

N° 2018 AGPT 0003

Doctorat AgroParisTech

THÈSE

pour obtenir le grade de docteur délivré par

**L'Institut des Sciences et Industries
du Vivant et de l'Environnement
(AgroParisTech)**

Spécialité : Sciences du Bois et des Fibres

présentée et soutenue publiquement par

Lara CLÍMACO DE MELO

le 21 Août 2018

Error propagation in forest growth models in the context of regional forecasts

Directeur de thèse : **Mathieu Fortin**

Co-encadrement de la thèse : **Robert Schneider**

Jury

M. Benoît Courbaud,	Directeur de recherche, IRSTEA, Grenoble, France	Président du jury & Rapporteur
Mme Anikka S. Kangas,	Directrice de recherche, LUKE, Helsinki, Finland	Rapporteur
Mme Celine Meredieu,	Chargée de recherche, INRA, Bordeaux, France	Examinatrice
M. Mathieu Fortin,	Ingénieur de recherche, AgroParisTech, Nancy, France	Directeur de thèse
M. Robert Schneider,	Professeur, UQAR, Québec, Canada	Co-encadrant de thèse

AgroParisTech

UMR Silva 1092 AgroParisTech-INRA, 14 rue Girardet, 54042, Nancy, France

Acknowledgments

I would like to express my sincere gratitude to my supervisor, Mathieu Fortin. Over the years, thanks to his encouragement, patience and advices, I have overcome many difficulties – and crossed many oceans. Thanks for showing that leadership and generosity can go together. I also thank my co-supervisor, Robert Schneider, for all the discussions and useful comments, for contributing to my learning experience, and for the great internship in Rimouski. I am also thankful to Rubén Manso for his participation in my thesis committee, and especially for sharing his knowledge so willingly, and for being always available to help me.

I would also like to thank the external members of my defense committee, Benoît Courbaud, Annika Kangas and Celine Meredieu for your insightful comments and suggestions, as well as for the kindness in the discussion.

I could not have undertaken this PhD without the scholarship funded by the Brazilian National Council for Scientific and Technological Development (CNPq). Many thanks for this opportunity. I also thank the available funds from Labex Arbre, which gave me the opportunity to present my research in various conferences.

I would like to say a big thank you to those who participated indirectly in this thesis by sharing many chats, laughs, beers and coffees, and friendship. I was really fortunate to meet amazing people from different places, from whom I have learned about diversity, identity, and humanity. I truly thank the folks from the Capoeira. My energies were renewed weekly by sharing music and joy with such beloved people. Also, I'm definitely a better person thanks to the learning I had with so strong, truth and inspirational women. Despite the distance, I had a lot of love and support from my dearest friends that I grew up with in Brazil. Thanks for existing!

To my beloved family, from whom I have been apart, thanks for your understanding. Foremost thanks go to my mother, who among many things, taught me about discipline, persistence, and will. These were fundamental internal resources to stand alone pursuing my project. Finally, I would like to dedicate this work to my nephew, André Luiz.

Contents

1	Introduction	11
1	Techniques and applications of growth models	12
2	Uncertainties in forest growth forecasts	14
2.1	Sources of uncertainties	14
2.2	Methods for estimating uncertainty	15
3	Objectives	16
4	Thesis context	18
4.1	Forest dynamics in Quebec	18
4.2	Quebec's Forest Inventory	21
4.3	ARTEMIS-2009	21
4.4	Simulation framework	23
2	Estimating model- and sampling-related uncertainty in large-area growth predictions	25
1	Introduction	26
2	Material and methods	28
2.1	Growth model	28
2.2	Input dataset	30
2.3	Simulation framework	32
2.4	Hybrid estimator	32
3	Results & Discussion	33
4	Conclusions	44
3	Using survival analysis to predict the harvesting of forest stands in Quebec, Canada	45
1	Introduction	46
2	Material and methods	48
2.1	Dataset	48

2.2	Statistical development	51
2.3	Model evaluation	54
3	Results	56
4	Discussion	60
4	The effect of natural and anthropogenic disturbances on the uncertainty of growth forecasts	65
1	Introduction	66
2	Material and methods	68
2.1	ARTEMIS growth model	68
2.2	Uncertainty estimation	70
2.3	Study area and dataset	71
2.4	Forecasting	73
3	Results	74
4	Discussion	78
5	Conclusions	82
5	Discussion and Perspectives	85
1	Research problems	86
2	Framework	87
3	Estimation of uncertainties in large-area growth forecasts	88
4	Perspectives	92
A	Corrected bootstrap variance estimator	95
B	Appendix of Paper II	101
C	Appendix of Paper III	103
	Bibliography	105
	Bibliography	105
	Abstract	123

List of Figures

2.1	Structure of ARTEMIS-2009 simulation system considering an iterative process. Rectangles are for models.	29
2.2	Distribution of the permanent plots that compose the network of Quebec's provincial forest inventory in Bas-Saint-Laurent.	31
2.3	Mean predicted basal area (m^2ha^{-1}) illustrated for each ecotype. The growth curves are related with the full stochastic and the partial stochastic predictions (disabling diameter, mortality and recruitment sub-models).	34
2.4	Contribution of the model and the sampling to the total variance of the full stochastic predictions in basal area.	36
2.5	Variance contribution of each sub-model that compose the ARTEMIS growth model by ecotype. This contribution is calculated as the relative decrease in the total variance (%) once the stochasticity of the sub-model has been disabled.	39
3.1	Distribution of the 12,570 permanent plots that compose the network of Quebec's provincial forest inventory covering the northern temperate zone and the boreal zone.	49
3.2	Observed proportions and predicted probabilities of events evaluated by the Hosmer-Lemeshow test.	57
3.3	Simulation of the AAC effects on harvest probability for Bas-Saint-Laurent. The dots represent the values of AAC. The line represents the cumulative harvest probability. The reference values are: BA = $18 \text{ m}^2\text{ha}^{-1}$; slope = 4% to 8%; stem density = $725 \text{ m}^2\text{ha}^{-1}$; and dynamics class = mixed.	59

3.4	Effect of the model variables on the predicted harvest probabilities for a 10-year interval. (a) basal area effect; (b) stem density effect; (c) dynamics class effect; (d) slope class effect. The reference values are: BA = 18 m ² ha ⁻¹ ; slope = 4% to 8%; stem density = 725 m ² ha ⁻¹ ; and dynamics class = mixed.	60
4.1	Flowchart of ARTEMIS-2009 considering its iterative process. Dark gray boxes represent the dynamic sub-models. Dotted light gray boxes are the static sub-models.	69
4.2	Distribution of the 393 permanent plots in Bas-Saint-Laurent. The plots are located in the ecological regions classified according to the MFWP: Appalachian Hills (4f); Baie des Chaleurs Coastline (4g); Gaspé Coastline (4h); Mountains of Gaspé Peninsula (5h); Highlands of Gaspé Peninsula (5i).	72
4.3	Mean predicted volumes (m ³ ha ⁻¹) and their 0.95 confidence interval for the Bas-Saint-Laurent region. The confidence intervals rely on the assumption of a normal distribution. The solid line represents the scenarios without harvesting, while the dashed line represents the scenario including harvesting.	75
4.4	Model, sampling and total variances illustrated per growth scenario. Model contribution: gray dashed line; Sampling contribution: dark gray dotted line; Total variance: black solid line.	76
C.1	Contribution of the model- and sampling-related variances to the total variance of volume predictions by ecotypes in the Bas-Saint-Laurent region. The simulations were run under the baseline scenario, i.e. the scenario without disturbances.	104

List of Tables

2.1	Summary of study area and sample plots. Area is the total area occupied for each one of the three ecotypes. The values in parentheses show the range of the variable.	31
2.2	Model- and sampling-related variance contribution (m^4ha^{-2}) in the full stochastic predictions per ecotype. The relative contribution appears in parentheses.	35
2.3	Variance values (m^4ha^{-2}) of the full stochastic prediction and the respective disabled growth sub-models represented per ecotype. . . .	43
3.1	Summary of some plot-level characteristics in the intervals between 1988 and 2014 for each dynamics class (B: broadleaved; M: mixed; C: coniferous; n : number of intervals; the range of the variables appears in parentheses). A particular interval belongs to a period when its final date falls into the range.	51
3.2	Annual allowable cut volumes (AAC) by administrative regions since 1988. The values of AAC and area represent the mean during the period, including public and private lands.	52
3.3	The goodness-of-fit for the tested models.	56
3.4	Maximum likelihood estimates of the parameters in the final model (Model 8) with their associated standard errors and approximate t -values.	58
4.1	Summary of 393 plots in the dataset. Attributes were broken down for the most abundant species. The minimum and maximum values are shown in parentheses.	73
4.2	Model and sampling-related variance contribution (m^2ha^{-1}), as well as the total variance estimated for each one of the four scenarios. The percentage contribution appears in parentheses.	77

Chapter 1

Introduction

Contents

1	Techniques and applications of growth models	12
2	Uncertainties in forest growth forecasts	14
2.1	Sources of uncertainties	14
2.2	Methods for estimating uncertainty	15
3	Objectives	16
4	Thesis context	18
4.1	Forest dynamics in Quebec	18
4.2	Quebec's Forest Inventory	21
4.3	ARTEMIS-2009	21
4.4	Simulation framework	23

1 Techniques and applications of growth models

Forest growth models are helpful tools due to their ability to transform long-term databases into future scenarios through simulation techniques (Vanclay, 1994, p. 4). When faced with different management options, models provide answers about the potential consequences of specific objectives (Pretzsch et al., 2008). Over time, new goals have boosted model developments in order to meet societal, economic and environmental demands. Starting with simple techniques, these developments have now reached complex computational procedures. Along with environmental data, growth models serve the purpose of prediction and guide (Vanclay, 1994, p. 4).

Several classifications of modeling approaches exist in the literature. In this paper, we summarize the concepts and theories that compose the 200 years of growth model history (Pretzsch, 2009). We chose to present the growth model classifications found in Vanclay (1994). Only the spatial level (stand level, size class and tree level), and the model structure (empirical or mechanistic) classifications are described.

Stand-level models are generally simple and robust. They predict the change in stem density, basal area and volume depending on the initial stand characteristics. They can be applied to mixed forests (e.g., Pothier and Auger, 2011) but are normally limited to monospecific forests (Pretzsch et al., 2008). Size-class models are considered by Vanclay (1994, p. 34) as an intermediate approach between stand-level models and tree-level models. Size-class models provide the changes in frequencies per tree class between an initial and a final date, considering growth, removal and mortality (Pretzsch, 2009, p. 433).

This thesis strictly focused on the third category: tree-level models. The approach consists of following individual trees through a temporal and, in some cases, spatial frame. A great advantage of this approach lies in its capacity to consider both pure and mixed stands of any structure (Pretzsch et al., 2008). In terms of applicability, Vanclay (1994, p. 57) mentioned that the tree-level models:

1. be coupled to harvest and conversion simulators;
2. account for competition effects;
3. take physiological processes into account.

Depending on whether or not tree-level models take the spatial distribution of the trees into account, they can be further distinguished as distance-dependent or distance-independent models (Porté and Bartelink, 2002). Distance-dependent

models include explicit spatial competition indices, whereas distance-independent models rely on non-spatial competition indices. For growth and yield predictions, distance-independent models are recommended, whereas distance-dependent models are often considered for forest succession issues ([Porté and Bartelink, 2002](#)).

Regarding the model structure, growth models can be classified as either empirical or mechanistic. An empirical model is built on the relationships obtained from measurements of target variables in sample plots ([Pretzsch et al., 2008](#)). For [Porté and Bartelink \(2002\)](#), the application of these models requires a validation of their empirical relationships, which represents a drawback. Mechanistic models, which are also known as process-based models, are meant to predict forest development through its basic physiological and ecological laws. Managers who are interested in forest development in the context of environmental changes acknowledge the importance of these models ([Pretzsch et al., 2008](#)). In terms of uncertainty, process-based models are considered to be more uncertain than empirical models due to the large number of parameters and input data that represent all the processes of the system ([Adams et al., 2013](#)).

Beyond the structure and the approach, modelers and users should determine if the model meets accuracy requirements in order to predict growth in a reliable way. Some essential steps are necessary: models should be evaluated through the statistical assessment of bias and accuracy ([Vanclay and Skovsgaard, 1998](#)). Additionally, the residual error distribution is helpful to identify potential unacceptable patterns ([Bokalo et al., 2013](#)). Moreover, through sensitivity analysis, the variability of model predictions is investigated in relation to the variability in the input data, estimated coefficients and the interaction between the sub-models. Sensitivity analysis can guide model enhancements ([Soares and Tomé, 2007](#)). Finally, evaluating forest resources through models also includes uncertainties that can be estimated using different methods. Communicating those uncertainties is helpful for risk management, to enhance forest inventory designs and to improve model predictions ([Holopainen et al., 2010](#)).

Whenever a harvest model is available, it is often integrated into growth models. This allows for business-as-usual scenarios that can support forest planning ([Mäkinen, 2010](#)). In this respect, an important task of forest planning is to manage harvest activities. In Quebec, sustainable forest management is enforced by law ([MFFP, 2003](#)). Among different objectives and commitments, this concept encompasses a mandatory calculation of the maximum volume that can be sustainably harvested, better known as annual allowable cut volumes (AAC) ([MFFP, 2003](#)).

Integrating a harvest model into a growth model results in a system that can be helpful for AAC calculations.

The forest science community contributes to these issues by adapting approaches and developing methods and models flexible enough to support decision-making in some contexts. These types of initiatives can be found in [Thurnher et al. \(2011\)](#), [Fortin and Deblois \(2007\)](#), [Mäkinen \(2010\)](#), [Antón-Fernández and Astrup \(2012\)](#) and [Gaudreau et al. \(2016\)](#).

The prediction of what would be the future forest resource is associated with uncertainties. Consequently, the risk related to unwanted outcomes from decisions in forest management has to be based on the assessment of these uncertainties ([Paré et al., 2016](#)). In this context, assuming that growth models are useful tools for forest management, a natural question arises as to their application: How uncertain are growth forecasts? Despite the development of techniques and knowledge, tree-level models remain complex systems. Growth forecasts are by nature uncertain, and such uncertainties should be estimated. The next section presents concepts and approaches to understand and treat uncertainties in growth forecasts.

2 Uncertainties in forest growth forecasts

2.1 Sources of uncertainties

Uncertainty analysis is normally required when working with models, especially complex models. Tree-level models that explain and predict forest growth belong to these complex models. In this section, we review some concepts about the field of uncertainty analysis. Because this analysis comprises a wide variety of fields and application areas, the purpose is not to provide a review of the existing theories and practices. Still, some basic and common concepts are a starting point: risk, sensitivity and uncertainty analyses.

A risk analysis is seen as an arrangement of the information about uncertainties ([Aven, 2003](#), p. 13). The appropriate characterization, propagation and representation of uncertainty is an important aspect of a risk analysis. Risk management includes measures that help avoid the occurrence of hazards/threats or that reduce their potential damage. More specifically, this type of analysis provides estimates of the risks expressed by probabilities and statistically expected values ([Aven, 2003](#), p. 13).

Sensitivity analysis aims at indicating how sensitive the model output is with

respect to individual inputs such as parameters or assumptions made (Helton et al., 2006). Closely related to sensitivity but still distinct, uncertainty can be seen as the natural variability in the system under study, which is also known as aleatory uncertainty (Aven et al., 2014, p. 61). Uncertainty can also be defined as ignorance or the lack of knowledge about the true value. In this case, it is referred to as epistemic uncertainty. Both types refer to the uncertainty observed in the output that arises from inputs such as data, methods or models used (Aven et al., 2014, p. 61). Kangas and Kangas (2004) add that epistemic uncertainty might be reduced by advancing knowledge, but aleatory uncertainty remains independent of knowledge advancement.

A quantitative analysis of uncertainty first identifies the sources that impact the system under study. In the forestry literature, the sources that are often reported arise from sampling and model errors (Kangas and Kangas, 2004). Model uncertainty is due to parameter estimation and residual errors that are linked to the aleatory uncertainty. The conceptual model and the implementation of the model are also elements of the accumulated uncertainty associated with models (Walker et al., 2003).

Sampling uncertainty comes from the fact that only a part of the population is sampled (Gertner, 1990). It can also be interpreted as the degree to which the sample is representative (Walker et al., 2003). These interpretations mean that we do not know the entire population and, as a consequence, it is reasonable to assume that an error will arise (Condés and McRoberts, 2017).

In growth forecasts, exogenous sources can also contribute to uncertainties. For instance, climate change can be seen as an important source of uncertainty when forecasting growth (Petr et al., 2014; Xu et al., 2009). Moreover, the implementation of natural hazards in growth modeling can greatly contribute to the model-related uncertainties as well. That is because of the stochastic nature of both natural disturbances and climate change.

2.2 Methods for estimating uncertainty

Uncertainty estimation has been considered under various frameworks and definitions (e.g., Fu et al., 2017; Kangas, 1999; Mäkinen, 2010; Yanai et al., 2012). It can be estimated through a quantity such as variance that makes it possible to express the variability associated with growth forecasts. Uncertainties that stem from the forecasts are normally propagated either analytically or through simulation. Error

propagation provides insights into the error in the predictions that are due to errors in the model inputs (Berger et al., 2014; Versteegen, 2016).

The type of uncertainty - epistemic or aleatory- can be used to guide and define the appropriate methods to treat them. According to Aven et al. (2014, p. 61), both epistemic and aleatory uncertainties can be treated through a probabilistic approach. Much of the recent development to treat uncertainties has focused on probabilistic techniques. However, there are cases for which the representation of epistemic uncertainty by these probabilistic techniques is limited (Pedroni and Zio, 2012). To address these cases, fuzzy and possibility theories are recommended alternatives (Aven et al., 2014, p. 64).

The complexity of the issues and the model can also guide the selection of the appropriate method. According to Bakr and Butler (2005), simple cases such as linear or nonlinear models can have their errors propagated through analytical methods such as Taylor series expansions (e.g., Gertner, 1990). Taylor series consider moments of the probability distribution of the input quantities (e.g., mean or variance) to provide moments of the output quantities. In cases for which the entire probability distributions of the output are of interest, or when the model is too complex, Taylor series no longer apply and the Monte Carlo technique (MC) is a recommended alternative (Aven et al., 2014, p. 63).

The MC technique (Rubinstein and Kroese, 2007) is a suitable tool for supporting uncertainty estimation and assessment. It consists of drawing random deviates from distributions to account for the variability in the input data and model parameters. It then traces the distribution of model outputs from the MC realizations (Baraldi and Zio, 2008). A great advantage of the MC technique is its flexibility, especially regarding the assumptions about probability distributions or correlations. The drawback is that it is time-consuming (Refsgaard et al., 2007).

An option to represent uncertainties consists of using variance estimators based on resampling methods (Berger et al., 2014). Recently, Fortin et al. (2018) developed a bootstrap variance estimator that can be applied in the hybrid inference context (Corona et al., 2014), i.e., the inference that relies on both model and sampling uncertainties. This is a common context in forestry.

3 Objectives

Over the last decades, the implementation of sustainable forest management strategies and national carbon reporting require information, criteria and regulations that

can support current international agreements (e.g., [Food and Agriculture Organization of the United Nations, 2010](#); [Service, 1995](#); [United Nations Conference on Environment and Development, 2010](#)). This has led to an increasing demand for predictions of future forest conditions over large areas. These predictions are obtained using growth models, which are fitted to sample-plot data and not to a complete census ([Condés and McRoberts, 2017](#)). Upscaling these local models to obtain large-area forecasts has practical implications in terms of uncertainties and has to follow a particular inference scheme. The hybrid inference scheme is a concept that incorporates uncertainties from both model and sampling sources ([Corona et al., 2014](#)). Recently, some studies have provided examples of uncertainty estimation in the context of hybrid inference in forestry (e.g., [Corona et al., 2014](#); [Fortin et al., 2016, 2018](#); [Fu et al., 2017](#); [McRoberts and Westfall, 2014, 2016](#); [Ståhl et al., 2014](#)).

Estimating uncertainties represents an essential step forward in this period of demand for large-area growth forecasts. The ability to address the uncertainties and associated risks can be beneficial over a wide range of decisions in forestry. Omitting uncertainties can lead to poor risk management, i.e., directly impact the choice of management alternatives, the forest product market or carbon projects ([Paré et al., 2016](#)). The importance of these issues was the driving force behind this study.

This Ph.D. is centered on the estimation of uncertainties in regional growth forecasts using a hybrid inference scheme. The effects of large-scale disturbances were also considered. We addressed these issues through three main objectives, each of them constituting a chapter of the thesis: (i) to estimate uncertainty arising from the model and the sampling in regional growth forecasts; (ii) to develop a harvest model based on survival analysis to predict the probability of harvest under a business-as-usual scenario; and (iii) to analyze the effect of large-scale disturbances on the uncertainties in regional growth forecasts. The growth simulator ARTEMIS-2009 ([Fortin and Langevin, 2010](#)), which applies to most forest types in Quebec, Canada, was taken as a case study. For each main objective, we delineated some specific goals, described below:

(i) Considering the major sources of uncertainty in large-area forecasts, our specific goal was to estimate the contribution of the model and the sampling to the total variances of regional growth forecasts. We also assessed the contribution of the different sub-models that composed the growth model to the total variance. The analysis combines the time frame and the forest types as factors influencing the behavior of the variance.

(ii) The survival analysis technique presents a structural flexibility that makes it possible to account for regional variables and time-varying covariates, which were assumed to influence harvest occurrence. The specific goal was to develop a survival model to predict harvest probabilities at the plot level. The model was to be integrated into ARTEMIS-2009, which would make it possible to account for current management practices over large areas. Such an implementation was required for the next objective.

(iii) The occurrence of major disturbances is known to increase the uncertainties of growth forecasts. This chapter explored the extent to which anthropogenic (harvest) and natural (spruce budworm outbreaks) disturbances affected the variances of regional growth forecasts. The latter analysis made it possible to determine the disturbance agent that contributed the most to the variance. An important outcome of this chapter was to provide guidelines as to how to reduce those variances.

The specific goals address current issues in forest growth modeling. They were structured as stand-alone manuscripts, referred to as “Papers”, and followed by roman numerals. Following is the list of publications:

Paper I. Melo, L.C., Schneider, R., and Fortin, M. (2018). Estimating model- and sampling-related uncertainty in large-area growth predictions. *Ecological Modelling*. 390: 62-69.

Paper II. Melo, L.C., Schneider, R., Manso, R. and Fortin, M. (2017). Using survival analysis to predict the harvesting of forest stands in Quebec, Canada. *Canadian Journal of Forest Research*. 47: 1066–1074.

Paper III. Melo, L.C., Schneider, R., and Fortin, M. (2018). The effect of natural and anthropogenic disturbances on the uncertainty of large-area forest growth forecasts. *Environmental Modelling and Software* *Submitted*.

4 Thesis context

4.1 Forest dynamics in Quebec

Modeling forest dynamics is related to the difficult task of tracking the variables that explain recruitment, growth and mortality, as well as their response to disturbances (Fraver et al., 2009). The intensity of ongoing environmental changes makes this

task more complex and requires an understanding of the underlying patterns. This is valuable knowledge to be considered in flexible management plans (Bartels et al., 2016). In Quebec, in particular, growth models are widely used to predict forest dynamics to support the design of ecosystem-based forest management strategies and to estimate the AAC (MFFP, 2013).

Forest dynamics are dependent on forest composition, among other factors. A brief description of Quebec’s forests is presented in this paper. A hierarchical ecological classification adopted by the Ministry of Forests, Wildlife and Parks (MFFP, 2016) divides Quebec’s forests between the boreal and temperate zones.

According to Grondin et al. (2009), boreal forests in Quebec are predominantly composed of balsam fir (*Abies balsamea* Mill.), white birch (*Betula papyrifera* Marsh.), white spruce (*Picea glauca* Voss) and black spruce (*Picea mariana* Britton). Temperate forests are mainly composed of sugar maple (*Acer saccharum* Marsh.), American basswood (*Tilia americana* L.), yellow birch (*Betula alleghaniensis* Britton) and ash (e.g., *Fraxinus americana* L.).

The main natural disturbances that shape Quebec’s boreal forests are fires (Kneeshaw and Bergeron, 1998; Pham et al., 2004) and insect outbreaks such as those by the spruce budworm (*Choristoneura fumiferana* (Clem.), SBW) (Grondin et al., 2009). Since the fire regime is largely influenced by regional climate, its recurrence is variable across the provinces (Boucher et al., 2003). Regions with drier climates are subject to fire cycles shorter than 100 years. In regions with humid climates such as Eastern Quebec, the estimated fire cycles are longer than 300 years (Bergeron, 2000). SBW outbreaks have great impacts since the insects rapidly defoliate the trees and create openings (Morin, 1994). Outbreak history shows an average recurrence of 35 years (Boulanger and Arseneault, 2004).

Temperate forests are characterized by frequent natural canopy gap dynamics (Fraver et al., 2009). Episodes of severe disturbances such as storms, insect outbreaks and forest dieback events might have an influence but are quite rare (Cohen et al., 2016). Still, Payette et al. (2016) reported an influence of wildfires on the development of maple stands.

Harvest activities were reported as the main anthropogenic agent disturbing the temperate and boreal ecosystems of Quebec (Bergeron, 2000; Bergeron et al., 2000). These practices are a concern for sustainable forest management because they can alter forest dynamics. For instance, harvest activities were reported as being responsible for the accelerated succession in mixed forests in central Canada (Taylor et al., 2013).

The administrative region of Bas-Saint-Laurent (BSL) is a specific case study in this thesis. The region is located in Eastern Quebec, at the northern limit of the Great Lakes and Saint Lawrence regions (Rowe, 1972). BSL forests are located at the transition point between the temperate deciduous and the boreal coniferous forest zones (Kneeshaw et al., 2011).

The literature describes the forest composition and dynamics of BSL with respect to the pre- and post-industrial periods. According to Boucher et al. (2009b), forests were dominated by conifers during the pre-industrial period. At that time, maples and birches were mostly in patches, except for elevated areas where maples were abundant (Boucher et al., 2009b). Forest dynamics were influenced by insect outbreaks, which were punctual but of great intensity (Bouchard and Kneeshaw, 2007). Spruce budworm outbreaks had a major influence (Boucher et al., 2009b). Small-scale disturbances such as tree-fall gaps and windthrows were also observed (Payette et al., 1990).

Post-industrial forests are marked by intense harvesting (Boucher et al., 2009a,b). Successive periods of harvest activities have had a significant impact by shifting a greater proportion of forests from coniferous to mixed cover types (Archambault et al., 1998; Boucher et al., 2009b). Comparing the pre- and post-industrial forest composition, Dupuis et al. (2011) revealed increases in terms of tree density for maple, poplar (*Populus* spp.) and white birch. The current disturbance regime of this region is one of the harvest activities punctuated by outbreaks of spruce budworm (Boulanger and Arseneault, 2004). Regardless of these changes, the Provincial Forest Inventory data revealed that BSL forests are still mainly composed of balsam fir and white birch, followed by sugar maple, white spruce, red maple (*Acer rubrum* L.) and black spruce.

Identifying the impacts of harvest and SBW outbreaks on the Bas-Saint-Laurent forests has been the focus of some studies. Regarding harvest practices, Archambault et al. (1998) observed that clear-cutting in balsam fir-yellow birch ecosystems caused major changes to the original species composition. Archambault et al. (2006) observed short-term changes in the proportion of hardwood species under management, such as a greater increase of birch species abundance. Concerning natural disturbances, the cyclic spruce budworm outbreaks are responsible for the high mortality of balsam fir and white spruce (Duchesne and Ouimet, 2008). According to Bouchard and Kneeshaw (2007), post-outbreak stands were invaded by broadleaved species in the southern mixed-wood forests. The authors claimed that even with normal recruitment of balsam fir, constant mortality due to successive outbreaks

could reduce its ability to compete in the long term. The direct consequence is more favorable conditions for non-host species such as white birch.

4.2 Quebec's Forest Inventory

Quebec's forests were the focus of this thesis. The dataset used to meet the proposed objectives consisted essentially of field measurement data that were obtained from the network of permanent plots of Quebec's Ministry of Forests, Wildlife and Parks. This network has been monitored for almost 50 years now. In the early 1970s, the network was designed with the main objective of estimating growth, forested areas and their volumes. Over the years, this objective was redefined to include ecological characteristics. These permanent plots are distributed over an area of 588,200 km², covering public and private forests. Measurements are taken in the northern temperate zone, dominated by broadleaved and mixed stands, and in the boreal zone, in which coniferous stands are predominant (MFFP, 2014).

The network comprises a total of 12 570 permanent plots that were randomly distributed according to a stratified random design (MFFP, 2015b). In these circular plots with an area of 400 m², dendrometric, geographical and ecological data were collected for monitoring individual trees with a minimum diameter of 9.1 cm. Plots were meant to be re-measured every 10 or 15 years. However, logistic issues resulted in some irregular measurement intervals. The compilation of the inventory data provides information on the tree and plot levels, such as tree status (i.e., alive, dead or harvested), diameter, height, species composition, basal area and commercial volumes, among others (MFFP, 2015a).

Depending on the individual objectives of the thesis, the network of permanent plots was screened. Figures illustrating the plot distribution in relation to the specific objectives are included in each paper. In Papers I and III, growth forecasts were carried out using only the 2003 measurements in the Bas-Saint-Laurent forests as a reference. In Paper II, we developed a harvest model using the measurements from all over Quebec. For Paper II specifically, it was necessary to gather other sources of information at the provincial scale, precisely, the annual allowable cut volumes.

4.3 ARTEMIS-2009

In this thesis, we worked with the growth model ARTEMIS-2009, which is a distance-independent tree-level model (Fortin and Langevin, 2010). ARTEMIS was parame-

terized using the network of permanent plots of the Quebec provincial forest inventory. The model was designed to provide growth forecasts of public forests, which represent 92% of the forested areas in Quebec (MFFP, 2015a). In these forests, the most common potential vegetation was modeled individually. As a consequence, ARTEMIS contains 25 versions, each one representing a potential vegetation.

Originally, the growth model was composed of four dynamic and two static sub-models, which predicted the mortality probability, the diameter increment, the number of recruits and the diameter of these recruits. A fifth dynamic sub-model was integrated into the context of this thesis to predict plot-level harvest probabilities. The two other sub-models also available in ARTEMIS predict tree height (Fortin et al., 2009) and commercial volume (Fortin et al., 2007). It is worth mentioning that the impact of spruce budworm defoliation is taken into consideration through the mortality sub-model, which adapts the probabilities whenever an outbreak occurs. More details about the model types and developments can be found in Fortin and Langevin (2010).

ARTEMIS uses forest inventory information at the tree and plot levels. Tree species, stem density (tree ha⁻¹), basal area (m²ha⁻¹) and slope class, among others, are some of the explanatory variables in the model. Moreover, ARTEMIS also considers climatic variables. Mean annual precipitation (mm) and temperature (°C) over the 1981-2010 period are entries in some sub-models. These variables are estimated using BioSIM software (Régnière et al., 2010).

The modeling approach of ARTEMIS is based on 10-year growth steps. Growth forecasts for longer time periods can be obtained by re-inserting the output of a given step as the starting point of the next step. The simulations can be run in a deterministic or stochastic mode. The stochastic mode relies on the MC technique (Rubinstein and Kroese, 2007).

The random effects and the parameter estimates follow Gaussian distributions. The distributions of the residual errors depend on the sub-model. For example, the residual error in the diameter increment sub-model follows a Gaussian distribution. However, that of the mortality sub-model follows a uniform distribution bounded by 0 and 1. For multivariate distributions such as those of parameter estimates, a Cholesky factorization is used to maintain the covariance between the estimates. The reader is referred to the study of Fortin and Langevin (2010) for an extensive description of ARTEMIS-2009.

The model outputs are tree-level predictions that are aggregated into plot-level predictions. Using some upscaling techniques such as direct extrapolation methods

([Wu et al., 2006](#)), plot-level predictions can subsequently be upscaled into large-area predictions.

ARTEMIS is composed of various sub-models that interact with each other and, as such, is an example of a complex model ([Fortin et al., 2018](#)). Those systems include complex and highly non-linear interactions between the inputs and the parameters that describe the model outputs ([Willems, 2012](#)). Such complexity has important implications in the error propagation.

4.4 Simulation framework

In this thesis, all simulations were run on the CAPSIS platform ([Dufour-Kowalski et al., 2012](#)), which, among other models, includes ARTEMIS-2009. CAPSIS is a project focused on forest growth and yield modeling, and has been designed to host different types of forest growth models.

In our study, the stochastic simulations required a great amount of time and memory. We worked with a server where 16 GB of memory were made available for CAPSIS. The plot-level predictions were exported in comma separated files, which were then imported into RStudio ([RStudio Team, 2015](#)). In this environment, the estimator that allowed for uncertainty estimation was programmed.

