



Université de Montréal

**Modélisation de répartition d'espèces aviaires et de feux en forêt boréale du Québec dans un  
contexte de changement climatique**

Par

Jonathan Gaudreau

Département de Géographie

Faculté des Arts et des Sciences

Mémoire présenté à la Faculté des Arts et des Sciences

en vue de l'obtention du grade de Maîtrise

en géographie

18 Décembre 2015

©Jonathan Gaudreau, 2015

Université de Montréal  
Faculté des études supérieures et postdoctorales

Ce mémoire intitulé :

Modélisation de répartition d'espèces aviaires et de feux en forêt boréale du  
Québec dans un contexte de changement climatique

Présenté par :

Jonathan Gaudreau

évalué par un jury composé des personnes suivantes :

Thora Herrmann Martina, présidente-rapporteuse  
Liliana Pérez, directrice de recherche  
Pierre Drapeau, co-directeur  
Julie Talbot, évaluatrice externe

## Résumé

Les changements climatiques prennent une importance grandissante dans l'étude des phénomènes spatiaux à grande échelle. Plusieurs experts affirment que les changements climatiques seront un des principaux moteurs de changement écologique dans les prochaines décennies et que leurs conséquences seront inévitables. Ces changements se manifesteront sur le milieu physique par la fonte des calottes glaciaires, le dégel du pergélisol, l'instabilité des versants montagneux en zone de pergélisol, l'augmentation de l'intensité, de la sévérité et de la fréquence des événements climatiques extrêmes tels les feux de forêt. Les changements climatiques se manifesteront aussi sur le milieu biologique, tel la modification de la durée de la saison végétative, l'augmentation des espèces exotiques invasives et les changements dans la distribution en espèces vivantes. Deux aspects sont couverts par cette étude : 1) les changements dans la répartition spatiale de 39 espèces d'oiseaux et 2) les modifications dans les patrons spatiaux des feux, en forêt boréale québécoise, tous deux dans l'horizon climatique de 2100. Une approche de modélisation statistique démontre que la répartition spatiale des oiseaux de la forêt boréale est fortement liée à des variables bioclimatiques ( $R^2_{adj} = 0.53$ ). Ces résultats permettent d'effectuer des modélisations bioclimatiques pour le gros-bec errant et la mésange à tête noire qui voient une augmentation de la limite nordique de distribution de l'espèce suivant l'intensité du réchauffement climatique. Finalement, une modélisation spatialement explicite par automate cellulaire permet de démontrer comment les changements climatiques induiront une augmentation dans la fréquence de feux de forêt et dans la superficie brûlée en forêt boréale du Québec.

**Mots-clés** : Écologie, analyse multivariée, biogéographie, Québec, changements climatiques, feux, automates cellulaire, modélisation, systèmes complexes, forêt boréale

## Abstract

Climate is getting more important in the study of broad-scale spatial phenomena. Experts agree on the fact that climate change will likely be one of the main drivers of ecological change in the upcoming decade and that its consequences are inevitable. These changes induce consequence on the physical aspects, by ice caps melting, permafrost thawing, slope and soil instability in mountainous permafrost areas and increase of intensity, severity and frequency of extreme weather events such as wildland fires. Moreover, climate change also causes impacts on the biological aspects, with modification in the growing season, increase in exotic invasive species, and changes in spatial distribution of butterfly and bird species. Two of these aspects are covered in this study: 1) changes in the spatial distribution of 39 bird species and 2) modifications in spatial patterns of wildland fires, in the boreal forest of Quebec, up to year 2100. A statistical modelling approach showed that the spatial distribution of boreal birds is strongly linked to bioclimatic variables (53%). These results enables the bioclimatic modelling of two bird species, the evening grosbeak and black-capped chickadee. In both cases, the model shows an increase of the northernmost limit of distribution of each species following the rate of climate change. Finally, a spatially-explicit cellular automata model showed that climate change will likely induce an increase in wildfire frequency and total area burned for the boreal forest of Quebec.

**Keywords:** Ecology, multivariate analysis, biogeography, Quebec, climate change, wildfires, cellular automata, modelling, complex systems, boreal forest

# Table des matières

Résumé .....	iii
Abstract.....	iv
Liste des tableaux .....	viii
Liste des figures .....	ix
Lexique et abréviations.....	x
Remerciements.....	xii
Introduction .....	1
Fondements .....	3
Réponses écologiques aux changements climatiques.....	3
Méthodes de modélisation.....	4
Introduction aux systèmes complexes .....	4
Modélisation basée sur l'agent.....	5
Les automates cellulaires.....	6
Bibliographie.....	7
Premier article .....	9
Présentation de l'article.....	9
Accord des coauteurs .....	9
Identification des variables expliquant la distribution spatiale d'oiseaux de la forêt boréale et modélisation de tendances futures: une approche multivariée .....	10
Introduction .....	12
Matériel et préparation des données.....	13
Aire d'étude et sites échantillonnés .....	13
Variables réponses: abondances d'oiseaux .....	14
Variables explicatives.....	15
Méthodes.....	18
Analyses multivariées pour identifier les variables responsables de la répartition spatiale des oiseaux de la forêt boréale du Québec.....	18
Modélisation bioclimatique pour modéliser les distributions d'espèces.....	21
Résultats .....	24
Sélection de variables .....	24
Partitionnement de variation .....	28

Modélisation bioclimatique .....	29
Discussion .....	33
Conclusion.....	35
Remerciements.....	35
Codage des espèces d’oiseaux :.....	36
Codage des variables bioclimatiques (Hijmans <i>et al.</i> , 2005) .....	38
Codage des autres variables .....	39
Bibliographie .....	40
Deuxième article.....	44
Présentation de l’article.....	44
Accord des coauteurs .....	44
<i>BorealFireSim</i> : A GIS-based Cellular Automata Model of Wildfires for the Boreal Forest of Quebec in a Climate Change Paradigm.....	46
Abstract.....	46
Highlights .....	46
Keywords: .....	46
Introduction .....	47
Material and methods .....	50
Study Area.....	52
Model input data .....	53
Model behavior.....	57
Model implementation.....	61
Model outputs .....	62
Contrasting <i>BorealFireSim</i> outputs.....	63
Results and discussion .....	64
Resulting maps.....	64
Summary of the model’s outputs .....	68
Validation process.....	71
3.4. Contrasting model outputs with DHI forecasts .....	72
Assumptions and limitations of <i>BorealFireSim</i> .....	74
Conclusion.....	74
Acknowledgments .....	75
References .....	76

Conclusion générale.....80



## Liste des tableaux

Tableau I. Algorithmes de classification et de régression utilisés pour la génération des modèles de distribution d'espèces.....	23
Tableau II. Résultat du test de permutation sur les prédicteurs .....	27
Tableau III. Résultats du test de permutation sur les axes canoniques .....	28
Table IV. Analogues between new and previous climate scenarios (Snover et al, 2013) .....	55
Table V. Cell states and possible transitions.....	58
Table VI. Scores given to the cells for the algorithms.....	60
Table VII. Increase in mean area burned annually, RCP4.5 and RCP8.5.....	69
Table VIII. Increase in Mean Burning chance of cells .....	70

## Liste des figures

Figure 1. Sites échantillonnés. ....	14
Figure 2. Tendances dans l'évolution des concentrations de GES selon les 4 profils d'évolution de la concentration.....	17
Figure 3. Diagramme de Venn pour partitionnement de la variation de Y en fonction de deux groupes de variables explicatives. ....	21
Figure 4. Adapté de Berteaux et al., 2010. Représentation simplifiée de la méthode de modélisation bioclimatique utilisée.....	24
Figure 5. Résultats de la RDA sur les variables non-colinéaires .....	26
Figure 6. Résultats de la RDA sur les variables sélectionnées .....	26
Figure 7. Partitionnement de la variation des données des 39 espèces d'oiseaux (log) par deux groupes de variables explicatives : les variables bioclimatiques et les variables ressource .....	29
Figure 8. Distribution actuelle et prévue de la mésange à tête noire pour 2070 selon les quatre scénarios RCP .....	31
Figure 9. Distribution actuelle et prévue du gros-bec errant pour 2070 selon les quatre scénarios RCP .....	32
Figure 10. Flowchart of the BorealFireSim model .....	51
Figure 11. Study Area.....	52
Figure 12. Components of the Fire Weather Index (FWI) .....	54
Figure 13. Trends in the evolution of concentration of greenhouse gases for the four RCP scenarios. ....	56
Figure 14. Neighborhood representation used by the BorealFireSim.....	58
Figure 15. Graphical User Interface of BorealFireSim .....	62
Figure 16. Burning chance in 2050 and 2080 for the scenarios RCP4.5 and RCP8.5 .....	64
Figure 17. Cells state after 200 repetitions.....	66
Figure 18. Change in burning chance after 200 repetitions .....	67
Figure 19. BorealFireSim forecast area burned (square-km) .....	69
Figure 20. Overlap of historical fires with the cell state map for present (2010) .....	72
Figure 21. Combination of DHI and burning chance forecasts for 2080 .....	73

## Lexique et abréviations

ABM	<i>Agent-Based Model</i> , méthode de modélisation basée sur l'agent ou sur l'individu
AUC	Aire sous la courbe, ou <i>Area Under the Curve</i> . Sert à la validation de modèles statistiques.
BAU	<i>Business As Usual</i> : expression utilisée pour qualifier les modèles d'émissions de gaz à effet de serre reposant sur une augmentation constante des émissions
CA	Automate cellulaire, méthode de modélisation itérative reposant sur une grille et un voisinage
FWI	Indice Canadien de sévérité des feux ( <i>Fire Weather Index</i> )
GES	Gaz à effet de serre
GIEC/IPCC	Groupe d'Experts Intergouvernemental sur l'Évolution du Climat/ <i>Intergovernmental Panel on Climate Change</i>
MERN	Ministère de l'Environnement et des Ressources Naturelles du Québec
MFFP	Ministère de la Faune, des Forêts et des Parcs du Québec
<i>Raster</i>	Couche de données géographiques de type matriciel, tel une image satellitaire, un modèle numérique d'élévation ou une carte des précipitations
RCP	Profils d'évolution de la concentration en gaz à effet de serre. Il s'agit de la nouvelle famille de scénarios climatiques formulées par le GIEC en 2013, remplaçant les scénarios « SRES » ou « Special Report on Emission Scenarios ».

RDA	Analyse de redondance canonique (méthode statistique multivariée) servant à identifier des liens entre des variables réponses et variables explicatives
Vecteur	Couche de données géographiques de type vectoriel, c'est-à-dire représentant des entités discrètes (points, lignes, polygones)
VIF	Facteur d'inflation de la variance, sert à détecter la colinéarité dans un ensemble de données.
WGS84	Système de coordonnées géodésiques mondial de 1984, ou <i>World Geodesic System 1984</i> . C'est le système de coordonnées utilisé habituellement pour tout ce qui traite d'un continent ou d'une très grande région d'étude. Les coordonnées sont en degrés décimaux, utilisant donc un système sphérique.

*Je dédie ce mémoire à ma femme, Joanie, qui a supporté mes nombreuses soirées d'absence à programmer au Second Cup et qui n'en peut plus de m'entendre parler de modélisation, et à mes parents qui, ne sachant toujours pas « ce qu'on fait en géologie [sic] », m'ont toujours soutenu et poussé à viser les sommets!*

## Remerciements

J'aimerais remercier ma directrice de recherche, Liliana, qui m'a donné l'opportunité d'effectuer un virage à 180 en me donnant de judicieux conseils pour m'aider à me tailler une place en géomatique et en modélisation. Je remercie aussi chaleureusement mes grands amis : Catalina, Julie, Lisa, Guillaume et tous les autres que ne peux pas mentionner. Je tiens aussi à remercier tous les chercheurs qui m'ont consacré un peu de leur temps à diverses étapes de ma recherche: Pierre Legendre, Pierre Drapeau, François Girard, Aurélie Terrier, Sylvie Gauthier, Christopher Bone, sans oublier le personnel technique du Centre d'Études de la Forêt (CEF): Luc Lauzon, Mélanie Desrochers et Daniel Lesieur. Finalement, je voudrais remercier la Fédération des Associations Étudiantes du Campus de l'Université de Montréal (FAÉCUM) par son programme PARSECS, le département de Géographie, le CEF, Liliana Pérez, ESRI Canada et la Faculté des Études Supérieures et Postdoctorales qui m'ont tous soutenus financièrement, à différents degrés, durant ces trois années.

## Introduction

La géographie, sous toutes ses disciplines, s'intéresse de près à l'évolution du climat et aux répercussions des potentiels changements sur les grands systèmes qui composent la terre, soient la biosphère, la lithosphère, la cryosphère, l'hydrosphère et l'atmosphère. Les géographes, de toutes spécialisations, mettront l'accent sur l'interaction entre ces cinq sphères, l'espèce humaine et les phénomènes terrestres à macro ou à micro échelles. La division essentiellement « géomatique et technique » de la géographie, porte avant tout, mais non exclusivement, sur les méthodes computationnelles servant à représenter ces phénomènes. C'est particulièrement le cas en modélisation de systèmes complexes, où les phénomènes terrestres sont considérés dans leur ensemble, c'est-à-dire un ensemble d'interactions entre les domaines physiques, biologiques, anthropiques et parfois économiques. C'est dans cet esprit holistique que s'inscrit cette recherche où deux thèmes, ou phénomènes, sont mis de l'avant au travers de deux articles scientifiques portant sur un aspect des changements climatiques prévus. Dans un premier temps, la présente recherche vise à étudier les réponses de la faune aviaire aux changements climatiques, via l'utilisation de données d'abondances d'espèces aviaires, de méthodes géomatiques, statistiques et de modélisation bioclimatique et, dans un second temps, à se questionner sur l'évolution des patrons spatiaux des feux en forêt boréale du Québec, par le biais d'une méthode issue de la théorie des systèmes complexes, soit la modélisation par automate cellulaire.

La recherche proposée vise à examiner comment les changements climatiques sont passibles d'influencer la répartition spatiale d'espèces d'oiseaux dans la forêt boréale du Québec et de patrons spatiaux des feux de forêt dans un horizon climatique compris entre 2010 et 2100. Celle-ci sera réalisée via la combinaison des connaissances en écologie animale, en sciences géomatiques et en théorie des systèmes complexes. Nous émettons l'hypothèse que la distribution spatiale des oiseaux de la forêt boréale et le risque d'incendies en forêt boréale seront fortement corrélé au climat. Ainsi, les changements climatiques prévus induiront une migration des oiseaux vers de plus hautes latitudes, ou de plus hautes altitudes dans le cas de régions montagneuses, ainsi qu'une intensification des événements de feux plus importante au Nord qu'au Sud de la province. Ce mémoire se compose de deux articles scientifiques comprenant tous deux une revue de littérature et une description méthodologique destinés à

faciliter la compréhension lors de la lecture par une audience généraliste. Le premier article a été publié dans une revue européenne généraliste en géographie (CyberGéo) tandis que le second sera publié en Mars 2016 dans une revue d'informatique et d'écologie (*Ecological Informatics*) s'adressant à un public varié. En ce sens, seuls les aspects non couverts par les deux articles apparaissent dans la section *Fondements*, c'est-à-dire une revue de littérature générale des réponses écologiques aux changements climatiques ainsi qu'une brève introduction aux méthodes de modélisation issues des systèmes complexes. Les sujets étant bien distincts, chaque section comporte sa propre bibliographie. Finalement, les deux articles sont suivis d'une conclusion générale.

## Fondements

Cette section consiste en une revue de littérature couvrant les principaux thèmes propres à cette recherche. Tout d'abord, l'aspect écologique sera traité, avec une brève revue de littérature des réponses écologiques aux changements climatiques. Par la suite, une seconde partie portera principalement sur les méthodes de modélisation de systèmes complexes utilisées dans le cadre de cette recherche, soit la modélisation bioclimatique et la modélisation par automate cellulaire. Certains éléments ne sont pas couverts dans cette partie puisqu'ils sont couverts largement à l'intérieur des articles scientifiques, tel que les méthodes de statistiques multivariées présentes dans le premier chapitre ou les dynamiques de feux de forêts présentées dans le second chapitre.

## Réponses écologiques aux changements climatiques

La très importante revue de littérature dressée par Parmesan (2006), fait état des multiples conséquences des changements climatiques anthropiques sur la faune et sur la flore. Ces conséquences sont regroupées par l'auteur sous trois catégories. La première grande catégorie traite des changements phénologiques, notamment avec la modification de la durée de la saison végétative en Amérique du Nord, les changements majeurs dans la période de floraison des cerisiers au Japon et la récolte de plus en plus précoce dans les vignobles européens (Parmesan, 2006). La seconde catégorie de changements écologiques comprend les altérations au niveau trophique dus à l'adaptation à des vitesses différentes entre les espèces, puisque la végétation s'adapte habituellement plus rapidement aux changements climatiques que les animaux. Finalement, la troisième catégorie, la plus pertinente dans le cadre de cette recherche, décrit les changements dans l'étendue spatiale et dans l'abondance d'espèces. Forister *et al.* (2010), traitant des effets des changements climatiques sur les altérations dans les habitats de papillons, concluent que les espèces spécialisées sont très souvent incapables de s'adapter aux changements climatiques rapides, de par leur dépendance à certains éléments de leur environnement, qui sont aussi affectés par les changements climatiques. Bien que les espèces alpines à l'étude subissaient de très fortes réductions démographiques suivant le rythme du réchauffement et de l'assèchement de l'habitat, l'accroissement de la diversité des milieux de haute altitude a entraîné une augmentation dans l'abondance et dans la richesse en espèces de papillons non-alpins (Forister *et al.*, 2010). Dans un même ordre d'idées, l'étude d'Hurlbert et



Liang (2012) portant sur la variation spatiotemporelle dans la migration aviaire conclue que les espèces d'oiseaux spécialistes seront davantage vulnérables face aux changements climatiques que les espèces généralistes. En effet, en plus d'affecter leur répartition spatiale, les changements climatiques affectent aussi les parcours migratoires des oiseaux. Les espèces généralistes ont ainsi une meilleure réponse phénologique puisqu'elles ne dépendent pas d'éléments uniques à un milieu. Hurlbert et Liang (2012) sont également arrivés à la conclusion que la variable influençant le plus la réponse aux changements climatiques chez les oiseaux migrateurs était la vitesse de migration; les oiseaux rapides pouvant bénéficier d'une plus grande flexibilité que les migrateurs lents en cas d'événements climatiques extrêmes (Hurlbert and Liang, 2012). Dans une étude similaire réalisée en 2009, Lawler *et al.* ont fait une modélisation des changements d'espèces pour l'entièreté de l'Hémisphère Ouest, combinant les données de 3576 espèces d'oiseaux, de 1561 espèces de mammifères et de 1616 espèces d'amphibiens. Ils concluent qu'il faudra prévoir d'ici 2100 une perte nette de 11% d'espèces, selon le scénario de changement climatique B1 (diminution nette des émissions de GES) et une perte nette de 17% selon le scénario A2 (scénario « business as usual », avec augmentation constante des émissions de GES). Leurs résultats ont été calculés via l'algorithme *RandomForest*, qui est utilisé dans la première partie de cette recherche (Lawler *et al.*, 2009; Breiman, 2001).

## Méthodes de modélisation

### Introduction aux systèmes complexes

La question des changements climatiques, afin d'être adressée de façon adéquate, doit être considérée dans toute sa complexité. Ainsi, les méthodes présentées sont issues de la théorie de la complexité. Avant de traiter de complexité, il est bien important de définir celle-ci. On compte trois types de complexité, soit 1) la *complexité algorithmique*, 2) la *complexité déterministe* et 3) la *complexité globale* (O'Sullivan, 2004). La *complexité algorithmique* fait référence à la difficulté à résoudre des problèmes mathématiques et informatiques et est utilisée pour quantifier la quantité d'information dans des données. La *complexité déterministe*, réfère aux comportements imprévisibles des systèmes déterministes et est souvent résumée dans la culture populaire à la "théorie du chaos" ou à "l'effet papillon", signifiant qu'un battement des ailes d'un papillon à l'Ouest des États-Unis pourrait générer une tempête à Londres. Finalement, la *complexité globale*, la plus importante en géographie et en écologie, est l'étude des

phénomènes caractérisés par les interactions entre plusieurs composantes d'un système et entre les composantes et le système. Dans un système complexe, les interactions entre les composantes du système et entre le système et son environnement ne peuvent pas être expliquées uniquement par la somme des composantes. Les quelques caractéristiques des dynamiques complexes suivantes peuvent en expliquer la raison. L'*auto-organisation* et l'*émergence* signifient qu'un système possède l'habilité à s'organiser par lui-même, sans directives ou intervention externes. Les systèmes sont tous différents, allant des systèmes hautement structurés et ordonnés aux systèmes complètement chaotiques et aléatoires. La plupart des systèmes se situent entre ces deux extrêmes, étant structurés tout en démontrant de la non-linéarité, des équilibres dynamiques et des boucles de rétroaction (O'Sullivan, 2004). L'*auto-organisation*, en comparaison à la *théorie du chaos*, présuppose que de multiples conditions initiales différentes tendent à converger vers des résultats similaires, alors que la *théorie du chaos* présuppose que des changements infinitésimaux dans un système vont mener à une grande différenciation. Un autre concept pertinent dans l'étude des systèmes complexes est la *dépendance au sentier* ou *path-dependence*, signifiant que l'état actuel d'un système est influencé par ses états passés.

Parmi les méthodes utilisées en modélisation de systèmes complexes, les deux plus pertinentes à l'étude des phénomènes à caractère spatiotemporel sont la modélisation multi-agents ou *agent-based modelling* (ABM) et les automates cellulaires ou *cellular automata* (CA).

### **Modélisation basée sur l'agent**

Une méthode utilisée en modélisation des systèmes complexes est la modélisation basée sur l'individu (terme utilisé en écologie), parfois appelée modélisation multi-agent (en génie et en informatique), modèles cognitif computationnel (en science de la cognition) ou modèle basé sur les agents (en sciences économiques). L'ensemble de ces termes est regroupé, en anglais, sous l'expression *Agent-Based Model* (ABM) (Tang and Bennett, 2010). La modélisation multi-agents est une approche computationnelle permettant de créer des systèmes adaptatifs complexes dans lesquels on observe des caractéristiques d'auto-organisation, de non-linéarité et de comportement adaptatif des agents (Tang and Bennett, 2010). Les agents sont des représentations simplifiées du monde réel, réagissant à des stimuli provenant de leur environnement et des autres

agents. Les agents sont des objets de programmation soumis à des règles, capables de réagir, d'apprendre et de résoudre des problèmes.

### **Les automates cellulaires**

Une seconde méthode, la modélisation par automates cellulaires consiste en des modèles dynamiques comprenant cinq composantes, soit: 1) une grille de cellules, 2) un état pour chaque cellule, 3) un voisinage, les deux plus communs étant ceux de Moore (8 voisins) et de Von Neumann (4 voisins), 4) des règles de transition déterminant comment le voisinage affectera la cellule du centre à toutes les itérations et 5) le nombre et l'importance des itérations (Bone, Dragicevic, et Roberts, 2006). Les automates cellulaires sont utilisés pour mesurer des dynamiques spatio-temporelles, en combinaison aux systèmes d'information géographique (SIG), qui sont plutôt statiques. En géographie, les automates cellulaires basés sur les SIG ont été utilisés pour plusieurs applications, notamment les dynamiques urbaines (Kocabas et Dragicevic, 2007; Li et Yeh, 2002; Singh, 2003; Ward, Murray, and Phinn, 2000; Walsh *et al.*, 2006), les espèces invasives (Perez et Dragicevic, 2012; Dragičević, 2010; Bone *et al.*, 2013; Bone, Dragicevic, et Roberts, 2006) et les feux de forêt (Alexandridis *et al.*, 2008; Yassemi et Dragičević, 2008). Un automate cellulaire d'évolution de feux de forêt est utilisé dans le second article de ce mémoire.

