Associations between cyanobacteria and indices of secondary production in the western basin of lake erie. *Limnology and Oceanography*, 63(S1):S232–S243.
- Lavender S, Pinkerton M, Moore G, 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 & Vincent W (2007a). Les fleurs d’eau de cyanobactéries: Document d’information vulgarisée. *INRS*, 917:20.

- Lavoie I, Laurion I, Warren A & Vincent W (2007b). Les fleurs d'eau de cyanobactéries: Revue de littérature. Institut National de la Recherche Scientifique (INRS)-Centre Eau, Terre et Environnement.
- Lee JH, Huang Y, Dickman M & Jayawardena AW (2003). Neural network modelling of coastal algal blooms. *Ecological Modelling*, 159(2-3):179–201.
- Li L, Li L & Song K (2015). Remote sensing of freshwater cyanobacteria: An extended iop inversion model of inland waters for partitioning absorption coefficient and estimating phycocyanin. *Remote Sensing of Environment*, 157:9–23.
- Lilover MJ & Stips A (2008). The variability of parameters controlling the cyanobacteria bloom biomass in the baltic sea. *Journal of Marine Systems*, 74:S108–S115.
- Liu Y, Wang Z, Guo H, Yu S & Sheng H (2013). Modelling the effect of weather conditions on cyanobacterial bloom outbreaks in lake dianchi: a rough decision-adjusted logistic regression model. *Environmental Modeling & Assessment*, 18(2):199–207.
- Lubac B & Loisel H (2007). Variability and classification of remote sensing reflectance spectra in the eastern english channel and southern north sea. *Remote Sensing of Environment*, 110(1):45–58.
- Lunetta RS, Schaeffer BA, Stumpf RP, Keith D, Jacobs SA & Murphy MS (2015). Evaluation of cyanobacteria cell count detection derived from meris imagery across the eastern usa. *Remote Sensing of Environment*, 157:24–34.
- Maier HR, Dandy GC & Burch MD (1998). Use of artificial neural networks for modelling cyanobacteria anabaena spp. in the river murray, south australia. *Ecological Modelling*, 105(2-3):257–272.
- Martins JV, Tanré D, Remer L, Kaufman Y, Mattoo S & Levy R (2002). Modis cloud screening for remote sensing of aerosols over oceans using spatial variability. *Geophysical Research Letters*, 29(12):MOD4–1.
- Matthews MW (2014). Eutrophication and cyanobacterial blooms in south african inland waters: 10years of meris observations. *Remote Sensing of Environment*, 155:161–177.
- Mehnert G, Leunert F, Cirés S, Jöhnk KD, Rucker J, Nixdorf B & Wiedner C (2010). Competitiveness of invasive and native cyanobacteria from temperate freshwaters under various light and temperature conditions. *Journal of Plankton Research*, 32(7):1009–1021.
- Mekis É & Vincent LA (2011). An overview of the second generation adjusted daily precipitation dataset for trend analysis in canada. *Atmosphere-Ocean*, 49(2):163–177.
- MELCC (2000). *Portrait global de la qualité des eaux au Québec*. <http://www.mddep.gouv.qc.ca/eau/sys-image/global/global2.htmchloro>.
- MELCC (2002). *Bilan des lacs et cours d'eau touchés par les fleurs d'eau d'algues bleu-vert au Québec*. <http://www.mddep.gouv.qc.ca/eau/algues-bv/gestion/index.htm#resultats>. Ministère du Développement durable, de l'Environnement et des Parcs.
- MELCC (2011). *Les plans d'eau touchés par une fleur d'eau d'algues bleu-vert - Bilan final de la gestion des épisodes en 2010*. <http://collections.banq.qc.ca/ark:/52327/bs2062380>. Direction du suivi de l'état de l'environnement.

- MELCC (2013). *Les plans d'eau touchés par une fleur d'eau d'algues bleu-vert - Bilan final de la gestion des épisodes en 2012*. http://www.grobec.org/pdf/action/Bilan2012_aIgues_bleu-vert.pdf. Direction du suivi de l'état de l'environnement.
- MELCC (2014a). *Bilan de la gestion des épisodes de fleurs d'eau d'algues bleu-vert, de 2007 à 2012*. http://www.environnement.gouv.qc.ca/EAU/algues-bv/bilan/Bilan_ABV2007-2012.pdf. Direction du suivi de l'état de l'environnement.
- MELCC (2014b). *Bilan de la gestion des épisodes de fleurs d'eau d'algues bleu-vert en 2013 - Résultats pour les plans d'eau et les installations municipales de production d'eau potable*. <http://www.environnement.gouv.qc.ca/eau/algues-bv/bilan/saison2013/algues-bilan-2013.pdf>. Direction du suivi de l'état de l'environnement.
- MELCC (2014c). *Plan d'intervention détaillé sur les algues bleu-vert 2007-2017*. http://www.environnement.gouv.qc.ca/eau/algues-bv/plan_intervention2007-2017.pdf.
- MELCC (2015). *Algues bleu-vert: Gestion des épisodes de fleurs d'eau*. <http://www.environnement.gouv.qc.ca/eau/algues-bv/gestion/index.htm>.
- Mesinger F, DiMego G, Kalnay E, Mitchell K, Shafran PC, Ebisuzaki W, Jović D, Woollen J, Rogers E, Berbery EH *et al.* (2006). North american regional reanalysis. *Bulletin of the American Meteorological Society*, 87(3):343–360.
- Michalak AM (2016). Study role of climate change in extreme threats to water quality. *Nature News*, 535(7612):349.
- Michalak AM, Anderson EJ, Beletsky D, Boland S, Bosch NS, Bridgeman TB, Chaffin JD, Cho K, Confesor R, Daloğlu I *et al.* (2013). Record-setting algal bloom in lake erie caused by agricultural and meteorological trends consistent with expected future conditions. *Proceedings of the National Academy of Sciences*, 110(16):6448–6452.
- Mitchell TM *et al.* (1997). Machine learning. 1997. *Burr Ridge, IL: McGraw Hill*, 45(37):870–877.
- Mooij WM, Janse J, Domis LDS, Hülsmann S & Ibelings BW (2007). Predicting the effect of climate change on temperate shallow lakes with the ecosystem model pclake. *Shallow Lakes in a Changing World*, Springer, pages 443–454.
- Moss B (2012). Cogs in the endless machine: lakes, climate change and nutrient cycles: a review. *Science of the Total Environment*, 434:130–142.
- Muttill N & Chau Kw (2006). Neural network and genetic programming for modelling coastal algal blooms. *International Journal of Environment and Pollution*.
- Ndong M, Bird D, Nguyen-Quang T, De Boutray ML, Zamyadi A, Vinçon-Leite B, Lemaire BJ, Prévost M & Dorner S (2014). Estimating the risk of cyanobacterial occurrence using an index integrating meteorological factors: Application to drinking water production. *Water research*, 56:98–108.
- Nicolas J, Deschamps P, Loisel H & Moulin C (2005). *Algorithm Theoretical Basis Document, POLDER-2/Ocean Color/Atmospheric corrections*.
- Nordkvist K, Loisel H & Gaurier LD (2009). Cloud masking of seawifs images over coastal waters using spectral variability. *Optics express*, 17(15):12246–12258.

