Environ Ecol Stat (2015) 22:145–160
DOI 10.1007/s10651-014-0290-7

Influence of prior distributions and random effects on count regression models: implications for estimating standing dead tree abundance

Matthew B. Russell

Received: 16 December 2013 / Revised: 1 April 2014 / Published online: 7 May 2014

Abstract The presence and abundance of standing dead trees (SDTs) in forests is typically characterized by an excess number of zeros and high variation. The variability inherent in SDT data naturally leads to the assessment of novel quantitative methods to represent SDT populations and their role in contributing to forest ecosystem structure. This analysis assessed the performance of count regression methods fit with Bayesian mixed-effects models that estimate SDTs (all dead trees with a diameter at breast height $\geq$12.7 cm found on plots 0.07-ha in size) on over 17,000 forest inventory plots across the US Lake States (Michigan, Minnesota, and Wisconsin). Random effects models that used the independent variables basal area, mean annual temperature, and plot ownership (i.e., publicly- or privately-owned) as fixed effects and forest type as a random effect outperformed a method which used fixed-effects, alone. Standard and zero-inflated negative binomial models were effective in accounting for overdispersion present in the data (variance/mean ratio for SDT counts was 72.6), whereas Poisson count models were not. Random effects were calibrated on a new population of SDTs. Employing informative prior distributions from the developed model led to improved estimates of SDT abundance by reducing root mean square error by seven percent. This analysis highlights an approach that uses existing models for representing population averages while calibrating their local random effects with new data to arrive at improved estimates of SDTs across the region.

Handling Editor: Pierre Dutilleul.

Electronic supplementary material The online version of this article (doi:10.1007/s10651-014-0290-7) contains supplementary material, which is available to authorized users.

M. B. Russell (✉)
Department of Forest Resources, University of Minnesota, St. Paul, MN 55108, USA
e-mail: russellm@umn.edu

Keywords Biomass · Carbon · Coarse woody debris · Forest inventory · Markov chain Monte Carlo (MCMC) · Zero-inflation

1 Introduction

Standing dead trees (SDTs) provide a heterogeneous forest structure (Harmon et al. 1986), play an important role in forest carbon dynamics (Harmon et al. 2011), and provide habitat for deadwood-dependent organisms (Stokland et al. 2012). Managing for SDTs and their associated attributes (e.g., their presence, abundance, and size distribution) has become a requisite for many forest certification programs (e.g., Sustainable Forestry Initiative 2004). From a global perspective, estimates of SDT biomass and carbon are needed for countries that report greenhouse gas emissions from land use, land-use change, and forestry sectors to the United Nations Framework Convention on Climate Change (Woodall et al. 2012). Based on these ecological benefits and certification and reporting requirements, accurately depicting SDT presence and abundance is important.

Estimating standing deadwood attributes using traditional forest inventory information is inherently difficult given the rarity of SDTs relative to live tree abundance. The shortfalls of models depicting SDT abundance was recently highlighted by Woodall et al. (2012), who found that models estimating carbon stocks for SDTs in the US were overestimated by 100 % when compared to field observations. Despite the hardships in modeling the resource, researchers have observed success in estimating SDT abundance using physiographic information (Eskelson et al. 2009; Temesgen et al. 2008a), measures of stand density (Russell et al. 2012; An and MacFarlane 2013), and remote sensing technologies (Bater et al. 2009; Eskelson et al. 2012). Common in many analyses with SDTs is assessing the degree to which the data exhibit zero-inflation (ZI), defined as proportions of zeros in excess of that expected under a given count distributional assumption. Hence, count distributions such as the Poisson and negative binomial are useful in describing the stochastic nature of the number of SDTs (Eskelson et al. 2009). To quantify both structural and stochastic sources, ZI count models are commonly employed. In the case for estimating SDT abundance, one model component estimates the probability of SDT occurrence using a logistic equation while a subsequent component estimates SDT abundance from a count distribution.

The use of Bayesian methodologies in modeling forest resources will likely increase in the future (Weiskittel et al. 2011). Despite their advantages for characterizing and interpreting model output, utilization of Bayesian methods in the forest sciences have been relatively minimal when compared to other fields (Li et al. 2011a). Although parameter estimates and their associated confidence intervals are not likely to differ between maximum likelihood (ML) methods and Bayesian techniques that employ noninformative priors (e.g., Table 4 in Ellison 2004; Fig. 3 in Li et al. 2011a), there are several benefits in employing a Bayesian methodology. Two such benefits are that (1) posterior distributions may be subsequently used to assess model estimation uncertainty and (2) posteriors can be updated as new data become available. This is a tremendous advantage to modelers that rely on national forest inventory data, as new information is commonly collected either by remeasuring existing samples (i.e., remeasurement of the count of SDTs on a permanent sample plot) or collecting data from new samples (i.e., measuring SDTs in a previously non-measured forest stand).

