Fitting truncated generalized beta distributions to stem density data from permanent sample plots in Quebec, Canada

Gregory Paradis *

Département des sciences du bois et de la forêt, Faculté de foresterie, de géographie et de géomatique, Pavillon Abitibi–Price, Université Laval, Québec, QC G1K 7P4, Canada.
gregory.paradis.1@ulaval.ca

Luc LeBel

Département des sciences du bois et de la forêt, Faculté de foresterie, de géographie et de géomatique, Pavillon Abitibi–Price, Université Laval, Québec, QC G1K 7P4, Canada.
luc.lebel@sbf.ulaval.ca

September 27, 2016

Abstract

We fit truncated distributions from the generalized beta family to a large dataset of stems inventoried from permanent fixed-area plots in the province of Quebec, Canada. We describe a two-stage parameter-fitting methodology that produces reasonable estimates of parameter estimation error and correlations for input data with bounded domain. We present best-fit results for 10 species groups and 3 cover types.

• why do we do this ?
• what problem needed to be solved ?
• what benefits can we yield from these results ?

~~2~~ 3 contributions

1) visc a connaître la distribution des tiges par groupe d'essences pour trois type de forêt ?

*Corresponding authour. Contact telephone number (418) 456-2208.

2) la méthodologie

3) test 25 familles —

* Général, mentalité afin de montrer notre ouverture…

1 Introduction

Long-term wood supply optimization models typically use a highly aggregated representation of harvesting activities. A typical wood supply model objective function maximizes the sum of merchantable harvest volume induced by applying one of several possible prescriptions to each forest stratum. Hierarchical forest management (HFM) planning should be implemented with effective linkages between planning levels (Paradis et al., 2013). In practice, these linkages are often ineffective or altogether absent. Lower HFM planning levels will typically be modelled on a shorter horizon than wood supply models, using a more detailed representation of reality. Financial performance indicators typically play an important role in short-term planning, whereas financial indicators are typically absent from wood supply models. Linking HFM planning levels, to verify coherence of the overall HFM planning process, requires comparison or decisions expressed at different aggregation levels. In other words, we must be able to *disaggregate* output unit of the upper level model to match the input unit of the lower level model.

This is rather vague.

Defining such disaggregation functions is not trivial—for example, we may need to convert species-group–wise merchantable harvest volume (output from a long-term wood supply model) to species-wise assortments of stems, binned into several diameter at breast height (DBH) size classes. Disaggregating volume into diameter size classes presupposes availability of a stem diameter distribution model that is valid for the forest of interest. The literature contains many examples of stem diameter distribution models [insert appropriate citations here], however the published models are all specific a certain combination of species, stand structure, geographic area, and inventory sampling method. To the best of our knowledge, no diameter distribution modelling methodology has been published that can handle any combination of species and cover type in Quebec, Canada. Furthermore, none of the published studies clearly document how to correctly estimate best-fit parameter uncertainty and correlations for the common case where observed diameter data has *a priori* bounded domain (e.g. only merchantable stems of a certain minimum diameter are inventoried and trees never grow beyond a certain maximum diameter). The present study fills this gap in the literature.

to

This is problem statement, what you attempt to solve? I think we must make a stronger case. Restate ~~what about adding~~ what we / others could do with it?

ok, but at this point the reader does not know why you focus on Québec, Canada — what about else where?

We fit truncated distributions from the generalized beta family to a large dataset of stems from government-compiled permanent fixed-area plots in the province of Quebec, Canada. We describe a two-stage parameter-fitting methodology that correctly handles parameter estimation error and correlations for input data with bounded domain. We present best-fit distributions for 30 combinations of species group and cover type.

The remainder of this paper is organized as follows. We describe our methodology in §2. Results are presented in §3, followed by discussion in §4 and concluding remarks in §5.

2 Methods

Ducey and Gove (2015) document three parent distributions in the generalized beta family that can be used to derive several other distributions. These parent distributions are the generalized beta distribution of the first kind (GB1), the generalized beta distribution of the second kind (GB2), and the generalized gamma distribution (GG). The probability density functions (PDF) of GB1 and GB2 distributions have the following forms (adapted from Ducey and Gove, 2015)

$$\mathrm{GB1}(x; a, b, p, q) = \frac{|a| x^{ap-1} \left[1 - (x/b)^a\right]^{q-1}}{b^{ap} B(p, q)}, \qquad 0 < y^a < b^a, b > 0, p > 0, q > 0 \tag{1}$$

and

$$\mathrm{GB2}(x; a, b, p, q) = \frac{|a| x^{ap-1} x^{q-1}}{b^{ap} B(p, q) \left[1 - (x/b)^a\right]^{p+q}}, \qquad a > 0, b > 0, p > 0, q > 0 \tag{2}$$

defined for $x > 0$, where $\mathrm{B}(p, q)$ represents the beta function (not to be confounded with the beta, or generalized beta, distributions), which is given by

$$\mathrm{B}(p, q) = \int_0^1 t^{p-1}(1-t)^{q-1} dt. \tag{3}$$

The PDF of the generalized gamma GG distribution has the following form

$$\mathrm{GG}(x; a, \beta, p) = \frac{a x^{ap-1} e^{-\left(\frac{x}{\beta}\right)^a}}{\beta^{ap} \Gamma(p)}, \qquad a > 0, \beta > 0, q > 0 \tag{4}$$

defined for $x > 0$, where $\Gamma(p)$ represents the gamma function (not to be confounded with the gamma, or generalized gamma, distributions), which is given by

$$\Gamma(p) = \int_0^\infty x^{p-1} e^{-x} dx. \tag{5}$$

We can define the PDFs for 22 different distributions in the generalized beta family in terms of one of the three parent distributions, as follows (adapted from Ducey and Gove, 2015)

x = ?

a = ?

p = ?

$$\begin{aligned}
\mathrm{IB1}(x;b,p,q) &= \mathrm{GB1}(x;-1,b,p,q) && (6)\\
\mathrm{UG}(x;b,d,q) &= \lim_{a\to\infty} \mathrm{GB1}(x;a,b,d/a,q) && (7)\\
\mathrm{B1}(x;b,p,q) &= \mathrm{GB1}(x;1,b,p,q) && (8)\\
\mathrm{B2}(x;b,p,q) &= \mathrm{GB2}(x;1,b,p,q) && (9)\\
\mathrm{SM}(x;a,b,q) &= \mathrm{GB2}(x;a,b,1,q) && (10)\\
\mathrm{Dagum}(x;a,b,p) &= \mathrm{GB2}(x;a,b,p,1) && (11)\\
\mathrm{Pareto}(x;b,p) &= \mathrm{GB1}(x;-1,b,p,1) && (12)\\
\mathrm{P}(x;b,p) &= \mathrm{GB1}(x;1,b,p,1) && (13)\\
\mathrm{LN}(x;\mu,\sigma) &= \lim_{a\to 0} \mathrm{GG}(x;a,(\sigma^2a^2)^{1/a},(a\mu+1)/(\sigma^2a^2)) && (14)\\
\mathrm{GA}(x;\beta,p) &= \mathrm{GG}(x;1,\beta,p) && (15)\\
\mathrm{W}(x;a,\beta) &= \mathrm{GG}(x;a,\beta,1) && (16)\\
\mathrm{F}(x;u,v) &= \mathrm{GB2}(x;1,v/u,u/2,v/2) && (17)\\
\mathrm{L}(x;b,q) &= \mathrm{GB2}(x;1,b,1,q) && (18)\\
\mathrm{IL}(x;b,p) &= \mathrm{GB2}(x;1,b,p,1) && (19)\\
\mathrm{Fisk}(x;a,b) &= \mathrm{GB2}(x;a,b,1,1) && (20)\\
\mathrm{U}(x;b) &= \mathrm{GB1}(x;1,b,1,1) && (21)\\
\tfrac{1}{2}\mathrm{N}(x;0,\sigma) &= \mathrm{GG}(x;2,\sigma^2,1/2) && (22)\\
\chi^2(x;p) &= \mathrm{GG}(x;1,2,p) && (23)\\
\mathrm{EXP}(x;\beta) &= \mathrm{GG}(x;1,\beta,1) && (24)\\
\mathrm{R}(x;\beta) &= \mathrm{GG}(x;2,\beta,1) && (25)\\
\tfrac{1}{2}\mathrm{t}(x;df) &= \mathrm{GB2}(x;2,\sqrt{df},1/2,df/2) && (26)\\
\mathrm{LL}(x;b) &= \mathrm{GB2}(x;1,b,1,1) && (27)
\end{aligned}$$

We use a weighted non-linear least squares algorithm to fit target distribution parameters to data binned into 26 size classes of uniform width w.

The objective function value of the weighted non-linear least squares (NLSQ) problem minimizes the sum of squares of the residual terms

$$Z\left(f(x;\hat{\boldsymbol{P}})\right) = \min \quad \sum_{i\in I} e\left(f(x_i;\boldsymbol{P}),\hat{y}_i\right)^2 \tag{28}$$

with

$$e\left(f(x_i;\boldsymbol{P}),\hat{y}_i\right) = f(x_i;\boldsymbol{P}) - \hat{y}_i. \tag{29}$$

where x_i is the diameter corresponding to the center of bin $i \in I$, $f(x_i;\boldsymbol{P})$ is the value of the PDF of the target distribution at $x_i \in \boldsymbol{X}$ (given a vector of parameters $\boldsymbol{P}$), and $\hat{y}_i \in \hat{\boldsymbol{Y}}$ is the observed value in bin i.

q cm!