Chapter 2

Estimating model- and sampling-related uncertainty in large-area growth predictions

L.C. Melo, R. Schneider, M. Fortin

Contents

1	Introduction	26
2	Material and methods	28
2.1	Growth model	28
2.2	Input dataset	30
2.3	Simulation framework	32
2.4	Hybrid estimator	32
3	Results & Discussion	33
4	Conclusions	44

1 Introduction

Knowledge of forest growth is important for assessing sustainable forest management (Peng, 2000), understanding the effects of climate change on forest dynamics (Ameztegui et al., 2015) and establishing changes in ecosystem productivity and biomass accumulation (Paré et al., 2016; ?). Given those needs, it is not surprising that forest growth models have become indispensable tools for forest managers and policy makers. Models are a simplification of complex phenomena and as such they are subject to prediction errors. Model evaluation is then required to ensure user confidence. A thorough evaluation involves examinations of model design, fitting and implementation (Vanclay and Skovsgaard, 1998).

Since the mid-20th century, forest growth models have increased in complexity in order to meet stakeholder demands, including not only economical aspects but ecological and social issues as well (Porté and Bartelink, 2002). Multi-objective models have been developed. Nevertheless, this increasing complexity has non-negligible consequences (Pretzsch, 2009). A greater number of model parameters, random effects and interactions will increase prediction uncertainty (Walker et al., 2003).

Uncertainty is described as the lack of knowledge and random variation that arises from multiple sources of errors (Aven et al., 2014). Error propagation provides insights into uncertainties in predictions. It computes how much of output predictions are uncertain due to error propagation from the input model (Berger et al., 2014; Sexton et al., 2015). Previous studies about uncertainties in forest growth models have already highlighted the fact that sufficient information is frequently missing or incomplete, especially when it comes to propagating the errors from the tree to stand, regional or national levels (Phillips et al., 2000).

Over the last two decades, there has been an increasing demand for large-area predictions at the regional and national levels mainly due to international agreements concerning climate change (Ciais et al., 2008; Groen et al., 2013). Uncertainty estimation of large-area predictions is not straightforward for several reasons. Two major sources of uncertainty can be identified: the model and the sampling. Model uncertainty arises from model misspecification, the estimation of model coefficients and the residual variation (Kangas, 1999). Sampling uncertainty is due to the fact that the initial forest conditions are usually unknown. Instead, they are estimated from a forest inventory, leading to the propagation of sampling errors when large-area predictions are carried out (Breidenbach et al., 2014).

The aforementioned context is one of hybrid inference. The term hybrid was coined by [Corona et al. \(2014\)](#) and refers to the fact that the inferential process relies on both a model and a probability design ([McRoberts et al., 2016](#)). It arises when the variable of interest is not measured or not measurable, such as future forest growth, and when the explanatory variables are only available from samples and not for the whole population ([Fortin et al., 2016](#)). Some authors have already applied hybrid estimators in the context of large-area estimation of volume, biomass and carbon (e.g., [Healey et al., 2012](#); [McRoberts et al., 2016](#); [Saarela et al., 2015](#); [Ståhl et al., 2011, 2016](#)). When working with complex models such as tree-level growth models, current analytic hybrid estimators, i.e., those based on algebra (e.g., [Ståhl et al., 2011](#)), can rarely be applied. An alternative consists in using bootstrap hybrid estimators ([Fortin et al., 2018](#)).

Quantifying sampling and model errors of large-area growth predictions is essential since it can help us identify which issues need to be addressed in order to reduce the uncertainty of the predictions. Previous studies on large-area estimates of volume and biomass have shown that the major source of uncertainty originated from the sampling ([Breidenbach et al., 2014](#); [McRoberts and Westfall, 2014](#); [Ståhl et al., 2014](#)). However, the use of growth models in the context of hybrid inference adds a temporal variability. [Kangas \(1999\)](#) reported that uncertainty increased along with projection length. Thus, the contribution of both sampling and models to prediction uncertainty may also change along the projection length. To the best of our knowledge, this temporal variability has not been addressed, with the notable exception of [Condés and McRoberts \(2017\)](#) who worked on short-term predictions. Moreover, tree-level growth models are complex when compared to stand-level models since they include many sub-models. This system of sub-models follows the dynamic of individual tree development through a temporal and spatial frame ([Pretzch et al., 2008](#)). Because mortality and recruitment are highly stochastic ([Sheil and May, 1996](#)), it could be anticipated that they would contribute more to the total uncertainty. Identifying which one of the sub-models contributes more uncertainty could help to improve models, understand forest dynamics and reduce prediction uncertainty. As far as we know, this has not been addressed yet.

The aim of this study was to estimate uncertainty in growth predictions at the level of a large area. To do this, we worked on two hypotheses: (i) over long simulation periods, i.e., a 100-year prediction, model uncertainty becomes greater than sampling uncertainty; (ii) among the model components, mortality is the major contributor to prediction uncertainty, then followed by recruitment. To confirm

or invalidate these two hypotheses, we aimed at decomposing the total prediction variance into a model and sampling component that was made possible by using a bootstrap hybrid estimator. Secondly, through a variance decomposition approach, the prediction uncertainty was also decomposed into sub-models (i.e., growth, mortality and recruitment) in order to assess which one contributed the most to the variance of large-area growth predictions. We worked with the ARTEMIS tree-level growth model (Fortin and Langevin, 2012), which was used to generate large-area predictions for the Bas-Saint-Laurent region in Quebec, Canada.

2 Material and methods

2.1 Growth model

The 2009 version of the ARTEMIS distance-independent tree-level growth model was built, fitted and evaluated using data from the network of permanent plots of Quebec’s provincial forest inventory. This network consists of 12,570 randomly located sample plots that were established in Quebec’s commercial forests and that has been measured since the 1970s. ARTEMIS takes the vast majority of the forest types in Quebec into account (Fortin and Langevin, 2010, 2012).

The model consists of four dynamic and two static sub-models (Fig. 2.1). The dynamic parts are those typical of population dynamic models: a mortality sub-model, a diameter increment sub-model for survivor trees, and two sub-models that predict the number and the diameter at breast height (DBH, 1.3 m in height) of the recruits, respectively. The static sub-models predict tree height and commercial volume. All sub-models are of the linear or generalized linear type. The model is based on 10-year growth intervals. Longer predictions are obtained through an iterative procedure, the result of the previous interval being re-inserted in the model. Readers are referred to Fortin and Langevin (2010, 2012) for further details on the model.

ARTEMIS growth predictions are based on a wide range of explanatory variables that can be retrieved from the compilation of forest inventories in Quebec: tree species, harvest occurrence (yes/no), spruce budworm defoliation (yes/no), stem density (tree ha⁻¹), basal area (m²ha⁻¹) and forest type. ARTEMIS also makes use of climatic variables, such as the 1971-2000 mean annual precipitation (mm) and mean annual temperature (°C). These climate variables can be estimated using the BioSIM application (Régnière et al., 2010).

ARTEMIS can predict forest growth either in a deterministic or stochastic fashion (Fortin and Langevin, 2010), depending on the user’s decision. Stochastic predictions rely on the Monte Carlo technique (Rubinstein and Kroese, 2007) and they can be either fully or partially stochastic. The full stochastic mode assumes that all the sub-models are stochastic. In contrast, the partial stochastic mode makes it possible to disable the stochasticity in the selected sub-models. Predictions are generated at tree level, and the plot-level outcome can be obtained by aggregating tree-level predictions. The model was also designed to handle many plots at the same time since the usual input data were expected to come from forest inventories. ARTEMIS has been used for different applications, from simple productivity assessment to the comparison of different silviculture options (e.g., Fortin, 2014; Fortin and Langevin, 2012; Laliberté et al., 2016).

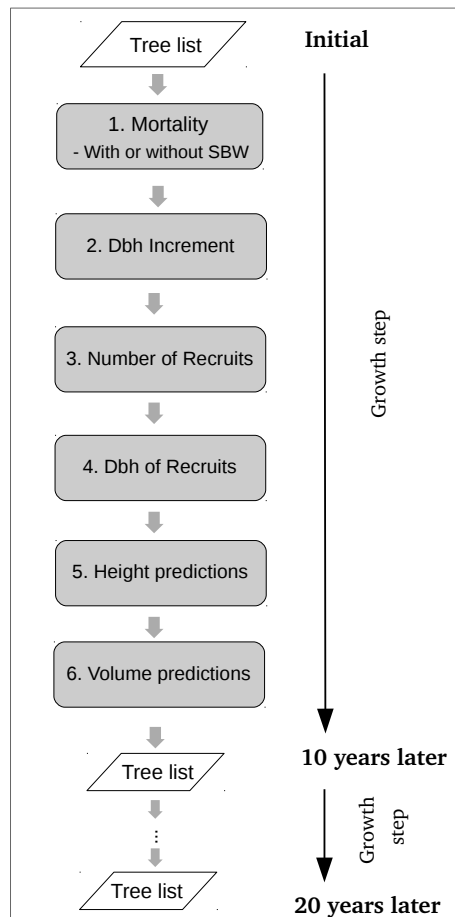


Figure 2.1: Structure of ARTEMIS-2009 simulation system considering an iterative process. Rectangles are for models.

2.2 Input dataset

The input data we used to make large-area predictions also came from the provincial network of permanent plots of Quebec’s Ministry of Forests, Wildlife and Parks (MFWP). Since our focus was to generate regional-level growth predictions, we kept only the measurements from the Bas-Saint-Laurent region, which covers an area of 28,401 km² and two vegetations zones: the northern temperate and the boreal zones (Poirier et al., 2013). The forest composition of the Bas-Saint-Laurent region made it possible to perform predictions for different forest types, since it encompasses broadleaved, mixed and coniferous stands. Broadleaved and mixed forests are mainly composed of sugar maple (*Acer saccharum* Marsh.), yellow birch (*Betula alleghaniensis* Britton), balsam fir (*Abies balsamea* Mill.) and white spruce (*Picea glauca* Voss). Coniferous stands are dominated by balsam fir, white and black spruce (*Picea mariana* Britton) with a minor component of white birch (*Betula papyrifera* Marsh.) and trembling aspen (*Populus tremuloides* Michx.) (Poirier et al., 2013). Moreover, this region has been exploited for timber since the beginning of the 19th century, and for this reason, it is of historical importance for forestry in Quebec Boucher et al. (2009b). This first region-based screening resulted in 1,572 plot measurements that covered the period from 1975 to 2012. We chose the year for which we had the largest sample size, 2003, with a total of 393 plots.

The subsequent screening took the ecotype into account. We used the current ecological classification system used by the MFWP, which is based on the physical characteristics of the site, forest dynamics and its structural elements (Saucier et al., 2009, p. 186-205). Since there were too many ecotypes, we decided to keep three ecotypes that represented the diversity of forest stand composition and for which we had the largest sample sizes. We therefore worked on the following three ecotypes: sugar maple-yellow birch, balsam fir-white birch and balsam fir-white cedar. For convenience, we will refer to these three ecotypes as the broadleaved, mixed and coniferous ecotypes, respectively. The final dataset contained 188 plots.

In each of these 400-m² plots, all trees with DBH equal to or greater than 9.1 cm were tagged for individual monitoring. All explanatory variables required by ARTEMIS were available from the compilation of the input dataset of Quebec’s MFWP. A summary of the dataset and the study area is provided in Table 2.1. The spatial distribution of the plots is shown in Fig. 2.2.

CHAPTER 2. ESTIMATING MODEL- AND SAMPLING-RELATED UNCERTAINTY IN LARGE-AREA GROWTH PREDICTIONS

Table 2.1: Summary of study area and sample plots. Area is the total area occupied for each one of the three ecotypes. The values in parentheses show the range of the variable.

Ecotype	Attribute	Values
Broadleaved	Number of plots	70
	Area ($\times 10^6$ ha)	5.524
	Basal area (m^2ha^{-1})	19.7 (1.0-38.9)
	Stem density (tree ha^{-1})	451 (125-1350)
Mixed	Number of plots	96
	Area ($\times 10^6$ ha)	39.735
	Basal area (m^2ha^{-1})	28.2 (0.2-47.0)
	Stem density (tree ha^{-1})	994 (35-2600)
Coniferous	Number of plots	22
	Area ($\times 10^6$ ha)	6.045
	Basal area (m^2ha^{-1})	30.6 (0.6-46.0)
	Stem density (tree ha^{-1})	800 (66 - 1691)

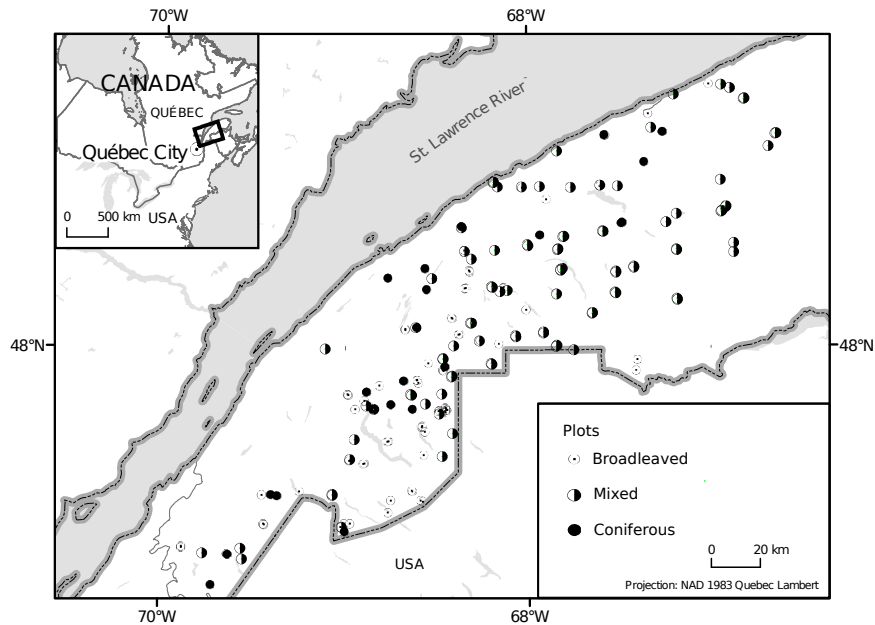


Figure 2.2: Distribution of the permanent plots that compose the network of Quebec's provincial forest inventory in Bas-Saint-Laurent.

2.3 Simulation framework

We chose to quantify the variance of large-area growth predictions in terms of basal area. The simulation framework consisted of 100-year growth predictions running from 2003 to 2103 based on 10,000 Monte Carlo realizations. These predictions were generated for each one of the three aforementioned ecotypes and excluded all exogenous disturbances such as harvesting, pest outbreaks and fires. Additionally, we considered that the mean temperature and the mean precipitation would increase by 2°C and 5%, respectively, over the 21st century.

First, a large-area prediction was generated for each ecotype using the full stochastic mode, i.e., by considering the stochasticity of all ARTEMIS sub-models. The bootstrap variance estimator that is described in the next section made it possible to split the total variance of large-area predictions into a sampling- and a model-related component.

We then generated a series of large-area predictions using the partial stochastic mode, with the aim of decomposing the total variance. Using the above framework, we alternately disabled the stochasticity of the mortality, diameter increment and recruitment sub-models, while keeping the other sub-models in a stochastic mode. The same bootstrap variance estimator was used.

Basal area predictions at the plot level were obtained by aggregating the tree-level predicted basal areas since they were produced by ARTEMIS. The predictions were run on the CAPSIS platform ([Dufour-Kowalski et al., 2012](#)).

2.4 Hybrid estimator

[Fortin et al. \(2018\)](#) proposed a bootstrap variance estimator that can be used in the context of hybrid inference with complex models, such as tree-level growth models. This variance estimator is the one we used and it is briefly described in the next paragraphs. Further details of this estimator are available in [Appendix A](#).

Under the assumption of a simple random sampling design without replacement with equal inclusion probabilities, the sample mean is an unbiased estimator of the mean of the population (μ):

$$\hat{\mu} = \frac{1}{n} \sum_{i \in s} y_i \quad (2.1)$$

where s is the sample, y_i is the basal area in plot i , and n is the sample size. The design variance of this estimator is:

$$\hat{\mathbb{V}}_d(\hat{\mu}) = \left(1 - \frac{n}{N}\right) \frac{\sum_{i \in s} (y_i - \hat{\mu})^2}{n(n-1)} \quad (2.2)$$

where N is the number of units in the population.

Because future basal areas are unknown by definition, we relied on ARTEMIS to obtain basal area predictions ($\hat{y}_i$). Propagating the errors in the parameter estimates, the random effects and the residual errors in the two previous estimators is not straightforward and must be done using the Monte Carlo technique. For a particular realization b , random deviates are generated to account for these different sources of model-related uncertainty and realized plot basal areas are obtained. An estimate of the mean and its variance can then be obtained from Eqs. 2.1 and 2.2. After generating a large number of realizations, the bootstrap estimator of the mean is:

$$\hat{\mu}_{\text{BS}} = \frac{1}{B} \sum_{b=1}^B \hat{\mu}_b \quad (2.3)$$

where $\hat{\mu}_b$ is the sample mean obtained from realization b , and B is the total number of realizations.

Consistent with Fortin et al. (2018), an unbiased bootstrap variance estimator is:

$$\hat{\mathbb{V}}(\hat{\mu}_{\text{BS}}) = \frac{\sum_{b=1}^B (\hat{\mu}_b - \hat{\mu}_{\text{BS}})^2}{B} + 2\hat{\mathbb{V}}_d(\hat{\mu}_{\bar{y}}) - \frac{\sum_{b=1}^B \hat{\mathbb{V}}_d(\hat{\mu}_b)}{B} \quad (2.4)$$

where $\hat{\mathbb{V}}_d(\hat{\mu}_{\bar{y}})$ can be obtained by substituting $\bar{y}_i = \sum_{b=1}^B \frac{y_{i,b}}{B}$ and $\hat{\mu}_{\text{BS}}$ for y_i and $\hat{\mu}$, respectively, in the variance estimator found in Eq. 2.2. The term $\hat{\mathbb{V}}_d(\hat{\mu}_{\bar{y}})$ represents the contribution of the sampling to the total variance. The model contribution can be calculated as $\hat{\mathbb{V}}(\hat{\mu}_{\text{BS}}) - \hat{\mathbb{V}}_d(\hat{\mu}_{\bar{y}})$. Equations 2.3 and 2.4 were used to obtain the model- and sampling-related variance components of our large-area growth predictions.

3 Results & Discussion

This study focused on estimating uncertainties arising from large-area growth predictions based on tree-level growth models. First, the predicted basal areas are shown in Fig. 2.3 considering the full and partial stochastic modes for the three

ecotypes. For all ecotypes, the basal area predictions revealed differences depending on whether they were fully or partially stochastic. These differences were only perceptible after 2050 and slightly increased until 2103.

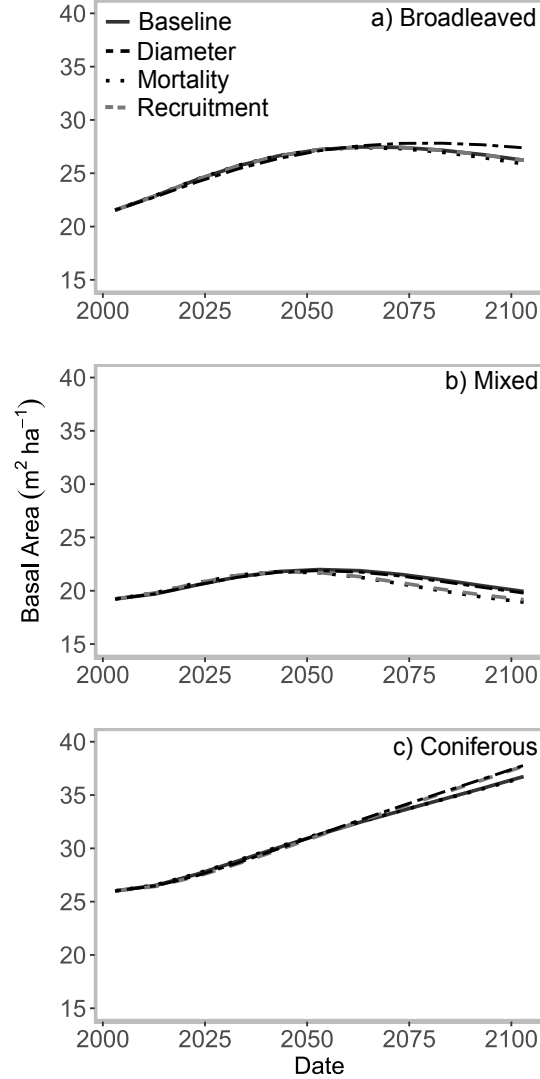


Figure 2.3: Mean predicted basal area (m^2ha^{-1}) illustrated for each ecotype. The growth curves are related with the full stochastic and the partial stochastic predictions (disabling diameter, mortality and recruitment sub-models).

The conifer ecotype presented the highest growth, followed by the broadleaved and the mixed ecotypes. For the broadleaved and mixed ecotypes, the predicted basal areas showed a quadratic pattern with an optimum around 2050 and 2060. While the broadleaved ecotype remained close to the optimum until 2103, the mixed

ecotype slightly declined. For coniferous ecotype, a steady growth was predicted.

The contribution of the model- and sampling-related variance components in the full stochastic predictions are shown in Table 2.2 and Fig. 2.4. The results were consistent for the three ecotypes. The total variance sharply decreased during the first half of the projection and then stabilized until the end. For the broadleaved and the coniferous ecotypes, the total variance at the end of the projection was half that of the beginning. For the mixed ecotype, the variance decreased by 95% over the projection length.

Table 2.2: Model- and sampling-related variance contribution (m^4ha^{-2}) in the full stochastic predictions per ecotype. The relative contribution appears in parentheses.

Ecotype	Year	Model-related	Sampling-related	Total Variance
Broadleaved	2003	0.000 (0.0%)	1.186 (100.0%)	1.186
	2013	0.008 (1.1%)	0.742 (98.9%)	0.750
	2023	0.024 (4.2%)	0.550 (95.8%)	0.574
	2033	0.044 (8.9%)	0.451 (91.1%)	0.495
	2043	0.064 (13.8%)	0.398 (86.2%)	0.461
	2053	0.094 (20.4%)	0.368 (79.6%)	0.462
	2063	0.118 (25.2%)	0.349 (74.8%)	0.467
	2073	0.146 (30.5%)	0.333 (69.5%)	0.479
	2083	0.175 (35.9%)	0.312 (64.1%)	0.486
	2093	0.209 (42.4%)	0.284 (57.6%)	0.492
	2103	0.244 (49.5%)	0.248 (50.5%)	0.492
Mixed	2003	0.000 (0.0%)	2.133 (100.0%)	2.133
	2013	0.003 (0.3%)	1.327 (99.7%)	1.331
	2023	0.007 (0.9%)	0.847 (99.1%)	0.847
	2033	0.012 (2.3%)	0.531 (97.7%)	0.531
	2043	0.020 (5.8%)	0.339 (94.2%)	0.339
	2053	0.023 (9.8%)	0.234 (90.2%)	0.234
	2063	0.028 (15.0%)	0.186 (85.0%)	0.186
	2073	0.030 (18.5%)	0.132 (81.5%)	0.163
	2083	0.036 (23.9%)	0.113 (76.1%)	0.149
	2093	0.035 (27.6%)	0.092 (72.4%)	0.127
	2103	0.038 (36.2%)	0.067 (63.8%)	0.105
Coniferous	2003	0.000 (0.0%)	10.939 (100.0%)	10.939
	2013	0.037 (0.5%)	8.089 (99.5%)	8.126
	2023	0.129 (1.9%)	6.541 (98.1%)	6.670
	2033	0.232 (4.0%)	5.508 (96.0%)	5.740
	2043	0.377 (7.4%)	4.713 (92.6%)	5.090
	2053	0.557 (12.1%)	4.048 (87.9%)	4.606
	2063	0.808 (18.7%)	3.521 (81.3%)	4.329
	2073	1.053 (25.2%)	3.119 (74.8%)	4.173
	2083	1.317 (31.7%)	2.834 (68.3%)	4.151
	2093	1.654 (38.6%)	2.633 (61.4%)	4.287
	2103	2.031 (45.4%)	2.440 (54.6%)	4.471

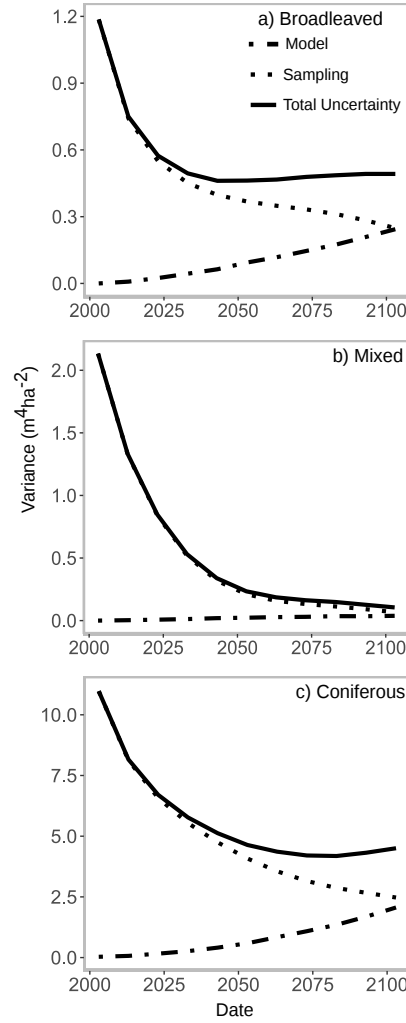


Figure 2.4: Contribution of the model and the sampling to the total variance of the full stochastic predictions in basal area.

The sampling showed the greatest contribution to the total variance. This contribution decreased all along the projection length and more sharply during the first half. It mainly explained the decrease in the total variance. The model-related variance represented a smaller proportion of the total variance, with no contribution at all at the beginning of the projection. However, this variance increased along the projection length until it represented almost half the total variance at the end of the projection. More precisely, the model-related variances represented 49.5%, 45.4% and 36.2% of the 2103 variance in the broadleaved, mixed and coniferous ecotypes, respectively.

As evoked in our first hypothesis, we expected that the model-related variance

would increase over time. Considering the increasing pattern for all ecotypes, there was no evidence that the model-related variance would not represent the greatest share of the prediction variance if the projection length was longer than 100 years. Kangas (1998) also described an increasing model-related variance in long-term volume predictions at the stand level.

The decreasing pattern of the sampling-related variance over time was the main reason for which the total variance also decreased. This is probably due to the convergence of the feed-forward coupled model, which balances itself and reaches a steady state in long-term predictions (Vanclay, 1994, p.46), especially in the absence of disturbances. Moreover, the sampling-related variance defined in Fortin et al. (2018) is actually the designed-based variance of the estimator of the mean calculated from mean predicted basal areas and not from observed basal areas. The model convergence effect yields plot-level predicted values that are more alike over time. However, this does not mean that the forest will be more homogeneous in the future. While the variance of the predicted basal areas decreases over time, the residual error in these predictions increases and this contributes to a greater heterogeneity among the plots.

It could be argued that these trends are not related to the convergence of the model but to the region instead. In order to check this assertion, we carried out some simulations for the administrative region of Mauricie, also located in the province of Quebec. We obtained similar results indicating that the model was truly responsible for the declining sampling-related variances.

An additional insight into the context of hybrid inference is related to the sample size. As shown in Eq. 4.2, the greater the sample size is, the smaller the variance of the mean estimate will be. Thus, sampling variance is affected by sample size. However, Fortin et al. (2016, 2018) showed that the model-related variance is almost insensitive to sample size. This has important implications for large-area predictions. For instance, large-area predictions based on national or regional forest inventories have larger sample sizes than those of this study. As a consequence, the sampling-related variance would probably be smaller, whereas the model-related variance would remain constant. In such conditions, the model-related variance could become the main source of uncertainty, not only in a long-term prediction (i.e., 100 years), but sooner. If we double the sample size for the broadleaved ecotype, for example, as of the year 2050 when sampling-related variance really begins to decrease, model-related variance could become significant at an earlier date.

It is plausible to assume that variance behavior is dependent on the inherent

variability of the ecotypes. Because mixed forests are more heterogeneous, we could expect that growth patterns would show greater variability. When comparing the sampling-related variances across the ecotypes, the sample size may hide the inherent variability. To eliminate the sample size effect, we compared the estimates of the initial population variance, i.e. that of 2003. The mixed and coniferous ecotypes were close, with population variances estimated at 200 and 240 m⁴ha⁻², respectively. Consequently, the larger sampling variance of the coniferous ecotype (Fig. 2.4) was essentially due to the smaller sample size and not so much to a greater population variance. Again, it must be stressed that 2003 is the only year for which the population variance can be derived from the sampling-related variance because this latter accounts for all the variability. For all the other years, the population variance derived from the sampling-related variance represents the variance of the predicted basal areas. The residual error in the predictions must be taken into account in the estimation of the population variance. However, the current implementation of the bootstrap variance estimator in Fortin et al. (2018) does not allow for this estimation and further developments are required.

The variance estimates obtained in the partial stochastic predictions revealed that the mortality sub-model induced the greatest share of variance in basal area predictions for all ecotypes (Fig. 2.5 and Table 2.3). When the stochasticity of this sub-model was disabled, the total variance decreased by 35% to 60% at the end of the projections. Disabling the stochasticity of the diameter increment or recruitment sub-models decreased the variance by 10% to 25% in the broadleaved and coniferous ecotypes. The mixed ecotype showed a different pattern. Disabling the stochasticity of the diameter increment sub-model resulted in slightly higher variances when compared to the full stochastic predictions. More precisely, the total variance was increased by 6% in the last two decades of the projection.

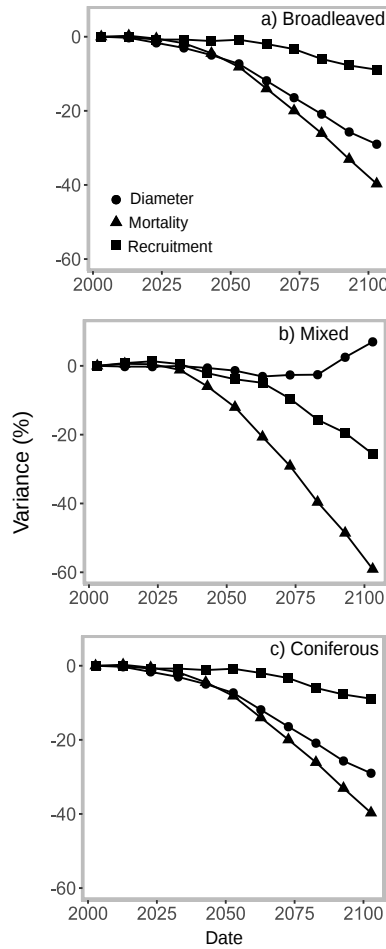


Figure 2.5: Variance contribution of each sub-model that compose the ARTEMIS growth model by ecotype. This contribution is calculated as the relative decrease in the total variance (%) once the stochasticity of the sub-model has been disabled.

In the light of these results, we could only partially validate our second hypothesis. The differences between the different growth components were relatively small for all ecotypes during the first decades, and it is only after 2050 that the departures increased (Fig. 2.5). On the long term, we observed that mortality accounted for the greatest share of variance. The recruitment sub-model, however, did not bring as much variance as we had expected on the long term. Tree mortality is complex and stochastic, leading to large variations over time. Indeed, it remains one of the most difficult growth components to explain (Jr. Hamilton, 1986). Kangas (1998) reported mortality as the main source of error in stand volume predictions. In our study, the contribution of mortality to prediction variance could have been emphasized by the fact that we worked with basal area. Since larger trees are more

likely to die (Vanclay, 1994, p.189), they induce a greater variability in basal area predictions. If we had been working with stem density, this would not have been the case. The same rationale applies to the recruitment. Since ARTEMIS recruits small trees, the impact on basal area predictions is rather small.

In this study, we used a Monte Carlo technique. Although it is computationally intensive and time-consuming, a large-scale application of this method made it possible to propagate errors from multiple sources into ARTEMIS. The differences we observed between the predictions in fully or partially stochastic modes (Fig. 2.3) were also highlighted by some authors. Omitting some errors leads to different predictions (Fortin and Langevin, 2012; Zhou and Buongiorno, 2004). In addition, stochastic models turned out to be more realistic and theoretically superior than deterministic models, e.g., it provides information on the variability of model sources (Fox et al., 2001).

As a matter of fact, most single-tree models are often iterative, i.e., predictions are reinserted into the model in order to obtain long-term growth projections. As a consequence, in deterministic simulations, some tree- and stand-level predictors are no longer observed after the first growth step. They are instead predicted and, thereby, become random variables.

Studies on the uncertainty assessment of large-area forest predictions exist (e.g., Eyvindson and Kangas, 2016; Kangas, 1999; Mäkinen, 2010; Yang et al., 2016). Most studies focus on either model or sampling uncertainty, but rarely on both sources. Furthermore, it seems that model uncertainty has attracted more attention than sampling uncertainty. The fact that sampling uncertainty has remained overlooked in most studies on large-area growth predictions can be related to the complex error propagation in forest growth models.