All of the benefits of Bayesian methods can be exploited if count regression models fitted with mixed models are analyzed under a Bayesian framework. Under a mixed-effects model in general, fixed-effects parameters are specified to represent population average responses, while random-effects parameters are selected to represent local responses. By calibrating the random effects with new data, improved local estimates can result. When estimating random effects under an ML framework, there are multiple options for choosing the appropriate selection criteria. For example, Temesgen et al. (2008b) recommended local calibration of a nonlinear mixed model by employing a best linear unbiased prediction, but only if a subsample of measurements were available. An assessment of how best to subsample from a SDT dataset to localize parameters using an ML framework may become multifaceted given the large number of variables common to traditional forest inventories and how random terms are specified within the model. A benefit of Bayesian methods is that in the case of estimating SDT presence and abundance, local parameters can be estimated for new data by using posterior distributions as prior distributions in an estimation phase. As the majority of the world's forests are not inventoried for deadwood (Woodall et al. 2009), the ability for robust models to characterize rare ecosystem components such as SDTs is essential for determining forest structure, assessing biodiversity potential, and quantifying forest carbon stocks.

The primary goal of this study was to use a Bayesian approach for quantifying SDT abundance across the US Lake States using various count regression strategies. Specific objectives were to: (1) estimate SDT presence and abundance on 0.07-ha sample plots using forest inventory and climate information and Markov chain Monte Carlo methods, and (2) compare methodologies for estimating random effects using Bayesian methods on a new population of SDTs.

2 Methods

2.1 Study area

Forests of the Lake States region of the United States (including the states of Michigan, Minnesota, and Wisconsin) contain diverse forest types and are comparable to many of the world's forests found in cool-to-cold temperate zones (Frelich 2002). Soils supporting these forests range from deep sandy and silt loams to coarse loams to shallow soils underlain by bedrock. Natural disturbances occurring in these various forest types include fire, wind events, and herbivory (Frelich 2002). Sixty-three forest types, as identified by the USDA Forest Service's Forest Inventory and Analysis (FIA) program (Woudenberg et al. 2010), were identified in the study area. The most common forest type groups were observed in the aspen, sugar maple-beech-yellow birch, and white oak-red oak-hickory types (n = 3,087, 2,126, and 1,426, respectively).

2.2 Data

Tree and plot records were obtained from the USDA Forest Service's FIA program (Forest Inventory and Analysis 2012). Measurements on live and SDTs $\geq$ 12.7 cm in diameter at breast height (DBH) on a main plot were collected on four circular subplots with a 7.2-m radius spaced 36.6 m apart. Hence, the main plot totaling 0.07 ha was the spatial scale investigated in this analysis, and only plots with four forested subplots were considered. Data collected on plots between 2000 and 2011 were compiled. Many of these plots were remeasured once or twice during this time span, while others were measured only once. Thirty-year (1961–1990) average climate data were obtained for each FIA plot by specifying latitude, longitude, and elevation of each plot location to a model developed from climate station data across forests of North America (Rehfeldt 2006; USDA Forest Service 2013). Mean annual temperatures of the FIA plots ranged from 1.3 to 9.7 °C and precipitation from 46 to 100 cm. Only plots with no visible disturbance and no observable stand treatments since the last remeasurement (or in the last 5 years for plots that were initially measured) were used in this analysis. To avoid the potential temporal autocorrelation between remeasured plots, only the most recent measurement of a plot was used throughout this analysis.

A main objective of this analysis was to use an existing model as a framework for updating model parameters with independent (i.e., "newer") data not used in the fitting procedure: hence, plot data collected from 2000 through 2009 were considered the model fitting (MF) dataset, while plots measured in 2010 and 2011 were treated as the model update (MU) dataset. Plot-level variables using live and SDT tree records were summarized for each plot, and grouped by ownership type (i.e., public versus private).

2.3 Estimating SDT presence and abundance

Variations of the Poisson (P) and negative binomial (NB) random variables were employed to test their effectiveness in accounting for the variability of SDT presence and abundance across the region. These count regression models are useful for estimating the number of SDTs found in a fixed area of land (Eskelson et al. 2009; Russell et al. 2012; Temesgen et al. 2008a). The Poisson regression model is the benchmark model for count data but becomes restrictive when estimating attributes other than the mean (Winkelmann 2008). Negative binomial regression models are count models that include an overdispersion parameter, making them more flexible than Poisson models and suitable for estimating SDT abundance (Eskelson et al. 2009). To account for data with a high proportion of zeros, zero-modified count models are often applied. From these kinds of models, zero-inflated Poisson (ZIP) and NB (ZINB) models are two types of mixture models that estimate structural and stochastic zeroes (Welsh et al. 1996; Gray 2005). Hence, comparing the Poisson and negative binomial family of models is an evaluation of the degree of overdispersion in the data, unobserved heterogeneity, and excess zero values (Winkelmann 2008; p. 174). Random effects can also be incorporated into count models to account for unexplained variation inherent in the data (e.g., Li et al. 2011b).

2.3.1 The Poisson family

A P probability is estimated by the mass function

$$f_P(y) = \frac{\lambda^y e^{-\lambda}}{y!} \tag{1}$$

where y denotes an SDT count and λ the SDT count mean. In the ZIP model, zero counts are estimated by those derived from a binomial or Poisson distribution. A ZIP probability is estimated by the mass function

$$f_{ZIP}(y) = \begin{cases} \pi + (1-\pi)\,e^{-\lambda}, & y = 0 \\ (1-\pi)\,f_P(y), & y = 1, 2, 3, \ldots \end{cases} \tag{2}$$

where π is the probability of zero occurrence, λ is estimated using independent variables, and $f_P(y)$ is the right-hand side of Eq. 1. In the case of the Poisson, the variance is equal to its mean.

