

DÉPARTEMENT DE GÉOMATIQUE APPLIQUÉE  
Faculté des lettres et sciences humaines  
Université de Sherbrooke

# Rapport de stage

Sous le thème

Évaluation d'un indice de qualité d'habitat des plants de bleuets  
sauvages

Effectué par :

Donard Dzeutouo Zapa

Au



Supervisé par

Marc-André Goderre

Professeur responsable :

Richard Fournier

Mickaël Germain

SHERBROOKE  
Novembre 2016

## Table des matières

Liste des figures .....	i
Liste des tableaux .....	i
Résumé.....	1
1. Introduction .....	2
2. Méthodologie.....	6
4.1 Zone d'étude .....	7
4.2 Protocole d'échantillonnage et collecte des données terrains .....	8
4.3 Traitement des données .....	9
4.4 L'intégration dans l'interface cartographique .....	10
4.5 L'analyse statistique .....	10
1.5.1 Test d'identification des variables pertinentes .....	10
1.5.2 Régression logistique .....	15
1.5.3 Évaluation de modèles .....	18
1.5.4 Classement des modalités. ....	21
3. Évaluation et recommandation.....	22
References .....	27
4. Annexes.....	30
4.1 Annexes 1: Cartographie des sous zone de végétation .....	30
4.2 Annexe 2 : Cartographie des domaines bioclimatiques .....	30
4.3 Annexe 3 : Fiche de relevé terrain.....	31
4.4 Annexe 4 : Indice V de cramer - Intensité du lien entre les variables.....	32

## Liste des figures

Figure 1: Production annuelle du bleuets sauvage Saguenay-Lac-Saint-Jean .....	5
Figure 2: Schéma méthodologique pour l'évaluation d'un IQH des plants de bleuets.....	7
Figure 3: Carte montrant les limites de la zone d'étude.....	8
Figure 4: Résultat du modèle 1 sous R.....	17
Figure 5: Résultat du modèle 2 sous R.....	18
Figure 6: Résultat du modèle 3 sous R.....	18
Figure 7: Représentation graphique des rapports de cotes du modèle 2.....	23
Figure 8: Variabilité du pourcentage de recouvrement moyen des plants en fonction du type de perturbation .....	24
Figure 9: Variabilité du pourcentage de recouvrement moyen des plants en fonction du type écologique .....	24
Figure 10: Variabilité du pourcentage de recouvrement moyen des plants en fonction du drainage .....	25
Figure 11: Variabilité du pourcentage de recouvrement moyen des plants en fonction de la densité du couvert forestier.....	25
Figure 12: Variabilité du pourcentage de recouvrement moyen des plants en fonction de l'âge de la perturbation .....	26

## Liste des tableaux

Tableau 1: Superficies aménagées en bleuetières – 2015.....	5
Tableau 2: Tableau de contingence de la classe de pente et du recouvrement en plant de bleuets sauvage .....	12
Tableau 3: Effectif théorique entre la classe de pente et du recouvrement en plant de bleuets sauvage en cas d'indépendance des variables .....	12
Tableau 4: Tableau de contingence de la classe de pente et du recouvrement en plant de bleuets sauvage après regroupement.....	13
Tableau 5: Test de significativité des variables.....	14
Tableau 6: Interprétation de l'indice V de Cramer .....	16
Tableau 7: Pouvoir prédictif du modèle 1 pour 38 sites d'échantillonnages.....	19
Tableau 8: Pouvoir prédictif du modèle 2 pour 38 sites d'échantillonnages.....	19
Tableau 9: Pouvoir prédictif du modèle 3 pour 38 sites d'échantillonnages.....	20
Tableau 10: Comparaison des modèles selon le critère AIC et les prédictions incorrectes .....	20

## Résumé

Ce rapport s'inscrit dans le cadre d'un stage académique effectué en été 2016 au Centre de géomatique du Québec (CGQ) à Chicoutimi. C'est un organisme parapublic à but non lucratif qui possède le statut de centre collégial de transfert de technologie (CCTT) conféré par le ministère de l'Éducation et de l'Enseignement supérieur (MEES) et par le ministère de l'Économie, des Sciences et de l'Innovation (MESI). La mission du CGQ est de servir de levier à l'innovation et à la performance chez ses partenaires par l'utilisation et le développement de la géomatique. En partenariat avec le Syndicat des producteurs de bleuets du Québec (SPBQ), l'Agence de gestion intégrée des ressources et l'Université de Sherbrooke, le CGQ a voulu par ce stage utiliser des outils statistiques pour l'évaluation des indices de qualité d'habitat (IQH) qui sous-tendent le portail cartographique web de valorisation des produits forestiers non ligneux (PFNL). Ces statistiques visent à répondre à deux questionnements majeurs qui sous-tendent le développement et la fiabilité des IQH de bleuets sauvages à savoir :

1. les variables environnementales les plus pertinentes pour la présence des plants de bleuets sauvages
2. les modalités qui favorisent ou limitent le potentiel de présence des plants de bleuet pour chacune des variables pertinentes.

Pour cela une campagne d'échantillonnage nous a permis de noter les variables environnementales de 157 sites, qui ont permis de bâtir et d'évaluer les modèles de régression logistique. À l'issue de ces analyses, les variables telles que la densité du couvert forestier, le drainage, le type écologique, l'âge depuis la dernière perturbation et le type de perturbation se sont avérées pertinentes pour la présence ou l'absence des plants de bleuets sauvages. Bien que la densité du couvert forestier et le type de perturbation ne semblaient pas significatifs au seuil de 5% dans les modèles, nous avons constaté que le meilleur modèle prédictif s'est avéré être celui qui contenait ces deux variables. De plus, les analyses ont révélé que contrairement aux croyances actuelles, les types écologiques RE\* (RE1, RE2, RE3) sont meilleurs pour les plants de bleuets que les autres types écologiques.

## 1. Introduction

Le bleuet sauvage aussi connu sous le nom de bleuet nain, bleuet feuille-dentelée ou airelle à feuilles étroite est une espèce d'arbustes à petits fruits de la famille des éricacées et du genre *Vaccinium*. Cette plante a été décrite pour la première fois par le botaniste écossais William Aiton en 1789. Les bleuets sauvages sont fréquemment trouvés dans les zones tropicales à haute altitude, mais aussi dans les régions froides et tempérées. Elles poussent sur toutes les terres acides de l'hémisphère nord composées de 25% de matière organique, ils abondent spécialement dans les régions montagneuses où ils colonisent de vastes espaces, principalement sur une latitude supérieure à 44° dans le nord des États-Unis et au Canada (Ketata, 2012). Le Québec, le Nouveau-Brunswick, la Nouvelle-Écosse, et l'Île-du-Prince-Édouard sont les principales provinces productrices de bleuets sauvages au Canada. Toutefois, le Québec se démarque avec plus de 40 % de la production canadienne, concentrée principalement dans la région du Saguenay-Lac-Saint-Jean, la Côte-Nord et l'Abitibi-Témiscamingue (SPBQ, 2014). Selon Chiasson et Agrall (1996), il existe cinq espèces de bleuets qui poussent à l'état sauvage au Canada, dont les trois plus répandues sont : (1) l'airelle à feuilles étroites (*Vaccinium angustifolium* Ait), (2) airelle fausse myrtille (*Vaccinium myrtilloides* Michx) et (3) l'airelle en corymbe (*Vaccinium corymbosum* L.). Comme dans le cas de toutes les plantes, la croissance et le développement des plants de bleuet sauvage sont étroitement liés aux conditions environnementales. Plusieurs études ont mis en relief l'influence de certains facteurs environnementaux sur la croissance et le développement des plants de bleuet sauvage. C'est par exemple le cas de l'étude menée à l'échelle des forêts de France par Coudun et Gégout (2007). Ces auteurs ont démontré que le pH du sol et le rapport carbone azote (C/N) permet de prédire la distribution et la présence des plants de bleuets sauvages. Plus encore, ils ont démontré que les plants de bleuets sauvages ont une préférence pour les sites très froids (en particulier les montagnes) et les sols pauvres en élément nutritif (faible pH avec un rapport C/N élevé). Cette dépendance des plants de bleuets aux propriétés chimiques du sol a également été relevée par Farooque et al. (2012) dans une étude en Nouvelle-Écosse. D'après cette étude, les valeurs de l'azote inorganique du sol, de matière organique, de la conductivité électrique, de l'argile, et de l'humidité volumétrique étaient généralement plus élevées dans les zones de fort potentiel de présence des plants de bleuets et vice versa. Contrairement au sable et au limon donc la teneur augmentait avec la diminution du

potentiel de présence des plants bleuets. En plus de ces facteurs, Hester et al, (1991); Atlegrim et Sjöberg, (1996); Kim et al, (2011); Haijun et al, (2013) ont établi que les caractéristiques de croissance et de reproduction telles que le nombre de fleurs, le taux de mise en fruits par bourgeon et le rendement en fruits augmentent de façon significative avec l'augmentation de l'éclairage solaire.

Au Canada en général, et au Québec en particulier, on dispose de très peu d'informations formelles sur les besoins environnementaux du bleuet sauvage en milieu forestier, ceci en raison du fait que les rares études qui traitent du sujet sont plus portées sur les bleuets sauvages en milieu de production. Parmi les études faites au Québec, on peut citer celle réalisée dans le cadre des travaux de Forêt modèle du Lac-Saint-Jean, sur l'évaluation comparative de la productivité du bleuet sauvage en milieu naturel. Cette étude a rapporté que le type écologique correspondant à une pessière noire à mousse, sur des sols très pauvres semble favoriser le développement de plants de bleuets en milieu forestier, tandis que la récolte en bois court qui nuit à leur développement. Les auteurs pensent que cette nuisance s'explique par le fait que les résidus de coupe créent un paillis trop épais qui limite la croissance des nouvelles tiges (Coopérative de solidarité forestière de la rivière aux saumons, 2011). En plus de cela, Drummond et al., 2008; Holloway, (2006); Duchesne et Wetzel, (2004); Waldrop *et al.*, (1992); Duval, (1993) précise que les plants de bleuets apparaissent naturellement en abondance après un feu de forêt ou une coupe, car c'est une plante qui survit généralement dans les sous-bois jusqu'à ce qu'un feu ou une coupe se produise. Après une telle perturbation, le sol bénéficie de plus de rayonnement solaire et d'un apport important de minéraux provenant des cendres des arbres et des débris végétaux. Ces facteurs permettent alors une forte croissance des plants de bleuets, qui sont en fait une espèce de transition dans les écosystèmes forestiers de conifères. C'est une plante colonisatrice qui exige beaucoup de lumière. Peu à peu, de quelques années à une dizaine selon la fertilité des sites, cette plante fait place à d'autres espèces de transition qui l'ombragent (framboisiers, bouleaux) et qui le font reprendre sa place restreinte comme espèce de sous-bois. Néanmoins, bien que le feu soit documenté comme un facteur propice au développement des plants de bleuets, certains auteurs comme Vickery *et al.*, (2005); Kautz, (1987) ; Neilson et Wood, (1985) ont démontré que ce facteur n'est bénéfique que si le feu a lieu quand les plants sont en dormance tard à l'automne ou tôt au printemps, et quand le sol est encore humide. Car en période sèche, la matière organique du sol ainsi que

les rhizomes souterrains des plants de bleuets peuvent être gravement endommagés par feu. Duchesne et Wetzell, (2004) ont ajouté que le potentiel de développement des plants de bleuets après un feu était également dépendant de l'intensité du feu et de la profondeur de pénétration du feu. Cette étude réalisée en Ontario, a démontré que les faibles feux donc l'intensité varie entre 597 KW/m et 1268 KW/m, sont propices au développement des plants de bleuets, contrairement aux feux de moyens et de fortes intensités (intensité supérieure à 1268 KW/m). Ainsi, il ressort que le potentiel de présence des plants de bleuets est soumis à une multitude de contraintes environnementales liées aux facteurs naturels comme l'altitude, la pente, le PH du sol, l'éclairage solaire, la texture du sol, le drainage, les feux et aux facteurs anthropiques comme la coupe forestière et les feux d'origines humaines.