- Obenour DR, Gronewold AD, Stow CA & Scavia D (2014). Using a bayesian hierarchical model to improve lake erie cyanobacteria bloom forecasts. *Water Resources Research*, 50(10):7847–7860.
- of Canada Working Group SL *et al.* (2010). Soil landscapes of canada version 3.2. *Agriculture and Agri-Food Canada.(digital map and database at 1: 1 million scale)*.
- Oh HM, Ahn CY, Lee JW, Chon TS, Choi KH & Park YS (2007). Community patterning and identification of predominant factors in algal bloom in daechung reservoir (korea) using artificial neural networks. *ecological modelling*, 203(1-2):109–118.
- Onderka M (2007). Correlations between several environmental factors affecting the bloom events of cyanobacteria in liptovska mara reservoir (slovakia)—a simple regression model. *Ecological Modelling*, 209(2):412–416.
- Paerl HW, Hall NS & Calandrino ES (2011). Controlling harmful cyanobacterial blooms in a world experiencing anthropogenic and climatic-induced change. *Science of the Total Environment*, 409(10):1739–1745.
- Paerl HW & Huisman J (2009). Climate change: a catalyst for global expansion of harmful cyanobacterial blooms. *Environmental microbiology reports*, 1(1):27–37.
- Paerl HW & Paul VJ (2012). Climate change: links to global expansion of harmful cyanobacteria. *Water research*, 46(5):1349–1363.
- Palmer S, Odermatt D, Hunter P, Brockmann C, Présing M, Balzter H & Tóth V (2015). Satellite remote sensing of phytoplankton phenology in lake balaton using 10 years of meris observations. *Remote Sensing of Environment*.
- Peel MC, Finlayson BL & McMahon TA (2007). Updated world map of the köppen-geiger climate classification. *Hydrology and earth system sciences discussions*, 4(2):439–473.
- Pick FR (2016). Blooming algae: a canadian perspective on the rise of toxic cyanobacteria. *Canadian Journal of Fisheries and Aquatic Sciences*, 73(7):1149–1158.
- Pickett ST (1989). Space-for-time substitution as an alternative to long-term studies. *Long-term studies in ecology*, Springer, pages 110–135.
- Platnick S, Meyer KG, King MD, Wind G, Amarasinghe N, Marchant B, Arnold GT, Zhang Z, Hubanks PA, Holz RE *et al.* (2017). The modis cloud optical and microphysical products: Collection 6 updates and examples from terra and aqua. *IEEE Transactions on Geoscience and Remote Sensing*, 55(1):502–525.
- Pospisil T & Lee AB (2018). Rfcde: Random forests for conditional density estimation. *Department of Statistics Data Science, Carnegie Mellon University, Pittsburgh, PA 15289, USA*.
- Priedītis G, Smits I, Dāgis S, Paura L, Krumins J, Dubrovskis D *et al.* (2015). Assessment of hyperspectral data analysis methods to classify tree species. *Research for Rural Development. International Scientific Conference Proceedings (Latvia)*, Latvia University of Agriculture.
- Przytulska A, Bartosiewicz M & Vincent WF (2017). Increased risk of cyanobacterial blooms in northern high-latitude lakes through climate warming and phosphorus enrichment. *Freshwater biology*, 62(12):1986–1996.

- Ptacnik R, Andersen T & Tamminen T (2010). Performance of the redfield ratio and a family of nutrient limitation indicators as thresholds for phytoplankton n vs. p limitation. *Ecosystems*, 13(8):1201–1214.
- Racault MF, Le Quéré C, Buitenhuis E, Sathyendranath S & Platt T (2012). Phytoplankton phenology in the global ocean. *Ecological Indicators*, 14(1):152–163.
- Radhakrishna Rao C (1952). *Advanced statistical methods in biometric research*. A Division Of Macmillan Publishing Co, Inc New York; Collier-Macmillan Publishers; London.
- Ralston DK, Keafer BA, Brosnahan ML & Anderson DM (2014). Temperature dependence of an estuarine harmful algal bloom: Resolving interannual variability in bloom dynamics using a degree-day approach. *Limnology and oceanography*, 59(4):1112–1126.
- Rantajarvi E, Olsonen R, Hallfors S, Leppanen J & Raateoja M (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.
- Rao CR (1948). The utilization of multiple measurements in problems of biological classification. *Journal of the Royal Statistical Society. Series B (Methodological)*, 10(2):159–203.
- Rashidan K & Bird D (2001). Role of predatory bacteria in the termination of a cyanobacterial bloom. *Microbial Ecology*, 41(2):97–105.
- Ratté-Fortin C, Chokmani K & El-Alem A (2018). A novel algorithm of cloud detection for water quality studies using 250 m downscaled modis imagery. *International Journal of Remote Sensing*, pages 1–12.
- Recknagel F, Bobbin J, Whigham P & Wilson H (2002). Comparative application of artificial neural networks and genetic algorithms for multivariate time-series modelling of algal blooms in freshwater lakes. *Journal of Hydroinformatics*, 4(2):125–133.
- Recknagel F, French M, Harkonen P & Yabunaka KI (1997). Artificial neural network approach for modelling and prediction of algal blooms. *Ecological Modelling*, 96(1-3):11–28.
- Reichwaldt ES & Ghadouani A (2012). Effects of rainfall patterns on toxic cyanobacterial blooms in a changing climate: between simplistic scenarios and complex dynamics. *water research*, 46(5):1372–1393.
- Reichwein Zientek L & Thompson B (2006). Commonality analysis: Partitioning variance to facilitate better understanding of data. *Journal of Early Intervention*, 28(4):299–307.
- Reynolds C & Walsby A (1975). Water-blooms. *Biological reviews*, 50(4):437–481.
- Reynolds CS (1984). *The ecology of freshwater phytoplankton*. Cambridge University Press.
- Richter R (1996). A spatially adaptive fast atmospheric correction algorithm. *International Journal of Remote Sensing*, 17(6):1201–1214.
- Ricketts TH, Dinerstein E, Olson DM, Eichbaum W, Loucks CJ, Kavanaugh K, Hedao P, Hurley P, DellaSalla D, Abell R *et al.* (1999). *Terrestrial ecoregions of North America: a conservation assessment*. volume 1. Island Press.