2.3.2 The negative binomial family

The primary difference in the NB compared to the P distribution is the incorporation of an overdispersion parameter α. An NB probability, obtained as a gamma mixture of P distributions, is estimated by the mass function

$$f_{NB}(y) = \frac{\Gamma\,(y + 1/\alpha)}{\Gamma\,(y+1)\,\Gamma\,(1/\alpha)} \left(\frac{1}{1+\lambda\alpha}\right)^{1/\alpha} \left(\frac{\lambda\alpha}{1+\lambda\alpha}\right)^{y} \tag{3}$$

where y denotes an SDT count and λ the SDT count mean. The variance of the NB defined above (Eq. 3) is $\mu + \alpha\mu^2$. Analogous to the ZIP model, a ZINB probability is estimated by the mass function

$$f_{ZINB}(y) = \begin{cases} \pi + (1-\pi)\left(\frac{1}{1+\lambda\alpha}\right)^{1/\alpha}, & y = 0 \\ (1-\pi)\,f_{NB}(y), & y = 1, 2, 3, \ldots \end{cases} \tag{4}$$

where $f_{NB}(y)$ is the right-hand side of Eq. 3 and all other variables are as previously defined. For the λ parameter in the models, each were related to a system of linear independent variables $\mathbf{X}\boldsymbol{\beta}$, where $\mathbf{X}$ is a vector of explanatory variables and $\boldsymbol{\beta}$ is a vector of regression coefficients to be estimated.

The factors which influenced SDT abundance (i.e., the independent variables included in $\mathbf{X}$) were related to stand structure, climate, and ownership type of the FIA plot. Measures of stand density have been found to be correlated with SDT abundance in other regions (An and MacFarlane 2013; Russell et al. 2012; Temesgen et al. 2008a). Climate variables have been shown to be useful surrogates of SDT abundance across large regions (Eskelson et al. 2012) and have similarly been shown to be related to downed deadwood populations (Woodall and Liknes 2008a). Researchers have found an increase in SDT abundance on publicly compared to privately-owned lands (Eskelson et al. 2012; Kennedy et al. 2008; Ohmann et al. 2007), making ownership type an important variable to consider when characterizing SDT populations across large geographic regions. Based on previous studies and the FIA data structure, three explanatory variables were chosen for estimating the presence and abundance of SDTs per hectare: (1) total basal area for live trees > 12.7 cm (BA; $\mathrm{m^2 ha^{-1}}$); (2) mean annual temperature of the FIA plot (MAT; °C); and, (3) a variable PUB indicating whether the FIA plot was located on public (PUB = 1) or private (PUB = 0) land.

2.4 Model construction

Log odds and log-link functions were used to model variation in the means of SDT absence and positive counts (abundance), respectively. Because of variability in SDTs across forest types (Eskelson et al. 2009, 2012), a forest type-level random effects parameter was specified on the intercept term of the linear equation estimating SDT abundance. Hence, for the P and ZIP models, the log-link function describes λ as

$$\ln(\lambda) = \beta_0 + b_i + \beta_1 \mathrm{BA} + \beta_2 \mathrm{MAT} + \beta_3 \mathrm{PUB} \tag{5}$$

where β_i's are parameters to be estimated using Bayesian methods for the P or ZIP approach and b_i is the random intercept term for the ith forest type. Similarly, for the NB and ZINB models, the log-link function describes λ as

$$\ln(\lambda) = \chi_0 + c_i + \chi_1 \mathrm{BA} + \chi_2 \mathrm{MAT} + \chi_3 \mathrm{PUB} \tag{6}$$

where χ_i's are parameters to be estimated using Bayesian methods for the NB or ZINB approach and c_i is the random intercept term for the ith forest type. The π parameter (Eqs. 2 and 4) representing the binomial process in the ZIP and ZINB models similarly employed the random intercept term, although the MAT variable consistently included zero in the estimated 95 % credible interval, so was omitted. Hence, the logistic equation characterizing the π (Eqs. 2 and 4) parameter was

$$\pi = \frac{1}{1 + \exp\left(-\left(\gamma_0 + v_i + \gamma_1 \mathrm{BA} + \gamma_2 \mathrm{PUB}\right)\right)} \tag{7}$$

where γ_i's are parameters to be estimated with Bayesian methods and v_i is the random intercept term for the ith forest type.

