Modeling annualized occurrence, frequency, and composition of ingrowth using mixed-effects zero-inflated models and permanent plots in the Acadian Forest Region of North America

Rongxia Li, Aaron R. Weiskittel, and John A. Kershaw, Jr.

Abstract: Forest tree ingrowth is a highly variable and largely stochastic process. Consequently, predicting occurrence, frequency, and composition of ingrowth is a challenging task but of great importance in long-term forest growth and yield model projections. However, ingrowth data often require different statistical techniques other than traditional Gaussian regression, because these data are often bounded, skewed, and non-normal and commonly contain a large fraction of zeros. This study presents a set of regression models based on discrete Poisson and negative binomial probability distributions for ingrowth data collected from permanent sample plots in the Acadian Forest Region of North America. Models considered here include regular Poisson, zero-inflated Poisson (ZIP), zero-altered Poisson (ZAP; hurdle Poisson), regular negative binomial (NB), zero-inflated negative binomial (ZINB), and zero-altered negative binomial (ZANB; hurdle NB). Plot-level random effects were incorporated into each of these models. The ZINB model with random effects was found to provide the best fit statistics for modeling annualized occurrence and frequency of ingrowth. The key explanatory variables were stand basal area per hectare, percentage of hardwood basal area, number of trees per hectare, a measure of site quality, and the minimum measured diameter at breast height of each plot. A similar model was developed to predict species composition. All models showed logical behavior despite the high variability observed in the original data.

Résumé : Le recrutement d'arbres en forêt est un processus très variable et fortement stochastique. Par conséquent, la prévision de l'occurrence, de la fréquence et de la composition du recrutement d'arbres est une tâche difficile, mais d'une grande importance pour les projections à long terme des modèles de croissance et de production forestière. Cependant, les données de recrutement nécessitent souvent des techniques statistiques autres que la traditionnelle régression gaussienne parce que ces données sont généralement délimitées, asymétriques et non normales en plus de contenir couramment une grande proportion de valeurs nulles. Cette étude présente une série de modèles de régression basés sur des distributions de probabilité discrètes de Poisson et binomiales négatives étalonnés à partir de données de recrutement provenant de placettes échantillons permanentes établies dans la région forestière acadienne de l'Amérique du Nord. Les modèles considérés dans cette étude incluent la distribution de Poisson régulière, de Poisson à excès de zéros (PEZ), de Poisson tronquée à zéro (PTZ; Poisson à obstacle), binomiale négative régulière (BN), binomiale négative à excès de zéros (BNEZ) et binomiale négative tronquée à zéro (BNTZ; BN à obstacle). Les effets aléatoires à l'échelle de la placette ont été introduits dans chacun de ces modèles. Le modèle BNEZ avec effets aléatoires a produit les meilleurs ajustements statistiques pour modéliser l'occurrence et la fréquence annualisées du recrutement. Les variables explicatives les plus importantes étaient la surface terrière du peuplement à l'hectare, la proportion de la surface terrière en feuillus, le nombre d'arbres à l'hectare, une mesure de la qualité de la station et le DHP minimal mesuré dans chaque placette. Un modèle similaire a été mis au point pour prédire la composition en espèces. Tous les modèles ont produit un comportement logique malgré la grande variabilité observée dans les données originales.

[Traduit par la Rédaction]

Introduction

Ingrowth is defined as trees in a sample plot that have grown into a required threshold size (usually measured by tree height or diameter at breast height (dbh)) over a certain period (Beers 1962; von Gadow et al. 2006). Modeling tree ingrowth is of great importance for forest growth simulations, particularly long-term projections, as it represents one of four key components of forest development: survivor growth, ingrowth, mortality, and harvest (Beers 1962; Shifley et al. 1993).

There are generally two main approaches to model stand ingrowth: static and dynamic (Weiskittel et al. 2011b). Static models predict a constant amount of recruitment with little

consideration of stand characteristics, whereas dynamic models use site and stand conditions as independent variables to predict future stand recruitment. Given that the average amount of recruitment is often assumed to be constant over a long growth period (Mendoza and Setyarso 1986), the static approach is commonly used in stand table projection and matrix models (Vanclay1992). However, such simplified assumptions are often not satisfactory when initial stand conditions vary greatly or accurate growth predictions are sought for short- to medium-term projections.

Dynamic models, on the other hand, provide relatively more accurate predictions by utilizing stand characteristics as explanatory variables in statistical regression models. A number of studies have developed one single linear or nonlinear equation (e.g., Hann 1980; Shifley et al. 1993; Adams and Ek 1974) to predict amount of recruitment at the end of each simulation cycle. However, these models will always predict ingrowth to occur, even if it has not. Consequently, two-stage models were proposed and demonstrated as a better approach in many studies (e.g., Schweiger and Sterba 1997; Vanclay 1992; Adame et al. 2010). In a two-stage model, the first equation estimates probability of ingrowth occurrence based on a set of covariates, and a second equation estimates amount of ingrowth based on the same or a different set of covariates, given that ingrowth has occurred.

Regardless of the modeling approach used, it is challenging to accurately predict number of ingrowth stems over a certain period, because the response variable representing stand ingrowth, namely number of ingrowth trees in a sample plot, is count data by nature and often quite stochastic. As a result, traditional Gaussian models are not appropriate in modeling ingrowth due to the unique distributional features frequently present in the data such as skewness and dispersion. Added to this is the difficulty in dealing with a large proportion of zero counts for the response variable. The standard method to model count data is through a Poisson distribution. However, a regular Poisson model lacks the capability to fully account for the dispersion and heterogeneity in most ingrowth data. Relative to Poisson models, ingrowth data generally exhibit overdispersion as it shows a larger than expected zero fraction and a longer than expected right tail (Affleck 2006).

To overcome the difficulty in modeling dispersed count data, the aforementioned two-stage modeling approach is often adopted to separately estimate zero stem counts and positive stem counts. Fortin and DeBlois (2007) adapted this two-stage modeling into zero-inflated models for hardwood stands in southern Quebec by combining two separate estimation processes into a joint distribution of probabilities. Such zero-inflated models have obvious benefits over commonly used two-part lognormal models with respect to reduction of error propagation. However, relatively few studies have adopted this zero-inflated modeling approach in predicting stand recruitment, and the work of Fortin and DeBlois (2007) is limited to only a few hardwood species. In addition, the hierarchical nature of their data was not taken into account with the use of random effects.

Forest stand ingrowth is a complicated stochastic process influenced by several stand characteristics and climatic and geographical factors at a range of spatial and temporal scales. Complex interactions often exist between these factors too.

As such, it is nearly impossible to accurately capture all (or even a substantial proportion) of the observed variability in empirical ingrowth data. Incorporation of random effects to account for unidentified and unexplainable variation is a logical step forward when longitudinal observations are available. In addition, ingrowth from the same permanent sample plot tends to show a certain amount of correlation, which can also be accounted for by using a mixed-effects model. Although a mixed-effects modeling approach has been adopted in various areas of forest modeling applications such as height–age modeling (Fang and Bailey 2001; Wang et al. 2007; Zhao et al. 2005; Calegario et al. 2005), tree stem taper modeling (Garber and Maguire 2003; Leites and Robinson 2004; Li and Weiskittel 2010), and others (Grégoire et al. 1995; Hall and Clutter 2004), we have not found any examples in the forestry literature that utilized random effects to model ingrowth data.

The aim of this study was to find the best modeling approach for annualized ingrowth count data for mixed species and mixed cohort stands in the Acadian Forest Region of North America. Zero-inflated models with the incorporation of random effects to enhance model predictive abilities were developed and used to assess ingrowth variation due to stand characteristics and site quality. In addition, models for ingrowth species composition were also developed.