Dans la région du Saguenay-Lac-Saint-Jean, les superficies aménagées pour la production du bleuet sauvage représentent près de 82% de l'ensemble des bleuetières de la province (Tableau 1), faisant de la production du bleuet, la deuxième production agricole en importance, après la production laitière, devançant ainsi la production céréalière de cette région. De plus, cette production représente environ 95 % de la production provinciale du bleuet sauvage (Olivier Paré, 2013). Par ailleurs, la contribution du bleuet sauvage cueilli en forêt dans la production totale du bleuet reste extrêmement faible dans cette région (Figure 1). Le manque d'outil efficace pour identifier les zones de présence des plants de bleuet en forêt est présentement une limite importante pour le développement de cette industrie, car on estime que moins de 15% du bleuet disponible en forêt est cueilli chaque année. D'où la nécessité de développer un modèle local qui intègre les variables locales pertinentes à l'identification des zones de présence de bleuets sauvages en milieu forestier.

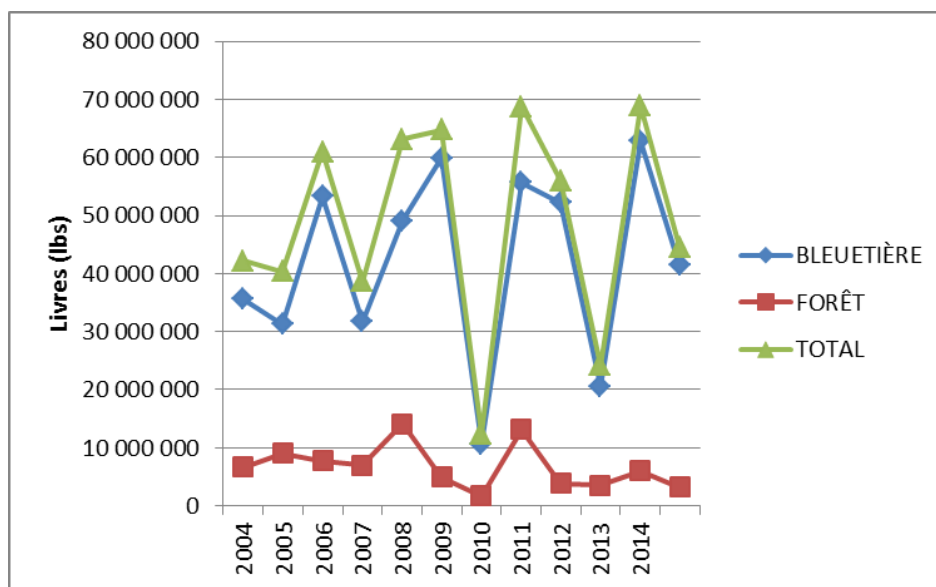


Figure 1: Production annuelle du bleuets sauvage Saguenay-Lac-Saint-Jean  
Source : Syndicat des producteurs de bleuets

Régions	Superficies aménagées en bleuetières – 2015	
	Hectares	%
Saguenay-Lac-Saint-Jean	29 030	82 %
Côte-Nord	4 595	13 %
Abitibi-Témiscamingue	983	3 %
Autres régions	971	3 %
<b>Total</b>	<b>35 579</b>	<b>100 %</b>

Tableau 1: Superficies aménagées en bleuetières – 2015  
Source : MAPAQ, Direction régionale du Saguenay-Lac-Saint-Jean, mars 2016

Au cours des deux dernières années, un grand nombre de modèles locaux de prédiction de présence des bleuets sauvages ont été développés sur la base des IQH. Les IQH sont des modèles théoriques qui permettent de juger, sur une échelle relative normalisée de 0 à 1, la valeur d'un milieu à répondre aux besoins en habitat d'une espèce (Julie, 2012). Introduit il y a plusieurs années au Québec, le concept d'IQH est souvent utilisé pour évaluer globalement la qualité d'habitat et intégrer les préoccupations relatives à une espèce dans le développement de certains projets, tels que la planification de l'aménagement forestier. Dans la pratique, la cote relative d'un milieu (sa valeur d'IQH) est fonction de quelques



variables que l'on sait importantes pour l'espèce cible, à laquelle on accole une valeur relative. Ces variables qui s'additionnent, se multiplient ou se divisent entre elles pour arriver à une cote finale, sont généralement issues des connaissances de l'écologie de l'espèce et d'opinions d'experts (Crête, 2003). En plus de refléter une simplification du milieu naturel, des facteurs autres que la qualité de l'habitat, lesquels ne sont pas considérés dans ces modèles, peuvent influencer leur robustesse à une échelle locale, comme la compétition avec d'autres espèces végétales (Schamberger et O'Neil, 1986). Puisque ces modèles d'IQH ne sont basés que sur des connaissances théoriques, ils doivent donc être validés pour établir une concordance avec une variable qui atteste de la présence de l'espèce dans le milieu, sans quoi un modèle d'IQH n'aurait aucune signification pour l'espèce cible. Selon Roloff et Kernohan, (1999), la validation d'un modèle d'IQH devrait porter sur sept critères à savoir : la précision des données (les entrants), les variables du modèle, la validité des tests comparatifs, l'échelle de la validation, les classes de valeurs de l'IQH, la durée de la validation et la densité de l'espèce.

Cette étude vise comme objectif principal l'évaluation des IQH de bleuets sauvages en milieu forestier par régression logistique. Il s'agit donc principalement d'utiliser les outils statistiques pour établir un lien entre le recouvrement en plant de bleuet et les variables environnementales pertinentes. Ce lien permettra de : (1) valider les variables environnementales actuellement impliquées dans les IQH de potentiel de présence des bleuets sauvages et (2) d'attribuer un ordre d'importance à chacune des modalités de ces variables. Cet objectif principal sera accompli selon le traitement d'un objectif spécifique qui est l'identification des variables environnementales qui permettent de prédire la présence des plants de bleuets sauvages en milieu forestier.

## 2. Méthodologie

Cette section présente les méthodes, les traitements, ainsi que les techniques utilisées pour l'atteinte des objectifs du projet. La figure 2 ci-dessous présente le schéma méthodologique global qui peut se décomposer en cinq parties principales à savoir : (1) la collecte des données qui traite du protocole d'acquisition des mesures liées aux différentes variables environnementales, (2) le traitement des données qui traite (3) l'intégration dans l'interface cartographique qui traite de la visualisation des sites échantillonnés sur le portail cartographique web, (4) l'analyse statistique pour établir les liens entre les variables de l'environnement et la potentiel de présence des plants et (5) l'évaluation de l'IQH existant.

Afin d'offrir une description complète de cette méthodologie, ces étapes sont détaillées dans les prochaines sections.

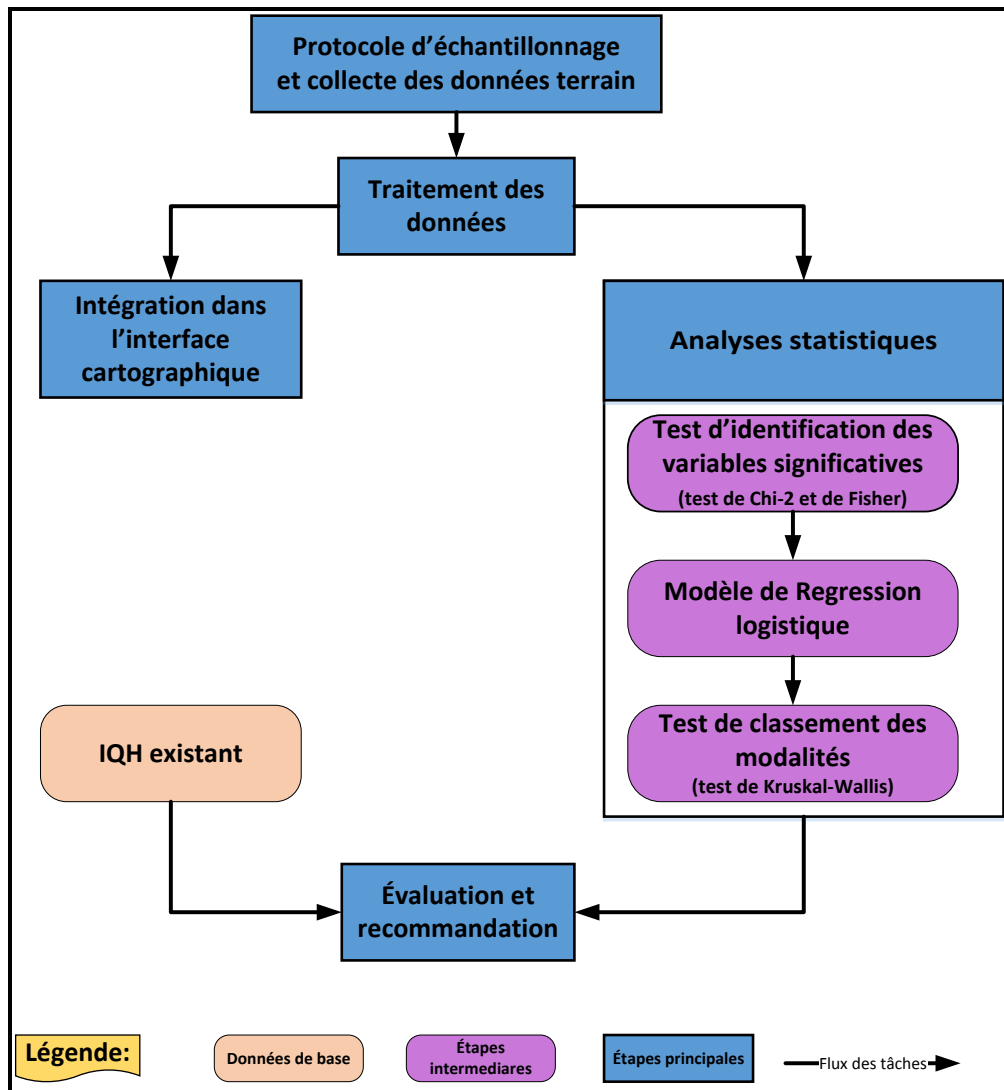


Figure 2: Schéma méthodologique pour l'évaluation d'un IQH des plants de bleuets.

#### 4.1 Zone d'étude

L'étude est localisée dans la forêt modèle du Lac-Saint-Jean (FMLSJ), située dans la région du Saguenay-Lac-Saint-Jean, qui est la région administrative 02 du Saguenay-Lac-Saint-Jean de la province de Québec. Cette région est située entre la latitude 48° et 49°N et les longitudes 71° et 73°O (figure 3). Elle s'étend sur plus de 95 892 km<sup>2</sup> dont 90 % sont constitués de terres publiques forestières gérées par le biais des contrats d'aménagement et d'approvisionnement forestier et 10 % de territoire municipalisé (MFFP, 2015). La FMLSJ fait partie de la grande forêt boréale canadienne (annexe 1) et est incluse dans les domaines bioclimatiques de la sapinière à bouleau jaune, de la sapinière à bouleau blanc et plus au

nord de la pessière à mousses (annexe 2). Bien que cette région soit majoritairement couverte de till épais, il est cependant intéressant à noter que le sud de cette zone est couvert par des dépôts sableux de type fluvioglaciale, propice à la culture du bleuets. Le réseau hydrographique est développé et composé de nombreux lacs, ainsi que de plusieurs petits cours d'eau. Les essences forestières principales sont le sapin baumier (*Abies balsamea*), l'épinette noire (*Picea mariana*), l'épinette blanche (*Picea glauca*), le pin gris (*Pinus banksiana*), le mélèze laricin (*Larix laricina*), le peuplier faux-tremble (*Populus tremuloides*) et le bouleau blanc (*Betula*) (FMLSJ, 2016).