- Robarts RD & Zohary T (1987). Temperature effects on photosynthetic capacity, respiration, and growth rates of bloom-forming cyanobacteria. *New Zealand Journal of Marine and Freshwater Research*, 21(3):391–399.
- Robinson W, Franz B, Patt F, Bailey S & Werdell P (2003). Masks and flags updates. *Algorithm updates for the fourth Sea-WiFS data reprocessing, NASA Technical Memorandum*, 206892:34–40.
- Roche M (1963). Hydrologie de surface: Gauthier-villars editeur. *Paris*, 431p.
- Rolland A, Bird DF & Giani A (2005). Seasonal changes in composition of the cyanobacterial community and the occurrence of hepatotoxic blooms in the eastern townships, québec, canada. *Journal of Plankton Research*, 27(7):683–694.
- Rolland DC, Bourget S, Warren A, Laurion I & Vincent WF (2013). Extreme variability of cyanobacterial blooms in an urban drinking water supply. *Journal of plankton research*, 35(4):744–758.
- Rosen BH, Davis TW, Gobler CJ, Kramer BJ & Loftin KA (2017). Cyanobacteria of the 2016 lake okeechobee and okeechobee waterway harmful algal bloom. US Geological Survey.
- Sapiano M, Brown C, Schollaert Uz S & Vargas M (2012). Establishing a global climatology of marine phytoplankton phenological characteristics. *Journal of Geophysical Research: Oceans (1978–2012)*, 117(C8).
- Sasaoka K, Chiba S & Saino T (2011). Climatic forcing and phytoplankton phenology over the subarctic north pacific from 1998 to 2006, as observed from ocean color data. *Geophysical Research Letters*, 38(15).
- Schwarz G *et al.* (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464.
- Seaburg KG, Parked BC, Wharton RA & Simmons GM (1981). Temperature-growth responses of algal isolates from antarctic oases1. *Journal of Phycology*, 17(4):353–360.
- Silva T, Vinçon-Leite B, Tassin B, Petrucci G, Seidl M, Lemaire B & Nascimento N (2011). An integrated approach for urban water modelling, linking a watershed hydrological model and a cyanobacteria dynamics model in urban lakes. *12th International Conference on Urban Drainage Proceedings*, Porto Alegre, Brazil.
- Silva TFdG (2014). *Suivi et modélisation de la dynamique des cyanobactéries dans les lacs urbains au sein de leur bassin versant*. Thèse de doctorat, Université Paris-Est.
- Simiyu B, Oduor S, Rohrlack T, Sitoki L & Kurmayer R (2018). Microcystin content in phytoplankton and in small fish from eutrophic nyanza gulf, lake victoria, kenya. *Toxins*, 10(7):275.
- Sklar M (1959). *Fonctions de répartition à n dimensions et leurs marges*. Université Paris 8.
- Steffen MM, Belisle BS, Watson SB, Boyer GL & Wilhelm SW (2014). Status, causes and controls of cyanobacterial blooms in lake erie. *Journal of Great Lakes Research*, 40(2):215–225.
- Steffensen DA (2008). Economic cost of cyanobacterial blooms. *Cyanobacterial harmful algal blooms: state of the science and research needs*, Springer, pages 855–865.
- Steinmetz F, Deschamps PY & Ramon D (2011). Atmospheric correction in presence of sun glint: application to meris. *Optics express*, 19(10):9783–9800.

- Takamura N, Kadono Y, Fukushima M, Nakagawa M & KIM BH (2003). Effects of aquatic macrophytes on water quality and phytoplankton communities in shallow lakes. *Ecological Research*, 18(4):381–395.
- Tang EP, Tremblay R & Vincent WF (1997). Cyanobacterial dominance of polar freshwater ecosystems: Are high-latitude mat-formers adapted to low temperature? 1. *Journal of Phycology*, 33(2):171–181.
- Taranu ZE, Gregory-Eaves I, Leavitt PR, Bunting L, Buchaca T, Catalan J, Domaizon I, Guilizzoni P, Lami A, McGowan S *et al.* (2015). Acceleration of cyanobacterial dominance in north temperate-subarctic lakes during the anthropocene. *Ecology Letters*, 18(4):375–384.
- Tarrant P & Neuer S (2009). Monitoring algal blooms in a southwestern u.s. reservoir system. *Eos, Transactions American Geophysical Union*, 90(5):38–39. DOI:10.1029/2009EO050002.
- Thackeray S, Jones I & Maberly S (2008). Long-term change in the phenology of spring phytoplankton: species-specific responses to nutrient enrichment and climatic change. *Journal of Ecology*, 96(3):523–535.
- Thomalla S, Fauchereau N, Swart S & Monteiro P (2011). Regional scale characteristics of the seasonal cycle of chlorophyll in the southern ocean. *Biogeosciences*, 8(10):2849–2866.
- Thompson B (1991). A primer on the logic and use of canonical correlation analysis. *Measurement and Evaluation in Counseling and Development*.
- Townhill BL, Tinker J, Jones M, Pitois S, Creach V, Simpson SD, Dye S, Bear E & Pinnegar JK (2018). Harmful algal blooms and climate change: exploring future distribution changes. *ICES Journal of Marine Science*, 75(6):1882–1893.
- Trishchenko A, Luo Y, Khlopenkov K & Park W (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. *Enhancing Resilience in a Changing Climate. Earth Sciences Sector Canada Centre for Remote Sensing (CCRS). Natural Resources Canada*.
- Trudgill D, Honek A, Li D & Van Straalen N (2005). Thermal time—concepts and utility. *Annals of Applied Biology*, 146(1):1–14.
- Van Tongeren O, Van Liere L, Gulati R, Postema G & Boesewinkel-De Bruyn P (1992). Multivariate analysis of the plankton communities in the loosdrecht lakes: relationship with the chemical and physical environment. *Restoration and Recovery of Shallow Eutrophic Lake Ecosystems in The Netherlands*, Springer, pages 105–117.
- Vincent LA, Wang XL, Milewska EJ, Wan H, Yang F & Swail V (2012). A second generation of homogenized canadian monthly surface air temperature for climate trend analysis. *Journal of Geophysical Research: Atmospheres*, 117(D18).
- Vollenweider RA (1982). *Eutrophication of waters: monitoring, assessment and control*. Organisation for Economic Co-operation and Development.
- Wagner C & Adrian R (2009). Cyanobacteria dominance: quantifying the effects of climate change. *Limnology and Oceanography*, 54(6part2):2460–2468.