## Bibliographie

- Alexandridis, a., D. Vakalis, C.I. Siettos, and G.V. Bafas. 2008. "A Cellular Automata Model for Forest Fire Spread Prediction: The Case of the Wildfire That Swept through Spetses Island in 1990."
- Bone, C., Dragicevic, S., and Roberts, A. 2006. "A Fuzzy-Constrained Cellular Automata Model of Forest Insect Infestations." *Ecological Modelling* 192 (1-2) (February): 107–125.
- Bone, C., White, J., Wulder, M., Robertson, C., Nelson, T. 2013. "Impact of Forest Fragmentation on Patterns of Mountain Pine Beetle-Caused Tree Mortality." *Forests* 4 (2) (April 29): 279–295.
- Breiman, L. 2001. "Random Forests." *Machine Learning*: 1–33.
- Dragičević, S. 2010. "Modeling the Dynamics of Complex Spatial Systems Using GIS, Cellular Automata and Fuzzy Sets Applied to Invasive Plant Species Propagation." *Geography Compass* 6: 599–615.
- Forister, M. L., McCall A.C., Sanders, N.J., Fordyce, J.A., Thorne, J.H., O'Brien, J., Waetjen, D.P., Shapiro, A.M. 2010. "Compounded Effects of Climate Change and Habitat Alteration Shift Patterns of Butterfly Diversity." *Proceedings of the National Academy of Sciences of the United States of America* 107 (5) (February 2): 2088–92.
- Hurlbert, A.H., and Liang, Z. 2012. "Spatiotemporal Variation in Avian Migration Phenology: Citizen Science Reveals Effects of Climate Change." *PloS One* 7 (2) (January): e31662.
- Kocabas, V., and Dragicevic, S. 2007. "Enhancing a GIS Cellular Automata Model of Land Use Change: Bayesian Networks, Influence Diagrams and Causality." *Transactions in GIS* 11 (5) (October): 681–702.
- Lawler, J.J., Shafer, S.L., White, D., Kareiva, P., Maurer, E.P., Blaustein, A.R., and Bartlein, P.J. 2009. "Projected Climate-Induced Faunal Change in the Western Hemisphere." *Ecology* 90 (3) (March): 588–97.
- Li, X., and Yeh A. 2002. "Neural-Network-Based Cellular Automata for Simulating Multiple Land Use Changes Using GIS." *International Journal of Geographical Information Science* 16 (4) (June): 323–343.
- O'Sullivan, D. 2004. "Complexity Science and Human Geography." *Transactions of the Institute of British Geographers* 29 (3) (September): 282–295.
- Parmesan, C. 2006. "Ecological and Evolutionary Responses to Recent Climate Change." *Annual Review of Ecology, Evolution, and Systematics* 37 (1) (December): 637–669.
- Perez, L, and Dragicevic, S. 2012. "Landscape-Level Simulation of Forest Insect Disturbance: Coupling Swarm Intelligent Agents with GIS-Based Cellular Automata Model." *Ecological Modelling* 231 (April): 53–64.
- Singh, A.K. 2003. Modelling land use land cover change using cellular automata in a geo-spatial environment.

- Tang, W., and Bennett, D.A. 2010. "Agent-Based Modeling of Animal Movement: A Review." *Geography Compass* 4 (7) (July 2): 682–700.
- Walsh, S.J., Entwisle, B., Rindfuss, R.A., and Page, P.H. 2006. "Spatial Simulation Modelling of Land Use/land Cover Change Scenarios in Northeastern Thailand: a Cellular Automata Approach." *Journal of Land Use Science* 1 (1) (June 30): 5–28.
- Ward, D.P., a.T. Murray, and S.R. Phinn. 2000. "A Stochastically Constrained Cellular Model of Urban Growth." *Computers, Environment and Urban Systems* 24 (6) (November): 539–558.
- Yassemi, S, and S Dragičević. 2008. "Web Cellular Automata: A Forest Fire Modeling Approach and Prototype Tool." *Cartography and Geographic ...* 35 (2): 103–115.

## Premier article

Jonathan Gaudreau, Liliana Perez et Pierre Legendre, « Identification des variables expliquant la distribution spatiale d'oiseaux de la forêt boréale et modélisation de tendances futures : une approche multivariée », *Cybergeo : European Journal of Geography* [En ligne], Systèmes, Modélisation, Géostatistiques, document 722, mis en ligne le 11 mai 2015. URL : <http://cybergeo.revues.org/26969> ; DOI : 10.4000/cybergeo.26969

## Présentation de l'article

Cet article a été écrit dans le contexte du cours BIO6077 : Analyse quantitative des données biologiques et a permis d'explorer les relations entre des variables explicatives et des variables réponses. Dans le cas présent, les variables explicatives sont des variables bioclimatiques, d'élévation et de perturbations et les variables réponses sont des abondances de 39 espèces d'oiseaux de la forêt boréale. L'article se divise en deux grandes familles de méthodes : 1) l'analyse statistique multivariée, incluant les régressions, le partitionnement de variation et la RDA et 2) la modélisation de répartition d'espèces, ou modélisation bioclimatique. Les résultats démontrent que les variables bioclimatiques sont responsables à 53% dans la variation d'abondances d'espèces d'oiseaux de la forêt boréale et que des changements climatiques induiront un déplacement vers le nord pour les deux espèces modélisées, soient la mésange à tête noire et le gros-bec errant.

## Accord des coauteurs

Cet article a été écrit conjointement par Jonathan Gaudreau, premier auteur, Liliana Pérez, second auteur et Pierre Legendre, troisième auteur. Jonathan Gaudreau a été responsable de la rédaction, du choix des méthodes, des analyses, de l'interprétation des résultats, et de l'écriture complète du manuscrit. Liliana Pérez a fourni la révision méthodologique, la correction, et a aidé à la structuration de l'article, tandis que Pierre Legendre a fourni des conseils et commentaires concernant les méthodes employées et a aidé à orienter le projet alors qu'il était au stade de projet de session pour le cours BIO6077, dont Dr. Legendre est le professeur.

À titre de coauteur, j'autorise Jonathan Gaudreau à présenter l'article *Identification des variables expliquant la distribution spatiale d'oiseaux de la forêt boréale et modélisation de tendances futures : une approche multivariée* dans son mémoire de maîtrise.

Liliana Pérez

Co-auteur

Pierre Legendre

Co-auteur

# Identification des variables expliquant la distribution spatiale d'oiseaux de la forêt boréale et modélisation de tendances futures: une approche multivariée

Jonathan Gaudreau <sup>1</sup>, Liliana Perez <sup>1</sup>, Pierre Drapeau <sup>2</sup>

<sup>1</sup>Département de Géographie, Université de Montréal, Canada

<sup>2</sup>Département de Sciences Biologiques, Université du Québec à Montréal

## Résumé

Les changements climatiques prennent une importance grandissante dans l'étude de la dynamique des populations animales. Plusieurs experts s'entendent pour affirmer que les changements climatiques seront un des principaux moteurs de changement écologique dans les prochaines décennies. L'objectif de cette recherche consiste en l'identification des principaux facteurs responsables de la distribution spatiale d'oiseaux de la forêt boréale du Québec, afin de proposer des modèles de distribution d'espèces suivant les changements climatiques prévus. Deux approches multivariées sont employées, l'analyse de redondance canonique (RDA) et le partitionnement de variation. Au total, 39 espèces d'oiseaux sont sélectionnées, en plus de variables bioclimatiques, anthropiques, écoforestières et d'élévation. Les variables bioclimatiques sont responsables de 53% de la variation dans la répartition spatiale des oiseaux étudiés, les variables ressources, c'est-à-dire l'altitude et le pourcentage de milieux humides, de 5%, et les variables anthropiques de moins de 1%. Les résultats des modèles démontrent que les deux espèces modélisées verront leur répartition spatiale fortement modifiée par les changements climatiques et se déplaceront vers les latitudes plus septentrionales ou en altitude, suivant l'intensité du réchauffement.

## Abstract

Climate is getting more significant in the study of animal population dynamics. Experts agree on the fact that climate change will likely be one of the main drivers of ecological change in upcoming decades. The goal of this research is to identify the main drivers of Québec boreal bird species distribution, in order to generate models of future spatial distributions under different climate scenarios. For this purpose two multivariate approaches are employed (Redundancy Canonical Analysis -RDA- and variation partitioning). A total of 39 bird species are selected as well as bioclimatic variables, anthropic disturbances, forest cover and elevation. Bioclimatic variables explain 53% of the variation in species distribution, while resource variables, comprising elevation and percentage of wet areas are responsible of 5% and anthropic variables, of 1%. Model results for two species support the hypothesis that the spatial distribution of boreal birds

will be deeply modified by climate change and that the birds will likely move towards higher latitude or altitude, following the warming intensity.

**Mots-clés** : écologie, analyse multivariée, biogéographie, Québec, changements climatiques

**Keywords** : ecology, multivariate analysis, biogeography, Québec, climate change



## Introduction

Les changements climatiques prennent une importance grandissante dans la recherche en environnement et pourraient être un des moteurs principaux de changement écologique dans les prochaines décennies (Parmesan, 2006; Lawler *et al.*, 2009; Virkkala *et al.*, 2008). Dans son rapport de 2013, le Groupe d'Experts Intergouvernemental sur l'Évolution du Climat (GIEC) conclut que la plupart des systèmes naturels seront touchés à différentes intensités (IPCC, 2013). Les écosystèmes terrestres doivent s'adapter aux printemps hâtifs et la faune et la flore migrent, dans la plupart des cas, vers de plus hautes latitudes ou altitudes (Forister *et al.*, 2010; Hurlbert & Liang, 2012; IPCC, 2013; Walther *et al.*, 2002). Les conclusions du GIEC, basées sur plus de 29 000 tableaux de données provenant de 75 études, affirment qu'un nombre important d'écosystèmes seraient perturbés (IPCC, 2013). Les impacts des changements climatiques sur le milieu physique comprennent, entre autres, la fonte des calottes glaciaires, le dégel du pergélisol et la vulnérabilité et l'instabilité des versants montagneux en zones de pergélisol. Parmesan (2006) fait état des multiples conséquences des changements climatiques anthropiques sur la faune et sur la flore. Parmi celles-ci se retrouvent les modifications dans la répartition spatiale des espèces.

Des études similaires, réalisées en Amérique du Nord, convergent et concluent que des espèces migrent vers de plus hautes latitudes ou en altitude, lorsque la topographie le permet (Forister *et al.*, 2010; Hurlbert and Liang, 2012; Walther *et al.*, 2002). Cette hypothèse est corroborée par des recherches portant sur l'Antarctique, l'Arctique, ainsi que sur les milieux tempérés de l'hémisphère nord (Parmesan, 2006; Virkkala *et al.*, 2010). Ainsi, l'augmentation des températures, en particulier dans les régions boréales et polaires, induit-elle une pression considérable sur les écosystèmes forestiers?

La province de Québec, située dans le Centre-Est du Canada, montre un gradient latitudinal important et comprend de nombreux domaines bioclimatiques, allant des forêts tempérées à la toundra arctique. Environ 50% de son territoire est couvert par la forêt boréale, qui comprend quatre domaines bioclimatiques distincts (MERN, 2013). La forêt boréale du Québec, s'étendant sur plus d'un million de kilomètres carrés et abritant de 150 à 300 espèces d'oiseaux migrateurs, sera fortement affectée dans les prochaines années (MERN, 2013).

Ainsi, afin de prévoir les répercussions des changements climatiques sur la répartition spatiale de la faune aviaire boréale, il importe d'identifier les principaux facteurs associés à cette répartition. Cette recherche vise donc, dans un premier temps, l'identification des variables explicatives influençant la répartition spatiale de 39 espèces d'oiseaux de la forêt boréale *via* l'utilisation de méthodes de modélisation statistique appliquées aux données multivariées, c'est-à-dire l'analyse de redondance canonique (RDA) et le partitionnement de la variation. Suite à ces analyses, deux espèces voient leur répartition spatiale modélisée *via* des approches de modélisation bioclimatique. De plus, la présente recherche vise à approfondir les connaissances sur les réponses fauniques aux changements climatiques, dans le but de mieux adapter les politiques de conservation. L'hypothèse émise est que la répartition spatiale des espèces est fortement corrélée au climat, plus particulièrement aux précipitations et à la température moyenne annuelle. Ainsi, les projections climatiques devraient favoriser une migration des espèces boréales vers de plus hautes latitudes ou en altitude en présence de topographie variable.

## Matériel et préparation des données

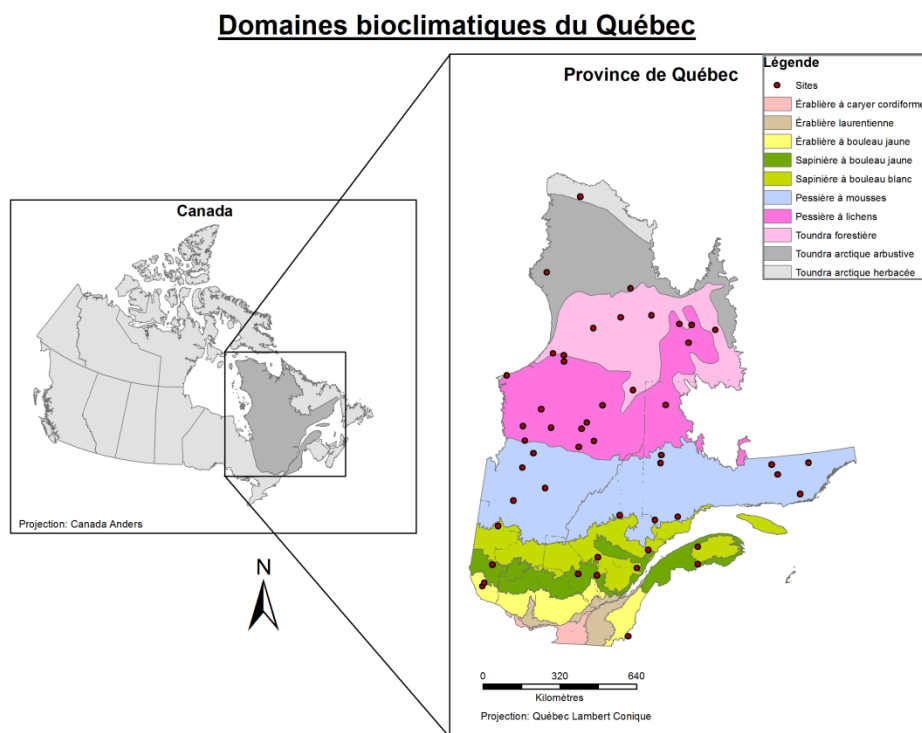
Les données nécessaires aux analyses proviennent toutes de bases de données ouvertes d'organismes (Global Forest Watch, 2009; Gouvernement du Canada, 2000; Hijmans *et al.*, 2005; CanFI, 2001; eBird, 2013). Les outils nécessaires à l'analyse sont, pour la plupart, le langage de programmation et de statistiques R 3.1.0 (R Core Team, 2014) et la suite ESRI ArcInfo 10.1 (ESRI, 2012) pour certaines transformations cartographiques.

## Aire d'étude et sites échantillonnés

La recherche porte sur la forêt boréale du Québec, Canada (44.99°S, 62.58°N, 79.75°O, 56.93°O) (*figure 1*). Les deux types de modèles nécessitent deux types d'intrants différents. La modélisation par analyse canonique de redondance (RDA) nécessite un tableau de données dont les observations sont répertoriées par site. Ainsi, puisque les données explicatives sont disponibles pour l'ensemble du Québec en format *raster*, un échantillonnage aléatoire est effectué, afin de réduire le nombre de points d'observation. Au total, cinquante sites sont aléatoirement sélectionnés, dans l'ensemble des domaines bioclimatiques du Québec (*figure 1*), en utilisant l'outil de création de points aléatoires distribués de façon uniforme d'ArcInfo (ESRI, 2012). La table d'attribut, à laquelle sont liées les valeurs des abondances d'oiseaux et des variables explicatives, est ensuite exportée et utilisée dans le logiciel statistique R (R Core Team, 2014) pour les analyses statistiques multivariées. Pour la seconde

partie, c'est-à-dire la projection via des approches de modélisation bioclimatique, les données pour l'ensemble du territoire étudié seront directement utilisées comme intrants. En effet, la librairie *biomod2* (Thuiller *et al.*, 2013) exige, en intrant, des couches *raster*.

Figure 1. Sites échantillonnés.



Variables réponses: abondances d'oiseaux

Les données d'abondance d'oiseaux proviennent de la base de données libre *eBird* (*eBird*, 2013), qui contient, pour le Québec uniquement, plus de 1.3 millions d'entrées. Bien que les données d'*eBird* ne fassent pas l'objet de standardisation et de vérification rigoureuse, la quantité phénoménale de données disponibles permet de minimiser l'impact de ce biais sur la qualité des données. Les trente-neuf espèces d'oiseaux ont été sélectionnées en parcourant une liste des oiseaux de la forêt boréale (BorealBirds, 2015). Le seul critère de sélection était que l'espèce soit résidente de la forêt boréale. Par souci de simplification, des codes à quatre lettres sont utilisés pour représenter les espèces d'oiseaux. Consultez l'annexe pour les noms correspondant aux sigles. Pour chaque observation, plusieurs informations sont disponibles, notamment l'heure de l'observation, l'espèce, le sexe, la localisation, entre autres. Seuls les champs concernant les coordonnées spatiales, soit la latitude et longitude, et le

nombre d'observations, sont conservés pour l'analyse. Les analyses statistiques multivariées (RDA et partitionnement de variation) nécessitent une étape supplémentaire. En effet, afin de permettre un échantillonnage aléatoire pour l'ensemble des espèces en utilisant les mêmes sites, l'aire maximale de répartition est d'abord interpolée pour chaque espèce. Cette interpolation produit une carte *raster* d'abondance de chaque espèce basée sur des observations réelles, simulant un polygone de répartition convexe, mais fonctionnant dans un espace matriciel, c'est-à-dire composé de cellules. La méthode utilisée est l'interpolation des points d'observations, en abondances, via la méthode *nearest neighbors* d'ArcInfo (ESRI, 2014). Le résultat donne une couche *raster* représentant l'aire maximale de distribution de l'espèce, faite en pixélisant l'abondance plutôt que la présence/absence. L'exercice est répété pour chacune des espèces choisies, totalisant 39 cartes matricielles. Cette interpolation des valeurs est requise pour l'échantillonnage aléatoire des sites nécessaire aux analyses statistiques multivariées, qui demandent des sites communs pour les 39 espèces. Évidemment, lors de la seconde étape, c'est-à-dire la modélisation bioclimatique, les observations réelles de la base de données sont entrées directement dans le modèle sous forme de coordonnées latitude/longitude, puisqu'une seule espèce à la fois est utilisée comme variable réponse.

### Variables explicatives

Le choix des variables explicatives, ou variables indépendantes, ou variables prédictives, est primordial à ces analyses. En effet, le but des analyses multivariées est l'identification de variables spatialisées influençant la répartition spatiale des oiseaux de la forêt boréale. De ces variables, seules les variables significatives sont utilisées dans les prévisions. Les variables explicatives sont projetées selon le système géodésique mondial de 1984, ou *WGS84* et consistent en cartes matricielles, ou *raster*, de 2112 lignes par 2721 colonnes. Le cadastre s'étend de 51.1 degrés ouest à 79.775 degrés ouest en longitude et de 44.98 degrés nord à 62.58 degrés nord en latitude. La résolution est de 0.0083 degré décimal par cellule, ou 787m par 787m, soit de moins d'un kilomètre carré. Les données bioclimatiques proviennent de la base de données *WorldClim* et s'étendent de BIO1 à BIO19 (voir l'annexe 1 pour la signification des sigles et des noms de variables) (Hijmans *et al*, 2005). Ces données, de nature quantitative discrète et à faible résolution spatiale (<1km), furent dérivées de températures et précipitations mensuelles pour engendrer des variables plus pertinentes en écologie, notamment la température moyenne estivale, l'étendue annuelle des précipitations, la saisonnalité des précipitations, etc. Les 19 variables bioclimatiques furent considérées initialement, afin de déterminer

lesquelles semblaient associées aux distributions d'espèces d'oiseaux. Lors de la sélection de variables, certaines furent abandonnées de par leur faible contribution aux modèles. Les données représentant les conditions climatiques actuelles consistent en la moyenne de la période de référence de 1950 à 2000. Les données futures représentent l'an 2070 (moyenne de la période 2061-2080) pour les quatre profils d'évolution de la concentration, ou scénarios *RCP (boîte 1)*, établis par le Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC), dans son cinquième rapport (IPCC, 2013).

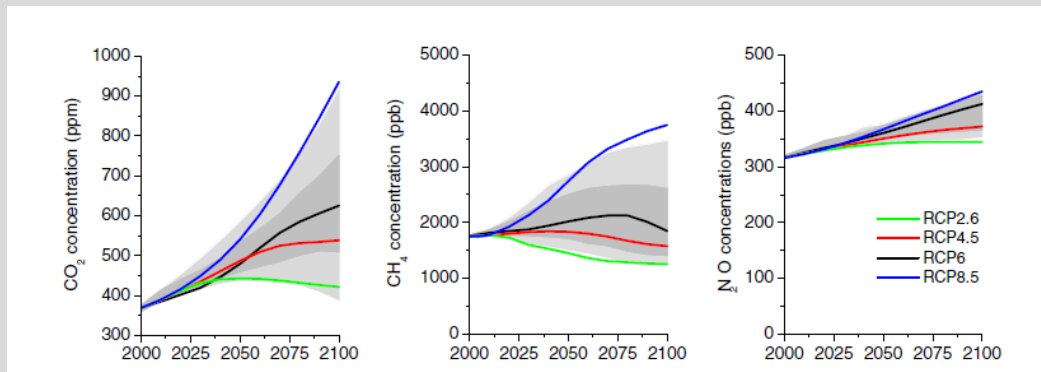
### Boîte 1. Scénarios RCP

À l'occasion de son 5<sup>e</sup> rapport, le GIEC définit un nouvel ensemble de scénarios climatiques, portant le nom de *Representative Concentration Pathways*, ou *RCPs*. Ces scénarios sont nommés par référence à leur forçage radiatif total pour l'année 2100 par rapport à 1750, année de référence préindustrielle. Les scénarios, nommés RCP2.6, RCP4.5, RCP6.0 et RCP8.5 représentent respectivement des forçages de  $2.6\text{Wm}^{-2}$ ,  $4.5\text{Wm}^{-2}$ ,  $6.0\text{Wm}^{-2}$  et de  $8.5\text{Wm}^{-2}$ . Les simulations (*figure 2*) furent effectuées avec des concentrations de  $\text{CO}_2$  atmosphérique, pour l'année 2100, de 421 ppm (RCP2.6), 538 ppm (RCP4.5), 670 ppm (RCP6.0) et 936 ppm (RCP8.5) (IPCC, 2013).

Figure 2. Tendances dans l'évolution des concentrations de gaz à effet de serre selon les 4 profils d'évolution de la concentration.

*Figure 2. Tendances dans l'évolution des concentrations de GES selon les 4 profils d'évolution de la concentration*

*De gauche à droite, on retrouve : le dioxyde de carbone ( $\text{CO}_2$ ) le méthane ( $\text{CH}_4$ ), et les oxydes nitreux ( $\text{N}_2\text{O}$ ) (Vuuren et al. 2011)*



Les données d'élévation proviennent de la Base Nationale de Données Topographiques du Canada (BNDT) (Gouvernement du Canada, 2000). Cette carte matricielle affiche l'élévation, en mètres, pour l'ensemble du Québec. Aux berges, rives, océans et estuaires est attribuée une valeur de zéro. Le niveau marin moyen est assigné selon le système de référence altimétrique géodésique du Canada de 1928 et la résolution verticale est d'un mètre (Gouvernement du Canada, 2000). Les données de perturbations anthropiques du Québec proviennent de Global Forest Watch Canada et sont composées

de quatre différents types de perturbation : 1) linéaires (routes, pipelines, lignes de haute tension), 2) de forme polygonale (forêt exploitée, puits de pétrole/gaz, terres agricoles), 3) réservoirs et 4) mines actives (Global Forest Watch, 2009). Aucun réservoir, aucune mine et aucune perturbation de forme polygonale ne se trouve sur les sites aléatoirement sélectionnés, si bien que seules les perturbations linéaires sont conservées pour l'analyse. Finalement, les données de foresterie proviennent de l'inventaire forestier canadien de 2001 et indiquent le pourcentage de territoire boisé par hectare, en plus de contenir les informations relatives à la proportion des différentes essences d'arbres (CanFI, 2001). Les données sont de nature quantitative discrète (CanFI, 2001). De cette carte multicouche, deux rasters sont extraits, soient : 1) une carte montrant le pourcentage de forêt par cellule et 2) une carte représentant le pourcentage de milieux humides par cellule. Finalement, ces variables peuvent être regroupées en trois catégories : les variables ressources, comprenant les données forestières et d'élévation, les variables bioclimatiques, BIO1 à BIO19, et les variables anthropiques, comprenant les perturbations de type linéaire.