For the 2000–2009 dataset, parameters were estimated for Eqs. 5–7 using the MCMC procedure available in the SAS/STAT® software system (SAS Institute Inc 2011). For each of the equations, 200,000 iterations were run through a Markov chain to achieve convergence and estimation of posterior distributions using the Metropolis sampling algorithm. These iterations were run following a 20,000 iteration burn-in, where sample estimates were discarded. To reduce autocorrelations between successive Markov chain samples, the thinning parameter was set to three. Noninformative priors were chosen because little information on these parameters existed: a normal distribution was selected as the prior for each of the parameters such that $\sim N(0.0, 1.0e^4)$. Priors for variance terms were set as $(1/\sigma) \sim$ Gamma(0.01,0.01), a well-suited distribution given the non-negative nature of the inverse Gamma. Assessment of the convergence of the MCMC output was accomplished using an approach outlined by McCarthy (2007, pp. 252–253). This included confirming that trace plots from the Markov chain were well-mixed, inspecting graphs to ensure that successive Markov chains were effectively noncorrelated, and ensuring that the 95 % highest posterior density (HPD) interval excluded the value zero (akin to the frequentist approach of accepting a value of a parameter at the $\alpha = 0.05$ level). Effective sample sizes (Kass et al. 1998) for each of the parameters were examined to assess Markov chain mixing.

In addition to assessing the posterior distributions of parameters, relative model fit was evaluated using the deviance information criterion (DIC; Spiegelhalter et al. 2002).

2.5 Estimating SDTs for a new population

With the existing set of models for the 2000–2009 data, alternative strategies were employed to assess the performance of those models in estimating SDT abundance in the 2010–2011 data. Specifically, three strategies were used to determine the effectiveness of estimating new random effects for the 2010–2011 using the existing models (Eqs. 5–7).

The first method applied each of the four models to estimate SDTs using the appropriate fixed-effects parameters (termed method Fixed). In the second method, the 2010–2011 data were entered into a MCMC run to estimate the posterior distributions for the random effects parameters. Here, noninformative priors (identical to what is outlined in the previous section) were specified to initiate the MCMC (termed method NonInform). The final method was similar to the second, but informative priors, identified by using the estimated mean and standard deviation for the posterior estimates from the 2000–2009 data (i.e., Eqs. 5–7), were specified to initiate the MCMC (termed method Inform; Online Resource 1).

Uncertainty in SDT estimates from these four methods was assessed after computing root mean square error (RMSE), mean bias (MB), and mean absolute bias (MAB). These measures were computed in this analysis as:

$$\begin{aligned} \mathrm{RMSE} &= \sqrt{\sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2 / n} \\ \mathrm{MB} &= \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right) / n \\ \mathrm{MAB} &= \sum_{i=1}^{n} \left|\left(y_i - \hat{y}_i\right)\right| / n \end{aligned} \tag{8}$$

where y_i is the observed number of SDTs, $\hat{y}_i$ is the estimated number of SDTs, and n is the number of observations.

Table 1 Summary of plot conditions for data acquired across the US Lake States from the Forest Inventory and Analysis database

Attribute	2000–2009 data				2010–2011 data			
	Mean	SD	Min	Max	Mean	SD	Min	Max
	Public lands ($n = 6{,}384$ plots; 13 % with zero SDTs)				Public lands ($n = 2{,}295$ plots; 13 % zero SDTs)			
SDT (count ha^{-1})	78.2	76.2	0.0	595.0	84.0	76.8	0.0	550.0
BA (m^2 ha^{-1})	16.6	11.3	0.2	78.0	16.6	11.4	0.2	66.2
MAT (°C)	4.7	1.5	1.3	9.6	4.4	1.4	1.3	9.4
	Private lands ($n = 11{,}253$ plots; 21 % with zero SDTs)				Private lands ($n = 3{,}148$ plots; 18 % zero SDTs)			
SDT (count ha^{-1})	57.7	15.6	0.0	609.0	66.2	67.3	0.0	520.0
BA (m^2 ha^{-1})	15.6	10.5	0.2	78.0	14.4	10.0	0.2	63.9
MAT (°C)	5.8	1.6	1.6	9.7	5.4	1.6	1.6	9.6

Abbreviations are standing dead trees (SDTs); basal area per hectare (BA; live trees $\geq$ 12.7cm); mean annual temperature (MAT)

3 Results

3.1 Model comparison

From the 17,637 FIA plots contained in the MF dataset, 13 and 21 % of plots contained zero SDTs on public and private lands, respectively. Average number of SDTs (count ha^{-1}) was 65.1 $\pm$ 68.7 (mean $\pm$ SD; Table 1). Observed variance/mean ratio for SDT counts was 72.6, indicating overdispersion relative to a Poisson assumption was present in the data. Locations of plots spanned approximately 8° of latitude and 15° of longitude (Fig. 1).

For models, that incorporated a ZI component, results showed that the probability of finding at least one SDT on a plot increased if the plot was found on public lands with relatively large BA. Using the ZIP output, at a BA of 10 and 30 $m^2\,ha^{-1}$ on publicly-owned land, models indicated that the probability of at least one SDT being present on an FIA plot was 0.48 and 0.96, respectively. The MAT variable was not effective in detecting SDT presence when investigating the ZINB count model, as indicated by the value zero contained in the Bayesian HPD interval for the estimated parameter. Model parameters generally suggested that the number of SDTs would increase on plots found on public lands with increasing BA and decreasing MAT. For example, at a BA of 30 m^2ha^{-1} and a MAT of 8 °C, 16 % more SDTs ha^{-1} were estimated to occur on public compared to privately owned lands when considering a ZINB model (Table 2).