- Wan H, Wang XL & Swail VR (2010). Homogenization and trend analysis of canadian near-surface wind speeds. *Journal of Climate*, 23(5):1209–1225.
- Wang L, Zhang Jy, Wang Xc, Peng Zr & Wang Y (2008). Multianalysis between chlorophyll-a and environmental factors in dianshan lake water. *Journal of Shanghai Fisheries University*, 1:010.
- Wang M & Shi W (2006). Cloud masking for ocean color data processing in the coastal regions. *Geoscience and Remote Sensing, IEEE Transactions on*, 44(11):3196–3105.
- Wang M, Son S & Shi W (2009). Evaluation of modis swir and nir-swir atmospheric correction algorithms using seabass data. *Remote Sensing of Environment*, 113(3):635–644.
- Wang X, Lu Y, He G, Han J & Wang T (2007). Multivariate analysis of interactions between phytoplankton biomass and environmental variables in taihu lake, china. *Environmental monitoring and assessment*, 133(1-3):243–253.
- Wang Y, Ma H, Sheng D & Wang D (2012). Assessing the interactions between chlorophyll a and environmental variables using copula method. *Journal of Hydrologic Engineering*, 17(4):495–506.
- Warren A (2011). *Suivi des cyanobactéries en milieu lacustre par fluorimétrie in vivo*. Thèse de doctorat, Université du Québec.
- Watkinson A, o’Neil J & Dennison W (2005). Ecophysiology of the marine cyanobacterium, *lyngbya majuscula* (oscillatoriaceae) in moreton bay, australia. *Harmful Algae*, 4(4):697–715.
- 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*, 38:68–75.
- Winder M, Berger SA, Lewandowska A, Aberle N, Lengfellner K, Sommer U & Diehl S (2012). Spring phenological responses of marine and freshwater plankton to changing temperature and light conditions. *Marine biology*, 159(11):2491–2501.
- Winder M & Schindler DE (2004). Climate change uncouples trophic interactions in an aquatic ecosystem. *Ecology*, 85(8):2100–2106.
- Winder M & Sommer U (2012). Phytoplankton response to a changing climate. *Hydrobiologia*, 698(1):5–16.
- Winter JG, DeSellas AM, Fletcher R, Heintsch L, Morley A, Nakamoto L & Utsumi K (2011). Algal blooms in ontario, canada: increases in reports since 1994. *Lake and Reservoir Management*, 27(2): 107–114.
- World Health Organization (2003). *Guidelines for safe recreational water environments: Coastal and fresh waters*. volume 1. World Health Organization.
- Wynne TT, Stumpf RP, Tomlinson MC & Dyble J (2010). Characterizing a cyanobacterial bloom in western lake erie using satellite imagery and meteorological data. *Limnology and Oceanography*, 55(5):2025–2036.
- Xia J, Du P, He X & Chanussot J (2014). Hyperspectral remote sensing image classification based on rotation forest. *IEEE Geoscience and Remote Sensing Letters*, 11(1):239–243.

- Yue S (2000). Joint probability distribution of annual maximum storm peaks and amounts as represented by daily rainfalls. *Hydrological Sciences Journal*, 45(2):315–326.
- Yue S & Rasmussen P (2002). Bivariate frequency analysis: discussion of some useful concepts in hydrological application. *Hydrological Processes*, 16(14):2881–2898.
- Zamyadi A, McQuaid N, Prévost M & Dorner S (2012). Monitoring of potentially toxic cyanobacteria using an online multi-probe in drinking water sources. *Journal of Environmental Monitoring*, 14(2): 579–588.
- Zhai L, Platt T, Tang C, Sathyendranath S & Walls RH (2011). Phytoplankton phenology on the scotian shelf. *ICES Journal of Marine Science: Journal du Conseil*, 68(4):781–791.
- Zhang L & Singh V (2007). Bivariate rainfall frequency distributions using archimedean copulas. *Journal of Hydrology*, 332(1):93–109.
- Zhang M, Duan H, Shi X, Yu Y & Kong F (2012). Contributions of meteorology to the phenology of cyanobacterial blooms: Implications for future climate change. *Water Research*, 46(2):442 – 452. DOI:<http://dx.doi.org/10.1016/j.watres.2011.11.013>.
- Zhang Ym (2007). The influence of lake hydrodynamics on blue algal growth. *CHINA ENVIRONMENTAL SCIENCE-CHINESE EDITION*-, 27(5):707.

Sixième partie

ANNEXES

Annexe A

Imputation des données manquantes

Plusieurs travaux portant sur l'étude des FEA ont réduit la fréquence de données manquantes en travaillant à une résolution temporelle plus faible. Notons par exemple l'utilisation de données satellitaires à une résolution temporelle de 8 jours (Cole *et al.*, 2012; Brody *et al.*, 2013) ou même mensuelle (Edwards & Richardson, 2004). D'autres ont utilisé des méthodes d'interpolation temporelle afin de limiter les données manquantes (Sasaoka *et al.*, 2011; Zhai *et al.*, 2011; Racault *et al.*, 2012; Sapiano *et al.*, 2012). Notons par ailleurs que certaines études n'ont montré peu ou pas de différence entre des données reconstruites et des données originales, par exemple pour ce qui est de la date d'initiation d'une FEA, (Brody *et al.*, 2013). Matthews (2014) ont déterminé que seuls les lacs montrant une couverture nuageuse maximale de 95 % de sa superficie étaient considérés dans la base de données, pour des lacs ayant des superficies variant entre 6.7 km^2 et 346.6 km^2 à une résolution spatiale entre 260 m et 290 m. Duan *et al.* (2014) ont pour leur part considéré dans leur base de données seulement les images MODIS du lac Taihu (Chine) possédant une couverture nuageuse maximale de 25 % de la superficie du lac.

Par ailleurs, Ben Aissia (2014) a présenté une revue des méthodes d'imputation de données dans un cadre univarié et multivarié. Les techniques univariées sont séparées en deux domaines, soit celui des séries chronologiques et celui de l'analyse fréquentielle. Dans le cas des séries chronologiques, les techniques d'imputation incluent l'utilisation (1) de la moyenne de la série (donnée manquante remplacée par la moyenne), (2) de la moyenne d'une partie de la série (donnée manquante remplacée par la moyenne d'un sous-groupe de données), (3) des modèles de séries temporelles permettant de déterminer le modèle et les paramètres correspondants et estimer la donnée manquante, (4) l'interpolation spatiale et/ou temporelle permettant d'interpoler 2 points, l'une directement avant et l'autre après la donnée manquante, (5) la régression permettant d'estimer les paramètres de la régression et par la suite la donnée manquante, (6) l'imputation hot deck permettant de remplacer la donnée manquante par une valeur provenant d'un cas similaire et (7) la pondération inverse à la distance permettant d'estimer la donnée manquante par interpolation spatiale en utilisant des poids sur les points voisins.