## **Méthodes**

Afin d'être en mesure de voir la répartition spatiale future d'espèces d'oiseaux de la forêt boréale, une première étape consiste à identifier les variables associées à cette répartition spatiale communes à la plupart des espèces. La première partie consiste en des analyses multivariées destinées à identifier les prédicteurs communs des 39 espèces d'oiseaux sélectionnées. Lorsque ces variables sont sélectionnées, une seconde partie consiste en la prévision de la répartition spatiale future pour deux espèces sélectionnées parmi les 39 d'origine. Les deux espèces sélectionnées furent sélectionnées en fonction de leur relation avec les variables sélectionnées, afin que les modèles bioclimatiques soient plus significatifs. Deux espèces seulement ont été sélectionnées afin de limiter le temps de calcul (+- quatre heure par espèce, par scénario climatique), mais d'autres auraient aussi pu être modélisées.

### **Analyses multivariées pour identifier les variables responsables de la répartition spatiale des oiseaux de la forêt boréale du Québec**

#### **Transformation des données**

Les résultats de l'interpolation des points d'observation des oiseaux donnent des abondances allant de 0 (absence) à 720 (plus grande observation répertoriée, il s'agit de macreuses à tête blanche observées au sanctuaire d'oiseaux de l'île aux Basques). Un test de normalité de Shapiro-Wilks, dont l'hypothèse nulle consiste à affirmer que les données suivent une distribution normale et où

l'hypothèse alternative implique une distribution asymétrique ou anormale, permet de conclure que les observations sont de distribution très asymétrique (Royston, 1982). Une transformation logarithmique des données réponses est donc effectuée afin de réduire l'asymétrie des données et, ainsi, d'améliorer la puissance des tests statistiques.

L'analyse de redondance canonique est d'abord utilisée sur l'ensemble des données (Legendre et Legendre, 2012). Par la suite, les variables colinéaires sont successivement retirées du modèle, afin d'atteindre un facteur d'inflation de la variance (VIF) inférieur à 10 pour chacune des variables retenues. Afin de conserver les variables les plus représentatives, une sélection progressive bidirectionnelle est ensuite menée. Ce type d'approche effectue une sélection des variables pertinentes en fonction du critère d'information d'Akaike, ou *AIC* et de la représentativité (*p*-value) (Oksanen *et al.*, 2013). Une sélection progressive par ajout, en fonction du R-carré ajusté, est aussi utilisée, afin de comparer le résultat de la sélection (Oksanen *et al.*, 2013). Ces trois analyses sont répétées jusqu'à ce que seules les variables significatives soient intégrées au modèle.

#### 1) Analyse de redondance canonique

Une analyse de redondance canonique est un type d'analyse asymétrique, c'est-à-dire où on retrouve un tableau réponse *Y* et un tableau explicatif *X*. Les données subissent d'abord une régression linéaire multiple, afin de calculer la matrice des valeurs ajustées  $\hat{Y}$ , avec l'équation suivante (Legendre et Legendre, 2012. eq.11.11):

$$\hat{Y} = X[X'X]^{-1}X'Y \quad (1)$$

où *X* représente le tableau explicatif et *Y*, le tableau réponse. Par la suite, les valeurs ajustées font l'objet d'une analyse en composantes principales (ACP) et sont projetés dans un plan cartésien (ter Braak, 1986). La RDA permet de projeter trois types d'objets, soient les sites, les variables réponses et les variables explicatives, dans un seul graphique d'ordination, rendant la lecture beaucoup plus simple et permettant une meilleure compréhension des relations entre *Y* et *X* (Legendre et Legendre, 2012). Plusieurs tests sont effectués sur les résultats de la RDA afin de vérifier la signification des résultats. Les



axes canoniques et les prédictors peuvent être soumis à des tests par permutations pour en calculer la statistique  $F$  et la probabilité associée.

## 2) Réduction de la colinéarité par suppression de variables

Afin de bien identifier les variables expliquant la répartition spatiale d'espèces d'oiseaux, il est nécessaire de bénéficier de variables explicatives indépendantes les unes des autres. Ainsi, il est crucial de réduire la colinéarité. On parle de colinéarité lorsque deux ou plusieurs variables explicatives sont corrélées, partageant une relation linéaire. Une des mesures utilisées pour identifier la colinéarité est le facteur d'inflation de la variance (ou  $VIF$ ), qui se calcule en menant, successivement, des régressions linéaires de chacune des variables explicatives sur l'ensemble des variables explicatives. Les variables bioclimatiques offertes par WorldClim étant basées sur des données de température et de précipitation, elles sont donc toutes construites sur les mêmes données, et donc très corrélées. Le facteur d'inflation de la variance est obtenu avec l'équation suivante (Legendre et Legendre, 2012. Eq. 10.7):

$$VIF_j = \frac{1}{1 - R_j^2} \quad (2)$$

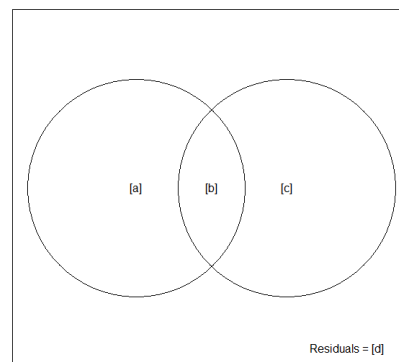
où ( $j$ ) est la variable explicative dont on désire connaître le VIF. Le seuil maximal est fixé à 10, puisqu'un seuil de 5 ne permet pas, dans le cas présent, de conserver assez de variables environnementales. Le retrait des variables à haut facteur d'inflation de la variance permet de réduire la colinéarité en supprimant les variables colinéaires, permettant ainsi de bien identifier l'influence individuelle des variables explicatives. Les variables abandonnées sont celles ayant le plus haut VIF et étant le moins pertinentes du point de vue écologique.

## 3) Partitionnement de variation pour séparer les variables en groupes

Afin d'identifier les facteurs influençant davantage la répartition spatiale des oiseaux de la forêt boréale, les variables sont regroupées en trois groupes, soient *birds.envClim*, comprenant les variables bioclimatiques, *birds.envRes*, comprenant les variables ressource, soit l'élévation, le pourcentage de forêt et le pourcentage de milieu humide et *birds.envAnth*, comprenant les perturbations anthropiques de type linéaire (Oksanen *et al.*, 2013).

Le partitionnement de variation (Borcard *et al.*, 1992) est représenté par un diagramme de Venn, tel que présenté à la figure 3. Ce diagramme permet d'identifier la contribution individuelle de chaque groupe de variables. La partie [a] représente la part de la variation expliquée exclusivement par le groupe de variables représenté dans le cercle de gauche, [c] représente la part de la variation expliquée uniquement par le groupe de variables représenté dans la sphère de droite. Finalement, la partie [b] représente l'intersection entre les deux groupes, soit la portion de la variation expliquée conjointement par les deux groupes de variables.

*Figure 3. Diagramme de Venn pour partitionnement de la variation de Y en fonction de deux groupes de variables explicatives.*



### **Modélisation bioclimatique pour modéliser les distributions d'espèces**

Des approches de modélisation bioclimatique sont utilisées pour produire des modèles de distribution d'espèces (de l'anglais *Species Distribution Models, SDMs*), en utilisant la librairie R *biomod2*, une plateforme de modélisation, de calibration et d'évaluation de modèles de distribution d'espèces, ou de niche (boîte 2) permettant d'avoir recours à plus de dix algorithmes de classification et de régression (Thuiller *et al.*, 2013).

## Boîte 2. Postulat d'équilibre et concept de niche.

Il est important de définir les deux concepts de ces modèles statiques : le postulat d'équilibre et le concept de niche. Le postulat d'équilibre est défini comme suit : *puisque les données environnementales et d'abondance d'espèces sont normalement échantillonnées sur une période de temps limitée ou sur des sites d'étude limités, les modèles ne reflètent qu'un cliché temporel de la relation attendue entre les conditions environnementales et la répartition des espèces. Ainsi, pour plus de commodité, on assume que les espèces modélisées se trouvent en pseudo-équilibre avec leur environnement* (Guisan *et al.*, 2005, traduit de l'anglais) Le second concept à définir est celui de niche. Les modèles de distribution d'espèce, plus particulièrement les cartes résultantes, reposent sur une combinaison entre la *niche écologique* de Grinnell, où une espèce se trouve partout où les conditions environnementales sont favorables à sa reproduction et la *niche réalisée* de Hutchinson, où une espèce est exclue de sa niche fondamentale par prédation ou compétition (Guisan *et al.*, 2005). Cette combinaison est due au fait que les modèles de distribution d'espèce, ou modèles de niche, utilisent des observations réelles pour identifier la niche écologique d'une espèce. Or, les observations réelles sont le fruit de la compétition et de la prédation des espèces; la *niche réalisée* sert donc indirectement à la conception des modèles.

Pour la modélisation des distributions des deux espèces sélectionnées, cinq algorithmes furent utilisés, soit : RF, MARS, MAXENT, FDA et GLM (*tableau I*).

Tableau I. Algorithmes de classification et de régression utilisés pour la génération des modèles de distribution d'espèces

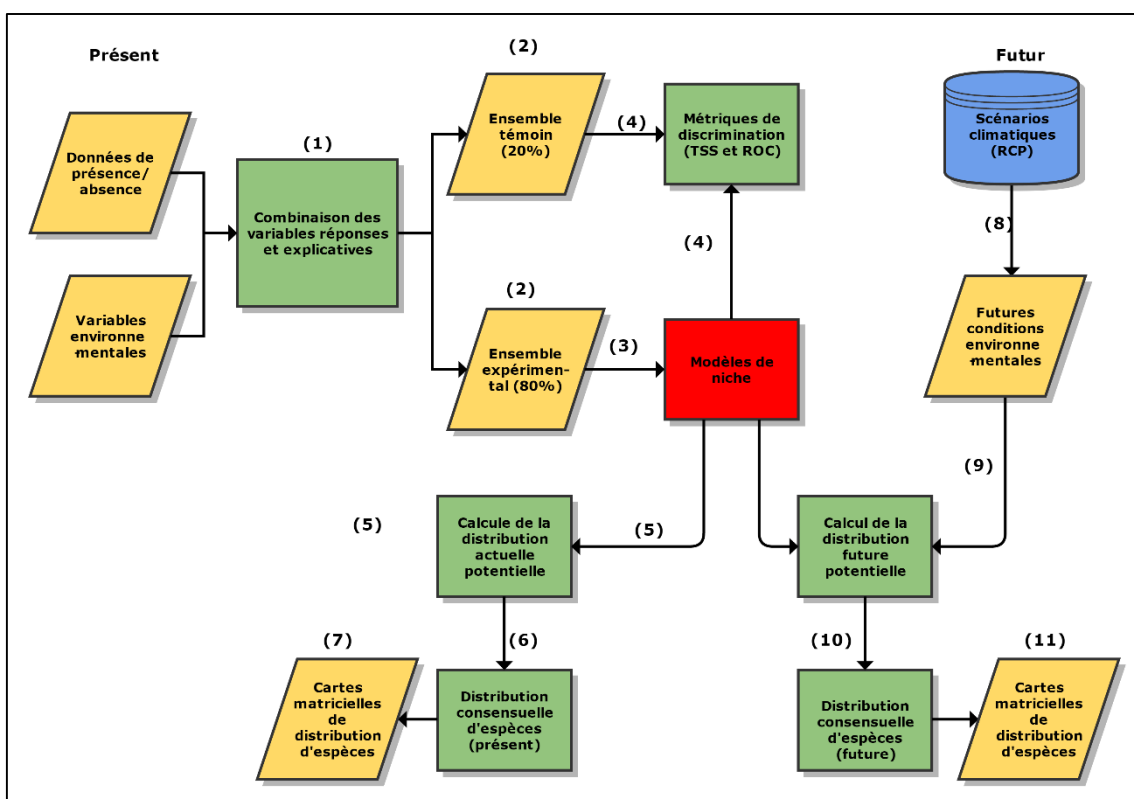
Sigle	Signification	Description	Source
RF	Forêt d'arbres décisionnels	Mesure l'importance des variables via de multiples permutations.	Breiman, 2001
MARS	Régression multivariée par spline adaptive	Permet de générer des modèles non-linéaires comprenant des interactions.	Friedman, 1991
MAXENT	Entropie maximale	Permet la modélisation d'absences seules. Approche d'apprentissage automatique, ou machine-learning.	Phillips et al., 2004
FDA	Analyse flexible discriminante	Algorithme de classification non-linéaire par score optimal.	Hastie et al., 1994
GLM	Régressions linéaires généralisées	Permet la régression linéaire en tenant compte de la stochasticité et d'une fonction de lien, binaire dans le cas présent.	McCullagh et Nelder, 1983

Tout d'abord, pour chaque espèce, les données de présence sont combinées à 10 000 pseudo-absences générées de façon aléatoire sur l'ensemble du territoire. Puis, ces présences/absences sont superposées aux variables environnementales et divisées en un ensemble contrôle (80%), utilisée pour générer un modèle de niche, et en un ensemble témoin (20%) servant à valider le modèle de niche par le biais de métriques de discrimination. Pour éviter un trop important taux d'erreur, le seuil minimal d'aire sous la courbe ROC (de l'anglais *area under the receiver operating characteristic curve (AUC)*) est défini à 0.7 (*rappel : une classification aléatoire aura un score AUC de 0.5 et une classification parfaite aura un score AUC de 1*). Puis, les futures conditions environnementales sont ajoutées au modèle pour remplacer les conditions actuelles et créer des cartes de distribution futures. Pour définir la présence ou l'absence d'espèce, une approche de modélisation par ensemble (de l'anglais *Ensemble Modeling*) est utilisée (Thuiller *et al.*, 2013). Cette méthode permet de calculer la moyenne des projections des différents algorithmes statistiques utilisés. Par la suite, un algorithme réalise une décision par comité, (de l'anglais *committee averaging*), où les probabilités des différents modèles subissent une transformation en données binaires selon le seuil d'AUC > 0.7. Ainsi, les cellules ayant une valeur située entre 0 et 0.7 se voient attribuer la valeur de zéro et les cellules ayant une valeur supérieure ou égale à 0.7 se voient attribuer la valeur de un. Puis, chaque méthode vote en faveur ou en défaveur de la présence de l'espèce. Finalement, ce score est ramené entre 0 (absence unanime) et 1000 (présence unanime) et est attribué comme valeur finale. Ainsi, cette méthode permet de concilier les décisions binaires de présence/absence et l'incertitude liée aux projections dans une même carte, témoignant de l'incertitude inhérente aux différents algorithmes de modélisation bioclimatique (Pearson et Thuiller

2006). Une valeur de 500 signifiera que 50% des modèles voient une présence et 50%, une absence (Georges et Thuiller, 2014). La figure 4 présente une version simplifiée du processus de modélisation bioclimatique utilisé pour une espèce.

Figure 4. Adapté de Berteaux et al., 2010. Représentation simplifiée de la méthode de modélisation bioclimatique utilisée.

Les chiffres entre parenthèses représentent les étapes dans l'exécution du code. Les parallélogrammes représentent des entrées/sorties, les rectangles, des processus et le cylindre bleu, une base de données. Le rectangle rouge représente le coeur du modèle, soit le modèle de niche.



## Résultats

### Sélection de variables

Une première RDA menée sur l'ensemble des variables explicatives permet d'identifier plusieurs choses. Tout d'abord, le calcul des facteurs d'inflation de la variance permet d'identifier les variables les plus colinéaires et de les retirer manuellement. En incluant toutes les variables explicatives, certaines variables montrent un VIF extrêmement élevé, proche de 2000 (bio1, ou

température moyenne annuelle, bio11, ou température moyenne du trimestre le plus frais et bio6, ou température minimale du mois le plus frais). Le R-carré est de 0.854, mais le R-carré ajusté est de 0.735. Le test par permutation effectué sur les variables indique que les variables *pct\_for*, bio13 (précipitations du mois le plus arrosé), bio2 (étendue diurne moyenne), bio3 (isothermalité) et bio5 (température maximale du mois le plus chaud) sont significatives ( $p < 0.05$ ). Le test par permutation effectué sur les axes canoniques indique que seuls les axes 1, 7, 11 et 15 sont significatifs ( $p < 0.01$ ) après 299 permutations.

Afin de réduire le nombre de variables et d'augmenter la puissance des analyses, les variables colinéaires sont supprimées une à une, et ce de façon manuelle. Les variables conservées sont l'altitude, les perturbations linéaires, le pourcentage de forêt, le pourcentage de milieu humide et les variables bioclimatiques bio7, bio11, bio15 et bio16, représentant respectivement l'étendue annuelle des températures, les températures moyennes du trimestre le plus frais, la saisonnalité des précipitations et les précipitations du trimestre avec le plus de précipitations. Le VIF moyen des huit variables sélectionnées est de 5.24, avec 11.7 comme valeur maximale (*bio11*), ce que nous jugeons acceptable, puisque bio11 est légèrement au-dessus de seuil fixé, bien que la littérature recommande parfois des seuils de 15, voire de 20 (Legendre et Legendre, 2012). La RDA est ensuite effectuée une seconde fois, cette fois sur les variables non-colinéaires (*figure 5*). Le R-carré ajusté du modèle est de 0.58.

Par la suite, une sélection bidirectionnelle de variables mène à la suppression des variables *linear* et *pct\_for*. L'analyse finale est effectuée sur six variables, soient les quatre variables bioclimatiques (*bio7*, *bio11*, *bio15* et *bio16*), le pourcentage de milieu humide et l'élévation (*figure 6*). Sur ce graphique, il devient clair que la variable bio11, soit la température moyenne du trimestre le plus frais, est la variable qui influence la distribution spatiale de la majorité des espèces étudiées. La seconde variable en importance est la variable *bio16*, ou les précipitations du trimestre le plus arrosé. À l'opposé, deux variables influencent négativement la présence et l'abondance des espèces: *bio15* et *bio7*, soient la variation saisonnière des précipitations et l'étendue annuelle des températures. Ces graphiques permettent de représenter les gradients de température et de précipitations retrouvés au Québec. Pour plus d'information à propos des sites (*row1* à *row50*) voir la table « Sites Échantillonnés » en annexe, où, pour chaque site, les coordonnées latitude/longitude ainsi que le domaine bioclimatique dans lequel le site se trouve sont donnés. Un des éléments intéressants pouvant être tiré de l'observation des graphiques de RDA et le groupement des espèces et des sites.



En effet, dans le cadran bas-droite, on retrouve des espèces liées aux milieux humides ou à l'eau, telles la macreuse à ailes-blanches (WWSC), la macreuse à bec jaune (BLSC) ou la macreuse à front blanc (SUSC) et le site row42, se trouvant en pessière à lichens, plus particulièrement près du réservoir hydroélectrique La Grande 2. Un test par permutations effectué sur les variables explicatives permet de tester la signification des six variables sélectionnées (*Tableau II*). Toutes les variables intégrées au modèle final sont hautement significatives ( $p=0.01$ ), à l'exception de *pct\_wet*, qui est faiblement significative ( $p=0.07$ ).

*Tableau II. Résultat du test de permutation sur les prédicteurs*

<b>Variables</b>	<b>Df</b>	<b>Var</b>	<b>F</b>	<b>N.perm</b>	<b>Pr(&gt;F)</b>	<b>VIF</b>
<b>elevation</b>	1	49.512	23.8522	99	0.01	3.359
<b>pct_wet</b>	1	4.443	2.1403	99	0.07	1.232
<b>bio11</b>	1	67.141	32.3447	99	0.01	9.217
<b>bio15</b>	1	10.6	5.1063	99	0.01	7.199
<b>bio16</b>	1	14.793	7.1266	99	0.01	5.181
<b>bio7</b>	1	9.725	4.685	99	0.01	1.779
<b>résidus</b>	43	89.259				

Finalement, le même test, mais réalisé sur les axes canoniques, indique que les trois premiers axes sont hautement significatifs ( $p<0.005$ ) (*Tableau III*) et que le quatrième axe est significatif ( $p=0.071$ ).



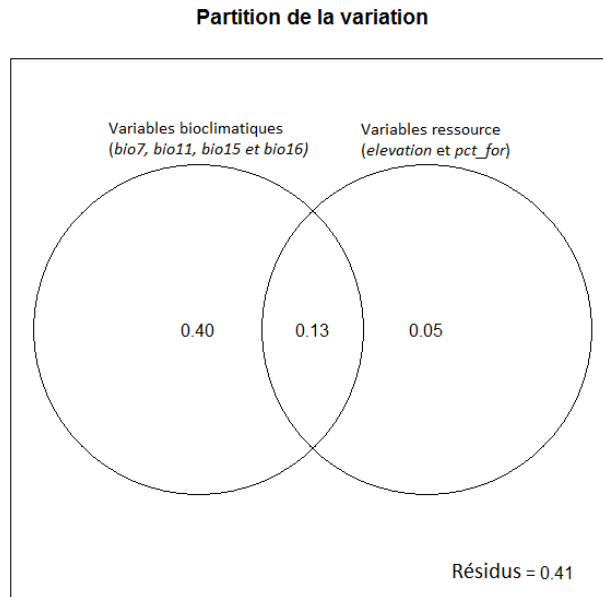
Tableau III. Résultats du test de permutation sur les axes canoniques

Axes	Df	Var	F	N.perm	Pr(>F)
<b>RDA1</b>	1	117.189	56.455	199	0.005
<b>RDA2</b>	1	18.856	9.0835	199	0.005
<b>RDA3</b>	1	13.101	6.3114	199	0.005
<b>RDA4</b>	1	4.093	1.972	999	0.071
<b>RDA5</b>	1	2.017	0.9718	99	0.44
<b>RDA6</b>	1	0.958	0.4614	99	0.88
<b>résidus</b>	43	89.259			

### Partitionnement de variation

Pour le partitionnement de variation, les variables furent séparées en groupes. Les quatre variables bioclimatiques forment le groupe de variables bioclimatiques et les variables *pct\_wet* et *elevation* formèrent un groupe de variables ressources, c'est-à-dire des variables indirectement liées à l'alimentation des espèces. Puisque les perturbations linéaires sont rejetées du modèle, le groupe de variables anthropiques est abandonné. Les variables bioclimatiques sont responsables de 53% de la variation totale, soit directement (40%), soit dans l'information expliquée conjointement avec les variables ressources (13%) (Figure 7). Les variables ressources, quant à elles, ne sont responsables que de 18% de la variation totale, soit directement (5%), soit *via* l'information expliquée conjointement avec les variables bioclimatiques (13%). Finalement, les résidus représentent 41% de la variation.