We normalize our binned data, such that $\sum_{i\in I} w\hat{y}_i = 1$. The domain of input data is bounded, such that $a \le x_1 - w/2$ and $x_{|I|} + w/2 \le b$, where $a > 0$ and $b < \infty$. Our inventory dataset only includes merchantable stems (i.e. $a = 9$). Our inventory dataset contains very few stems of diameter greater than 61 cm. Estimation error of stem density for these stems is high, which does not respect the assumption of homogeneous error implicit to the fitting algorithm objective function (28). We therefore exclude these large stems from our analysis (i.e. $b = 61$).

The integral of the standard forms of the PDFs described above over the interval $[0, \infty]$ sum to 1 for any given vector of input parameters $\boldsymbol{P}$, that is

$$\int_0^\infty f(x; \boldsymbol{P}) dx = 1. \tag{30}$$

Fitting the standard forms of f to the normalized binned data will generally produce poor fits, as the sum of residuals will be positively biased (i.e. $\sum_{i\in I} e_i > 1$), with quality of fit inversely proportional to $b - a$. We can obtain a better fit using an augmented PDF $f'(x; \boldsymbol{P'}) = sf(x; \boldsymbol{P})$. The global scaling parameter s effectively relaxes the unity constraint on the integral of f. Using f', we would obtain similar quality fits for any scaling of bin value vector $\hat{\boldsymbol{Y}}$.

The variance $\mathrm{var}(\hat{p}_{p_j})$ of best-fit parameter estimator $\hat{p}_j \in \hat{\boldsymbol{P}}$ corresponds to element j of the diagonal of the covariance matrix. The covariance matrix, which automatically calculated by most software implementations of the NLSQ algorithm, corresponds to the inverse of the negative of the expected values of the Hessian matrix $-E[H(\hat{\boldsymbol{P}})]$, where the Hessian $H(\hat{\boldsymbol{P}})$ is the matrix of second derivatives of the likelihood function $\mathcal{L}$ with respect to $\hat{\boldsymbol{P}}$. Standard error $\sigma_{\hat{p}_j}$ of parameter $\hat{p}_j \in \hat{\boldsymbol{P}}$ is simply the square root of the variance (i.e. $\sigma_{\hat{p}_j} = \sqrt{\mathrm{var}(\hat{p}_{p_j})}$).

Note that variance estimates are only correct asymptotically. In practice, fitting algorithms will use numerical approximations of Hessian matrix values. Quality of finite approximations of the second derivatives of $\mathcal{L}$ will tend to be proportional to sample size $|\hat{\boldsymbol{Y}}|$, inversely proportional to distance from parameter constraint boundaries, and inversely proportional to the number of parameters $|\hat{\boldsymbol{P}}|$.

Parameter estimation error for augmented function $f'(x; \boldsymbol{P'})$ can be improved, without deteriorating fit quality, by solving the fitting problem in two stages. In the first stage, we determine $\hat{\boldsymbol{P'}}$ by solving for $Z(f'(x; \hat{\boldsymbol{P'}}))$. For the best-case scenario, where $f'(x; \boldsymbol{P'})$ is fitted to an infinitely large sample $\hat{\boldsymbol{Y}}$ randomly drawn from $f'(x; \hat{\boldsymbol{P'}})$, the estimated value of scaling parameter $\hat{s} \in \hat{\boldsymbol{P'}}$ will completely eliminate the bias in the sum of residuals (i.e. $\sum_{i\in I} e(f(x_i; \hat{\boldsymbol{P}}), \hat{y}_i) = 0$), such that $\int_a^b f(x; \bar{\boldsymbol{P'}}) dx = \sum_{i\in I} w\hat{y}_i$. In the second stage, we solve for $Z(f''(x; \hat{\boldsymbol{P}}, \hat{s}))$, where f'' corresponds to our augmented distribution f' with the scaling parameter value fixed at $s = \hat{s}$ (i.e. only the original vector of parameters $\boldsymbol{P}$ is optimized by the fitting algorithm).

The shape distributions from both stages are equivalent, such that

$$Z(f'(x;\hat{P}')) \simeq Z(f''(x;\hat{P},\hat{s})). \tag{31}$$

However, error vector $\boldsymbol{\sigma}_{\hat{P}}$ and parameter covariance (which can be estimated from off-diagonal elements of the covariance matrix) estimated in the second stage will tend to be more reliable.

Our computational experiment dataset consists of 52 192 stems extracted from a database of permanent sample plot (PSP) data, collected from public forests in Quebec (Canada). This data was collected by the *Ministère de la forêt, de la faune et des parcs* (MFFP) as part of the official government inventory program[1], which operates on a 10-year cycle.

Data was collected throughout the province of Quebec, using 11.28 meter radius circular fixed-area plots. The full dataset contains 1 685 233 stems, sampled from 12570 permanent sample plot locations. However, this includes repeated measures from four decennial inventory cycles, collected from 7 different PSP networks. We filtered data to include only stems from the most recent inventory cycle, which ensures that we are not tallying repeated measure on the same plots We further filtered data to include only stems from the largest of the 7 PSP networks (codename *BAS1*), which ensures uniform data-collection standards for all stems.

Our ultimate goal (i.e. beyond the scope of this paper) is to link a long-term wood supply optimization model with a short-term fibre-procurement optimization model. Thus, we are interested in modelling diameter distribution of merchantable stems in mature (operable), undisturbed stands. We therefore applied a series of other filters to our stem dataset to exclude plots in disturbed or immature stands, unmerchantable stems (with DBH ¡ 9cm), very large stems (with DBH ¿ 61cm), dead or otherwise unmerchantable stems.

A Jupyter Notebook containing Python code implementing these filter and detailed explanations is available from the corresponding author upon request. Although we do not have permission to distribution the PSP dataset, it is possible to request a copy from the *Ministère des forêts, de la faune et de parcs* (see footnote for URL).

We segmented ~~for~~ the 52 192 stems in our filtered PSP dataset into 30 sub-datasets, representing combinations of 10 species groups and 3 cover types. More detailed information on species groups is provided an appendix. For each of 30 sub-datasets $d \in D$, we ran our two-stage fitting method on each of 22 target distributions $f \in F$. We fit each of the 22 distributions $f \in F$ listed earlier, using ran our two-stage fitting algorithm, to each of the 30 sub-datasets. Akaike information criterion (AIC) and Bayesian information criterion (BIC) goodness-of-fit metric were calculated for each combination of $d \in D$ and $f \in F$. For each sub-dataset $d \in D$, we ranked distributions $f \in F$ in decreasing order of the sum of AIC and BIC, and reported best-fit distribution f^*_d, parameter vector $\hat{P}_{f^*_d}$, AIC, and BIC.

[1]Detailed information on the MFFP inventory program is available from the MFFP web site (`http://www.mffp.gouv.qc.ca/forets/inventaire/`), including technical documentation on inventory methods, data standards, and contact information to request a copy of the data.

3 Results

4 Discussion

5 Conclusion

6 Acknowledgements

This study was supported by funding from the *FORAC Research Consortium.*

References

Ducey, M. J. and Gove, J. H. (2015). Size-biased distributions in the generalized beta distribution family, with applications to forestry. *Forestry*, 88(1):143–151.

Paradis, G., LeBel, L., D'Amours, S., and Bouchard, M. (2013). On the risk of systematic drift under incoherent hierarchical forest management planning. *Canadian Journal of Forest Research*, 43(5):480–492.

Appendix

Table 1 lists common and Latin names of species in the species groups used to segment our PSP data.

Table 1: Mapping of species group names to species common and Latin names. Alternate names are shown in parentheses.