Annexe B

Portrait de la qualité de l'eau des plans d'eau du Québec

Dans son plan de gestion des épisodes de FEA, le MELCC présente un portrait de la qualité des plans d'eau selon par région administrative et par zone de gestion intégrée de l'eau. Le MDDELCC a subdivisé le Québec en 40 zones de gestion intégrée des ressources en eau (ZGIE), cette opération découlant du plan d'intervention sur les algues bleu-vert 2007-2017 (MELCC, 2014c) et ayant comme objectif de considérer tous ces bassins versants dans le processus de gouvernance de l'eau. Dans son rapport sur la gestion des FEA (MELCC, 2014a), le gouvernement fait état de 36 ZGIE touchées annuellement par les FEA sur l'ensemble des 40 ZGIE entre 2007 et 2012 (voir tableau B.1). Les zones les plus touchées sont les bassins versants des rivières l'Assomption, Saint-François, du Nord, Rouge-Petite-Nation-Saumon, et des Sept. À elles seules, ces zones correspondent à 45% de l'ensemble des plans d'eau touchés et à 41% de l'ensemble des FEA répertoriées.

Tableau B.1 – Nombre de plans d'eau touchés selon la zone de gestion intégrée de l'eau, de 2007 à 2012 (MELCC, 2014a).

ZGIE	Nombre de plan d'eau touchés	Nombre de FEA
1 Abitibi-Jamésie	11	22
2 Baie Missisquoi	5	15
3 Batuscan-Champlain	5	20
4 Bayonne	2	3
5 Bécancour	10	25
6 Capitale	2	8
7 Charlevoix-Montmorency	2	6
8 Châteauguay	2	6
9 Chaudière	16	26
10 Chêne, du	0	0
11 Côte-du-Sud	6	11
12 Duplessis	1	5
13 Etchemin	3	4
14 Gaspésie - Nord	0	0
15 Gaspésie - Sud	0	0
16 Haute-Côte-Nord	1	1
17 Jacques-Cartier	3	5
18 Kamouraska-L'Islet-du-Loup	1	2
19 L'Assomption	35	68
20 Lac-St-Jean	14	39
21 Lièvre, du	20	44
22 Loup-Yamachiche	6	9
23 Manicouagan	1	1
24 Maskinongé	5	16
25 Matapédia-Ristigouche	4	5
26 Mille-Îles	5	6
27 Nicolet	4	5
28 Nord, du	39	61
29 Nord-Est du Bas-Saint-Laurent	8	22
30 Richelieu	3	6
31 Rouge-Petite-Nation-Saumon	45	85
32 Saguenay	13	31
33 Sainte-Anne	4	4
34 Saint-François	34	80
35 Saint-Jean, Fleuve	7	19
36 Saint-Maurice	22	38
37 Sept, des	32	52
38 Témiscamingue	21	38
39 Vaudreuil-Soulanges	0	0
40 Yamaska	12	56

Annexe C

Types de données climatiques et physiographiques

- Données de stations

Une station météorologique est composée d'un ensemble de capteurs qui enregistrent et fournissent des données climatiques à une échelle locale. Dans la plupart des cas, la prise de données nécessite un observateur humain. Cette contrainte augmente le risque de données manquantes. De plus, lors de remplacement des instruments de mesures, on peut obtenir des « sauts » qui correspondent à des journées non-mesurées. Ces données peuvent également être affectées par la proximité de perturbations (chaleur urbaine, masse d'eau). Ainsi, les données observées permettent d'obtenir une information locale à partir d'instruments mais nécessitent la disponibilité de gens sur le terrain, une maintenance du site et des instruments de mesures. Environnement Canada fournit ces données en tableaux mensuels et permet un accès à de longues séries sur leur portail DAI. Agrometeo Québec fournit ces données en temps réel.

- Données homogénéisées

Le processus d'homogénéisation permet de combler les données manquantes d'une série de données brutes, et ainsi, d'obtenir une série de données continue. De plus, il permet de détecter et de retirer des données erronées dans la série. L'information reste locale et les tendances obtenues sont plus représentatives de la réalité. Les séries de précipitation homogénéisée sont fortement corrélées au pas de temps quotidien avec la série du pluviomètre automatique et ne contiennent pas de données manquantes. Cependant, peu de stations sont disponibles et peu d'information concernant l'incertitude des méthodes d'homogénéisation n'est disponible également. Environnement Canada est le principal fournisseur de ce type de données via les travaux de Wan *et al.* (2010)(vitesse du vent), Mekis & Vincent (2011)(précipitation) et Vincent *et al.* (2012) (température de l'air de surface).

- Données interpolées sur grille

Ces séries de précipitations montrent une très bonne corrélation avec les données du pluviomètre automatique mais elles ont de la difficulté à bien représenter les extrêmes. Le CRU produit plusieurs produits d'interpolations (CRUTEM, HadCRUT, HadSST, gu23wld, etc.) à une résolution temporelle mensuelle à partir de 1850 et à une résolution spatiale de +/- 500 km sur le globe. CANGRD

est une interpolation des données homogénéisées d'Environnement Canada à une résolution temporelle mensuelle à partir de 1900 et résolution spatiale de 50 km sur le territoire canadien. Ressource Naturelle Canada produit une grille interpolée sur le Canada avec la méthode ANUSPLIN à une résolution temporelle journalière à partir de 1950 et une résolution spatiale de 10 km. Ressource Naturelle Canada calcule plusieurs indicateurs bioclimatiques, comme le nombre de jours de la saison de croissance, la précipitation, la température moyenne, minimale et maximale.

- Réanalyse

Ces données sont issues d'un modèle climatique mais montrent une très bonne corrélation avec les séries de données quotidiennes d'une station automatique. La réanalyse consiste à utiliser un maximum de données climatiques observées (toutes les 6 heures) de manière satellitaire. Ces données servent à réajuster les données de la série brute. Les réanalyses modernes montrent une résolution spatiale pouvant aller jusqu'à 40 km et une résolution temporelle jusqu'aux heures. Elles utilisent un maximum de données observées pour initialiser le modèle plusieurs fois par jour.

- Modèle climatique régional

Les modèles climatiques régionaux peuvent être pilotés par une réanalyse: dans ce cas, les données sont issues d'un modèle numérique et montrent une très bonne corrélation avec les séries de données quotidiennes du pluviomètre automatique. Autrement dit, cette technique consiste à limiter les estimations du modèle climatique à une zone connue (appelée « tuile »). Cependant, pour initialiser le modèle, une réanalyse est effectuée sur les estimations et celle-ci se limite aux frontières de la tuile. Le reste des données obtenues sont issues du modèle régional qui demeure moins corrélé à une série de données brutes. Ils peuvent également être pilotés par un modèle climatique global: les données se retrouvent sur une grille relativement fine mais ne montrent aucune corrélation quotidienne avec les données du pluviomètre automatique.