Figure 7. Partitionnement de la variation des données des 39 espèces d'oiseaux (log) par deux groupes de variables explicatives : les variables bioclimatiques et les variables ressource



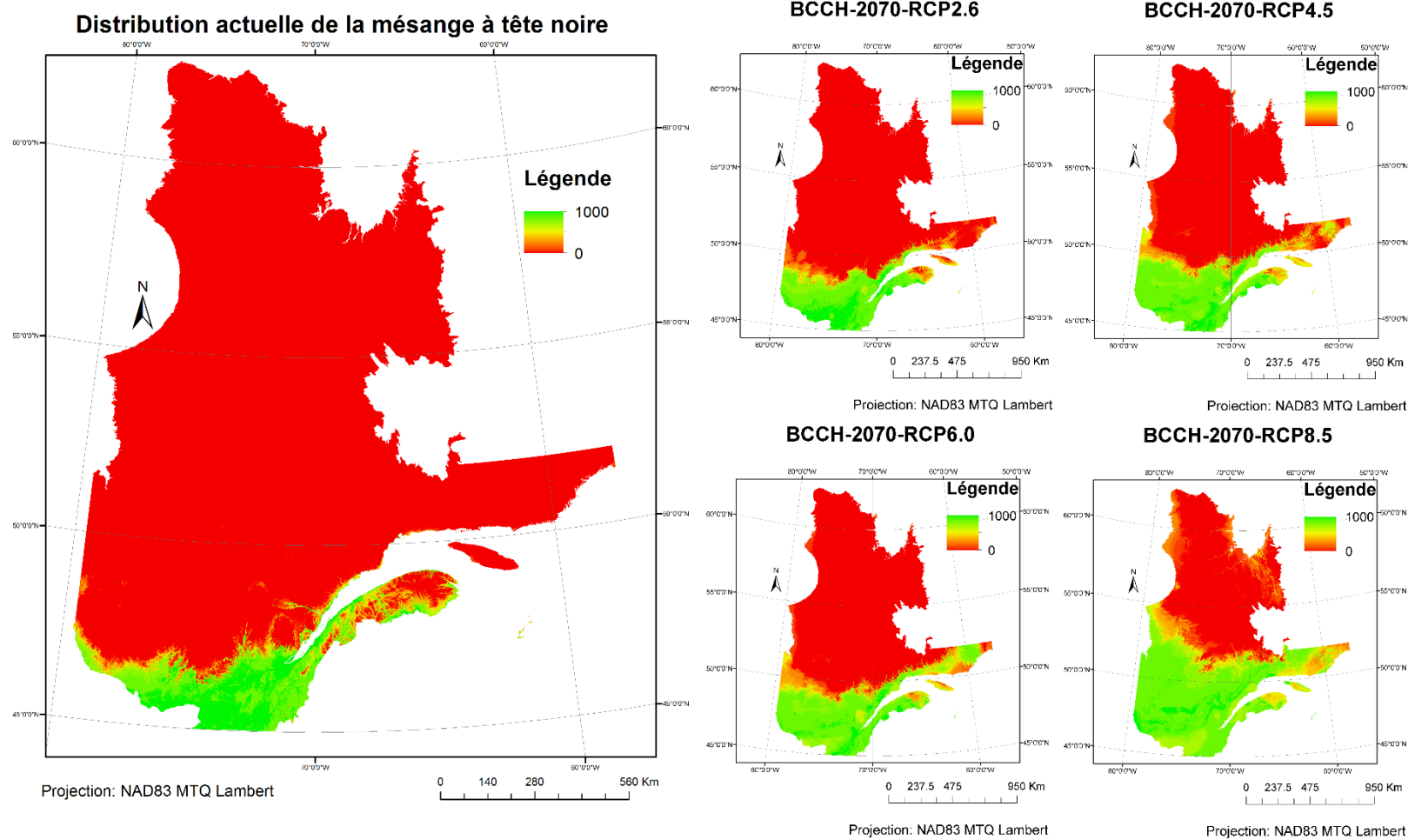
### Modélisation bioclimatique

Les deux espèces sélectionnées pour les projections sont la mésange à tête noire (BCCH) et le gros-bec errant (EVGR), de par leur très haut R-carré ajusté ( $R^2_{\text{adj}} = 0.79$  et  $R^2_{\text{adj}} = 0.79$  respectivement) obtenus lors de régressions linéaires multiples sur les variables explicatives sélectionnées lors de la RDA. Dans les deux cas, les espèces verront leur limite nordique s'étendre vers de plus hautes latitudes à une intensité différente selon le scénario climatique utilisé (figures 8 et 9).

Les cartes sont obtenues par consensus entre les différentes techniques. Ainsi, une valeur de 1000 représente l'accord total sur la présence de l'espèce, une valeur de zéro le désaccord total et toutes les valeurs intermédiaires représentent le degré d'incertitude. Ces cartes peuvent être lues comme des cartes binaires de présence et absence, mais tenant compte des incertitudes dans les projections. Les cartes démontrent, pour la mésange à tête noire, un déplacement rapide vers le nord, surtout dans l'Ouest de la province, et plus lent vers l'Est, soit les vers régions de la péninsule gaspésienne et la Côte-Nord (AUC = 0.982). Pour les gros-becs errants, les résultats démontrent un déplacement de l'aire de répartition vers le nord (AUC = 0.959). La limite nordique augmente pour les

scénarios 2.6, 4.5, 6.0 et augmente fortement pour le scénario 8.5. Pour les scénarios 6.0 et 8.5, les résultats démontrent un déplacement de la limite méridionale vers le nord, plus particulièrement dans le sud-est du Québec et dans la péninsule gaspésienne. Toutefois, l'incertitude pour ces deux secteurs laisse sous-entendre que certains modèles voient une présence et d'autres, majoritaires, voient une absence.

Figure 8. Distribution actuelle et prévue de la mésange à tête noire pour 2070 selon les quatre scénarios RCP





## Discussion

Les résultats de l'analyse de redondance canonique (RDA) indiquent que la répartition spatiale des 39 espèces d'oiseaux de la forêt boréale échantillonnés, serait liée en grande partie à des facteurs bioclimatiques, notamment les températures maximale et minimale, ainsi que les précipitations du trimestre ayant le plus de précipitations. Alors que l'analyse initiale inclut plusieurs variables, telles les perturbations anthropiques, la couverture forestière et les milieux humides, seules quatre variables bioclimatiques, l'élévation et la présence de milieux humides sont responsables de l'abondance d'oiseaux en forêt boréale. Les variables bioclimatiques expliquent 53% de la distribution spatiale des espèces aviaires et les variables ressources, 5%. Ainsi, il est possible d'affirmer que, pour nos données, les perturbations anthropiques semblent n'avoir aucun lien avec la distribution spatiale des oiseaux de la forêt boréale. Toutefois, il importe de rester vigilant avec ces résultats, puisque les observations d'oiseaux sont habituellement prises en bordure des routes ou en milieu anthropisé. Ainsi, nous ne bénéficions pas ou très peu d'observations pour le Nord du Québec, ni pour les endroits situés loin des villes et villages. De plus, les données provenant de la base de données eBird, bien que volumineuses, sont fortement biaisées par un manque de standardisation de la prise de données terrain. Par ailleurs, il serait plus avisé de trouver de meilleures variables représentant l'impact anthropique, notamment des données quant à l'utilisation du sol et aux occupations humaines. De plus, il faut rester vigilant à la sensibilité des analyses à la résolution spatiale des données. En effet, l'échantillonnage aléatoire donne des sites aux propriétés différentes selon la résolution des données. Ces résultats permettent toutefois de considérer les variables bioclimatiques comme important moteur influençant la répartition spatiale d'oiseaux de la forêt boréale québécoise.

Ainsi, pour la seconde étape de ce projet, soit la modélisation de distribution spatiale d'espèces d'oiseaux, seules les variables bioclimatiques et l'élévation sont utilisées. Toutefois, des régressions linéaires multiples calculées pour chacune des espèces de façon individuelle permettent d'identifier des variables différentes selon l'espèce. Certaines espèces, comme le viréo aux yeux rouges (REVI), a un  $R^2_{adj}$  de 0.78 pour quatre variables (elevation, bio11, bio16 et bio7), alors que d'autres, tel le roitelet à couronne rubis (RCKI) a un  $R^2_{adj}$  de 0.26, avec une seule variable significative (bio11). Les résultats des modèles appuient l'idée que les changements climatiques favoriseraient un changement dans la répartition spatiale de ces deux espèces d'oiseaux de la forêt boréale. Cette idée est en accord avec le paradoxe de biodiversité nordique, qui consiste à affirmer que bien que les changements climatiques seront une grande cause d'extinction d'espèces, les régions boréales, elles, subiront un

accroissement dans la richesse des espèces et dans la biodiversité (Bertheaux *et al.*, 2010). Les résultats d'une étude similaire démontrent que pour 80% des scénarios envisagés, l'ensemble de l'Amérique du Nord devrait subir une perte nette d'espèces animales de 11% sous les scénarios climatiques B1 (conservateurs) ainsi qu'une perte d'espèces animales de 17% sous les scénarios A2 (moyennes à hautes émissions de gaz à effet de serre) (Lawler *et al.* 2009).

Les modèles de distribution d'espèces tels que produits dans le cadre de cette recherche sont statiques et reposent sur un postulat de pseudo-équilibre, tel que défini par Guisan et Zimmermann (2000). En effet, toute modélisation statique repose sur la prémisse que les patrons de distribution des espèces modélisées sont en équilibre ou pseudo-équilibre avec l'environnement, puisque les modèles statiques ne permettent pas de gérer les conditions de déséquilibre ou d'équilibre dynamique que démontrent les écosystèmes réels. Les approches de modélisation de type dynamique et stochastique issues des théories des systèmes complexes, telle la modélisation basée sur l'individu ou l'automate cellulaire, permettent de pallier à ces problèmes, puisqu'elles reposent sur une approche ascendante (*bottom-up*), modélisant à l'échelle de l'individu et permettant les rétroactions et la non-linéarité, alors que les approches mobilisées ici reposent sur une approche descendante (*top-down*), où les aires de répartition d'espèces sont identifiées par des statistiques et par des variables environnementales et non par les traits comportementaux et préférences de ces espèces.

La méthode dressée ici, consistant en une combinaison d'analyses multivariées et de modélisation de distribution d'espèces pourrait aisément être adaptée à d'autres espèces et/ou à d'autres régions d'étude. La première étape, incluant l'analyse de redondance canonique, la suppression de colinéarité, la sélection bidirectionnelle puis le partitionnement de variation, permet de sélectionner les variables explicatives pertinentes et de les inclure dans la seconde étape, plutôt que de produire des modèles incluant plusieurs variables inutiles.

## Conclusion

Les analyses multivariées permettent d'identifier les variables explicatives responsables de la distribution spatiale de 39 espèces d'oiseaux de la forêt boréale du Québec. Les variables identifiées comme étant des variables clés comprennent l'étendue annuelle des températures, les températures moyennes du trimestre le plus frais, la saisonnalité des précipitations, les précipitations du trimestre le plus arrosé et l'élévation. La variable partageant le rapport le plus fort avec les deux espèces modélisées est bio11, ou les températures moyennes du trimestre le plus frais. Ceci s'explique aisément par le fait que les deux espèces modélisées sont des espèces boréales. Les résultats des analyses multivariées démontrent que les espèces d'oiseaux sélectionnées pour l'étude ont une distribution spatiale fortement corrélée au climat. Ainsi, comme le démontrent les résultats des modèles bioclimatiques, les changements climatiques risquent-ils d'induire des changements dans la répartition spatiale de ces espèces. Des modèles mécanistes et stochastiques prenant en considération la fragmentation forestière et les perturbations naturelles et anthropiques seraient en mesure de rendre justice à la complexité inhérente aux écosystèmes forestiers et, ainsi, de prendre en considération les interactions et rétroactions possibles avec d'autres espèces ou avec le milieu.

## Remerciements

Les auteurs voudraient remercier le Centre d'Étude de la Forêt (CEF) et l'Université de Montréal, qui ont offert leur support financier, les ornithologues amateurs et professionnels ayant mis au point la très complète base de données *eBird*, tous les organismes ayant fourni des données, et ce, complètement gratuitement. Les auteurs voudraient aussi remercier particulièrement tous les chercheurs contribuant à la démocratisation des outils scientifiques, libres et participatifs. Les auteurs sont aussi très reconnaissants pour les commentaires fournis par deux réviseurs anonymes dans une version précédente de ce manuscrit.



Codage des espèces d'oiseaux :

Code	Nom Latin	Nom Anglais	Nom Français
AMRE	<i>Setophaga ruticilla</i>	American Redstart	Paruline Flamboyante
AMRO	<i>Turdus migratorius</i>	American Robin	Merle d'Amérique
BBWA	<i>Setophaga castanea</i>	Bay-breasted Warbler	Paruline à Poitrine Baie
BBWO	<i>Picoides arcticus</i>	Black-backed Woodpecker	Pic à dos noir
BCCH	<i>Poecile atricapillus</i>	Black-capped Chickadee	Mésange à tête noire
BOCH	<i>Poecile hudsonicus</i>	Boreal Chickadee	Mésange à tête brune
BLSC	<i>Melanitta americana</i>	Black Scoter	Macreuse à bec jaune
BAWW	<i>Mniotilta varia</i>	Black-and-white Warbler	Paruline noir-et-blanc
CORA	<i>Corvus corax</i>	Common Raven	Grand corbeau
CORE	<i>Acanthis flammea</i>	Common Redpoll	Sizerin flammé
DEJU	<i>Junco hyemalis</i>	Dark-eyed Junco	Junco ardoisé
EVGR	<i>Hesperiphona vespertina</i>	Evening Grosbeak	Gros-bec errant
FOSP	<i>Passerella iliaca</i>	Fox Sparrow	Bruant fauve
GCKI	<i>Regulus satrapa</i>	Golden-crowned Kinglet	Roitelet à couronne dorée
GRAJ	<i>Perisoreus canadensis</i>	Gray Jay	Mésangeai du Canada
LISP	<i>Melospiza lincolni</i>	Lincoln's Sparrow	Bruant de Lincoln
NSHR	<i>Lanius excubitor</i>	Northern Shrike	Pie-grièche grise
OSFL	<i>Contopus cooperi</i>	Olive-sided Flycatcher	Moucherolle à côtés-olive
PUFI	<i>Haemorhous purpureus</i>	Purple Finch	Roselin pourpré
PIGR	<i>Pinicola enucleator</i>	Pine Grosbeak	Durbec des sapins
PIWA	<i>Setophaga pinus</i>	Pine Warbler	Paruline des pins
RUBL	<i>Euphagus carolinus</i>	Rusty Blackbird	Quiscale rouilleux
RBNU	<i>Sitta canadensis</i>	Red-breasted Nuthatch	Sitelle à poitrine rousse
RCKI	<i>Regulus calendula</i>	Ruby-crowned Kinglet	Roitelet à couronne rubis
RECR	<i>Loxia curvirostra</i>	Red Crossbill	Bec-croisé des sapins
REVI	<i>Vireo olivaceus</i>	Red-eyed Vireo	Viréo aux yeux rouges
RUGR	<i>Bonasa umbellus</i>	Ruffed Grouse	Gélinotte huppée
SPGR	<i>Falcapennis canadensis</i>	Spruce Grouse	Tétras du Canada
SOSA	<i>Tringa solitaria</i>	Solitary Sandpiper	Chevalier solitaire
SPSA	<i>Actitis macularius</i>	Spotted Sandpiper	Chevalier grivelé
SUSC	<i>Melanitta perspicillata</i>	Surf Scoter	Macreuse à front blanc
SWSP	<i>Melospiza georgiana</i>	Swamp Sparrow	Bruant des marais
SWTH	<i>Catharus ustulatus</i>	Swainson's Thrush	Grive à dos olive
VEER	<i>Catharus fuscescens</i>	Veery	Grive fauve
WCSP	<i>Zonotrichia leucophrys</i>	White-crowned Sparrow	Bruant à couronne blanche
WHIM	<i>Numenius phaeopus</i>	Whimbrel	Courlis corlieu
WWCR	<i>Loxia leucoptera</i>	White-winged Crossbill	Bec-croisé bifascié
WWSC	<i>Melanitta deglandi</i>	White-winged Scoter	Macreuse à ailes blanches
YBFL	<i>Empidonax flaviventris</i>	Yellow-bellied Flycatcher	Moucherolle à ventre jaune

## Sites échantillonnés

Nom	Latitude	Longitude	Domaine Bioclimatique
row0	55.8447	-76.2316	Toundra forestière
row1	48.6628	-69.6323	Sapinière à bouleau blanc
row2	46.8135	-78.6928	Érablière à bouleau jaune
row3	51.6501	-59.9633	Pessière à mousses
row4	47.6502	-73.5156	Sapinière à bouleau jaune
row5	49.9449	-71.3324	Pessière à mousses
row6	53.0530	-75.8372	Pessière à lichens
row7	52.4532	-77.3685	Pessière à lichens
row8	50.1640	-77.6107	Pessière à mousses
row9	55.8104	-75.5142	Pessière à lichens
row10	47.9788	-70.2513	Sapinière à bouleau blanc
row11	49.1433	-78.3106	Sapinière à bouleau blanc
row12	54.6786	-70.7768	Toundra forestière
row13	57.3887	-71.8150	Toundra forestière
row14	47.6650	-78.3112	Sapinière à bouleau jaune
row15	50.5063	-60.6872	Pessière à mousses
row16	45.4110	-70.6361	Érablière à bouleau jaune
row17	50.7644	-75.8232	Pessière à mousses
row18	52.0158	-76.7739	Pessière à mousses
row19	49.9230	-67.9408	Pessière à mousses
row20	54.8004	-79.0797	Pessière à lichens
row21	58.4981	-71.2060	Toundra arctique arbustive
row22	48.7750	-66.8054	Sapinière à bouleau blanc
row23	53.7060	-76.5934	Pessière à lichens
row24	52.4247	-74.0392	Pessière à lichens
row25	51.9439	-68.9568	Pessière à mousses
row26	55.5781	-75.4543	Pessière à lichens
row27	56.4672	-67.1192	Pessière à lichens
row28	51.3266	-61.8963	Pessière à mousses
row29	52.6880	-73.1066	Pessière à lichens
row30	57.1851	-67.7253	Pessière à lichens
row31	56.9220	-65.2494	Toundra forestière
row32	51.4333	-77.3326	Pessière à mousses
row33	57.1362	-66.8704	Pessière à lichens
row34	58.8476	-77.3046	Toundra arctique arbustive
row35	51.7193	-62.2129	Pessière à mousses
row36	54.1290	-68.6697	Pessière à lichens
row37	48.3239	-72.4839	Sapinière à bouleau blanc

row38	52.9978	-77.6026	Pessière à lichens
row39	52.2419	-68.9173	Pessière à mousses
row40	53.3667	-73.6408	Pessière à lichens
row41	49.8007	-69.2638	Pessière à mousses
row42	54.0512	-72.7190	Pessière à lichens
row43	48.1273	-66.8305	Sapinière à bouleau blanc
row44	61.7936	-75.4240	Toundra arctique herbacée
row45	53.1172	-73.9403	Pessière à lichens
row46	46.9410	-78.6283	Érablière à bouleau jaune
row47	57.5095	-69.6575	Toundra forestière
row48	56.9296	-73.6560	Toundra forestière
row49	47.6405	-72.4772	Sapinière à bouleau jaune

### Codage des variables bioclimatiques (Hijmans *et al.*, 2005)

BIO1 = Température moyenne annuelle

BIO2 = Étendue diurne moyenne (Moyenne mensuelle de (tmax – tmin)

BIO3 = Isothermalité (BIO2/BIO7) (\* 100)

BIO4 = Saisonnalité des températures (écart \* 100)

BIO5 = Température maximale du mois le plus chaud

BIO6 = Température minimale du mois le plus frais

BIO7 = Étendue annuelle des températures (BIO5-BIO6)

BIO8 = Température moyenne du trimestre le plus arrosé

BIO9 = Température moyenne du trimestre le plus sec

BIO10 = Température moyenne du trimestre le plus chaud

BIO11 = Température moyenne du trimestre le plus frais

BIO12 = Précipitations annuelles

BIO13 = Précipitations du mois le plus arrosé

BIO14 = Précipitations du mois le plus sec

BIO15 = Saisonnalité des précipitations (coefficient de variation)

BIO16 = Précipitations du trimestre le plus arrosé

BIO17 = Précipitations du trimestre le plus sec

BIO18 = Précipitations du trimestre le plus chaud

BIO19 = Précipitations du trimestre le plus frais

### Codage des autres variables

elevation = élévation en mètres

pct\_for = pourcentage de forêt par cellule

pct\_wet = pourcentage de milieu humide par cellule

linear = perturbations de type linéaires

## Bibliographie

- Bertaux, D., de Blois, S., Angers, J.-F., *et al.*, 2010. The CC-Bio Project : Studying the Effects of Climate Change on Quebec Biodiversity. *Diversity* 2:1181-1204.
- Breiman, L. 2001. Random forests. *Machine Learning* 45:15-32.
- Borcard, D., Legendre, P., Drapeau, P. (1992) Partialling out the Spatial Component of Ecological Variation. *Ecology* 73 (3): 1045-1055
- Boreal Birds, 2015. Comprehensive Boreal Bird Guide. <http://www.borealbirds.org/comprehensive-boreal-bird-guide>. En ligne.
- Canada's Forest Inventory (CanFI). 2001. Canada's Forest Inventory for the boreal eco-region - percent forested, Quebec.
- eBird Basic Dataset. Version: ebd\_CA-QC\_198701\_201412\_relNov-2013. Cornell Lab of Ornithology, Ithaca, New York. Nov 2013.
- ESRI 2012. ArcGIS Desktop: Release 10.1. Redlands, CA: Environmental Systems Research Institute.
- Forister, M.L., McCall, A. C., Sanders, N.J. , Fordyce, J.A., Thorne, J.H., O'Brien, J., Waetjen, D.P., and Shapiro, A. M. 2010. Compounded Effects of Climate Change and Habitat Alteration Shift Patterns of Butterfly Diversity. *Proceedings of the National Academy of Sciences of the United States of America* 107 (5) (February 2): 2088–92.
- Friedman, J. H. 1991. Multivariate Adaptive Regression Splines. *The Annals of Statistics* 19 (1): 1-67
- Georges, D., Thuiller, W., Engler, R. 2014. BIOMOD\_EnsembleModeling {biomod2}: Create and evaluate an ensemble set of models and predictions. En Ligne. [http://www.inside-r.org/packages/cran/biomod2/docs/BIOMOD\\_EnsembleModeling](http://www.inside-r.org/packages/cran/biomod2/docs/BIOMOD_EnsembleModeling)
- Global Forest Watch Canada. 2009. Cumulative Anthropogenic Access, Eastern Canada.
- Gouvernement du Canada, Ressources Naturelles Canada, Centre d'information topographique. (2000) Données numériques d'élévation du Canada. <http://www.geobase.ca/geobase/en/data/cded/description.html> En ligne. Sherbrooke (Québec), Canada.

- Guisan, A., N.E. Zimmermann. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling* 135:147-186.
- Guisan, A., W. Thuiller. (2005). Predicting species distribution: offering more than simple habitat models. *Ecology letters* 8: 993-1009
- Hastie, T., Tibshirani, R., Buja, A. 1994. Flex Discriminant Analysis by Optimal Scoring. *Journal of the American Statistical Association* 89 (428): 1255-1270.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. and Jarvis, A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25: 1965-1978.
- Hurlbert, A.H., and Liang, Z. 2012. Spatiotemporal Variation in Avian Migration Phenology: Citizen Science Reveals Effects of Climate Change. *PloS One* 7 (2) (January): e31662. doi:10.1371/journal.pone.0031662.
- IPCC, 2013: Summary for Policymakers. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Oksanen, J., F. Blanchet, G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R. B., Simpson, G. L., Solymos, P, Henry, M., Stevens, H. and Wagner, H. (2013). *vegan: Community Ecology Package*. R package version 2.0-10. <http://CRAN.R-project.org/package=vegan>
- Lawler, Joshua J, Sarah L Shafer, Denis White, Peter Kareiva, Edwin P Maurer, Andrew R Blaustein, and Patrick J Bartlein. 2009. "Projected Climate-Induced Faunal Change in the Western Hemisphere." *Ecology* 90 (3) (March): 588–97.
- Legendre, P., Eugène D. Gallagher. 2001. Ecologically meaningful transformations for ordination of species data. *Oecologia* 129:271-280
- Legendre, P, L. Legendre, 2012. *Numerical Ecology*, Third English Edition. *Elsevier*. *Developments in Environmental Modelling*, 24.
- Lescop-Sinclair, Kateri, and Serge Payette. 1995. "Recent Advance of the Arctic Treeline Along the Eastern Coast of Hudson Bay" *83 (6): 929–936.*

- McCullagh, P., J.A. Nelder. 1983. Generalized linear models. Monographs on Statistics and Applied Probability. London, England. Chapman and Hall. 261p.
- MERN, 2013. "Bref Portrait de La Forêt Boréale Au Québec." <http://www.MERN.gouv.qc.ca/presse/feux-grands.jsp>.
- Nelson, T., Coops, N., Wulder, M., Perez, L., Fitterer, J., Powers, R., Fontana, F., 2014. Predicting Climate Change Impacts to the Canadian Boreal Forest. *Diversity* 6, 133–157.
- Parmesan, C. 2006. "Ecological and Evolutionary Responses to Recent Climate Change." *Annual Review of Ecology, Evolution, and Systematics* 37 (1) (December): 637–669. doi:10.1146/annurev.ecolsys.37.091305.110100. <http://www.annualreviews.org/doi/abs/10.1146/annurev.ecolsys.37.091305.110100>.
- Phillips, S.J., Dudik, M., Shapire, R.E. 2004. A Maximum Entropy Approach to Species Distribution Modeling. *Proceeding of the Twenty-First International Conference on Machine Learning*: 655-662
- R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Royston, P. 1982. An extension of Shapiro and Wilk's *W* test for normality to large samples. *Applied Statistics* 31, 176-180.
- ter Braak, C. J. F., 1986. Canonical Correspondance Analysis: A new eigenvector technique for multivariate direct gradient analysis. *Ecology* 67, 1167-1179
- Thuiller, W., Georges, D., and Engler, R. (2013). biomod2: Ensemble platform for species distribution modeling. R package version 3.1-25. <http://CRAN.Rproject.org/package=biomod2>
- Virkkala, R., Heikkinen, R.K., Leikola, N., and Luoto, M. 2008. Projected Large-Scale Range Reductions of Northern-Boreal Land Bird Species Due to Climate Change. *Biological Conservation* 141 (5) (May): 1343– 1353.
- Virkkala, R., Marmion, M., Heikkinen, R.K., Thuiller, W., and Luoto, M. 2010. Predicting Range Shifts of Northern Bird Species: Influence of Modelling Technique and Topography. *Acta Oecologica* 36 (3) (May): 269–281.