Some studies have relied on upscaling techniques based on model input aggregation to produce large-area predictions (e.g., Fischer et al., 2014; Kangas, 1996, 1998). The idea is to create an average plot for which growth is then predicted using the model (Wu et al., 2006). Although this approach has some advantages (Bugmann et al., 2000), averaging explanatory variable values over a greater scale with nonlinear models potentially leads to biases, a phenomenon that is known as Jensen's (1906) inequality and that is seldom mentioned in ecological studies (Ruel and Ayres, 1999). In our study, we managed to avoid Jensen's inequality by averaging prediction through the direct extrapolation method (Wu et al., 2006). The method consists of first running the model locally and subsequently averaging model outputs to a greater unit of area (Wu et al., 2006).

Model- and sampling-related variances in large-area estimates of biomass have been predicted using design-based variance estimators and considering the model-related variance through a first-order approximation (e.g., [Fu et al., 2017](#); [Ståhl et al., 2014](#)). However, first-order approximations imply an additional source of error and provide estimates of the statistical moments, not a distribution ([Dettinger and Wilson, 1981](#)). In the presence of a complex system such as nonlinear and time-varying errors, traditional uncertainty methods based on algebra are not adequate for the corrected estimation ([Wilson and Smith, 2013](#)). As an alternative to approximations, [Breidenbach et al. \(2014\)](#) and [McRoberts and Westfall \(2014\)](#) proposed the use of a bootstrap variance estimator. In the context of hybrid inference, this bootstrap estimator can be biased and a corrected estimator has been suggested by [Fortin et al. \(2018\)](#), which is the one we used in this study.

Decomposing the variance of complex models requires advanced techniques ([Xu and Gertner, 2008](#)). In this case study, we found smaller variances in the partial stochastic predictions compared with those in full stochastic mode for one sub-model, namely the diameter increment sub-model. We thought that this could be due to unstable variance estimates. We therefore increased the number of Monte Carlo realizations and tested whether the variance estimates had stabilized. After 10,000 realizations, both sampling- and model-related variance components were found to be stable. The model implementation was also checked. From then on, we considered that these unforeseen results were reliable.

The variance decomposition approach assumes that uncertainties from the different sources are independent. Nevertheless, they may not be independent due to model complexity and highly nonlinear interactions propagated throughout the system. Such uncertainties may occur in an unpredictable manner in the future ([Willems, 2012](#)). Part of the prediction uncertainty in ARTEMIS is also a consequence of its architecture. For instance, if the order of the sub-models had been different from that shown in [Fig. 2.1](#), we would be dealing with different interactions between the sub-models and different predictions of forest growth. Thus, the variance contribution could have been affected.

In a preliminary trial, we also considered the variance decomposition the other way around. We compared deterministic predictions with partial stochastic predictions, where we alternately enabled the stochasticity of one sub-model at a time. Results similar to those presented here were observed and are thus not presented.

In many studies on uncertainty in forestry, the main variable is the volume per hectare. In this study, we chose to estimate the variance in the predictions in basal

area. This was motivated by the fact that we wanted to avoid additional sources of uncertainty from the static sub-models of height-diameter relationship and volume.

CHAPTER 2. ESTIMATING MODEL- AND SAMPLING-RELATED
UNCERTAINTY IN LARGE-AREA GROWTH PREDICTIONS

Table 2.3: Variance values (m^4ha^{-2}) of the full stochastic prediction and the respective disabled growth sub-models represented per ecotype.

Ecotype	Year	Full Stochastic	Recruitment disabled	Mortality disabled	Diameter disabled
Broadleaved	2003	1.186	1.186	1.186	1.186
	2013	0.750	0.755	0.749	0.752
	2023	0.574	0.574	0.561	0.570
	2033	0.495	0.502	0.467	0.482
	2043	0.461	0.479	0.420	0.449
	2053	0.462	0.481	0.397	0.436
	2063	0.467	0.484	0.381	0.450
	2073	0.479	0.484	0.370	0.462
	2083	0.486	0.501	0.353	0.471
	2093	0.492	0.498	0.328	0.480
	2103	0.492	0.483	0.296	0.488
Mixed	2003	2.133	2.133	2.133	2.133
	2013	1.331	1.328	1.339	1.341
	2023	0.847	0.845	0.851	0.858
	2033	0.531	0.530	0.524	0.533
	2043	0.339	0.337	0.319	0.332
	2053	0.234	0.231	0.206	0.225
	2063	0.186	0.180	0.147	0.177
	2073	0.163	0.158	0.115	0.147
	2083	0.149	0.145	0.090	0.125
	2093	0.127	0.130	0.065	0.102
	2103	0.105	0.113	0.043	0.078
Coniferous	2003	10.939	10.939	10.939	10.939
	2013	8.106	8.147	8.126	8.115
	2023	6.591	6.636	6.670	6.618
	2033	5.646	5.639	5.740	5.696
	2043	5.025	4.863	5.090	5.031
	2053	4.560	4.230	4.606	4.568
	2063	4.215	3.723	4.329	4.244
	2073	4.070	3.340	4.173	4.033
	2083	4.039	3.069	4.151	3.902
	2093	4.164	2.870	4.287	3.955
	2103	4.413	2.696	4.471	4.074

4 Conclusions

In this study, we managed to estimate both model- and sampling-related variance of growth predictions at a regional level. This uncertainty estimation required a bootstrap variance estimator for hybrid inference (Fortin et al., 2018) as well as a tree-level growth model that allowed for full stochastic predictions. Additionally, the variances induced by the mortality, diameter increment and recruitment sub-models were individually estimated through a variance decomposition approach. Our study provides insights into how model- and sampling-related variances change along projection length in the context of hybrid inference, as does the growth model component that contributes the most to prediction uncertainty.

The originality of our work includes the temporal perspective in the context of hybrid inference, the variance decomposition information and flexible methods, adaptable to different ecotypes and complex growth models. Sampling-related variance was reported as the most important component of large-area estimates by many authors (e.g., Breidenbach et al., 2014; McRoberts and Westfall, 2014; Ståhl et al., 2014). In our study, the time scale turned out to be an original angle since it indicated that the extent of sampling- and model-related variance is a function of time. Sampling error is the most important source of variance in short-term predictions, while in long-term predictions, the model contribution is as important as that of the sampling. Consequently, efforts should focus on increasing the sampling size and developing an efficient sampling design in order to decrease the overall variance. Similar recommendations were made by McRoberts and Westfall (2014). In long-term predictions, the model should be the target of attention. By decomposing the variance, we were able to point out that the mortality sub-model should first and foremost be improved since it has a greater contribution to the prediction variance.

As a future perspective, we recognize the importance of natural or anthropogenic disturbances, which were deliberately omitted in our predictions for the sake of simplicity. Further investigations should be conducted to take these into account. Finally, since forest management planning is a long-term process, we conclude that efforts on modeling techniques should be considered in order to reduce total uncertainty in growth predictions.

Chapter 3

Using survival analysis to predict the harvesting of forest stands in Quebec, Canada

L.C. Melo, R. Schneider, R. Manso, J-P.
Saucier, M. Fortin

Contents

1	Introduction	46
2	Material and methods	48
2.1	Dataset	48
2.2	Statistical development	51
2.3	Model evaluation	54
3	Results	56
4	Discussion	60

1 Introduction

Understanding forest dynamics is essential to the development of long-term strategies to ensure the sustainable use of natural resources and conservation. Sustainability is closely dependent on forest management practices. The concept of sustainability has evolved to encompass a more complex understanding of the diversity of values, resources and ecological services that forests represent (Sample, 2004). For a long time, forest management plans only focused on timber production and economic concerns in many countries around the world (Dong et al., 2015). Today, forest managers address new challenges, taking environmental protection and social demands linked to economical interests into consideration (Hernandez et al., 2014). The inclusion of these new factors has led to increasingly complex decision-making procedures, which, in turn, require decision models to meet and support this new approach (Hernandez et al., 2014).

Forest management planning consists of making decisions about what to do in the future given what is known about the past and the present. This type of planning is based on a hierarchical structure that consists of both strategic and tactical decisions. One of the goals of the strategic level is to help with decisions related to sustainable harvests that take large areas, long-term time horizons, silvicultural policies and legislation issues into account (Martell et al., 1998). At the tactical level, the objective is to conduct harvest operations on a short-term horizon for specific areas (Martell et al., 1998). Since these strategic decisions are often the result of a modeling process for long-term horizons, we have decided to focus on this aspect. e Different methodologies have been developed to address the complex choice of which trees or plots should be harvested in long-term growth projections (Fortin, 2014). They can be grouped into two categories: harvest algorithms and harvest models. Mostly of available harvest algorithms are based on user-defined rules that can be used to specify which trees should be harvested considering stand characteristics such as the diameter classes and stand basal area (e.g., Miina, 1996; Pukkala et al., 1998). Depending on the objective function, the application of harvest algorithms may be a difficult task. When the strategy includes uneven-aged and mixed forests, this complexity is even greater (Thurnher et al., 2011).

An alternative to harvest algorithms has been the development of statistical models that predict the tree- or plot-level probability of being harvested (e.g., Antón-Fernández and Astrup, 2012; Fortin et al., 2013; Sterba et al., 2000; Thurnher et al., 2011). These models all use logistic regressions. However, the logistic regression has

some limitations in a context of harvest planning, especially when the exact date of the harvest is unknown, which is the case in many inventories. The harvest model is then fitted on time intervals bounded by two successive measurements, the first one providing the initial conditions, whereas the second one gives the information as to whether or not the plot or the tree has been harvested in the meantime. If the sampling intensity changes over time or if the intervals overlap, the maximum likelihood estimates of the model parameters may be biased ([Firth, 1993](#); [McCullagh, 2008](#)). Additionally, uneven time intervals are an issue for the logistic regression since this technique does not effectively use temporal information ([Wang et al., 2013](#)).

A different option could be the use of survival analysis. Survival analysis comprises many methods to deal with lifetime data, which is defined as the time elapsed until the change of state of an individual (e.g., death), referred to as “event”. In our context, this event would be the harvesting of a plot or some trees in that plot. Survival analyses are especially meant to predict the probability that a harvest occurs within a given interval of time, although the exact moment is unknown. These data are commonly referred to as interval-censored data ([Lawless, 2003](#), p. 10). As mentioned by [Wang et al. \(2013\)](#), survival analysis makes better use of temporal information, minimizing uncertainties in the estimation process.

Beyond the issue of interval-censored data, it is worth mentioning that the values of some potential explanatory variables may change within the time intervals, a typical issue that is known as time-varying explanatory variables and that can also be tackled using survival analysis ([Allison, 1982](#)). Time-varying variables may be richer since they provide additional information that better explain the results.

The flexibility of survival analysis methods also makes it possible to accommodate multiple levels of explanatory variables, which are defined at the geographic scale at which they are measured. In addition to the usual plot-level explanatory variables, regional- and national-level variables may also play a relevant role. [Antón-Fernández and Astrup \(2012\)](#) observed that the changes in economy or legislation were a limiting factor in their study. Likewise, [Thurnher et al. \(2011\)](#) also reported that changes in global management practices had considerable effects on predictions from harvest models. The survival analysis approach could prove to be more effective by considering the effects of variables, in contrast to the traditional logistic regression.

In forestry, some authors have already used the survival analysis approach to fit models of individual tree mortality (e.g., [Manso et al., 2015](#); [Rose et al., 2006](#)).

However, to the best of our knowledge, the survival analysis approach has not yet been applied in harvest modeling. Thus, our main objective was to test its applicability in a context of plot-level harvest occurrence. On the basis of three hypotheses, we attempted to determine whether or not survival analysis can overcome some limitations reported in the forest harvest literature. More specifically, our hypotheses were: (1) The survival analysis approach provides unbiased predictions of harvest occurrence (Hypothesis 1); (2) Regional variables have an effect on harvest occurrence (Hypothesis 2) and; (3) Time-varying covariates contribute to increasing the model likelihood (Hypothesis 3). To test these three hypotheses, we used the data from a network of permanent plots in the province of Quebec, Canada, as a real-world case study.

2 Material and methods

2.1 Dataset

The dataset we used in this study was taken from the network of permanent plots of Quebec’s Ministry of Forests, Wildlife and Parks (MFWP). This permanent plot network covered the two vegetation zones in Quebec, which also reflected major climate subdivisions: the northern temperate zone, dominated by broadleaved and mixed stands, and the boreal zone, in which coniferous stands were predominant.

The network comprised 12,570 randomly located sample plots that were established in the early 1970s to monitor forest dynamics (Fig. 3.1). The sampling intensity was set to one plot every 26 km² in the temperate zone and to one plot every 259 km² in the boreal zone, mainly because broadleaved and mixed stands were more heterogeneous (MFFP, 2014). The plots were supposed to be measured every 10 years but, for practical reasons, the intervals between the measurements were uneven in most cases, which resulted in two to six measurements per plot. In these 400-m² circular plots, trees with diameter at breast height (dbh: 1.3 m in height) equal to or greater than 9.1 cm were tagged for individual monitoring. At each measurement, their diameter and status, i.e., alive, dead or harvested, were recorded.

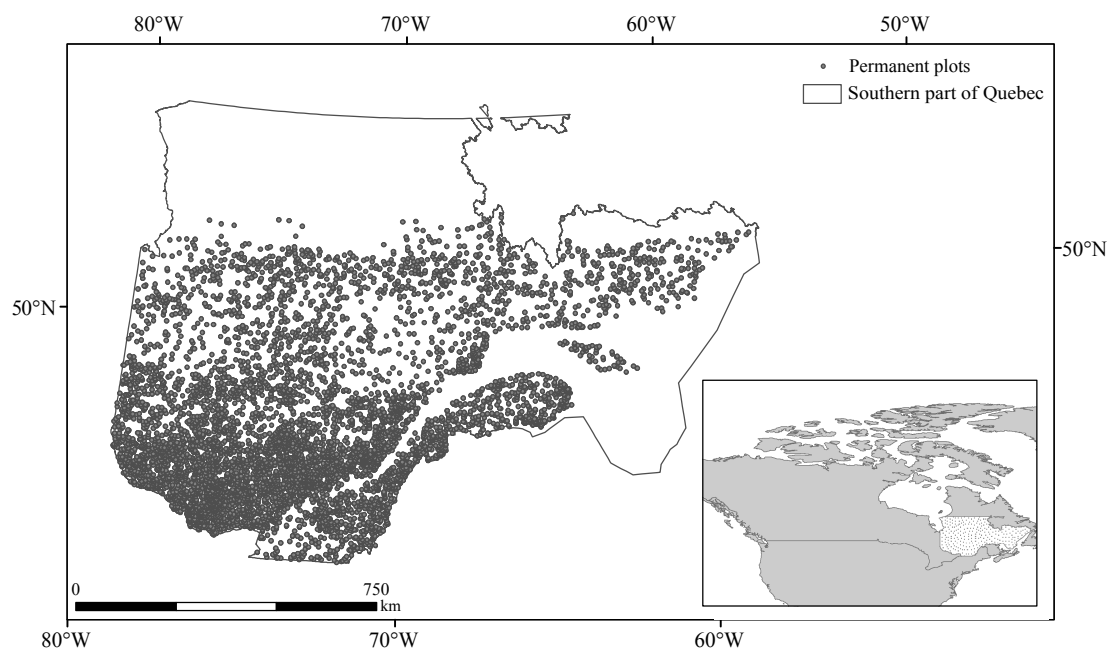


Figure 3.1: Distribution of the 12,570 permanent plots that compose the network of Quebec’s provincial forest inventory covering the northern temperate zone and the boreal zone.

Our objective was to model the harvest occurrence at the plot level. For this purpose, the measurements of the same plot were grouped into non-overlapping successive intervals. For example, if a plot had three measurements, measurements one and two were paired to create a first interval, whereas measurements two and three were used to create a second interval. In these non-overlapping successive intervals, we considered that the first measurement represented the initial conditions of the plot. The second measurement that composed the interval provided the response variable, namely whether the plot had been harvested or not and so on. A plot was considered to be harvested if at least one tree had been harvested during the interval.

Different plot-level variables were tested in the model as possible predictors of the harvest occurrence. These variables were: basal area (m^2ha^{-1}); stem density (stem ha^{-1}); length of the interval (years); and slope class, which discriminated between accessible and non-accessible forest stands.

The ecological type, which is a classification based on the physical characteristics of the site, forest dynamics and structural elements of the vegetation (Saucier et al., 2009, p. 186-205), was also available for each plot. However, there were too many to

consider them all in a model and, consequently, they were grouped into three classes based on forest dynamics: B, M and C, for broadleaved, mixed and coniferous ecological types, respectively. These forest dynamics classes actually represented the forest composition of the late successional stage. The dominant forest dynamics class was coniferous, followed by the mixed and the broadleaved classes ([Boulay, 2015](#)).

Regarding our second hypothesis on the effect of regional variables, we investigated some political and economic factors relevant to our case study. For instance, the American countervailing duty on Canadian wood was a result of actions by the U.S. government to restrict Canadian lumber products since 1982 and that had a considerable impact on the harvest regime and trade in the early 2000s ([Descôteaux and Martin, 2009](#); [Rahman and Devadoss, 2002](#)). The currency exchange rate could also be an important factor because it affects the profitability of Canadian lumber wood exports towards the United States. In Canada, the drop in the market in the early 2000s, with a sharp decrease in 2005, could have impacted the harvest occurrence.

At the regional level, the annual allowable cut volumes (AAC) are major drivers in forest planning in Quebec and it could be reasonably assumed that they would affect the plot-level harvest occurrence. Since 1988, Quebec's forest law prescribes a mandatory calculation of a maximum volume that can be sustainably harvested. The allowable cut volumes are recalculated every 5 years. For further details about this calculation, the reader is referred to [MFFP \(2003, 2013\)](#). Regional values of AAC were retrieved from the annual reports on the state of Quebec's forest ([Boulay, 2013, 2015](#); [Parent, 1988, 1990, 1992, 1993, 1994, 1996, 1999, 2009, 2010](#); [Parent and Fortin, 2000, 2002, 2003, 2004, 2005, 2006, 2007, 2008](#); [Parent et al., 2012](#)).

This value is usually expressed as the potential in terms of cubic meters in the different administrative regions. However, since the regions do not share the same forested area, we reported these values on a hectare basis (m^3ha^{-1}). Given the anticipated incidence of the AAC, we kept only the intervals for which this variable was available. After this screening, 12,596 intervals out of 29,013 were kept, corresponding to the 1988-2014 period. A summary of this subset is shown in [Tables 3.1 and 3.2](#).

Table 3.1: Summary of some plot-level characteristics in the intervals between 1988 and 2014 for each dynamics class (B: broadleaved; M: mixed; C: coniferous; n : number of intervals; the range of the variables appears in parentheses). A particular interval belongs to a period when its final date falls into the range.

Dynamics class	Period	n	Basal area (m ² ha ⁻¹)	Stem density (stems ha ⁻¹)	Proportion harvested
B	1988-1998	296	21.8 (0.2-63.8)	599 (25-2200)	20.9%
	1999-2009	1442	19.1 (1.9-45.9)	596 (25-1900)	26.6%
	2010-2014	290	21.0 (0.2-45.9)	604 (25-1700)	29.0%
M	1988-1998	450	18.0 (0.2-63.3)	746 (25-3575)	7.8 %
	1999-2009	3623	20.8 (0.2-62.6)	793 (25-3975)	13.4%
	2010-2014	1147	19.5 (0.19-62.6)	872 (25-3975)	12.2%
C	1988-1998	426	16.3 (0.2-64.1)	818 (25-3050)	7.5%
	1999-2009	3059	14.5 (0.2-56.3)	870 (25-3450)	10.8%
	2010-2014	1863	15.2 (0.2-57.8)	881 (25-3750)	7.9%

2.2 Statistical development

The underlying concepts of the survival analysis approach are fully described in [Allison \(1982\)](#) and [Lawless \(2003\)](#). In order to illustrate the approach, let T be a non-negative random variable that represents the time at which the event of interest precisely occurred. The cumulative distribution function (cdf) given by $F(t) = \Pr[T \leq t]$ and the probability density function (pdf) designated by $f(t)$ are mathematical specifications of the distribution of T . Given the cdf, the survival function, which represents the probability that the event has not yet occurred at time t , can be defined as:

$$S(t) = \Pr[T > t] = 1 - F(t) \quad (3.1)$$

The hazard function $h(t)$ is the instantaneous rate of occurrence at time t , provided that the event has not yet occurred, such that $h(t) = \frac{f(t)}{S(t)}$. Analogously, the cumulative hazard function is defined as the sum of all the hazards accumulated since the beginning of the experiment:

$$H(t) = \int_0^t h(z)dz. \quad (3.2)$$

The survival function can be linked to the hazard function as follows:

$$S(t) = e^{-\int_0^t h(z)dz} \quad (3.3)$$

*CHAPTER 3. USING SURVIVAL ANALYSIS TO PREDICT THE
HARVESTING OF FOREST STANDS IN QUEBEC, CANADA*

Table 3.2: Annual allowable cut volumes (AAC) by administrative regions since 1988. The values of AAC and area represent the mean during the period, including public and private lands.

Region	Period	Area (km ²)	AAC (m ³ ha ⁻¹)
Bas-Saint-Laurent - Gaspésie	1988-1998	34, 530	3.62 (2.82-4.08)
	1999-2009	34 775	3.07 (2.98-3.98)
	2010-2014	34 740	3.40 (3.38-3.45)
Saguenay-Lac-Saint-Jean	1988-1998	81, 769	2.96 (2.22-3.53)
	1999-2009	76 667	2.63 (2.45-3.01)
	2010-2014	74 996	2.43 (2.34-2.78)
Québec	1988-1998	26, 117	3.35 (2.87- 3.94)
	1999-2009	26 502	3.20 (3.00-3.38)
	2010-2014	26 577	2.75 (2.67-3.07)
Trois-Rivières	1988-1998	35, 047	2.97 (2.56-3.20)
	1999-2009	34 835	3.12 (2.77-3.61)
	2010-2014	34 847	3.12 (2.90-4.01)
Estrie	1988-1998	7,733	3.11 (2.61-3.27)
	1999-2009	7670	3.13 (2.87-3.70)
	2010-2014	7668	3.04 (2.82-3.93)
Montréal	1988-1998	30, 239	3.57 (2.98-4.96)
	1999-2009	31 426	2.93 (2.71-3.55)
	2010-2014	26 577	2.70 (2.54-3.34)
Outaouais	1988-1998	28, 931	3.92 (2.84-4.12)
	1999-2009	27 245	3.80 (3.18-4.02)
	2010-2014	27 240	3.63 (3.45-4.34)
Abitibi-Témiscamingue	1988-1998	51, 950	2.92 (2.02-3.44)
	1999-2009	48 518	3.33 (2.01-3.69)
	2010-2014	48 595	2.99 (2.69-3.11)
Côte-Nord	1988-1998	153, 174	2.03 (1.30-4.23)
	1999-2009	103 367	2.34 (1.89-3.73)
	2010-2014	80 293	1.96 (1.96-1.99)

In the discrete case, the survival function can be alternatively defined as (e.g., [Lawless, 2003](#), p. 10):

$$S(t) = e^{-\sum_{z=0}^t h(z)} \quad (3.4)$$

There are many different formulations for hazard functions ([Willet and Singer, 1993](#)), including the proportional hazard function. This formulation assumes that some covariates increase or decrease a common hazard that is referred to as the baseline. If we define i , j , and k as the indices of the clusters of plots, plots and intervals, respectively, then a proportional hazard function can be expressed as follows:

$$h_{ijk}(t) = h_0(t, \mathbf{g}_{ijk}\boldsymbol{\gamma})e^{\mathbf{x}_{ijk}\boldsymbol{\beta}} \quad (3.5)$$

where h_0 is the baseline, $\mathbf{x}_{ijk}$ and $\mathbf{g}_{ijk}$ are row vectors of explanatory variables, and $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ are two column vectors of unknown parameters. Note that vectors $\mathbf{g}_{ijk}$ and $\boldsymbol{\gamma}$ define the baseline, whereas vectors $\mathbf{x}_{ijk}$ and $\boldsymbol{\beta}$ define the proportional effect on the baseline. If some covariates in $\mathbf{g}_{ijk}$ change within a given interval of time and their values are known for each time t , then it is possible to use these updated values in the hazard function, which is actually one method for considering time-varying covariates in the model.

Considering the context of our study, we did not know the exact time of harvest occurrence. We only knew that a particular plot had survived until the beginning of the interval t_1 and that it had been harvested or not at the end of the interval, namely at time t_2 . The harvest probability is thus conditional on the survival up to the beginning of the interval. Let us define a binary variable y_{ijk} that adopts the value 1 if the plot has been harvested during interval k in cluster i and plot j , and 0 otherwise. The marginal probability of an occurrence during such an interval is thus given by:

$$\Pr_{\text{marg}} = \Pr[t_1 < T \leq t_2] = S_{ijk}(t_1) - S_{ijk}(t_2) \quad (3.6)$$

The second issue is that we do not know the whole plot history. The main consequence of this is that actual calendar times corresponding to t_1 and t_2 are also unknown. This concept is known as left-truncation (Lawless, 2003, p. 68). To take this left-truncation into account, the probability in Eq. ?? has to be conditional on the survival up to t_1 :

$$\Pr[y_{ijk} = 1] = \Pr[t_1 < T \leq t_2 | T > t_1] = \frac{S_{ijk}(t_1) - S_{ijk}(t_2)}{S_{ijk}(t_1)} \quad (3.7)$$

Incorporating the hazard function 3.5 into Eq. 3.7 yields:

$$\Pr[y_{ijk} = 1] = 1 - \frac{e^{-\sum_{z=0}^{t_2} h(z)}}{e^{-\sum_{z=0}^{t_1} h(z)}} = 1 - e^{\mathbf{x}_{ijk}\boldsymbol{\beta} - \sum_{z=t_1}^{t_2} h_0(z, \mathbf{g}_{ijk}\boldsymbol{\gamma})} \equiv f(\mathbf{x}_{ijk}, \mathbf{g}_{ijk}, \boldsymbol{\beta}, \boldsymbol{\gamma}) \quad (3.8)$$

The likelihood of the parameters with respect to the whole dataset will then be:

$$L(\boldsymbol{\beta}, \boldsymbol{\gamma} \mid \mathbf{X}, \mathbf{G}, \mathbf{y}) = \prod_i \prod_j \prod_k f(\mathbf{x}_{ijk}, \mathbf{g}_{ijk}, \boldsymbol{\beta}, \boldsymbol{\gamma})^{y_{ijk}} (1 - f(\mathbf{x}_{ijk}, \mathbf{g}_{ijk}, \boldsymbol{\beta}, \boldsymbol{\gamma}))^{1-y_{ijk}} \quad (3.9)$$

where the rows of $\mathbf{X}$, $\mathbf{G}$ and $\mathbf{y}$ contain the $\mathbf{x}_{ijk}$, $\mathbf{g}_{ijk}$ and y_{ijk} , respectively. All parameters can be estimated by maximizing the likelihood function with respect to γ and β .

It could be reasonably assumed that spatial correlation existed in the plot-level occurrence of harvest. These spatial correlations were treated as random effects on the baseline. Some preliminary analyses showed that the harvest occurrence was positively correlated in space and it could be reasonably assumed that these correlations decreased with increasing distances between two plots. To assess the correlation pattern, we first defined plot clusters at different scales.

These clusters were set according to the hierarchical mapping system in Quebec, which is actually composed of four nested grids of different resolutions. One grid square covered approximately 256,000 km², 16,000 km², 1,000 km² and 250 km², depending on the resolution of the grid. These squares defined by the different grids are referred to as map elements and we used the indices 1 to 4 for the highest to the lowest resolution.

The cruise line could be interpreted as a fifth resolution, representing a local scale. It referred to a straight line transverse to a slope in which data are collected. According to the sampling design, it was established to contain two plots separated by 425 m. The approximate area represented by a cruise line was 0.57 km², which actually represented the highest resolution. Cruise line and map elements at the different scales were successively tested in the model baseline as grouping factors for cluster random effects.

From a statistical perspective, the specification of random effects in the model makes the parameter estimation more complex. Function 3.8 has to be redefined as $f(\mathbf{x}_{ijk}, \mathbf{g}_{ijk}, \beta, \gamma, u_i)$ where u_i is the cluster random effect. Since these random effects are unobserved, the model likelihood has to be marginalized over the distribution of the random effects. We omit the mathematical developments, but the reader is referred to [Pinheiro and Bates \(2000, p. 62\)](#) for further details about the estimation. Such marginal likelihood functions can be optimized using the NLMIXED procedure available in SAS ([SAS, 2015, Ch. 82](#)). An example of the implementation of this type of model is shown in [Appendix B](#).

2.3 Model evaluation

The evaluation of the models was carried out taking variable significance and the Akaike Information Criterion (AIC) into account. Once the model is fitted, the

expectation of the harvest probability conditional on the cluster random effect u_i is given by function $f(\mathbf{x}_{ijk}, \mathbf{g}_{ijk}, \boldsymbol{\beta}, \boldsymbol{\gamma}, u_i)$. In practice, these random effects are unobserved. The prediction based on the fixed effects only, i.e., $u_i = 0$, are actually not population-averaged predictions when the model is nonlinear (McCulloch et al., 2008, p.190).

In fact, the population-averaged prediction is the conditional expectation of y_{ijk} marginalized over the distribution of u_i , which is given by the following integral:

$$E[y_{ijk} \mid \mathbf{x}_{ijk}, \mathbf{g}_{ijk}] = \int f(\mathbf{x}_{ijk}, \mathbf{g}_{ijk}, \boldsymbol{\beta}, \boldsymbol{\gamma}, u_i) \text{pdf}(u_i, \sigma_u^2) du_i$$

where $\text{pdf}(u_i, \sigma_u^2)$ is the probability density function of a normal distribution with mean 0 and variance σ_u^2 .

This integral has no closed-form solution. However, the Gauss-Hermite quadrature can be used to approximate integrals of functions by a weighted average of the integrand over a pre-determined grid (Pinheiro and Bates, 1995). An example of this technique can be found in Fortin (2013). All tests were run on these population-averaged predictions.

In order to evaluate the model, we ran a 10-fold cross-validation. The intervals were split into 10 groups and we fitted the model 10 times, with one of the 10 groups successively omitted. At the end, each group had its own predictions. Then, the Hosmer-Lemeshow test was carried out. This test evaluates the difference between the fitted probabilities and the observed probabilities divided by the number of observations in the groups. The test provides a χ^2 statistic under the null hypothesis that there is no lack of fit. We also obtained the Receiving Operation Characteristics (ROC) by plotting sensitivity (rate of correctly classified events) against specificity (rate of correctly classified non-events) for different cut-offs (critical probability of event occurrence to discriminate events from non-events). The ROC allowed us to calculate the Area Under the Curve (AUC).

In order to illustrate model predictions, we generated an average plot based on the values observed in the dataset. To show the effect of a particular variable, its value was changed within the range observed in the dataset, while the other variables were kept constant. Hence, we simulated short-term forecasts (10-year intervals) of harvest probabilities. These outputs were then used to highlight model strengths and weaknesses.

3 Results

The fitting was carried out from a simpler model to the most complex, following the steps shown in Table 3.3. The first model was fitted with some of the plot-level variables under the assumption of even time intervals. The model was then refined by adding different plot and regional variables as well as the interval length. The best model with no consideration for spatial correlation issues included basal area, stem density, slope classes, dynamics classes and AAC. It had an AIC of 8691.2 and an AUC of 0.7582.

Table 3.3: The goodness-of-fit for the tested models.

Model type	Model designation	AIC	Δ AIC	Hosmer-Lemeshow		AUC
				χ^2	p value	
Fixed-effect	1	9180.0	0	9.56584	0.29683	0.7015
	2	9097.3	82.7	13.3898	0.099122	0.7141
	3	8691.2	405.8	5.93247	0.6548	0.7582
Mixed-effects	4	8497.3	193.9	8.57676	0.37925	0.7581
	5	8516.5	19.2	17.0651	0.029438	0.7578
	6	8577.4	60.9	18.7245	0.016404	0.7767
	7	8619.4	42	9.85327	0.27548	0.7644
	8	8383.6	235.8	10.9099	0.20686	0.7583

- 1: Model considering stand variables, except slope classes. The hazard is constant over time.
2: Model 1 including slope classes.
3: Model 2 with a variation of hazard over time.
4: Model 3 with random effect of map element level 1
5: Model 3 with random effect of map element level 2
6: Model 3 with random effect of map element level 3
7: Model 3 with random effect of map element level 4
8: Model 3 with random effect of cruise line
 Δ AIC: Difference between the AIC value of the successive models.