Methods

Data

Data used in this study came from an extensive regional database of fixed-area permanent plots compiled from a variety of data sources (Weiskittel et al. 2010). Some important sources of data were the US Forest Service (USFS) Forest Inventory and Analysis (FIA), the USFS Penobscot Experimental Forest, and permanent sample plot (PSP) data from several Canadian provinces. Sample plots covered the majority of Maine and southeastern Canada, including Quebec, Nova Scotia, and New Brunswick. Forests in this region are generally mixed-species forests, representing a unique transition zone from boreal forests in the north to the broadleaf hardwood forests in the south (Braun 1950). The primary conifer species in this region include balsam fir (*Abies balsamea* (L.) Mill), red spruce (*Picea rubens* (Sarg.)), white spruce (*Picea glauca* (Moench) Voss.), white pine (*Pinus strobus* L.), eastern hemlock (*Tsuga canadensis* (L.) Carr.), and northern white-cedar (*Thuja occidentalis* L.). Hardwoods commonly found in the region include red maple (*Acer rubrum* L.), paper birch (*Betula papyrifera* Marsh.), yellow birch (*Betula alleghaniensis* Britt.), and aspen (*Populus* Michx.).

Multiple measurements, which included plot size, total basal area, and number of trees, were obtained from each sample plot. The time interval between two measurements varied among plots, with most having 5-year remeasurement intervals (61%). The number of observed ingrowth trees in a sample plot was divided by measurement length to obtain annualized ingrowth counts (rounded to the nearest integer and standardized on a per-hectare basis). A list of measurements based on data sources are summarized in Table 1. Because site index was rarely measured on these plots and detailed soil maps do not exist for much of the region, a site quality

Table 1. Ingrowth data attributes by data source.

No.	Source	No. of plots	Mean plot size (ha)	Mean measurement interval (year)	Mean total basal area ($m^2 \cdot ha^{-1}$)	Mean stem density ($no. \cdot ha^{-1}$)	Climate site index (m)	No. of ingrowth observations ($trees \cdot ha^{-1} \cdot year^{-1}$)		Minimum dbh (cm)
								Mean	Range	
1	AFERP	175	0.050	5	39.8	1294	16.3	27	(0, 212)	2.3
2	CFRU GIS	365	0.020	1	30.5	959	13.4	7	(0, 250)	11.9
3	CTRN	85	0.081	1	25.9	1837	14.2	5	(0, 278)	4.1
4	FIA	438	0.075	11	16.4	305	14.6	10	(0, 67)	12.8
5	FIA	4457	0.014	5	27.0	898	14.4	24	(0, 297)	6.0
6	New Brunswick	1999	0.027	4	25.4	2156	11.9	27	(0, 299)	4.3
7	Nova Scotia	2754	0.040	5	18.9	826	10.7	22	(0, 270)	9.2
8	Maine PEF	198	0.074	4	17.4	920	16.0	52	(0, 298)	1.6
9	Maine PEF	275	0.081	5	27.0	2045	16.0	50	(0, 239)	1.5
10	Maine PEF	32	0.076	4	0.1	275	15.9	27	(0, 204)	0.5
11	Quebec BAS1	1523	0.040	10	19.5	884	12.4	24	(0, 192)	4.4
12	Quebec BAS2	498	0.032	11	15.8	833	11.6	33	(0, 225)	4.0
13	Quebec FEDE	116	0.037	9	16.8	887	13.0	42	(0, 175)	4.5
14	Quebec PACA	15	0.040	10	24.5	950	12.6	19	(0, 48)	4.1
15	Quebec SCOF	115	0.037	10	7.7	303	13.0	60	(0, 190)	4.0
16	Quebec SPIM	339	0.034	10	21.4	965	12.7	29	(0, 157)	4.4
17	Quebec UNLA	162	0.040	5	20.8	1264	11.8	44	(0, 220)	3.9

Note: Mean total basal area and mean stem density were measured when the plots were initially established. Sources: AFERP, Acadian Forest Ecosystem Research Program; CFRU GIS, Cooperative Forestry Research Unit Growth Impact Study; CTRN, Commercial Thinning Research Network; FIA, Forest Inventory and Analysis; PEF, Penobscot Experimental Forest; BAS1, first plot network established from 1970 to 1977; BAS2, second plot network established in 1989; FEDE, Fédération des producteurs de bois du Québec; PACA, Parks Canada; SCOF, Service de la Comptablité Forestière; SPIM, Service de la protection des insectes et des maladies; UNLA, University of Laval. dbh, diameter at breast height.

index variable was derived based on geographical location of sample plots. This index is based on 1 km^2 climate normals from 1960–1991 and empirically derived relationships with observed site index (Weiskittel et al. 2011*a*). The model was developed using random forests, a nonparametric technique, and the final model explained 73.6% of the original variation in site index and had a root mean squared error of 1.04 m (Weiskittel et al. 2011*a*).

In addition to the stochastic nature of ingrowth, another difficulty in this analysis was the use of different threshold diameters for determining ingrowth. In this analysis, the threshold diameter varied from 0.1 to 11.4 cm according to different data sources. Although Shifley et al. (1993) attempted to develop a method for estimating forest ingrowth at multiple threshold diameters, the precision was quite low and other factors were likely more influential. In our study, we included the threshold diameter (minimum dbh for each plot) as a predictor variable to enhance model performance. However, it is worth noting that the majority of the data had a threshold diameter of <11.7 cm (90% of observations) with a median of 9 cm.

Data Analysis

Zero-inflated Poisson (ZIP) and zero-altered Poisson (ZAP)

A regular Poisson probability mass function is defined as

$$[1] \quad f_{\mathrm{P}}(y) = \frac{\lambda^y \mathrm{e}^{-\lambda}}{y!}$$

where *y* refers to the random variable of ingrowth count, and λ is the parameter that represents both mean and variance of *y*. As mentioned earlier, applying a regular Poisson distribution to ingrowth tree count data cannot adequately account for the excess zeroes present in the data.

Zero-inflated Poisson (ZIP) and zero-altered Poisson (ZAP, also called hurdle Poisson) models are two commonly used methods to include excess zeroes in the modeling process. The difference between these two modeling approaches is how the probability of a zero occurrence is measured. In a ZIP model, zero count data are divided into two parts: (*i*) zero counts caused by a binomial mechanism and (*ii*) zero counts caused by a Poisson process. In a ZAP model, zeroes can be fully modeled as a binomial process with a probability π for presence of a zero and $1 - \pi$ for nonzeroes. Consequently, nonzero counts are modeled slightly differently. In ZIP models, nonzero counts with a probability of $1 - \pi$ are formed under the assumption of a Poisson distribution. In ZAP models, because the Poisson distribution does not enter the process of forming zeroes, the nonzero counts are described by a truncated Poisson distribution, i.e., a Poisson distribution without a probability of a zero occurring. In reality, the difference between ZIP and ZAP models are often subtle in many situations. The probability mass functions of ZIP and ZAP are defined as

$$[2] \quad f_{\mathrm{ZIP}}(y) = \begin{cases} \pi + (1-\pi)\mathrm{e}^{-\lambda} & y = 0 \\ (1-\pi)\dfrac{\lambda^y \mathrm{e}^{-\lambda}}{y!} & y > 0 \end{cases}$$

$$[3] \quad f_{\mathrm{ZAP}}(y) = \begin{cases} \pi & y = 0 \\ (1-\pi)\dfrac{f_P(y)}{1 - f_P(y=0)} & y > 0 \end{cases}$$

where the parameter λ is estimated through measured explanatory variables, and π is the probability of zero occurrence in a binomial process.