- Walther, G.-R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T.J.C., 2002. Fromentin, J.-M., Hoegh- Guldberg, O., Bairlein, F. 2002. Ecological responses to recent climate change.
- Pearson, RG, and W Thuiller. 2006. "Model-based Uncertainty in Species Range Prediction." *Journal of ...* <http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2699.2006.01460.x/full>.
- Vuuren, D.P., Edmonds, J., Kainuma, M., Riahi,K., Thomson, A., Hibbard, K., Hurrt, G.C., *et al.* 2011. "The Representative Concentration Pathways: An Overview." *Climatic Change* 109 (1-2): 5–31. doi:10.1007/s10584-011-0148-z.



## Deuxième article

Gaudreau, J., L.Perez & P. Drapeau. 2016. *BorealFireSim : A GIS-based Cellular Automata Model of Wildfires for the Boreal Forest of Quebec in a Climate Change Paradigm*. Ecological Informatics 32 : pp 12-27.

### Présentation de l'article

Le premier article traite essentiellement de l'interaction entre changement climatique et distribution spatiale d'espèces. La première partie, l'analyse statistique multivariée, a permis d'identifier les variables bioclimatiques responsables de la distribution d'oiseaux de la forêt boréale. Puis, ces variables bioclimatiques ayant fait l'objet d'une actualisation par des données futures ont permis d'entrevoir les potentielles répercussions des changements climatiques sur la répartition spatiale de certaines espèces d'oiseaux. Ce second article traite toujours de changements climatiques (scénarios RCP), mais porte cette fois sur l'évolution des patrons spatiaux de feux de forêt. La question motivant la démarche entreprise lors de la rédaction de cet article est la suivante : « Comment les changements climatiques induiront des changements dans la répartition spatiale des feux en forêt boréale québécoise? ». Afin de répondre à cette question, un modèle, *BorealFireSim*, fut créé pour identifier ces changements de 2010 à 2100 pour l'entièreté du Québec. Le modèle consiste en une modélisation spatialement explicite par automate cellulaire (*GIS-based CA model*) qui permet d'identifier les zones où les changements climatiques seront susceptibles de causer une augmentation dans le risque d'incendies et donc, de la superficie brûlée annuellement. Plutôt que d'être un modèle focalisant sur un événement de feu ou un front de feu, il s'agit d'un modèle à très petite échelle d'une résolution cellulaire de 20km<sup>2</sup>.

### Accord des coauteurs

Cet article a été écrit conjointement par Jonathan Gaudreau, premier auteur, Liliana Pérez, second auteur et Pierre Drapeau, troisième auteur. Jonathan Gaudreau a été responsable de la rédaction, du choix des méthodes, des analyses, de l'interprétation des résultats, et de l'écriture complète du document. Liliana Pérez a aidé à la conceptualisation du modèle, à la révision méthodologique, à la correction, et a aidé à la structuration de l'article, tandis que Pierre Drapeau a fourni des conseils et commentaires concernant les feux et la forêt, en plus d'aider à sélectionner les variables à intégrer au modèle.

À titre de coauteur, j'autorise Jonathan Gaudreau à publier l'article *BorealFireSim : A GIS-based Cellular Automata Model of Wildfires for the Boreal Forest of Quebec in a Climate Change Paradigm* dans son mémoire de maîtrise.

Liliana Pérez

Co-auteur

Pierre Drapeau

Co-auteur

# ***BorealFireSim*: A GIS-based Cellular Automata Model of Wildfires for the Boreal Forest of Quebec in a Climate Change Paradigm.**

Jonathan Gaudreau<sup>1\*</sup>, Liliana Perez<sup>1</sup>, Pierre Drapeau<sup>2</sup>

1. Laboratory of Environmental Geosimulation (LEDGE), Geography Department, University of Montreal, 520 Chemin de la Côte-Sainte-Catherine, H3C 3J7, Quebec, Canada.\*Corresponding author

2. Centre d'étude de la forêt, Département des sciences biologiques, Université du Québec à Montréal, C.P. 8888, Succursale Centre-Ville, Montréal, QC H3C 3P8, Canada.

## **Abstract**

Wildfires are the main cause of forest disturbance in the boreal forest of Canada. Climate change studies forecast important changes in fire cycles, such as increases in fire intensity, severity, and occurrence. The geographical information system (GIS) based cellular automata model, *BorealFireSim*, serves as a tool to identify future fire patterns in the boreal forest of Quebec, Canada. The model was calibrated using 1950-2010 climate data for the present baseline and forecasts of burning probability up to 2100 were calculated using two RCP scenarios of climate change. Results show that, with every scenario, the mean area burned will likely increase on a provincial scale, while some areas might expect decreases with a low emission scenario. Comparison with other models shows that areas forecasted to have an increase in fire likelihood, overlap with predicted areas of higher vegetation productivity. The results presented in this research aid identifying key areas for fire-dependent species in the near future.

## **Highlights**

- Climate Change will likely alter wildfire patterns in the boreal forest of Quebec
- *BorealFireSim* can model those wildfire patterns
- Forecasts show increases in fire likelihood
- Important changes in fire patterns for the northeastern part of the boreal forest
- Fire likelihood and vegetation productivity forecasts are overlapping

## **Keywords:**

Wildfires; Complexity; Modelling; GIS; Climate Change; Cellular Automata

## Introduction

Fire is the main source of disturbance in the boreal forest of Quebec (Natural Resources Canada, 2014). Wildfires are essential for forest regrowth of tree species such as jack pine or black spruce that depend on extreme heat to reproduce, as well as for insects or bird species depending on dead trees, or snags (Bonnot *et al.*, 2009; Nappi *et al.*, 2003). For the last decade, fire suppression costs in Canada ranged from 500 million to 1 billion dollars per year and burned more than 2.3 million hectares annually (Natural Resources Canada, 2014). Moreover, natural wildfires burning more than 200ha account for 97% of the total area burned across Canada and represent 3% of the wildfires (Natural Resources Canada, 2014). Wildland fires are caused mainly by four factors: fuels, climate-weather, ignition agents, and people (Flannigan *et al.*, 2009). Numerous studies have concluded that upcoming climate change will be a major driver of ecological change (Flannigan *et al.*, 2000; IPCC, 2013). Dale *et al.* (2008) states how the interactions between climate, disturbances and forest systems are critical to determine climate change impacts on forests. ). Among the biological impacts of climate change, variations in migration patterns of animals, increasing prevalence of wildfires and massive insect outbreaks are the most relevant (IPCC, 2013) Moreover, Mantyka-Pringle *et al.* (2015) demonstrated, in a study on the interplay between climate change and land-cover change, that adding climate change to land-cover change could increase the impacts of land-cover changes by up to 43% for birds and 24% for mammals. Fire frequency, size and seasonality would likely also be affected by climate change (IPCC, 2013). ). Keane *et al.* (2008), showed that predicted future climate change will likely cause major shifts in landscape vegetation dynamics and this shift is likely to be enhanced by independent changes in biophysical conditions. Changes in fire behavior will affect forest value for wildlife habitat as well as for the industry. Additionally, fire ignition and spread depend on the amount and frequency of precipitation, the type of forest cover and different conditions, such as thunderstorms, topography and wind speed, amongst others; thus these variables should be included within wildfire models (Dale *et al.*, 2001).

Modelling fire behaviour and spatiotemporal patterns enable better understanding of the feedbacks and interactions occurring in forested landscapes. Fire propagation models are usually deterministic and based on linear statistics; examples of such type of models are the Canadian Wildland Fire Effects Model (CanFIRE) (de Groot, 2012). CanFIRE is used by the Canadian Forest Service to predict the physical and ecological impacts of fires. Another widely used model is FARSITE (Finney, M.A., Andrews, 1999), a GIS-based fire growth model that is used to produce maps of fire behavior on a fire

event. Even though FARSITE is extremely powerful and couples statistical decision making to GIS, this model is not meant for large scale spatiotemporal fire dynamics, but for fire spread across landscapes (Finney, M.A., Andrews, 1999). While FARSITE and other FARSITE-based models like Fire-BGC (Green *et al.*, 1995) – a spatially-explicit fire succession model designed to investigate long-term trends in landscape pattern under historical and future fire regimes – focus on fire spread across a landscape, *BorealFireSim* works at a provincial scale, and focuses on a long term spatio-temporal changes in wildfire patterns in the boreal forest, dealing with many fire events in space and in time (Finney, M.A., Andrews, 1999; Keane *et al.*, 1999). Given that wildfire ignition can be caused by diverse interacting conditions, such as climate, elevation, dryness, tree species, weather, and presence of wet areas, the complex dynamics between these conditions give rise to spatial patterns of burned areas, emerging from local interactions to global scale patterns through time. The dynamic behaviour of wildfire processes can be studied by complex systems theory, which takes into account non-linearity of processes and feedbacks with the environment. The term complexity is used in this research, to represent the process by which identical initial conditions in an environment will give rise to different outcomes if the experiment is repeated multiple times (Batty and Torrens, 2001). Researchers often integrate complex behaviours into simple models using stochastic and dynamic modelling approaches. Among these approaches, cellular automata (CA) models have been proven effective to reproduce non-linear processes (Wolfram, 1994).

Cellular automata are models comprising a grid of cells where each one has a finite number of states. The state of a cell is influenced by the neighboring cells via transition rules. These transitions rules are applied to each cell for a certain number of time steps. In CA models, the state of a cell can be summarized with the following equation:

$$S_{i,j}(t + 1) = f\{N_{i,j}(t), S_{i,j}(t), \Delta T\} \quad (3)$$

where the state (S) of a cell  $i,j$  at a time (t+1) is a function of its neighborhood  $N_{i,j}(t)$ , and its state at the previous time  $S_{i,j}(t)$  within a discrete time step  $\Delta T$ . The major advantage of this approach is that instead of running simulations on the whole system with complex mathematical equations, simple rules are imposed on cells that can only interact with their neighbors. During and after the simulation, spatial patterns emerge from these local interactions between cells (Li and Magill, 2001).

The approach used in this research is a GIS-based cellular automata, which allows us to model dynamic, complex and non-linear interactions on large spatial and temporal scales. When coupled to

GIS, CA models make powerful tools for simulating complex spatiotemporal phenomena such as wildfire. While most of cellular automata models represent abstract or virtual environments, adding actual *georeferenced* map layers lets us model complex dynamics taking into account real landscapes and that is especially why GIS-based CA have been used in numerous fields. Examples of CA and GIS-based CA models can be found in multiple studies of dynamic processes such as land use/cover change and urban dynamics (Kocabas & Dragicevic, 2006, 2007; Ward *et al.*, 2000; Yeh & Li, 2003), invasive species (Bone *et al.*, 2006; Perez & Dragicevic, 2012) and forest fires (Alexandridis *et al.*, 2008, 2011; Yassemi *et al.*, 2008), to cite only few of the applications. In forest fires studies, Alexandridis (2011) showed the power of GIS-based cellular automata combined with meteorological data as a way to efficiently predict the evolution of fire front on forest landscapes. Alexandridis (2011) also included the *spotting effect* which is a phenomenon where burning material is transported by wind to areas not adjacent to the fire front, sometimes causing the ignition of a new, independent, fire event. Even though spotting could be important for fire front evolution models, this phenomenon is not relevant on a provincial scale, where the spatial resolution doesn't allow these short range (100 meters approximately) dynamics.

This research presents a novel GIS-based CA modeling approach named *BorealFireSim*, where the importance of model variables and transition rules are based on literature and on a thorough sensitivity analysis. Moreover, the *BorealFireSim* uses province wide information to simulate the probability of annual forest wildfires under current and future climate scenarios. In general, fire models consist of fire front spread models, that is, the evolution of a single fire depending on various variables, such as bush density, bush flammability and wind speed and direction (Li & Magill, 2001). Alternatively, *BorealFireSim* model aims to simulate probable fire patterns on a provincial scale under climate change and does not take into account fine scale fire spread behaviour as much as a fire front model would.

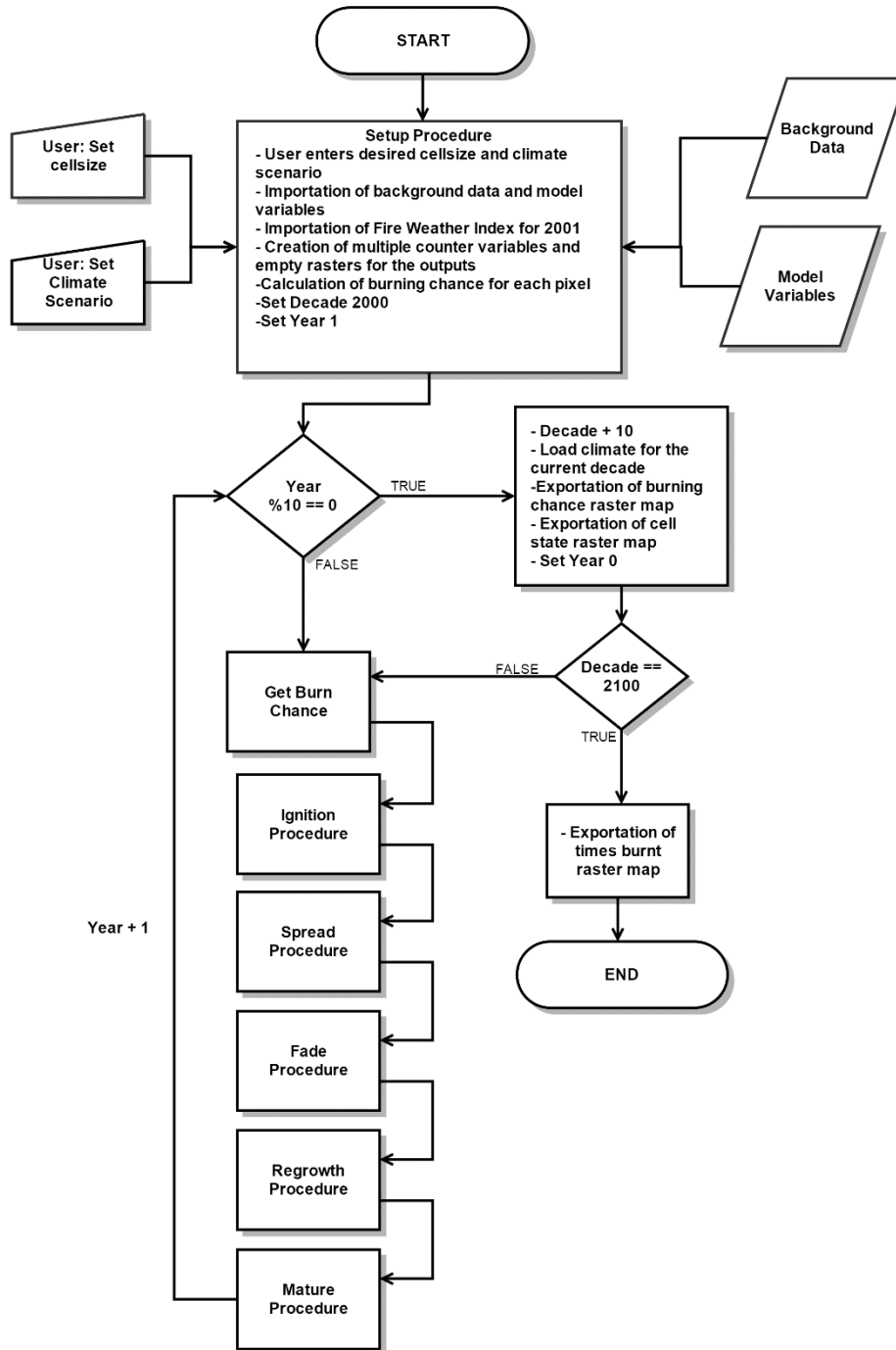
The main goal of this study is to simulate the complexities of wildfire processes and to identify changes in their spatial patterns throughout Quebec's boreal forest. Fire spread and post-fire regrowth behaviour in *BorealFireSim* are based on climate and environmental variables. Furthermore, after calibrating the model to reproduce current fire patterns, simulations are made using future climate data (CMIP5) for different representative concentration pathways (RCPs) (IPCC, 2013; Hijmans *et al.*, 2005). The following manuscript is organized in five sections. The material and methods (input data, model's transition rules, model flowchart, variables, algorithms and outputs) are detailed in the

following section. The results and discussion, section three, show maps of burning chance, cell states after 100 years, fire risk maps and statistical analysis. Likewise, results are compared to forecasted dynamic habitat index maps for 2050 and 2080 (Nelson *et al.*, 2014) to identify relevant areas for species conservation. Finally, limitations and conclusions come back on the assumptions and limitations of *BorealFireSim* and summarize the model results.

## **Material and methods**

The conceptual model representation of *BorealFireSim* is synthesized in a flowchart (Fig. 10) that illustrates all the model steps and algorithms with their attributes, functions and relationships. Following sub-sections explain the model behavior, variables, transition rules and the two algorithms developed to represent fire spread and forest regrowth processes, as expressed in the flowchart.

Figure 10. Flowchart of the BorealFireSim model





## Study Area

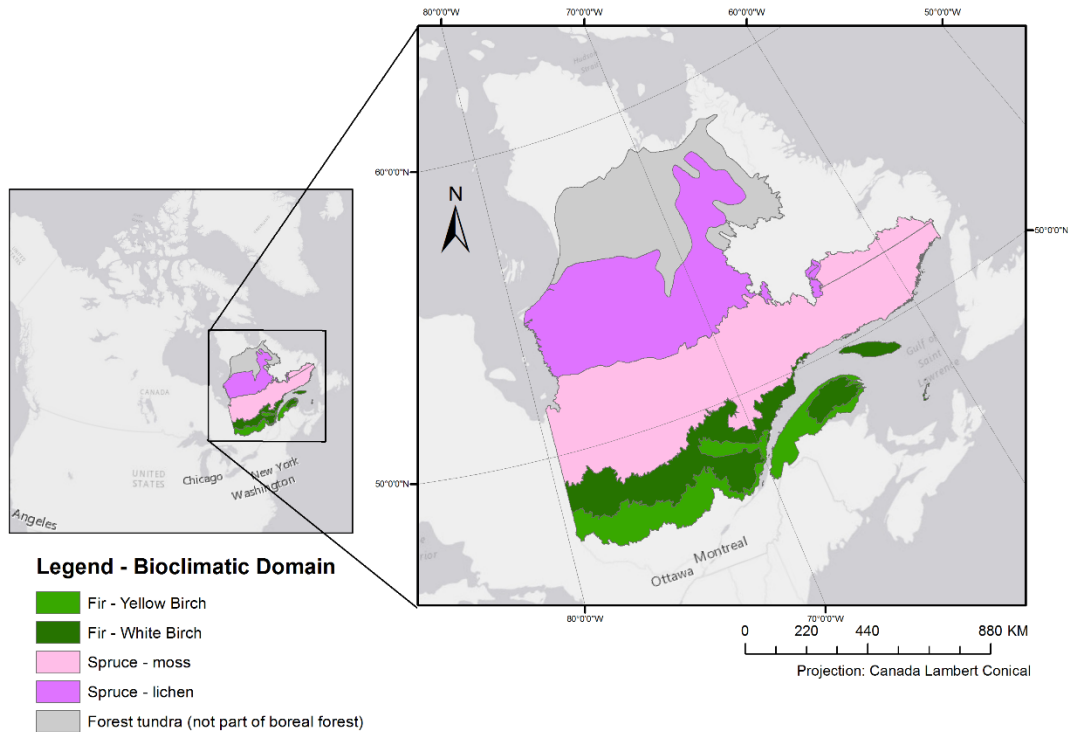
*BorealFireSim* was programmed to be functional on any boreal area, as long as the input data are available. However, this research focuses on the boreal forest of Quebec. The boreal forest of Quebec is the main vegetation domain in the province of Quebec, Canada, covering more than 560 000 km<sup>2</sup>, or a third of the province area. It is composed of a 75% coniferous cover, mainly black spruce (*Picea mariana*), jack pine (*Pinus banksiana*), balsam fir (*Abies balsamea*), white spruce (*Picea glauca*) and tamarack (*Larix laricina*) (Énergie et Ressources Naturelles Québec, 2013). Wildfires are an important source of natural disturbance in the boreal region of Quebec, an average of 928 fires is recorded each year, with a total of 42.2 million ha burned since 1973 (Forêt, Faune et Parcs Québec, 2013). From 2008 to 2014, fires prevention and suppression cost the government of Quebec \$677,900,000 and affected 2,705,400 ha of forest (with 1,753,900 in 2013 only), with a total of 4308 fires (MFFP, 2015). The boreal forest is very important for Quebec's economy; from 2004 to 2012, forest industry represented an average annual income of 2.24G\$ (MFFP, 2015). However, forest is not only important on the economic side, but also to the animal species that form its ecosystems. Wildfires are important for species such as the black-backed woodpecker (*Picoides arcticus*), which feeds on wood-boring beetles and are directly linked to burned forests (Nappi *et al.*, 2003).

The extent covered by *BorealFireSim* for this research is 45.375°N to 59.491°N to 79.757°W to 57.112°W (*Fig. 11*) and its spatial resolution is 4.48x4.48 km or ~20km<sup>2</sup>. The model includes the boreal forest of Quebec, as well as the forest tundra. The forest tundra was included in the model to prevent edge effects and to take into account shifts in the boreal forest northern limit with climate change.