The fit statistics revealed a substantial improvement in the maximum likelihood when random effects were included. The best fit was obtained with a cruise line random effect, which yielded an AIC of 8383.6 and an AUC of 0.7583. The final model had the following form:

$$\Pr(y_{ijk} = 1) = 1 - e^{-e^{\beta_1 \ln(BA_{ijk}) + \beta_2 N_{ijk} + \beta_3 s + \beta_4 v} \sum_{z=t_1}^{t_2} e^{\gamma_0 + \gamma_1 AAC_z + u_i}} \quad (3.10)$$

where BA_{ijk} is the basal area (m^2ha^{-1}) of plot j in cluster i during interval k ; N is the stem density (stem ha^{-1}); s is the index for the six different slope classes; v is the index for the three dynamics classes; AAC_z is the regional annual allowable cut

volume (m^3ha^{-1}) for year z ; and u_i is the cruise line random effect. For this model, the Hosmer-Lemeshow test indicated no evidence of under- or overestimation and no violation of the model assumptions at $p < 0.01$. The predicted probabilities and the observed proportions for the 10 groups were plotted (Fig. 3.2).

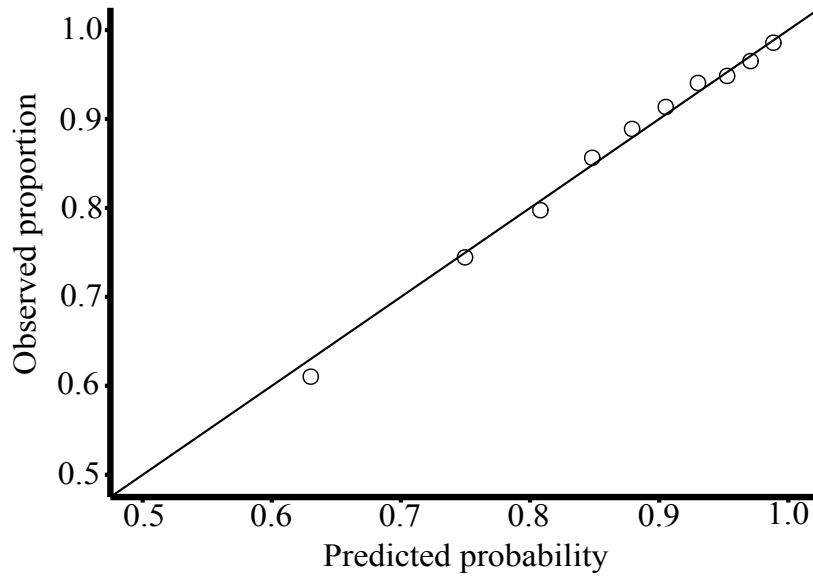


Figure 3.2: Observed proportions and predicted probabilities of events evaluated by the Hosmer-Lemeshow test.

The resulting parameter estimates are shown in Table 3.4. The sign of the estimates indicates the influence on the predicted probabilities, with a positive sign indicating an increase in the probability of harvest and *vice versa*. In this respect, basal area increased the harvest probability. In contrast, stem density and slope classes decreased it. In addition, dynamics classes also showed a significant effect. Greater harvest probabilities were associated with the broadleaved dynamics class, followed by the mixed and the coniferous classes. Our results also indicated that reductions of the AAC induced a decrease in the harvest probabilities.

The Bas-Saint-Laurent administrative region was chosen to illustrate the predictions of the final model based on an average plot (Fig. 3.3). AAC values from 2000 to 2010, which ranged from 1.49 to 1.97 m^3ha^{-1} , were used for the baseline. For the lowest value of AAC, the corresponding annual harvest probability was 0.013. For the highest AAC, this probability was 0.022. The effect of the AAC drop in 2006 could be detected in the cumulative probability since the increase for that year was

*CHAPTER 3. USING SURVIVAL ANALYSIS TO PREDICT THE
HARVESTING OF FOREST STANDS IN QUEBEC, CANADA*

Table 3.4: Maximum likelihood estimates of the parameters in the final model (Model 8) with their associated standard errors and approximate t -values.

Effect	Class	Parameter	Estimate	Standard Error	t -values	Pr > $ t $
Proportional part						
Basal area		β_1	1.0584	0.06413	16.5	< .0001
Stem density		β_2	-6.8×10^{-4}	8.9×10^{-5}	-7.6	< .0001
Slope class ¹	4% to 8%	$\beta_{3,1}$	-0.135	0.08943	-1.51	0.1313
	9% to 15%	$\beta_{3,2}$	-0.2086	0.09108	-2.29	0.0221
	16% to 30%	$\beta_{3,3}$	-0.4445	0.09978	-4.45	< .0001
	31% to 40%	$\beta_{3,4}$	-0.7786	0.1694	-4.6	< .0001
	> 41%	$\beta_{3,5}$	-1.3747	0.2565	-5.36	< .0001
Dynamics class	broadleaved	$\beta_{4,1}$	0.4171	0.09437	4.42	< .0001
	mixed	$\beta_{4,2}$	0.1214	0.07356	1.65	< .0001
Baseline						
Intercept		γ_0	-9.206	0.223	-41.29	< .0001
AAC		γ_1	1.3442	0.08023	16.75	< .0001
Variance random effect		σ_u^2	1.4332	0.1234	11.62	< .0001

¹: Inclination

smaller than those of previous years. The AAC increase in 2010 had exactly the opposite effect on the cumulative probability. Over the 10-year interval, that plot had a predicted probability of being harvested of close to 0.18.

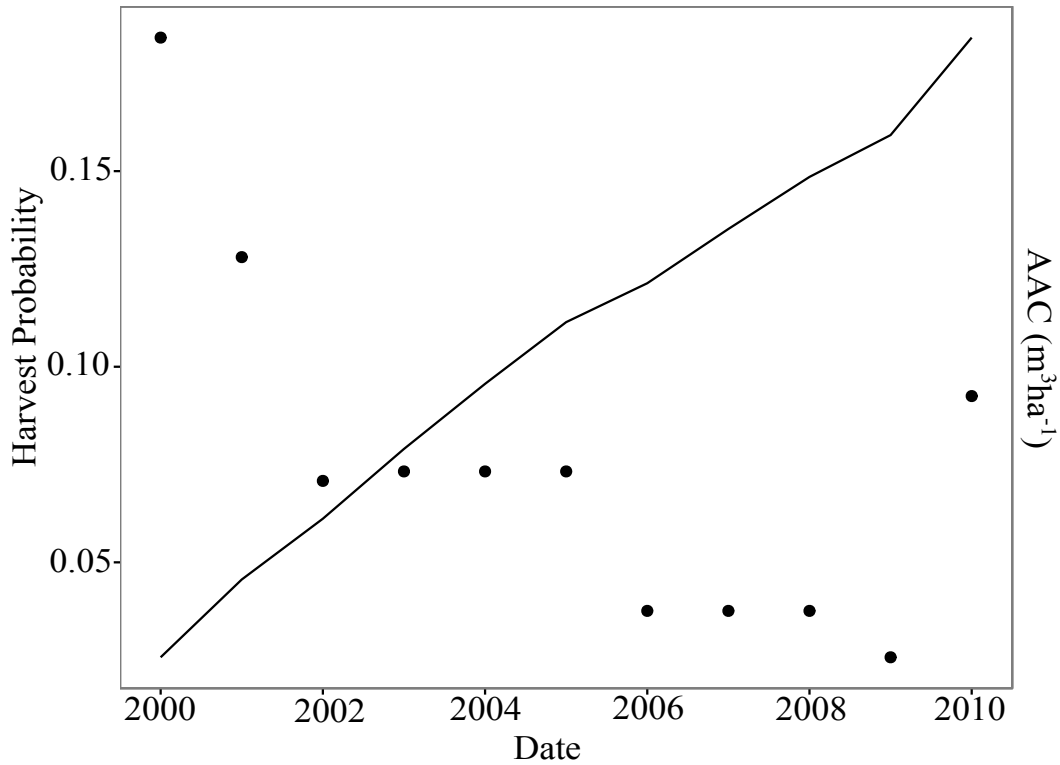


Figure 3.3: Simulation of the AAC effects on harvest probability for Bas-Saint-Laurent. The dots represent the values of AAC. The line represents the cumulative harvest probability. The reference values are: BA = $18 \text{ m}^2\text{ha}^{-1}$; slope = 4% to 8%; stem density = $725 \text{ m}^2\text{ha}^{-1}$; and dynamics class = mixed.

The effects of the other covariates on the predictions is shown in Fig. 3.4. Basal area had a considerable influence over the predicted harvest probabilities. The probability associated with the largest basal area was twice that of the lowest basal area, with values decreasing from 0.22 to 0.10 (Fig. 3.4a). An increase in stem density decreased the harvest probability, with a difference of about 0.06 between the highest and the lowest densities (Fig. 3.4b). The predicted probabilities for the broadleaved dynamics class were almost two times higher than those of the coniferous class, which exhibited the lowest probability with 0.16. For the mixed class, the estimated harvest probability was 0.20 (Fig. 3.4c). The steeper the slope, the lower the probabilities of harvest were (Fig. 3.4d).

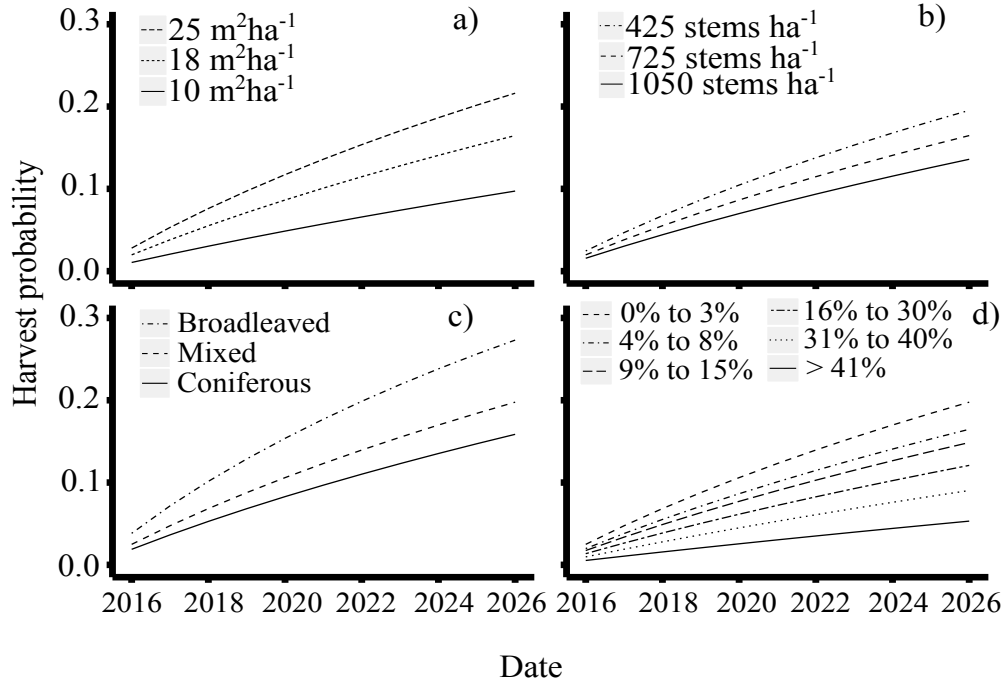


Figure 3.4: Effect of the model variables on the predicted harvest probabilities for a 10-year interval. (a) basal area effect; (b) stem density effect; (c) dynamics class effect; (d) slope class effect. The reference values are: BA = 18 m²ha⁻¹; slope = 4% to 8%; stem density = 725 m²ha⁻¹; and dynamics class = mixed.

The different random effect specifications based on the map elements showed similar variance estimates. A small difference was found for the highest resolution, which had a variance estimate of 0.53. The variance of the cruise line random effect was larger by far than those of the map elements, with an estimated value of 1.43. For all tested models, neither the countervailing duty nor the exchange rates had significant contributions and, consequently, they were discarded from the model.

4 Discussion

In this study we used plot- and regional-level variables combined with spatial random effects to predict the plot-level harvest probability using a survival analysis approach. Our results showed the potential of survival analysis to provide annual predictions of harvest occurrence using a hazard function. Additionally, this approach allowed

us to overcome limitations reported in the harvest literature, such as taking changes in economic conditions or legislation into account ([Antón-Fernández and Astrup, 2012](#)), as well as changes in management practices and length of intervals ([Thurnher et al., 2011](#)). Through this hazard function, a survival analysis can deal with uneven intervals and time-varying regional variables. It is definitely an easier and cheaper alternative than investing in additional field measurements. Thus, our modeling approach and the nature of the variables make this study an original contribution.

Previous works on harvest predictions were primarily focused on harvest algorithms (e.g., [Heaps, 2015](#); [Liu et al., 2006](#); [Nelson and Finn, 1990](#)). More recently, the logistic regression was introduced as an effective method for modeling this kind of variable. These studies considered tree- and plot-level variables (e.g., [Antón-Fernández and Astrup, 2012](#); [Thurnher et al., 2011](#)) but economic or political covariates acting at the regional level had been ignored so far. Moreover, the logistic regression shows some shortcomings when dealing with interval-censored data. It suffers from sample bias ([Wang et al., 2013](#)) if the sampling intensity changes over time. Also, it does not account for within-interval time-varying covariates. The market and the management strategy may change during the intervals and these changes cannot be taken into account in the logistic regression. Our results showed that the survival analysis approach provides enough flexibility to allow for these regional variables and can generate unbiased predictions of harvest occurrence, confirming our first and second hypotheses.

The hazard function used to generate plot harvest probabilities has two components: the baseline and the proportional part ([Lawless, 2003](#), p. 36). We assumed that the covariates affecting the baseline were at the regional level and were allowed to vary within the intervals on an annual basis, whereas the plot-level variables belonged to the exponential part of the hazard function. In order to test our second and third hypotheses, we modeled the baseline as a function of the AAC, which allows the hazard to change over time within a given interval. This is the time-dependent part of the hazard. In contrast, the plot-level variables were considered as having a constant impact on the hazard over the interval.

In strategic planning that includes the assessment of sustainability, forestry actions have long-lasting effects and cover larger areas when compared to tactical planning ([Kangas et al., 2001](#)). In order to accomplish the strategic level decisions, our harvest model should be coupled with a growth model, which would allow the manager to account for business-as-usual scenarios and management variations. We therefore recommend our model as capable of predicting the harvesting probability

on a long-term planning horizon. The general principles underlying our methods may be applicable to a wide range of forest types.

Model evaluation based on cross-validation has been criticized by some authors (e.g., [Kozak and Kozak, 2003](#)). It turns out that the cross-validation does not provide much more information than what could be obtained with the whole dataset. We do agree with this statement. However, there is another benefit of using cross-validation, which is the detection of overfitted models. From our experience, it often happens that complex mixed-effects models converge with the whole dataset but not during the cross-validation. This is essentially due to a particular fixed effect that is strongly related to a few plots. This happened with a preliminary version of the model. When this fixed effect was eliminated, the model then converged for all the runs of the cross-validation. This process ensured that the final model we obtained was the most parsimonious.

With respect to the tested covariates, we managed to quantify the importance and intensities of their effects. As expected, an increase in basal area induced an increase in harvest probability. In Quebec, dense stands are often submitted to different intensities of harvest to meet various objectives, e.g., to increase timber yield, to promote natural succession, to regulate competition levels and to improve stand vigor ([Majcen et al., 2003](#)). Our plot-level harvest model reflected these management practices. Additionally, the results are also in accordance with [Sterba et al. \(2000\)](#) who found a similar relationship between the harvest probability and the natural logarithm of basal area in a clearcut regime in Austria.

The harvest probabilities decreased with increasing stem densities. This result can be explained by the fact that lower stem densities are found in mature stands ([Bose et al., 2013](#)). The influence of maturity on harvest probability was also reported in the study of [Antón-Fernández and Astrup \(2012\)](#). Moreover, for a given basal area, more stems means trees with lower individual volumes, which is less interesting from a commercial point of view.

Lower harvest probabilities were found in the plots with steeper slopes, showing the obvious limitations related to harvest operations. In Quebec, harvesting on slopes of more than 40% is unusual and is authorized only for preventive measures ([MNR, 1998](#)).

The model predictions also confirmed that the dynamics class was a determining factor in the harvest occurrence. The broadleaved class showed greater probabilities of harvest. This pattern seems to be consistent with the management strategy in Quebec. Broadleaved species are more abundant in Southern Quebec, whereas

coniferous species are dominant in the northern part of the commercial forest. There are obvious differences in the harvest intensity depending on the dynamics class. Until the 1990s, clearcutting and diameter-limit cutting were common practices but were found to be unsuitable for many vegetation types, which led to an evolution of management practices ([Archambault et al., 1998](#)). Since the 1990s, the most common treatment in the stands of the broadleaved dynamics class is selection cutting based on a cycle of 20 to 25 years ([MFFP, 2013](#)). For coniferous stands, the final cut at the end of the rotation, which lasts 50 years at best, is usually carried out, implying a harvesting of 90 to 99% of the merchantable volume ([MFFP, 2013](#)).

As regards regional-level variables, we also tested the exchange rate and countervailing duty as regional variables in the model. However, neither of them proved to have a significant effect on the harvest occurrence. This could be due to confounding effects with the AAC. As a matter of fact, the AAC dropped in the mid 2000s coincides with a drop in the exchange rate and the implementation of the countervailing duty.

Spatial correlation has been extensively studied as well as its effects. One of the benefits of considering spatial correlation is the better consistency in parameter estimates ([Bhat and Sener, 2009](#)). It also plays an important role in harvest scheduling since different spatial patterns influence some economic and conservation objectives ([BoWang and Gadow, 2006](#)).

In our model, we used random effects to account for these spatial correlations. The cluster random effect resulted in an improvement of the model fit. This random effect can be interpreted as a trend for plots belonging to the same cruise line to jointly deviate from the expected value. This is not surprising if we compare the area covered by a cruise line with the area of cut blocks. The correlations between the map elements at different resolutions were smaller but non-negligible. These correlations over larger areas could be explained by management practices. Until recently, forest companies planned their annual operation by sectors, which means that the stands of these sectors were more likely to be harvested at the same time as those of other sectors. If the harvest model was to be used for tactical planning, it could eventually be improved by integrating the geographical information of the tactical plans, which is not really possible in strategic planning over large areas.

Multiple random effects with the cruise lines nested in the map elements were also tested. In spite of our efforts to fit this model, convergence problems were a limiting aspect. An alternative approach to random effects would be the direct modeling of spatial correlations. The inclusion of a copula could be an example.

Thus, further improvements in terms of the described limitations still remain as possible points of investigation.

In conclusion, this study proposed an alternative approach to model plot-level harvest occurrence. Our approach combined survival analysis with time-varying variables, taking uneven intervals and censored observation data into account.

Regarding our hypotheses, we can conclude that the survival analysis approach can be used to obtain unbiased predictions of plot-level harvest occurrence. Secondly, the inclusion of variables at the regional level, such as AAC, had a significant effect on harvest probabilities. The inclusion of such a time-varying explanatory variable in the model significantly increased its likelihood, which was the third hypothesis. In the context of forest management, such a multi-level approach could be useful in strategic planning. Coupled to a growth model, it makes it possible to generate large-area growth predictions for forests under management.

Chapter 4

The effect of natural and anthropogenic disturbances on the uncertainty of growth forecasts

L.C. Melo, R. Schneider, M. Fortin

Contents

1	Introduction	66
2	Material and methods	68
2.1	ARTEMIS growth model	68
2.2	Uncertainty estimation	70
2.3	Study area and dataset	71
2.4	Forecasting	73
3	Results	74
4	Discussion	78
5	Conclusions	82

1 Introduction

The influence of disturbances on forest ecosystems has been given special attention over the last decades due to anticipated environmental changes (Turner, 2010). The disturbance regime plays a dominant role in shaping forest dynamics, such as influencing structure and composition (Bouchard and Pothier, 2011), as well as determining temporal and spatial patterns (Didion et al., 2007). This important role has triggered efforts to include disturbances in forest management plans (Daniel et al., 2017) and in growth forecasts (Turner, 2010).

Natural disturbances along with anthropogenic activities are the major agents that shape the landscape. In European forests, the most common large-scale disturbances are storms, followed by fires and insect outbreaks (Schelhaas et al., 2003). In the Canadian boreal forests, the natural disturbances are mainly fires and insect outbreaks such as forest tent caterpillar (*Malacosoma disstria*), jack pine budworm (*Choristoneura pinus*) and spruce budworm (*Choristoneura fumiferana* (Clem.); SBW) (Brandt et al., 2013). They are known to affect up to millions of hectares (Gauthier et al., 2015). Among other anthropogenic disturbances such as agriculture and roads, harvesting activities have become a key driver of forest dynamics (Venier et al., 2014). Approximately 40% of the boreal forest is under management and these managed areas are more disturbed by harvesting than by natural disturbances (Venier et al., 2014).

Including disturbances in growth models is necessary to properly simulate forest dynamics over large areas (Seidl et al., 2011). It generates more realistic growth forecasts, which are of great interest in practical forestry, ecology and climate change mitigation activities (Ståhl et al., 2016). Nevertheless, the process of simulating forest growth over large areas implies propagating errors. Such uncertainties arise from the model and the sampling. Model errors are a result of parameters estimation and the structure of the model, among others factors (Refsgaard et al., 2007; Walker et al., 2003). Sampling errors are due to the upscaling of forest variables to a higher level (Breidenbach et al., 2014).

Because large-area growth forecasts are based on both the model and the sampling design (Ståhl et al., 2016), they represent a typical case of what is known as hybrid inference (Corona et al., 2014). This context of hybrid inference arises when: (i) the variable of interest, such as growth, is not observed but predicted using a model; and (ii) the explanatory variables of the model are observed in the sample only and not throughout the entire population (Fortin et al., 2016; McRoberts and

Westfall, 2014). This requires special estimators that account for both sources of uncertainty (McRoberts and Westfall, 2016). Hybrid estimators applied with forest growth models propagate errors from the plot to the regional or national level and they represent an implementation of an upscaling method known as the direct extrapolation method (Wu et al., 2006).

Uncertainty assessment of forest growth forecasts has been studied by many authors (e.g., Horemans et al., 2016; Kangas, 1999; Xu and Gertner, 2008). In some cases, natural and anthropogenic disturbances were taken into account. However, to the best of our knowledge, the uncertainty they induce in large-area growth forecasts has not been fully addressed. In the very few cases where the uncertainty due to the disturbances was addressed, it was either for anthropogenic or natural disturbances, but not for both. Moreover, the uncertainty that stemmed from the sampling was overlooked (e.g., Bergeron et al., 2017). This conjecture motivated this study.

The impact of a particular type of disturbance on growth forecast uncertainty can be assumed to be closely related to its spatial and temporal patterns. Some authors who studied these patterns with regard to population dynamics found that different populations from the same species present a synchronicity (e.g., Williams and Liebhold, 2000), i.e., one population is likely to occur simultaneously with other populations. This correlated population fluctuation has been detected in various taxa and over many spatial scales (Liebhold et al., 2004). For insects acting as a disturbance in forests, the synchronicity of outbreaks can impact growth forecasts due to the coincident changes in forest attributes. At the landscape level, the extent of these changes can lead to variability.

Given the influence of disturbances on forests, our main objective was to take them into consideration in large-area growth forecasts and to estimate their contribution in terms of uncertainty. To do this, we worked on a real-world case study: the administrative region of Bas-Saint-Laurent, Quebec, Canada. More specifically, we focused on spruce budworm outbreaks and harvesting, which are the two major disturbances in that region. As a natural component, spruce budworm outbreak is of concern since it occurs on a large scale with return intervals of a few decades, and has a great impact on forest productivity (Boulanger et al., 2012). The region is presently facing a SBW outbreak.

Motivated by the spatial synchrony theory, we first hypothesized that spruce budworm outbreaks have a greater impact than harvesting on the uncertainty of large-area growth forecasts. Furthermore, in the context of hybrid inference, our second hypothesis was that disturbances induced more uncertainty than the sam-

pling in these forecasts. Data from the provincial network of permanent plots in Quebec, Canada, and the ARTEMIS growth model (Fortin and Langevin, 2012) were used to generate those large-area growth forecasts for the Bas-Saint-Laurent region.

2 Material and methods

2.1 ARTEMIS growth model

We worked with the distance-independent individual-based growth model, ARTEMIS (Fortin and Langevin, 2010, 2012). The 2009 version of the model was designed and fitted using the network of permanent plots of the Quebec provincial forest inventory. This network is composed of 12,570 sample plots established in the 1970s (MFFP, 2009, 2015c). Briefly, ARTEMIS is composed of seven sub-models, with five of them being dynamic and the other two static (Fig. 4.1). The dynamic sub-models predict the harvest probability, the mortality probability, the diameter increment, the number of recruits and the diameter of these recruits, respectively. The harvest module works in two steps. It first predicts the probability that a particular plot is harvested considering the elapsed time, the annual allowable cut volumes (MFFP, 2003) and some plot-level variables such as the slope inclination (Melo et al., 2017). Whenever a plot is harvested, the second part of this sub-model predicts the probability that trees within this plot are harvested given a particular management regime (Fortin, 2014). The two static sub-models make it possible to predict tree height and commercial volume based on other characteristics of the trees and the plot. More details are available in Fortin and Langevin (2010, 2012).

The model uses 10-year growth steps. The output of a given step is re-inserted in the model in order to obtain forecasts over longer time periods. Users may run growth simulations in a deterministic or stochastic fashion. The stochastic mode relies on the Monte Carlo technique (Rubinstein and Kroese, 2007). In such a mode, three types of errors are simulated: the errors in the parameter estimates, the plot or interval random effects and the residual errors. The model provides tree-level predictions. Plot-level predictions are obtained through the aggregation of the predictions at the tree level.

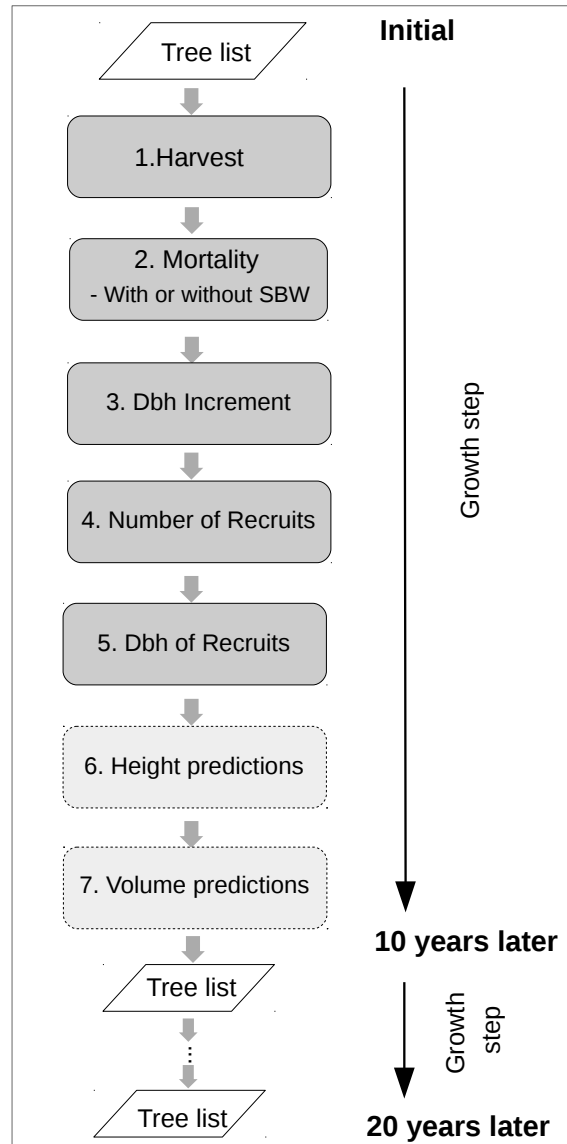


Figure 4.1: Flowchart of ARTEMIS-2009 considering its iterative process. Dark gray boxes represent the dynamic sub-models. Dotted light gray boxes are the static sub-models.

ARTEMIS relies on a large array of explanatory variables. At the plot and tree levels, the model considers the tree species, harvest occurrence (yes/no), stem density (tree ha⁻¹) and basal area (m²ha⁻¹), which is the sum of the cross-section areas at 1.3 m in height. Basal area is a variable that is commonly used by foresters as a competition index (Stadt et al., 2007). These aforementioned variables are considered as endogenous variables, i.e., influenced by factors within the system (Rao and Toutenburg, 1995). ARTEMIS also considers the potential vegetation. The potential vegetation refers to the composition at a late successional stage (Grondin

et al., 2009). Thirty-two potential types of vegetation exist in the province of Quebec (Saucier et al., 2015) and ARTEMIS was designed to work with the 25 most frequent ones. Each potential vegetation type was modeled individually, thus resulting in 25 versions of the model (Fortin and Langevin, 2010).

ARTEMIS also takes the impact of spruce budworm defoliation into consideration. It does not predict the occurrence of an outbreak, but whenever an outbreak occurs, the mortality sub-model adapts through a binary variable. The predicted probabilities of mortality for spruce (*Picea* spp.) and balsam fir (*Abies balsamea* (L.) Mill.) then increase. The recurrence of spruce budworm outbreaks is set by the user and represents an exogenous variable. An annual probability of occurrence can be derived from the recurrence, under the assumption that the occurrences are independent in time.

Finally, mean annual precipitation (mm) and temperature (°C) are predictors in ARTEMIS. They are both entries in the mortality and recruitment sub-models. The diameter increment sub-model considers only precipitation, whereas the sub-models predicting recruit diameter and tree height consider only the mean temperature. These variables are estimated using BioSIM, a software program that predicts climate variables for a particular geographical location based on the data of the nearest climate stations (Régnière et al., 2010).

2.2 Uncertainty estimation

A key step in our study is to estimate uncertainty in the predicted volumes in the context of hybrid inference, that is, inference relying on both the model and the sampling design (Corona et al., 2014). To do this, we used a hybrid variance estimator based on the bootstrap method (Fortin et al., 2018). The mathematical developments behind the estimator are presented in the Appendix A.

If we consider a design of simple random sampling without replacement with even inclusion probabilities, an unbiased estimator of the population mean is the sample mean:

$$\hat{\mu} = \frac{1}{n} \sum_{i \in s} y_i \quad (4.1)$$

where s is the sample, y_i is the variable of interest in plot i and n is the sample size.

The design variance of this estimator is, in turn, estimated as:

$$\hat{V}(\hat{\mu}) = \left(1 - \frac{n}{N}\right) \frac{\sum_{i \in s} (y_i - \hat{\mu})^2}{n(n-1)} \quad (4.2)$$

where N is the number of units in the population.

When y_i is not available, a model can be used to obtain a prediction that is denoted as $\hat{y}_i$. Substituting $\hat{y}_i$ for y_i in Eq. 4.1 still yields an unbiased estimator of the mean provided that the model has no lack of fit. However, the adaptation of the variance estimator requires further developments, namely propagating errors from different sources within the model. This error propagation can be carried out using the Monte Carlo technique (Rubinstein and Kroese, 2007). The technique consists of drawing random deviates to account for the errors in the parameter estimates, the random effects and the residual errors. A single simulation based on a particular set of deviates provides a realization of the estimated mean and the estimated variance shown in Eqs. 4.1 and 4.2. After a great number of realizations, the bootstrap estimator of the mean is:

$$\hat{\mu}_{BS} = \frac{1}{B} \sum_{b=1}^B \hat{\mu}_b \quad (4.3)$$

where $\hat{\mu}_b$ is the sample mean obtained from realization b , and B is the total number of realizations.

Consistent with Fortin et al. (2018), an unbiased bootstrap variance estimator is:

$$\hat{V}(\hat{\mu}_{BS}) = \frac{\sum_{b=1}^B (\hat{\mu}_b - \hat{\mu}_{BS})^2}{B} + 2\hat{V}_d(\hat{\mu}_{\bar{y}}) - \frac{\sum_{b=1}^B \hat{V}_d(\hat{\mu}_b)}{B} \quad (4.4)$$

where $\hat{V}_d(\hat{\mu}_{\bar{y}})$ can be obtained by substituting $\bar{y}_i = \sum_{b=1}^B \frac{y_{i,b}}{B}$ and $\hat{\mu}_{BS}$ for y_i and $\hat{\mu}$, respectively, in the variance estimator found in Eq. 4.2. The sampling contribution to the total variance is obtained through $\hat{V}_d(\hat{\mu}_{\bar{y}})$, while the model contribution is calculated as $\hat{V}(\hat{\mu}_{BS}) - \hat{V}_d(\hat{\mu}_{\bar{y}})$.

2.3 Study area and dataset

The inventory data were a subset of the provincial network of permanent plots of Quebec's Ministry of Forests, Wildlife and Parks (MFWP). We limited our analysis to the regional level. Thus, our dataset included only the plot measurements from the Bas-Saint-Laurent administrative region. Covering a surface of 22,185 km², the forest composition is representative of broadleaved, mixed and coniferous vegetation.