Species Group	Common Name	Latin Name
Other Hardwoods	(white, American) ash	*Fraxinus americana*
Other Hardwoods	black ash	*Fraxinus nigra*
Other Hardwoods	(green, red) ash	*Fraxinus pennsylvanica*
Other Hardwoods	(North) American beech	*Fagus grandifolia*
Other Hardwoods	(American, white, water) elm	*Ulmus americana*
Other Hardwoods	slippery elm	*Ulmus rubra*
Other Hardwoods	(rock, cork) elm	*Ulmus thomasii*
Other Hardwoods	American hophornbeam	*Ostrya virginiana*
Other Hardwoods	American linden (basswood)	*Tilia americana*
White Birch	grey birch	*Betula populifolia*
White Birch	(white, paper) birch	*Beutul papyrifera*
Yellow Birch	yellow birch	*Betala alleghaniensis*
Oak-Hickory	(bitternut, swamp) hickory	*Carya cordiformis*
Oak-Hickory	shagbark hickory	*Carya ovata*
Oak-Hickory	([wild, mountain] black, rum) cherry	*Prunus serotina*
Oak-Hickory	white oak	*Quercus alba*
Oak-Hickory	swamp white oak	*Quercus bicolor*
Oak-Hickory	bur oak	*Quercus macrocarpa*
Oak-Hickory	(northern, eastern) red oak	*Quercus rubra*
Oak-Hickory	(butternut, white walnut)	*Juglans cinerea*
Spruce-Pine-Fir	white spruce	*Picea glauca*
Spruce-Pine-Fir	black spruce	*Picea mariana*
Spruce-Pine-Fir	Norway spruce	*Picea abies*
Spruce-Pine-Fir	red spruce	*Picea rubens*
Spruce-Pine-Fir	hybrid larch	*Larix X marschlinsii*
Spruce-Pine-Fir	Japanese larch	*Larix leptolepis*
Spruce-Pine-Fir	([eastern, American] larch, tamarack)	*Larix larincina*
Spruce-Pine-Fir	European larch	*Larix decidua*
Spruce-Pine-Fir	pitch pine	*Pinus rigida*
Spruce-Pine-Fir	([eastern, black] jack, grey, scrub) pine	*Pinus banksiana*
Spruce-Pine-Fir	Scots pine	*Pinus sylvestris*
Spruce-Pine-Fir	balsam fir	*Abies balsamea*
Other Maples	(silver, silverleaf) maple	*Acer saccharinum*
Other Maples	black maple	*Acer nigrum*
Other Maples	red maple	*Acer rubrum*
Sugar Maple	(sugar, rock) maple	*Acer saccharum*
Poplar	balsam poplar	*Populus balamifera*
Poplar	eastern cottonwood (poplar)	*Populus deltoides*
Poplar	(large-tooth, big-tooth) aspen	*Populus grandidentata*
Poplar	hybrid poplar	*Populus sp X P. sp.*
Poplar	([quaking, trembling] [aspen, poplar]	*Populus tremuloides*
White Pine	white pine	*Pinus strobus*
Red Pine	red pine	*Pinus resinosa*
Hemlock-Cedar	(eastern, Canadian) hemlock	*Tsuga canadensis*
Hemlock-Cedar	(eastern, northern) white-cedar	*Thuja occidentalis*

Compiling disaggregation coefficients to link long- and short-term planning models

Gregory Paradis *

Département des sciences du bois et de la forêt, Faculté de foresterie, de géographie et de géomatique, Pavillon Abitibi–Price, Université Laval, Québec, QC G1K 7P4, Canada.
gregory.paradis.1@ulaval.ca

Luc LeBel

Département des sciences du bois et de la forêt, Faculté de foresterie, de géographie et de géomatique, Pavillon Abitibi–Price, Université Laval, Québec, QC G1K 7P4, Canada.
luc.lebel@sbf.ulaval.ca

October 9, 2016

Abstract

It has been shown that credibility of wood supply optimization models can be improved by using a bilevel formulation that anticipates industrial fibre consumption. The upper level model corresponds to the standard long-term wood supply optimization model, and the lower level corresponds to a short-term network flow optimization model. Linking the two levels requires disaggregation of upper-level output using a matrix of disaggregation coefficients. These disaggregation coefficients are not typically available, and compiling them involves complex manipulations of large amounts of data. We described a methodology for compiling such a matrix of disaggregation coefficients, using readily-available data.

describe output

*Corresponding authour. Contact telephone number (418) 456-2208.

and others

1 Introduction

Long-term wood supply optimization models typically use a highly aggregated representation of harvesting activities. A typical wood supply model objective function maximizes the sum of merchantable harvest volume induced by applying one of several possible prescriptions to each forest stratum. Hierarchical forest management (HFM) planning should be implemented with effective linkages between planning levels (Paradis et al., 2013). In practice, these linkages are often ineffective or altogether absent. Lower HFM planning levels will typically be modelled on a shorter horizon than wood supply models, using a more detailed representation of reality. Financial performance indicators typically play an important role in short-term planning, whereas financial indicators are generally ~~typically~~ absent from wood supply models. Linking HFM planning levels, to verify coherence of the overall HFM planning process, requires comparison or decisions expressed at different aggregation levels. In other words, we must be able to *disaggregate* output unit of the upper level model to match the input unit of the lower level model. We described a methodology for compiling vectors of disaggregation coefficients, using readily-available data.

Our goal is to disaggregate harvest volume output from a wood supply model into discrete 2-centimeter diameter at breast height (DBH) stem size classes. We start with the volume $u_{cst}(Z)$ of species group s harvested from cover type c with a treatment type t, induced by applying solution Z to the wood supply model for a given management unit. If we have a function $p_{cst}(x)$ defining the proportion of each unit of volume of species s, harvested from cover type c with treatment t, which we can assume will be in stems in size class x, we can define a disaggregated volume function $v_{cst}(z,x)$ by taking the dot product of $u_{cst}(z)$ (scalar value) and $p_{cst}(x)$ (vector).

$$v_{cst}(z,x) = u_{cst}(z) \cdot p_{cst}(x), \quad \forall c \in C, \forall s \in S, \forall t \in T \tag{1}$$

where C is the set of cover types, S is the set of species groups, and T is the set of treatment types.

$p_{cst}(x)$ is defined as the dot product of three vector components, as follows

$$p_{cst}(x) = f_{cs}(x) \cdot g_{cst}(x) \cdot h_{cs}(x) \tag{2}$$

$f_{cs}(x)$ defines the probability distribution of stem sizes of standing inventory. We will compile $f_{cs}(x)$ from permanent sample plot data, using the methodology described in Paradis and LeBel (2016).

$g_{cst}(x)$ defines the probability that a stem of species s and size x will be harvested from cover type c under treatment t. We compile $g_{cst}(x)$ using a statistical model published by Fortin (2014).

$h_{cs}(x)$ defines normalized form factor vectors for each combination of cover type $c \in C$ and species $s \in S$. We compile $h_{cs}(x)$ from regional form factor models published by the Quebec government.

give a peek view of results and what is their practical, fondamental value. What can be done w/ them.

The remainder of this paper is organized as follows. We describe our methodology in §2. Results are presented in §3, followed by discussion in §4 and concluding remarks in §5.

2 Methods

We compile disaggregation coefficients for management unit UA 064-51 in Quebec, Canada.

We use the methodology described in Paradis and LeBel (2016) to compile $f_{cs}(x)$. This methodology uses a weighted nonlinear least squares (NLLS) algorithm to find best-fit parameters for a number of candidate distributions, and selects the best distribution based on the small-sample Akaike information criterion (AICc). Our dataset consists of 52 192 stems extracted from a database of PSP data[1]. These stems are filtered from the full PSP dataset to include only live, merchantable stems from the most recent inventory cycle, from mature, undisturbed stands, for which there was valid data in all fields. This dataset was divided into 30 sub-datasets, corresponding to combinations 10 species groups and 3 cover types. We fit generalized gamma (GG), gamma (GA), Weibull (W), and exponential (EXP) distributions to each of these 30 sub-datasets and use the distribution with lowest AICc to compile $f_{cs}(x)$.

Next, we compile $g_{cst}(x)$ using the statistical model published by Fortin (2014) to determine the bin-wise harvest probabilities for selective cutting ($t = 2$) and commercial thinning ($t = 3$) treatment types. All stems in the standing inventory are removed during a final cutting ($t = 1$), so we simply assign a harvest probability of 1 to all bins for this case. Note that we use the post-2004 selection cutting model from Fortin (2014), and that we normalize the vector values, such that $\sum_{x \in X} g_{cst}(x) = 1, \forall c \in C, s \in S, t \in T$.

Fortin (2014) use a species aggregation scheme consisting of 12 species groups. However, we need to compile $g_{cst}(x)$ in terms of our 10 species groups. Using the relative frequency distribution of the 12 Fortin species groups in our permanent sample plot dataset, we generate a set of weight coefficients for each combination of species group s and cover type s. We then use these weight coefficients to compile Fortin model results in terms of our set of species groups S. The Fortin model requires us to provide an estimate of stem density (i.e., stem count per hectare) as one of the input parameters. We use the permanent sample plot data to compile mean stem density estimates for each of the three cover types.

c

Next, we compile $h_c s(x)$. Local form factor models have been compiled by the government authorities for several regions in Quebec[2]. The downloadable

[1]Detailed information on the PSP inventory program under which our test data was collected is available from the MFFP web site (http://www.mffp.gouv.qc.ca/forets/inventaire/), including technical documentation on inventory methods, data standards, and contact information to request a copy of the data.

[2]Detailed information on the form factor model is available from the MFFP web site (https://www.mffp.gouv.qc.ca/forets/inventaire/fiches/tarif-cubage.jsp), including technical documentation and links to download the model.

package includes documentation (PDF format), species-wise regression parameters and form factors for the LIN3 stem volume model (DBF format), and some spatial data (Shapefile format) delineating the boundaries of the local form factor zones. The form factor model is defined for individual species (i.e., 1-to-1 mapping with the species codes in our permanent sample plot dataset). We compile weighted average form factors, for a given management unit, by cover type c and species group s. We aggregate species into species groups using the relative frequency distribution of species in our permanent sample plot dataset as weight coefficients. Several form factor zones can overlap the boundaries of a given management unit. We use a geographic information system to intersect form factor zones and management unit boundary, to determine a set of zone-wise weights that can be used to blend multiples zone models to derive weighted-average form factors for the management unit of interest.

Finally, we compile $p_{cst}(x)$ from the three vector components. We show results for management unit UA 064-51, however our methodology can be applied generically to any of the 71 management units in Quebec using the same input data sources.

A Jupyter Notebook containing Python code implementing our methodology and detailed explanations is available from the corresponding author upon request.