### *Figure 11. Study Area*

*The study area is located in Quebec, Canada, and comprises five bioclimatic domains. The four domains in color represent the boreal forest of Quebec and the one in grey represents the forest tundra.*

## Area covered by BorealFireSim



## Model input data

### Climate variables

Climate information was calculated with BioSim10, which is a software designed to assist in pest management via temperature-driven simulation models (Saint-Amant & Régnière, 2015). Even though its main purpose is the forecast of events in the seasonal biology of pests, it is a powerful software for forecasting climate information, using Environment Canada's weather stations. The only climate variable present in *BorealFireSim* is the Fire Weather Index (FWI) mean of the fire season, for instance from May to September. Present climate data consists of the mean for the 1981-2010 period and future values were calculated with climate normals provided by OURANOS, a consortium on regional climatology and adaptation to climate change located in Montreal, Québec, which are available on the BioSim normals database (Saint-Amant & Régnière, 2015). Present data consist of a 30 year average while future data extend up to 2100. This is explained by the fact that we suppose a pseudo-equilibrium

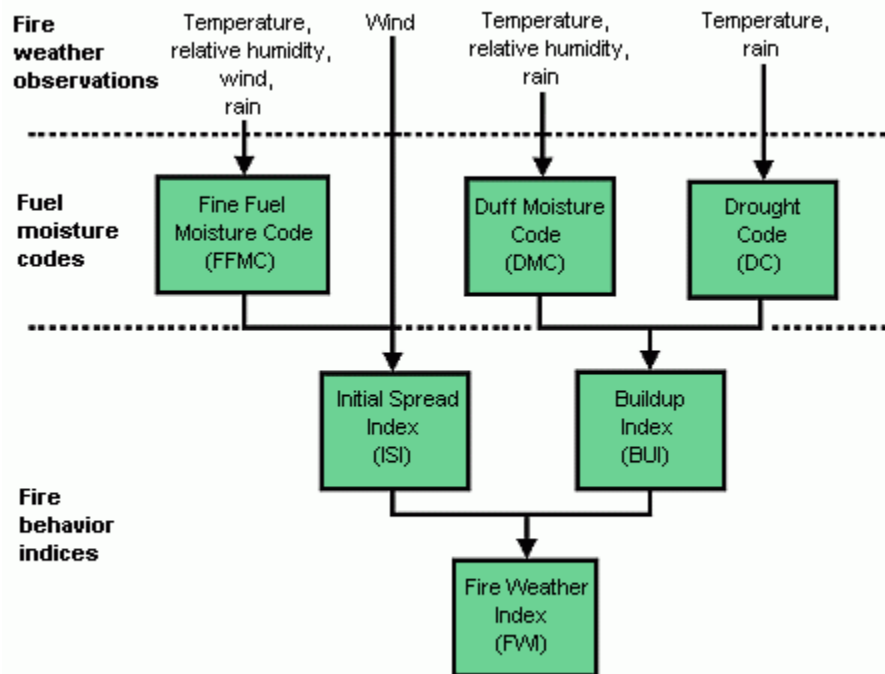
between actual fire occurrences and climate. Once the link is identified between present climate and present fire risk or occurrence, we replace climate data by future climate data and see how the modelled system reacts. The model aims more at getting a better understanding of regional fire patterns than at trying to predict fire events. The FWI is an index of the fire severity and was originally conceived as a way to represent fire intensity, as defined by Byram (1959) in the equation:

$$I = HWR \quad (4)$$

where the energy output rate per unit length of fire front **I** is the combination of the heat of combustion **H**, the weight of fuel consumed per unit area **W** and the rate of advance **R** (Van Wagner, 1987). However, the FWI system developed in Canada integrates multiple components and is based on daily observations of temperature, relative humidity, wind speed and 24-hour rainfall (Natural Resources Canada, 2008). We use FWI as our only climate variable because it is calculated with many variables relevant to fires, such as wind speed, temperatures and precipitation, or Drought Code (DC). Figure 12 shows the components of the FWI system.

Figure 12. Components of the Fire Weather Index (FWI)

(Source: Natural Resources Canada, 2008)



Calculations of the FWI are made for each decade, between 2000 and 2100 and results are exported to raster layers using the GeoTIFF format. Climate forecasts were used for two different climate scenarios: RCP4.5 and RCP8.5 (IPCC, 2013; Saint-Amant & Régnière, 2015). The representative concentration pathways (RCPs) consists of the latest climate scenarios (AR5<sup>1</sup>), replacing the AR4-SRES<sup>2</sup> scenarios and are named after their radiative forcing, defined by the Intergovernmental Panel on Climate Change (IPCC) as a “cumulative measure of human emissions of greenhouse gases from all sources expressed in Watts by square meter” (IPCC, 2013). For example, concentration pathway 4.5 represents a forcing of 4.5W/m<sup>2</sup> in 2100. For the purpose of clarification and further explanation on RCPs, table IV presents a comparison between RCPs and previous scenarios and figure 13 shows the different greenhouse gases emission curves.

*Table IV. Analogues between new and previous climate scenarios (Snober et al, 2013)*

<b>New Scenarios (RCP)</b>	<b>Description</b>	<b>Comparison to SRES scenarios</b>	<b>Term used in this paper</b>
<b>RCP 2.6</b>	An extremely low scenario that reflects aggressive greenhouse gas reduction and sequestration efforts	No analogue in previous scenarios	“Very low”
<b>RCP 4.5</b>	A low scenario in which greenhouse gas emission stabilize by mid-century and fall sharply thereafter	Very close to B1 by 2100, but higher emissions at mid-century	“Low”
<b>RCP 6.0</b>	A medium scenario in which greenhouse gas emissions increase gradually until stabilizing in the final decades of the 21 <sup>st</sup> century.	Similar to A1B by 2100, but closer to B1 at mid-century	“Medium”
<b>RCP 8.5</b>	A high scenario that assumes	Nearly identical to A1F1.	“Business As

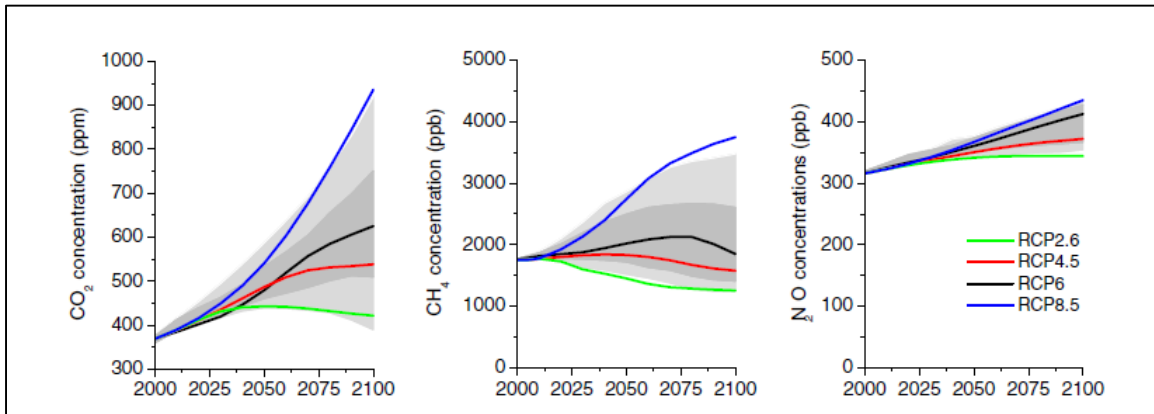
<sup>1</sup> IPCC Assessment Report 5 (IPCC, 2013)

<sup>2</sup> Special Report on Emissions Scenarios (SRES). In AR4 most common scenarios were A2, B1 and A1B (IPCC, 2007)

	continued increases in greenhouse gas emissions until the end of 21 <sup>st</sup> century	The A2 scenario would be placed between RCP6.0 and RCP8.5	Usual (BAU)"
--	---	---	--------------

Figure 13. Trends in the evolution of concentration of greenhouse gases for the four RCP scenarios.

Left to right: carbon dioxide (CO<sub>2</sub>), methane (CH<sub>4</sub>) and nitrous oxides (N<sub>2</sub>O) (Vuuren *et al.*, 2011)



### Ecological variables

Ecological variables include species-specific variables such as mean age of stands, dominant species, percentage of wet areas, forest density and bioclimatic domain. These variables, except for the bioclimatic domains (Forêts, Faune et Parcs Québec, 2015), were extracted from the Canadian forest inventory of 2001 database, or *CanFI* (Canadian Forest Service, 2001). In general, stands with age over 80 years are considered as mature forest and they are more prone to fire because their multiples layers of tree heights serving as a fire ladder (Barrett *et al.*, 2010). Dominant species are used to categorize fire-prone conifers, such as black spruce and jack pine, while giving less chance of burning to hardwoods such as birch or maple (British Columbia Forest Service, 2014). The density and type of vegetation are the most important factors for fire spread. No matter the weather conditions, if there is no connectivity between forests patches, fire will be unable to spread, except in the presence of very strong wind (Li & Magill, 2001). Forest density layer was derived from the Canadian Forest Inventory (CanFI) of 2001 and was converted to a raster grid with values from 0 to 100 representing the percentage of every cell that is covered by forest (Canadian Forest Service, 2001). Fire has greater chance to spread between cells of higher forest density.

## Model behavior

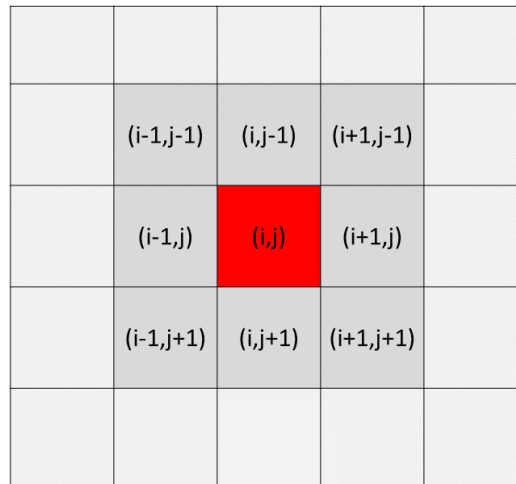
*BorealFireSim* takes into account multiple predictors identified by studies on forest spread (Barrett *et al.*, 2010; Bergeron *et al.*, 2001), post-fire regrowth (Johnstone and Chapin III, 2006), fire ignition (Li and Magill, 2001) and on other CA-based fire models (Yassemi *et al.*, 2008). Once the model is initialized during the setup process, roads, lakes and rivers act as background data and are turned off (given a state of 0, or null), while all other variables are considered into the model. When all the GIS layers are correctly imported, the setup ends. The user can now use the main (go) process to launch the model. The temporal resolution of *BorealFireSim* being one year, this means that each iteration represents one fire season. The go procedure will repeat itself until the number of iteration reaches 10. Once the number of iterations reaches 10, this means that the fire model will have run for 10 years and that the climate variables must be updated with the climate variables of the next decade. For example, the first ten iterations are made with FWI from 2000. After ten years, this variable is replaced by the forecast values for 2010, and so on, until the model reaches 2100. At the end of one decade simulation, Geotiff maps representing the risk of fire of every cell and the state of every cell are exported. Within the model each cell can have five different states: non-flammable, mature forests, burning forests, burnt forests and in regrowth. Table V shows the possible states and the transitions of a cell state in the *BorealFireSim* model. Non-flammable cells are cells that are turned off in the model and they consist of roads, rivers and lakes and cells that are not part of the study area or have *nodata* values; to summarize, these cells are non-suitable to spread the fire and therefore can never change state. Cells with a state value of 1 are cells with unburned vegetation, without discrimination to the dominant species, the density or the percentage of wet areas; hence they all have a chance of burning. These cells can either stay unburned to be ignited by lightning or by fire spreading. When a cell is burned, its state changes to 2 for the current time step and will be set to 3 at the end of the same time step. Since the temporal resolution of the model is a year, a cell that is ignited will pass from state 1 to state 2 to state 3 in the same iteration. Once a cell passes to state 3, it can either stay burned or start regrowing. Finally, cells with state 4 will eventually change to state 1 as they mature if they are not burned once again.

Table V. Cell states and possible transitions

Description	Cell state	Possible transitions
Non-flammable	0	[0] ==> [0]
Vegetation (fuel)	1	[1] ==> [1] [1] ==> [2]
Burning	2	[2] ==> [3]
Burned forest	3	[3] ==> [3] [3] ==> [4]
In regrowth	4	[4] ==> [1]

Every iteration, each cell evaluates its eight neighbors and the algorithms determine if the state of the cell should change. The Moore neighborhood was selected since it is more representative of real processes of fire propagation (Fig. 14).

Figure 14. Neighborhood representation used by the BorealFireSim



The flow of the model consists of five different actions, also called *procedures*. Every year, from 30 to 50 fires, which are the calculated average number of fires of more than 2km<sup>2</sup> burning each year in Quebec, are generated randomly throughout the landscape. Then, for every neighbor of a burning cell, the risk of fire spreading, between 0 and 1, is calculated. Each of the eight neighbors will compare its risk of spreading with a random number between 0 and 1. If the fire spreading chance is superior to

the random number, the fire will spread towards that neighbor. After having spread to the neighboring patches, fire will fade and leave burnt cells. These cells can stay burned from five to fifteen years, after which they start to be in regrowth for 20 to 60 years, until they consist of mature stands. These processes are executed by five procedures called: 1) ignition, 2) spreading, 3) fading, 4) regrowth and 5) maturation, and are repeated on a yearly basis. One year is represented by an iteration of the model. After the simulation runs for an iteration, a map representing the state of the cells and representing the burning chance is exported in a raster format, and will become the input for the following iteration.

### Algorithms

The core of the model consists of two algorithms; the fire spreading algorithm, responsible for the calculation of burning chance (BC) and the second one in charge of the regrowth chance (RC) calculation. The importance of every variable was determined using literature. The following subsection explains the different variables and their weights in the fire spreading algorithm.

#### Fire spread algorithm

Every cell for every variable is given a score (table VI). This score is then divided by the highest score of all cells, giving a value between 0 and 1, representing the probability of burn or regrowth for each cell. Table VI shows the different scores attributed to the different variables. An initial score of 2 is given to every cell because wildfires can happen anywhere in the forest and to reduce the model sensitivity to the variables. The score of 2 was determined by running the model several times and comparing the number of fires and the number of burned square kilometers with the actual values for the boreal forest of Quebec (MFFP, 2014). A score was added to the cells dominated by black spruces or by jack pines, since these two conifer species are rich in oils and highly flammable, while leafy vegetation is less prone to fire, because of its high foliar moisture content (Li & Magill, 2001; Drever *et al.*, 2006). Another variable added to the model is the percentage of wet areas. Fire spreads at its fastest rate when two conditions are met: drought and high temperatures. Wet areas are less prone to fire than regular forests because of the humidity of the soil and of the vegetation. Furthermore, wet areas facilitate the formation of permafrost, keeping the soil moist and low in temperature (Barrett *et al.*, 2010). The next variable to be included in the model is the Fire Weather Index (FWI) (Van Wagner, 1987). The FWI classes in the model represent the six danger classes defined by Van Wagner (1987), from Extreme to Very Low danger.

Finally the last variable to be added to the model is the forest density. Li and Magill (2001), identified *Bush density* as the most important factor of all for CA-based fire models. The minimum



density threshold for a cellular automaton with a 4-cell neighborhood, or Von Neumann neighborhood (North, East, West and South neighbors) is 59% (Resnick, 1994). Under 59% of density, the fire will not be able to spread on large areas. However, for an 8-cell neighborhood, or a Moore neighborhood, the density threshold is 41% (Li & Magill, 2015). Three classes of density were made: when the density was 0 percent, the burning chance was reduced to 0, when it was from 0 to 41 percent, the score was negative (-0.5) and when the density was higher than 41%, the score was positive (0.5). All the scores determined by the values presented in *Table VI* will be combined by the Fire Spread Algorithm (FSA).

*Table VI. Scores given to the cells for the algorithms*

<b>Variables</b>	<b>Possible values</b>	<b>Possible scores</b>
Initial Burn chance	Constant	2
Dominant species	black spruce or jack pine	+ 0.25
% wet areas (wetlands)	[>50, >20 and <= 50]	[-1, -0.5]
Fire Weather Index (Van Wagner, 1987)	30 + 17 – 29 9 – 16 5– 8 2 – 4 0 – 1	Extreme (+2) Very High (+1) High (+0.5) Moderate (+0.25) Low (+0) Very Low (-0.25)
Bush density (Li & Magill, 2001)	[0, >0 and < 41, >41]	[0, -0.5, 0.5]

After all the scores are defined for a cell, a burning chance (BC) is calculated for every cell, by dividing the sum of its scores by the maximum sum reached by a cell. This gives every cell a BC from 0 to 1 representing the chance of fire spreading over it, if one of its neighbors is previously ignited. The calculation of burning chance can be summarized with the following equation:

$$BC_{(i,j)} = (2+DS_{(i,j)}+ D_{(i,j)}+FWI_{(i,j)}+W_{(i,j)}) / MbC \quad (5)$$

where BC is calculated with a minimal chance of 1 for every cell, the dominant species score (DS), density score (D), fire weather index (FWI) and wet areas score (W) divided by the maximal BC of all cells (MbC). A cell with 0.85 BC would have an 85% chance of being ignited, while a cell with 0.25 would likely be spared. At every iteration, the burning chance is calculated for every cell. Therefore, it is

influenced by the state of the cell at the previous time step. If a cell has been burned in a previous step, it cannot be burned at the next step.

### Regrowth algorithm

A regrowth algorithm is used to determine the timing of post fire regrowth. This algorithm uses three values for the regrowth decision: 1) the presence of burning neighbors, 2) the presence of living neighbors and 3) the species found on the cell prior to ignition. For instance, if jack pines and black spruces were present before the fire, regrowth will be accelerated because of their serotinous and semi-serotinous cones that sprout when subject to intense heat, as fire (Barrett *et al*, 2010; Mansuy *et al*, 2012; Vasiliauskas & Chen, 2004). Regrowth chance (RC) is calculated with the following equation:

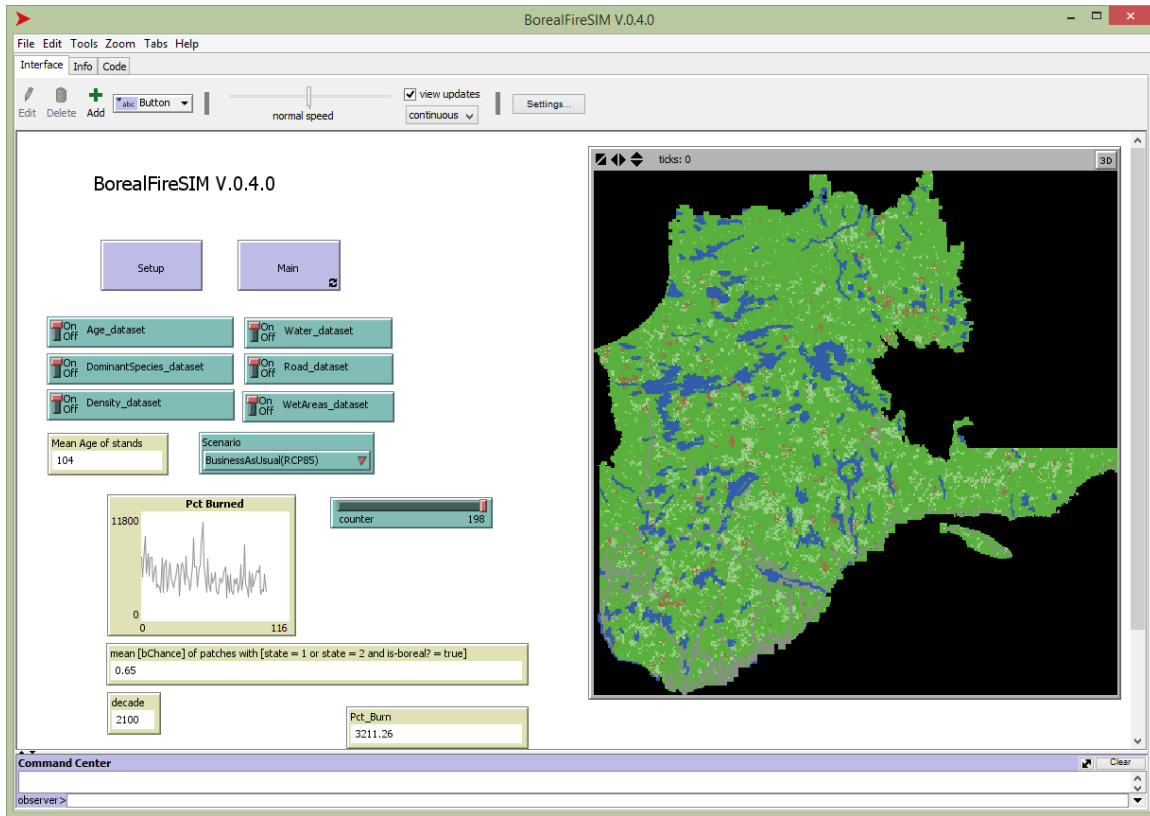
$$RC_{(i,j)} = (FN_{(i,j)} + LN_{(i,j)} + DS_{(i,j)}) / MrC \quad (6)$$

where the presence of burning neighbors, the presence of living neighbors and pre-fire dominant species scores are divided by the maximal RC of all cells.

### Model implementation

*BorealFireSim* model was programmed in NetLogo (Wilensky, 1999), a programmable modeling environment, and consists of a grid of 500x500 cells representing the boreal forest of Quebec. Its spatial resolution is 4.48 x 4.48 km, or ~20km<sup>2</sup> and the temporal resolution is one year. A graphical user-interface (GUI) was implemented to allow users to vary parameters, climate scenarios and datasets. The GUI (*Fig. 15*) was implemented directly in NetLogo and can be editable by the user to fit his needs. It also displays relevant information such as mean age of stands, area burned (km<sup>2</sup>), mean BC of flammable cells, current decade and a plot of the area burned for every year. Six variables can be switched on and off by the user to choose desired parameters: the age dataset, the dominant species dataset, the density dataset, the water bodies dataset, which is a shapefile containing the major lakes and rivers, the road dataset and the wet areas dataset. In the current version, two climate scenarios can be used: Low Emissions (RCP45) and Business As Usual (RCP85), which correspond to RCP4.5 and RCP8.5 (Table IV).

Figure 15. Graphical User Interface of BorealFireSim



Prior to the setup, users will define the cell size, climate scenario and the number of iterations desired. We recommend splitting the area in 500x500 cells, for a cell size of ~15km<sup>2</sup>. However, users with access to powerful computers could increase the number of cells for better results.

### Model outputs

Diverse information is exported by *BorealFireSim* during its execution, depending on the time step. At the end of every iteration, or every year, the model writes the current year, the current decade and the number of cells and area (km<sup>2</sup>) that has been burned during the current year to a comma separated text file (CSV). Every ten years, when the decade changes and the climate values for that decade replace the climate values of the preceding decade, the burning chance and cell state variables are both converted to a georeferenced raster dataset in *ASCII* format. Finally, at the end of the execution, when the model reaches 2100, a fire risk map is created and its values are the number of times each cell was burned. For instance, a cell that was burned twice over 100 years would have the value 2.

To assess of the stochasticity of the model, simulations with both climate scenarios were repeated 50, 100, 150 and 200 times. Since fire is ignited at a random location, two executions of the model could result in two different outputs. Moreover, since the values at time  $t$  are based on the values of  $t-1$ , the tiniest change in the first years could result in major differences at the final time step. By repeating the simulation for multiple times, we can reduce the uncertainties inherent to stochastic models.

### Contrasting *BorealFireSim* outputs

To assess the predictive ability of *BorealFireSim*, simulations ( $n=200$ ) were overlapped by a Boolean fire occurrence matrix (0 = unburned, 1 = burned). This dataset, provided by the Ministère des Forêts, de la Faune et des Parcs du Canada (MFFP), contains every single wildfire that was reported in Quebec from 2008 to 2014 (MFFP, 2015). The validation dataset is completely independent and was not used for calibration of the model. The model was calibrated according to literature, varying the parameters weights to have an annual average of 2800km<sup>2</sup> burned, with a standard deviation of 4900km<sup>2</sup>, being comparable to calculated annual sum of burned area in Quebec using the MFFP fire dataset.

In order to evaluate the implications of the *BorealFireSim* model outcomes by using future climate change scenarios, future fire likelihood maps were compared with forecasts from Nelson *et al.* (2014). In their study, Nelson *et al.* (2014) used climate data and an indirect indicator of biodiversity called dynamic habitat index (DHI) (Perez *et al.*, 2015) to forecast, by means of random forest algorithms, future spatial distribution of vegetation productivity based on climate change conditions. In this section we use both, the outputs from the *BorealFireSim* model and the outputs from Nelson *et al.* (2014), to visually assess the relationship of the two sets of outcomes using the same climate scenarios into the future.

Fire likelihood maps and maps from a DHI component (cumulative greenness) for 2050 and 2070 with low, medium and business as usual (BAU) scenarios were overlapped (Table IV) to compare spatial patterns. Cumulative greenness is an indication of overall vegetation productivity and the potential landscape productivity at any given location. It was expected that areas with high greenness would overlap with areas showing high fire likelihood, since both studies are based on climate variables.