Zero-inflated negative binomial (ZINB) and zero-altered negative binomial (ZANB)

Although ZIP and ZAP models can adequately model data containing a large number of zeroes, they generally fail to accurately fit data when the data show a pattern of overdispersion in the positive part of the count data, because the Poisson distribution requires a strict equality of mean and variance. Unlike Poisson models, negative binomial (NB) models have an additional parameter to account for overdispersion. The simplistic version of the probability mass function of the NB distribution is defined as

$$[4] \quad f_{\mathrm{NB}}(y) = \frac{\Gamma\left(y + \frac{1}{k}\right)}{\Gamma\left(\frac{1}{k}\right)y!}\left(\frac{1}{\mu k + 1}\right)^{1/k}\left(\frac{\mu k}{\mu k + 1}\right)^y$$

where *y* is the random variable of ingrowth count, and *k* and μ are parameters to be estimated (Zuur et al. 2009). The variance of the NB distribution is $\mathrm{Var}(y) = \mu + k\mu^2$. When $k \to 0$, the NB distribution converges to a Poisson distribution (Johnson et al. 2005). Similar to ZIP and ZAP, zero-inflated negative binomial (ZINB) and zero-altered negative binomial (ZANB, also called hurdle NB) mass functions are defined as

$$[5] \quad f_{\mathrm{ZINB}}(y) = \begin{cases} \pi + (1-\pi)\left(\dfrac{1}{\mu k + 1}\right)^{1/k} & y = 0 \\ (1-\pi)f_{\mathrm{NB}}(y) & y > 0 \end{cases}$$

$$[6] \quad f_{\mathrm{ZANB}}(y) = \begin{cases} \pi & y = 0 \\ (1-\pi)\dfrac{f_{\mathrm{NB}}(y)}{1 - f_{\mathrm{NB}}(y=0)} & y > 0 \end{cases}$$

where f_{NB} refers to the NB probability mass function (eq. 5). All other notation is the same as previously defined.

Systematic linear predictor

In a generalized linear model (GLM), three components are essential: (*i*) the distribution of the response variable *Y*; (*ii*) a link function; and (*iii*) a systematic linear predictor $\mathbf{X}\boldsymbol{\beta}$ (Zuur et al. 2009, p. 209). The distribution for the response variable, in our case, is a (truncated) Poisson or (truncated) NB for the positive count part and a binomial or combination of binomial and truncated Poisson or truncated NB for the zero count part. A link function is a function that connects the linear predictor (explanatory variables) with the expected value of the response variable. The logarithm was used as the link function for the nonzero counts in both the Poisson and NB models. The systematic linear predictor is a multiplication of a parameter vector $\boldsymbol{\beta}$ and an explanatory variable design matrix $\mathbf{X}$.

Key factors that influence the annualized number of ingrowth trees are related to a variety of stand and site conditions. Understocked stands have potential growing space for ingrowth trees (Shifley et al. 1993), so stand density should be a good indicator for the number of future ingrowth trees. Shifley et al. (1993) also pointed out that tree size, species composition, stand shading conditions, and other stochastic events (e.g., weather, disturbance) may all contribute to the probability of occurrence of ingrowth trees and number of ingrowth trees in a certain area. Based on the data availability and desired use in future projections, four explanatory variables for predicting number of ingrowth trees per hectare were selected: (*i*) total basal area (m²·ha⁻¹); (*ii*) hardwood basal area percentage; (*iii*) number of trees per hectare; (*iv*) the site quality index described above; and (*v*) minimum measured dbh of each plot. Therefore, in the above standard Poisson, ZIP, and ZAP models, we have

$$[7] \quad \ln(\lambda) = \mathbf{X}\boldsymbol{\beta} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5$$

For standard negative binomial, ZINB, and ZANB models, the logarithm link function connects the parameter μ with the same four explanatory variables as

$$[8] \quad \ln(\mu) = \mathbf{X}\boldsymbol{\beta}$$
$$= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5$$

The parameter π in the binomial process for zero count estimation in ZIP, ZAP, ZINB, and ZANB models can be estimated through a logistic model:

$$[9] \quad \pi = \left(\frac{1}{1 + e^{-\mathbf{x}\boldsymbol{\gamma}}}\right)^L$$
$$= \left(\frac{1}{1 + e^{-(\gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 x_3 + \gamma_4 x_4 + \gamma_5 x_5)}}\right)^L$$

where the explanatory variables x_1, x_2, x_3, x_4, and x_5 refer to total basal area, hardwood basal area percentage, number of trees per hectare, site quality index, and minimum measured dbh of the plot, respectively; the βs and γs are the parameters to estimate; and L represents the time interval between two measurements. With increasing L, the probability of zero occurrence gradually decreases (Yao et al. 2001).

We added a plot-level random-effects parameter on the intercept of the linear predictors for the estimation of positive ingrowth count. Thus, the plot-to-plot variability was taken into account, and the model likelihood was expected to improve.

Ingrowth species composition model

Ingrowth is a species-dependent process, because resource needs differ among species. Thus, each species contributes differently to the probability of ingrowth occurrence (Shifley et al. 1993). Moreover, species growing in a mixed stand may interact differently through the interspecies competition for resources. To predict ingrowth species composition, a system of equations was developed that used percentage of ingrowth tree basal area for each species as the dependent variable. The independent variables were stand total basal area, percentage of basal area for each species, and the site

quality index variable. A logistic model was used, but was constrained to force additivity (Parresol 2001). The original ingrowth data contained over 50 different individual species. However, balsam fir and spruce accounted for over 50% of the observations. Consequently, the species in this analysis were grouped into the following categories: birch (8.9%), balsam fir (26.1%), red maple (8.8%), spruce (24.5%), white pine (1.4%), other hardwood (21.7%), and other softwood (8.6%). The system of equations is given as

$$[10] \quad y_i = \frac{1}{1 + \exp(-(b_{i0} + b_{i1} x_1 + b_{i2} x_2 + b_{i3} x_3 + b_{i4} x_4))},$$
$$i = 1, \ldots, 7$$

where y_i is the percentage of ingrowth trees for species i; b_{i0}, b_{i1}, b_{i2}, b_{i3}, and b_{i4} are parameters for species i; x_1, x_2, x_3, and x_4 represent stand total basal area, percentage of basal area for each species, site quality index, and minimum dbh of each plot, respectively. Species index i corresponds to the seven species groups defined above.

For the parameter estimation of the annualized total number of ingrowth trees through maximum likelihood process, we used the SAS/STAT NLMIXED procedure (SAS Institute Inc. 2008). For the ingrowth species composition estimation, the system of equations was simultaneously fit by the SAS/STAT MODEL procedure (SAS Institute Inc. 2008).

Evaluation criteria

Akaike's (1973) information criterion (AIC), Bayesian information criterion (BIC; Schwarz 1978), and –2log-likelihood (–2logL) were used to compare model performance. Smaller values of AIC, BIC, and –2logL indicate a better fit. For nested models, the likelihood ratio test was used to compare whether one model was statistically significantly superior to the other one at the level of 0.05. In our study, the standard Poisson model can be considered as nested in the standard negative binomial model as it had an additional parameter to explain the extent of nonzero count dispersion. The likelihood ratio test can give evidence as to whether this dispersion parameter is needed for the data. Similarly, the ZIP versus ZINB, as well as ZAP versus ZANB, can be tested using the likelihood ratio test. Additionally, the standard Poisson model can be tested against ZIP or ZAP, and standard negative binomial model can be tested against ZINB or ZANB to examine whether the binomial part with a parameter π was necessary in modeling excess zeroes.

For non-nested fitted models that had similar AIC, BIC, or –2logL values, the Vuong (1989) test was employed to further test for significant differences. The Vuong (1989) test has been widely used to compare and test categorical model performance in many fields (e.g., Rose et al. 2006; Zorn 1996) and is defined as

$$[11] \quad \mathrm{V} = \frac{\mathrm{LR}_n(\widehat{\boldsymbol{\theta}}_n, \widehat{\boldsymbol{\gamma}}_n)}{\sqrt{n}\widehat{\varpi}}$$

where n is the number of observations, $\widehat{\boldsymbol{\theta}}_n$ and $\widehat{\boldsymbol{\gamma}}_n$ represent fitted parameter vectors from two comparable models, and $\widehat{\varpi}$ is the standard deviation of the log-likelihood difference between two models, which is given by

Table 2. Fit statistics of Poisson, zero-inflated Poisson (ZIP), zero-altered Poisson (ZAP), negative binomial (NB), zero-inflated negative binomial (ZINB), and zero-altered negative binomial (ZANB) models with and without random effects.