3 Results

Figures 1 and 2 show results of applying our methodology to compile disaggregation functions for management unit UA 064-51 in Quebec, Canada. Species group is fixed for a given row of subfigures, and cover type is fixed for given column of subfigures. Treatment type 1 (solid line) corresponds to clearcut harvesting, treatment type 2 (dashed line) corresponds to selection cut, and treatment type 3 (dotted line) corresponds to commercial thinning.

4 Discussion

5 Conclusion

6 Acknowledgements

This study was supported by funding from the *FORAC Research Consortium*.

References

Fortin, M. (2014). Using a segmented logistic model to predict trees to be harvested in forest growth forecasts. *Forest Systems*, 23(1):139–152.

Paradis, G. and LeBel, L. (2016). Fitting truncated generalized beta distributions to stem density data from permanent sample plots in Quebec, Canada. [unpublished manuscript].

Paradis, G., LeBel, L., D'Amours, S., and Bouchard, M. (2013). On the risk of systematic drift under incoherent hierarchical forest management planning. *Canadian Journal of Forest Research*, 43(5):480–492.

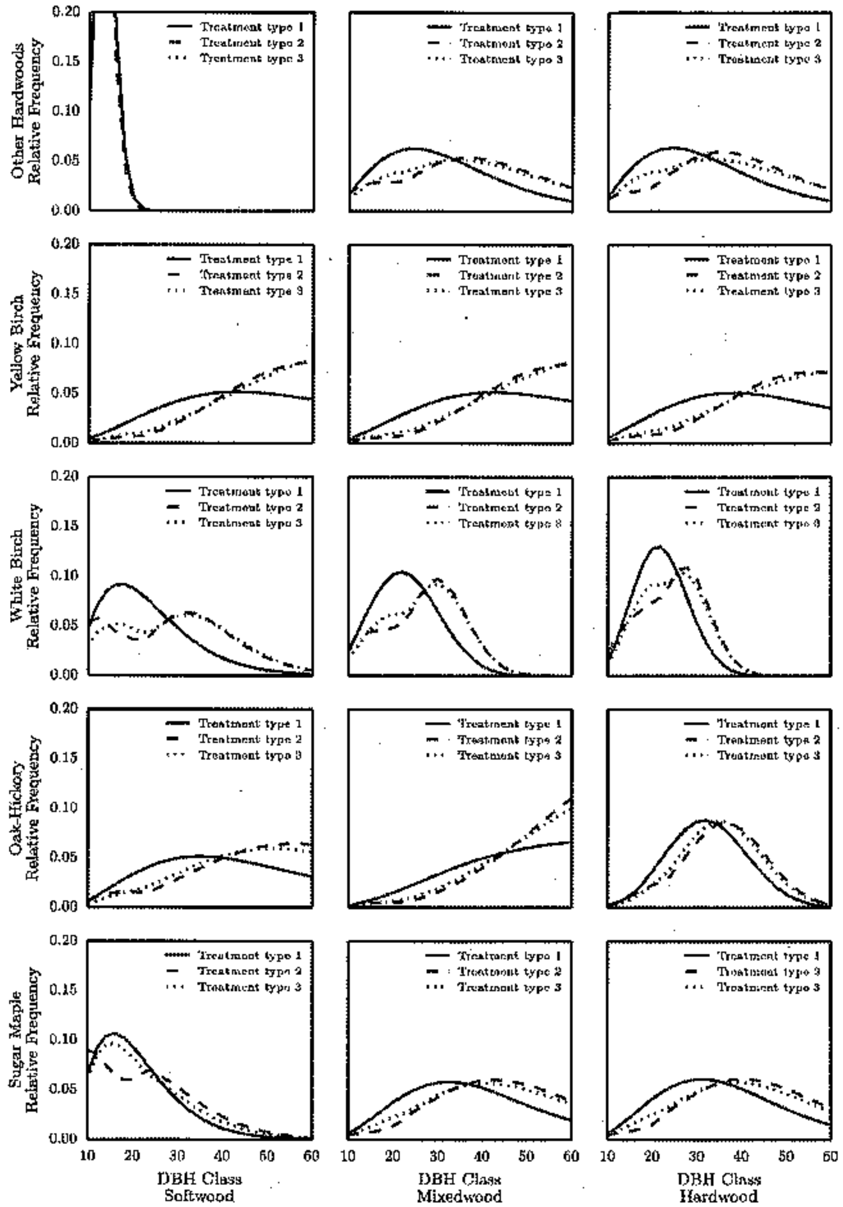

Figure 1: Example of disaggregation functions compiled for management unit UA 064-51 in Quebec, Canada. Species group is fixed for a given row of subfigures, and cover type if fixed for given column of subfigures. Treatment type 1 (solid line) corresponds to clearcut harvesting, treatment type 2 (dashed line) corresponds to selection cut, and treatment type 3 (dotted line) corresponds to commercial thinning.

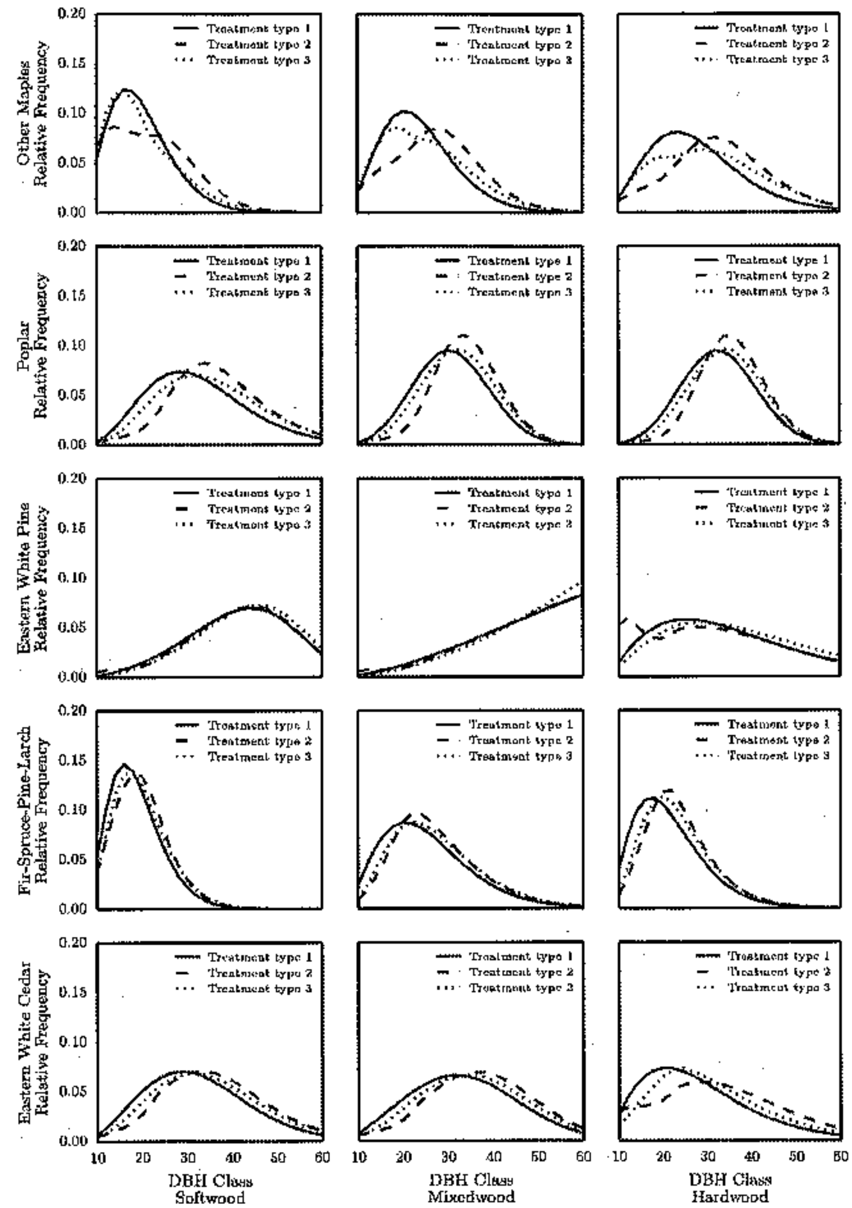

Figure 2: [Continued from Figure 1] Example of disaggregation functions compiled for management unit UA 064-51 in Quebec, Canada. Species group is fixed for a given row of subfigures, and cover type if fixed for given column of subfigures. Treatment type 1 (solid line) corresponds to clearcut harvesting, treatment type 2 (dashed line) corresponds to selection cut, and treatment type 3 (dotted line) corresponds to commercial thinning.

Rapport Technique CIRRELT.

Le titre ne va pas de soi. Il soulève cependant la curiosité...