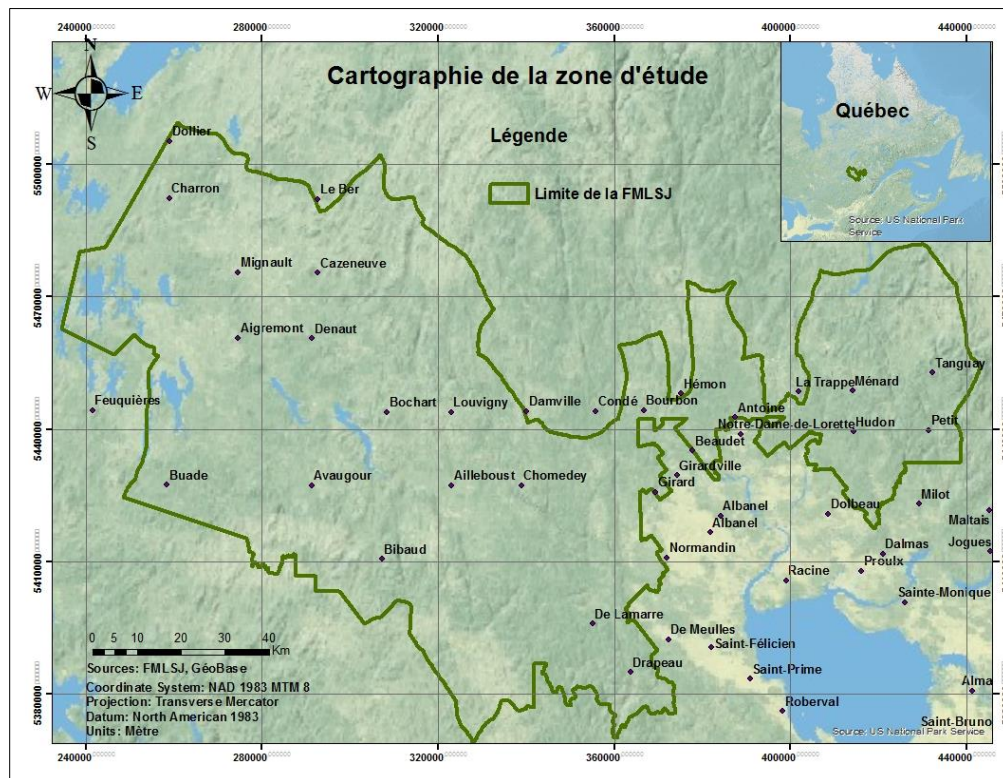


Figure 3: Carte montrant les limites de la zone d'étude.

## 4.2 Protocole d'échantillonnage et collecte des données terrains

La modélisation du potentiel de présence des bleuets sauvage en milieu forestier requiert l'acquisition des données qui caractérisent l'environnement, par exemple la pente, le type de peuplement forestier, le drainage. Ces données servent à valider la présence ou l'absence de plants de bleuets, et à identifier les variables qui influencent leurs présences. Ces données d'échantillonnages ont été collectées sur 140 placettes circulaires de 20 m de rayon identifiées dans un sous-ensemble d'environ 60 sites de 150 m de rayon. Ces placettes étaient distribuées spatialement et de façon aléatoire et stratifiées selon le taux de présence des plants de bleuets dans la zone d'étude. L'IQH du bleuets sauvage tel que

développé dans une autre étude portant sur la même zone sera utilisé pour répartir ces 140 placettes en quatre classes d'IQH (Dupras-Tessier, 2016). Soit 30 placettes pour chacune des classes de présence : « nulles », « faibles », « moyennes » et 50 placettes pour la classe « élevée ». Le formulaire des mesures terrain est procuré à l'annexe 3, de même que la codification des différentes variables. Les données relevées à chaque parcelle sont : (1) l'évaluation du pourcentage de recouvrement en plant de bleuet, (2) l'évaluation du pourcentage de recouvrement en fruit de bleuet, (3) l'estimation du potentiel de cueillette, (4) le pourcentage de recouvrement de la compétition (herbacé, arbustif, arborescent), (5) la composition du couvert, (6) type et âge de la perturbation, (7) le drainage, (8) le dépôts de surface, (9) la pente, (10) le type écologique, et (11) le groupe d'espèces indicatrices. Au total, les 11 variables issues des campagnes d'échantillonnages, sont des variables candidates à l'explication de la variabilité spatiale du potentiel de présence des plants de bleuets sauvages en milieu forestier.

#### 4.3 Traitement des données

Le domaine du relevé des données environnementales est très complexe et souvent influencé par le jugement personnel des membres de l'équipe d'échantillonnage. Puisque les résultats du traitement ne peuvent être ni meilleurs, ni fiables que les données relevées, il est donc primordiale de s'assurer de la qualité de ces données. Ceci est encore plus vrai dans le cas où ces données sont collectées par plusieurs équipes. Dans le cas de cette étude, les formulaires d'enregistrement sont d'abord nettoyés des valeurs aberrantes de codification des variables et des enregistrements contradictoires. Par la suite, ils sont compilés dans un tableau Excel, puis convertis au format csv pour faciliter : (1) l'intégration dans un système d'information géographique (SIG) et (2) l'analyse statistique par le logiciel R. Une fois intégrées dans un SIG, les données d'échantillonnages sont converties au format GeoJSON pour faciliter l'affichage des points sur le portail cartographique web. Le format GeoJSON est un format ouvert d'encodage d'ensemble de données géospatiales simples utilisant la norme JSON (JavaScript Object Notation). Il permet de décrire des données de type point, ligne, chaîne de caractères, polygone, ainsi que des ensembles et sous-ensembles de ces types de données et d'y ajouter des attributs d'information qui ne sont pas spatiale (Butler *et al*, 2008).

#### 4.4 L'intégration dans l'interface cartographique

Dans les projets antérieurs de l'Université de Sherbrooke, Dupras-Tessier, (2016) a mis au point un portail cartographique web pour la valorisation des PFNL. Il est donc question dans cette section d'intégrer les points d'échantillonnages dans ce portail, pour permettre aux utilisateurs d'accéder aux caractéristiques physiques des sites visités et d'avoir accès à une photographie de ces sites. Pour cela, les méthodes «new ol.layer.Vector» et «new ol.Overlay.Popup()» de la librairie «OpenLayers 3» sont utilisées respectivement pour : (1) intégrer le format «GeoJSON» des points d'échantillonnages dans le portail et (2) l'affichage des caractéristiques physiques de chaque site aux cliques de l'utilisateur. L'issue de ces traitements est disponible à l'adresse : [http://igeomedia.com/~odupras/data/data/portail\\_v1.0.1/www/](http://igeomedia.com/~odupras/data/data/portail_v1.0.1/www/) ou les utilisateurs peuvent interroger les points d'échantillonnages pour la visualisation des caractéristiques. La librairie OpenLayers 3 utilisée est une librairie libre, publiée sous licence BSD (*Berkeley Software Distribution License*), c'est à dire permet de réutiliser tout ou une partie de la librairie sans restriction, qu'il soit intégré dans un logiciel libre ou propriétaire. Cette librairie constituée de fonctions JavaScript bénéficie des dernières technologies web comme WebGL et permet la mise en place d'applications clientes Web cartographiques fluides dans n'importe quel navigateur moderne (Gratier *et al*, 2015).

#### 4.5 L'analyse statistique

Dans cette étude, les traitements statistiques des données sont réalisés à l'aide du logiciel R version 3.1.1, qui propose aux utilisateurs un langage et un environnement logiciel open source pour les calculs statistiques et graphiques. Trois types d'analyse statistique sont effectués pour répondre aux objectifs de cette étude à savoir : (1) une analyse de dépendance pour identifier les variables environnementales qui sont statistiquement liées à la présence ou à l'absence des plants de bleuets, (2) une régression logistique pour établir un lien entre le potentiel de présence des plants de bleuets et les variables environnementales significatives et (3) une analyse de variance appliquée à chacune des variables pertinentes, pour identifier les valeurs les plus propices pour les plants de bleuets

##### 1.5.1 Test d'identification des variables pertinentes

Parmi toutes les variables enregistrées lors des campagnes d'échantillonnages, le pourcentage de recouvrement est celle qui atteste du potentiel de présence des plants de bleuets dans un site. L'identification des variables pertinentes se résume donc à l'étude de la dépendance entre le pourcentage de recouvrement des plants et les autres variables. Bien

qu'il existe plusieurs tests permettant de mesurer le degré de significativité de la relation de dépendance entre deux caractères qualitatifs discrets, le test de khi-2 est le plus fréquemment utilisé et le mieux adapté à la plupart des situations où le tableau de contingence a un nombre de lignes ou de colonnes quelconques (Grasland, 2000). Le tableau de contingence est une méthode de représentation qui situe un effectif selon deux critères, dont un en ligne et l'autre en colonne (tableau 2).

L'idée générale du test du Chi-2 est de quantifier l'écart global entre les effectifs observés dans le tableau de contingence et les effectifs théoriques que l'on observerait dans le cas de l'indépendance entre les deux variables (variable en ligne et variable en colonne), puis de comparer la valeur de cette statistique à sa probabilité p (p-value) d'apparition dans le cas d'une série de tirages au sort effectués de façon aléatoire en tenant compte de la taille du tableau (Barnier, 2008). Pour un tableau de contingence de N cases, la statistique de Khi-2 noté habituellement  $\chi^2$  est défini par

$$\chi^2 = \sum_{i=1}^N \frac{(\text{Effectif observé} - \text{Effectif théorique})^2}{\text{Effectif théorique}}.$$

Cependant, Barnier, (2008) a également démontré que le découpage des classes, le regroupement des modalités, et les effectifs influencent grandement le résultat d'un test de khi-2. Autrement dit, la dépendance ou l'indépendance de deux variables est dictée par les effectifs et par la façon de classer les modalités. Pour remédier à cette limitation, le statisticien britannique Ronald Aylmer Fisher a mis au point un test alternatif au nom de test exact de Fisher. Ce test consiste à générer les milliers voire les millions de tableaux possibles ayant les mêmes totaux de lignes et de colonnes que le tableau de données observées, puis de calculer la proportion (p-value) exacte (d'où son nom) de ceux qui sont encore plus éloignés de l'hypothèse nulle que nos résultats (Poinot, 2005). Si cette probabilité est faible, on conclut que le hasard n'a pas pu faire ça tout seul, et qu'il y a donc un effet réel entre les variables. En pratique, ce test est une alternative au test de Khi-2, mais qui vise à tester la même hypothèse et s'interprète exactement de la même manière. C'est-à-dire, si la probabilité p obtenue est inférieure à un seuil défini à l'avance (en général  $p = 0,05$ ), on peut rejeter l'hypothèse d'indépendance des deux variables. La principale différence est qu'il s'agit d'un test exact, et non d'une approximation tirée d'une loi statistique. De plus, il ne nécessite aucune contrainte de taille d'échantillon (Barnier, 2008). Dans le cas de cette étude, le test de Khi-2 sera appliqué conjointement avec le test exact de Fisher pour consolider les conclusions issues du test de Khi-2. Donc ces deux tests nous

permettront de conclure quant à l'existence d'une relation statistiquement significative entre le recouvrement en plant de bleuets, et chacune des variables environnementales ciblées, sous l'hypothèse d'indépendance des variables. Par exemple, pour la variable classe de pente, l'hypothèse nulle ( $H_0$ ) se formule comme suit: «Le recouvrement en plant de bleuet d'un site est indépendant de la pente moyenne du site ».

Dans R, la fonction «table ( ) » permet d'obtenir le tableau de contingence suivant entre le recouvrement de plant et la classe de pente.