## Results and discussion

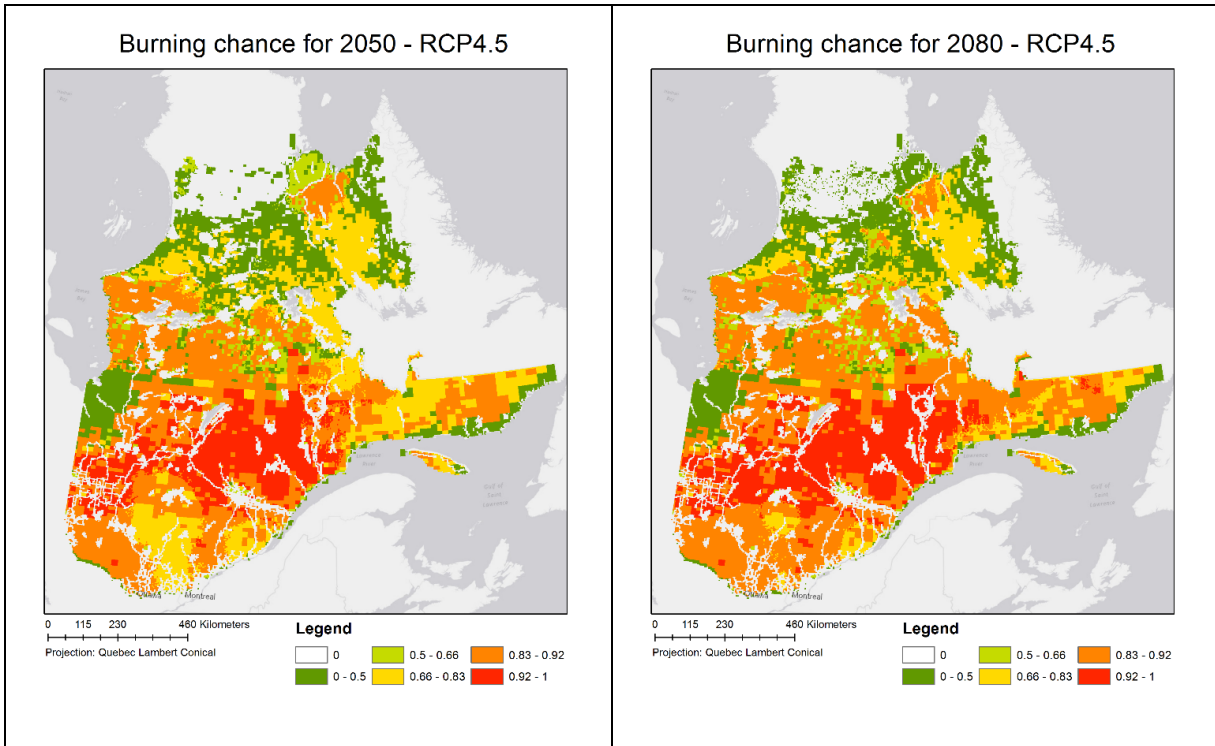
### Resulting maps

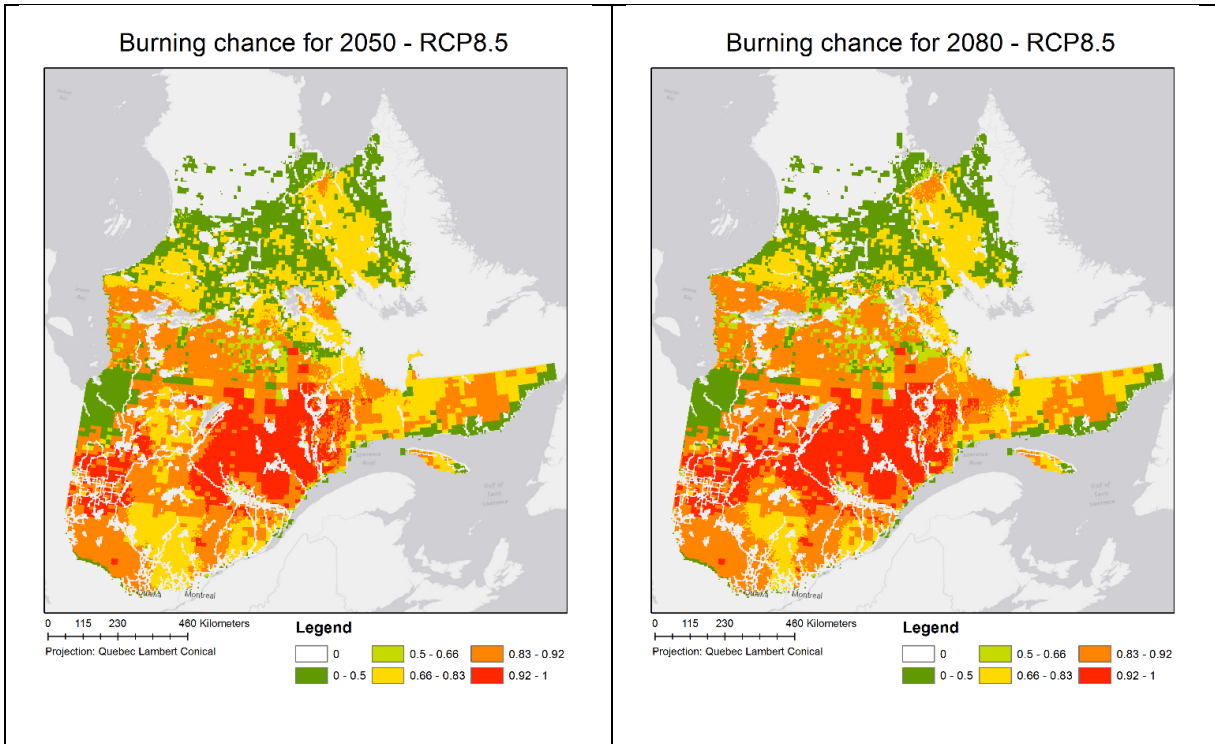
For the following maps, only the simulations with 200 repetitions for the years 2010, 2050 and 2080 are displayed. However, the maps for the 50, 100 and 150 repetitions are available in supplementary materials for every decade from 2000 to 2100.

### Burning chance

Figure 16 shows the burning chance calculated by the spreading algorithm. These maps show that the burning chance does not change drastically, even with a high emission scenario. The areas with the highest changes in burning chance are northwest and the center-east. Since the model is stochastic, the burning chance does not automatically drive fire events. If no fire is ignited on a cell with a high burning spread chance, there will be no fire. Since very scarce information can be retrieved from these maps, it was decided to create maps with the average state of the cells with multiple repetitions.

*Figure 16. Burning chance in 2050 and 2080 for the scenarios RCP4.5 and RCP8.5*

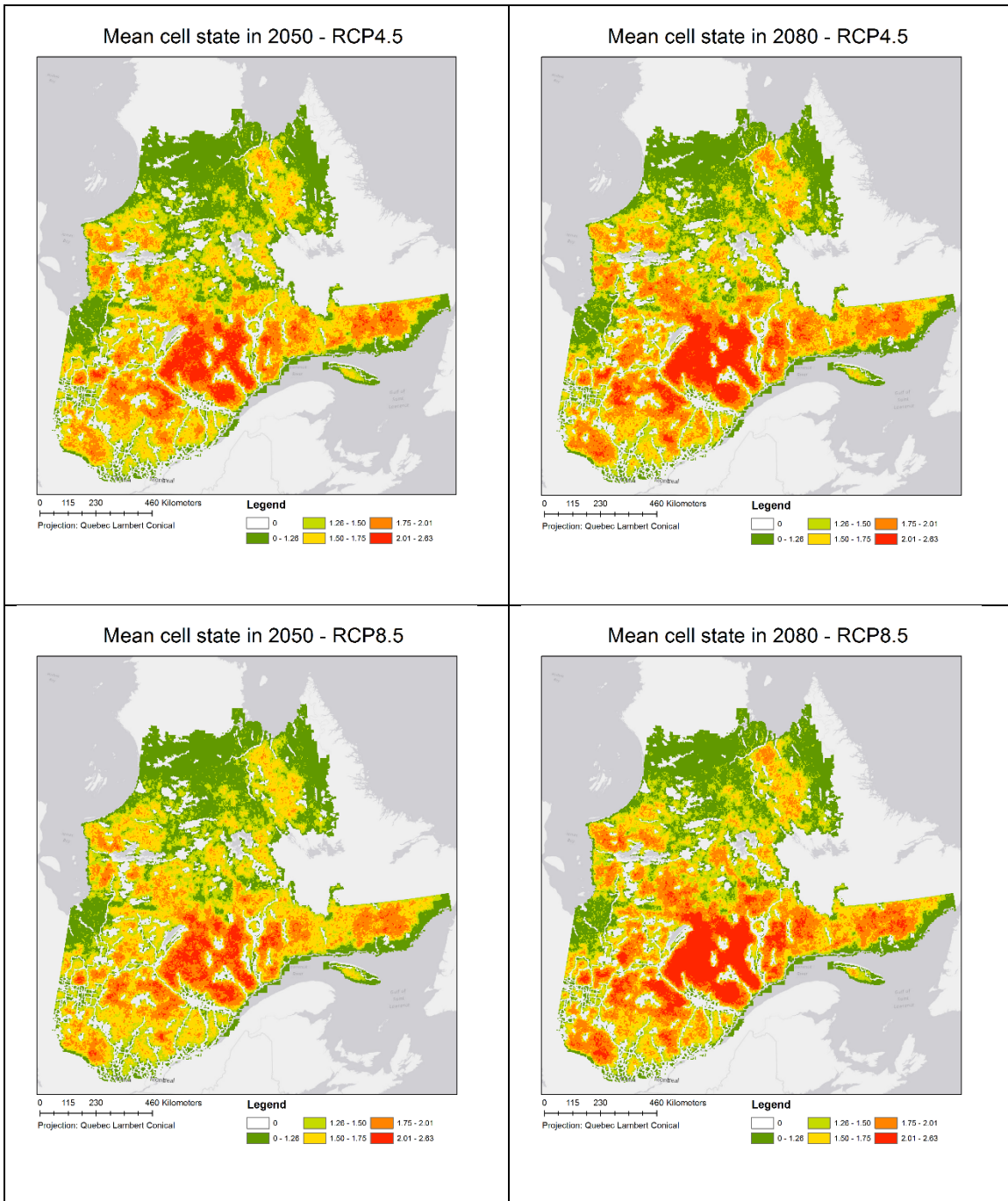




### Cell state change

Figure 17 shows the mean state of cells at years 2050 and 2080. Cells with a value near to 1 were unburned in most cases. The interesting part about the maps in figure 17 is the cells with values from 2 to 2.83. These cells were subject frequent change, while cells with a value of 0 or 1 did not change state often during the hundred iterations. On these maps, it is clearly identifiable that the northeast, the east and the center areas of the province will likely experience more fires.

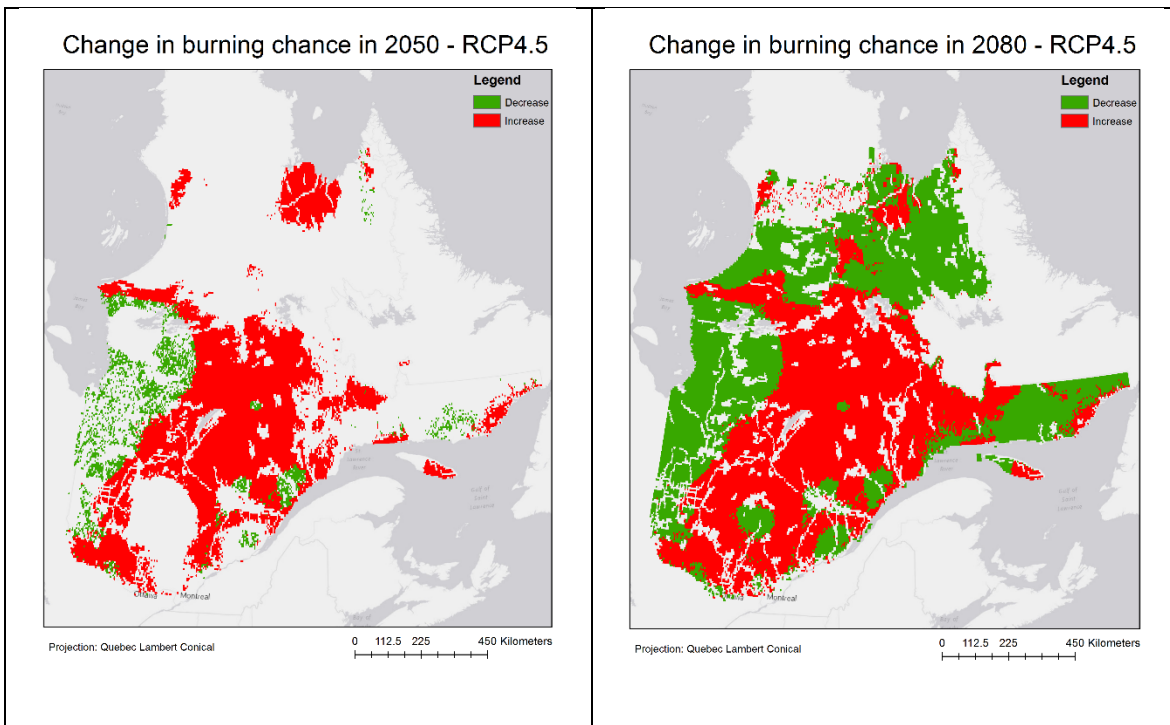
Figure 17. Cells state after 200 repetitions



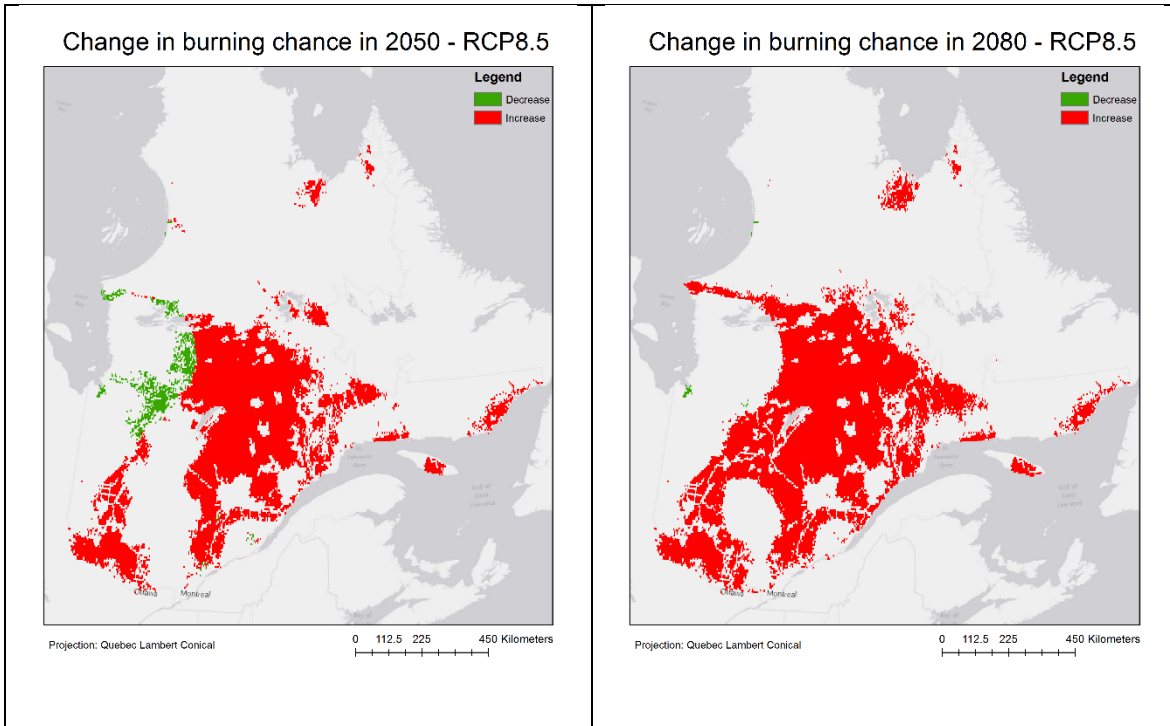
### Change maps

Maps presented in figure 18 allow better identifying the changes in fire risk. They represent the change in burning chance from 2010 to 2050 and from 2010 to 2080. We highlight that results for 2080 are based on results from 2010 to 2070 as well, since the model is dynamic. Areas in green are areas where the burning chance decreased, without regards to the intensity of the change, and areas in red are areas where the burning chance increased.

*Figure 18. Change in burning chance after 200 repetitions*





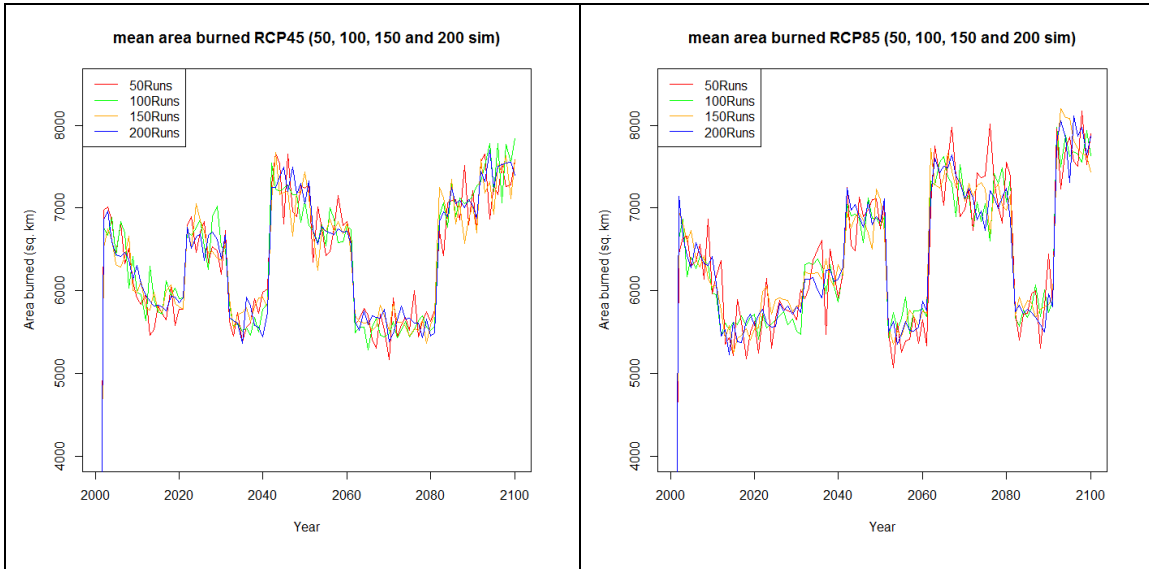


These maps show that with a low emission climate scenario such as the RCP4.5, continental areas are likely to expect an increase in fire spreading chance, while coastal and/or wet areas are likely to expect a decrease in fire spreading chance, because of the decrease in FWI. However, for a BAU scenario such as RCP8.5, while some areas might expect a decrease in fire spreading chance for 2050, there either no change or an increase everywhere for 2080, except for the James Bay lowlands, which are part of the greater Hudson's Bay lowlands, the third largest wetland region and the largest peatlands system in Canada, where more than 50% of the surface is covered in water (Dyke and Sladen, 2010)

### Summary of the model's outputs

Figure 19 shows the burned area per year with the RCP4.5 and RCP8.5 climate scenarios. The values are the mean values of the 50, 100, 150 and 200 repetitions. The simulations with 50 repetitions show more inter-annual variability, while simulations with 200 repetitions show less inter-annual variations.

Figure 19. BorealFireSim forecast area burned (square-km)



The mean area burned with RCP45 and RCP85 are very alike. However, by taking the values and calculating the percentage of increase, we get a clearer picture of the importance of climate change on changes in fire risk. Table VII shows

Table VII. Increase in mean area burned annually, RCP4.5 and RCP8.5

Mean area burned annually (KM2) - RCP4.5			
n runs	2010	2100	Change (%)
50	5897,767	7291,376	23,63
100	5872,539	7510,123	27,89
150	5756,418	7301,29	26,84
200	5855,098	7410,082	26,56

Mean area burned annually (KM2) - RCP8.5			
n runs	2010	2100	Change (%)
50	5795,127	7528,929	29,92
100	5743,225	7529,932	31,11
150	5836,994	7621,494	30,57
200	5857,536	7627,264	30,21

In all cases, change in area burned annually between 2010 and 2100 is more important with a business as usual scenario. However, burning chance is not the same in every part of the study area. In order to better represent the interprovincial climate variability, we calculated the mean burning chance of cells by bioclimatic domain for 200 repetitions of *BorealFireSim* for RCP4.5 and RCP8.5. Table VIII shows the mean burning chance of cells by bioclimatic domain in 2010 and 2100, as well as the percentage of increase.

*Table VIII. Increase in Mean Burning chance of cells*

<b>Mean burning chance by bioclimatic domain (RCP4.5)</b>			
<b>Domain</b>	<b>2010</b>	<b>2100</b>	<b>Change (%)</b>
<b>Forest tundra</b>	0.446	0.459	2.91
<b>Spruce-lichen</b>	0.739	0.776	5.01
<b>Spruce-moss</b>	0.764	0.78	2.09
<b>fir-white birch</b>	0.868	0.897	3.34
<b>fir-yellow birch</b>	0.844	0.878	4.03

<b>Mean burning chance by bioclimatic domain (RCP8.5)</b>			
<b>Domain</b>	<b>2010</b>	<b>2100</b>	<b>Change (%)</b>
<b>Forest tundra</b>	0.446	0.497	11.43
<b>Spruce-lichen</b>	0.739	0.789	6.77
<b>Spruce-moss</b>	0.764	0.78	2.09
<b>fir-white birch</b>	0.868	0.904	4.15
<b>fir-yellow birch</b>	0.844	0.885	4.86

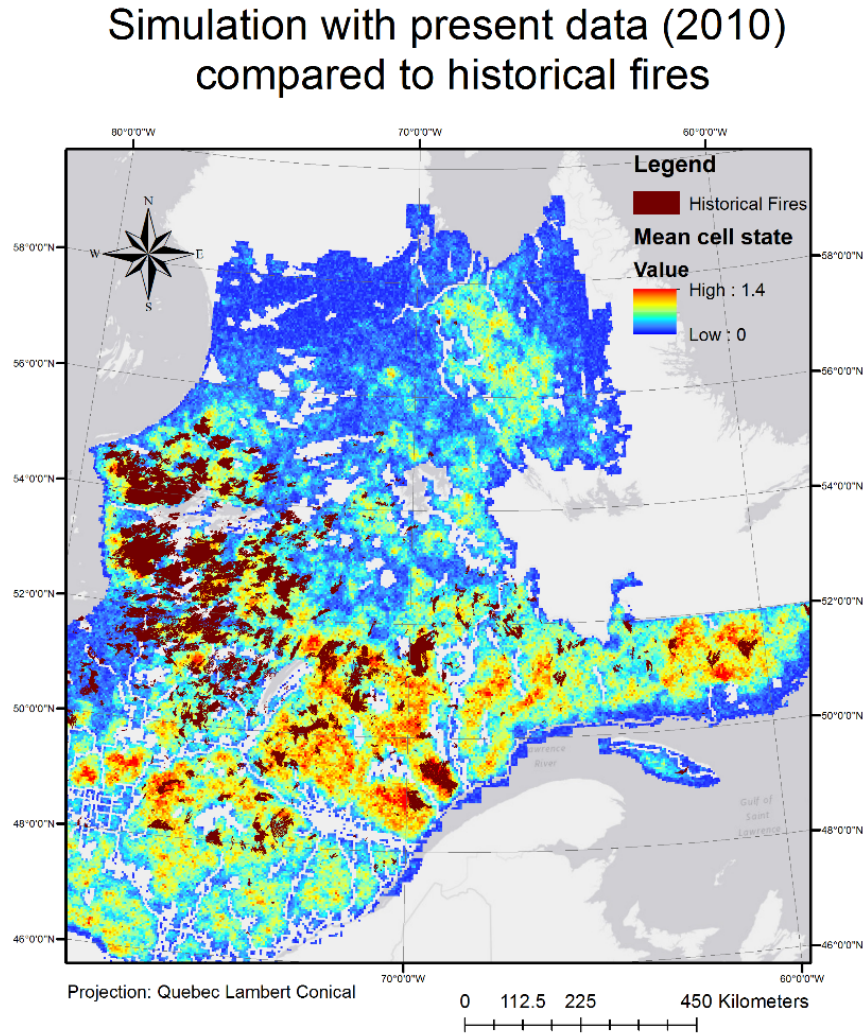
By comparing the change between RCP4.5 and RCP8.5, it is clear that forest tundra and, with lesser importance spruce-lichen domains will likely be more affected by climate change than the spruce-moss, fir-white birch and fir-yellow birch domains. If we compare table VII with table VIII, it is quite surprising that, for the entire study area, in a BAU scenario, the mean burning chance went from 0.732 to 0.771, which represents an increase of only 5.32%, however, it represented an increase of 30% in the area burned, passing from 5857 to 7627ha. This shows that a minor increase in burning chance for the individual cells leads to major increases in the spatial patterns of fire. When burning chance is high, a small increase leads to catastrophic results. However, an increase in burning chance does not automatically lead to more fires. Forest tundra had an increase of 11.43% of mean burning chance but the values passed from 0.446 to 0.497, still being very low. In a cellular automaton, more specifically in

*BorealFireSim*, a chance of spreading of 0.446 means that, for every neighbor of a burning cell, a random number between 0 and 5 will have to be under 0.446 for the fire to spread. The number 5 was decided during the calibration of the model. Multiple runs with different values were executed and the simulations with an interval of 0 to 5 exhibited a comparable number of fire events and of area burned annually with historical fires.