Model	$-2\log L$	AIC	BIC	Pearson's χ^2/df
No random effect				
Standard Poisson	1 050 000	1 050 000	1 050 000	34.38
ZIP	662 455	662 479	662 580	5.67
ZAP	662 456	662 480	662 581	5.67
Standard NB	254 239	254 253	254 312	0.81
ZINB	234 474	234 500	234 610	0.91
ZANB	234 559	234 585	234 695	1.52
With random effect				
Standard Poisson	526 497	526 511	526 564	11.08
ZIP	342 879	342 905	343 003	1.73
ZAP	343 013	343 039	343 137	1.72
Standard NB	253 824	253 840	253 900	0.58
ZINB	231 158	231 186	231 291	0.58
ZANB	231 302	231 330	231 435	0.82

$$[12] \quad \varpi^2 = \text{var}\left(\log\frac{P^1(\boldsymbol{Y}|\mathbf{X};\widehat{\boldsymbol{\theta}}_n)}{P^2(\boldsymbol{Y}|\mathbf{X};\widehat{\boldsymbol{\gamma}}_n)}\right)$$

where P^1 and P^2 are two likelihood functions for two comparable models 1 and 2, and $\text{LR}_n(\widehat{\boldsymbol{\theta}}_n, \widehat{\boldsymbol{\gamma}}_n)$ is the sum of the difference of log-likelihood between two non-nested models. Its mathematical definition is given as

$$[13] \quad \text{LR}_n(\widehat{\boldsymbol{\theta}}_n, \widehat{\boldsymbol{\gamma}}_n) = L_n^1(\widehat{\boldsymbol{\theta}}_n) - L_n^2(\widehat{\boldsymbol{\gamma}}_n)$$
$$= \sum_{i=0}^n \log\frac{P^1(\boldsymbol{Y}|\mathbf{X};\widehat{\boldsymbol{\theta}}_n)}{P^2(\boldsymbol{Y}|\mathbf{X};\widehat{\boldsymbol{\gamma}}_n)}$$

The Vuong (1989) statistic (V) asymptotically follows a standard normal distribution. Therefore, $|V| < 1.96$ suggests no significant difference at 0.05 significance level between models 1 and 2, whereas if $V < -1.96$, model 2 is significantly better than model 1, and $V > 1.96$ favors model 1.

Pearson's chi-square statistics were also calculated to test the models' goodness of fit, defined as

$$[14] \quad \chi_p^2 = \sum_{i=1}^n \frac{\left(y_i - \text{E}(y_i)\right)^2}{\text{Var}(y_i)}$$

where y_i is the observed response variable (the number of ingrowth trees), and $\text{E}(y_i)$ and $\text{Var}(y_i)$ are expected value and variance, respectively, of the response variable under the model probability assumption. A general indication of a good model fit is if the ratio of the model chi-square statistic to its degree of freedom is close to 1.

Results

Ingrowth occurrence and frequency

Model comparison

Of the 33 587 observations available for analysis, 30.1% of them were zeros. When it did occur, the average ingrowth was 22.8 ± 34.1 counts·ha^{-1}·year^{-1} (mean ± SD), with a range between 1 and 299 counts·ha^{-1}·year^{-1}. The observed variance to mean ratio of the response variable was 51.1,

which suggested the possibility of an overdispersion pattern existing in the data. Additional justification for the inclusion of an overdispersion parameter in the NB models was provided by the fit statistics in Table 2. All the NB models (including standard NB, ZINB, and ZANB) had significantly lower –2logL, AIC, and BIC values compared with the Poisson models (standard Poisson, ZIP, and ZAP). Finally, the formal likelihood ratio tests of standard NB, ZINB, and ZANB against standard Poisson, ZIP, and ZAP models, respectively, gave highly significant differences (p value < 0.0001). This clearly indicates that the dispersion parameters in the NB, ZINB, and ZANB models were needed. The goodness-of-fit Pearson χ^2 statistic also suggested that NB models provided a better fit than Poisson models as they were close to 1 when divided by the degrees of freedom (Table 2).

To determine whether excess zeros in the data should be accounted for through a binomial process, the NB was tested against ZINB, as well as the ZANB. The results showed that the ZINB and ZANB models were statistically superior over the NB model (p value < 0.0001). Comparing models with and without random effects, all models were improved when random effects were included (Table 2). Overall, the ZINB with random effects was shown to be the best model tested (p value < 0.0001), which was also confirmed by the Vuong (1989) statistic (results not shown).

Model fit

When the observed and predicted annualized ingrowth frequencies were compared, it was evident that regular Poisson, ZIP, and ZAP models lacked the ability to capture the trend for small values (<40) (Fig. 1; Table 3), especially when models did not include random effects. Predicted frequencies with the NB, ZINB, and ZANB models corresponded relatively well to the observed frequencies. However, the standard NB model could not accurately predict excess zeros in the data. Both ZINB and ZANB with random effects described the data relatively well.

In the final ZINB model with random effects, all parameter estimates were statistically significant (Table 4). Stand basal area, hardwood basal area percentage, and threshold diameter

Fig. 1. The observed and predicted frequencies for the annualized number of ingrowth trees per hectare for the standard Poisson, zero-inflated Poisson (ZIP), zero-altered Poisson (ZAP), standard negative binomial (NB), zero-inflated negative binomial (ZINB), and zero-altered negative binomial (ZANB) models with and without random effects.
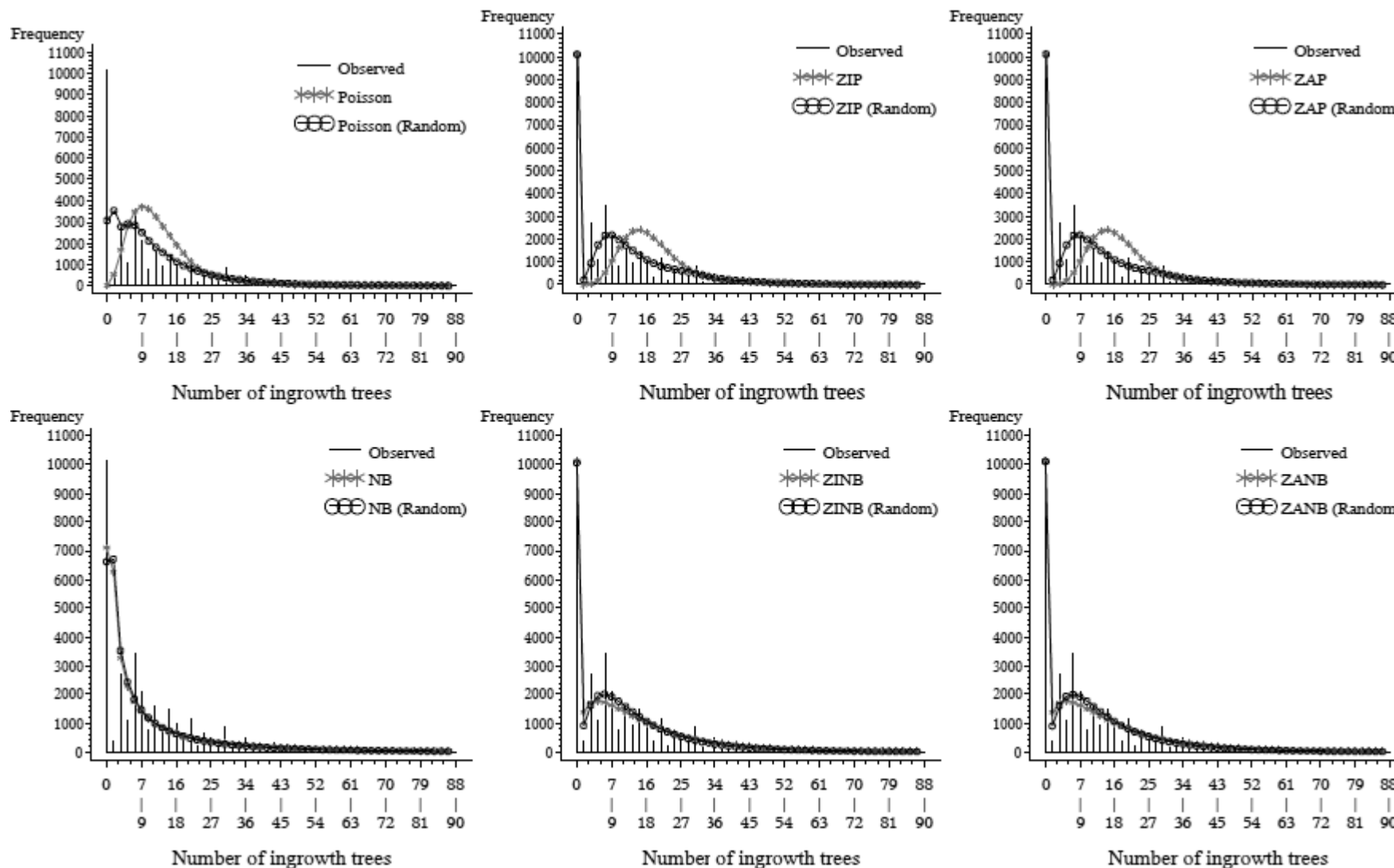
Table 3. The observed frequency versus predicted frequency using fitted Poisson, zero-inflated Poisson (ZIP), zero-altered Poisson (ZAP), negative binomial (NB), zero-inflated negative binomial (ZINB), and zero-altered negative binomial (ZANB) models with random effects.