Overdispersion present in the SDT dependent variable suggested the need for assessing the performance of NB and ZINB models in addition to the Poisson family of models. The lowest values for DIC were reported for the NB and ZINB mod-

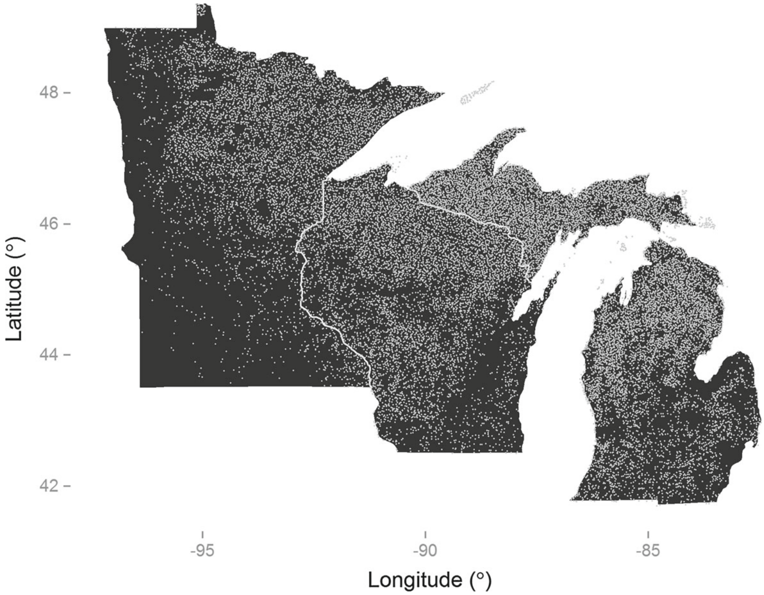

Fig. 1 Approximate locations of forest inventory plots across the US Lake States, 2000–2011

Table 2 Parameter estimates and 95 % highest posterior density (HPD) intervals fit with random intercept effects estimating the number of standing dead trees across the US Lake States for Poisson, negative binomial (NB), and zero-inflated Poisson (ZIP) and negative binomial (ZINB) models

Parameter	Mean	SD	HPD_{LOW}	HPD_{HIGH}	Mean	SD	HPD_{LOW}	HPD_{HIGH}
	Poisson [a]				NB [a]			
β_0	3.930	0.13	3.6934	4.1211	4.00	0.056	3.89	4.11
β_1	0.0229	0.000089	0.0227	0.0231	0.0285	0.00085	0.0268	0.0302
β_2	−0.0944	0.00071	−0.0957	−0.093	−0.0917	0.0055	−0.102	−0.0805
β_3	0.1281	0.0021	0.1242	0.1322	0.153	0.017	0.122	0.187
	ZIP [b,c]				ZINB [b,c]			
χ_0	4.28	0.034	4.21	4.34	3.93	0.057	3.85	4.04
χ_1	0.0132	0.000091	0.0130	0.0133	0.0285	0.00061	0.0273	0.0297
χ_2	−0.0664	0.00072	−0.0678	−0.0650	−0.0905	0.0039	−0.0979	−0.0827
χ_3	0.0887	0.0020	0.0845	0.0925	0.152	0.012	0.128	0.174
γ_0	0.0500	0.088	−0.129	0.219	0.0425	0.19	−0.236	0.329
γ_1	−0.0957	0.0029	−0.102	−0.0903	0.352	0.18	−0.000920	0.611
γ_2	−0.445	0.047	−0.539	−0.352	–	–	–	–

[a] $\ln(\lambda) = \beta_0 + b_i + \beta_1 \mathrm{BA} + \beta_2 \mathrm{MAT} + \beta_3 \mathrm{PUB}$

[b] $\ln(\lambda) = \chi_0 + c_i + \chi_1 \mathrm{BA} + \chi_2 \mathrm{MAT} + \chi_3 \mathrm{PUB}$

[c] $\pi = \frac{1}{1+\exp(-(\gamma_0+v_i+\gamma_1 \mathrm{BA}+\gamma_2 \mathrm{PUB}))}$

Table 3 Deviance information criteria (DIC) of models fitted to the standing dead tree data using noninformative prior distributions with and without random effects for Poisson, negative binomial (NB), zero-inflated Poisson (ZIP), and zero-inflated negative binomial (ZINB) models

Model	DIC	
	Without random effects	With random effects
Poisson	1,135,272	1,055,880
NB	180,989	179,834
ZIP	761,356	713,191
ZINB	196,181	193,762

els. After comparing models both with and without random effects, all models were improved when a random effects term was included on the intercept parameter of a component equation (Table 3). Preliminary analyses which included an FIA plot-level random effects parameter on the intercept term indicated poor mixing of Markov chains found in trace plots and significant autocorrelation following the simulation run as found in lag plots. After comparing observed and estimated SDTs, standard P and NB models did not appear to account for the excess number of zeros present in the data. For both the standard and ZI forms, models fit with random effects generally outperformed those without random effects, especially when moderate counts of SDTs were observed (i.e., between 41 and 120 SDTs ha^{-1}). Random ZIP and ZINB models overestimated the number of plots with zero SDTs, but the ZINB model fit with random effects performed fair when moderate counts of SDTs were observed (Fig. 2).