	Classe de pente				
Classe de recouvrement en plant	A	B	C	D	E
1	11	9	5	6	1
2	5	6	11	8	2
3	19	14	14	6	0
4	7	10	10	6	1
5	1	0	1	1	0

Tableau 2: Tableau de contingence de la classe de pente et du recouvrement en plant de bleuets sauvage

Sous hypothèse d'indépendance, les effectifs théoriques obtenus via la commande «chisq.test( tableau contingence)\$expected» sont :

	Classe de pente				
Classe de recouvrement de plant	A	B	C	D	E
1	8.9350649	8.1038961	8.5194805	5.610390	0.83116883
2	8.9350649	8.1038961	8.5194805	5.610390	0.83116883
3	14.7987013	13.4220779	14.1103896	9.292208	1.37662338
4	9.4935065	8.6103896	9.0519481	5.961039	0.88311688
5	0.8376623	0.7597403	0.7987013	0.525974	0.07792208

Tableau 3: Effectif théorique entre la classe de pente et du recouvrement en plant de bleuets sauvage en cas d'indépendance des variables

Bien que le test de Chi-2 soit relativement simple à mettre en œuvre, il ne peut cependant être utilisé rigoureusement que si les effectifs théoriques des cases du tableau de contingence sont supérieurs à 5 dans 80% des cases du tableau (Butler, 1985; Cochran, 1954; De Cock, 1998; Nelson *et al.*, 2002; Woods *et al.*, 1986). Dans le cas où cette



condition n'est pas respectée, on peut regrouper les modalités en fonction des connaissances que l'on a du phénomène, ceci dans le but de rendre le tableau éligible au test de Khi-2 (Nelson *et al.*, 2002; Poinso, 2004). Dans le cas présent, les classes de pente supérieure à 31% (classe D et E) sont regroupées en une classe D. De même pour les classes de recouvrement de plant 4 et 5 (classe 4). On obtient après ce regroupement le tableau de contingence suivant (tableau 4) qui respecte les conditions d'applications du test de Khi-2.

Classe de recouvrement en plant	Classe de pente			
	A	B	C	D
1	11	9	5	7
2	5	6	11	10
3	19	14	14	6
4	8	10	11	8

Tableau 4: Tableau de contingence de la classe de pente et du recouvrement en plant de bleuets sauvage après regroupement

En exécutant le test de Khi-2 via la fonction « `chisq.test()` » qui prend comme argument principal le tableau de contingence, on obtient :

Pearson's Chi-squared test

data: contin

X-squared = 11.051, df = 9, p-value = 0.2722

Donc la valeur de la statistique Khi-2 est 11,051 (X-squared) et la probabilité d'avoir un tel résultat pour 9 degré de liberté est de moins de 0,2722 (p-value). Puisque la p-value > 0,05, on accepte hypothèse nulle, c'est-à-dire que sur les sites d'échantillonnages, il n'existe pas de lien statistique entre le pourcentage de recouvrement des plants de bleuets et la pente du milieu. Par le test exact de Fisher, on obtient :

`fisher.test(contin,workspace=2e8)`

Fisher's Exact Test for Count Data

data: contin

p-value = 0.4173

Ce qui conduit à la même conclusion, car la p-value = 0,4 > 0,05.

En suivant la même méthodologie pour toutes les autres variables environnementales, on obtient les résultats du tableau ci-dessous (tableau 5).



<b>Variables environnementales</b>	<b>Test de Khi-2 (p-value)</b>	<b>Test exact de Fisher (p-value)</b>	<b>Significativité à 5%</b>	<b>Significativité à 10%</b>
altitude	0,3527	0,3299		
Pente	0,27	0,41		
Orientation	0,90	0,95		
Position	0,3732	0,3366		
Forme	0,06445	0,0699		x
versant	0,6475	0,7095		
Drainage	0,07	0,03	x	x
Type perturbation	0,03	0,03	x	x
âge perturbation	0,02	0,002	x	x
Type écologique	0,08242	0,01056	x	x
Type de couvert	0,1363	0,09771		x
Densité du couvert	4,733e <sup>-05</sup>	4,488e <sup>-05</sup>	x	x
Hauteur du couvert	1,206e <sup>-06</sup>	3,9e <sup>-06</sup>	x	x
Texture	0,4451	2,226e <sup>-05</sup>	x	x
Gr_esp_ind	0,02451	1,817e <sup>-05</sup>	x	x
esp_arb1	0,06312	0,03862	x	x
esp_herba1	0,001583	0,001395	x	x
eau500	0,8325	0,6881		

Tableau 5: Test de significativité des variables

Dans l'interprétation, une variable est considérée significative à 5% si la valeur obtenue pour sa p-value est inférieure à 0,05. Il en est de même pour une significativité à 10% (p-value < 0,1). On peut donc dire au vu de ces résultats que le type de perturbation, l'âge depuis la dernière perturbation, la densité du couvert forestier, la hauteur du couvert, le type écologique, la texture du sol, le groupe d'espèce indicatrice, l'espèce d'arbuste

prédominant, et l'espèce herbacée prédominante sont les variables environnementales qui présentent une dépendance statistique significative (à 5%) avec le pourcentage de recouvrement des plants de bleuets sauvages. Toutefois, dans l'optique de raffiner cette interprétation, plusieurs modèles de régressions logistiques seront établis entre le pourcentage de recouvrement des plants (variable réponse) et les variables significatives. Après une procédure de sélection de modèles, seules les variables restantes dans le meilleur modèle prédictif seront retenues.

### 1.5.2 Régression logistique

La régression logistique est une technique prédictive. Elle vise à construire un modèle permettant de prédire les valeurs prises par une variable cible qualitative (variable dépendante) à partir d'un ensemble de variables explicatives (variables indépendantes) quantitatives ou qualitatives (Rakotomalala, 2011). Dans le cas où la variable cible est binaire, on parle de régression logistique binaire ; si elle possède plus de 2 modalités, on parle de régression logistique polytomique. Contrairement à d'autres méthodes de régression, la régression logistique n'exige pas que les variables indépendantes soient distribuées normalement, linéaires ou qu'ils possèdent une variance égale entre chaque groupe (Tabachnick et Fidell, 2000). Toutefois, cette technique nécessite des échantillons de grande taille pour atteindre un bon niveau de stabilité. Un nombre minimal de 50 par variable est suggéré. De plus, les prédicteurs (variables indépendantes) choisis doivent être spécifiques à un groupe ou l'autre, puisqu'il s'agit d'une variable dichotomique. En d'autres mots, les catégories doivent être mutuellement exclusives et exhaustives. Cette technique s'avère également très sensible à la multicolinéarité entre les prédicteurs, celle-ci se vérifiant à l'aide d'une matrice de corrélation. Ainsi, il est nécessaire d'examiner les corrélations entre les prédicteurs avant de procéder à l'élaboration du modèle. Lorsque certains prédicteurs sont fortement corrélés entre eux, il est préférable d'en éliminer quelques-uns puisqu'il s'agit probablement de variables redondantes (Desjardins, 2005). Cette technique bien connue dans les sciences de la santé et en sciences humaines, où la variable à prédire est la présence ou l'absence d'une maladie, sera appliquée pour dire la présence ou l'absence des plants de bleuets sauvages. Pour cela, la variable dépendante qui est le pourcentage de recouvrement des plants de bleuets sera regroupée en deux modalités à savoir : «absent» ou «0» si le recouvrement est inférieur à 10 % et «présent» ou «1» si le recouvrement est supérieur est 10%. En d'autres termes, les sites qui présentent moins de

10 % de recouvrement des plants de bleuets seront considérés comme non propices à présence des plants de bleuets, tandis que ceux ayant plus de 10% de recouvrement en plant seront considérés comme propices aux plants de bleuets. Puisque les variables indépendantes sont de type qualitatif, une analyse d'association sera effectuée afin d'éliminer la multicolinéarité (liaison) entre les prédicteurs. Pour cela l'indice V de Cramer, publié en 1946 par Harald Cramér, qui est l'équivalent du coefficient de régression linéaire dans le cas des variables qualitatives sera utilisé. Il est défini par la racine carrée du  $\chi^2$  divisé par le  $\chi^2$  max. Ce  $\chi^2$  max théorique est égal à l'effectif multiplié par le plus petit côté du tableau (nombre de lignes ou de colonnes) moins 1.

$$V = \sqrt{\frac{\chi^2}{\text{Effectif total} \times (\text{nombre de ligne} - 1, \text{nombre de colonnes} - 1)}}$$
. Où  $\chi^2$  représente la valeur du Khi-2.

Il s'interprète de la même façon qu'un coefficient de corrélation linéaire, c'est-à-dire plus le V est élevé plus la dépendance entre les deux variables est forte, plus le V est faible, plus les variables se rapprochent de l'indépendance. Les cas extrêmes sont  $V = 0$ , dans le cas où les deux variables sont parfaitement indépendantes, et  $V = 1$ , dans le cas où les variables sont identiques (Barnier, 2008). Le V de Cramer varie dans l'intervalle  $[0, 1]$ , et le tableau 6 ci-dessous présente les intervalles d'interprétation.

Valeur	Force du lien statistique
0	Absence de relation
Entre 0,05 et 0,10	Très faible
Entre 0,10 et 0,20	Faible
Entre 0,20 et 0,40	Modérée
Entre 0,40 et 0,80	Forte
Entre 0.80 et 1	Louche (Colinéarité)

Tableau 6: Interprétation de l'indice V de Cramer

En appliquant cet indice à l'ensemble des variables indépendantes (Annexe 4), il ressort une dépendance forte entre :

- le drainage, la texture, le groupe d'espèce indicatrice
- la texture, le groupe d'espèce indicatrice, le type écologique, le type de couvert forestier, les espèces arbustes dominants, les espèces herbacées dominants

Ainsi, le premier modèle de régression logistique (figure 4) est formulé à partir des six variables indépendantes non liées qui sont le drainage, le type écologique, le type de

perturbation, l'âge depuis la dernière perturbation, la densité du couvert forestier et la hauteur.

Coefficients:				
	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-3.4972	2.0417	-1.713	0.08674 .
drainageImp_Mauvais	-4.6910	1.6379	-2.864	0.00418 **
drainageModere	-2.0229	1.0626	-1.904	0.05694 .
Type_pertuFeu	0.9744	1.0395	0.937	0.34853
Age_pertu2 ans ou 13 à 20ans	4.7906	1.7872	2.680	0.00735 **
Age_pertu3 à 8 ans	23.6211	2176.7202	0.011	0.99134
Age_pertu9 à 12 ans	4.0983	1.7201	2.383	0.01719 *
type_EcoloRE	3.7771	1.5793	2.392	0.01677 *
type_EcoloRS2	2.1476	1.8737	1.146	0.25171
den_couvNormale_Forte	-2.2972	1.1641	-1.973	0.04847 *
den_couvTres_Faible	-0.9795	1.1949	-0.820	0.41236
haut_couv2m et moins	-1.5678	1.3400	-1.170	0.24201
haut_couv4 à 17m	0.2834	1.1414	0.248	0.80391
---				
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
(Dispersion parameter for binomial family taken to be 1)				
Null deviance: 76.767 on 86 degrees of freedom				
Residual deviance: 41.540 on 74 degrees of freedom				
AIC: 67.54				