| Count	Observed frequency	Predicted frequency					
		Poisson	ZIP	ZAP	NB	ZINB	ZANB
0	10 118	3 077	10 109	10 118	6 630	10 071	10 118
1–3	354	3 546	213	213	6 713	927	909
4–6	2 716	2 792	952	949	3 535	1 635	1 611
7–9	1 098	2 922	1 747	1 746	2 443	1 960	1 939
10–12	3 441	2 858	2 164	2 162	1 849	2 019	2 006
13–15	2 123	2 528	2 180	2 179	1 469	1 931	1 924
16–18	795	2 125	1 976	1 976	1 202	1 772	1 771
19–21	1 585	1 821	1 735	1 735	1 005	1 588	1 590
22–24	947	1 580	1 508	1 508	854	1 404	1 407
25–27	1 478	1 349	1 292	1 292	734	1 230	1 234
28–30	1 013	1 136	1 097	1 097	637	1 073	1 078
31–33	335	962	942	942	558	934	939
34–36	1 142	825	827	827	492	813	817
37–39	190	710	742	742	436	707	711
40–42	655	611	677	676	389	616	619
43–45	513	526	619	619	348	538	540
46–48	362	455	561	561	313	470	472
49–51	859	394	497	497	283	411	413
52–54	118	342	429	428	256	361	363
55–57	309	297	363	363	232	318	319
58–60	470	259	306	306	212	280	281
61–63	89	227	260	260	193	247	248
64–66	244	200	225	225	177	219	219
67–69	113	176	197	197	163	194	194
70–72	307	156	175	175	150	172	173
73–75	204	139	155	155	138	154	154
76–78	42	124	138	138	127	137	137
79–81	131	111	123	123	118	123	123
82–84	127	100	109	109	109	110	110
85–87	107	90	98	98	101	99	99
88–90	139	81	88	88	94	89	89
91–93	48	73	79	79	87	80	80
94–96	157	66	72	72	81	72	72
97–99	23	59	65	65	76	65	65
100–102	156	54	58	58	71	59	59
103–105	72	49	52	52	66	54	53
106–108	43	45	46	46	62	49	49
109–111	68	41	42	42	58	44	44
112–114	23	38	37	37	55	40	40
115–117	43	35	34	34	51	37	37
118–120	65	33	31	31	48	34	34
121–123	8	31	29	29	45	31	31
124–126	48	29	27	27	43	28	28
127–129	15	27	25	25	40	26	26
130–132	51	26	24	24	38	24	24
133–135	35	24	23	23	36	22	22
136–138	14	23	22	22	34	20	20
139–141	30	21	21	21	32	19	19
142–144	20	12	10	10	30	17	17
145–147	11	2	2	2	29	16	16

had a negative effect on the number of nonzero ingrowth tree counts, whereas number of trees per hectare and site quality index had a positive influence. The effect of stand basal area on ingrowth was much more pronounced compared with the other factors (Fig. 2). Once stand basal area was greater than 60 m^2·ha^{-1}, percentage of hardwood basal area or number of

Table 4. The estimated parameters for zero-inflated negative binomial model (ZINB) with and without random effects.

Parameter	No random effects			With random effects		
	Estimate	SE	p value	Estimate	SE	p value
γ_0	–0.2116	0.0659	0.0013	–0.1596	0.0626	0.0108
γ_1	0.0255	0.0008	<0.0001	0.0253	0.0008	<0.0001
γ_2	–0.1396	0.0274	<0.0001	–0.1241	0.0261	<0.0001
γ_3	–0.0054	0.0011	<0.0001	–0.0583	0.0107	<0.0001
γ_4	0.0433	0.0044	<0.0001	0.0419	0.0043	<0.0001
γ_5	0.0409	0.0028	<0.0001	0.0393	0.0027	<0.0001
β_0	3.8982	0.0414	<0.0001	4.0303	0.0491	<0.0001
β_1	–0.0257	0.0005	<0.0001	–0.0277	0.0005	<0.0001
β_2	–0.3668	0.0166	<0.0001	–0.3654	0.0200	<0.0001
β_3	0.0002	7.3E-6	<0.0001	0.0002	6.9E-6	<0.0001
β_4	0.0216	0.0028	<0.0001	0.0159	0.0034	<0.0001
β_5	–0.0514	0.0019	<0.0001	–0.0642	0.0023	<0.0001
$\hat{k}$	0.6603	0.0067	<0.0001	0.3906	0.0052	<0.0001
$\hat{\sigma}^2$	—	—	—	0.2825	0.0075	<0.0001

Fig. 2. The expected annual recruitment given ingrowth occurrence under four site quality indices (5, 10, 15, and 25 m), two hardwood composition structures (20% and 80%), and two tree stem densities (1000 and 1500 stems·ha^{-1}) with a threshold dbh = 9 cm, based on the fitted random zero-inflated negative binomial (ZINB) model.
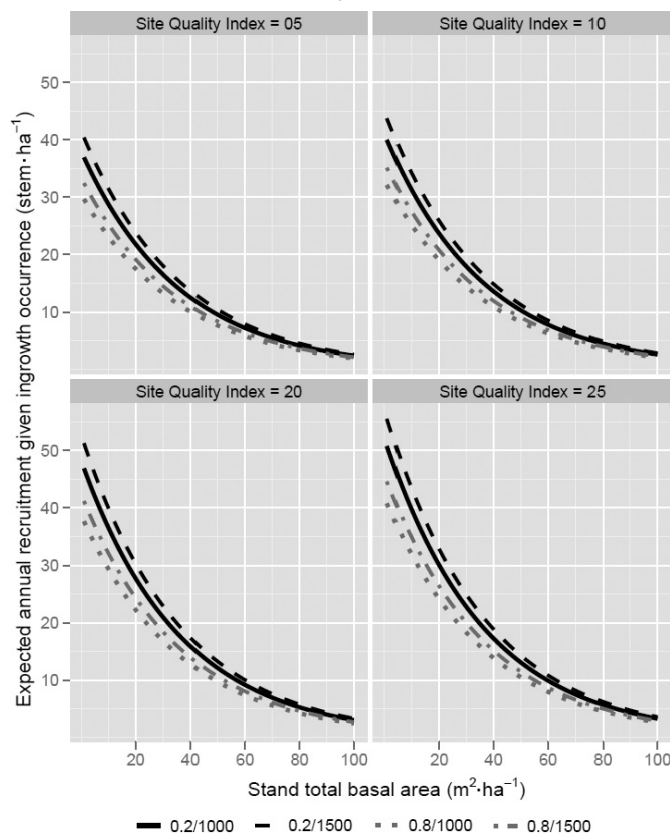
stems per hectare had little effect on ingrowth. However, the data show that 99% of the sample plots had a stand basal area below 60 m^2·ha^{-1}.