Retro-fitting value-creation potential indicators to long-term wood supply models

Gregory Paradis *

Département des sciences du bois et de la forêt, Faculté de foresterie, de géographie et de géomatique, Pavillon Abitibi–Price, Université Laval, Québec, QC G1K 7P4, Canada.
gregory.paradis.1@ulaval.ca

Luc LeBel

Département des sciences du bois et de la forêt, Faculté de foresterie, de géographie et de géomatique, Pavillon Abitibi–Price, Université Laval, Québec, QC G1K 7P4, Canada.
luc.lebel@sbf.ulaval.ca

October 10, 2016

Abstract

It has been shown that credibility of wood supply optimization models can be improved by using a bilevel formulation that anticipates industrial fibre consumption. The upper level model corresponds to the standard long-term wood supply optimization model, and the lower level corresponds to a short-term network flow optimization model. The lower-level model maximises profit sale of primary forest products. To compile such a model, we must be able to estimate value-creation potential of harvest volume output from the upper-level model. In Quebec, Canada, a branch of the government responsible for marketing fibre harvested form public forests has compiled a simulation model (MERIS) capable of estimating the value-creation potential of a range of silviculture treatments to a given stand. Input data for MERIS is a stand table, describing stand inventory in terms of the 45 merchantable species codes and 26 stem diameter size classes. We describe a methodology for retro-fitting ~~value-creation potential~~ indicators to the official ~~the~~ wood ~~supply models~~ used to determine annual allowable cut in Quebec, based on information extracted from the database underlying the MERIS model. Our methodology greatly simplifies the otherwise onerous task of compiling bilevel wood supply models from available data, and can be applied to any forest management unit in Quebec.

Where would we get this [illegible] [illegible] published

Explain what is "retrofitting" in the context of this project.

*Corresponding authour. Contact telephone number (418) 456-2208.

1 Introduction

Paradis et al. (2016) describe a bilevel wood supply model formulation that reduces risk of wood supply failures. They demonstrate the potential usefulness of their formulation using a realistic synthetic dataset. Linking upper- and lower-level models involves disaggregating harvest volume from the upper-level model into logs of different size classes, which are then dispatched to different processing units in the lower-level model. The lower-level model must then simulate processing these logs into primary forest products, and finally selling these products to end customers. The objective function of the lower-level model maximised total network profit, which is estimated from the sum of revenue from sale of products net of fibre procurement cost, processing cost, and transportation cost. Implementing their bilevel model in practice therefore presupposed available of a methodology for compiling disaggregation coefficients for upper-level volume output, as well as a methodology for compiling unit value-creation coefficients to estimate unit profit for all possible fibre flow paths through the lower-level network.

A branch of the Quebec government responsible for marketing fibre harvested from public forests (*Bureau de mise en marché des bois*, or BMMB) has published a simulation model (MERIS) capable of estimating the value-creation potential of a range of silviculture treatments, given a stand table describing stand inventory in terms of 10 species groups and 26 stem diameter size classes. ? describe a methodology for compiling disaggregation coefficients that can be used to explode volume from wood supply models into the same 10 species groups and 26 stem diameter size classes used by the MERIS model. Using these disaggregation coefficients, one can potentially map wood supply model output to unit value-creation-potential data in the database underlying the MERIS model, thereby enabling estimation of value-creation-potential of the optimal solution of the wood supply model. However, neither the structure of the wood supply models nor the database underlying MERIS are well documented. The task of linking these models, to compile a bilevel model as described is Paradis et al. (2016), is not trivial. To help overcome this obstacle, we developed a replicable methodology for retro-fitting unit value-creation-potential data from the MERIS database to the wood supply models used to determine annual allowable cut (AAC) in Quebec.

thoroughly ?

Determining AAC in Quebec is the responsibility of the *Bureau du forestier en chef* (BFEC). BFEC use the Woodstock modelling platform to model long-term wood supply. Woodstock does not feature a scripting interface, so any changes to the model must be applied manually (i.e. via keyboard and mouse). Furthermore, Woodstock models in Quebec are automatically generated by an in-house model-compiling interface (Horizon CPF), which results in relatively verbose models. For example, the Woodstock model for management unit UA 064-51, which we use as a test dataset to illustrate our methodology, contains over 600 000 lines of code. The task of retro-fitting value-creation-potential indicators (compiled using our methodology) to these models by manually editing the Woodstock code is too time-consuming to be practical. To make the retro-fitting task more tractable, we implemented a Python software module (which we refer to as *wood supply simulation system*, or WS3) capable of importing and interpreting the code from any Woodstock model (assuming generic model structure generated by Horizon CPF) and simulating application of the optimal solution stored in the model. The WS3 module is highly scriptable, making *post hoc* injection of value-creation-potential (or any other) indicators a manageable task.

Our methodology can be applied with relative ease (compared to an *ad hoc* approach) to any forest management unit in Quebec, thereby making the bilevel wood supply modelling approach described by Paradis et al. (2016) much more accessible. Our methodology could also be used to support compilation of unit value-creation-potential coefficients as input for stand-alone fibre processing network flow optimisation models, for example the LogiLab model described in Jerbi et al. (2012). Alternatively, our methodology could be used to compile *a priori* value-creation indicators for long-term wood supply models (as opposed to *post hoc* injection of these indicators into the optimal solution, as we show here).

Note that we developed this methodology in the context of a larger research agenda, whose goal is to explore innovative business models to realize value-creation-potential from Quebec forests~~—our government and industry partners expect concrete, implementable solutions to relevant problems~~. We mention this to explain why we tailored our methodology so specifically to the MERIS database and the Woodstock model format used by the BFEC in Quebec. Notwithstanding the Quebec-specific details, we hope that the methodology presented here will be a useful framework for researchers and practitioners in other jurisdictions wishing to link long- and short-term models.

As an example, we compile value-creation potential of AAC volume from the optimal solution of the Woodstock model for management unit UA 064-51, in terms of the 10 species groups, 3 cover types, 3 treatment types, and 26 stem size classes used in Paradis and LeBel (2016).

The remainder of this paper is organized as follows. We describe our methodology in §2. Results are presented in §3, followed by discussion in §4 and concluding remarks in §5.

2 Methods

As mentioned in the introduction, neither the code structure of the BFEC Woodstock models nor the MERIS database are very well documented, and both are quite complex. The methodology we developed to link these two models is necessarily complex. We endeavoured to keep the description as short as possible, while providing sufficient detail to facilitate replication of our methods. Neither of these models were designed to be compatible with the other. However, they are both designed specifically to model growth and harvesting of stands from public forests in Quebec, albeit at different scales. Thus, some of the information in both models is *conceptually* compatible, but the data used to

placer comme dernière phrase du paragraphe?

represent this information is stored at different aggregation levels. Many of the data-wrangling challenges we faced while developing the methodology presented here involved interfacing data stored at incompatible aggregation levels, while avoiding unnecessary introduction of noise and bias in the output.

The rest of this section is divided into two subsections. The first subsection provides an overview of the main steps in the methodology, and the second subsection describes each step in more detail.

2.1 Overview

In summary, our methodology extracts information from the MERIS database and *post hoc* injects value-creation-potential performance indicators into the development types harvested in the optimal solution of a Woodstock wood supply model. This methodology represents one phase in a larger modelling process—our ultimate goal is to compile a network flow model that can simulate total network-level value-creation-potential of alternative fibre supply and demand scenarios, then use game-theoretic benefit-sharing strategies to determine the maximal stable subset of potential benefit that can be realized under collaborative planning.

Our methodology can be broken down into two stages. The first stage involves compiling volume disaggregation coefficients, and mapping these coefficients to each decision in the wood supply model optimal solution. The second stage involves importing data from the MERIS database and retro-fitting the wood supply model with value-creation potential performance indicators.

In the first stage of our methodology, we compile 90 disaggregation coefficient vectors using the methodology described in Paradis and LeBel (2016). These vectors of coefficients allow us to disaggregate harvest volume into the same 26 stem size class bins used to store value-creation-potential data in the MERIS database. Each of these vectors maps to one of the 90 cases of an intermediate aggregation scheme. This intermediate aggregation scheme was chosen such that *(a)* it is compatible with the Woodstock model format, *(b)* it is compatible with the MERIS database format, and *(c)* data is available to compile reasonable disaggregation coefficients for each case of this scheme. This intermediate data aggregation level is composed of 90 combinations of 10 species groups (documented in an appendix in Paradis and LeBel (2016)), 3 cover types (softwood, mixedwood, hardwood), and 3 harvest treatment types (clearcut, selection cut, commercial thinning).

In the stage of our methodology, we map value-creation potential data from the MERIS database onto our disaggregated volumes, and re-aggregate the data to produce the value-creation indicators we need for subsequent phases of our fibre supply modelling project. The MERIS database actually contains two distinct value-creation models. The first value model in MERIS represents financial value-creation potential, from the perspective of an industrial facility that procures raw fibre from public forests, transforms this fibre into one or more primary forest products (and co-products, such as chips), and sells these products to end-customers in external markets at exogenously-defined prices.

This is the model we use, when importing data from MERIS. Henceforth, we will be referring to this model, unless otherwise specified. The second value model in MERIS represents economic value-creation potential, from the perspective of a government steward managing fibre flow from public forests for the benefit of society as a whole. The first part of the methodology we describe here (i.e. disaggregation of wood supply model harvest volume, via an intermediate aggregation scheme disaggregation coefficient vectors) could potentially be used to map wood supply models to the economic value model in MERIS, although we have not tested this. I don't understand this.