Figure 4: Résultat du modèle 1 sous R

Au vu de ce résultat, la hauteur du couvert forestier et le type de perturbation ne semblent pas significatifs au seuil de 5%, ni au seuil de 10 %. Puisque la hauteur du couvert forestier semble plus éloignée de la significativité (plus grande p-value) que le type de perturbation, elle est exclue de la liste des variables significative pour donner lieu à un second modèle donc les résultats sont présentés dans la figure 5. Dans ce modèle, les variables les plus significatives (seuil de 5%) sont le drainage, l'âge depuis la dernière perturbation, et le type écologique, suivis par la densité du couvert forestier (seuil de 10%) et le type de perturbation. Dans l'objectif de conserver les meilleurs prédicteurs de l'ensemble des variables échantillonnées, un troisième modèle de régression logistique automatisée est généré. C'est une méthode de régression itérative qui inclut d'abord dans le modèle la variable qui contribue le plus à réduire la variabilité résiduelle du modèle à chaque étape. Alternativement, toutes les variables sont entrées dans le modèle et les variables sont progressivement exclues, en fonction de celles qui contribuent le moins au modèle (Graham, 2003). En appliquant cette méthode, on obtient le troisième modèle qui génère les résultats présentés dans la figure 6.

```

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    -3.2442     1.8703   -1.735  0.08281 .
drainageImp_Mauvais -4.2829     1.4563   -2.941  0.00327 **
drainageModere   -1.9453     0.9831   -1.979  0.04786 *
Type_pertuFeu     0.9253     0.9788    0.945  0.34448
Age_pertu2 ans ou 13 à 20ans  4.4081     1.6654    2.647  0.00812 **
Age_pertu3 à 8 ans  22.5593    2207.1412    0.010  0.99184
Age_pertu9 à 12 ans  3.5139     1.4562    2.413  0.01582 *
type_EcoloRE      3.4481     1.4574    2.366  0.01799 *
type_EcoloRS2     2.4864     1.7496    1.421  0.15529
den_couvNormale_Forte -1.7646     1.0109   -1.746  0.08089 .
den_couvTres Faible -0.8895     1.1862   -0.750  0.45333
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 76.767  on 86  degrees of freedom
Residual deviance: 43.371  on 76  degrees of freedom
AIC: 65.371

```

Figure 5: Résultat du modèle 2 sous R

```

Call:
glm(formula = reponse ~ Age_pertu + drainage + type_Ecolo, family = binomial,
    data = tableau1)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.60743   0.00003   0.27832   0.40528   1.52227

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    -3.8426     1.7141   -2.242  0.02498 *
Age_pertu2 ans ou 13 à 20ans  4.1473     1.5867    2.614  0.00896 **
Age_pertu3 à 8 ans  22.7936    2179.0060    0.010  0.99165
Age_pertu9 à 12 ans  4.0136     1.4566    2.755  0.00586 **
drainageImp_Mauvais  -3.6371     1.2711   -2.862  0.00422 **
drainageModere     -2.0008     0.9623   -2.079  0.03761 *
type_EcoloRE       3.0607     1.2619    2.425  0.01529 *
type_EcoloRS2      2.2871     1.5641    1.462  0.14367
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 76.767  on 86  degrees of freedom
Residual deviance: 47.927  on 79  degrees of freedom
AIC: 63.927

```

Figure 6: Résultat du modèle 3 sous R

Selon ce modèle, les seules variables pertinentes pour la présence des plants de bleuets sauvages sont le drainage, le type écologique et l'âge depuis la dernière perturbation.

### 1.5.3 Évaluation de modèles

Dans l'optique de sélectionner le modèle qui a le plus grand pouvoir prédictif, une étape d'évaluation de modèle est faite sur la base de 38 sites d'échantillonnages (12 sites sans les plants de bleuets et 26 sites avec des plants de bleuets) qui n'ont pas été utilisés pour la



construction des modèles. Pour le premier modèle basé sur drainage, le type écologique, le type de perturbation, l'âge depuis la dernière perturbation, la densité du couvert forestier et la hauteur, on obtient 8 prédictions incorrectes sur les 38 (tableau 7). Parmi les 12 sites où les plants sont absents, le modèle trouve à tort la présence des plants dans 4 sites et à raison leur absence dans 8 sites. Par contre, dans les 26 sites où les plants sont présents, le modèle trouve à tort l'absence des plants dans 4 sites et à raison leur présence dans 22 sites.

	Valeurs réelles		
Prediction	Absent	Present	Total
Absent	8	4	12
Present	4	22	26
Total	12	26	

Tableau 7: Pouvoir prédictif du modèle 1 pour 38 sites d'échantillonnages

Pour le deuxième modèle basé sur le drainage, le type écologique, le type de perturbation, l'âge depuis la dernière perturbation et la densité du couvert forestier, on obtient 3 prédictions incorrectes sur les 38 (tableau 8). Parmi les 12 sites où les plants sont absents, le modèle trouve à tort la présence des plants dans 2 sites et à raison leur absence dans 10 sites. Par contre, dans les 26 sites où les plants sont présents, le modèle trouve à tort l'absence des plants dans 1 site et à raison leur présence dans 25 sites.

	Valeurs réelles		
Prediction	Absent	Present	Total
Absent	10	1	11
Present	2	25	27
Total	12	26	

Tableau 8: Pouvoir prédictif du modèle 2 pour 38 sites d'échantillonnages

Pour le troisième modèle basé sur le drainage, le type écologique et l'âge depuis la dernière perturbation, on obtient 5 prédictions incorrectes sur les 38 (tableau 9). Parmi les 12 sites où les plants sont absents, le modèle trouve à tort la présence des plants dans 3 sites et à raison leur absence dans 9 sites. Par contre, dans les 26 sites où les plants sont présents, le modèle trouve à tort l'absence des plants dans 2 sites et à raison leur présence dans 24 sites.

	Valeurs réelles		
Prediction	Absent	Present	Total
Absent	9	2	11
Present	3	24	27
Total	12	26	

Tableau 9: Pouvoir prédictif du modèle 3 pour 38 sites d'échantillonnages

Dans la pratique, il est fréquent d'être confronté à des données provenant de campagne terrain dans lesquelles les variables explicatives doivent souvent être relevées pour leur rôle connu ou supposé dans le phénomène étudié, sans qu'il soit possible de les contrôler ou de s'assurer de leur effet réel. On se retrouve alors confronté à de nombreux modèles possibles, correspondant aux différentes combinaisons de ces variables explicatives. Plusieurs méthodes statistiques existent pour la sélection des modèles, le plus connu et le plus utilisé est le critère d'information d'Akaike (AIC). Ce critère repose sur un compromis entre la qualité de l'ajustement et la complexité du modèle, en pénalisant les modèles qui ont un grand nombre de paramètres (Akaike, 1987). Or en général, augmenter le nombre de paramètre améliore nécessairement la qualité de l'ajustement, raison pour laquelle le critère de sélection dans le cas de ce projet sera basé uniquement sur le pouvoir prédictif de chaque modèle (qualité d'ajustement). Selon ce critère, le modèle 2 basé sur drainage, le type écologique, le type de perturbation, l'âge depuis la dernière perturbation et la densité du couvert forestier sera retenu comme modèle de prédiction du potentiel de présence des plants de bleuets sauvage en milieu forestier (tableau 10).

Critère	AIC	Prediction incorrectes
Modèle 1	67,54	8 sur 38
Modèle 2	65,37	3 sur 38
Modèle 3	63,92	5 sur 38

Tableau 10: Comparaison des modèles selon le critère AIC et les prédictions incorrectes

#### 1.5.4 Classement des modalités.

Les analyses effectuées dans les sections précédentes ont permis d'identifier les variables pertinentes pour la prédiction du potentiel de présence des plants de bleuets sauvages. Dans cette section, il sera question de mettre en relief pour chacune de ces variables, les valeurs qui favorisent ou limitent la présence des plants de bleuets. Par exemple, dans le cas du type de perturbation, la question revient à savoir si une coupe forestière est meilleure qu'un feu pour les plants de bleuets sauvages. Pour cela, l'interprétation des coefficients du modèle (rapport de cotes) de régression logistique permettent de répondre à cette question. Les rapports de cotes s'interprète de façon similaire au risque relatif. Un rapport de cote de 1 signifie l'absence d'effet, un rapport de cote largement supérieur à 1 correspond à une augmentation du phénomène étudié tandis qu'un rapport de cote largement inférieur à 1 correspond à une diminution du phénomène étudié. Ici, les résultats sont présentés sous forme graphique pour faciliter l'interprétation (figure 7).

Pour la variable type de perturbation, la modalité de référence est la coupe forestière, puisque la valeur du rapport de cote associé à la modalité feu (2,5) est supérieure à 1, on peut donc dire qu'un site perturbé par un feu est meilleur pour les plants de bleuet qu'un site perturbé par une coupe forestière. Ce résultat est également confirmé par la figure 8 qui représente le recouvrement moyen des plants en fonction du type de perturbation. La moyenne de recouvrement dans les sites perturbés par une coupe est de 46 % contre 48% pour les sites perturbés par un feu. Un test de Kruskal-Wallis montre que cette différence est significative au seuil de 5% avec une p-value de 0,007.

Pour la variable type écologique, la modalité de référence est la sapinière à bouleau blanc (MS2), puisque les valeurs des rapports de cotes associées aux deux autres modalités (sapinière à épinette noire et pessière noire à mousse ou à éricacées) sont supérieures à 1, on peut donc dire que les types écologiques sapinière à épinette noire et pessière noire à mousse ou à éricacées sont meilleurs pour les plants de bleuets que la sapinière à bouleau blanc. Ce résultat est également confirmé par la figure 9 qui représente le recouvrement moyen des plants en fonction du type écologique. En suivant la même logique pour la variable drainage, on peut dire que les bons drainages et les drainages rapides sont meilleurs pour les plants de bleuets que les autres types (imparfait, modéré, mauvais) de drainage (figure 10). Pour la densité du couvert forestier, les sites ayant une faible densité de couvert sont plus propices aux plants que les sites avec une densité forte et normale (figure 11). Pour la variable âge de la perturbation, les sites perturbés depuis plus de 3ans,



et moins de 12 ans sont meilleurs pour les plants de bleuet que les autres sites perturbés (figure 12).

### 3. Évaluation et recommandation

Les IQH du bleuet sauvage utilisés actuellement sont basés sur trois variables environnementales donc le drainage, les dépôts de surfaces, le type de couvert forestier et deux variables de perturbation donc le type de perturbation et l'âge de la perturbation. Par ailleurs, il existe un lien très significatif entre le type de couvert forestier et le type écologique (annexe 4). Donc on peut d'un point de vue statistique remplacer l'un par l'autre et vice versa. Vu ainsi, la densité du couvert forestier se trouve être la seule variable significative pour les plants de bleuet sauvage, mais absent dans les IQH existants. De plus, sachant que le bleuet sauvage est une espèce de transition et une plante colonisatrice qui poussent en abondance après une perturbation, et qui peu à peu, de quelques années à une dizaine selon la fertilité des sites, fait place à d'autres espèces de transition qui l'ombragent (framboisiers, bouleaux) et qui le font reprendre sa place restreinte comme espèce de sous-bois, on devrait intégrer des variables pertinentes comme la densité du couvert forestier dans les IQH pour rendre compte de cette caractéristique environnementale.

Par ailleurs, en évaluant l'intensité de la dépendance entre le recouvrement des plants et les variables pertinentes, on s'aperçoit que la densité du couvert est la variable qui a le plus fort lien avec le recouvrement des plants, suivie du type écologique et du drainage (annexe 4). Donc, on devrait s'attendre à avoir une pondération qui obéit à cette hiérarchie.