The dominant species in this region are sugar maple (*Acer saccharum* Marsh.), yellow birch (*Betula alleghaniensis* Britton), balsam fir, white spruce (*Picea glauca* Voss) and black spruce (*Picea mariana* Britton). The plots are located in five different ecological regions, the result of a classification established by the MFWP to characterize the composition and dynamics of the vegetation (MFWP, 2016). The plot distribution in the different ecological regions is shown in Fig. 4.2.

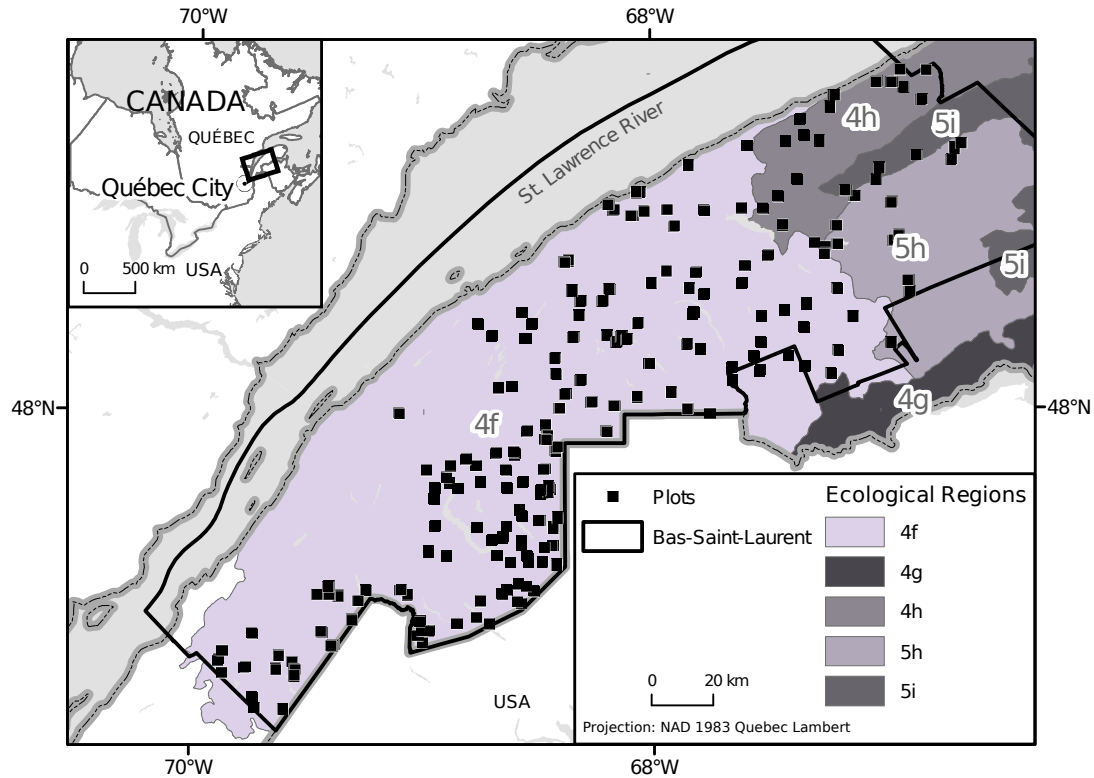


Figure 4.2: Distribution of the 393 permanent plots in Bas-Saint-Laurent. The plots are located in the ecological regions classified according to the MFWP: Appalachian Hills (4f); Baie des Chaleurs Coastline (4g); Gaspé Coastline (4h); Mountains of Gaspé Peninsula (5h); Highlands of Gaspé Peninsula (5i).

Historically, the Bas-Saint-Laurent region has been subject to anthropogenic and natural disturbances. The region was affected by severe SBW outbreaks during the last century, which, as a consequence, triggered salvage cutting (Boulanger and Arseneault, 2004). Moreover, silvicultural practices have deeply transformed forest composition (Boucher et al., 2009b). The current regional forest planning guidelines prescribe silvicultural practices adapted to the different forest types (Gagnon et al., 2015): selection cutting in shade-tolerant broadleaved forests; irregular and regular

shelterwood cutting in mixed stands; and harvest with protection of regeneration and soils for most coniferous forests and some mixed stands.

There were 514 permanent plots located in the Bas-Saint-Laurent region, each with an area of 400 m² in area. Due to logistic constraints, these plots were not all measured at the same time. It was actually in 2003 when the greatest number of plots were measured. Thus, our final dataset consisted of 393 plots measured in 2003. In these plots, all trees with diameter at breast height (DBH, 1.3 m in height) equal to or greater than 9.1 cm were tagged for individual monitoring. A summary of the dataset is provided in Table 4.1.

Table 4.1: Summary of 393 plots in the dataset. Attributes were broken down for the most abundant species. The minimum and maximum values are shown in parentheses.

Plot-level	n	Basal area (m ² ha ⁻¹)	Stem density (trees ha ⁻¹)
Sugar Maple	393	2.5 (0 - 29.7)	68 (0 - 1266)
Red Maple	393	1.2 (0 - 13.7)	57 (0 - 900)
Balsam fir	393	5.1 (0 - 42.6)	256 (0 - 2350)
White spruce	393	1.5 (0 - 28.2)	61 (0 - 1850)
Black spruce	393	0.7 (0 - 21.5)	48 (0 - 1800)
White birch	393	1.6 (0 - 18.53)	95 (0 - 875)
All species	393	17.8 (0 - 61.2)	778 (25 - 2550)
Tree-level	n	DBH (cm)	Height (m)
Sugar Maple	1,124	20.4 (9.1 - 78.3)	17.5 (7.2 - 27.1)
Red Maple	901	14.9 (9.1 - 68.3)	14.6 (9.2 - 24.1)
Balsam fir	4,072	15.1 (9.1 - 49.3)	13.3 (3.8 - 24.5)
White spruce	983	17.0 (9.1 - 54.5)	13.0 (5.0 - 24.5)
Black spruce	762	13.3 (9.1 - 32.5)	10.7 (5.2 - 21.0)
White birch	1,492	14.2 (9.1 - 42.8)	13.2 (6.5 - 19.8)
All species	12,451	16.2 (9.1 - 98.8)	14.5 (3.8 - 27.7)

2.4 Forecasting

We built a framework to predict forest growth for the Bas-Saint-Laurent region, taking harvest and SBW outbreak effects into account. These projections were carried out for a period of 100 years (2003-2103), considering a 2°C temperature increase and a 5% precipitation increase over the 21st century, which roughly corresponds to the representative concentration pathway (RCP) 4.5 provided by the IPCC (2013, p. 1335). Since the initial year of our forecasts was 2003, the observed disturbance

history up to 2018 is known. We first configured the forecasts to update the plot status, i.e., to take the management of the first two decades (2003-2023) and the SBW outbreak initiated in 2013 into account. Once this initial condition was established, we then tested four scenarios: (i) a baseline scenario with no disturbances; (ii) a harvest scenario, in which plots were harvested according to the current level of annual cut volume allowance and the prescribed treatments for each forest type; (iii) a SBW scenario, in which we considered an average outbreak recurrence of once every 35 years, according to [Boulanger and Arseneault \(2004\)](#); and (iv) a scenario including both harvest and SBW outbreaks, structured as in the second and third scenarios but acting simultaneously here. The simulations were run on the CAPSIS platform ([Dufour-Kowalski et al., 2012](#)). We ran a total of 10,000 Monte Carlo realizations to account for the variability induced by disturbances for each scenario. It is worth mentioning that the forecasts include stochasticity from disturbances, as well as from the parameter estimates, the random effects and the residual errors.

ARTEMIS provides tree-level predictions. The individual predicted volumes were aggregated at the plot level. The hybrid bootstrap estimators shown in Eqs. 4.3 and 4.4 were then used to perform the upscaling to the regional level, as proposed in the direct extrapolation method ([Wu et al., 2006](#)).

3 Results

Long-term volume forecasts for the Bas-Saint-Laurent region are shown with their confidence intervals in Fig. 4.3. The baseline scenario, in which no disturbance was considered, resulted in an increasing volume that reached $220 \text{ m}^3\text{ha}^{-1}$ in 2103 (Fig. 4.3a). When the disturbances were taken into account, similar growth patterns were observed but predicted volumes were smaller. More precisely, when SBW outbreaks were included in the forecasts, the volume in 2103 was $20 \text{ m}^3\text{ha}^{-1}$ lower than that of the baseline (Fig. 4.3b). When considering harvest occurrence only, volume for the same period was $45 \text{ m}^3\text{ha}^{-1}$ lower compared to the baseline (Fig. 4.3a). For the scenario in which harvest and spruce budworm outbreaks occurred simultaneously, predicted volumes for 2103 were $60 \text{ m}^3\text{ha}^{-1}$ smaller than the baseline (Fig. 4.3b).

The confidence intervals provide an assessment as to how future growth can vary in the Bas-Saint-Laurent region under disturbances. Considering the predicted lower limit of the interval for the scenario considering both SBW outbreaks and harvesting (Fig. 4.3b), it is very unlikely that the mean volume per hectare at the end of the

21st century will be smaller than what it was in 2003.

Growth forecasts were characterized by a total variance that increased over time (Fig. 4.4). The magnitude of the increase was dependent on the scenario. The increase was steep for these scenarios including SBW (Fig. 4.4c,d). At the end of the time horizon, the variances of these two scenarios were greater than $100 \text{ m}^6\text{ha}^{-2}$, whereas the variances of the scenarios without SBW were smaller than $50 \text{ m}^6\text{ha}^{-2}$.

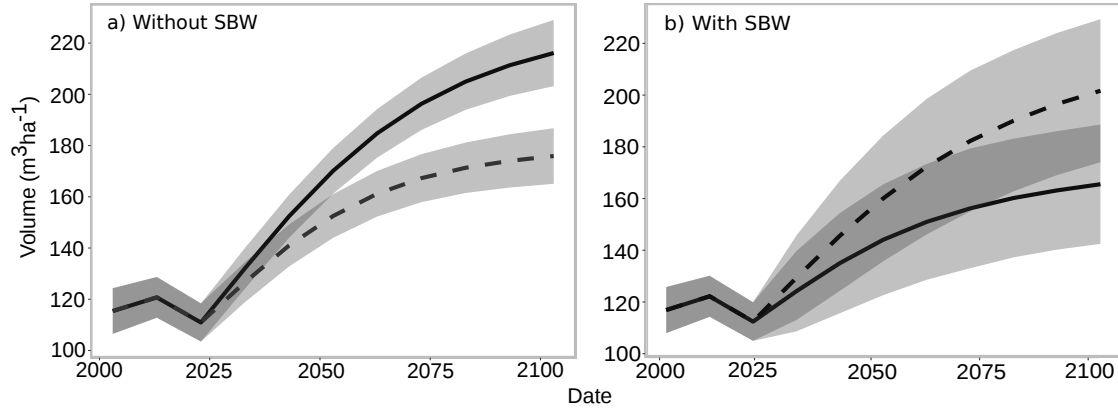


Figure 4.3: Mean predicted volumes (m^3ha^{-1}) and their 0.95 confidence interval for the Bas-Saint-Laurent region. The confidence intervals rely on the assumption of a normal distribution. The solid line represents the scenarios without harvesting, while the dashed line represents the scenario including harvesting.

The scenarios including harvesting were characterized by smaller total variances. The total variance in the scenario considering harvesting only reached $30 \text{ m}^6\text{ha}^{-2}$ in 2103, whereas it was estimated at $43 \text{ m}^6\text{ha}^{-2}$ for the baseline scenario (Fig. 4.4a,b). Likewise, in the scenario including simultaneously harvest and SBW, total variance was estimated at $138 \text{ m}^6\text{ha}^{-2}$, compared to $199 \text{ m}^6\text{ha}^{-2}$ in the scenario with SBW only (Fig. 4.4c,d).

The sampling-related variances showed the same pattern across the scenarios. The variance slightly decreased in the first two decades and then remained stable or slowly increased over time. Model-related variances increased over time. Our results revealed two main trends. In the first case, for the baseline and harvest scenarios, the model-related variances increased steadily (Fig. 4.4a,b). The second case was related to the inclusion of SBW outbreaks, which already greatly inflated the model-related variance on the short-term (Fig. 4.4c,d). In both cases, the model-related variance mainly explained the patterns observed in the total variance. The absolute and relative variances related to all four scenarios are presented in Table 4.2.

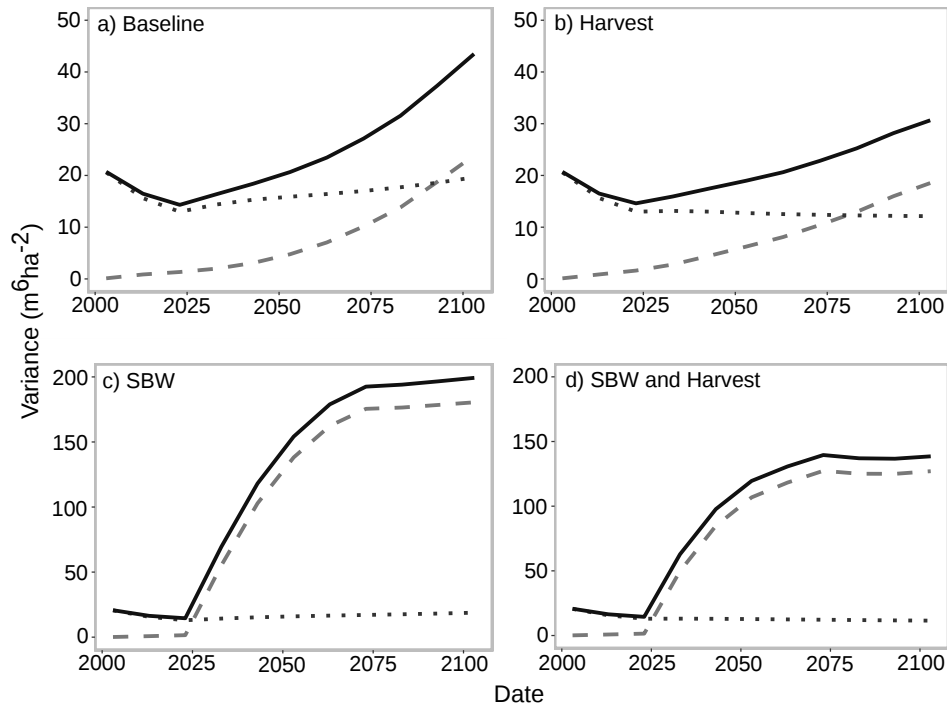


Figure 4.4: Model, sampling and total variances illustrated per growth scenario. Model contribution: gray dashed line; Sampling contribution: dark gray dotted line; Total variance: black solid line.

CHAPTER 4. THE EFFECT OF NATURAL AND ANTHROPOGENIC DISTURBANCES ON THE UNCERTAINTY OF GROWTH FORECASTS

Table 4.2: Model and sampling-related variance contribution (m^2ha^{-1}), as well as the total variance estimated for each one of the four scenarios. The percentage contribution appears in parentheses.

Scenarios	Year	Model-related	Sampling-related	Total Variance
Baseline	2003	0.12 (0.6%)	20.59 (99.4%)	20.70
	2013	0.86 (5.3%)	15.61 (94.7%)	16.47
	2023	1.34 (9.3%)	12.98 (90.7%)	14.32
	2033	1.99 (12.1%)	14.39 (87.9%)	16.38
	2043	3.09 (16.8%)	15.31 (83.2%)	18.40
	2053	4.77 (23.1%)	15.89 (76.9%)	20.66
	2063	7.05 (30.1%)	16.41 (69.9%)	23.46
	2073	10.14 (37.4%)	17.00 (62.6%)	27.14
	2083	13.82 (43.8%)	17.72 (56.2%)	31.54
	2093	18.76 (50.2%)	18.59 (49.8%)	37.35
Harvest	2103	23.86 (54.8%)	19.62 (45.2%)	43.48
	2003	0.11 (0.5%)	20.58 (99.5%)	20.70
	2013	0.87 (5.3%)	15.63 (94.7%)	16.50
	2023	1.61 (11.0%)	12.99 (89.0%)	14.61
	2033	2.78 (17.5%)	13.15 (82.5%)	15.93
	2043	4.47 (25.6%)	13.00 (74.4%)	17.48
	2053	6.26 (32.9%)	12.73 (67.1%)	18.99
	2063	8.10 (39.3%)	12.53 (60.7%)	20.63
	2073	10.43 (45.7%)	12.39 (54.3%)	22.82
	2083	12.96 (51.4%)	12.28 (48.6%)	25.24
Harvest and SBW	2093	15.98 (56.7%)	12.19 (43.3%)	28.17
	2103	18.53 (60.4%)	12.13 (39.6%)	30.66
	2003	0.13 (0.6%)	20.59 (99.4%)	20.72
	2013	0.80 (4.8%)	15.61 (95.2%)	16.41
	2023	1.47 (10.1%)	12.98 (89.9%)	14.45
	2033	49.77 (79.2%)	13.04 (20.8%)	62.81
	2043	84.78 (86.8%)	12.94 (13.2%)	97.71
	2053	106.72 (89.3%)	12.72 (10.7%)	119.54
	2063	118.23 (90.5%)	12.46 (9.5%)	130.70
	2073	127.34 (91.3%)	12.21 (8.7%)	139.54
SBW	2083	125.08 (91.3%)	11.94 (8.7%)	137.02
	2093	125.00 (91.4%)	11.69 (8.6%)	136.70
	2103	127.03 (91.7%)	11.51 (8.3%)	138.54
	2003	0.13 (0.6%)	20.58 (99.4%)	20.71
	2013	0.77 (4.7%)	15.61 (95.3%)	16.38
	2023	1.53 (10.5%)	12.99 (89.5%)	14.52
	2033	55.24 (79.4%)	14.28 (20.6%)	69.52
	2043	102.84 (87.1%)	15.27 (12.9%)	118.11
	2053	138.14 (89.6%)	15.99 (10.4%)	154.13
	2063	162.35 (90.7%)	16.56 (9.3%)	178.91
	2073	175.46 (91.1%)	17.07 (8.9%)	192.53
	2083	176.47 (90.9%)	15.59 (9.1%)	194.06
	2093	178.45 (90.8%)	18.14 (9.2%)	196.59
	2103	180.48 (90.6%)	18.80 (9.4%)	199.28

4 Discussion

This study focused on the uncertainty of large-area volume forecasts under the effect of harvesting and SBW outbreaks. It turns out that both disturbance types affect the volume yield in predictions and their variances. It is obvious that omitting disturbances leads to overestimating growth (Valle et al., 2006). In our study, we managed to estimate this bias. Harvesting is accounted for in most simulations, while natural disturbances are often omitted due to their stochastic nature. In the Bas-Saint-Laurent region, omitting SBW outbreaks caused an overestimation of 7.4% in volume at the end of the 21st century (Fig. 4.3).

Uncertainty estimation was performed in the context of hybrid inference at the regional scale. This was possible because: (i) a hybrid bootstrap variance estimator was available, and (ii) the model benefited from a full stochastic implementation, which is a requirement for the use of the estimator (Fortin et al., 2018). Using this framework, we reproduced the variance patterns arising from the model and the sampling, and checked how they were affected by SBW outbreaks and harvest activities. Such a comprehensive consideration for the different sources of uncertainty in growth forecasts contributes to the originality of our study compared to past efforts.

Our first hypothesis was that SBW outbreaks induced more uncertainty in volume forecasts than harvesting. The scenarios including SBW outbreaks led to a variance that was three to six times greater than those in the scenarios without outbreaks (Fig. 4.4). In our forecasts, enabling the occurrence of SBW outbreaks generated some realizations where all plots that contained host species were suddenly affected by greater mortality rates, whereas the other realizations were only subject to regular mortality. In contrast, harvesting affected all realizations, and in each of them, only a few plots were harvested while the others were left untouched. The clear consequence is a greater variability from the model in the scenarios including SBW outbreaks.

A surprising result was that the scenario including harvest was slightly less uncertain than the baseline scenario. Although unexpected, it can be reasonably assumed that the endogenous nature of the harvest sub-model implies less variability. In ARTEMIS, harvest probabilities are based on some plot-level variables that are predicted by the model. For instance, the larger the basal area is, the greater the probability of harvest will be (Melo et al., 2017). Regardless of the realizations, plots with greater basal areas are then more prone to be harvested. As a consequence, there are fewer plots with large basal areas, and the population tends to be

more homogeneous than in the baseline scenario. As outlined in [Kneeshaw et al. \(2011\)](#), harvest activities are likely to produce similar structural forests when compared with others natural disturbances such as spruce budworm outbreaks. Given this homogenizing effect of the harvesting, we cannot entirely validate our second hypothesis, which was that disturbances were expected to induce a greater deal of uncertainty in the forecasts than the sampling. This was true for SBW outbreaks, but not for harvesting activities.

The sampling-related variance did not show a decreasing trend in long-term predictions as it did in the study of [Melo et al. \(2018\)](#). It must be stressed that the sampling-related variance as estimated through the hybrid estimator of [Fortin et al. \(2018\)](#) is actually the variance of the mean plot-level predicted values. As the projection length increases, these plot-level predicted values tend to be similar due to model convergence. As reported in [Melo et al. \(2018\)](#), the population variance cannot be estimated from this sampling variance because it overlooks the increasing contribution of the residual errors. In other words, the flat trends we observed for the sampling-related variances cannot be interpreted as a constant degree of heterogeneity between the plots all along the projection. This plot-to-plot heterogeneity actually increases because of the model residual errors.

In the study of [Melo et al. \(2018\)](#), model- and sampling-related variances of basal area predictions were reported per ecotype for the Bas-Saint-Laurent region. To check if the different patterns in the sampling-related variances were a matter of ecotype, we also estimated model- and sampling-related variances of volume predictions per ecotype. We obtained trends similar to those observed by [Melo et al. \(2018\)](#), even if we were working with volumes, which allowed us to rule out any variable representation effect. This led us to consider the implications of an ecotype stratification on the sampling variance. The differences observed in the behavior of sampling uncertainty herein and in [Melo et al. \(2018\)](#) could probably be due to inter-ecotype variance. Further details about these additional results can be found in the Appendix C.

[Melo et al. \(2018\)](#) also outlined the impact of sample size in sampling variance. Despite the greater sample size in this study, the estimated sampling variances were still large. Again, the inter-ecotype variance could be an explanation for this trend that we observed in our simulations.

It is known that a stratified sampling design can decrease the variance of the estimates. The decrease in the variance is linked to the homogeneity within the strata ([Gregoire and Valentine, 2008](#), p. 127). In forestry, [McRoberts et al. \(2012\)](#)

demonstrated that a stratification based on LiDAR data reduced the variance of mean volume estimates of growing stock. [McRoberts and Westfall \(2016\)](#) also obtained smaller variance estimates when using a stratified estimator in the context of individual tree volume. Building on this, we can reasonably assume that our sampling variance would decrease if we used a stratification based on the ecotypes, for example. However, this remains to be tested.

Finally, in our simulations, we observed that there is uncertainty related to sampling, but more importantly, there is greater uncertainty in modeling growth when SBW outbreaks are included. Previous studies ([Breidenbach et al., 2014](#); [McRoberts and Westfall, 2014](#); [Ståhl et al., 2014](#)) concluded that the efforts to reduce sampling uncertainty were justified because it was the greatest source of uncertainty. In our study, priority should be given to reducing the uncertainties that stemmed from SBW outbreaks when forecasting growth. We do not advocate that sampling uncertainty should not be considered at all, but it clearly is smaller than the uncertainty from SBW outbreaks.

Existing research in growth forecast uncertainties under SBW is limited. The recent studies that are available support our findings. [Boulanger et al. \(2016\)](#) argued that model-specification uncertainty should be the focus of research assessing future pest outbreak dynamics. [Gray \(2017\)](#) suggested that future outbreak forecasts could be improved by building models with more precise data.

In this respect, an obvious question arises as to whether or not other exogenous disturbances such as fire and wind have the same effect on growth forecasts as those we observed for SBW. Introducing exogenous disturbances is subject to high levels of uncertainty ([Artés et al., 2013](#); [Cencerrado et al., 2015](#)). Relying on the spatial synchrony theory ([Williams and Liebhold, 2000](#)), it is reasonable to assume that large forest fires and severe windstorms would greatly impact some realizations, while others would remain untouched. In the study of [Bergeron et al. \(2017\)](#), in which forest age classes were assessed in relation to fire and harvest activities, the scenarios with the greatest variability were those that considered fire occurrence. In addition, [Pichancourt et al. \(2018\)](#) also reported an increase in the variance of carbon predictions when considering windstorms.

However, the comparison between the three types of disturbances - fire, storms and insect outbreaks - in terms of uncertainty contribution is not so simple and remains to be tested. The vulnerability of forest stands is dependent on the type of disturbances. For example, forest fires are more likely to occur in old boreal stands ([Bernier et al., 2016](#)). When windstorms occur, the tallest trees are more prone

to damage than the smaller ones (Schmidt et al., 2010). For insect outbreaks, the number of host species is often limited and, for this reason, the damage is highly dependent on the species composition. In the case of SBW outbreaks, the host-tree species are: balsam fir, black spruce, white spruce and red spruce (Gray, 2017).

The variance of large-area forecasts is closely related to the severity of the damage when the disturbance occurs. In our case study, the damage of SBW was severe because the host species were abundant at the regional level. The three host species represented 30% of the basal area of Bas-Saint-Laurent forests, with balsam fir alone representing 20.5% of the basal area at the regional level (Table 4.1).

Bergeron et al. (2017) assumed that all stands had an equal probability to be burned, regardless of their age or changes in vegetation composition. Likewise, in our study, we assumed that SBW outbreaks had equal probabilities of occurrence over time. In other words, the probability that a SBW outbreak occurs is not impacted by previous outbreaks. This can have an effect on the estimated variance. As a matter of fact, the probabilities of SBW outbreak occurrence are probably not independent of previous outbreaks. Candau and Fleming (2005) modeled SBW outbreak occurrences, and reported that the frequency and defoliations exhibited a spatial pattern that is influenced by climate and forest composition. This is a more complete approach in modeling, and using it would probably reduce the estimated variances we obtained in our study.

Estimating uncertainty arising from disturbances can provide important insights. In past studies, the focus was generally on model development or model uncertainty, while the perspective of hybrid inference was missing. In terms of approach, we chose to run the model at the plot level and to then scale the predictions up to a greater spatial level. This approach is known as the direct extrapolation method and is recommended to reduce errors arising from nonlinearity, such as Jensen's inequality (Wu et al., 2006). Furthermore, variance estimates based on the Monte Carlo technique, like those in this study, are preferred to analytical methods since they apply to complex and nonlinear models (Wilson and Smith, 2013), such as ARTEMIS.

The scenario in which harvest and SBW outbreaks could occur simultaneously resulted in smaller volume forecasts. In reality, the estimates in this particular scenario can be underestimated. In the event of an outbreak, salvage cuttings normally take place (Boulanger and Arseneault, 2004). However, ARTEMIS does not consider this possibility, which means that some plots that are harvested by the model are actually spared. Moreover, a major driver of the harvest model is the annual

allowable cut (AAC) volume, that is estimated by a government agency. In our simulation, this AAC volume is constant, whereas in practice, it is re-estimated every 5 years.

Even if harvest activities did not have the greatest contribution to the forecast variance, some authors reported the impacts of uncertainties related to harvest in forest planning. For example, [Pasalodos-Tato et al. \(2013\)](#) observed economic losses on harvest scheduling due to errors in forest inventory. [Makinen et al. \(2012\)](#) also found that errors on growth predictions and forest inventory had a critical impact on harvest scheduling planning problems. As recommended by [Robinson et al. \(2016\)](#) and [Daniel et al. \(2017\)](#), efforts to assess uncertainties in harvest activity should be done with respect to management planning. Recent developments integrated uncertainties into forest management planning. Non probabilistic methods, such as programming analysis, were developed in [Eyvindson and Kangas \(2016\)](#) and [Eyvindson and Kangas \(2017\)](#). This issue of management planning is beyond the scope of this paper, but our framework may serve as a basis to facilitate the implementation of these methods.

5 Conclusions

Estimating uncertainties in forest growth forecasts under disturbances can provide insights to decision-makers. Volume forecasts for the Bas-Saint-Laurent region are more uncertain when including SBW outbreaks. This natural disturbance proved to be the most important source of uncertainty against harvest and sampling variance. We therefore suggest that forest management would be more realistic if it accounted for the uncertainties that stem from natural disturbances.

In order to reduce the uncertainty of large-area growth forecasts in the Bas-Saint-Laurent region, the understanding and prediction of SBW dynamics is a crucial issue. An essential step would be to take the relationships between the variables that explain the occurrence of disturbance events into account. Along with what was proposed by [Gray \(2017\)](#), we also suggest that efforts should be made to gather reliable datasets that could be used to create or improve existing models of SBW dynamics.

Even though the sampling variance was not the greatest contributor to the forecast variances, it could be easily improved through the use of stratified estimators. The bootstrap estimator developed by [Fortin et al. \(2018\)](#) can be adapted to the stratified sampling design. This remains to be tested.

Finally, due to the greater contribution of model variance to the uncertainties of forecasts, a deep uncertainty analysis that applies in the context of model-based decision could provide complementary and valuable insights as discussed in [Maier et al. \(2016\)](#).

*CHAPTER 4. THE EFFECT OF NATURAL AND ANTHROPOGENIC
DISTURBANCES ON THE UNCERTAINTY OF GROWTH FORECASTS*

Chapter 5

Discussion and Perspectives

Contents

1	Research problems	86
2	Framework	87
3	Estimation of uncertainties in large-area growth forecasts .	88
4	Perspectives	92

This last section presents a unified vision of my thesis. As a starting point, I review the research problems I worked on, and I then present an overview of how we approached those problems. I also discuss the major contributions of the thesis and, finally, the perspectives for future studies.

1 Research problems

Tree-level growth models provide insights into how forest dynamics can vary in a short- or long-term interval, under natural and anthropogenic disturbances. Whether it is for management purposes or for scientific research, the use of these models requires a quantitative assessment of their associated uncertainties.

The structure of the growth model that allowed for growth forecasts in this thesis was also the target of attention. Among the sub-models that compose a growth model, developing a harvest model was a specific goal. Harvest models are essential to forecast growth and to assist management plans ([Baskent and Keles, 2005](#)). Gaps were reported in the modeling approach, which should be flexible enough to predict harvest probabilities by including multiple levels of explanatory variables and changes in the economy, legislation and management practices ([Antón-Fernández and Astrup, 2012](#); [Thurnher et al., 2011](#)). The survival analysis approach is meant to tackle censored data, uneven time intervals, time-varying variables, and regional- and national-level variables ([Wang et al., 2013](#)). In our study, survival analysis did not only make it possible to overcome these limits but, in a broader perspective, it also made it possible to run business-as-usual scenarios when integrated into ARTEMIS-2009.

Over the past years, researchers have outlined the interest in estimating uncertainties in growth modeling (e.g., [Breidenbach et al., 2014](#); [Daniel et al., 2017](#); [Gertner, 1984](#); [Kangas and Kangas, 2004](#); [Ståhl et al., 2014](#)). However, uncertainty estimation is a complex issue, which is related to time and space. Most studies present simplified contributions, probably due to such a complexity. Hence, it is natural that unanswered questions remain. For instance, some studies treat only a single source of uncertainty such as model error (e.g., [Holopainen et al., 2010](#)). Some authors who carried out uncertainty estimations in the context of hybrid inference addressed the issues with simple models and not with complex models. Additionally, the time horizon of forecasts was a missing point (e.g., [McRoberts and Westfall, 2014](#)). Moreover, uncertainty estimation is often based on analytical approaches (e.g., [Fu et al., 2017](#)), which are of limited applicability for complex

models. Finally, the effects of disturbances on the uncertainty of growth forecasts have often been overlooked (e.g., [Gray, 2017](#)).

The central research issue of this thesis was focused on the uncertainty that stemmed from forest growth forecasts at a regional scale. The way we chose to answer some of these questions is presented below.

2 Framework

Uncertainty estimation of large-area growth forecasts was first addressed in Paper I, where we estimated the variance of basal area forecasts for broadleaved, mixed and coniferous ecotypes in the Bas-Saint-Laurent (BSL) region. We focused on the model and the sampling contribution to the total estimated variance. Thus, since Quebec's forests are disturbed by harvest, among other factors, we proposed an original approach for modeling plot-level harvest probabilities in Paper II. We used the survival analysis technique, which is meant to account for censored data ([Lawless, 2003](#), p. 10). The approach also includes time-varying variables. The model was meant to be integrated into ARTEMIS-2009. Finally, in Paper III, we placed our focus on the uncertainty that arises from disturbances - spruce budworm outbreaks and harvest - which are a relevant forest management issue in the BSL region. In this third chapter, large-area growth forecasts accounting for forest disturbances were carried out in terms of volume for the whole BSL region.