### Validation process

When overlapping *BorealFireSim* output with the MFFP Boolean fire dataset, the resulting map (fig. 20) demonstrates that *BorealFireSim* has very good predictive capacity for determining fire risks. Over the 52<sup>nd</sup> parallel, historical fires perfectly overlap with areas of high fire risk. However, under the 52<sup>nd</sup> parallel, there is a slight difference between forecast values and actual fires, and that is caused by the fact that fire suppressing activities are located under parallel 52. Statistics on the whole province and over parallel 52 were calculated. In order to validate the outcomes of *BorealFireSim*, the Map Comparison Toolkit (MCK) was used to calculate a fuzzy Kappa index (Visser & de Nijs, 2006). Kappa index is widely used metric to assess the similarity between two raster layers. The Kappa index results for the comparison between the simulation outcomes of *BorealFireSim* and the MFFP fire dataset showed a similarity between the two maps of 88.9% for the province and 94.9% over parallel 52. However, Kappa is getting more and more criticized and another metric, the *quantity disagreement and allocation disagreement* is considered as a solution (Pontius *et al.*, 2011). Being extremely minimalist, quantity disagreement and allocation disagreement relies on a contingency table for categorical variables. In the case of *BorealFireSim*, we reclassified the state map for 200 runs to zeros (0) and ones (1), for unburned and burned pixels respectively. This enables a comparison between *BorealFireSim* state map and the MFFP dataset. For unburned pixels, there is an agreement of 86% for the whole province and an agreement of 94% over parallel 52. For burned pixels, the agreement is of 50% for the whole province and 34% over parallel 52. These numbers show that *BorealFireSim* has power to identify areas not likely to burn, which account to 94% of the area with both the whole province and over parallel 52. Nevertheless, it is important to highlight that the difference between predicted burned cells and historical fires does not mean that *BorealFireSim* is wrong; it means that if there would be no anthropogenic effort to prevent and manage fires, we would likely see fire occurrences on these areas. In that sense, *BorealFireSim* would benefit from including a fire management sub model.

Figure 20. *Overlap of historical fires with the cell state map for present (2010)*



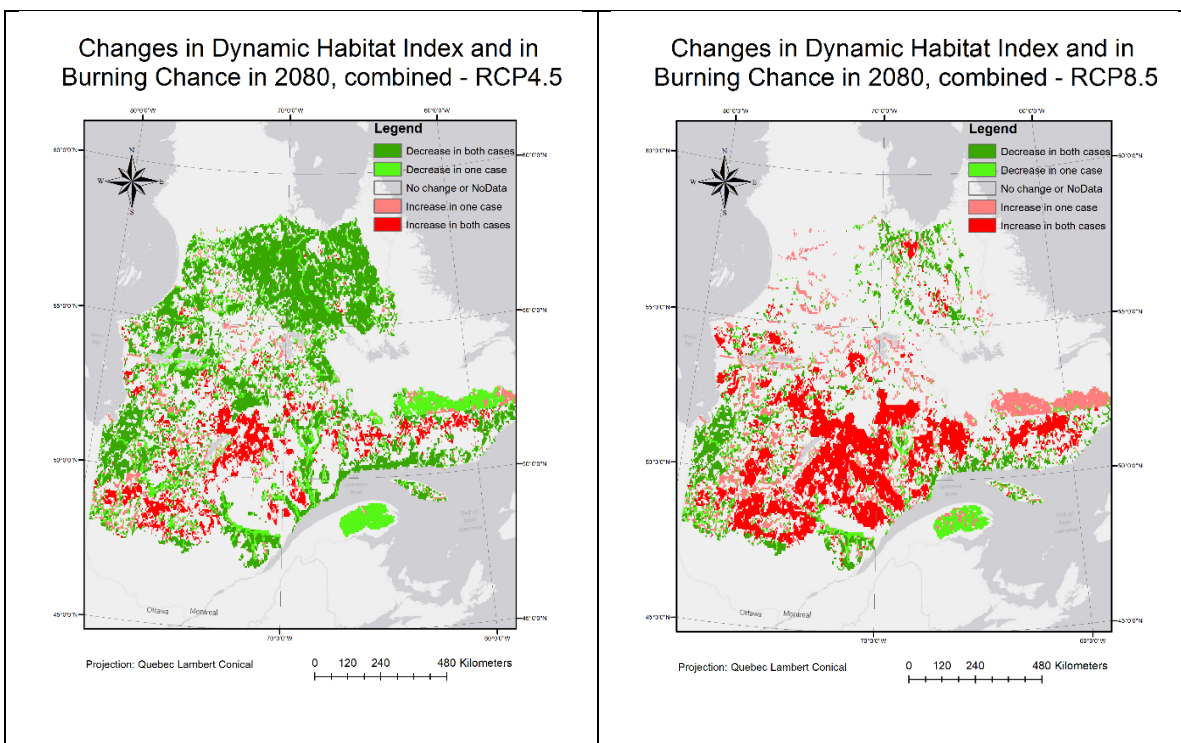
### 3.4. Contrasting model outputs with DHI forecasts

The comparison between dynamic habitat index and fire spread chance helped us identifying areas of good quality for fire specialist species. Dynamic habitat change was calculated by subtracting present DHI cell values to the 2080 DHI layer, for both scenarios, to get a value representing the change in DHI for each cell. The result gave the change in DHI, from -8500 to +8500 and the values under zero were reclassified as -1 and the values over zero, as 1. This gave binary maps of decrease/increase in

DHI for 2080. For the fire spreading chance maps, results for 2080 were reclassified as follow: zeros and NoDatAs were reclassified as 0, values under 0.7 were reclassified as -1 and values over 0.7, as 1. The two variables both having values of -1, 0 and 1, we summed the layers to get values from -2 to +2 (Figure 12). Calculations were made for 2080 with low emissions (RCP4.5 for the spreading chance and B1 for the DHI) and with Business As Usual (RCP8.5 for the spreading chance and A2 for the DHI)<sup>3</sup>.

The model shows that boreal regions are more likely to burn more frequently with climate change. Comparison with vegetation productivity forecasts offers many interesting insights. Figure 21 shows that areas with an increase in fire risks also exhibit an increase in greenness. *BorealFireSim* does not take into account changes in vegetation productivity. However, by comparing its results with greenness forecasts, possible feedbacks can be foreseen. On the opposite, areas with an increase in vegetation productivity do not mean an increase in fire risk, such as in wet areas and/or peatlands.

Figure 21. Combination of DHI and burning chance forecasts for 2080



The upcoming changes in fire disturbances modelled by *BorealFireSim* can be used in wildlife conservation. Numerous species depend on burned forest, or snags, either for feeding or habitat.

<sup>3</sup> As shown in table IV, RCP4.5 can be compared to B1 and RCP8.5 can be compared to A2.

Identifying future fire patterns give researchers foresights on the future distribution of fire dependent species, such as wood-boring beetles or bird species feeding on those beetles.

### Assumptions and limitations of *BorealFireSim*

*BorealFireSim* shows that climate change will likely induce changes in spatial patterns of wildland fires, and that these changes are not homogenous. Wildfire patterns will likely vary more at higher latitudes. Changes in wildfire occurrences might facilitate regeneration accidents, meaning that succession of disturbances, which can lead to a decrease in the age of stands in areas where fire risk, will increase. This could give rise to feedbacks and non-linear dynamics that would require more research. Interesting dynamics could also emerge with the inclusion of timber logging and post fire logging to wildfire dynamics. One of the limitations of *BorealFireSim* resides in the fact that the model does not aim to predict fire events at a given time or location. *BorealFireSim* aims to characterize the evolution of wildfires' spatial patterns. In that sense, kappa and traditional statistical validation are inadequate, because complex systems models like *BorealFireSim* are based on stochasticity where, in the present case, 200 simulations are executed, as opposed to reality, where there is only one timeline. Modellers in complex science, and users of *BorealForestSim* have to keep in mind that the model is a simplified vision of reality aiming to describe and provide insights on the evolution of wildfires and should in no case be used as an intelligent decision-making tool. Moreover, it would be interesting to take into account changes in forest composition during the forecast period in regards to climate change. These changes in tree species could give rise to interesting dynamics in the northernmost and southernmost edges of the boreal forest. In the future, land-use and forestry submodels could be added as add-ons to *BorealFireSim* so that fire suppression agents, in a similar manner to what was done by Alexandridis *et al.* (2011), could mimic Canada government efforts to extinguish fires located under parallel 52. By adding agents, *BorealFireSim* would benefit from more complexity and be more realistic.

### Conclusion

This research presents a novel GIS-based CA approached to simulate wildfires under climate change conditions. *BorealFireSim* is a powerful model for the identification of changes in wildfire dynamics in the boreal forest of Quebec, as it is based on numerous variables and uses a cellular automata approach, making it a dynamic and stochastic model. The results from *BorealFireSim* serve as a first step towards integrating the complexity of fire dynamics into species distribution models (SDM). Maps of future fire patterns can be used by researchers wanting to take into account future

disturbance dynamics into SDMs. The results show that, in all cases, the mean area of burned forest will likely increase with both a low emission of a high emission climate scenario, increasing by 30.2% by the end of the century with a Business As Usual (BAU) scenario. Knowing this, conservation agencies can plan their future policies by integrating forecasts from *BorealFireSim* and by focusing on protecting areas with higher fire risk and higher greenness, as they will be key areas for fire specialist species.

### **Acknowledgments**

The authors would like to thank the Center for Forest Research, Governments of Quebec and of Canada and experts who gave important feedbacks during the model early development phase. The authors would also like to thank Sylvie Gauthier and Aurélie Terrier whose expert advice helped parameterize *BorealFireSim*.



## References

- Alexandridis, a., Vakalis, D., Siettos, C.I., Bafas, G.V., 2008. A cellular automata model for forest fire spread prediction: The case of the wildfire that swept through Spetses Island in 1990. *Appl. Math. Comput.* 204, 191–201.
- Alexandridis, A., Russo, L., Vakalis, D., Bafas, G.V., Siettos, C.I. 2011. Wildland fire spread modelling using cellular automata: Evolution in large-scale spatially heterogeneous environments under fire suppression tactics. *International Journal of Wildland Fire* 20 (5): 633-647.
- Authors, D., Alexander, L., Allen, S., 2013. Working Group I Contribution to the IPCC Fifth Assessment Report Climate Change 2013: The Physical Science Basis Summary for Policymakers.
- Batty, M. and P. Torrens. (2001). *Modeling Complexity: The limits to Prediction*. CyberGeo: European Journal of Geography: Dossiers, 12<sup>th</sup> European Colloquium on Quantitative and Theoretical Geography, France, September 7-11 2001.
- Barrett, K. E.S. Kasischke, A.D.McGuire, M.R. Turetsky, E.S. Kane. (2010) Modeling fire severity in black spruce stands in the Alaskan boreal forest using spectral and non-spectral geospatial data. *Remote Sensing of Environment* 114: 1494-1503
- Bergeron, Y. S. Gauthier, V. Kafka, P. Lefort. D. Lesieur. (2006). *Natural fire frequency for the eastern Canadian boreal forest: consequences for sustainable forestry*. *Canadian Journal of Forestry* 31:384-391
- Bone, C., Dragicevic, S., Roberts, A., 2006. A fuzzy-constrained cellular automata model of forest insect infestations. *Ecol. Modell.* 192, 107–125.
- Bone, C., White, J., Wulder, M., Robertson, C., Nelson, T., 2013. Impact of Forest Fragmentation on Patterns of Mountain Pine Beetle-Caused Tree Mortality. *Forests* 4, 279–295.
- Bonnot, T. W., Millspaugh, J. J., & Rumble, M. a. (2009). Multi-scale nest-site selection by black-backed woodpeckers in outbreaks of mountain pine beetles. *Forest Ecology and Management*, 259, 220–228. doi:10.1016/j.foreco.2009.10.021
- Breiman, L. (2001). *Random Forests*. *Journal of Machine Learning* 45:1 pp 5-32.
- British Columbia Forest Service Wildfire Management Branch. (2014) Fire Resistant Trees. <http://bcwildfire.ca/Prevention/property/Landscape/fireresistantplants.htm>. Last access on 27/02/2015
- Byram, G.M. 1959. Combustion of forest fuels. Chapter 3 in K.P. Davis, ed. *Forest Fire: Control and Use*, McGraw-Hill, New York
- Canadian Forest Service (2001). Canada's Forest Inventory (CanFI) 2001 for the boreal eco-region - percent forested, Quebec. <http://databasin.org/datasets/a852f638bb424439a308b1aa0e8f1d6b>

- Dale, V. H., Joyce, L. A., McNulty, S., Neilson, R. P., Di, S., Ridge, O., & Box, P. O. (2008). The interplay between climate change, forests, and. *BioScience*, 201–204.
- Dale, V. H., Joyce, L. A., McNulty, S., Ronald, P., & Matthew, P. (2001). Climate Change and Forest Disturbances. *BioScience*, 51(9), 723–734.
- De Groot, W. (2012). *CanFIRE*. Natural Resources Canada. Canadian Forest Service. Great Lakes Forestry Centre. Sault Ste. Marie, Ontario. Frontline Express 62. 2p.
- Drever, C. R., C. Messier, Y. Bergeron, F. Doyon. (2006) *Fire and canopy species composition in the Great Lakes-St. Lawrence forest of Témiscamingue, Québec*. *Forest Ecology and Management* 231: 27-37.
- Dyke, L.D., Sladen, W.E. 2010. *Permafrost and Peatland Evolution in the Northern Hudson Bay Lowland, Manitoba*. *Arctic Institute of North America* 63: 4.
- Energie et Ressources Naturelle Québec (2013) *Bref portrait de la forêt boréale au Québec*. <http://www.mern.gouv.qc.ca/presse/feux-grands.jsp>. Online.
- Finney, M. A., Andrews, P. L. 1999. FARSITE: Fire Area Simulator—a program for fire growth simulation. *Fire Management Notes* 59(2): 13-15.
- Flannigan, M. D., Stocks, B. J., & Wotton, B. M. (2000). Climate change and forest fires. *The Science of the Total Environment*, 262(3), 221–229. doi:[http://dx.doi.org/10.1016/S0048-9697\(00\)00524-6](http://dx.doi.org/10.1016/S0048-9697(00)00524-6)
- Forêt, Faune et Parcs Québec (MFFP), 2013. *Superficies affectées par les feux de forêt*. <http://www.mffp.gouv.qc.ca/publications/enligne/forets/criteres-indicateurs/2/213/213.asp>. Online.
- Forêt, Faune et Parcs Québec (MFFP), 2015. Système hiérarchique de classification écologique du territoire: niveaux du domaine bioclimatique. ESRI Shapefiles <http://www.mffp.gouv.qc.ca/forets/inventaire/inventaire-systeme.jsp>
- Green, K., Finney, M. A., Campbell, J., Weinstein, D., Landrum, V. 1995. Fire! Using GIS to predict fire behavior. *Journal of Forestry* 93(5): 21-25.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25, 1965–1978.
- Keane, R.E., Morgan, P., White, J.D. 1998. Temporal patterns of ecosystem processes on simulated landscapes in Glacier National Park, Montana, USA. *Landscape Ecology* 14 (3): 311-329.
- IPCC, 2013: Summary for Policymakers. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Stocker, T.F., D. Qin, G.-K. Plattner, M.

Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)].  
Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Keane, R.E., Holsinger, L.M., Parsons, R.A., Gray, K. 2008. Climate change effects on historical range and variability of two large landscapes in western Montana, USA. *Forest Ecology and Management* 254 (3): 375-389.

Kocabas, V., Dragicevic, S., 2006. Coupling Bayesian networks with GIS-based cellular automata for modeling land use change. *Geogr. Inf. Sci.* 217–233.

Kocabas, V., Dragicevic, S., 2007. Enhancing a GIS Cellular Automata Model of Land Use Change: Bayesian Networks, Influence Diagrams and Causality. *Trans. GIS* 11, 681–702.

Li, X., Magill, W., 2001. Modeling fire spread under environmental influence using a cellular automaton approach 08, 1–14.

Mantyka-Pringle, C. S., Visconti, P., Di Marco, M., Martin, T.G., Rondinini, C., Rhodes, J.R. Climate change modifies risk of global biodiversity loss due to land-cover change, *Biological Conservation*, Volume 187, July 2015, Pages 103-111, ISSN 0006-3207

Ministère des Forêts, de la Faune et des Parcs (MFFP). 2015. Ressources et industries forestières: Portrait statistique, édition 2015.

Ministère des Ressources naturelles et de la Faune (MRN). 2009. Classification of Forest Stands as Fuels According to the Canadian Forest Fire Behavior Prediction (FBP) System.

Natural Resources Canada (2006). Authors: K. Power and M. Gillis. Canada's forest inventory 2001. Information Report BC-X-408

Natural Resources Canada (2008). Canadian Forest Fire Weather Index (FWI) System.  
<http://cwfis.cfs.nrcan.gc.ca/background/summary/fwi>. Last access: 11/07/2015

Nappi, A., Drapeau, P., Giroux, J., Savard, J., & Moore, F. (2003). Snag use by foraging black-backed woodpeckers (*Picoides arcticus*) in a recently burned eastern boreal forest. *The Auk*, 120(2), 505–511.

Natural Resources Canada. (2014). Forest Topics: Fire. <http://www.nrcan.gc.ca/forests/fire/13143>. Last access: 27/02/2015

Nelson, T., Coops, N., Wulder, M., Perez, L., Fitterer, J., Powers, R., Fontana, F., 2014. Predicting Climate Change Impacts to the Canadian Boreal Forest. *Diversity* 6, 133–157.

Perez, L., Dragicevic, S., 2010. Modeling mountain pine beetle infestation with an agent-based approach at two spatial scales. *Environ. Model. Softw.* 25, 223–236.

- Perez, L., Dragicevic, S., 2012. Landscape-level simulation of forest insect disturbance: Coupling swarm intelligent agents with GIS-based cellular automata model. *Ecol. Modell.* 231, 53–64.
- Perez, L., Nelson, T., Coops, N., Fontana, F., Drever, C.R., 2015. Characterization of spatial relationships between three remotely sensed indirect indicators of biodiversity and climate: a 21 year's data series review across the Canadian boreal forest. *Int. J. Digit. Earth* Accepted.
- Pontius Jr, R. G., Millones, M., 2011. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing* 32(15) 4407-4429.
- Resnick, M. (1994). *Turtles, Termites and Traffic jams*, The MIT press, Cambridge, Massachusetts.
- Rindfuss, R.R., Entwisle, B., Walsh, S.J., An, L., Badenoch, N., Brown, D.G., Deadman, P., Evans, T.P., Fox, J., Gutmann, M., Kelly, M., Linderman, M., Liu, J., George, P., Mena, C.F., Messina, J.P., Moran, E.F., Parker, D.C., Prasartkul, P., Robinson, D.T., Sawangdee, Y., Vanwey, L.K., 2010. Land use change : complexity and comparisons 3, 1–10.
- Ronnie, C. Å., & Mark, C. (2009). Effects of climate on occurrence and size of large fires in a northern hardwood landscape : historical trends , forecasts , and implications for climate change in Te, (Rowe 1972), 261–272.
- Singh, A.K., 2003. *Modelling Land Use Land Cover Changes Using Cellular Automata in a Geo-Spatial Environment*.
- Tymstra, C.; Bryce, R.W.; Wotton, B.M.; Taylor, S.W.; Armitage, O.B. 2010. *Development and structure of Prometheus: the Canadian Wildland Fire Growth Simulation Model*. *Nat. Resour. Can., Can. For. Serv., North. For. Cent., Edmonton, AB. Inf. Rep. NOR-X-417*.
- Vasiliauskas, S. H.Y.H. Chen, 2004. Post-fire Regeneration of Boreal Forests in Northeastern Ontario. *Parks Research Forum of Ontario Proceedings 2004 (2004 PRFO Proceedings)*.
- Vuuren, Detlef P., Jae Edmonds, Mikiko Kainuma, *et al.*, 2011, "The Representative Concentration Pathways: An Overview", *Climatic Change* vol. 109 No 1 and 2, 5–31.
- Ward, D.P., Murray, a. T., Phinn, S.R., 2000. A stochastically constrained cellular model of urban growth. *Comput. Environ. Urban Syst.* 24, 539–558.
- Yassemi, S., Dragičević, S., Schmidt, M., 2008. Design and implementation of an integrated GIS-based cellular automata model to characterize forest fire behaviour. *Ecol. Modell.* 210, 71–84.
- Yeh, A., Li, X., 2003. Error propagation and model uncertainties of cellular automata in urban simulation with GIS. *7th Int. Conf. GeoComputation, ...* 1–16.

## Conclusion générale

Les méthodes de modélisation statistiques permettent d'identifier l'étendue annuelle des températures, les températures moyennes du trimestre le plus frais, la saisonnalité des précipitations, les précipitations du trimestre le plus arrosé et l'élévation comme variables expliquant la distribution des oiseaux de la forêt boréale. Ces variables ont permis de modéliser l'aire de répartition de deux espèces d'oiseaux de la forêt boréale du Québec, soient le gros-bec errant et la mésange à tête noire qui, en présence de changements climatiques, verront leur limite nordique se déplacer vers le nord, suivant le rythme du changement climatique de façon plus exacerbée le long des côtes qu'en milieu continental. Outre les changements dans la répartition spatiale de certaines espèces d'oiseaux, les changements climatiques prévus induiront aussi des modifications dans les patrons spatiaux de feux de forêt. En effet, le modèle spatialement explicite d'automate cellulaire *BorealFireSim* permet d'identifier les changements dans les dynamiques de feux en forêt boréale du Québec entre 2010 et 2100. Les résultats de *BorealFireSim* permettent de prendre en considération l'importance des feux dans la génération de modèles de répartition d'espèces. Le modèle, en lien avec divers scénarios de changements climatiques, prévoit en moyenne une augmentation de 30% dans la superficie brûlée avec un scénario BAU. Finalement, ces résultats permettent de considérer les approches de modélisation de systèmes complexes comme outils préférentiels pour l'identification de tendances futures, qu'elles traitent d'espèces animales, tels les oiseaux, ou de phénomènes naturels tels les feux de forêt.

En prenant en considération les résultats de ces deux articles, les chercheurs et preneurs de décisions seront en mesure de considérer les multiples facettes des changements climatiques lors de la prise de décision, afin de mettre l'accent sur la protection de zones à haut risque de feux et faisant office de points chauds de renouvellement d'espèces.