Ingrowth composition

Except for the birch species group, all parameters in the species composition model were statistically significant at the 0.05 level (Table 5). The models fit well with mean squared errors ranging from <0.01 to 0.1. In general, balsam fir and spruce had a significantly higher probability to generate ingrowth trees across a range of stand densities and site indices (Fig. 3). White pine and red maple had the lowest probabilities of the species groups examined. Some species such as balsam fir and red maple showed a positive relationship with stand total basal area, whereas others such as spruce and white pine had a negative relationship.

Discussion

Traditionally, tree recruitment was modeled using a two-stage conditional linear distribution as suggested in many studies (e.g., Stage and Ferguson 1982; Ferguson et al. 1986; Vanclay 1994). In the first stage, the zero count is described by a binomial distribution, whereas the positive count data are modeled by a continuous linear function in the second stage. Due to the often highly skewed nature of recruitment data, a logarithm transformation is generally applied to the nonzero response variable. Essentially, the original error term is assumed to be lognormally distributed. Such two-part conditional lognormal models may occasionally fit a particular ingrowth data set well, but the assumption of a continuous distribution violates the nature of ingrowth data and may cause the probability of inconsistent zero counts (Fortin and DeBlois 2007). This study adopted a modeling framework that was fully based on generalized count modeling theory. Consequently, the fitted models predicted the frequency for each observed ingrowth tree count rather than predicting the mean response variable as in Gaussian models. Because the assumption of count models conforms to the discrete nature of tree recruitment, the results are easily interpreted without any transformation that can cause error propagation and bias.
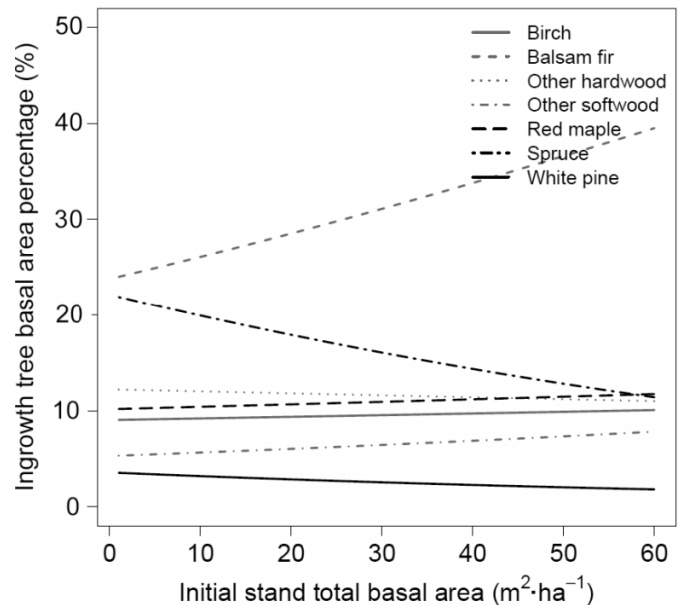
Although the theoretical framework of zero-inflated models was well established more than a decade ago (Lambert 1992; Hall 2000), their applications in forestry are limited due to the complexity and difficulty in obtaining convergence. Several recent efforts at using zero-inflated models to predict stand mortality as well as recruitment and regenera-

Table 5. The parameter estimates, standard error (SE), and *p* values for the fitted ingrowth species composition model by species group.

Parameter	Estimate	SE	*p* value
Birch			
b10	–2.5645	0.0917	<0.0001
b11	0.0020	0.0011	0.0554
b12	2.6624	0.0333	<0.0001
b13	–0.0010	0.0062	0.8704
b14	–0.0127	0.0042	0.0024
Balsam fir			
b20	–3.0291	0.0846	<0.0001
b21	0.0027	0.0010	0.0091
b22	2.7779	0.0342	<0.0001
b23	0.0211	0.0053	<0.0001
b24	0.0221	0.0040	<0.0001
Red maple			
b30	–0.6566	0.0661	<0.0001
b31	0.0123	0.0007	<0.0001
b32	1.7669	0.0174	<0.0001
b33	–0.0421	0.0045	<0.0001
b34	–0.0283	0.0030	<0.0001
Spruce			
b40	–1.2500	0.0679	<0.0001
b41	–0.0132	0.0007	<0.0001
b42	2.0470	0.0193	<0.0001
b43	–0.0514	0.0048	<0.0001
b44	0.0351	0.0030	<0.0001
White pine			
b50	–5.1074	0.0909	<0.0001
b51	–0.0117	0.0014	<0.0001
b52	3.8817	0.0562	<0.0001
b53	0.0501	0.0061	<0.0001
b54	0.0726	0.0057	<0.0001
Other hardwood			
b60	–2.9832	0.0681	<0.0001
b61	–0.0020	0.0008	0.017
b62	2.4837	0.0227	<0.0001
b63	0.0673	0.0045	<0.0001
b64	–0.0167	0.0031	<0.0001
Other softwood			
b70	–4.7182	0.0776	<0.0001
b71	0.0070	0.0008	<0.0001
b72	3.2269	0.0340	<0.0001
b73	0.1000	0.0049	<0.0001
b74	0.0188	0.0031	<0.0001

Fig. 3. The predicted ingrowth tree basal area percentage versus stand total basal area based on species composition model (covariates settings: species basal area percentage, 14.285% for each species category; site quality index, 12; threshold dbh, 9 cm).

tion are found in the forestry literature (e.g., Barry and Welsh 2002; Affleck 2006; Rathbun and Fei 2006; Fortin and DeBlois 2007). These studies generally focused on one or several species, and their predictions were periodically made at the end of one measurement cycle, which can restrict their usage. In addition, these previous studies did not often account for the hierarchical nature of their data. In this study, the zero-inflated models were shown to be largely improved with inclusion of random effects as random effects accounted for most of the variation between data sources and plot-to-plot variability. In addition, prediction of total annualized ingrowth for mixed-species stands provides a direct method to

incorporate ingrowth into existing forest growth and yield simulation tools.

The results indicated high plot-to-plot variability, which was unable to be explained. Graphical assessment of the plot-level random effects and residuals over various factors such as threshold diameter, plot size, soil drainage, and elevation did not show any obvious trends. However, plot size likely does influence both the probability and the amount of ingrowth as larger plots would generally have higher rates of ingrowth for a given set of stand conditions when compared with a smaller plot. Although an optimal plot size for measuring ingrowth in this region has yet to be established, the majority of observations in this study (73%) were obtained from plots that were >0.02 ha in size. Further studies need to be performed to investigate the effects of plot size on ingrowth.

In this analysis, ZIP regression did not perform well compared with ZINB models. This result partially agrees with those of Fortin and DeBlois (2007), in which their ZIP models performed worse than a zero-inflated discrete Weibull model. The lack of fit in ZIP models was caused by overdispersion, normally found in ingrowth data. A slight difference was observed between the ZINB and ZANB models. The predicted frequency was lower using ZANB than using ZINB for observed counts less than 20 except zero counts, whereas it was higher using ZINB for counts greater than 20 but less than 60 (Table 3). These differences were caused by underlying model assumptions. Both binomial and negative binomial processes govern the distribution of zero counts in ZINB, whereas the zero counts were modeled only by the binomial process for ZANB.