- Modèle climatique global

Les modèles climatiques modernes possèdent une résolution temporelle jusqu'aux trois heures et résolution spatiale jusqu'à 100 km. Ces données se retrouvent sur une grille relativement grossière et ne montrent aucune corrélation quotidienne avec les données du pluviomètre automatique.

- Données physiographiques

Ressources Naturelles Canada permet l'accès à des bases de données géospatiales sur l'ensemble du Canada, et plus précisément, à des données d'élévation et des données topographiques. Énergie et Ressources Naturelles Québec rend également accessible une base de données hydrographiques couvrant l'ensemble du Québec. Natural Earth permet l'accès à des bases de données topographiques et du milieu physique, et ce, à différentes échelles spatiales pour l'ensemble de la planète. Une base de données des cultures et du territoire agricole à l'échelle du Québec méridional a été mise en place par La Financière agricole du Québec.

Annexe D

Corrélations de Pearson

Tableau D.1 – Analyse des corrélations de Pearson entre les variables phénologiques et les variables environnementales.

		Frequency	Intensity	Extent	Beginning
Morphologie lac	Area	0.09	0.21	-0.01	-0.02
	Perimeter	0.16	0.54	-0.1	-0.05
	Gravelius coefficient	0.22	0.5	-0.18	-0.11
	Length of the Gravelius's rectangle	0.16	0.54	-0.1	-0.05
	Width of the Gravelius's rectangle	0.15	0.25	-0.01	-0.03
Morphologie bassin	Area	0.16	0.27	0.06	-0.05
	Perimeter	0.18	0.37	-0.03	-0.06
	Gravelius coefficient	0.21	0.34	-0.12	-0.12
	Length of the Gravelius's rectangle	0.18	0.37	-0.03	-0.06
	Width of the Gravelius's rectangle	0.21	0.46	-0.04	-0.05
	Slope - mean	0.1	0.04	0.09	-0.04
	Slope - standard deviation	-0.24	-0.09	-0.14	0.22
Physiographie	Land cover - Forest	0.16	0.28	0.04	-0.05
	Land cover - Settlement	0.17	0.13	0.16	-0.07
	Land cover - Cropland	0.2	0.16	0.21	-0.08
	Land cover - Forest (relative)	-0.33	-0.1	-0.21	0.17
	Land cover - Settlement (relative)	0.29	0.07	0.16	-0.22
	Land cover - Cropland (relative)	0.3	0.1	0.2	-0.13
	Population ecumene	0.32	0.06	0.25	-0.3
	Agriculture ecumene	0.28	0.09	0.19	-0.14
Climat	Total precipitation - annual	-0.23	-0.43	0.02	0.11
	Total precipitation - summer	-0.22	-0.35	0.01	0.09
	Mean temperature - annual	-0.13	-0.47	0.03	0
	Mean temperature - summer	-0.13	-0.47	0.03	0.01
	Wind speed - annual	0.02	0.02	0.08	-0.04
	Wind speed - summer	0.09	0.06	0.08	-0.08
	Degree days	0.61	0.16	0.35	-0.55

Tableau D.2 – Corrélations de Pearson entre les variables environnementales (gauche) et entre les variables phénologiques (droite).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1 L Area	1																										
2 L Perimeter	0.8	1																									
3 L Shape index	0	0.8	1																								
4 L Length	0	0.4	0	1																							
5 L Width	0	0.5	0	0.4	1																						
6 W Area	0	0.5	0.3	0.5	0.6	1																					
7 W Perimeter	0	0.4	0.6	0.4	0.6	0.9	1																				
8 W Shape index	0	0.4	0.6	0.4	0.2	0.4	0.7	1																			
9 W Length	0	0.5	0.7	0.5	0.9	0.9	0.9	0.9	1																		
10 W Width	0	0.4	0.1	0.4	0.4	0.4	0.4	0.4	0.4	1																	
11 Slope - mean	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1																
12 Slope - std	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1															
13 Forest	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1														
14 Settlement	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1													
15 Cropland	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1												
16 Forest (rel)	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1											
17 Settlement (rel)	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1										
18 Cropland (rel)	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1									
19 Pop ecumene	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1								
20 Agr ecumene	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1							
21 Pcp - annual	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1						
22 Pcp - summer	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1					
23 Temp - annual	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1				
24 Temp - summer	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1			
25 Wind - annual	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1		
26 Wind - summer	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	1
27 Degree days	0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1

Annexe E

Résultats supplémentaires

Cet annexe présente les résultats obtenues pour les deux variables phénologiques qui ne sont pas présentées dans le document, c'est-à-dire l'étendue et la date d'initiation des FEA.

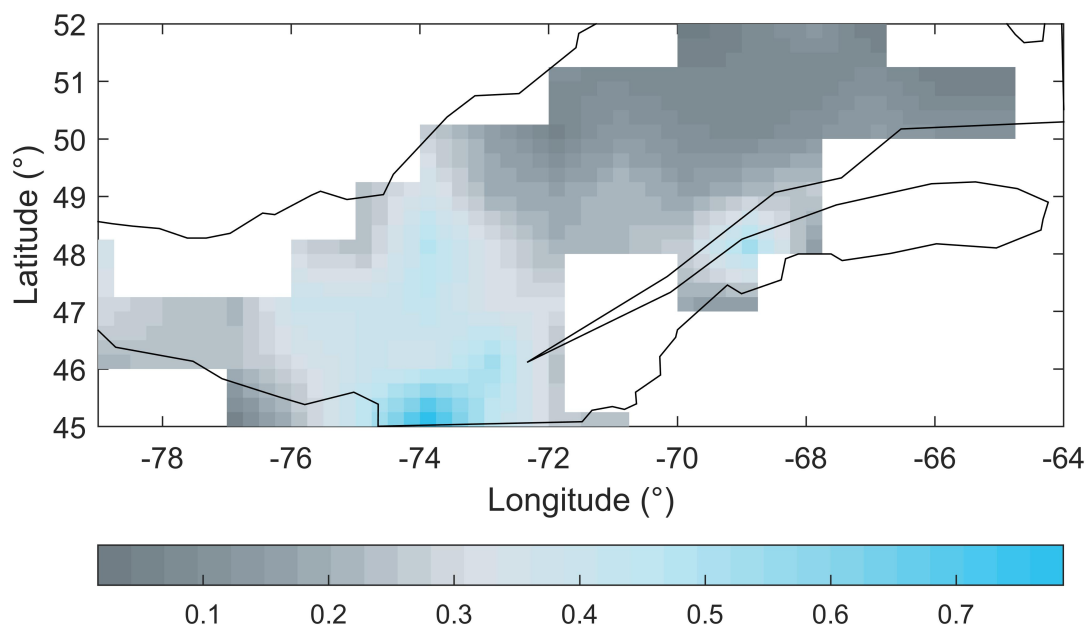


Figure E.1 – Répartition spatiale de l'étendue des FEA sur la région d'étude.