In order to estimate and rank the sources of uncertainty, the variance was considered to be the quantity of reference necessary to understand variability in forecasts. In the context of large-area forecasts, the estimator that enables the upscaling of plot predictions to larger areas relies on both the probabilistic sampling design and the model for inference. These hybrid estimators are often based on analytical approaches or the Monte Carlo technique ([Fortin et al., 2018](#)). Bootstrap methods are known to be flexible enough to tackle uncertainties from any distribution ([Wehrens et al., 2000](#)), while analytical estimators are sometimes limited when dealing with complex models ([Fortin et al., 2018](#)). Hybrid estimators have been used in contexts of large-area biomass, volume or carbon estimation but, to the best of our knowledge, they have never been applied in the context of growth forecasts.

Analytical hybrid estimators, i.e., estimators based on algebra, exist for linear, nonlinear and generalized linear models (e.g., [Fortin et al., 2016](#); [Saarela et al., 2015](#); [Ståhl et al., 2011, 2016](#)). The available analytical hybrid estimators rarely apply to coupled models used as decision-support tools because they are complex

cases. ARTEMIS falls into this category. An alternative to deal with complexity is the use of hybrid estimators based on Monte Carlo techniques. Inspired by [Rubin \(1987\)](#), [McRoberts and Westfall \(2014\)](#) and [McRoberts et al. \(2016\)](#) used bootstrap estimators for propagating uncertainty through individual tree volume. However, [Fortin et al. \(2018\)](#) reported that the available variance estimator might be biased in some circumstances. Thus, they presented a corrected bootstrap variance estimator that can be used in the context of hybrid inference, regardless of the model complexity. However, the model must implement the stochastic mode. This variance estimator is the one we used in this thesis.

The ARTEMIS-2009 model produces growth forecasts in a full stochastic manner at the tree level, which can be aggregated at the plot level and then upscaled at the regional level. Scaling is inevitable in the context of large-area forecasts since the growth model applies at the level at which it was fitted. According to [Wu et al. \(2006, p. 27, 30\)](#), different methods exist to bridge this scale gap between the model and the whole ecosystem. They can be grouped into two categories: (i) averaging the input data and parameters before the simulation, or (ii) averaging the outputs of the simulation. The first case is known as simple averaging and is generally recommended for linear models and deterministic simulations because of potential bias related to [Jensen's \(1906\)](#) inequality. In the second case, known as the direct extrapolation method, the model is run locally - in accordance with its spatial level - and respects the inputs and the parameter scale of nonlinear and complex models. This strategy avoids any potential bias caused by Jensen's inequality, which was already proven to impact forecasts in ARTEMIS-2009 ([Fortin and Langevin, 2012](#)). Due to the complexity of our forecasts, we had no choice but to work with the direct extrapolation method.

We depart from previous contributions by accounting for most sources of stochasticity in our growth forecasts. As a matter of fact, stochasticity was a requirement for applying the bootstrap hybrid variance estimator.

3 Estimation of uncertainties in large-area growth forecasts

In this section, we present three main aspects that highlight the essential results and contributions of this thesis.

Model and sampling contributions to the uncertainty of growth forecasts are relative to the time horizon and the spatial scale

In this thesis, model- and sampling-related variances were estimated in terms of basal area and per ecotype for the Bas-Saint-Laurent region (Paper I). The sampling-related variance turned out to be an important source of uncertainty when forecasting growth over a large area, which corroborated the results of previous studies (e.g., Breidenbach et al., 2014; McRoberts and Westfall, 2014). However, we observed that this was only true for short-term forecasts. Long-term growth forecasts (e.g., 100 years) resulted in model-related variances similar to sampling-related variances. Given the trends we observed, we could assume that in forecasts longer than 100 years, the model would be the most important contributor. Moreover, for larger areas, the sampling-related variance should probably be smaller due to a possibly larger sample size. In this case, the model-related variance could become the main contributor to the total variance, even in a mid-term forecast (e.g., 50 years).

In Paper III, we studied the contribution of disturbances in terms of uncertainty in volume forecasts at a regional level. The model-related variance turned out to be the greatest contributor to the variance before the end of the 100-year horizon. In the scenarios that accounted for SBW outbreaks, the contribution of the model to the total variance was already greater than that of the sampling after 30 years. In the scenario where only harvest activities were included, the model-related variance became the greater contributor after 70 years. At this regional scale, our sample size was four times greater than the sample sizes in Paper I. Nevertheless, the sampling variance, which was expected to be smaller, remained constant. In spite of their smaller sample sizes, the three populations in Paper I were less variable due to the ecotype-based stratification. We hypothesized that this stratification could reduce the sampling-related variances of these regional forecasts (Gregoire and Valentine, 2008, p. 127).

Concluding remark: To reduce uncertainties in growth forecasts in the short term, it is necessary to work on increasing the sample size and upgrading the sampling design. The use of stratified estimators could help reduce sampling-related variances in large-area forecasts, but in our case study, this remains to be tested. To reduce uncertainties in mid- and long-term forecasts, the model-related variance should be the target of improvements.

New and flexible approaches can be used to improve harvest and mor-

mortality models

The perspective of model uncertainty was further studied in Paper I by decomposing the total variance in its components. Among the sub-models that compose ARTEMIS-2009, mortality was the component that induced the greatest share of variance. By disabling the stochasticity of the mortality sub-model, the total variances decreased by 40% to 60%. There are major concerns related to mortality predictions in growth models. Among others, trees predicted as being dead by error in a growth step have no chance to become alive in the subsequent step. This can lead to underestimation and impact forest profits. The diameter increment sub-model had a maximum contribution of 25% to the total variance. Although many mortality models exist ([Bugmann, 2001](#)), our results suggest that the mortality sub-model should be given the priority in our efforts.

The potential of survival analysis techniques to model harvest was presented in Paper II. Annual predictions of harvest occurrence were obtained through a survival model, which is composed of a time-dependent part. In this part of the model, the time-varying and regional variable studied (annual allowable cut volumes, AAC), as well as the cluster random effect representing spatial correlation, resulted in an improvement of the model fit. Modeling harvest with respect to these time-varying variables was an original contribution of this paper. By coupling the harvest model with a growth model, the predictions would make it possible to account for business-as-usual forest management and could be helpful in these AAC calculations. A drawback to the harvest model is that the AAC was considered as constant in the future and, as a consequence, the harvesting schedules will remain constant in the future. Further model improvements should consider the changes in AAC calculations.

Since mortality data are also interval-censored, a promising path would be the use of survival analysis when modeling mortality. The flexibility of the technique would make it possible to address some issues in mortality models. For instance, an interaction between tree mortality and disturbances exists, which can be treated through a survival model (e.g., [Manso et al., 2015](#)). Modeling mortality through survival analysis was also previously done by [Rose et al. \(2006\)](#) and [Fortin et al. \(2008\)](#). However, none of these three aforementioned studies explored the time-varying component. For instance, climate variables, which are known to influence mortality probability, can be tackled through this time-varying part of the model.

Concluding remark: The management strategies or the occurrence of mortality

may change during the intervals in which the data are collected or the forecasts are carried out. Whether it is for modeling harvest or mortality, time-varying variables are allowed to vary within the intervals. This provides greater flexibility to overcome previous model limitations, such as accounting for changes in climate variables.

The impact of natural disturbances on growth forecast uncertainty is much greater when compared to anthropogenic disturbances and sampling

Paper III focused on estimating the contribution of disturbances to the total variance of growth forecasts. More specifically, the contributions of SBW outbreaks, harvest and sampling were compared. The stochastic character of induced SBW outbreaks increases the variance by 80% after the first 30 years of the time horizon. At the end of the horizon, the model-related variance accounted for 91% of the total variance. In light of this result, it can be reasonably assumed that other natural disturbances such as fires and windstorms would also greatly impact the variance. The estimated variances under harvest activities slightly decreased, probably because the harvest tends to homogenize forests.

This outlook proposed in Paper III outlined the model contribution to the total variance in large-area growth forecasts. From a practical point of view, such uncertainties can have important implications. For instance, the last SBW outbreak in the Bas-Saint-Laurent region had affected a total of 1.1 million ha in 2016, which represented 48% of the total area (MFFP, 2017). Considering the potential losses in the commercial value caused by SBW outbreaks, there would be a benefit in estimating and communicating uncertainties in growth forecasts related to this disturbance. This knowledge can actually help managers to adjust their management practices in the near future.

In Paper III, SBW outbreak recurrence represented an exogenous variable in the growth simulations. A sub-model accounting for the insect dynamics could be integrated into the growth model, allowing the occurrence to be part of the system. SBW occurrence would then become an endogenous variable.

By including SBW outbreaks in forest growth forecasts, the implications of time cycle processes and interactions with other disturbance agents should be taken into account. For example, SBW outbreaks have an indirect effect on harvest probabilities. As soon as an outbreak increases in severity, decision-makers normally promote harvest operations. In such a case, the harvest model should account for this response, which was not taken into account in our simulations.

Concluding remark: Uncertainties related to growth forecasts appeared to be greatly affected by natural disturbances. Hence, priority should be given to the prediction of the occurrence of disturbances. This could be done by working with better data or by improving modeling developments.

4 Perspectives

This thesis made it possible:

1. To evaluate the contributions of the model- and the sampling-related variances to the total variances of large-area growth forecasts;
2. To outline the potential of survival analysis in predicting harvest probabilities;
3. To estimate the impact of large-scale disturbances on the uncertainty of growth forecasts.

In light of the results, we suggest some perspectives that can provide further insights into forest growth modeling.

The harvest model based on the survival analysis approach can be adapted to other levels of management planning. This way, the spatial correlations that were observed in the occurrence of harvesting could be better handled than through a mere random effect. Also, the predictions could be improved by accounting for road networks, or even integrating the blocks to be harvested.

In complex models such as tree-level models, it is difficult to account for the stochasticity of all components and their interactions. We acknowledge that the realism of the simulations would be improved if the evolution of the annual allowable cut volumes was considered. Realism would also be improved by considering the interaction between climate variables and mortality or SBW recurrence.

This thesis focused on SBW outbreaks, which are a major natural disturbance in Bas-Saint-Laurent forests. Other natural disturbances such as fires and wind-storms should also be considered in future research. A new perspective would be the comparison of natural disturbances in terms of contribution to growth forecast uncertainty. Insects are more species-specific than forest fires. The extent and the intensity of the disturbances could therefore be different. As a consequence, the impact on the variance of growth forecasts could also be distinct. Thus, we might

be able to determine which disturbance induces more uncertainty, which could be a follow-up to this study.

As previously discussed in Papers I and III, the population variance cannot be directly derived from the sampling variance. It is also necessary to take the residual errors of the model into account. Hence, the consequence is that the exact variability among the population units remains unknown. In order to overcome this limitation, the bootstrap variance estimator should be further developed in order to decompose the share of uncertainty for each model-related source.

While this thesis was focused on the perspective of scientific research, the issues we worked on can be of interest to forest managers. We suggest that uncertainty estimation should be implemented in management plans. Some efforts have been made in this direction, as demonstrated in [Daniel et al. \(2017\)](#) and [Eyvindson and Kangas \(2016\)](#). Because they greatly impact growth forecast variances, there could also be an advantage to integrating natural disturbances into management plans.

Regarding the implications of this thesis on forest planning, the estimated uncertainties open up the discussion about the place of short-term actions in forest management. Relying on uncertain predictions to determine the long-term harvest schedule may not be as sustainable as would be expected. This is a common context in forestry, where managers often have to make decisions on the basis of uncertain information. This context, which can also be extended to other fields of study, has brought the focus on deep uncertainty analysis (e.g., [Dittrich et al., 2016](#)). Among others, a strategy to deal with uncertain futures would be to consider safety bounds or to opt for flexible options ([Hallegatte, 2009](#)).

We were faced with methodological issues related to the variance decomposition. The nonlinear interaction between sub-models could be better treated if new information or methods were made available. For instance, new approaches should consider the amount of variability that comes from the interaction dependency ([Willems, 2012](#)).

We were also faced with methodological issues related to the computational time of the simulations and uncertainty estimation. Because of limited memory capacity, simulating the total number of Monte Carlo realizations was not possible. We were therefore forced to break down this total number of realizations into four groups, and to then carry out the simulations in four batches. Variance estimation was carried out after the output realizations were re-grouped in R. All these intermediate steps were time-consuming and subject to eventual implementation errors. This was an important issue in our simulations. By directly implementing the bootstrap variance

hybrid estimator in ARTEMIS-2009, a great amount of time could be saved.

Another way to reduce the computational time would be the use of a hybrid approach, which is different from hybrid inference. A hybrid approach to propagate errors combines Monte Carlo and Taylor series, like the one used [Fu et al. \(2017\)](#). This way, it would be possible to avoid the limitations and take advantage of both methods.

Finally, bridging data gaps could help to improve the models, reducing potential errors arising from sampling and better predicting disturbances.

Appendix A

Corrected bootstrap variance estimator

Fortin et al. (2018) presented a bootstrap hybrid variance estimator for complex models. Complex models are defined as complex systems for which it is impossible to predict the behaviour without simulations. From a mathematical point of view, they are not differentiable and, consequently, common methods for error propagation such as Taylor series cannot be used. Individual-based models in ecology are typical examples of complex models (Railsback, 2001).

Fortin et al. (2018) provided a proof of the unbiasedness of their variance estimator in the context of random sampling without replacement with unequal inclusion probabilities. The proof assumed the use of the well-known Horvitz-Thompson (HT) estimator (Horvitz and Thompson, 1952) for the total of the population.

If the inclusion probabilities are equal for all population units, many authors prefer to work with an estimator of the mean instead of the HT estimator for the total. This was precisely the case in this study. The unbiasedness of the bootstrap hybrid variance estimator developed in Fortin et al. (2018) is demonstrated for the mean of the population in the next developments. For the sake of simplicity, we will use subscripts m and d to refer to model- and design-based components, respectively, while hybrid components will bear both subscripts.

The mean of the population (μ_y) is

$$\mu_y = \sum_{i=1}^N \frac{y_i}{N} \quad (\text{A.1})$$

where y_i is the value of population unit i and N is the total number of population units.

In the context of model-based inference, we assume the existence of a super-population model. For the sake of simplicity, we will further assume that the super-population model is linear: $y_i = \mathbf{x}_i\boldsymbol{\beta} + \epsilon_i$. Replacing y_i in Eq. A.1 by the right-hand side of the super-population model yields

$$\begin{aligned} \mu_y &= \sum_{i=1}^N \frac{\mathbf{x}_i\boldsymbol{\beta}}{N} + \sum_{i=1}^N \frac{\epsilon_i}{N} \\ &= \boldsymbol{\mu}_x\boldsymbol{\beta} + \sum_{i=1}^N \frac{\epsilon_i}{N} \end{aligned} \quad (\text{A.2})$$

where $\boldsymbol{\mu}_x$ is the mean of the auxiliary variables that enter into the super-population model. Even though the population parameters and the auxiliary variables are known for the entire population, the expression $\boldsymbol{\mu}_x\boldsymbol{\beta}$ remains the expected

tation of the mean of the population and the variance of the true population mean is

$$\begin{aligned}
 \mathbb{V}_m(\mu_y - \boldsymbol{\mu}_x \boldsymbol{\beta}) &= \mathbb{V}_m\left(\sum_{i=1}^N \frac{\epsilon_i}{N}\right) \\
 &= \frac{1}{N^2} \mathbb{V}_m\left(\sum_{i=1}^N \epsilon_i\right) \\
 &= \frac{1}{N^2} \left(\sum_{i=1}^N \mathbb{V}_m(\epsilon_i) + \sum_{i=1}^N \sum_{i' \neq i}^N \mathbb{COV}(\epsilon_i, \epsilon_{i'}) \right) \quad (\text{A.3})
 \end{aligned}$$

The context of hybrid inference arises when these auxiliary variables are not available for all the population units, but only for a sample of population units. If we assume that this sample (s) of auxiliary variables was drawn following a probabilistic design of simple random sampling without replacement, then $\boldsymbol{\mu}_x$ can be estimated through the sample mean:

$$\hat{\boldsymbol{\mu}}_x = \sum_{i \in s} \frac{\mathbf{x}_i}{n} \quad (\text{A.4})$$

where n is the sample size. Considering that the model parameters are also estimated, the hybrid point estimator of the mean is then

$$\hat{\mu}_y = \hat{\boldsymbol{\mu}}_x \hat{\boldsymbol{\beta}} \quad (\text{A.5})$$

with variance

$$\mathbb{V}_{md}(\hat{\boldsymbol{\mu}}_x \hat{\boldsymbol{\beta}} - \boldsymbol{\mu}_x \boldsymbol{\beta}) = \boldsymbol{\mu}_x \mathbb{V}_m(\hat{\boldsymbol{\beta}}) \boldsymbol{\mu}_x^T + \boldsymbol{\beta}^T \mathbb{V}_d(\hat{\boldsymbol{\mu}}_x) \boldsymbol{\beta} + \text{Tr}(\mathbb{V}_d(\hat{\boldsymbol{\mu}}_x) \mathbb{V}_m(\hat{\boldsymbol{\beta}})) \quad (\text{A.6})$$

The total variance in the context of hybrid inference is then the sum of the variance in Eqs. A.3 and A.6:

$$\begin{aligned}
 \mathbb{V}_{md}(\mu_y - \hat{\boldsymbol{\mu}}_x \hat{\boldsymbol{\beta}}) &= \boldsymbol{\mu}_x \mathbb{V}_m(\hat{\boldsymbol{\beta}}) \boldsymbol{\mu}_x^T + \boldsymbol{\beta}^T \mathbb{V}_d(\hat{\boldsymbol{\mu}}_x) \boldsymbol{\beta} + \text{Tr}(\mathbb{V}_d(\hat{\boldsymbol{\mu}}_x) \mathbb{V}_m(\hat{\boldsymbol{\beta}})) \\
 &\quad + \frac{1}{N^2} \left(\sum_{i=1}^N \mathbb{V}_m(\epsilon_i) + \sum_{i=1}^N \sum_{i' \neq i}^N \mathbb{COV}(\epsilon_i, \epsilon_{i'}) \right) \quad (\text{A.7})
 \end{aligned}$$

Bootstrap variance estimators have already been used in the context of hybrid

inference (e.g. [McRoberts and Westfall, 2016](#)). The method was inspired by [Rubin \(1987\)](#). It consists of drawing random deviates to account for the errors in the parameter estimates $(\epsilon_{\beta,b})$ and residual errors $(\epsilon_{i,b})$, with b being the realization index. Basically, for each realization b , the y_i in the design-based estimators of the mean and its variance are replaced by $\mathbf{x}_i\hat{\beta} + \mathbf{x}_i\epsilon_{\beta,b} + \epsilon_{i,b}$. An estimate of the mean $(\hat{\mu}_b)$ and its variance $(\hat{V}_d(\hat{\mu}_b))$ are then obtained for realization b . After repeating the process B times, a bootstrap estimator of the mean $(\hat{\mu}_{BS,y})$ and its variance $(\hat{V}_{BS}(\hat{\mu}_{BS,y}))$ are obtained as

$$\hat{\mu}_{BS,y} = \sum_{b=1}^B \frac{\hat{\mu}_b}{B} \quad (\text{A.8})$$

$$\hat{V}_{md}(\hat{\mu}_{BS,y}) = \hat{V}_m(\hat{\mu}_b) + \hat{\mathbb{E}}_m[\hat{V}_d(\hat{\mu}_b)] \quad (\text{A.9})$$

As the number of realizations B tends to infinity, the value of the estimator in Eq. A.8 converges to that of the estimator in Eq. A.5. However, it can be shown that the expectation of the variance estimator in Eq. A.9 overestimates the true variance (Eq. A.7). The expectation of the first term on the right-hand side of Eq. A.9 is

$$\begin{aligned} \mathbb{E}[\hat{V}_m(\hat{\mu}_b)] &= \mathbb{E} \left[\hat{V}_m \left(\sum_{i \in s} \frac{\mathbf{x}_i \hat{\beta} + \mathbf{x}_i \epsilon_{\beta,b} + \epsilon_{i,b}}{n} \right) \right] \\ &= \mathbb{E} \left[\hat{V}_m \left(\frac{\sum_{i \in s} \mathbf{x}_i \hat{\beta}}{n} + \frac{\sum_{i \in s} \mathbf{x}_i \epsilon_{\beta,b}}{n} + \frac{\sum_{i \in s} \epsilon_{i,b}}{n} \right) \right] \\ &= \mathbb{E} \left[\hat{V}_m(\hat{\mu}_x \epsilon_{\beta,b}) + \hat{V}_m \left(\frac{\sum_{i \in s} \epsilon_{i,b}}{n} \right) \right] \\ &= \mathbb{E} \left[\hat{\mu}_x \hat{V}_m(\hat{\beta}) \hat{\mu}_x^T + \frac{1}{n^2} \left(\sum_{i \in s} \hat{V}_m(\epsilon_i) + \sum_{i \in s} \sum_{i' \neq i} \widehat{\text{COV}}_m(\epsilon_i, \epsilon_{i'}) \right) \right] \\ &= \boldsymbol{\mu}_x \mathbb{V}_m(\hat{\beta}) \boldsymbol{\mu}_x^T + \text{Tr}(\mathbb{V}_d(\hat{\mu}_x) \mathbb{V}_m(\hat{\beta})) \\ &\quad + \frac{1}{Nn} \sum_{i=1}^N \mathbb{V}_m(\epsilon_i) + \frac{n-1}{Nn(N-1)} \sum_{i=1}^N \sum_{i' \neq i}^N \text{COV}_m(\epsilon_i, \epsilon_{i'}) \end{aligned} \quad (\text{A.10})$$

The expectation of the second term on the right-hand side of Eq. A.9 is

$$\begin{aligned}
 \mathbb{E}[\hat{\mathbb{E}}_m[\hat{\mathbb{V}}_d(\hat{\mu}_b)]] &= \mathbb{E} \left[\hat{\mathbb{E}}_m \left[\left(1 - \frac{n}{N} \right) \sum_{i \in s} \frac{(\mathbf{x}_i \hat{\beta} + \mathbf{x}_i \epsilon_{\beta,b} + \epsilon_{i,b} - \sum_{i \in s} \frac{\mathbf{x}_i \hat{\beta} + \mathbf{x}_i \epsilon_{\beta,b} + \epsilon_{i,b}}{n})^2}{n(n-1)} \right] \right] \\
 &= \mathbb{E} \left[\hat{\beta}^T \hat{\mathbb{V}}_d(\hat{\mu}_x) \hat{\beta} + \text{Tr}(\hat{\mathbb{V}}_d(\hat{\mu}_x) \hat{\mathbb{V}}_m(\hat{\beta})) + \frac{1-n/N}{n^2} \sum_{i \in s} \hat{\mathbb{V}}_m(\epsilon_i) + \frac{n/N-1}{n^2(n-1)} \sum_{i \in s} \sum_{i' \neq i} \widehat{\text{COV}}_m(\epsilon_i, \epsilon_{i'}) \right] \\
 &= \beta^T \mathbb{V}_d(\hat{\mu}_x) \beta + 2\text{Tr}(\mathbb{V}_d(\hat{\mu}_x) \mathbb{V}_m(\hat{\beta})) + \frac{1-n/N}{Nn} \sum_{i=1}^N \mathbb{V}_m(\epsilon_i) + \frac{n/N-1}{Nn(N-1)} \sum_{i=1}^N \sum_{i' \neq i}^N \text{COV}_m(\epsilon_i, \epsilon_{i'}) \quad (\text{A.11})
 \end{aligned}$$

The sum of the expectations shown in Eqs A.10 and A.11 yields the expectation of the bootstrap hybrid variance estimator in Eq. A.9:

$$\begin{aligned}
 \mathbb{E}[\hat{\mathbb{V}}_{md}(\hat{\mu}_{\text{BS},y})] &= \boldsymbol{\mu}_x^T \mathbb{V}_m(\hat{\beta}) \boldsymbol{\mu}_x + 3\text{Tr}(\mathbb{V}_d(\hat{\mu}_x) \mathbb{V}_m(\hat{\beta})) + \beta^T \mathbb{V}_d(\hat{\mu}_x) \beta \\
 &\quad + \frac{2-n/N}{Nn} \sum_{i=1}^N \mathbb{V}_m(\epsilon_i) + \frac{n/N+n-2}{Nn(N-1)} \sum_{i=1}^N \sum_{i' \neq i}^N \text{COV}_m(\epsilon_i, \epsilon_{i'}) \quad (\text{A.12})
 \end{aligned}$$

Given the true variance shown in Eq. A.7, the bias of this variance estimator is then:

$$\begin{aligned}
 \text{Bias}(\hat{\mathbb{V}}_{md}(\hat{\mu}_{\text{BS},y})) &= 2\text{Tr}(\mathbb{V}_d(\hat{\mu}_x) \mathbb{V}_m(\hat{\beta})) + \frac{2N/n-2}{N^2} \sum_{i=1}^N \mathbb{V}_m(\epsilon_i) \\
 &\quad + \frac{2(n-N)}{N^2 n(N-1)} \sum_{i=1}^N \sum_{i' \neq i}^N \text{COV}_m(\epsilon_i, \epsilon_{i'}) \quad (\text{A.13})
 \end{aligned}$$

Note that the bias presented by Fortin et al. (2018) in their Eq. 11 reduces to the bias shown in Eq. A.13 when the inclusion probability is assumed to be n/N for all population units and the estimated variance is divided by N^2 to ensure the conversion from the estimate of the total to the estimate of the mean.

The bias shown in Eq. A.13 can be corrected through the inclusion of a new term. Let $\hat{\mathbb{V}}_d(\hat{\mu}_{\bar{y}})$ be the design variance estimator calculated on $\bar{y}_i = \sum_{b=1}^B \frac{\mathbf{x}_i \hat{\beta} + \mathbf{x}_i \epsilon_{\beta,b} + \epsilon_{i,b}}{B}$ instead of y_i . As B tends to infinity, then

$$\lim_{B \rightarrow \infty} \bar{y}_i = \mathbf{x}_i \hat{\beta} \quad (\text{A.14})$$

Then the variance estimator $\hat{\mathbb{V}}_d(\hat{\mu}_{\bar{y}})$ becomes

$$\begin{aligned}\lim_{B \rightarrow \infty} \hat{\mathbb{V}}_d(\hat{\mu}_{\bar{y}}) &= \frac{1}{n(n-1)} \sum_{i \in s} \left(\mathbf{x}_i \hat{\boldsymbol{\beta}} - \sum_{i \in s} \frac{\mathbf{x}_i \hat{\boldsymbol{\beta}}}{n} \right)^2 \\ &= \hat{\boldsymbol{\beta}}^T \hat{\mathbb{V}}_d(\hat{\boldsymbol{\mu}}_x) \hat{\boldsymbol{\beta}}\end{aligned}\tag{A.15}$$

and its expectation is

$$\mathbb{E}[\hat{\mathbb{V}}_d(\hat{\mu}_{\bar{y}})] = \boldsymbol{\beta}^T \mathbb{V}_d(\hat{\boldsymbol{\mu}}_x) \boldsymbol{\beta} + \text{Tr}(\mathbb{V}_d(\hat{\boldsymbol{\mu}}_x) \mathbb{V}_m(\hat{\boldsymbol{\beta}}))\tag{A.16}$$

The bootstrap variance estimator shown in Eq. A.9 can be corrected as follows

$$\hat{\mathbb{V}}_{md, \text{CORR}}(\hat{\mu}_{\text{BS}, y}) = \hat{\mathbb{V}}_m(\hat{\mu}_b) + 2\hat{\mathbb{V}}_d(\hat{\mu}_{\bar{y}}) - \hat{\mathbb{E}}_m[\hat{\mathbb{V}}_d(\hat{\mu}_b)]\tag{A.17}$$

Considering the expectations shown in Eqs. A.10, A.11 and A.16, the expectation of this corrected variance estimator is then

$$\begin{aligned}\mathbb{E}[\hat{\mathbb{V}}_{md, \text{CORR}}(\hat{\mu}_{\text{BS}, y})] &= \boldsymbol{\mu}_x^T \mathbb{V}_m(\hat{\boldsymbol{\beta}}) \boldsymbol{\mu}_x + \boldsymbol{\beta}^T \mathbb{V}_d(\hat{\boldsymbol{\mu}}_x) \boldsymbol{\beta} + \text{Tr}(\mathbb{V}_d(\hat{\boldsymbol{\mu}}_x) \mathbb{V}_m(\hat{\boldsymbol{\beta}})) \\ &\quad + \frac{1}{N^2} \left(\sum_{i=1}^N \mathbb{V}_m(\epsilon_i) + \sum_{i=1}^N \sum_{i' \neq i}^N \text{COV}(\epsilon_i, \epsilon_{i'}) \right)\end{aligned}\tag{A.18}$$

which is precisely the true variance as shown in Eq. A.7.

This mathematical proof shows that the corrected hybrid variance estimator shown in Eq. A.17 is unbiased if the super-population model is linear. In the case of nonlinear or complex super-population models, the estimator is asymptotically unbiased. The simulation studies carried out with complex models in Fortin et al. (2018) confirmed this asymptotical property.

Appendix B

Appendix of Paper II

*Implementation of the survival model in
SAS*

```
*****
* Implementation in SAS: Survival Analysis using Proc NLMIXED
*****
```

Obs: The annual AACs are contained in the 1988-2014 period, therefore, a total time interval of 27 years.

```
data= data set
RE = random effect
parameterestimates = output parameters
T-V = time-varying variable
out = output dataset
beta = vector of parameters
x = vector of explanatory variables
EO = event occurrence

proc sort data=; by RE; run;
ods output parameterestimates = myEstimates;
proc nlmixed data= technique=NRRIDG gconv=;
  where year >= 1988;
  title "Model 8";
  parms beta1 = 1 beta2 = 1 beta3 = 1 beta4 = 1
        gamma0 = 1 gamma1 = 1;
  array AAC[27] AAC 1988-AAC2014;
  array dummy[27] d1988-d2014;
  SAnnee = 0;
  do Year = 1 to 27;
    hazard = SAnnee + exp(gamma0 + gamma1 AAC [27] + u) dummy[27];
  end;
  proportional = beta x ;
  SCond = exp(-exp(proportional)hazard);
  S = SCond**exp(u);
  Likelihood = EO * log(S) + (1 - EO) * log(1-S);
  model EO ~ general(Likelihood);
  random u ~ normal(0, s2u) subject= RE;
  predict SCond out= ;
run
```

Appendix C

Appendix of Paper III

*Uncertainty estimation of volume
predictions per ecotypes*

The contribution of the model- and sampling-related variance components to the total uncertainty of volume predictions in the Bas-Saint-Laurent region is presented in Fig. C.1. For all ecotypes studied, the sampling-related variance was the most important contributor to the total variance for short-term forecasts. However, since this contribution decreases over time, model-related variances increase until they become equal to or slightly higher than sampling-related variances in the long term. These results corroborate the trends that were observed by Melo et al. (2018).

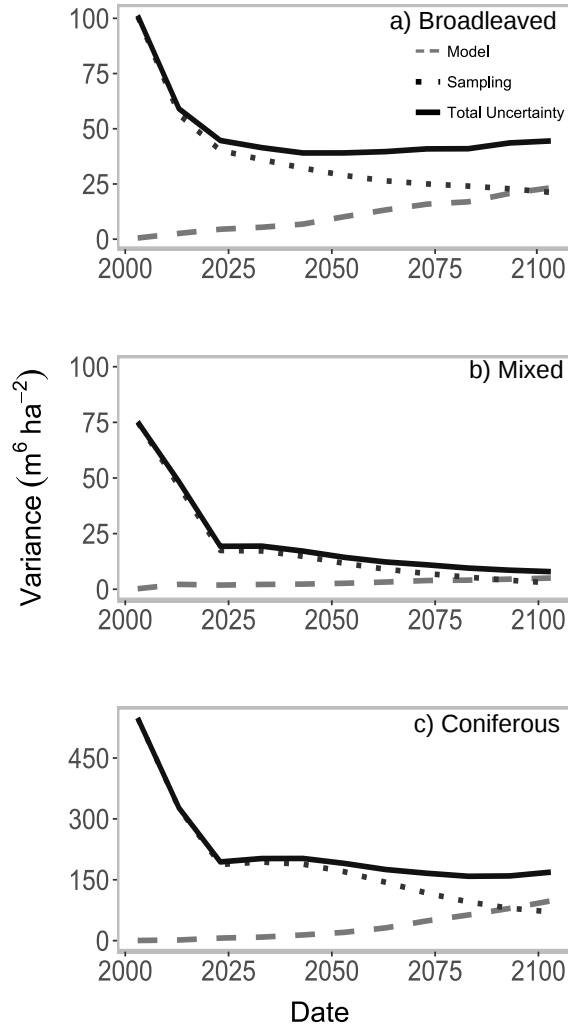


Figure C.1: Contribution of the model- and sampling-related variances to the total variance of volume predictions by ecotypes in the Bas-Saint-Laurent region. The simulations were run under the baseline scenario, i.e. the scenario without disturbances.