The value-creation potential model in MERIS is composed of six components: fixed cost, harvest cost, sylviculture credits, stumpage cost, transportation cost, and product values. The *fixed cost* value component in MERIS includes fixed costs related to fibre procurement (i.e. general administrative costs, access road planning and amortization costs)—it is expressed on an area (ha^{-1}) basis. The *harvest cost* value component in MERIS includes all variable costs associated with fibre extraction, including cost of loading logs onto trucks for transportation to processing facilities—it is expressed on a volume ($(m^3)^{-1}$) basis. In Quebec, cost of implementing prescribed non-commercial sylviculture treatments must be assumed by the same entity that harvests the fibre, however this cost is offset by a credit applied to future stumpage fees. ce n'est plus le cas? The *sylviculture credit* value component in MERIS models this credit—it is expressed on an area (ha^{-1}) basis. In Quebec, a stumpage fee must be paid for each unit of fibre harvested from public forest. This corresponds to the *stumpage cost* value component in MERIS—it is expressed on a volume ($(m^3)^{-1}$) basis. The *transportation cost* value component in MERIS estimates cost of transporting fibre from the harvesting site to the processing facility (not including loading cost, but including unloading cost)—it is expressed on a volume ($(m^3)^{-1}$) basis. The *product value* component in MERIS estimates revenue generated from sale of all primary products and co-products that will be produced from a given unit of fibre (specified in terms of species and stem size class), net of processing cost and cost of transporting products to markets.

The next subsection provides a more detailed description of both stages in our methodology.

2.2 Detailed Methods

2.2.1 Stage 1: Disaggregate harvest volume

We start by importing a Woodstock model into WS3, and applying the optimal solution stored in the model (i.e. the solution used to determine AAC).

In Woodstock lingo, a *development type* is equivalent to a forest stratum (i.e. unique combinations of stratification variables, or *themes*) and an age. A single decision in a Woodstock solution is composed of area of a given development type on which a given treatment is applied at a given period.

We want to aggregate our data by broad cover type (softwood, mixedwood, hardwood), but there is no attribute in the BFEC Woodstock models that

3

directly maps to this aggregation level. We can, however, use yield components yg_g_gf (hardwood basal area) and yg_g_gr (softwood basal area) to deduce the cover type of a given development type. We then inject a cover type attribute (i.e. *yield component* into each development type in the model. In Quebec, a softwood cover type must contain at least 75% softwood, a hardwood cover type must contain at least 75% hardwood, and the rest is mixedwood. We analyse basal area at optimal rotation age (i.e. maximum mean annual increment). This should correspond to the age at which the yagemat yield component equals 0 in our BFEC Woodstock models, but this attribute is unreliable so instead we use WS3 to compute maximum MAI age on the fly from the total volume yield component yv_s.

We need to aggregate treatments in the Woodstock dataset into 3 types (1: final cut, 2: selection cutting, 3: commercial thinning) so we can disaggregate harvest volume using function $p_{cst}(x)$, as described in Paradis and LeBel (2016). We add three aggregate actions to our model. Although some actions are difficult to classify, one must bear in mind that these values will be used to select the best volume disaggregation function for each treatment. The type 1 disaggregation profile assumes all stems are harvested, and so the stem size distribution matches standing inventory. Type 2 and 3 disaggregation profiles are based on a model published by Fortin (2014). Note that, for the softwood cover type, we restricted our stem size distribution analysis to high density (class A and B), high basal area (28+ m^3 per ha) stands. Applying commercial thinning treatments to mixedwood or hardwood cover types is just bad silviculture, so we are assuming that this will not come up in the model. 2

Next, we map the 45 species codes in the forest inventory to the 10 species group names used in the MERIS database. Fortunately (although not entirely coincidentally), there is a 1:1 match between species group definitions in MERIS and in the BFEC Woodstock models.

Finally, we compile 90 vectors of disaggregation coefficients, using the methodology described in Paradis and LeBel (2016).

2.2.2 Stage 2: Compiling value-creation indicators

We will now describe in more detail how data for each of the six value components is extracted from the database and compiled into a performance indicator that can be retro-fitted to the Woodstock model solution.

This stage is rather complex. The first few steps involve importing product distribution, stumpage fee, transportation cost, and product value data from MERIS database. Once all the required data has been imported from the MERIS database and loaded into memory, we begin compiling this data to create the 6 value components described earlier, which we inject directly into a live WS3 simulation.

The MERIS system defines two *profiles*, corresponding to hardwood and softwood sawmills. Some value component values vary depending on the profile selected by the user. We will mention how we deal with the profiles when this is relevant.

We import fixed cost data from both hardwood and softwood sawmill profiles, assigning values from the hardwood sawmill profile to hardwood development types and values from the softwood sawmill profile to softwood and mixedwood development types.

Next, we import stumpage rates data from the MERIS database. The stumpage model in Quebec is rather complex. First, stumpage rates vary by tarification zone. The geographical boundaries of these tarification zones do not always line up with management unit boundaries. Thus, each management unit typically overlaps several stumpage zones. For example, management unit UA 064-51, which we use here as a test case for our methodology, overlaps ~~six~~ different tarification zones. The BFEC Woodstock models include a theme (i.e. a stratification variable), that specifies the tarification zone of each development type. Within a stumpage tarification zone, rates are specified in terms of species group and product. The MERIS database defines empirical product distributions for each combination of 45 species and 26 stem size classes. 8 merchantable products are defined in MERIS: veneer logs, 4 hardwood sawlog grades (F1, F2, F3, F4, according the classification scheme defined in Petro and Calvert, 1976), 2 softwood sawlog grades (small sawlogs correspond to DBH size classes 10 through 14, and large sawlogs correspond to DBH size classes 16 and up), and pulpwood. 3 unmerchantable products are defined in MERIS: rot, unutilized, and other. Note that stumpage is defined by species (rather than species group), so we must compile species-wise weight parameter vectors for each of the 30 combination of species group and cover type from the permanent sample plot data used in step 1 (see Paradis and LeBel, 2016 for details).

Next, we import transportation cost data from the MERIS database. Transportation cost coefficients in MERIS are compiled, for each stumpage zone, using 9 product-species groupings (henceforth referred to as *commodities*). We provide detailed species-product–commodity mapping in an appendix. Note that we split white and yellow birch pulpwood into two commodities (we will need this to map to timber licence contract aggregation level further downstream, when compiling the network flow optimisation model), although both are species modeled as one commodity (birch pulpwood) in MERIS. Transportation cost coefficients in MERIS are compiled by stumpage tarification zone, using the mean transportation distance for the closest three processing facilities accepting a given commodity. *Maximum distance? Assumed load sizes? Driving speed by road class? Source of road class data and routes? Assumed truck rental rate?*

you don't know if you should mention this?

Next, we import product value data from the MERIS database. Product values are keyed on species and product class, with unit values for all combinations of species and product class specified for both hardwood and softwood sawmill profiles. The assumption is that a hardwood mill may accept softwood logs, but will pay a lower price for these logs than the softwood mill (to account for the trouble of having to store these logs until they can be dispatched to a softwood sawmill). The inverse goes for softwood sawmills. We plan to use these values to build a network flow optimisation model, which will only allow valid commodity-processor flows, so we import the higher of the profile-wise prices for each combination of species and product class.

Next, we compile harvest cost and sylviculture credit for each decision in the Woodstock optimal solution. Both harvest cost and sylviculture credit are derived from arithmetic functions (rather than being imported directly from the MERIS database).

Harvest cost is estimated using a predictive model, compiled by the BMMB for us in the MERIS system. The BMMB model is an amalgam of machine productivity functions, originally compiled from time study data by FERIC (now part of FPInnovations), combined with several assumptions regarding frequency ratio of machines that compose different systems, relative proportions of system utilization, intensity of harvest prescriptions (final cut versus partial cut), mean skidding distances, roadside sorting complexity, etc. These productivity assumptions are combined with rental rate assumptions, and machine utilization ratio assumptions. The BMMB model mixes productivity functions for feller-bunchers, single-grip harvesters, grapple skidders, forwarders, delimbers, and slashers. The end result is a prediction of unit harvest cost on a volume basis, as a function of cover type, harvest intensity, and mean piece size. The harvest cost function is presented in an appendix.

Sylviculture credit is estimated using the official arithmetic functions the published by government for the purpose of calculating silviculture credit. The sylviculture credit model is made up of 7 different functions. First, we classify (using expert judgement) each partial cut treatment in the woodstock model to one of three classes (used by government-defined criteria for selecting sylviculture credit formulas): progressive cut, selective cut, commercial thinning. We were able to simplify the model down to 4 functions (1, 2, 4, 7). For treatments classified as commercial thinning, we use function 1 for softwood and mixedwood cover types, and 2 for hardwood cover types. For treatments classified as selection cuts, we use function 4. For treatment classified as progressive cuts, we use function 7 for softwood and mixedwood cover types, and 4 for hardwood cover types. The formulas are a function of total standing volume before harvest, harvest volume, and mean piece size of harvested stems. The four formulas are presented in an appendix.

Mean piece size is estimated from Woodstock model yield curve data (quotient of total volume and stem density curves). Although they are accurate (confirmed with David Pothier), the stem density regressions are rather imprecise (i.e., we can expect a large random error, evenly distributed about the mean). The harvest cost prediction model has an overall inverse exponential shape (i.e. inverse J shape). Thus, underestimating mean stem size would tend to induce a relatively large increase in estimated harvest cost, whereas overestimating mean stem size would induce a relatively small increase in harvest cost. The BFEC Woodstock models use a combination of two different growth models (NATURA for even-aged stands, and ARTIMIS for uneven-aged stands). If we can estimate the error distributions of both total volume and stem density curves for NATURA and ARTEMIS models, then we can calculate the error distribution of the quotient of these random variates, which we can then use to calculate the expected value of the harvest cost function. We implemented a help class in `ws3.common.rvquot_gen`, which encapsulates functions from the pacal library for calculating the quotient of two normally distributed random variates. `rvquot_gen` subclasses `scipy.stats.rv_continuous`, so we can simply call `rvquot.expect(...)` on an instance of our class (which has been instantiated with appropriate scale and location parameters for the numerator and denominator random variates) to output the expected value of the harvest cost function.