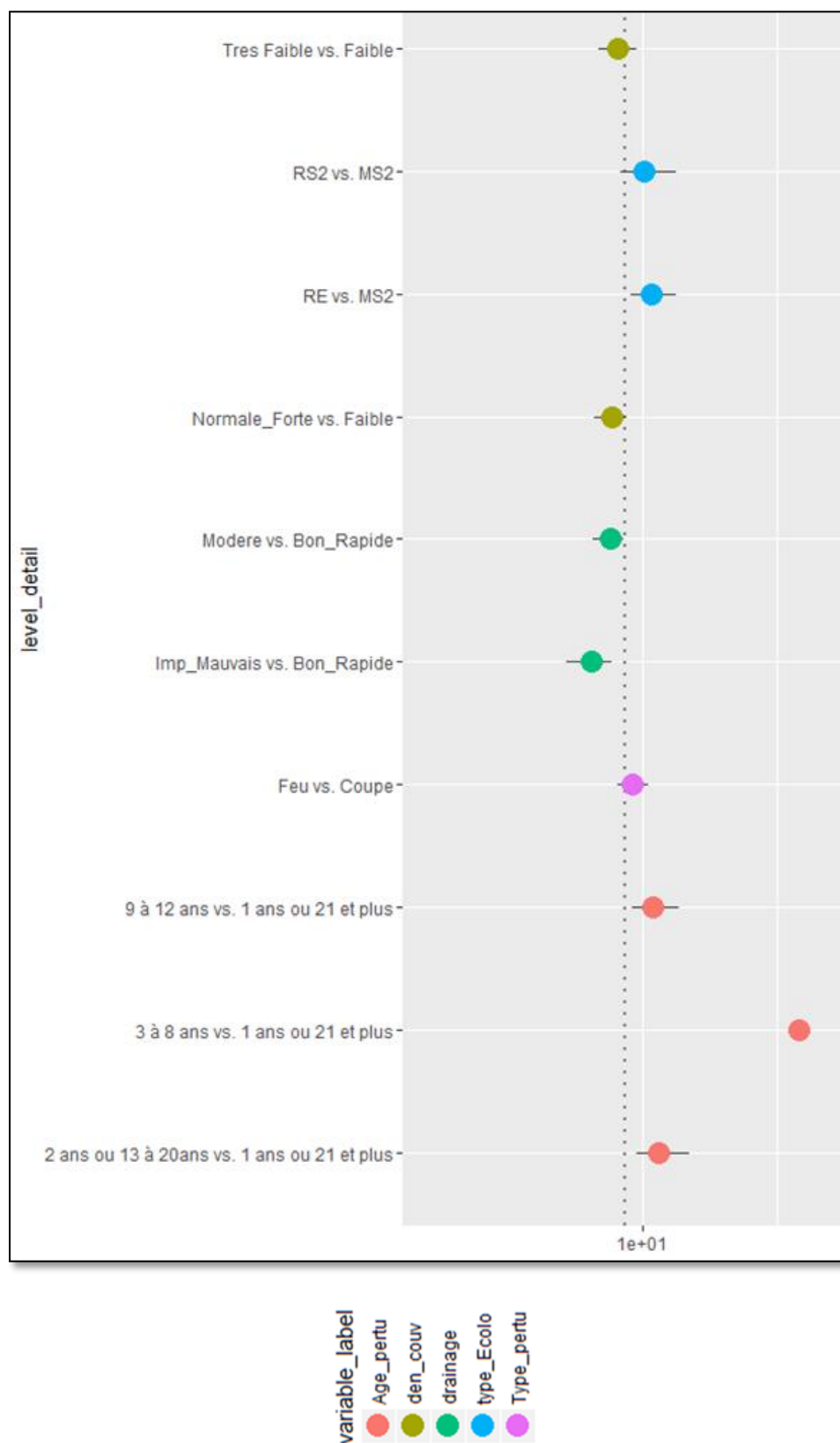


Figure 7: Représentation graphique des rapports de cotes du modèle 2

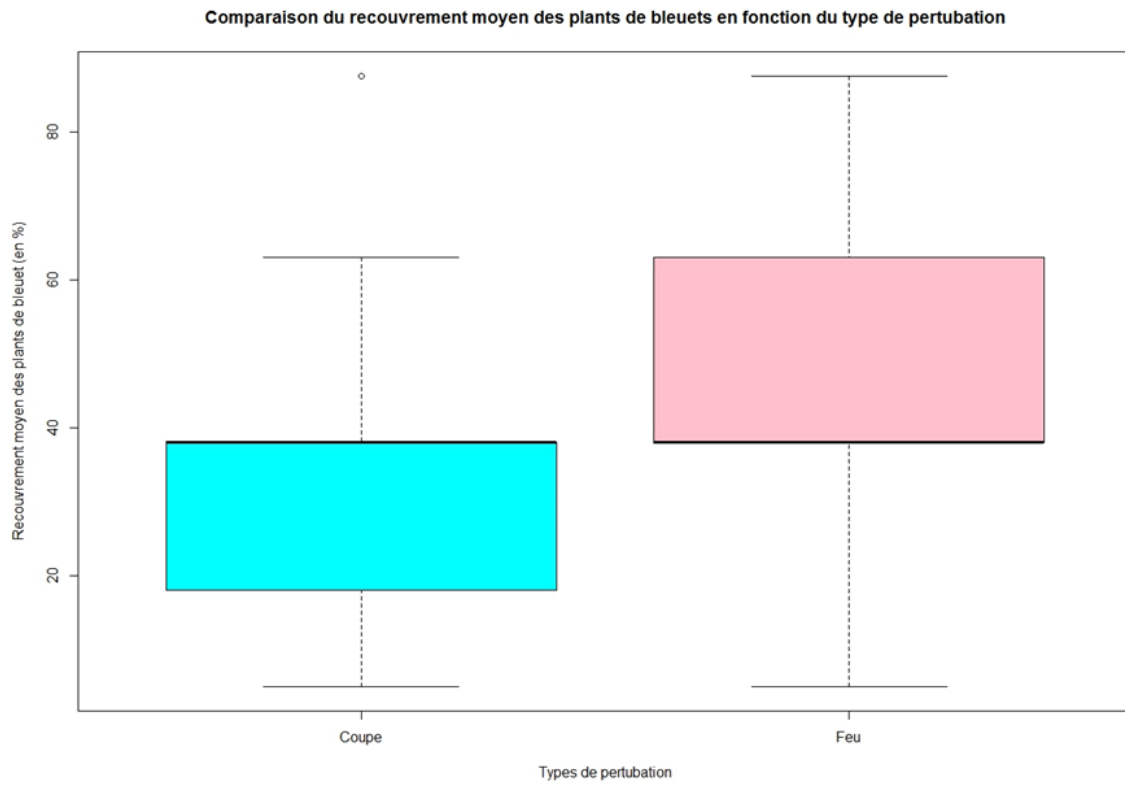


Figure 8: Variabilité du pourcentage de recouvrement moyen des plants en fonction du type de perturbation

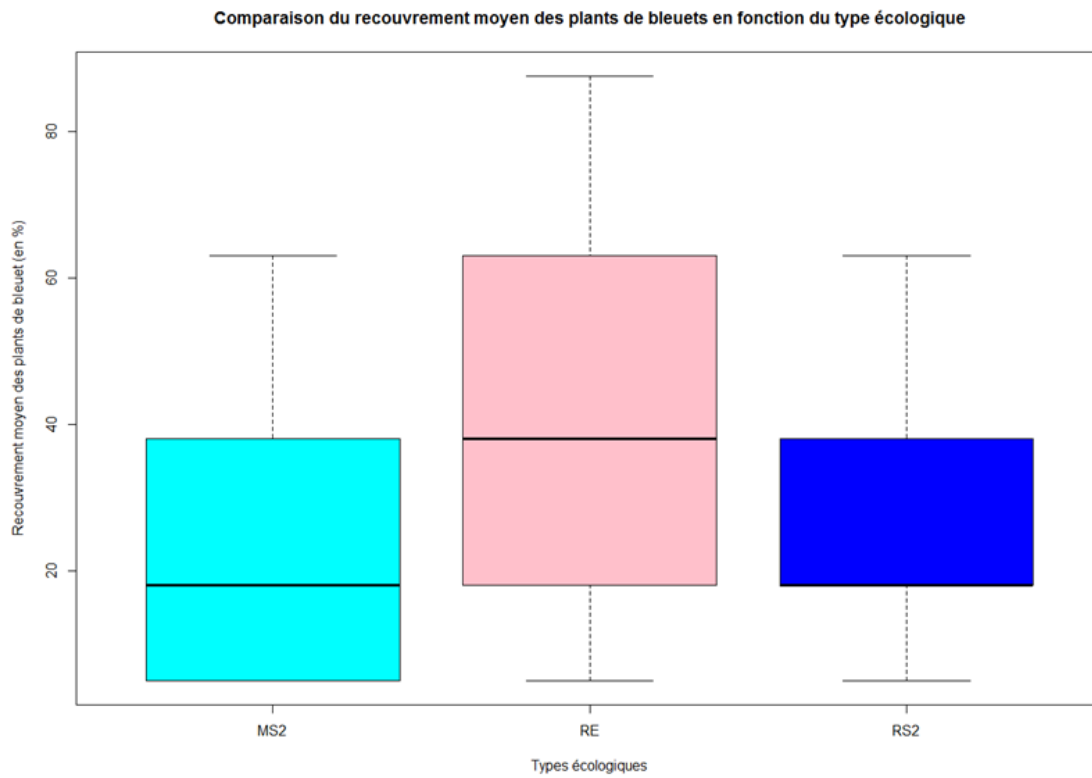


Figure 9: Variabilité du pourcentage de recouvrement moyen des plants en fonction du type écologique

Comparaison du recouvrement moyen des plants de bleuets en fonction du drainage

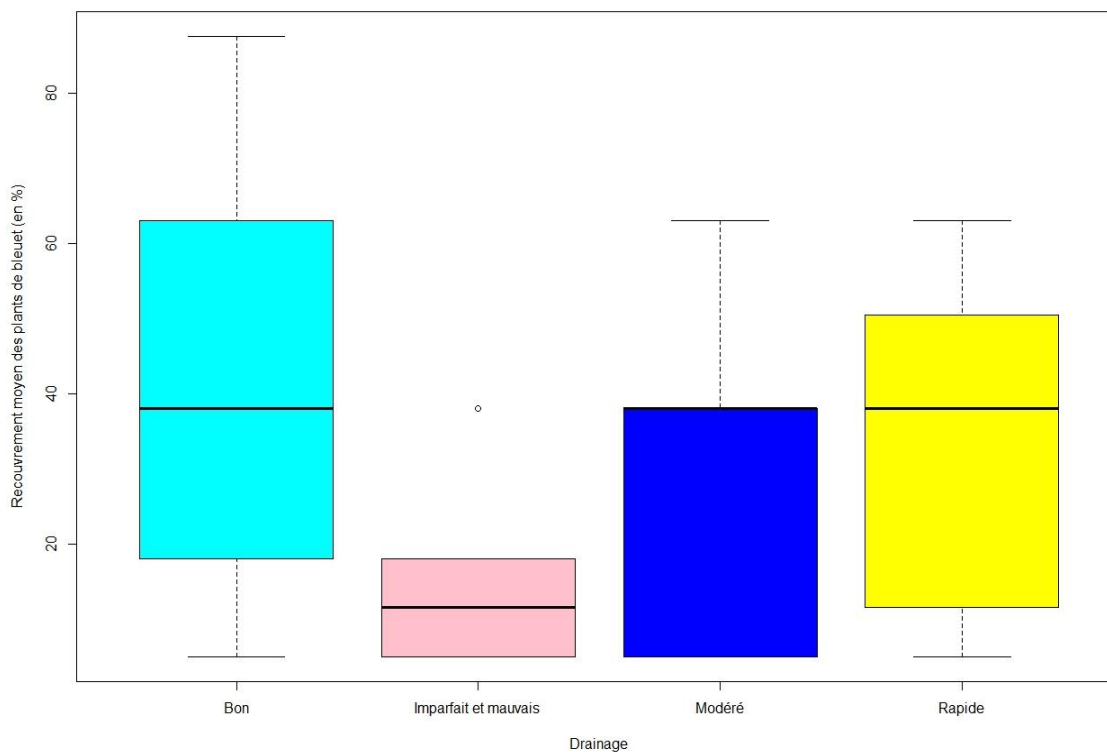


Figure 10: Variabilité du pourcentage de recouvrement moyen des plants en fonction du drainage

Comparaison du recouvrement moyen des plants de bleuets en fonction de la densité du couv