Bibliography

- Adams, H., Williams, A., Xu, C., Rauscher, S., Jiang, X., and McDowell, N. (2013). Empirical and process-based approaches to climate-induced forest mortality models. Frontiers in Plant Science, 4:1–4. [13](#)
- Allison, P. D. (1982). Discrete-time methods for the analysis of event histories. Sociological Methodology, 13:61–98. [47](#), [51](#)
- Ameztegui, A., Coll, L., and Messier, C. (2015). Modelling the effect of climate-induced changes in recruitment and juvenile growth on mixed-forest dynamics: The case of montane–subalpine pyrenean ecotones. Ecological Modelling, 313:84–93. [26](#)
- Antón-Fernández, C. and Astrup, R. (2012). Empirical harvest models and their use in regional business-as-usual scenarios of timber supply and carbon stock development. Scandinavian Journal of Forest Research, 27:379–392. [14](#), [46](#), [47](#), [61](#), [62](#), [86](#)
- Archambault, L., Deslile, C., Larocque, G., Sirois, L., and Belleau, P. (2006). Fifty years of forest dynamics following diameter-limit cuttings in balsam fir – yellow birch stands of the Lower St. Lawrence region, Quebec. Canadian Journal of Forest Research, 36:2745–2755. [20](#)
- Archambault, L., Morissette, J., and Bernier-Cardou, M. (1998). Forest succession over a 20-year period following clearcutting in balsam fir-yellow birch ecosystems of eastern Quebec, Canada. Forest Ecology and Management, 102:61–74. [20](#), [63](#)
- Artés, T., Cencerrado, A., Cortés, A., and Margalef, T. (2013). Relieving the effects of uncertainty in forest fire spread prediction by hybrid MPI-OpenMP parallel strategies. Procedia Computer Science, 18:2278–2287. [80](#)
- Aven, T. (2003). Foundations of Risk Analysis: A Knowledge and Decision-Oriented Perspective. John Wiley and Sons, United Kingdom. [14](#)
- Aven, T., Baraldi, P., Flage, R., and Zio, E. (2014). Uncertainty in Risk Assessment: the representation and treatment of uncertainties by probabilistic and non-probabilistic methods. John Wiley and Sons, United Kingdom. [15](#), [16](#), [26](#)
- Bakr, M. and Butler, A. (2005). Nonstationary stochastic analysis in well capture zone design using first-order Taylor’s series approximation. Water Resources Research, 41:1–12. [16](#)
- Baraldi, P. and Zio, E. (2008). A combined Monte Carlo and possibilistic approach to uncertainty propagation in event tree analysis. Risk Analysis, 28:1309–1325.

- Bartels, S. F., Chen, H., Wulder, M., and White, J. (2016). Trends in post-disturbance recovery rates of Canada's forests following wildfire and harvest. *Forest Ecology and Management*, 361:194–207. [19](#)
- Baskent, E. and Keles, S. (2005). Spatial forest planning: A review. *Ecological Modelling*, 188:145–173. [86](#)
- Berger, A., Gschwantner, T., McRoberts, R. E., and Schadauer, K. (2014). Effects of measurement errors on individual tree stem volume estimates for the Austrian national forest inventory. *Forest Science*, 60:14–24. [16](#), [26](#)
- Bergeron, Y., Vijayakumar, D. B. I. P., Ouzennou, H., Raulier, F., Leduc, A., and Gauthier, S. (2017). Projections of future forest age class structure under the influence of fire and harvesting: implications for forest management in the boreal forest of eastern Canada. *Forestry*, 90:485–495. [67](#), [80](#), [81](#)
- Bergeron, Y. (2000). Species and stand dynamics in the mixedwoods to Quebec's southern boreal forest. *Ecology*, 81:1500–1516. [19](#)
- Bergeron, Y., Leduc, A., Harvey, B. D., and Gauthier, S. (2000). Natural fire regime: A guide for sustainable management of the Canadian boreal forest. *Silva Fennica*, 36:81–95. [19](#)
- Bernier, P. Y., Gauthier, S., Pierre-Olivier, J., Manka, F., Boulanger, Y., Beaudoin, A., and Guindon, L. (2016). Mapping local effects of forest properties on fire risk across Canada. *Forests*, 7:1–11. [80](#)
- Bhat, C. and Sener, I. (2009). A copula-based closed-form binary logit choice model for accomodating spatial correlation across observational units. *J. Geogr Syst*, 11:243–272. [63](#)
- Bokalo, M., Stadt, K., Comeau, P., and Titus, S. (2013). The validation of the Mixedwood Growth Model (MGM) for use in forest management decision making. *Forests*, 4:1–27. [13](#)
- Bose, A., Harvey, B., Brais, S., Beaudet, M., and Leduc, A. (2013). Constraint to partial cutting in the boreal forest of Canada in the context of natural disturbance-based management: a review. *Forestry*, 87:11–28. [62](#)
- Bouchard, M. and Kneeshaw, D. and Messier, C. (2007). Forest dynamics following spruce budworm outbreaks in the northern and southern mixedwoods of central Quebec. *Canadian Journal of Forest Research*, 37:763–772. [20](#)
- Bouchard, M. and Pothier, D. (2011). Long-term influence of fire and harvesting on boreal forest age structure and forest composition in eastern Québec. *Forest*

- Ecology and Management, 261:811–820. [66](#)
- Boucher, D., Arsenaut, D., Sirois, L., and Blais, L. (2009a). Logging pattern and landscape changes over the last century at the boreal and deciduous forest transition in eastern Canada. Landscape Ecology, 24:171–184. [20](#)
- Boucher, D., de Grandpré, L., and Gauthier, S. (2009b). La forêt préindustrielle du Bas-Saint-Laurent et sa transformation (1820-2000) : implications pour l'aménagement écosystémique. The Forestry Chronicle, 79:318–328. [20](#), [30](#), [72](#)
- Boucher, D., de Grandpré, L., and Gauthier, S. (2003). Développement d'un outil de classification de la structure des peuplements et comparaison de deux territoires de la pessière à mousses du Québec. The Forestry Chronicle, 79:318–328. [19](#)
- Boulanger, Y. and Arseneault, D. (2004). Spruce budworm outbreaks in eastern Quebec over the last 450 years. Canadian Journal of Forest Research, 34:1035–1043. [19](#), [20](#), [72](#), [74](#), [81](#)
- Boulanger, Y., Arseneault, D., Morin, H., Jardon, Y., Bertrand, P., and Dagneau, C. (2012). Dendrochronological reconstruction of spruce budworm (*Choristoneura fumiferana*) outbreaks in southern Quebec for the last 400 years. Canadian Journal of Forest Research, 42:1264–1276. [67](#)
- Boulanger, Y., Gray, D. R., Cooke, B. J., and Degrandpré (2016). Model-specification uncertainty in future forest pest outbreak. Global Change Biology, 22:1595–1607. [80](#)
- Boulay, E. (2013). Ressources et industries Forestières - Portrait statistique - Édition 2013. Technical report, Ministère des Ressources naturelles, Gouvernement du Québec, Canada. [50](#)
- Boulay, E. (2015). Ressources et industries Forestières - Portrait statistique - Édition 2015. Technical report, Ministère des Forêts, de la Faune et des Parcs, Gouvernement du Québec, Canada. [50](#)
- BoWang, C. and Gadow, K. (2006). Timber harvest planning with spatial objectives, using the method of simulated annealing. European Journal of Forest Research, 121:25–34. [63](#)
- Brandt, J., Flannigan, M., Maynard, D., Thompson, I., and Volney, W. (2013). An introduction to Canada's boreal zone: ecosystem processes, health, sustainability, and environmental issues. Environmental Reviews, 21:207–226. [66](#)
- Breidenbach, J., Antón-Fernández, C., McRoberts, R., and Astrup, R. (2014). Quantifying the model-related variability of biomass stock and change estimates in the norwegian national forest inventory. Forest Science, 60:25–33. [26](#), [27](#), [41](#), [44](#), [66](#),

- [80](#), [86](#), [89](#), [126](#)
- Bugmann, H. (2001). A review of forest gaps models. Climatic Change, 51:259–305. [90](#), [127](#)
- Bugmann, H., Lindner, M., Lasch, P., Flechsig, M., Ebert, B., and Cramer, W. (2000). Scaling issues in forest succession modelling. Climatic Change, 44:265–289. [40](#)
- Candau, J.-N. and Fleming, R. (2005). Landscape-scale spatial distribution of Spruce Budworm defoliation in relation to bioclimatic conditions. Canadian Journal of Forest Research, 35:2218–2232. [81](#)
- Cencerrado, A., Artés, T., Cortés, A., and Margalef, T. (2015). Relieving uncertainty in forest fire spread prediction by exploiting multicore architectures. Procedia Computer Science, 51:1752–1761. [80](#)
- Ciais, P., Schelhaas, M. J., Zaehle, S., Piao, S. L., Cescatti, A., Liski, J., Luysaert, S. Le-Maire, G., Schulze, E.-D., Bouriaud, O., Freibauer, A., Valentini, R., and Nabuurs, G. J. (2008). Carbon accumulation in European forests. Nature Geoscience, 1:425–429. [26](#)
- Cohen, W., Yang, Z., Stehman, S., Schroeder, T., Bell, D., Masek, J., Huang, C., and Meigs, G. (2016). Forest disturbance across the conterminous united states from 1985–2012: The emerging dominance of forest decline. Forest Ecology and Management, 360:242–252. [19](#)
- Condés, S. and McRoberts, R. (2017). Updating national forest inventory estimates of growing stock volume using hybrid inference. Forest Ecology and Management, 400:48–57. [15](#), [17](#), [27](#), [124](#)
- Corona, P., Fattorini, L., Franceschi, S., Scrinzi, G., and C., T. (2014). Estimation of standing wood volume in forest compartments by exploiting airborne laser scanning information: model-based, design-based, and hybrid perspectives. Canadian Journal of Forest Research, 44:1303–1311. [16](#), [17](#), [27](#), [66](#), [70](#), [124](#)
- Daniel, C., Ter-Mikaelian, M., Wotton, B., Rayfield, B., and Fortin, M.-J. (2017). Incorporating uncertainty into forest management planning: Timber harvest, wild-fire and climate change in the boreal forest. Forest Ecology and Management, 400:542–554. [66](#), [82](#), [86](#), [93](#)
- Descôteaux, D. and Martin, P. (2009). Désunis dans l’adversité: Les consommateurs Américains et le conflit du bois d’œuvre entre le canada et les États-Unis. Revue Études Internationales, 40:373–394. [50](#)
- Dettinger, M. and Wilson, J. (1981). First order analysis of uncertainty in numerical

- models of groundwater flow part 1. mathematical development. Water Resources Research, 17:149–161. [41](#)
- Didion, M., Fortin, M.-J., and Fall, A. (2007). Forest age structure as indicator of boreal forest sustainability under alternative management and fire regimes: A landscape level sensitivity analysis. Ecological Modelling, 200:45–58. [66](#)
- Dittrich, R., Wreford, A., and Moran, D. (2016). A survey of decision-making approaches for climate change adaptation: Are robust methods the way forward? Ecological Economics, 122:79–89. [93](#)
- Dong, L., Bettinger, P., Z Liu, Z., and Qin, H. (2015). A comparison of a neighborhood search technique for forest spatial harvest scheduling problems: A case study of the simulated annealing algorithm. Forest Ecology and Management, 356:124–135. [46](#)
- Duchesne, L. and Ouimet, R. (2008). Population dynamics of tree species in southern Quebec, Canada: 1970–2005. Forest Ecology and Management, 255:3001–3012. [20](#)
- Dufour-Kowalski, S., Courbaud, B., Dreyfus, P., Meredieu, C., and de Coligny, F. (2012). Capsis: an open software framework and community for forest growth modelling. Annals of Forest Science, 69:221–233. [23](#), [32](#), [74](#)
- Dupuis, S., Arseneault, D., and Sirois, L. (2011). Change from pre-settlement to present-day forest composition reconstructed from early land survey records in eastern quebec, canada. Journal of Vegetation Science, 22:1–12. [20](#)
- Eyvindson, K. and Kangas, A. (2016). Evaluating the required scenario set size for stochastic programming in forest management planning: incorporating inventory and growth model uncertainty. Canadian Journal of Forest Research, 46:340–347. [40](#), [82](#), [93](#)
- Eyvindson, K. and Kangas, A. (2017). Comment on a simple way to incorporate uncertainty and risk into forest harvest scheduling. Forest Ecology and Management, 386:86–91. [82](#)
- Firth, D. (1993). Bias reduction of maximum likelihood estimates. Biometrika, 80:27–38. [47](#)
- Fischer, R. et al. (2014). Lessons learned from applying a forest gap model to understand ecosystem and carbon dynamics of complex tropical forests. Ecological Modelling, 326:139–152. [40](#)
- Food and Agriculture Organization of the United Nations (2010). Global Forest Resources Assessment. Food and Agriculture Organization of the United Nations.

[17](#), [124](#)

- Fortin, M. (2013). Population-averaged predictions with generalized linear mixed-effects models in forestry: an estimator based on gauss-hermite quadrature. Canadien Journal of Forest Research, 43:129–138. [55](#)
- Fortin, M. (2014). Using a segmented logistic model to predict trees to be harvested in forest growth forecasts. Forest Systems, 23:139–152. [29](#), [46](#), [68](#)
- Fortin, M., Bernier, S., Saucier, J.-P., and Labbé, F. (2009). Une relation hauteur-diamètre tenant compte de l'influence de la station et du climat pour 20 espèces commerciales du Québec. Direction de la recherche forestière, Ministère des Ressources naturelles et de la Faune du Québec, Mémoire de recherche forestière 153. [22](#)
- Fortin, M., Bédard, S., Deblois, J., and Meunier, S. (2008). Predicting individual tree mortality in northern hardwood stands under uneven-aged management in southern Québec, Canada. Annals of Forest Science, 65:205. [90](#), [127](#)
- Fortin, M. and Deblois, J. (2007). Modeling tree recruitment with zero-inflated models: The example of hardwood stands in southern Québec, Canada. Forest Science, 53:529–539. [14](#)
- Fortin, M., Deblois, J., Bernier, S., and Blais, G. (2007). Mise au point d'un tarif de cubage général pour les forêts québécoises: une approche pour mieux évaluer l'incertitude associée aux prévisions. The Forestry Chronicle, 754:754–765. [22](#)
- Fortin, M., Delisle-Boulianne, S., and Pothier, D. (2013). Considering spatial correlations between binary response variables in forestry: an example applied to tree harvest modeling. Forest Science, 59:253–260. [46](#)
- Fortin, M. and Langevin, L. (2010). ARTEMIS-2009: un modèle de croissance basé sur une approche par tiges individuelles pour les forêts du Québec, Canada. Direction de la recherche forestière, Ministère des Ressources naturelles et de la Faune du Québec, Mémoire de recherche forestière 156. [17](#), [21](#), [22](#), [28](#), [29](#), [68](#), [70](#), [124](#)
- Fortin, M. and Langevin, L. (2012). Stochastic or deterministic single-tree models: is there any difference in growth predictions? Annals of Forest Science, 69:271–282. [28](#), [29](#), [40](#), [68](#), [88](#)
- Fortin, M., Manso, R., and Calama, R. (2016). Hybrid estimation based on mixed-effects models in forest inventories. Canadian Journal of Forest Research, 46:1310–1319. [17](#), [27](#), [37](#), [66](#), [87](#), [124](#)
- Fortin, M., Manso, R., and Schneider, R. (2018). Parametric bootstrap estimators

- for hybrid inference in forest inventories. *Forestry*, 91:354–365. [16](#), [17](#), [23](#), [27](#), [32](#), [33](#), [37](#), [38](#), [41](#), [44](#), [70](#), [71](#), [78](#), [79](#), [82](#), [87](#), [88](#), [96](#), [99](#), [100](#), [124](#)
- Fox, J., Ades, P., and Bi, H. (2001). Stochastic structure and individual-tree growth models. *Forest Ecology and Management*, 154:261–276. [40](#)
- Fraver, S., White, A. S., and Seymour, R. S. (2009). Natural disturbance in an old-growth landscape of northern Maine, USA. *Journal of Ecology*, 97:289–298. [18](#), [19](#)
- Fu, Y., Lei, Y., Zeng, W., Hao, R., Zhang, G., Zhong, Q., and Xu, M. (2017). Uncertainty assessment in aboveground biomass estimation at the regional scale using a new method considering both sampling error and model error. *Canadian Journal of Forest Research*, 47:1095–1103. [15](#), [17](#), [41](#), [86](#), [94](#), [124](#)
- Gagnon, L., St-Hilaire, G., and Rioux, M. (2015). *Sommaire du plan d’aménagement forestier intégré tactique du Bas-Saint-Laurent, Québec, Canada*. Ministère des Ressources naturelles, Direction générale du Bas-Saint-Laurent. [72](#)
- Gaudreau, J., Perez, L., and Drapeau, P. (2016). BorealFireSim: A GIS-based cellular automata model of wildfires for the boreal forest of Quebec in a climate change paradigm. *Ecological Informatics*, 32:12–27. [14](#)
- Gauthier, S., Bernier, P., Kuuluvainen, T., Shvidenko, A. Z., and Schepaschenko, D. G. (2015). Boreal forest health and global change. *Science*, 349:819–822. [66](#)
- Gertner, G. (1990). The sensitivity of measurement error in stand volume estimation. *Canadian Journal of Forest Research*, 20:800–804. [15](#), [16](#)
- Gertner, G. Z. (1984). Control of sampling error and measurement error in a horizontal point cruise. *Canadian Journal of Forest Research*, 14:40–43. [86](#)
- Gray, D. R. (2017). Quantifying the sources of epistemic uncertainty in model predictions of insect disturbances in an uncertain climate. *Annals of Forest Science*, 74:1–13. [80](#), [81](#), [82](#), [87](#)
- Gregoire, T. and Valentine, H. (2008). *Sampling strategies for natural resources and the environment*. Chapman & Hall/CRC, New York. [79](#), [89](#), [126](#)
- Groen, T., Verkerk, P., Bottcher, H., Grassi, G., et al. (2013). What causes differences between national estimates of forest management carbon emissions and removals compared to estimates of large-scale models? *Environmental Science & Policy*, 33:222–232. [26](#)
- Grondin, P., Saucier, J.-P., et al. (2009). *Manuel de foresterie*, chapitre Écologie forestière, pages 169–291. Ordre des ingénieurs forestiers du Québec. [19](#), [69](#)
- Hallegatte, S. (2009). Strategies to adapt to an uncertain climate change. *Global*

- Environmental Change, 19:240–247. [93](#)
- Healey, S., Patterson, P., Saatchi, S., Lefsky, M., Lister, A., and Freeman, E. (2012). A sample design for globally consistent biomass estimation using lidar data from the geoscience laser altimeter system (GLAS). Carbon Balance and Management, 7:1–9. [27](#)
- Heaps, T. (2015). Convergence of optimal harvesting policies to a normal forest. Journal of Economic Dynamics and Control, 54:74–85. [61](#)
- Helton, J., Johnson, J., Sallaberry, C., and Storlie, C. (2006). Survey of sampling-based methods for uncertainty and sensitivity analysis. Reliability Engineering and System Safety, 91:1175–1209. [15](#)
- Hernandez, M., Gómez, T., Molina, J., León, M., and Caballero, R. (2014). Efficiency in forest management: A multiobjective harvest scheduling model. Journal of Forest Economics, 20:236–251. [46](#)
- Holopainen, M., Mäkinen, A., Rasinmäki, J., Hyytiäinen, K., Bayazidi, S., and Pietilä, I. (2010). Comparison of various sources of uncertainty in stand-level net present. Forest Policy and Economics, 12:377–386. [13](#), [86](#)
- Horemans, J., Bosela, M., Dobor, L., Barna, M., Bahyl, J., Deckmyn, G., Fabrika, M., Sedmak, R., and Ceulemans, R. (2016). Variance decomposition of predictions of stem biomass increment for european beech: Contribution of selected sources of uncertainty. Forest Ecology and Management, 361:46–55. [67](#)
- Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. Journal of American Statistical Association, 47:663–685. [96](#)
- IPCC (2013). Climate Change 2013: the Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the International Panel on Climate Change, chapter Annex I: Atlas of Global and Regional Climate Projections. Cambridge University Press, Cambridge, United Kingdom. [73](#)
- Jensen, J. L. W. V. (1906). Sur les fonctions convexes et les inégalités entre les valeurs moyennes. Acta mathematica, 30:175–193. [40](#), [88](#)
- Jr. Hamilton, D. (1986). A logistic model of mortality in thinned and unthinned mixed conifer stands of northern idaho. Forest Science, 32:989–1000. [39](#)
- Kangas, A. (1996). Small-area estimates using model-based methods. Canadian Journal of Forest Research, 26:758–766. [40](#)
- Kangas, A. (1998). Uncertainty in growth and yield projections due to annual variation of diameter growth. Forest Ecology and Management, 108:223–230. [37](#),

[39](#), [40](#)

- Kangas, A. (1999). Methods for assessing uncertainty of growth and yield predictions. Canadian Journal of Forestry Research, 29:1357–1364. [15](#), [26](#), [27](#), [40](#), [67](#)
- Kangas, A. and Kangas, J. (2004). Probability, possibility and evidence: approaches to consider risk and uncertainty in forestry decision analysis. Forest Policy and Economics, 6:169–188. [15](#), [86](#)
- Kangas, J., Kangas, A., Leskinen, P., and Pykalainen, J. (2001). MCDM methods in strategic planning of forestry on state-owned lands in finland: Applications and experience. Journal of Multi-Criteria Decision Analysis, 10:257–271. [61](#)
- Kneeshaw, D. and Bergeron, Y. (1998). Canopy gap characteristics and tree replacement in the southeastern boreal forests. Ecology, 79:783–794. [19](#)
- Kneeshaw, D., Harvey, B., Reyes, G., Caron, M.-N., and Barlow, S. (2011). Spruce budworm, windthrow and partial cutting: Do different partial disturbances produce different forest structures? Forest Ecology and Management, 262:482–490. [20](#), [79](#)
- Kozak, A. and Kozak, R. (2003). Does cross validation provide additional information in the evaluation of regression models? Canadien Journal of Forest Research, 33:976–987. [62](#)
- Laiberté, J., Pothier, D., and Achim, A. (2016). Adjusting harvest rules for red oak in selection cuts of Canadian northern hardwood forests. Forestry, 89:402–411. [29](#)
- Lawless, J. (2003). Statistical Models and Methods for Lifetime Data. John Wiley and Sons, Hoboken. [47](#), [51](#), [52](#), [53](#), [61](#), [87](#)
- Liebhold, A., Koenig, W. D., and Bjørnstad, O. N. (2004). Spatial synchrony in population dynamics. Annual Review of Ecology, Evolution, and Systematics, 35:467–490. [67](#)
- Liu, G., Han, S., Zhao, X., Nelson, J., Wang, H., and Wang, W. (2006). Optimisation algorithms for spatially constrained forest planning. Ecological Modelling, 194:421–428. [61](#)
- Maier, H., Guillaume, J., van Delden, H., Riddell, G., Haasnoot, M., and Kwakkel, J. (2016). An uncertain future, deep uncertainty, scenarios, robustness and adaptation: How do they fit together? Environmental Modelling & Software, 81:154–164. [83](#)
- Majcen, Z., Bédard, S., and Godbout, C. (2003). La forêt feuillue du Québec et la recherche en sylviculture. Technical report, Note de recherche forestière

- Gouvernement du Québec, Canada. [62](#)
- Makinen, A., Kangas, A., and Nurmi, M. (2012). Using cost-plus-loss analysis to define optimal forest inventory interval and forest inventory accuracy. Silva Fennica, 46:211–226. [82](#)
- Manso, R., Morneau, F., Ningre, F., and Fortin, M. (2015). Incorporating stochasticity from extreme climatic events and multi-species competition relationships into single-tree mortality models. Forest Ecology and Management, 354:243–253. [47](#), [90](#), [127](#)
- Martell, D., Gunn, E., and Weintraub, A. (1998). Forest management challenges for operational researchers. European Journal of Operational Research, 104:1–17. [46](#)
- McCullagh, P. (2008). Sampling bias and logistic models. Journal of Royal Statistic Society, 70:643–677. [47](#)
- McCulloch, C., Searle, S., and Neuhaus, J. M. (2008). Generalized, linear, and mixed models. John Wiley & Sons, New York. [55](#)
- McRoberts, R., Chen, Q., Domke, G., Stahl, G., Saarela, S., and Westfall, J. (2016). Hybrid estimators for mean aboveground carbon per unit area. Forest Ecology and Management, 378:44–56. [27](#), [88](#)
- McRoberts, R., Gobakken, T., and Naesset, E. (2012). Post-stratified estimation of forest area and growing stock volume using lidar-based stratifications. Remote Sensing Environmental, 128:157–166. [79](#)
- McRoberts, R. and Westfall, J. (2014). Effects of uncertainty in model predictions of individual tree volume on large area volume estimates. Forest Science, 60:34–42. [17](#), [27](#), [41](#), [44](#), [66](#), [80](#), [86](#), [88](#), [89](#), [124](#), [126](#)
- McRoberts, R. and Westfall, J. (2016). Propagating uncertainty through individual tree volume model predictions to large-area volume estimates. Annals of Forest Science, 73:625–633. [17](#), [67](#), [80](#), [98](#), [124](#)
- Melo, L., Schneider, R., and Fortin, M. (2018). Quantifying model- and sampling-related uncertainty in large area growth predictions. Ecological Modelling, Submitted:1. [79](#), [104](#)
- Melo, L., Schneider, R., Manso, R., Saucier, J.-P., and Fortin, M. (2017). Using survival analysis to predict the harvesting of forest stands in Quebec, Canada. Canadian Journal of Forest Research, 47:1357–1364. [68](#), [78](#)
- MFFP (2003). Manuel d'aménagement forestier. Technical report, Ministère des Forêts, de la Faune et des Parcs, Gouvernement du Québec, Canada. [13](#), [50](#), [68](#)
- MFFP (2009). L'inventaire écoforestier du Québec méridional - une approche par

- peuplement écoforestier. Technical report, Ministère des Forêts, de la Faune et des Parcs, Gouvernement du Québec, Canada. [68](#)
- MFFP (2013). Manuel de détermination des possibilités forestières 2013-2018. Technical report, Bureau du forestier en chef, Gouvernement du Québec, Canada. [19](#), [50](#), [63](#)
- MFFP (2014). Réseaux des placettes-échantillons permanentes du Québec méridional. Technical report, Direction des inventaires forestiers, Ministère des Forêts, de la Faune et des Parcs, Gouvernement du Québec, Canada. [21](#), [48](#)
- MFFP (2015a). Le 5^{ème} inventaire écoforestier du Québec méridional - Présentations des orientations. Technical report, Direction des Inventaires Forestiers, Ministère des Forêts, de la Faune et des Parcs, Gouvernement du Québec, Canada. [21](#), [22](#)
- MFFP (2015b). Norme de stratification écoforestière - le 4^{ème} inventaire écoforestier du Québec méridional. Technical report, Ministère des Forêts, de la Faune et des Parcs, Gouvernement du Québec, Canada. [21](#)
- MFFP (2015c). Rapport d'activité 2014-2015. Technical report, Ministère des Forêts, de la Faune et des Parcs, Gouvernement du Québec, Canada. [68](#)
- MFFP (2016). Niveaux supérieurs du système hiérarchique de classification écologique. Direction des Inventaires Forestiers, Ministère des Forêts, de la Faune et des Parcs du Québec. [19](#), [72](#)
- MFFP (2017). Insectes, maladies et feux dans les forêts du Québec en 2016. Technical report, Direction de la protection des Forêts - Ministère des Forêts, de la Faune et des Parcs, Gouvernement du Québec, Canada. [91](#), [128](#)
- Miina, J. (1996). Optimizing thinning and rotation in a stand of pinus sylvestris on a drained peatland site. Scandinavian Journal of Forest Research, 11:182–192. [46](#)
- MNR (1998). Guide des saine pratiques forestières dans les pentes du Québec. Technical report, Ministère des Ressources naturelles, Gouvernement du Québec, Canada. [62](#)
- Morin, H. (1994). Dynamics of balsam fir forests in relation to Spruce Budworm outbreaks in the boreal zone of Quebec. Canadian Journal of Forest Research, 24:730–741. [19](#)
- Mäkinen, A. (2010). Uncertainty in forest simulators and forest planning systems. PhD thesis, University of Helsinki. [13](#), [14](#), [15](#), [40](#)
- Nelson, J. and Finn, S. (1990). The influency of cut-block size and adjacency rules on harvest levels and road networks. Canadian Journal of Forest Research, 21:595–

600. [61](#)

- Paré, D., Gertner, G., Bernier, P., and Yanai, R. (2016). Quantifying uncertainty in forest measurements and models: approaches and applications. Canadian Journal of Forest Research, 46:v. [14](#), [17](#), [26](#), [124](#)
- Parent, B. (1988). Ressources et industries Forestières. Technical report, Ministère des Ressources naturelles, Gouvernement du Québec, Canada. [50](#)
- Parent, B. (1990). Ressources et industries Forestières. Technical report, Ministère des Ressources naturelles, Gouvernement du Québec, Canada. [50](#)
- Parent, B. (1992). Ressources et industries Forestières. Technical report, Ministère des Ressources naturelles, Gouvernement du Québec, Canada. [50](#)
- Parent, B. (1993). Ressources et industries Forestières. Technical report, Ministère des Ressources naturelles, Gouvernement du Québec, Canada. [50](#)
- Parent, B. (1994). Ressources et industries Forestières. Technical report, Ministère des Ressources naturelles, Gouvernement du Québec, Canada. [50](#)
- Parent, B. (1996). Ressources et industries Forestières. Technical report, Ministère des Ressources naturelles, Gouvernement du Québec, Canada. [50](#)
- Parent, B. (1999). Ressources et industries Forestières - Portrait statistique - Edition 1999. Technical report, Ministère des Ressources naturelles, Gouvernement du Québec, Canada. [50](#)
- Parent, B. (2009). Ressources et industries Forestières - Portrait statistique - Edition 2009. Technical report, Ministère des Ressources naturelles, de la Faune, Gouvernement du Québec, Canada. [50](#)
- Parent, B. (2010). Ressources et industries Forestières - Portrait statistique - Edition 2010. Technical report, Ministère des Ressources naturelles, de la Faune, Gouvernement du Québec, Canada. [50](#)
- Parent, B. and Fortin, C. (2000). Ressources et industries Forestières - Portrait statistique - Edition 2000. Technical report, Ministère des Ressources naturelles, Gouvernement du Québec, Canada. [50](#)
- Parent, B. and Fortin, C. (2002). Ressources et industries Forestières - Portrait statistique - Edition 2002. Technical report, Ministère des Ressources naturelles, Gouvernement du Québec, Canada. [50](#)
- Parent, B. and Fortin, C. (2003). Ressources et industries Forestières - Portrait statistique - Edition 2003. Technical report, Ministère des Ressources naturelles, de la Faune et des Parcs, Gouvernement du Québec, Canada. [50](#)
- Parent, B. and Fortin, C. (2004). Ressources et industries Forestières. Technical

- report, Ministère des Ressources naturelles, Gouvernement du Québec, Canada. [50](#)
- Parent, B. and Fortin, C. (2005). Ressources et industries Forestières - Portrait statistique - Edition 2004. Technical report, Ministère des Ressources naturelles, de la Faune, Gouvernement du Québec, Canada. [50](#)
- Parent, B. and Fortin, C. (2006). Ressources et industries Forestières - Portrait statistique - Edition 2005-2006. Technical report, Ministère des Ressources naturelles, de la Faune, Gouvernement du Québec, Canada. [50](#)
- Parent, B. and Fortin, C. (2007). Ressources et industries Forestières - Portrait statistique - Edition 2007. Technical report, Ministère des Ressources naturelles, de la Faune, Gouvernement du Québec, Canada. [50](#)
- Parent, B. and Fortin, C. (2008). Ressources et industries Forestières - Portrait statistique - Edition 2008. Technical report, Ministère des Ressources naturelles, de la Faune, Gouvernement du Québec, Canada. [50](#)
- Parent, G., Boulay, E., and Fortin, C. (2012). Ressources et industries Forestières - Portrait statistique - Edition 2012. Technical report, Ministère des Ressources naturelles, de la Faune, Gouvernement du Québec, Canada. [50](#)
- Pasalodos-Tato, M., Makinen, A., Garcia-Gonzalo, J., Borges, J., Lamas, T., and Eriksson, L. (2013). Assessing uncertainty and risk in forest planning and decision support systems: review of classical methods and introduction of innovative approaches. *Forest Systems*, 22:282–303. [82](#)
- Payette, S., Fillion, L., and Delwaide, A. (1990). Disturbance regime in a cold temperate forest as deduced from tree-ring patterns: The Tantaré ecological reserve, Quebec. *Canadian Journal of Forest Research*, 20:1228–1241. [20](#)
- Payette, S., Pilon, V., Couillard, P.-L., and Frégeau, M. (2016). Long-term fire history of maple (acer) forest sites in the central St. Lawrence Lowland, Quebec. *Canadian Journal of Forest Research*, 46:822–831. [19](#)
- Pedroni, N. and Zio, E. (2012). Empirical comparison of methods for the hierarchical propagation of hybrid uncertainty in risk assessment, in presence of dependences. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 20:509–557. [16](#)
- Peng, C. (2000). Understanding the role of forest simulation models in sustainable forest management. *Environmental Impact Assessment Review*, 20:481–501. [26](#)
- Petr, M., Boerboom, L., Ray, D., and van der Veen, A. (2014). An uncertainty assessment framework for forest planning adaptation to climate change. *Forest*