We have contacted the developers of the ARTEMIS model and confirmed that the error terms for both total volume and stem density are normally distributed. The standard deviations for these error terms are documented in the NATURA documentation. Tables 8 and 11 of the NATURA documentation list standard deviations (REMQ) for stem density and total volume, by *sous-domaine bioclimatique* (SDB). UA 064-51 is in SDB `3ouest`, for 3 groups of strata (by simulation horizon length). We calculated the weighted-average standard deviation, using normalized strata counts in each group as weight coefficients. The mean standard deviations, expressed as a proportion of estimator value, are $\sigma = 0.386$ for total volume and $\sigma = 0.245$ stem density. Note that these error values seem to contracticts annecdotal information we obtained from David Pothier (error on stem density estimate is higher than error on total volume). Unfortunately, these are not the errors we utimately seek, because the yield curves in our BFEC models are the result of aggregating several (NATURA or ARTEMIS, depending on the case) curves to form composite curves. Assuming that the error at all points on the BFEC composite curves is i.i.d. Gaussian distributed, and assuming that we treat each component curve as a single sample (i.e. use the value of component curves directly, ignoring that they are themselves i.i.d Gaussian distributed, as discussed above), we can estimate the standard error σ_i of a given composite curve at any age class $i \in I$ from the the values of its component curves $j \in J$ at the same age

$$\sigma_{\hat{y}_i} = \sqrt{\frac{\sum_{j \in J} (y_{ij} - \hat{y}_i)^2}{|\hat{\mathbf{Y}}| - 1}} \tag{1}$$

where $\hat{y}_i$ represents the value of the BFEC composite curve at age class i, and y_{ij} is the value of component curve j at age class i.

NATURA and ARTEMIS component curves can (allegedly) be obtained from DRF staff at the MFFP. However, we also need a mapping of component curves to composite curves, but have not been successful obtaining this information from BFEC staff. Without these mappings, we cannot estimate $\sigma_{\hat{y}_i}$ as described above. We use a conservative value of $\sigma = 0.5$ for to model both total volume and stem density as random variates, for the purposes of estimating the expected values of harvest cost and sylviculture credit functions.

Finally, we use all the data we just assembled to compile value-creation potential indicators that can be injected *post hoc* into the Woodstock model optimal solution. For each harvesting decision, we compile a total of 144 new indicators (6 value components plus net merchantable harvest volume, compiled at three aggregation levels [total, commodity-wise, and species-group–wise]). We compile net merchantable harvest volume indicators by multiplying total volume at harvest age by treatment-wise coefficients embedded in the Woodstock model code. At this point, we have not automated the process of extracting these coefficients from the Woodstock models (i.e. they must be manually extracted by an expert). These net merchantable harvest volume indicators will facilitate compilation of the network flow model in a subsequent phase.

A Jupyter Notebook containing Python code implementing our methodology and detailed explanations is available from the corresponding author upon request.

3 Results

As an example, we present result of applying our methodology to management unit UA 064-51.

4 Discussion

5 Conclusion

6 Acknowledgements

This study was supported by funding from the *FORAC Research Consortium.*

References

Fortin, M. (2014). Using a segmented logistic model to predict trees to be harvested in forest growth forecasts. *Forest Systems*, 23(1):139–152.

Jerbi, W., Gaudreault, J., D'Amours, S., Nourelfath, M., Lemieux, S., Marier, P., and Bouchard, M. (2012). Optimization/simulation-based framework for the evaluation of supply chain management policies in the forest product industry. In *2012 IEEE International Conference on Systems, Man, and Cybernetics (IEEE SMC 2012)*. IEEE (accepted).

Paradis, G., Bouchard, M., LeBel, L., and D'Amours, S. (2016). A bilevel model formulation for the distributed wood supply planning problem. [unpublished manuscript].

Paradis, G. and LeBel, L. (2016). Compiling disaggregation coefficients to link long- and short-term planning models. [unpublished manuscript].

Petro, F. and Calvert, W. (1976). How to grade hardwood logs for factory lumber. Technical Report Forestry technical report no. 6, Canadian Forestry Service.

Est-ce que le titre est assez évocateur ?

Estimating the value-creation potential of optimal wood supply plans by integrating

Gregory Paradis *

Département des sciences du bois et de la forêt, Faculté de foresterie, de géographie et de géomatique, Pavillon Abitibi–Price, Université Laval, Québec, QC G1K 7P4, Canada.
gregory.paradis.1@ulaval.ca

Luc LeBel

Département des sciences du bois et de la forêt, Faculté de foresterie, de géographie et de géomatique, Pavillon Abitibi–Price, Université Laval, Québec, QC G1K 7P4, Canada.
luc.lebel@sbf.ulaval.ca

October 11, 2016

Abstract

Current implementation of wood supply optimization models in Quebec, Canada, do not include financial performance indicators. We describe a methodology for compiling a hybrid simulation-optimization model that can be used to estimate the value-creation potential of any subset of species-wise annual allowable cut (AAC) volume. Our model retro-fits financial performance indicators to the optimal solution of the long-term wood supply optimization model, which we link to a network flow optimization model that simulates profit-maximizing fibre consumption behaviour of a network of primary processing facilities. Our methodology uses the official government wood supply models, uses only input data that is readily available to government analystical staff, and can be applied with relative ease to any of the 71 management units in Quebec. To the best of our knowledge, we use the best data currently available. Thus, we present a methodology that produces state-of-the-art value-creation-potential estimates, and could potentially be implemented immediately by government staff in Quebec. We run a number of scenarios on management unit UA 064-51, as an example, and report value-creation potential as a function of the proportion of AAC that is consumed. We show that net value-creation potential of harvesting and consuming the entire AAC is negative.

OK but if we aim for forest science we should emphasize how the methodology is relevant in other forest jurisdictions.

*Corresponding authour. Contact telephone number (418) 456-2208.

1 Introduction

Wood supply anlysis plays a central role in strategic forest management planning on public forests. An important output from the wood supply planning process is *annual allowable cut* (AAC), which sets an upper bound on species-wise timber licences that are attributed to industrial fibre consumers for the current planning period. In Quebec, Canada, as in all other Canadian provinces, the provincial government is responsible for natural resource planning, which includes the forest resource. AAC is determined in Quebec using the Woodstock modelling platform. Government analysts use Woodstock to formulate and solve a linear optimization model that maximizes species-wise AAC. The current generation of Woodstock models compiled by Quebec government analysts does not feature any financial performance indicators. Thus, the models estimate maximum sustainable bio-physical fibre output from the forest, regardless of the value-creation potential of this fibre.

To be fair should we mention that AAC is then reduced based on expert opinion regarding operational constraints...?

Paradis et al. (2013) show that the current wood supply model formulation fails to predict risk of long-term wood supply failures under certain circumstances, even when species-wise harvest volumes are substantially lower than AAC in all planning periods. More specifically, Paradis et al. (2013) link the risk to incoherence between bio-physical AAC models and an actual industrial fibre consumption, and conjecture that risk of wood supply could be mitigated if the wood supply model formulation were extended to explicitely anticipate industrial fibre consumption behaviour. Paradis et al. (2016) describe a bilevel wood supply model formulation that mitigates this risk, albeit at a relatively high cost in terms of reduced AAC. To anticipate industrial fibre consumption, they embed a linear network flow optimization model within the existing wood supply model. The network flow optimization model essentially simulates the behaviour of a network of profit-maximising fibre processing facilities, assuming that only fibre with a positive net value-creation potential (for which there is a demand, and matching processing capacity) will in fact be consumed.

In Quebec, the AAC planning cycle length is currently 5 years, and may be increase to 10 years after the end of the current planning cycle (ending in 2018). At this point, it is unlikely that the government could implement the bilevel wood supply model formulation described in Paradis et al. (2016) by 2018 (modelling for the current planning cycle is already well underway), so we cannot expect the aforementionned risk of wood supply failure to be addressed *a priori* in the wood supply planning process until at least 2028.

In the mean time, it may nonetheless be beneficial to implement the technical capabilty to estimate value-creation potential of solutions output by the wood supply models for the next two planning cycles. We therefore propose a methodology for compiling a hybrid simulation-optimization model that can be used to estimate the value-creation potential of any subset of species-wise annual allowable cut (AAC) volume, based on the current generation of wood supply models. Our simulation-optimization model makes it possible to perform a *post hoc* analysis of the volume gap between species-wise AAC and anticipated industrial fibre consumption, which we hope will provide government policy makers with heretofore unavailable information to guide the strategic forest management process in the interim.