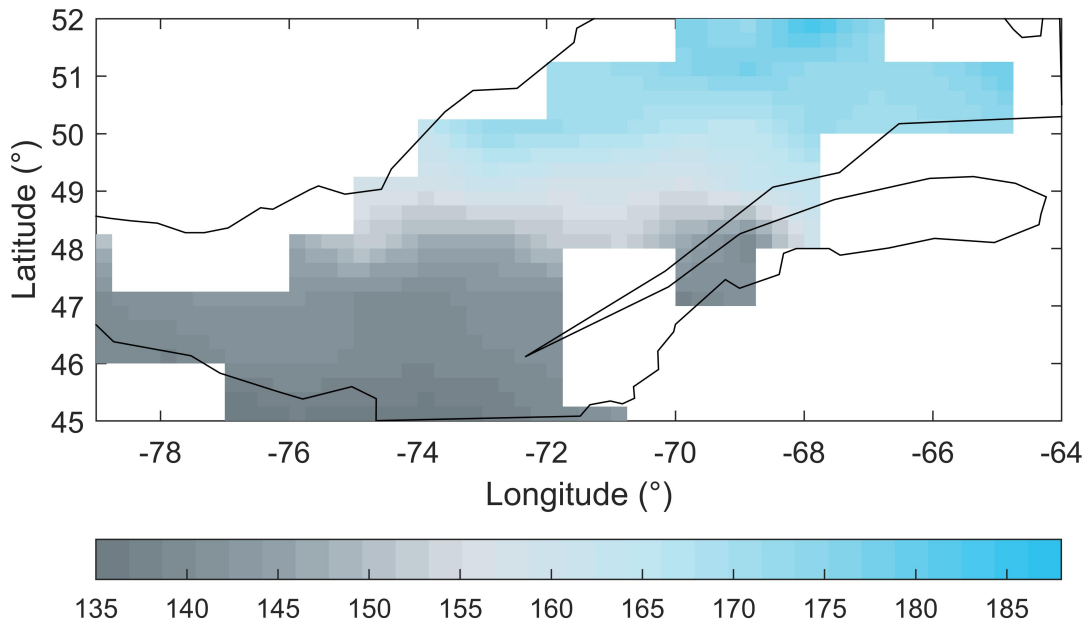


Figure E.2 – Répartition spatiale de la date de début des FEA sur la région d'étude.

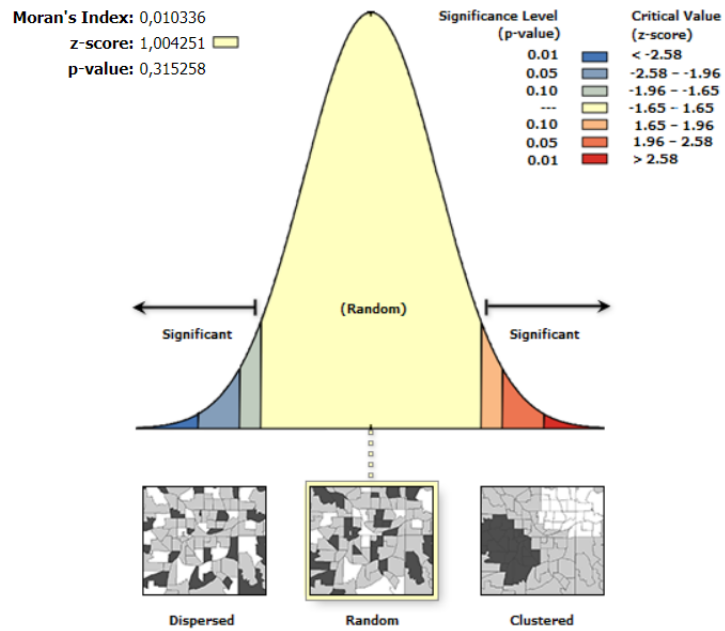


Figure E.3 – Test d'autocorrélation spatiale sur l'étendue maximale.

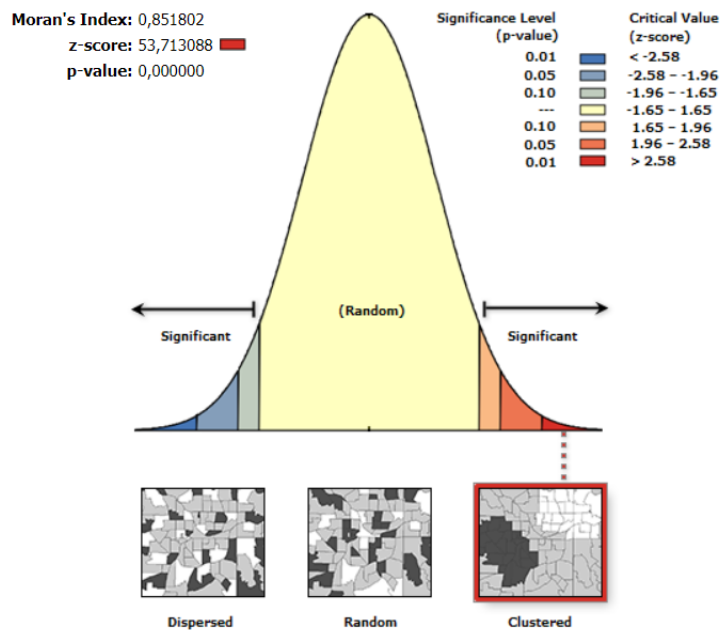


Figure E.4 – Test d'autocorrélation spatiale sur la date de début annuel d'épisodes.

Tableau E.1 – Variables phénologiques et leur corrélation avec les variables environnementales.