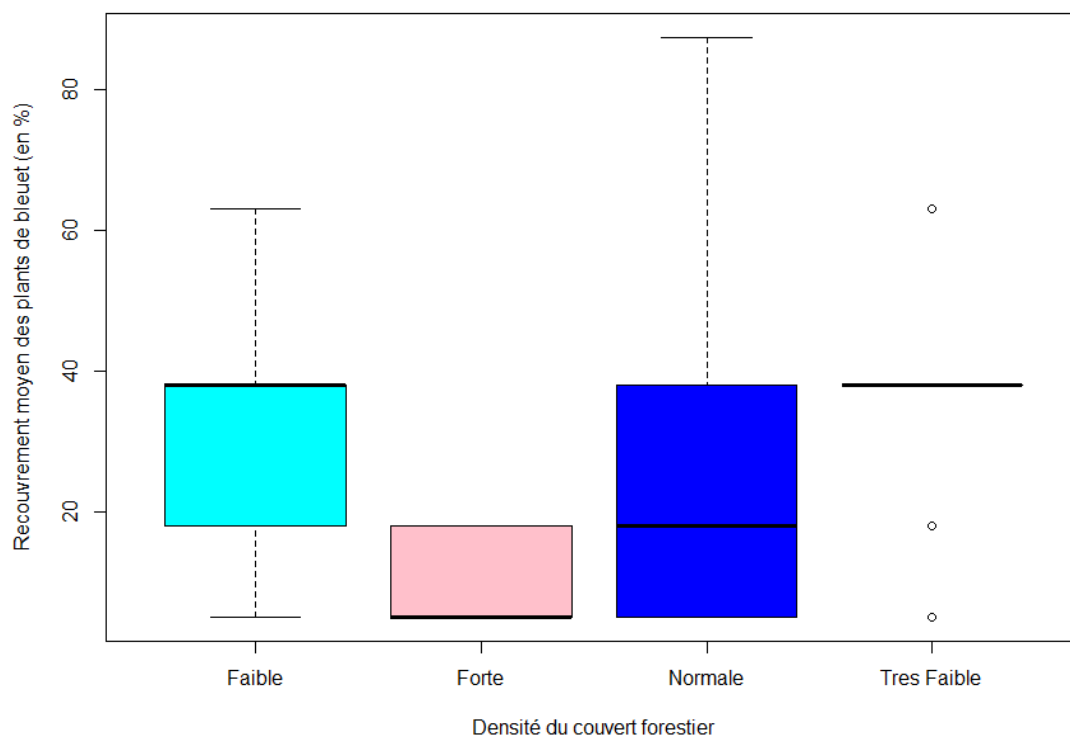


Figure 11: Variabilité du pourcentage de recouvrement moyen des plants en fonction de la densité du couvert forestier

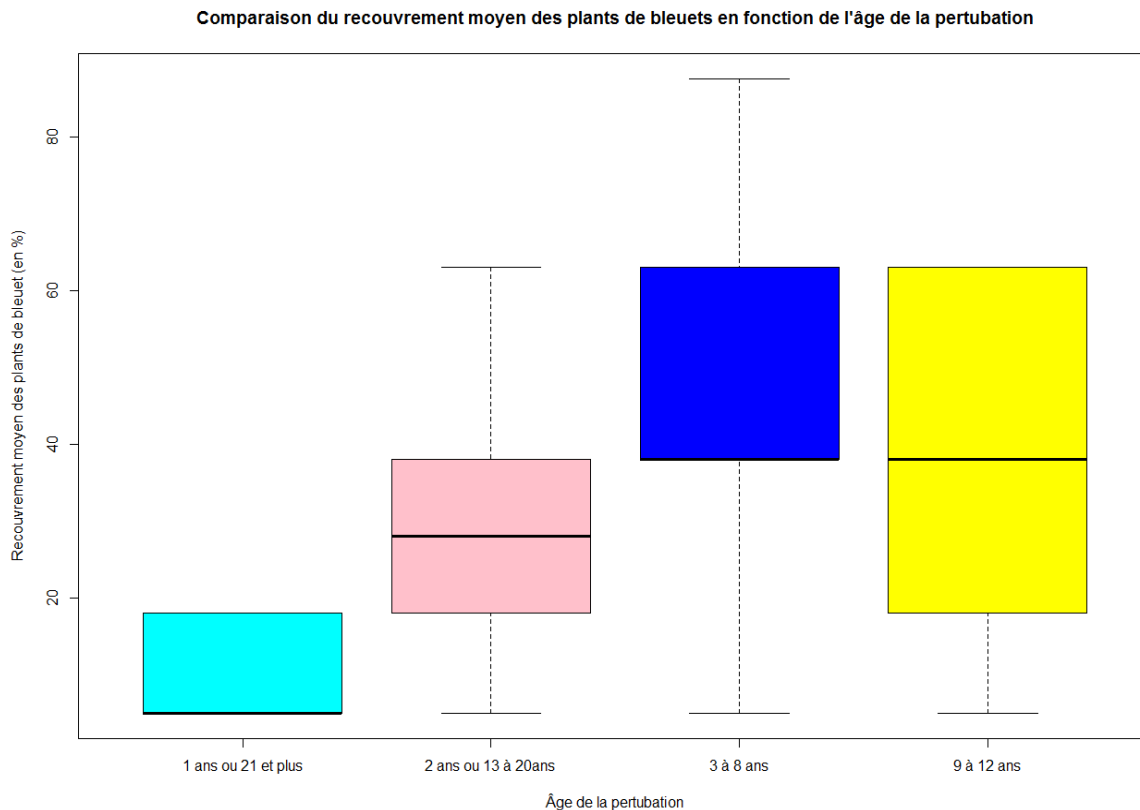


Figure 12: Variabilité du pourcentage de recouvrement moyen des plants en fonction de l'âge de la perturbation

En ce qui concerne les valeurs attribuées aux modalités des différentes variables pertinentes, tout semble conforme aux conclusions statistiques à l'exception du type écologique et du drainage. D'après la figure 9, le type écologique pessière noire à mousses ou à éricacées (RE1, RE2, RE3) est plus propice pour les plants de bleuets, suivis de la sapinière à épinette noire (RS2\*) et de la sapinière à bouleau blanc (MS2\*). Donc les valeurs associées à ces types écologiques devraient correspondre à cet ordre. Dans le cas du drainage, le bon drainage est la modalité qui favorise le plus le développement des plants de bleuets, suivis respectivement des drainages rapides, modéré, imparfait/mauvais (figure 10). Donc la pondération de cette variable peut se faire en quatre groupes, soit la valeur maximale 3 pour les bons drainages, 2 pour les drainages rapide et modéré car la variabilité est très faible, 1 pour les drainages imparfaits et mauvais et 0 pour les autres valeurs du drainage. À noter qu'aucun point d'échantillonnage ne présentait un drainage excessif, donc impossible d'évaluer statistiquement.

## References

- Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, 52(3), 317-332.
- Atlegrim, O. et Sjöberg, K. (1996). Response of bilberry (*Vaccinium myrtillus*) to clear-cutting and single-tree selection harvests in uneven-aged boreal *Picea abies* forests. *Forest Ecology and Management*, vol. 86, n°1-3, p. 39-50
- Barnier, J. (2008). Tout ce que vous n'avez jamais voulu savoir sur le  $\chi^2$  sans jamais avoir eu envie de le demander. Groupe de Recherche sur la Socialisation CNRS-UMR, 5040, 15.
- Butler, C. (1985) Statistics in linguistics.
- Butler, H., Daly, M., Doyle, A., Gillies, S., Schaub, T., & Schmidt, C. (2008). The GeoJSON format spécification. Rapport technique, 67.
- Cochran, W. G. (1954) Some Methods for Strengthening the Common  $\chi^2$  Tests. *Biometrics*, vol. 10, n°4, p. 417-451.
- Chiasson, G. et J. Agrall. (1996). Croissance et développement du bleuet sauvage. Ministère de l'Agriculture et de l'Aménagement rural du Nouveau-Brunswick. Feuille d'information A.2.0. [En ligne]. <http://www2.gnb.ca/content/dam/gnb/Departments/10/pdf/Agriculture/WildBlueberries-BleuetsSauvages/a20F.pdf> (Page consultée le 15 septembre 2010).
- Coopérative de solidarité forestière de la rivière aux saumons (2011) Les rapports de recherche. Évaluation comparative de la productivité du bleuet sauvage (*Vaccinium sp.*) en milieu naturel selon le type de récolte et d'aménagement sur différents types écologiques. La Doré, 71p.
- Coudun, C. et Gégout, J.-C. (2007). Quantitative prediction of the distribution and abundance of *Vaccinium myrtillus* with climatic and edaphic factors. *Journal of Vegetation Science*, vol. 18, n°4, p. 517-524.
- Crête, M. (2003). Les modèles d'indice de qualité de l'habitat : des outils utiles pour la gestion de l'habitat de la faune forestière? Société de la faune et des parcs du Québec, Direction du développement de la faune. 25 p.
- De Cock, S. (1998) A Recurrent Word Combination Approach to the Study of Formulae in the Speech of Native and Non-Native Speakers of English. *International Journal of Corpus Linguistics*, vol. 3, n°1, p. 59-80.
- Desjardins, J. (2005). L'analyse de régression logistique. Tutorial in quantitative methods for psychology, 1(1), 35-41.
- Duchesne, L. C. et Wetzel, S. (2004) Effect of Fire Intensity and Depth of Burn on Lowbush Blueberry, *Vaccinium angustifolium*, and Velvet Leaf Blueberry, *Vaccinium*

myrtilloïdes, Production in Eastern Ontario. The Canadian Field-Naturalist, vol. 118, n°2, p. 195-200.

Drummond, F., Smagula, J., Annis, S., & Yarborough, D. (2008). Organic wild blueberry production. Maine Agricultural and Forest Experiment Station Bulletin, 852.

Dupras-Tessier, O. (2016). Conception d'un portail participatif d'information géographique Web pour la mise en valeur des produits forestiers non ligneux d'après leurs potentiels de présence.

Farooque, A. A. 1. Z., Qamar U. 1. Schumann, Arnold W. 2. Madani, Ali 1; Percival, David C. (2012). Response of Wild Blueberry Yield to Spatial Variability of Soil. Soil Science, vol. 177(1), p. 56-68.

Forêt modèle Lac-Saint-Jean. (2016). Description du territoire de FMLSJ in À propos de FMLSJ. [http://www.foretmodeledulacsaintjean.ca/fr/page/description\\_du\\_territoire\\_de\\_fmlsj/](http://www.foretmodeledulacsaintjean.ca/fr/page/description_du_territoire_de_fmlsj/). Consulté le 12 août 2016.

Gratier, T., Spencer, P., & Hazzard, E. (2015). OpenLayers 3: Beginner's Guide. Packt Publishing Ltd.

Graham, M. H. (2003). Confronting multicollinearity in ecological multiple regression. Ecology, 84(11), 2809-2815.

Haijun, X., Guang, S., Lin, Z., & Yue, Z. (2013). Habitat Classification and Ecological Factors Analysis of *Vaccinium uliginosum* [J]. Journal of Northeast Forestry University, 1, 015.

Hester, A. J., Miles, J. et Gimingham, C. H. (1991) Succession from Heather Moorland to Birch Woodland. II. Growth and Competition Between *Vaccinium Myrtillus*, *Deschampsia Flexuosa* and *Agrostis Capillaris*. Journal of Ecology, vol. 79, n°2, p. 317-327.

Holloway, P. S. (2006) Managing Wild Bog Blueberry, Lingonberry, Cloudberry and Crowberry Stands in Alaska (Report). Natural Resources Conservation Service.

Jean, Duval. (1993) Les rapports de recherche. La régénération des bleuetières. Ecological Agriculture Projects, McGill University (Macdonald Campus), Ste-Anne-de-Bellevue, QC, H9X 3V9 Canada.

Julie, Labbé. (2012) Les rapports de recherche. Méthode performante d'évaluation de la qualité de l'habitat de l'original dans les zecs du Québec. Fondation de la faune du Québec.

Kautz, E. . (1987) Prescribed fire in blueberry management. Fire management notes United States Dept. of Agriculture, Forest Service.

Ketata, M. (2012). Effet des prétraitements cryogéniques sur la déshydratation osmotique des bleuets (Doctoral dissertation, Université Laval).