The current generation of models do not feature any financial performance indicators. The first step in our methodology consists in retro-fitting value-creation potential indicators to the output from the wood supply model, using a methodology described in Paradis and LeBel (2016). We then link the enhanced wood supply model to a network flow optimisation model that simulates fibre-consumption behaviour of a network of primary processing facilities. We solve the optimisation model repeatedly, constraining the model to consume an evenly-spaced range of proportions of total AAC. Naturally, our profit-maximizing network flow model consumes the subset of available fibre supply with the highest marginal value-creation potential at each iteration. Plotting output from such a simulation (i.e. value-creation potential versus proportion of AAC consumed) provides heretofore unavailable insight into the interaction between proposed wood supply and industrial fibre processing capacity, for a given management unit.

We deliberately designed our methodology around the latest generation of government wood supply models, using only input data that is readily available for the entire province and accessible to government analysts. Thus, our methodology could hypothetically be integrated into the government workflow in the short term with minimal requirement for further development. Furthermore, our methodology can be applied simulataneously model fibre supply from several management units, making regional (or provincial scale) analyses possible.

must highlight that presented methodology is useful / provide insights on how fibre supply models can be combined w/ supply chain models and provide the value creation information to decision makers.

As an example, we apply our methodology to management unit UA 064-51, and present simulation results for a number of scenarios.

The remainder of this paper is organized as follows. We describe our methodology in §2. Results are presented in §3, followed by discussion in §4 and concluding remarks in §5.

2 Methods

We present the methods in two subsections. In the first subsection we describe the formulation of the network flow optimisation model, and provide some information on data sources and methodology for compiling input data for the model. In the second subsection we describe a number of test scenarios, illustrating the application of our methodology to management unit UA 064-51 in Quebec.

2.1 Network flow optimisation model formulation and compilation

The MERIS database (compiled by experts staff at BMMB) contains all the data needed to estimate unit value-creation-potential of any standing tree stem in Quebec. For each forest management unit in Quebec, BFEC staff compile and solve a Woodstock model to determing species-wise AAC. We want to be able to map value-creation–potential data from the MERIS database to each component (i.e. action) in any BFEC Woodstock solution. Data in BFEC Woodstock models is more hightly aggregated than data in the BMMB MERIS database. For example, a harvesting action in a Woodstock model will specify the development type, the action type, and the treated area. Applying this action will generate an array of outputs, including net merchantable harvest volume (expressed in m^3, by species group). There are 11 standard species groups in BFEC Woodstock models. The MERIS database stores value data using several different levels of aggregation.

The finest level of data aggregation used in the MERIS database (for the data we need) expresses value-creation-potential of one unit (in m^3) of fibre in terms of a combination of one of 2 processor profiles (hardwood, softwood), 45 species codes, 26 stem size classes, and 6 product classes. In step 1 of phase 1 of this project, we compiled *disaggregation coefficients* that can be used to disaggregate (i.e. explode) each unit of harvested volume output from a BFEC Woodstock model solution into smaller volume sub-units. Each volume sub-unit exactly matches the finest aggregation level used in the MERIS database, thus allowing us to retrofit MERIS value-creation-potential data to BFEC Woodstock model solutions.

Need to have a schematic of complete methodology.

In this context, the *value* of the stem of a standing tree is defined from the perspective of a network of primary processing facilities seeking to maximize profit from sale of primary products (and primary co-products) to external markets. Given this definition of value, the unit value-creation–potential is equivalent to the sum of unit costs and revenues along a given stem's trajectory from standing tree to delivered primary product.

We designed a linear programming (LP) optimisation model formulation that can emulate profit-maximizing fibre consumption behaviour of any hypothetical network of primary fibre processing facilities. We use a network flow model design pattern: a directed graph of processing nodes connected by transportation arcs. Fibre flows through the network from *source* (`src`) nodes to emphsink (`snk`) nodes. We model three additional layers of nodes between source and sink nodes (*viz.*, *dispatch* (`dsp`), *commodity* (`cdt`), and *processor* (`prc`) nodes), for a total of five layers of nodes.

We model a 1:1 mapping between source nodes and harvesting decisions in the BFEC Woodstock model. In step 2 of phase 1, we retrofitted data from the MERIS database into an instance of our `ws3.woodstock.WoodstockModel` class, which allows us now to programatically interface the `WoodstockModel` object with the network flow model. There is a 1:1 mapping between source nodes and dispatch nodes. The number of source and dispatch nodes will vary, depending on the number of harvest actions in the simulated BFEC Woodstock model solution. Each dispatch node has one inbound arc. Flow along source-dispatch arcs represents *harvested area* (measured in hectares). By setting appropriate upper and lower flow capacity bounds on these arcs, we can simulate harvesting and consumption of any subset of the Woodstock solution. In step 2 of phase 1, we compiled a net value-creation-potential (NVCP) coefficient for each harvesting decision in the Woodstock model. A NVCP coefficient represents the sum of all revenues and costs, from standing tree to delivered primary product, expressed on the basis of harvested area. The objective function value of our network flow model is defined as the scalar product of source-dispatch flow and net value-creation-potential coefficient vectors.

Each dispatch node has one outbound arc connecting it to each of 12 commodity nodes. The notion of *commodity* is used in the MERIS database as an intermediate aggregation scheme combining species group and product class. Outbound flow from dispatch nodes represents *harvested volume* (measured in cubic metres). Thus, dispatch nodes convert harvested area commodity volumes. Unit stumpage cost in the MERIS database is defined as a function of stumpage zone and commodity, so we can map coefficients to dispatch-commodity flows representing unit stumpage cost.

stumpage zone allow us to know wood provenance.

Commodity nodes have one inbound arc for each dispatch node, and one outbound arc for each processing node. Outbound flows represent volume of merchantable fibre (in cubic metres). The number of processing nodes will vary from one instance to another, depending on the number of fibre processing facilities modelled. For our test dataset, we model fibre flows from management unit UA 064-51. We used publicly-available government documentation of timber licence (TL) agreements to compile the list of processing facilities that procure fibre directly from the target management unit. Since 2013, TL volumes in Quebec are attributed on a regional basis (i.e. for a set of management units). Our test dataset models fibre flows from a single management unit, so we need to disaggregate regional TL volumes to define appropriate upper bounds on flows between commodity and processing nodes. We use the most recent pre-2013 (management-unit-wise) TL volume data, for the set of management units in the target region, to estimate commodity-wise proportion of regional TL volumes corresponding to our target management unit, which we use as upper bounds on commodity-processor arc flows in our network model. Sawmill processing nodes generate a significant volume of chips as a co-product. For each sawmill processor node, we add one outbound arc for each pulpmill processor node.

We model one sink node for each processor node, with matching arcs. Facility capacity constraints can be defined as flow bounds on these arcs.

2.2 Description of illustrative test case and scenarios

To demonstrate the application of our methodology, we compiled a test model based on management unit UA 064-51. First provide some background information describing UA 064-51, summarize output from the Woodstock wood supply optimization model (i.e. the starting point for our simulations), and describe the processing facilities that constitute our test network (type, capacity, timber supply contracts, etc.). Next, we describe a number of scenarios that we ran using our hybrid simulation-optimization model.

2.2.1 Background

Describe UA 064-51.

Describe Woodstock optimal solution.

Describe processing facilities, GAs, etc.

2.2.2 Scenarios

We ran 12 different scenarios using our test model. We group some of these scenarios, to highlight that they are based on a common base scenario and represent a logical sequence of parameter changes. This list of scenarios is by no means definitive—our objective here is simply to show an application of our methodology on a real dataset, to illustrate potential usefulness of our model as a decision-support tool for strategic policy-makers.

Scenario 0 is the base model. We simply solve the network flow model to optimality.

Scenario 1 repeatedly solves the same model as scenario 0 for a range of harvest areas between 0 and 100% of available area.

Scenario 2 is based on scenario 1, adding commodity-wise subsidies. We simulate 3 sub-scenarios. Scenario 2.1 uses subsides from Belzille and Riopel (2015). Scenario 2.2 is based on scenario 2.1, and adds a $7\$/m^3$ subsidy for the SPFL commodity. Scenario 2.3 is based on scenario 2.2, and implements an iterative loop to estimate the minimal unit SPFL commodity subsidy such that the SPFL commodity breaks even when 100% of available area is harvested.

Scenario 3 is based on scenario 2.1. We simulate 2 sub-scenarios. Scenario 3.1 sets rente to 0. Scenario 3.2 is based on sub-scenario 3.1, and adds the minimal SPFL subsidy from scenario 2.3.

Scenario 4 is based on scenario 1, and adds area-based subsidies (based on the BMMB MAM policy).

Scenario 5 is based on scenario 1 We simulate 3 sub-scenarios. Scenario 5.1 sets upper-bound constraints on flow on commodity-processor arcs (based on GA volumes). Scenario 5.2 sets lower-bound constraints on flow on commodity-processor arcs (based on GA volumes). Scenario 5.3 implements scenario 5.2 as a goal-programming model (with penalties for violation of flow bounds).

3 Results

4 Discussion

5 Conclusion

6 Acknowledgements

This study was supported by funding from the *FORAC Research Consortium.*

References

Paradis, G., Bouchard, M., LeBel, L., and D'Amours, S. (2016). A bilevel model formulation for the distributed wood supply planning problem. [unpublished manuscript].

Paradis, G. and LeBel, L. (2016). Retro-fitting value-creation potential indicators to long-term wood supply models. [unpublished manuscript].

Paradis, G., LeBel, L., D'Amours, S., and Bouchard, M. (2013). On the risk of systematic drift under incoherent hierarchical forest management planning. *Canadian Journal of Forest Research*, 43(5):480–492.