- Policy and Economics, 41:1–11. [15](#)
- Pham, A. T., de Grandpré, L., Gauthier, S., and Bergeron, Y. (2004). Gap dynamics and replacement patterns in gaps of the northeastern boreal forest of Quebec. Canadian Journal of Forest Research, 34:353–364. [19](#)
- Phillips, D., Brown, S., Schroeder, P., and Birdsey, R. (2000). Toward error analysis of large-scale forest carbon budgets. Global Ecology and Biogeography, 9:305–313. [26](#)
- Pichancourt, J.-B., Manso, R., Albrecht, A., and Fortin, M. (2018). Are fossil-fuel to wood-based substitution policies that good to reach the french grand-est lulufc greenhouse gas emission targets? European Carbon Forest Policy, In prep.:1. [80](#)
- Pinheiro, J. and Bates, D. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. Journal of Computational and Graphical Statistics, 4:12–35. [55](#)
- Pinheiro, J. and Bates, D. (2000). Mixed effects models in S and S-PLUS. Springer, New York. [54](#)
- Poirier, C., Deschênes, J., Gosselin, C., Hardy, L., Lamoureux, J., and D., T. (2013). Portrait territorial Bas-Saint-Laurent - édition 2010. Technical report, Direction générale du Bas-Saint-Laurent, Ministère des Ressources naturelles, Gouvernement du Québec, Canada. [30](#)
- Porté, A. and Bartelink, H. (2002). Modelling mixed forest growth: a review of models for forest management. Ecological Modelling, 150:141–188. [12](#), [13](#), [26](#)
- Pothier, D. and Auger, I. (2011). NATURA-2009 : un modèle de prévision de la croissance à l'échelle du peuplement pour les forêts du Québec. Direction de la recherche forestière, Ministère des Ressources naturelles et de la Faune du Québec, Mémoire de recherche forestière 163. [12](#)
- Pretzch, H., Grote, R., Reineking, B., Rotzer, T., and Seifert, S. (2008). Models for forest ecosystem management: A european perspective. Annals of Botany, 101:1065–1087. [12](#), [13](#), [27](#)
- Pretzsch, H. (2009). Forest dynamics, growth and yield. From measurement to model. Spriger-Verlag, Berlin. [12](#), [26](#)
- Pukkala, T., Miina, J., Kurttila, M., and Kolstrom, T. (1998). A spatial yield model for optimizing the thinning regime of mixed stands of pinus sylvestris and picea abies. Scandinavian Journal of Forest Research, 13:31–42. [46](#)
- Rahman, O. and Devadoss, S. (2002). Economics of the US–Canada softwood lumber dispute: A historical perspective. Journal of International Law and Trade Policy,

- 52:29–45. [50](#)
- Railsback, S. F. (2001). Concepts from complex adaptative systems as a framework for individual-based modelling. Ecological Modelling, 139:47–62. [96](#)
- Rao, C. R. and Toutenburg, H. (1995). Linear Models - Least squares and alternatives. Springer series in Statistics, New York. [69](#)
- Refsgaard, J., van der Sluijs, J., Højberg, A., and Vanrolleghem, P. (2007). Uncertainty in the environmental modelling process - a framework and guidance. Environmental Modelling and Software, 22:1543–1556. [16](#), [66](#)
- Régnière, J., Saint-Amant, R., and Béchard, A. (2010). BioSIM 2010. Canadian Forest Service. [22](#), [28](#), [70](#)
- Robinson, A. P., McLarin, M., and Moss, I. (2016). A simple way to incorporate uncertainty and risk into forest harvest scheduling. Forest Ecology and Management, 359:11–18. [82](#)
- Rose, C. E., Hall, D. B., Shiver, D. B., Clutter, M. L., and Border, B. (2006). A multilevel approach to individual tree survival prediction. Forest Science, 52:31–43. [47](#), [90](#), [127](#)
- Rowe, J. S. (1972). Forest regions of Canada - n 1300. Technical report, Canadian Forestry Service, Ottawa. [20](#)
- RStudio Team (2015). RStudio: Integrated Development Environment for R. [23](#)
- Rubin, D. B. (1987). Multiple Imputation for Nonresponse in Surveys. John Wiley & Sons, New Jersey. [88](#), [98](#)
- Rubinstein, R. and Kroese, D. (2007). Simulation and the Monte Carlo Method. John Wiley & Sons, New Jersey. [16](#), [22](#), [29](#), [68](#), [71](#)
- Ruel, J. J. and Ayres, M. P. (1999). Jensen’s inequality predicts effects of environmental variation. TREE, 14:361–366. [40](#)
- Saarela, S., Schnell, S., Grafstrom, A., Tuominen, S., Nordkvist, K., Hyypä, J., Kangas, A., and Stahl, G. (2015). Effects of sample size and model form on the accuracy of model-based estimators of growing stock volume. Canadian Journal of Forest Research, 45:1524–1534. [27](#), [87](#)
- Sample, V. A. (2004). Sustainability in forestry: Origins, evolution and prospects. Technical report, Pinchot Institute for Conservation. [46](#)
- SAS (2015). SAS/STAT 14.1 User’s Guide. SAS Institute Inc., Cary, North Carolina, USA. [54](#)
- Saucier, J.-P., Bergeron, J.-F., Grondin, P., and Robitaille, A. (2015). Les régions écologiques du Québec Méridional: un des éléments du système hiérarchique de

- classification écologique du territoire mis au point par le ministère des ressources naturelles. Technical report, Ministère des Forêts, de la Faune et des Parcs, Gouvernement du Québec, Canada. [70](#)
- Saucier, J.-P., Robitaille, A., and Grondin, P. (2009). Cadre bioclimatique du Québec. Éditions Multimondes, Québec, Canada. [30](#), [49](#)
- Schelhaas, M.-J., Nabuurs, G.-J., and Schuck, A. (2003). Natural disturbances in the European forests in the 19th and 20th centuries. Global Change Biology, 9:1620–1633. [66](#)
- Schmidt, M., Hanewinkel, M., Kändler, G., Kublin, E., and Kohnle, U. (2010). An inventory-based approach for modeling single-tree storm damage — experiences with the winter storm of 1999 in southwestern germany. Canadian Journal of Forest Research, 40:1636–1652. [81](#)
- Seidl, R., Fernandes, P., Fonseca, T., et al. (2011). Modelling natural disturbances in forest ecosystems: a review. Ecological Modelling, 222:903–924. [66](#)
- Service, C. F. (1995). Criteria and Indicators for the Conservation and Sustainable Management of Temperate and Boreal Forests. Canadian Forest Service. [17](#), [124](#)
- Sexton, J., Noojipady, P., Anand, A., Song, W.-P., et al. (2015). A model for the propagation of uncertainty from continuous estimates of tree cover to categorical forest cover and change. Remote Sensing of Environment, 156:418–425. [26](#)
- Sheil, D. and May, R. M. (1996). Mortality and recruitment rate evaluations in heterogeneous tropical forests. Journal of Ecology, 84:91–100. [27](#)
- Soares, P. and Tomé, M. (2007). Model evaluation: From model components to sustainable forest management indicators. Cuadernos de la Sociedad Española de Ciencias Forestales, 23:27–34. [13](#)
- Stadt, K. J., Huston, C., Coates, K. D., Feng, Z., Dale, M. R. T., and Lieffers, V. J. (2007). Evaluation of competition and light estimation indices for predicting diameter growth in mature boreal mixed forests. Annals of Forest Science, 64:477–490. [69](#)
- Sterba, H., Golser, M., Moser, M., and Schadauer, K. (2000). A timber harvesting model for austria. Computers and Electronics in Agriculture, 28:133–149. [46](#), [62](#)
- Ståhl, G., Heikkinen, J., Petersson, H., Repola, J., and Holm, S. (2014). Sample-based estimation of greenhouse gas emissions from forests - a new approach to account for both sampling and model errors. Forest Science, 60:3–13. [17](#), [27](#), [41](#), [44](#), [80](#), [86](#), [124](#)
- Ståhl, G., Holm, S., Gregoire, T. G., Gobakken, T., Næsset, E., and Nelson, R.

- (2011). Model-based inference for biomass estimation in a lidar sample survey in hedmark county, norway. Canadian Journal of Forest Research, 41:96–107. [27](#), [87](#)
- Ståhl, G., Saarela, S., Schnell, S., Holm, S., Breidenbach, J., Healey, S. P., Patterson, P., Magnussen, S., Naesset, E., McRoberts, R. E., and Gregoire, T. G. (2016). Use of models in large-area forest surveys: comparing model-assisted, model-based and hybrid estimation. Forest Ecosystems, 3:1–11. [27](#), [66](#), [87](#)
- Taylor, A. R., Hart, T., and Chen, H. (2013). Tree community structural development in young boreal forests: A comparison of fire and harvesting disturbance. Forest Ecology and Management, 310:19–26. [19](#)
- Thurnher, C., Klopf, M., and Hasenauer, H. (2011). Forests in transition: a harvesting model for uneven-aged mixed species forests in Austria. Forestry, 84:517–526. [14](#), [46](#), [47](#), [61](#), [86](#)
- Turner, M. G. (2010). Disturbance and landscape dynamics in a changing world. Ecology, 91:2833–2849. [66](#)
- United Nations Conference on Environment and Development (2010). Agenda 21, Rio. Declaration of Forest Principles. United Nations Framework Convention on Climate Change. [17](#), [124](#)
- Valle, D., Schulze, M., Vidal, E., Grogan, J., and Sales, M. (2006). Identifying bias in stand-level growth and yield estimations: A case study in eastern Brazilian Amazonia. Forest Ecology and Management, 236:127–135. [78](#)
- Vanclay, J. (1994). Modelling forest growth and yield: applications to mixed tropical forests. Southern Cross University, CAB International. [12](#), [37](#), [40](#)
- Vanclay, J. and Skovsgaard, J. (1998). Evaluating forest growth models. Ecological Modelling, 98:1–12. [13](#), [26](#)
- Venier, L., Thompson, I., Fleming, R., et al. (2014). Effects of natural resource development on the terrestrial biodiversity of Canadian boreal forests. Environmental Review, 22:457–490. [66](#)
- Verstegen, J. A. (2016). Quantifying and reducing uncertainty in land use change model projections. PhD thesis, Utrecht University. [16](#)
- Walker, W., Harremoes, P., Rotmans, J., Van Der Sluijs, J., Van Asselt, M., Janssen, P., and Krayen Von Krauss, M. (2003). Defining uncertainty: A conceptual basis for uncertainty management in model-based decision support. Integrated Assessment, 4:5–17. [15](#), [26](#), [66](#)
- Wang, N., Brown, D., Yang, S., and Ligmann-Zielinska, A. (2013). Comparative performance of logistic regression and survival analysis for detecting spatial pre-

- dictors of land-use change. International Journal of Geographical Information Science, 27:1960–1982. [47](#), [61](#), [86](#)
- Wehrens, R., Putter, H., and Buydens, L. M. C. (2000). The bootstrap: a tutorial. Chemometrics and Intelligent Laboratory Systems, 54:35–52. [87](#)
- Willems, P. (2012). Model uncertainty analysis by variance decomposition. Physics and Chemistry of the Earth, 42-44:21–30. [23](#), [41](#), [93](#)
- Willet, J. and Singer, J. (1993). Investigating onset, cessation, relapse, and recovery: Why you should and how you can use discrete-time survival analysis to examine event occurrence. Journal of consulting and clinical psychology, 61:952–965. [52](#)
- Williams, D. and Liebhold, A. (2000). Spatial synchrony of spruce budworm outbreaks in Eastern North America. Ecology, 81:2753–2766. [67](#), [80](#)
- Wilson, B. M. and Smith, B. L. (2013). Taylor-series and Monte-Carlo-method uncertainty estimation of the width of a probability distribution based on varying bias and random error. Measurements Science and Technology, 24:1–12. [41](#), [81](#)
- Wu, J., Jones, K., Li, H., and Loucks, O. (2006). Scaling and Uncertainty Analysis in Ecology. Springer, Dordrecht, The Netherlands. [23](#), [40](#), [67](#), [74](#), [81](#), [88](#)
- Xu, C. and Gertner, G. (2008). Uncertainty and sensitivity analysis for models with correlated parameters. Reliability Engineering and System Safety, 93:1563–1573. [41](#), [67](#)
- Xu, C., Gertner, G., and Scheller, R. (2009). Uncertainties in the response of a forest landscape to global climatic change. Global Change Biology, 15:116–131. [15](#)
- Yanai, R., Levine, C., Green, M., and Campbell, J. (2012). Quantifying uncertainty in forest nutrient budgets. Journal of Forestry, 110:448–456. [15](#)
- Yang, Y., Yanai, R., Fatemi, F., Levine, C., Lilly, P., and Briggs, R. (2016). Sources of variability in tissue chemistry in northern hardwood species. Canadian Journal of Forest Research, 46:285–296. [40](#)
- Zhou, M. and Buongiorno, J. (2004). Nonlinearity and noise interaction in a model of forest growth. Ecological Modelling, 180:291–304. [40](#)

Abstract

Resumé complet

Au cours des dernières décennies, la mise en place des stratégies de gestion durable des forêts et les rapports nationaux sur le carbone ont exigé des informations, des critères et des règlements qui soutiennent les accords internationaux (par exemple, [Food and Agriculture Organization of the United Nations, 2010](#); [Service, 1995](#); [United Nations Conference on Environment and Development, 2010](#)). Ceci a mené à une demande croissante des prévisions des ressources forestières futures à grande échelle. Ces prévisions sont obtenues en utilisant des modèles de croissance, qui sont adaptés aux données des échantillons et non à un recensement complet ([Condés and McRoberts, 2017](#)). La mise en échelle de ces modèles locaux pour obtenir des prévisions à une échelle plus grande a des implications pratiques en termes d'incertitudes et doit suivre un schéma d'inférence particulier. L'inférence hybride est un concept qui incorpore des incertitudes tant du modèle que de l'échantillonnage ([Corona et al., 2014](#)). Récemment, quelques études ont fourni des exemples d'estimation des incertitudes dans le contexte d'inférence hybride pour la foresterie (par exemple., [Corona et al., 2014](#); [Fortin et al., 2016, 2018](#); [Fu et al., 2017](#); [McRoberts and Westfall, 2014, 2016](#); [Ståhl et al., 2014](#)).

Estimer ces incertitudes devient nécessaire face à ces demandes de prévisions de croissance à grande échelle. La prise en compte de ces incertitudes et des risques associés peut être avantageuse sur une vaste gamme de décisions dans la foresterie. La négligence de ces incertitudes peut mener à une mauvaise évaluation des risques. Elle peut, par exemple, avoir un impact direct sur les choix d'alternatives de gestion, du marché de produit forestier ou des projets sur le carbone ([Paré et al., 2016](#)). L'importance de ces questions a été l'élément moteur derrière cette étude.

Cette thèse de doctorat est centrée sur l'estimation des incertitudes dans les prévisions de la croissance régionale en utilisant un schéma d'inférence hybride. Les effets des perturbations à grande échelle ont aussi été considérées. Nous avons abordé ces questions par trois objectifs principaux, chacun d'entre eux constituant un chapitre de la thèse : (i) estimer les incertitudes résultant du modèle et de l'échantillonnage dans les prévisions de la croissance régionale; (ii) développer un modèle de récolte basé sur l'analyse de survie pour prévoir la probabilité de récolte sous un scénario "business as usual"; et (iii) analyser l'effet des perturbations à grande échelle sur les incertitudes dans les prévisions de la croissance régionale. Le modèle de croissance ARTEMIS-2009 ([Fortin and Langevin, 2010](#)), qui s'applique aux principaux types forestiers du Québec, Canada, a été utilisé comme étude de

cas. Pour chaque objectif principal, nous avons défini quelques volets spécifiques, décrits ci-dessous :

(i) En considérant les sources d'incertitude majeures dans les prévisions à grande échelle, notre objectif spécifique était d'estimer la contribution du modèle et de l'échantillonnage sur la variance totale des prévisions de la croissance régionale. Nous avons aussi évalué la contribution des différents sous-modèles qui composent le modèle de croissance à la variance totale. L'analyse combine l'horizon de temps et les types forestiers comme des facteurs influençant le comportement de la variance.

(ii) Les techniques d'analyse de survie présentent une flexibilité structurelle qui permet de tenir compte des variables régionales et des variables qui évoluent en fonction du temps, dont nous assumons l'influence sur l'occurrence de récolte. L'objectif spécifique était de développer un modèle de survie pour prévoir les probabilités de récolte au niveau de la placette. Le modèle devrait être intégré dans ARTEMIS-2009, afin de considérer les pratiques actuelles de gestion sur l'ensemble du territoire. Une telle intégration a été nécessaire pour répondre à l'objectif suivant.

(iii) La présence des perturbations majeures augmente les incertitudes de prévisions de la croissance. Ce chapitre a exploré l'effet des perturbations anthropique (la récolte) et naturelle (la tordeuse des bourgeons d'épinette, TBE) sur les variances des prévisions de la croissance régionale. Cette dernière analyse a permis de déterminer la perturbation qui a contribué le plus aux variances. Un important but de ce chapitre était de fournir des directives quant à la façon de réduire ces variances.

Ces objectifs spécifiques portent sur les enjeux actuels de la modélisation de la croissance forestière. Ils sont structurés en articles autonomes numérotés, appelés "Article". Voici la liste des publications :

Article I. Melo, L.C., Schneider, R., and Fortin, M. (2018). Estimating model- and sampling-related uncertainty in large-area growth predictions. *Ecological Modelling*. 390: 62-69.

Article II. Melo, L.C., Schneider, R., Manso, R. and Fortin, M. (2017). Using survival analysis to predict the harvesting of forest stands in Quebec, Canada. *Canadian Journal of Forest Research*. 47: 1066–1074.

Article III. Melo, L.C., Schneider, R., and Fortin, M. (2018). The effect of natural and anthropogenic disturbances on the uncertainty of large-area forest growth forecasts. *Environmental Modelling and Software* *Submitted*.

Concernant les résultats, les variances liées au modèle et à l'échantillonnage ont été évaluées en termes de surface terrière et d'écotype pour la région du Bas-Saint-Laurent (Article I). La variance de l'échantillonnage s'est révélée être une source importante d'incertitude en prédisant la croissance sur une grande échelle, ce qui a corroboré les résultats d'études précédentes (par exemple, [Breidenbach et al., 2014](#); [McRoberts and Westfall, 2014](#)). Cependant, nous avons observé que ceci était vrai seulement pour des prévisions à court terme. Des prévisions sur la croissance à long terme (par exemple, 100 ans) ont montré que les valeurs des variances liées aux modèles étaient semblables à celles de l'échantillonnage. Étant donné les tendances observées, nous pourrions supposer que pour des prévisions plus longues que 100 ans, le modèle deviendrait le contributeur le plus important. De plus, pour des grandes échelles, la variance liée à l'échantillonnage devrait probablement être plus petite en raison d'une probable taille d'échantillons plus grande. Dans ce cas, la variance liée au modèle pourrait devenir la source principale d'incertitude, même dans une prévision à moyen terme (par exemple, 50 ans).

Dans l'article III, nous avons étudié la contribution des perturbations en termes d'incertitudes dans les prévisions du volume pour la région du Bas-Saint-Laurent. La variance liée au modèle s'est révélée être le contributeur le plus important de la variance totale avant la fin de l'horizon de 100 ans. Pour les scénarios qui ont considéré les épidémies de TBE, la contribution du modèle à la variance totale était déjà plus grande que celle de l'échantillonnage après 30 ans. Dans un scénario où seulement les activités de récolte ont été incluses, la variance liée au modèle est devenu le contributeur le plus important après 70 ans. À cette échelle régionale, notre taille d'échantillon était quatre fois plus grande que les tailles d'échantillons dans l'article I. Néanmoins, la variance d'échantillonnage, que l'on s'est attendu être plus petit, est restée constante. Malgré les tailles d'échantillons plus petites, les trois populations dans l'article I ont été moins variables à cause de la stratification basée sur les écotypes. Nous avons posé l'hypothèse que cette stratification pourrait réduire les variances liées à l'échantillonnage de ces prévisions régionales ([Gregoire and Valentine, 2008](#), p.127).

La perspective de l'incertitude liée au modèle a été étudiée plus loin dans l'article I en décomposant la variance totale. Parmi les sous-modèles qui composent ARTEMIS-2009, la mortalité était le composant qui a induit la plus grande part de la variance. En désactivant la stochasticité du sous-modèle de mortalité, la variance totale a diminué de 40% à 60%. Il y a des implications majeures liées aux prédictions de

mortalité dans des modèles de croissance. Par exemple, les arbres prévus comme étant mort par erreur dans un pas de croissance n'ont aucune chance de devenir vivants dans les pas suivants. Ceci peut mener à une sous-estimation impactant les profits forestiers. Le sous-modèle d'accroissement en diamètre avait une contribution maximale de 25% sur la variance totale. Bien que beaucoup de modèles de mortalité existent ([Bugmann, 2001](#)), nos résultats suggèrent que le sous-modèle de mortalité doit être priorisé dans nos efforts.

Le potentiel des techniques d'analyse de survie pour modéliser la récolte a été présenté dans l'article II. Les prévisions annuelles d'occurrence de récolte ont été obtenues par un modèle de survie ayant une composante temporelle. Dans cette partie du modèle, la variable évoluant dans le temps et à l'échelle régionale (possibilité forestière, AAC), ainsi que l'effet aléatoire représentant la corrélation spatiale a résulté à une amélioration de l'ajustement du modèle. Modéliser la récolte en considérant ce variable évoluant dans le temps a été une contribution originale de ce papier. En combinant ce modèle de récolte avec un modèle de croissance, les prévisions permettrait de considérer une gestion des forêts de type "business as usual", et pourrait être utile dans les calculs d'AAC. Un inconvénient du modèle de récolte vient du fait que l'AAC a été considérée comme constante dans le temps et, en conséquence, les calendriers de récolte restent constants dans le futur. L'intégration de ce changement dans le modèle devrait être considéré afin d'améliorer les calculs d'AAC.

Puisque les données de mortalité sont aussi considérées comme censurées, un chemin prometteur serait l'utilisation d'analyse de survie pour modéliser la mortalité. La flexibilité de la technique permettrait d'aborder quelques limitations des modèles de mortalité. Par exemple, une interaction existant entre la mortalité des arbres et des perturbations peut être traitée par une analyse de survie (par exemple, [Manso et al., 2015](#)). La modélisation de la mortalité par l'analyse de survie a aussi été faite par [Rose et al. \(2006\)](#) et [Fortin et al. \(2008\)](#). Cependant, aucune de ces trois études mentionnées ci-dessus n'a exploré le composant évoluant dans le temps. Par exemple, des variables climatiques, connues pour influencer la probabilité de mortalité, peuvent être abordées par cette partie temporelle du modèle.

L'article III s'est concentré sur l'estimation de la contribution des perturbations à la variance totale de la prévision de la croissance. Plus spécifiquement, les contributions des épidémies de TBE, de la récolte et de l'échantillonnage ont été comparées. Le caractère stochastique de l'épidémie de TBE augmente la variance de 80% après 30 ans. À la fin de l'horizon temporel, la variance liée au modèle a représenté 91% de

la variance totale. À la lumière de ce résultat, il peut être raisonnablement assumé que d'autres perturbations naturelles comme des feux et des tempêtes de vents ont aussi un impact important sur la variance. La variance estimée sur les activités de récolte a légèrement diminué, probablement parce que la récolte a tendance à homogénéiser des forêts.

Cette perspective proposée dans l'article III a souligné la contribution du modèle à la variance totale dans des prévisions de la croissance à grande échelle. D'un point de vue pratique, telles incertitudes peuvent avoir des implications importantes. Par exemple, la dernière épidémie de TBE dans la région Bas-Saint-Laurent avait affecté un total de 1.1 millions ha en 2016, représentant 48% de la surface totale de la région (MFFP, 2017). Considérant les pertes potentielles de valeur commerciale causée par l'épidémie de TBE, il y aurait un avantage dans l'estimation et la communication d'incertitudes dans des prévisions de croissance liées à cette perturbation. Cette connaissance pourra aider des gestionnaires à ajuster leurs pratiques dans un proche avenir.

Dans l'article III, la récurrence des épidémies de TBE représentait une variable exogène dans les simulations de croissance. Un sous-modèle tenant compte de la dynamique des insectes pourrait être intégré au modèle de croissance, ce qui permettrait que ces occurrences fassent partie du système. L'occurrence de TBE deviendrait alors une variable endogène.

L'inclusion de l'épidémie de TBE dans des prévisions de croissance forestières devrait prendre en compte le processus cyclique et les interactions avec d'autres agents de perturbation. Par exemple, l'épidémie de TBE a un effet indirecte sur des probabilités de récolte. Aussitôt que la sévérité d'une épidémie augmente, les décideurs promeuvent des opérations de récolte. Dans un tel cas, le modèle de récolte devrait aussi considérer cette réponse, ce qui n'a pas été pris en compte dans nos simulations.

Finalement, afin de réduire les incertitudes dans la croissance prédite à court terme, il est nécessaire d'augmenter les tailles d'échantillon et d'améliorer les designs d'échantillonnage. L'utilisation d'estimateurs stratifiés pourrait aider à réduire la variance liée à l'échantillonnage pour des prévisions de grande échelle, mais dans notre étude de cas, cela reste à être testé. Pour réduire des incertitudes à moyen et long terme, la variance liée au modèle devrait être la cible d'amélioration principale. De plus, nous avons conclu que les stratégies de gestion ou les occurrences de mortalité peuvent changer pendant les intervalles sur lesquels les données sont collectées ou que les prévisions sont effectuées. Que ce soit pour modéliser la récolte ou la

mortalité, on doit permettre aux variables variant dans le temps d'évoluer durant ces intervalles. Ceci fournit la flexibilité nécessaire pour surmonter des limitations associés aux modèles précédentes, comme l'intégration de changements des variables climatiques. Finalement, les incertitudes liées aux prévisions de croissance ont semblé être grandement affectées par les perturbations naturelles. L'amélioration de la prévision de l'occurrence de ces perturbations devrait être priorisée, en travaillant notamment avec de meilleures données ou en améliorant les développements des modèles.

Abstract

In forestry, tree-level growth models provide predictions of forest dynamics and thereby, they support decision-making. Although they are widely used, the uncertainty of their predictions is rarely assessed. Understanding the sources of uncertainty and estimating their impact is an essential step forward in a period where large-scale forecasts are becoming more popular. This thesis addresses the issue of uncertainty estimation in regional growth forecasts. The effects of large-scale disturbances were also studied.

The growth model ARTEMIS-2009, which applies to most forest types in Quebec, Canada, was taken as a case study. A bootstrap hybrid estimator was used to estimate the model- and the sampling-related variances. The total variance was then decomposed to determine which model component induced the greatest share of variance in the forecasts. Then, the survival analysis approach was used to develop a harvest model based on plot and regional variables. This model was integrated into ARTEMIS so that harvesting combined with spruce budworm (SBW) outbreaks were accounted in the simulations. Then, their contributions in terms of uncertainty were estimated.

The results revealed that the sampling accounted for most of the variance in short-term forecasts. In long-term forecasts, the model contribution turned out to be as important as that of the sampling. The variance decomposition per model component indicated that the mortality sub-model induced the highest variability in the forecasts. A great deal of uncertainty was induced by the natural disturbances when they were accounted for in the projections. In particular, SBW showed to be the most important source of uncertainty compared to harvest activities and sampling. In the light of these results, our recommendations are that the effort to reduce uncertainty should focus on the sampling in short-term forecasts, and on the mortality sub-model and SBW occurrence in mid- and long-term forecasts.

Keywords: Monte Carlo simulations, lifetime analysis, hybrid inference, disturbances, harvest, forest management.

Resumo

Modelos de crescimento de árvores individuais fornecem previsões da dinâmica florestal, e portanto, permitem orientar as tomadas de decisões. Apesar dessa importância, as incertezas associadas às suas previsões são pouco conhecidas. Entender as fontes dessas incertezas e estimar seu impacto é essencial nesse momento onde as projeções de larga escala estão ganhando popularidade. Essa tese de doutorado trata das estimativas de incertezas relacionadas à projeção do crescimento florestal em uma escala regional. Os efeitos de perturbações de grande escala foram igualmente tratados.

O modelo de crescimento ARTEMIS-2009, que se aplica aos principais tipos florestais do Québec, Canadá, foi utilizado como estudo de caso. Um estimador híbrido do tipo bootstrap foi utilizado para estimar as variâncias do modelo e da amostragem. A variância total das projeções foi igualmente decomposta de modo à determinar qual sub-modelo induziu a maior variabilidade nas projeções. Em seguida, a abordagem da análise de sobrevida foi utilizada para desenvolver um modelo de probabilidade de colheita, que leva em consideração as variáveis da parcela e em escala regional. Este modelo de colheita foi integrado em ARTEMIS. À partir disso, a contribuição das perturbações em termos de incerteza foi levada em consideração nas simulações. Especificamente, nos concentramos na colheita e na epidemia de lagarta-dos-pinheiros (TBE).

Os resultados revelaram que a amostragem foi responsável pela maior parte da variância das projeções de curto prazo. Nas projeções de longo prazo, a contribuição do modelo foi tão importante quanto a da amostragem. A decomposição da variância pelo sub-modelo indicou que o sub-modelo de mortalidade foi o que induziu a maior variabilidade nas projeções. Quando as perturbações foram considerados nas projeções, a incerteza associada às projeções aumentou acentuadamente. Em particular, a TBE mostrou-se a principal fonte de incerteza quando comparada às atividades de colheita e amostragem. Através desses diferentes resultados, a recomendação final é que precisamos trabalhar no erro de amostragem para reduzir a incerteza em projeções do crescimento de curto prazo, e nos sub-modelo de mortalidade e ocorrência de perturbação natural nas projeções do crescimento de médio a longo prazo.

Palavras-chave: Simulação Monte Carlo, análise de sobrevida, inferência híbrida, perturbações, colheita, manejo florestal.

Résumé

Les modèles de croissance forestière à l'échelle de l'arbre fournissent des prévisions de la dynamique des forêts et permettent donc d'orienter la prise de décision. Malgré cette importance, les incertitudes qui sont associées à leurs prévisions sont rarement évaluées. Comprendre les sources de ces incertitudes et en estimer l'impact sont essentiels au moment où les prévisions à grande échelle gagnent en popularité. Cette thèse de doctorat traite de l'estimation des incertitudes des prévisions de la croissance à une échelle régionale. Les effets de perturbations à large échelle sont aussi traitées.

Le modèle de croissance ARTEMIS-2009, qui s'applique aux principaux types forestiers du Québec, Canada, a été utilisé comme étude de cas. Un estimateur hybride de type bootstrap a été utilisé pour estimer les variances provenant du modèle et de l'échantillonnage. La variance totale des prévisions a aussi été décomposée afin de déterminer quel sous-modèle induisait la plus grande part de variabilité dans les prévisions. Ensuite, une approche d'analyse de durée de vie a été utilisée pour élaborer un modèle d'occurrence de récolte qui tient compte de variables de placettes et de variables régionales. Ce modèle de récolte a été intégré à ARTEMIS de façon à pouvoir tenir en compte de l'effet de la récolte combinée aux épidémies de tordeuse des bourgeons de l'épinette (TBE). Après cela, la contribution des perturbations en termes d'incertitudes a été prise en considération dans les simulations.

Les résultats ont révélé que l'échantillonnage représentait la plus grande partie de la variance des prévisions à court terme. Dans les prévisions à long terme, la contribution du modèle s'est révélée aussi importante que celle de l'échantillonnage. La décomposition de la variance par sous-modèle a indiqué que le sous-modèle de mortalité était celui qui induisait la plus grande variabilité dans les projections. Lorsque les perturbations ont été prises en compte dans les projections, l'incertitude associée aux prévisions augmentait fortement. En particulier, la TBE s'est révélée être la principale source d'incertitude devant les activités de récolte et l'échantillonnage. En conclusion, notre recommandation est de concentrer les efforts pour réduire l'incertitude sur l'échantillonnage dans les prévisions à court terme, et sur le sous-modèle de mortalité et sur l'occurrence de la TBE dans les prévisions à moyen et à long terme. Mots-clés: Simulations Monte Carlo, analyse de la durée de vie, inférence hybride, perturbation, récolte, gestion forestière.

Mots-clés: Simulations Monte Carlo, analyse de la durée de vie, inférence hybride, perturbation, récolte, gestion forestière.