3.2 Updating model random effects

For the model update dataset (2009–2011), the performance of the various count models varied depending on whether fixed-effects were used or if random effects were locally estimated using noninformative or informative priors. In this approach, local posterior estimates of the intercept term varied depending on the FIA forest type. For example, sugar maple-beech-yellow birch forests displayed a similar shape and scale exhibited by the prior and global posterior distributions, while other forest types deviated somewhat from those distributions (e.g., aspen; Fig. 3). With the exception of the ZIP model, RMSE and MB were lowest for a method that used informative priors to update random effects to arrive at SDT estimates. For these models, improvement was minimal, yet RMSE was 7 % lower on average when using informative priors compared to a method which used fixed-effects alone for estimating the number of SDTs (Table 4).

4 Discussion

Similar to what was found by An and MacFarlane (2013) when analyzing data from various forest types in Michigan, models presented here showed that SDT presence

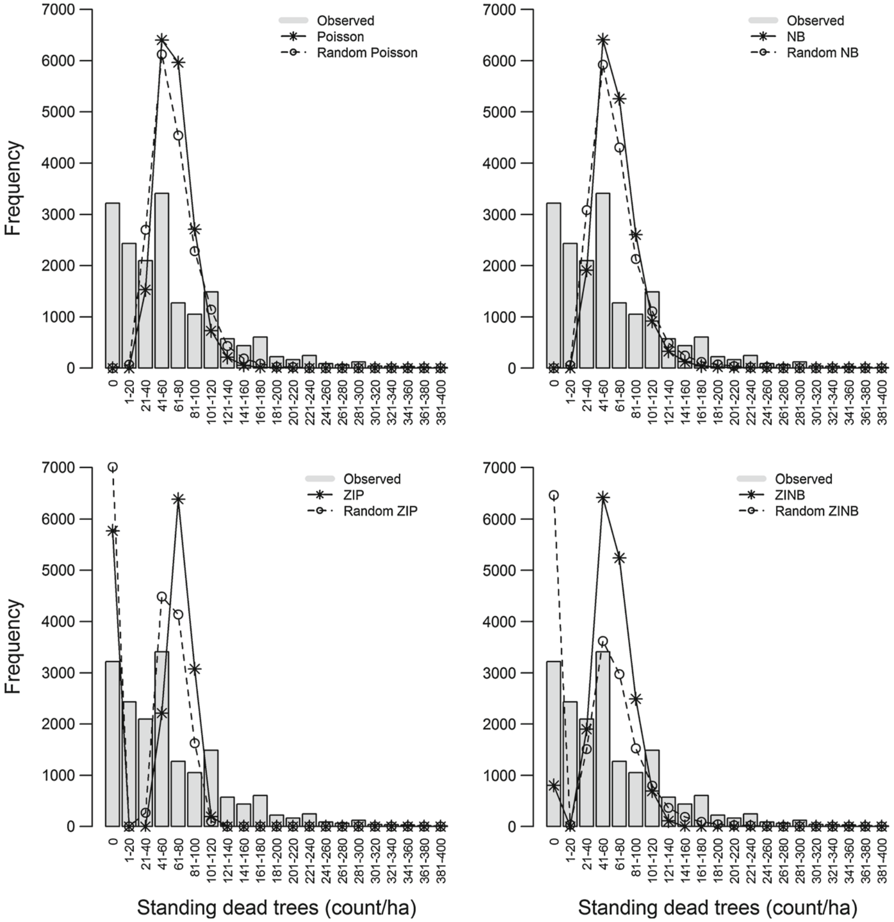

Fig. 2 Observed and estimated frequencies of the number of standing dead trees per hectare across the US Lake States for the Poisson, negative binomial (NB), zero-inflated Poisson (ZIP), and zero-inflated negative binomial (ZINB) models with and without random effects

in the US Lake States was dependent on forest type and the basal area of live trees found in the stand. Despite these relationships, ZI count models were not accurate when comparing the estimated number of SDTs when none were observed, a finding that has similarly been found in the US Pacific Northwest (Eskelson et al. 2012). Information on forest management practices (Russell et al. 2012) and disturbance history (Eskelson et al. 2012) may improve the estimation of the presence and/or abundance of SDTs, yet this kind of information would need to be used with caution when applied across large geographic regions and forest ownership types. Here, the probability of SDT presence was largely driven by BA and public/private ownership types. Models indicated decreasing SDT abundance as MAT increased, a finding also observed by Iwashita et al. (2013) when comparing SDT carbon stocks across a tropical montane forests and by Woodall and Liknes (2008a, b) across a latitudinal gradient in