Étendue	Début
Degrés-jour	Degrés-jour
(R=0.35)	(R=-0.55)
	Urbanisation
	(R=-0.30)

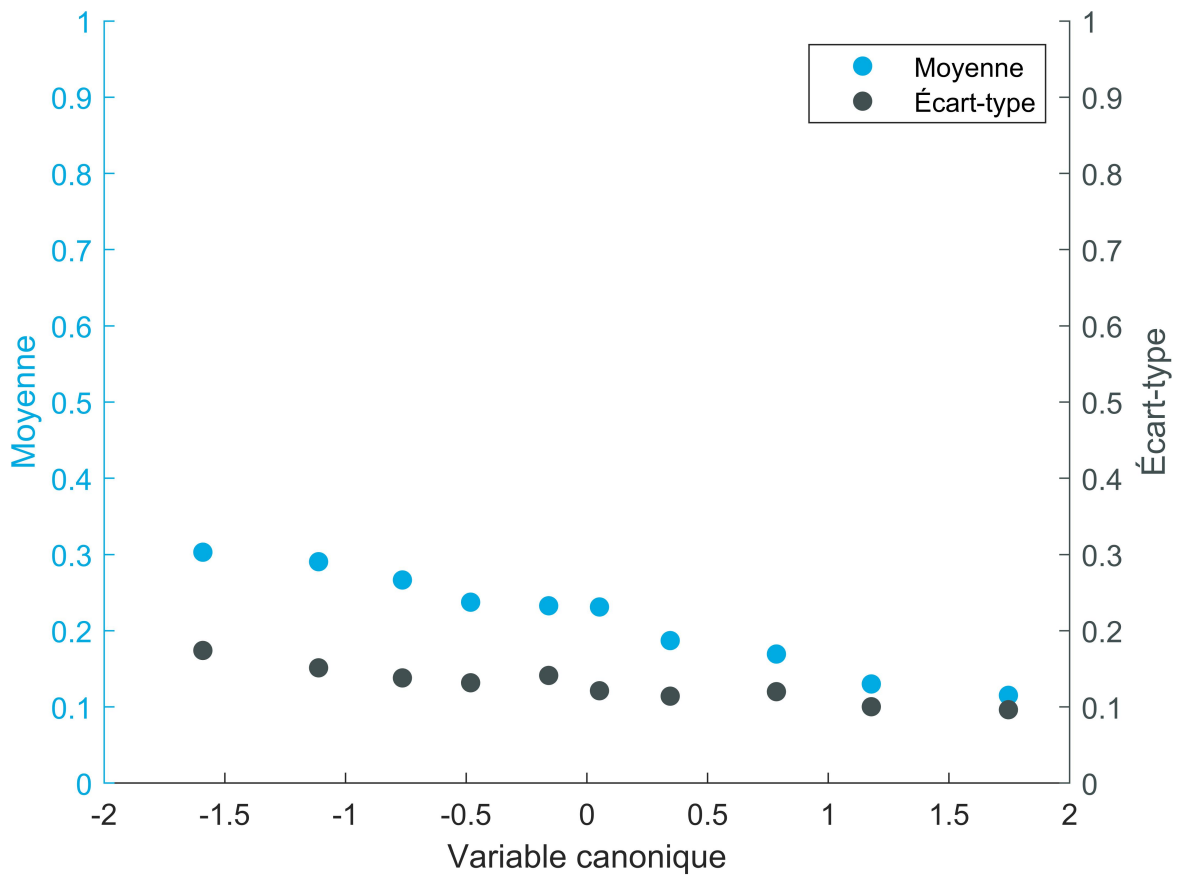


Figure E.5 – Moyennes et écart-types de la variation de l'étendue de FEA pour chaque classe de valeurs de la covariable.

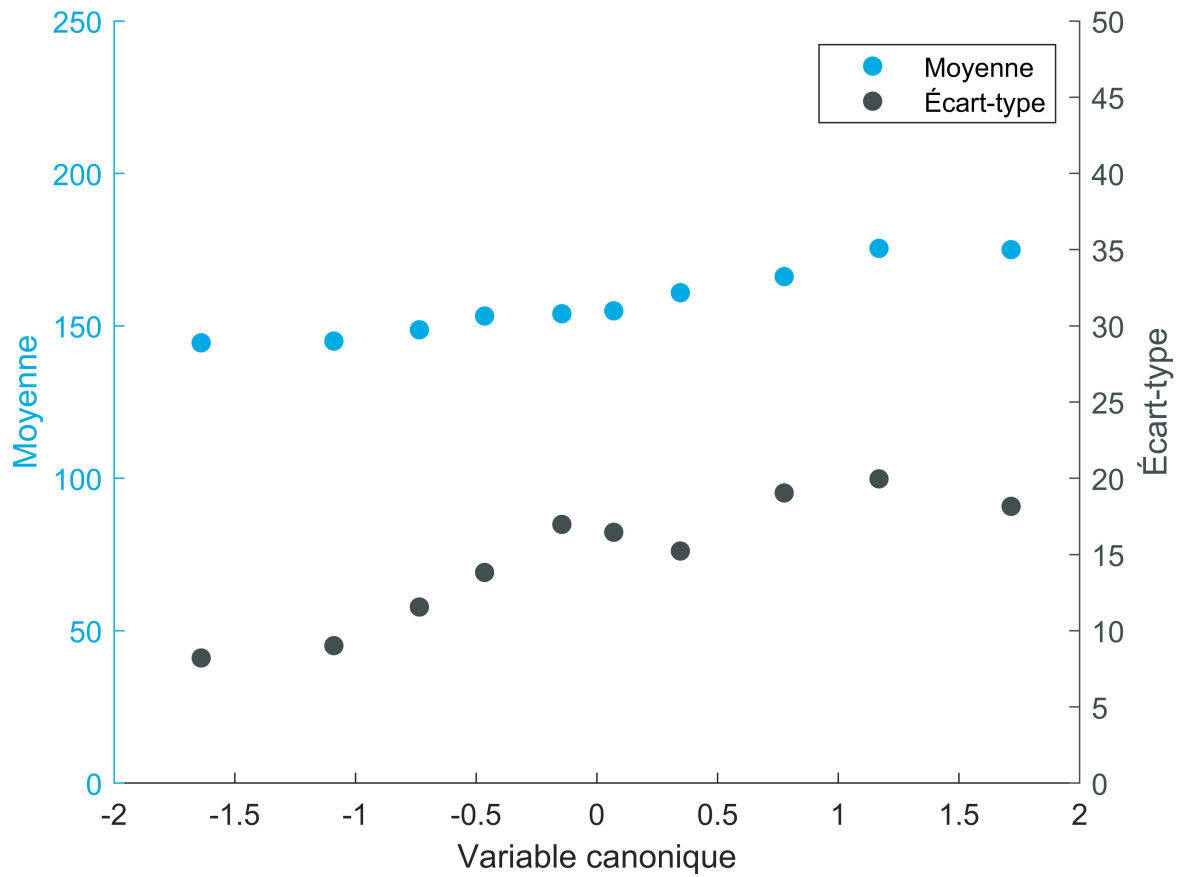


Figure E.6 – Moyennes et écart-types de la variation de date d’initiation annuelle de FEA pour chaque classe de valeurs de la covariable.