Ministère des forêts, faune et parcs du Québec. (2015). Gros plan sur Saguenay-Lac-Saint-Jean in Portrait territorial. <http://www.mffp.gouv.qc.ca/saguenay-lac-saint-jean/region/portrait.jsp>.

Neilson, W. T. A. et Wood, G. W. (1985) The blueberry maggot: distribution, economic importance, and management practices. *Acta Horticulturae*, n°165, p. 171-176.

Nelson, G., Wallis, S. et Aarts, B. (2002) Exploring Natural Language: Working with the British Component of the International Corpus of English. John Benjamins Publishing.

Olivier Paré. (2013) Les rapports de recherche. Perspectives sectorielles 2013-2015 Saguenay-lac-Saint-Jean. Service Canada Région du Saguenay-Lac-Saint-Jean. (26p)

Poinsot, D. (2004). Statistiques pour statophobes.(p. 175).

Poinsot, D. (2005). R pour les statophobes.

Rakotomalala, R. (2011). Pratique de la régression logistique. Régression Logistique Binaire et Polytomique, Université Lumière Lyon, 2

Roloff, G. J., & Kernohan, B. J. (1999). Evaluating reliability of habitat suitability index models. *Wildlife Society Bulletin*, 973-985.

Schamberger, M., & O'Neil, J. (1986). Concepts and constraints of habitat model testing.

Syndicat des producteurs de bleuets sauvages du Québec. (2014) Les rapports de recherche. Le secteur des bleuets sauvages : des retombées importantes et récurrentes pour le Québec et ses régions. 10p

Tabachnick, B. G., & Fidell, L. S. (2000). *Computer-assisted research design and analysis*. Allyn & Bacon, Inc..

Vickery, B. Z., A.I, J., W.g, S. et A.p, W. (2005) Influence of fire and other anthropogenic practices on grassland and shrubland birds in New England.

Waldrop, T. A., White, D. L. et Jones, S. M. (1992) Fire regimes for pine-grassland communities in the southeastern United States. *Forest Ecology and Management*, vol. 47, n°1, p. 195-210.

Woods, A., Fletcher, P. et Hughes, A. (1986) Statistics in Language Studies. Cambridge University Press.



## 4. Annexes

### 4.1 Annexes 1: Cartographie des sous zone de végétation

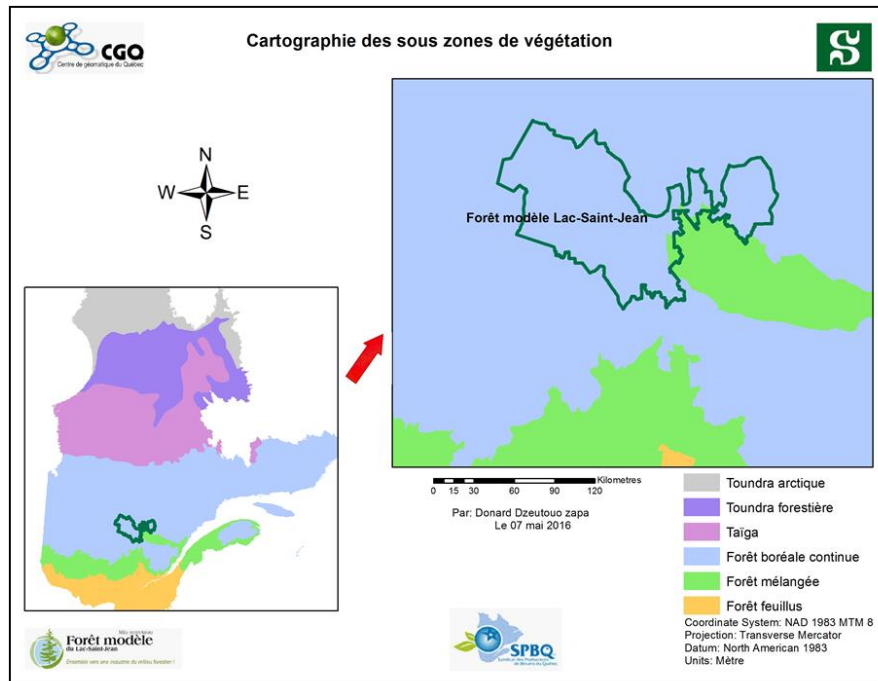


Figure A1: Représentation de la Forêt modèle du lac Saint-Jean dans les sous-zones de végétation forestière au Québec

### 4.2 Annexe 2 : Cartographie des domaines bioclimatiques

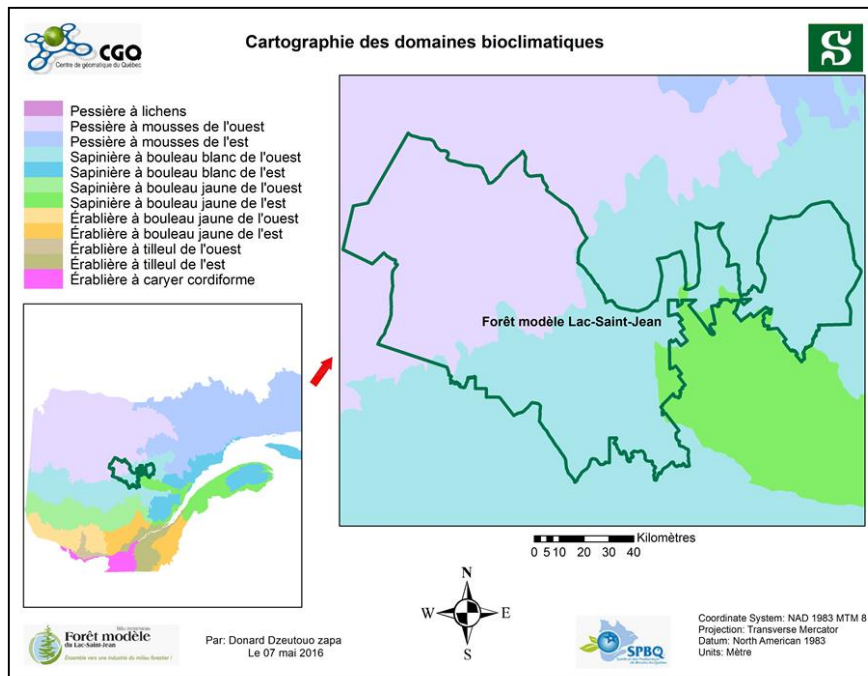


Figure A2: Représentation de la Forêt modèle du lac Saint-Jean dans les domaines bioclimatiques du Québec.

### 4.3 Annexe 3 : Fiche de relevé terrain


VALIDATION - IQH BLEUET										
<b>Parcelle #</b>					 Agence de gestion intégrée des ressources	<b>Secteur</b>				
<b>IQH</b>						<b>Coordonnées géographiques</b>				
<b>Date</b>						<b>Peuplement d'origine</b>				
<b>Évaluatrice / Évaluateur</b>						<b>Photo</b>				
<b>BLEUET</b>										
Évaluation du pourcentage de recouvrement (plante) <sup>(1)</sup>			Présence de fleurs - fruits (selon la saison)			Potentiel de cueillette observé				
0 - 10 % <input type="checkbox"/>			faible <input type="checkbox"/>			Faible <input type="checkbox"/>				
11 - 25 % <input type="checkbox"/>			moyenne <input type="checkbox"/>			Moyen <input type="checkbox"/>				
26 - 50 % <input type="checkbox"/>			forte <input type="checkbox"/>			Élevé <input type="checkbox"/>				
51 - 75 % <input type="checkbox"/>			très forte <input type="checkbox"/>							
76 - 100 % <input type="checkbox"/>										
<b>ÉVALUATION DE LA COMPÉTITION</b>										
<b>Évaluation du recouvrement de la strate strate herbacée <sup>(2)</sup></b>										
	recouvrement (%)									
<b>Espèce</b>	0 - 10	11 - 25	26 - 50	51 - 75	75 - 100					
1										
2										
3										
4										
<b>Évaluation du recouvrement de la strate arbustive <sup>(2)</sup></b>										
	recouvrement (%)					Hauteur moyenne de la strate				
<b>Espèce</b>	0 - 10	11 - 25	26 - 50	51 - 75	75 - 100					
1						4 - 7 m	<input type="checkbox"/>			
2						2 - 4 m	<input type="checkbox"/>			
3						1 - 2 m	<input type="checkbox"/>			
4						< 1 m	<input type="checkbox"/>			
<b>Composition du couvert forestier :</b>						/ / indiquer le couvert				
<b>Densité du couvert</b>	<b>Classe</b>			<b>Hauteur moyenne des arbres</b>	<b>Code</b>					
> 80 %	<b>A</b>	<input type="checkbox"/>		> 22 m	<b>1</b>	<input type="checkbox"/>				
60 - 80 %	<b>B</b>	<input type="checkbox"/>		17 - 22 m	<b>2</b>	<input type="checkbox"/>				
40 - 60 %	<b>C</b>	<input type="checkbox"/>		12 - 17 m	<b>3</b>	<input type="checkbox"/>				
25 - 40 %	<b>D</b>	<input type="checkbox"/>		7 - 12 m	<b>4</b>	<input type="checkbox"/>				
	<b>D</b>	<input type="checkbox"/>		4 - 7 m	<b>5</b>	<input type="checkbox"/>				
Absent (0 - 25 %)	<b>n/a</b>	<input type="checkbox"/>		2 - 4 m	<b>6</b>	<input type="checkbox"/>				
				< 2 m	<b>7</b>	<input type="checkbox"/>				
<b>Paramètres écologiques</b>										
<b>Topographie</b>										
Situation de la pente										
Exposition:										
Forme:										
Inclinaison:										
<b>Drainage</b>										
Excessif										
Rapide										
Bon										
Modéré										
Imparfait										
Mauvais, très mauvais										
<b>Texture</b>										
Gravier										
Sable										
Limon										
Argile										
Roc ou blocs										
Organique										
<b>Perturbation du sol</b>										
Aucune										
Scarifiage										
Abattage bois long										
Abattage bois court										
Érosion										
Ornière (présence)										
<b>Perturbation (Âge)</b>										
Feu <input type="checkbox"/>						3 à 8 ans				
Coupe <input type="checkbox"/>						9-12 ans				
						2 ans, 13-20 ans				
						21 ans +, 1 an				
<b>TYPE ÉCOLOGIQUE OBSERVÉ</b>										
<b>GROUPE D'ESPECES INDICATRICES (GEI)</b>										
<b>Remarques :</b>						1 - Rayon d'environ 20 m autour du point central de la parcelle = idée générale du site. 2 - Rayon de 20 m. Évaluer le recouvrement des principales herbacées et des arbustes. Voir liste en annexe.				

Tableau A4: Formulaire de relevé terrain utilisée par les équipes de ce projet

#### 4.4 Annexe 4 : Indice V de cramer - Intensité du lien entre les variables

	forme	drainage	texture	Type_pertu	Age_pertu	type_Ecolo	Gr_esp_ind	type_couv	den_couv	haut_couv	esp_arbl	esp_herbal
reponse	0,22	0,19	0,34	0,22	0,26	0,2	0,2	0,2	0,4	0,4	0,37	0,5
forme		0,2	NA	0,2	0,2	0,2	NA	0,2	0,2	0,2	NA	NA
drainage			0,5	0,3	0,2	0,2	0,5	0,1	0,2	0,2	0,3	0,4
texture				NA	NA	0,5	0,5	0,6	NA	0,3	0,5	0,6
Type_pertu					0,4	0,1	NA	0,1	0,4	NA	NA	NA
Age_pertu						0,3	NA	0,2	0,2	NA	NA	NA
type_Ecolo							0,5	0,6	0,3	0,2	0,4	0,5
Gr_esp_ind								0,5	0,4	0,4	0,5	0,4
type_couv									0,1	0,1	0,4	0,6
den_couv										0,4	NA	NA
haut_couv											0,3	0,4
esp_arbl												0,5

Tableau A5: Matrice de lien