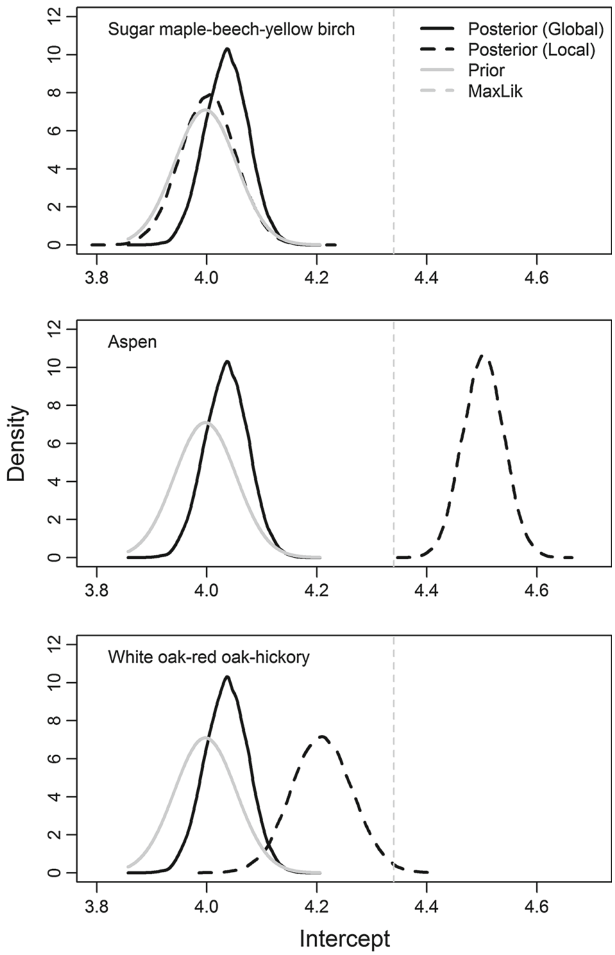

Fig. 3 Posterior and informative prior distributions estimated using global and local parameters with random effects from a model update dataset fit with a zero-inflated negative binomial model and a maximum likelihood (MaxLik) estimate from a model fitting dataset for the three most common forest types in the US Lake States

temperate US forests. By providing statistical models which are sensitive to climate variables, these analyses may help to provide insights into quantifying the role of SDT production and decomposition under future climate scenarios.

The count regression models examined in this study were improved if random effects were incorporated on the intercept parameter of the various components. Incorporating a forest-type level random effect permitted variability across the geographic region and generally resulted in a more accurate representation of SDT populations when compared to observed values (Fig. 2). For the ZI component, it is important to note that the presence/absence of SDTs is largely dependent on the sampling design from which the data are derived. In the case of the FIA data employed here, the minimum DBH threshold (12.7 cm), plot size (0.07-ha), and sampling design (fixed-area)

Table 4 Model validation statistics including root mean square error (RMSE), mean bias (MB), and mean absolute bias (MAB) for updating the random effects estimating the number of standing dead trees per hectare using Poisson, negative binomial (NB), zero-inflated Poisson (ZIP), and zero-inflated negative binomial (ZINB) models using fixed (Fixed), noninformative (NonInform) and informative (Inform) priors

Model	Method	Fit statistic		
		RMSE	MB	MAB
Poisson	Fixed	73.1	23.9	49.6
	NonInform	66.1	4.4	47.5
	Inform	66.0	2.1	47.9
NB	Fixed	70.7	13.0	49.3
	NonInform	66.7	4.2	47.7
	Inform	66.7	3.0	47.9
ZIP	Fixed	94.8	46.2	69.4
	NonInform	100.0	53.9	74.4
	Inform	101.0	56.2	75.5
ZINB	Fixed	71.3	16.6	49.2
	NonInform	67.0	8.6	47.1
	Inform	66.8	5.7	47.4

used will likely differ when compared to a disparate study that may employ an alternative sampling protocol. For example, a lower minimum DBH threshold and larger sample plot would increase both (1) the probability of encountering a SDT and (2) SDT abundance. Alternative sampling strategies such as point-relascope and N-tree sampling (Kenning et al. 2005) may be appropriate in some instances, yet would lead to different counts of SDTs than using fixed-area plots. The relatively poor performance of the ZIP model that included random effects in comparison to estimates using fixed effects alone (Table 4) could be due to the fact that overdispersion was not accounted for in the ZIP model and the difficulties in accurately estimating SDT presence. This poor performance of the ZIP model has similarly been shown by Fortin and DeBlois (2007) and Li et al. (2011b) using maximum likelihood methods in their analyses of forest ingrowth data. The general agreement in the performance of the standard NB and ZINB models both in terms model fit as measured by DIC (Table 3) and their effectiveness in estimating the number of SDTs for a new population (Table 4) was similarly observed by Gray (2005). Although this analysis only considered the most recent measurements from FIA plots, investigations into both spatial (e.g., Zhang and He 2013) and temporal autocorrelation structures that account for unequally spaced observations through time [e.g., the continuous time autoregressive process examined by Cruz-Mesía and Marshall (2006); variable occasion designs as discussed in Snijders and Bosker (2012)] and their role in the Bayesian random-effects model will likely need to be addressed in future modeling efforts.