When choosing which model to use in real applications, an understanding of the biological phenomenon of the nature of the study regarding zero occurrences (Zuur et al. 2009) is required. For example, if excess zeros in the data are generated

primarily by poor sampling design, measurement errors, and other unknown factors, so called "false" zeros (Martin et al. 2005), ZANB can be a good choice. However, if excess zeros are a mixture of false zeros and true zeros that are direct results of ecological or biological effects, the ZINB model should be selected. In our study, the ZINB regression provided better fit statistics than the ZANB model, and the Vuong (1989) test also indicated that ZINB performed relatively better. However, the ZANB model offered a better predicted frequency of zero counts than the ZINB model (Table 3). Although the ZINB was selected as the best model for this particular study, it would not be unexpected to find that ZANB models may perform equally well in other studies.

Tree regeneration and recruitment are complicated and stochastic processes (Vanclay 1992). The statistical models developed in this analysis are relatively simple as using stand conditions as the primary covariates can only partially account for the great variability exhibited in those processes. Much of the unexplained variability is due to the fact that ingrowth at any given time is a cumulative result of events and conditions over several previous years, including a variety of climatic, geographical, biological, and natural and human environmental factors and the interactions between them (Shifley et al. 1993). In addition, recruits generated from the same plot with similar site conditions, but measured in different years, are likely to be correlated. Such stochastic variations and correlations in ingrowth data have been realized and discussed in the literature, but the problem is difficult to resolve due to the complexity in reaching convergence (Fortin and DeBlois 2007). In this study, a random effect was added on the intercept of the systematic linear predictor in Poisson and negative binomial models (including standard and zero-inflated or zero-altered), and results presented a moderate improvement over those produced by no-random-effect models. In general, the fitted models performed quite well and logically despite the overall noisiness of the fitting data set.

In this analysis, the number of ingrowth trees decreased with greater stand density when expressed in terms of total basal area. As in this analysis, most other recruitment studies have found basal area to be the strongest predictor (e.g., Moser 1972; Ek 1974; Fortin and DeBlois 2007). Rather than basal area, crown competition factor (CCF) was initially assessed as a potential covariate, given its use by Shifley et al. (1993). However, CCF did not drastically improve model performance, and basal area was used instead for simplicity. Although stem density also is an indicator of stand density, the results showed it to have a positive effect on the number of ingrowth trees, unlike total basal area. This result also has been reported in several previous studies (e.g., Moser 1972, 1974; Ek 1974; Shifley et al. 1982) and most likely represents stage of stand development rather than competition as young dense stands are likely to have larger numbers of ingrowth trees, especially if the threshold diameter is large.

There is no consensus on whether site quality influences number of ingrowth trees. For example, Fortin and DeBlois (2007) did not include a measure of site quality in their recruitment model, whereas Ek (1974) found it to be nonsignificant. In contrast, Hann (1980) concluded that ingrowth was higher on better sites. The results of the present analysis agree with Hann (1980) as we found higher sites to have

more ingrowth occurrence. This is logical as better sites generally have better soil conditions or more available resources to support growth and development of regeneration. However, this ingrowth rate increase on better sites is highly dependent on ingrowth occurrence, i.e., when ingrowth is highly likely to occur in a plot, the site quality for this plot imposes a positive influence on producing the number of ingrowth trees. In this analysis, ingrowth rates were dependent on species composition in that stands dominated by hardwood species had a reduced annual ingrowth rates. This result may be caused by several factors, including the predominance of intolerant hardwood species in the Acadian Forest Region, the past site disturbance history, and the increased presence of balsam fir in the overstory (Olson and Wagner 2010).

A variety of approaches were used in the past to model species composition of ingrowth. For example, Ferguson et al. (1986) predicted the number of ingrowth species for a plot and then estimated a species probability of occurrence, whereas Hasenauer and Kindermann (2006) just predicted a species probability of occurrence similar to the approach used in this present analysis. In their model of species composition, Hasenauer and Kindermann (2006) used the same covariates utilized in predicting probability of recruitment, which included both stand competition and site quality measures. A similar approach was used in this analysis and was found to perform quite well.

In general, white pine and red maple had much lower ingrowth rates compared with balsam fir and spruce. The predominance of balsam fir ingrowth agrees with recent findings of Olson and Wagner (2010) in which balsam fir has dominated the understory in the last five decades across a wide range of silvicultural regimes. This is because balsam fir is relatively shade-tolerant, a prolific producer of seed, can grow on a range of habitats, and responds well to release (Bakuzis and Hansen 1965). In contrast, white pine is more shade-intolerant, seeding is more periodic, early growth is slow, and it has certain seedbed conditions (Burns and Honkala 1990). These species characteristics are clearly evident in the model predictions, particularly the response to changes in total stand basal area. For example, our model predicts percentage of balsam fir to increase as stand basal area increases, whereas the opposite is true for white pine. Likely, this is depicting changes in understory light conditions, and balsam fir would be favored in low light conditions.

In summary, this work represents a significant improvement in modeling tree recruitment by combining mixed-effects modeling techniques with zero-inflated distributions. The analysis provides a framework for modeling discrete, skewed, and non-Gaussian data in forestry, and the results are readily available to be incorporated into any forest growth simulator for the Acadian Forest Region of North America.

Acknowledgements

Nova Scotia Department of Natural Resources, and Québec Ministry of Natural Resources. Chris Hennigar, the Associate Editor, and two anonymous reviewers provided helpful feedback in improving a previous draft of the manuscript.

References

Adame, P., del Rìo, M., and Cañellas, I. 2010. Ingrowth model for pyrenean oak stands in north-western Spain using continuous forest inventory data. Eur. J. For. Res. **129**: 669–678.

Adams, D.M., and Ek, A.R. 1974. Optimizing the management of uneven-aged forest stands. Can. J. For. Res. **4**(3): 274–287. doi:10.1139/x74-041.

Affleck, D.L.R. 2006. Poisson mixture models for regression analysis of stand-level mortality. Can. J. For. Res. **36**(11): 2994–3006. doi:10.1139/x06-189.

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. *In* Proceedings of the 2nd International Symposium of Information Theory, Budapest, Hungary. *Edited by* B.N. Petrov and F. Csaki. Akademiai Kiado, Budapest, Hungary. pp. 267–281.

Bakuzis, E.V., and Hansen, H.L. 1965. Balsam fir: a monographic review. University of Minnesota, St. Paul, Minnesota.

Barry, S.C., and Welsh, A.H. 2002. Generalized additive modeling and zero inflated count data. Ecol. Model. **157**(2–3): 179–188. doi:10.1016/S0304-3800(02)00194-1.

Beers, T.W. 1962. Components of forest growth. J. For. **60**: 245–248.

Braun, E.L. 1950. Deciduous forests of eastern North America. Hafner, New York.

Burns, R.M., and Honkala, B.H. 1990. Silvics of North America. Vol. 1. Conifers. Agriculture Handbook 654, USDA Forest Service, Washington, D.C.

Calegario, N., Daniels, R.F., Maestri, R., and Neiva, R. 2005. Modeling dominant height growth base on nonlinear mixed-effects model: a clonal *Eucalyptus* plantation case study. For. Ecol. Manage. **204**(1): 11–21. doi:10.1016/j.foreco.2004.07.051.

Ek, A.R. 1974. Nonlinear models for stand table projection in northern hardwood stands. Can. J. For. Res. **4**(1): 23–27. doi:10.1139/x74-004.

Fang, Z., and Bailey, R.L. 2001. Nonlinear mixed effects modeling for slash pine dominant height growth following intensive silvicultural treatments. For. Sci. **47**: 287–300.

Ferguson, D.E., Stage, A.R., and Boyd, R.J. 1986. Predicting regeneration in the grand fir – cedar – hemlock ecosystem of the northern Rocky Mountains. For. Sci. Monogr. 26.

Fortin, M., and DeBlois, J. 2007. Modeling tree recruitment with zero-inflated models: the example of hardwood stands in southern Québec, Canada. For. Sci. **53**: 529–539.

Garber, S.M., and Maguire, D.A. 2003. Modeling stem taper of three central Oregon species using nonlinear mixed effects models and autoregressive error structures. For. Ecol. Manage. **179**(1–3): 507–522. doi:10.1016/S0378-1127(02)00528-5.

Grégoire, T.G., Schabenberger, O., and Barrett, J.P. 1995. Linear modeling of irregularly spaced, unbalanced, longitudinal data from permanent-plot measurements. Can. J. For. Res. **25**(1): 137–156. doi:10.1139/x95-017.

Hall, D.B. 2000. Zero-inflated Poisson and binomial regression with random effects: a case study. Biometrics, **56**(4): 1030–1039. doi:10.1111/j.0006-341X.2000.01030.x. PMID:11129458.

Hall, D.B., and Clutter, M. 2004. Multivariate multilevel nonlinear mixed effects models for timber yield predictions. Biometrics, **60**(1): 16–24. doi:10.1111/j.0006-341X.2004.00163.x. PMID:15032769.

Hann, D.W. 1980. Development and evaluation of an even- and

uneven-aged ponderosa pine/Arizona fesue stand simulator. USDA For. Serv. Res. Pap. INT-267.

Hasenauer, H., and Kindermann, G. 2006. Modeling regeneration in even- and uneven-aged mixed species forests. *In* Sustainable forest management: growth models for Europe. *Edited by* H. Hasenauer. Springer-Verlag, Berlin, Germany. pp. 167–193.

Johnson, N.L., Kotz, S., and Kemp, A.W. 2005. Univariate discrete distributions. 3rd ed. Wiley, New York.

Lambert, D. 1992. Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics, **34**(1): 1–14. doi:10.2307/1269547.

Leites, L.P., and Robinson, A.P. 2004. Improving taper equations of loblolly pine with crown dimensions in a mixed-effects modeling framework. For. Sci. **50**: 204–212.

Li, R., and Weiskittel, A. 2010. Comparison of model forms for estimating stem taper and volume in the primary conifer species of the North American Acadian Region. Ann. For. Sci. **67**(3), article 302. doi:10.1051/forest/2009109.

Martin, T.G., Wintle, B.A., Rhodes, J.R., Kuhnert, P.M., Field, S.A., Low-Choy, S.J., Tyre, A.J., and Possingham, H.P. 2005. Zero tolerance ecology: improving ecological inference by modeling the source of zero observation. Ecol. Lett. **8**(11): 1235–1246. doi:10.1111/j.1461-0248.2005.00826.x. PMID:21352447.

Mendoza, G.A., and Setyarso, A. 1986. A transition matrix forest growth model for evaluating alternative harvesting schemes in Indonesia. For. Ecol. Manage. **15**(3): 219–228. doi:10.1016/0378-1127(86)90068-X.

Moser, J.W. , Jr. 1972. Dynamics of an uneven-aged forest stand. For. Sci. **18**: 184–191.

Moser, J.W., Jr. 1974. A system of equations for the components of forest growth. *In* Growth models for tree and stand simulation. Proceedings of the 1973 IUFRO Meeting of Working Party S4.01-4. *Edited by* J. Fries. Royal College of Forestry, Stockholm, Res. Notes 30. pp. 260–287.

Olson, M.G., and Wagner, R.G. 2010. Long-term compositional dynamics of Acadian mixedwood stands under different silvicultural regimes. Can. J. For. Res. **40**(10): 1993–2002. doi:10.1139/X10-145.

Parresol, B.R. 2001. Additivity of nonlinear biomass equations. Can. J. For. Res. **31**(5): 865–878. doi:10.1139/x00-202.

Rathbun, S.L., and Fei, S. 2006. A spatial zero-inflated Poisson regression model for oak regeneration. Environ. Ecol. Stat. **13**(4): 409–426. doi:10.1007/s10651-006-0020-x.

Rose, C.E., Martin, S.W., Wannemuehler, K.A., and Plikaytis, B.D. 2006. On the use of zero-inflated and hurdle models for modeling vaccine adverse event count data. J. Biopharm. Stat. **16**(4): 463–481. doi:10.1080/10543400600719384. PMID:16892908.

SAS Institute Inc. 2008. SAS 9.2 help and documentation. SAS Institute Inc., Cary, North Carolina.

Schwarz, G.E. 1978. Estimating the dimension of a model. Ann. Stat. **6**(2): 461–464. doi:10.1214/aos/1176344136.

Schweiger, J., and Sterba, H. 1997. A model describing natural regeneration recruitment of Norway spruce (*Picea abies* (L.) Karst.) in Austria. For. Ecol. Manage. **97**(2): 107–118. doi:10.1016/S0378-1127(97)00092-3.

Shifley, S.R., Moser, J.W., Jr., and Brown, K.M. 1982. Growth and yield model for the elm–ash–cottonwood type in Indiana. USDA For. Serv. Res. Pap. NC-218.

Shifley, S.R., Ek, A.R., and Burk, T.E. 1993. A generalized methodology for estimating forest ingrowth at multiple threshold diameters. For. Sci. **39**: 776–798.

Stage, A.R., and Ferguson, D.E. 1982. Regeneration modeling as a component of forest succession simulation. *In* Forest Succession and Stand Development Research in the Northwest. Proceedings

of a Symposium at Oregon State University, Corvallis, Oregon, 26 March 1981. *Edited by* J.E. Means. Forest Research Laboratory, Oregon State University, Corvallis, Oregon. pp. 24–30.

Vanclay, J.K. 1992. Modelling regeneration and recruitment in a tropical rain forest. Can. J. For. Res. **22**(9): 1235–1248. doi:10.1139/x92-165.

Vanclay, J.K. 1994. Modelling forest growth and yield, applications to mixed tropical forest. CAB International, Wallingford, UK.

von Gadow, K.V., Pukkala, T., and Tomé, M. 2006. Forest inventory. Springer, Dordrecht, the Netherlands.

Vuong, Q.H. 1989. Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica, **57**(2): 307–333. doi:10.2307/1912557.

Wang, Y., LeMay, V.M., and Baker, T.G. 2007. Modeling and prediction of dominant height and site index of *Eucalyptus globulus* plantations using a nonlinear mixed-effects model approach. Can. J. For. Res. **37**(8): 1390–1403. doi:10.1139/X06-282.

Weiskittel, A.R., Wagner, R.G., and Seymour, R.S. 2010. Refinement of the forest vegetation simulator, northeastern variant growth and yield model: phase 1. *In* 2009 Annual Report. *Edited by* S.R. Mercier. University of Maine, Cooperative Forestry Research Unit, Orono, Maine. pp. 44–48.

Weiskittel, A.R., Wagner, R.G., and Seymour, R.S. 2011*a*. Refinement of the forest vegetation simulator, northeastern variant growth and yield model: phase 2. *In* 2010 Annual Report. *Edited by* W. Mercier. University of Maine Cooperative Forestry Research Unit, Orono, Maine.

Weiskittel, A.R., Hann, D.W., Kershaw, J.A., and Vanclay, J.K. 2011*b*. Forest growth and yield modeling. Wiley-Blackwell.

Yao, X., Titus, S.J., and MacDonald, S.E. 2001. A generalized logistic model of individual tree mortality for aspen, white spruce, and lodgepole pine in Alberta mixedwood forests. Can. J. For. Res. **31**: 283–291. doi:10.1139/x00-162.

Zhao, D., Wilson, M., and Borders, B.E. 2005. Modeling response curves and testing treatment effects in repeated measures experiments: a multilevel nonlinear mixed-effects models approach. Can. J. For. Res. **35**(1): 122–132. doi:10.1139/x04-163.

Zorn, C.J.W. 1996. Evaluating zero-inflated and hurdle Poisson specifications. Midwest Political Science Association Working Paper, 18–20 April 1996.

Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A., and Smith, G.M. 2009. Mixed effects models and extensions in ecology with R. Springer Science+Business Media, LLC.