Using process-based forest growth models, Oijen et al. (2013) concluded that the manner in which Bayesian models are initialized is key when evaluating their performance. This analysis similarly observed that model output varied depending on whether global fixed-effects or noninformative/informative prior distributions were specified during the estimation phase. As the Bayesian paradigm readily permits updating (Li et al. 2011a), this approach generally found that incorporating model coefficients as prior distributions to estimate random effects parameters for a new population outperformed an approach using fixed effects alone. There was a difference in model performance for this new population when comparing noninformative versus informative priors for some models and certain fit statistics, indicating that the likelihood of the data may have more of an influence on estimating the random effects than specifying prior distributions when a large number of samples are considered (e.g., $>$5,400 in this case). The finding that the regression coefficients from the Bayesian and maximum likelihood approaches were generally equivalent has been similarly observed by others (e.g., Ellison 2004; Li et al. 2011a). Although the Bayesian method is by design more computationally demanding, the benefits of the approach can not only result in improved estimates via calibrating the random effects as presented here, but one could also arrive at appropriate bounds of uncertainty surrounding mean estimates by drawing from the full posterior distribution. In the case that future modeling exercises assign prior distributions to equation parameters drawn from resampling (e.g., remeasurement of the FIA plots sampled here), an assessment of the biasedness or the performance in employing noninformative priors may need to be examined.

5 Conclusion

In the Bayesian model, a unified framework for inference and estimation is inherent (Cruz-Mesía and Marshall 2006). In the example presented here using SDT data, estimates on a new population are made using prior information from an existing model. Using a strategy that makes use of posterior distributions for global (i.e., fixed-effects) parameters, improvements in model estimates were observed when updating local (i.e., random-effects) parameters using a new SDT population rather than estimating them with global parameters alone.

The approach outlined here may be well-suited to ecological inventory data because new information is continually collected, whether by sampling more experimental units or remeasuring existing ones (e.g., permanent sample plots). Given that empirical forest growth and yield models are commonly parameterized using national forest inventory datasets [e.g., the Forest Vegetation Simulator (Crookston and Dixon 2005)], opportunities for using existing models to make estimates with new data are abundant. Although this analysis used SDT data, similar applications could be implemented in other ecological data that exhibit zero-inflation: the presence/abundance of invasive species and natural disturbance agents, tree mortality, and forest ingrowth are a few such examples. Results from this analysis show that if researchers and managers wish to use existing random effects models for quantifying ecosystem structure, global parameters may continue to be used to represent population averages while local random effects can be updated with new data to arrive at improved estimates for the subject considered.

Acknowledgments The author thanks Brian Gray for useful comments and conversation related to these analyses. Two anonymous reviewers helped to improve the content of this work.

References

An HS, MacFarlane DW (2013) Comparing a new model-based method to fixed-area sampling for estimating the abundance of standing dead trees. Forestry 86:231–239

Bater CW, Coops NC, Gergel SE, LeMay V, Collins D (2009) Estimation of standing dead tree class distributions in northwest coastal forests using lidar remote sensing. Can J For Res 39:1080–1091

Crookston NL, Dixon GE (2005) The forest vegetation simulator: a review of its applications, structure, and content. Comput Electron Agr 49:60–80

De la Cruz-Mesía R, Marshall G (2006) Non-linear random effects models with continuous time autoregressive errors: a Bayesian approach. Stat Med 25:1471–1484

Ellison AM (2004) Bayesian inference in ecology. Ecol Lett 7:509–520

Eskelson BNI, Temesgen H, Barrett TM (2009) Estimating cavity tree and snag abundance using negative binomial regression models and nearest neighbor imputation methods. Can J For Res 39:1749–1765

Eskelson BNI, Temesgen H, Hagar JC (2012) A comparison of selected parametric and imputation methods for estimating snag density and snag quality attributes. For Ecol Manage 272:26–34

Forest Inventory and Analysis (2012) FIA DataMart: FIADB version 5.1. http://apps.fs.fed.us/fiadb-downloads/datamart.html. Accessed 11 Oct. 2012

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

Frelich LE (2002) Forest dynamics and disturbance regimes. Cambridge University Press, Cambridge

Gray BR (2005) Selecting a distributional assumption for modelling relative densities of benthic macroinvertebrates. Ecol Model 185:1–12

Harmon ME, Franklin JF, Swanson FJ, Sollins P, Gregory SV, Lattin JD, Anderson NH, Cline SP, Aumen NG, Sedell JR, Lienkaemper GW, Cromack K, Cummins KW (1986) Ecology of coarse woody debris in temperate ecosystems. Adv Ecol Res 15:133–302

Harmon ME, Woodall CW, Fasth B, Sexton J, Yatkov M (2011) Differences between standing and downed dead tree wood density reduction factors: a comparison across decay classes and tree species. USDA For. Serv. Res. Pap. NRS-15

Iwashita DK, Litton CM, Giardina CP (2013) Coarse woody debris carbon storage across a mean annual temperature gradient in tropical montane wet forest. For Ecol Manag 291:336–343

Kass RE, Carlin BP, Gelman A, Neal R (1998) Markov chain Monte Carlo in practice: a roundtable discussion 52, 93–100. Amer, Stat 52

Kennedy RS, Spies TA, Gregory MJ (2008) Relationships of dead wood patterns with biophysical characteristics and ownership according to scale in Coastal Oregon, USA. Landscape Ecol 23:55–68

Kenning RS, Ducey MJ, Brissette JC, Gove JH (2005) Field efficiency and bias of snag inventory methods. Can J For Res 35:2900–2910

Li R, Stewart B, Weiskittel A (2011a) A Bayesian approach for modelling non-linear longitudinal/hierarchical data with random effects in forestry. Forestry 85:17–25

Li R, Weiskittel AR, Kershaw JA Jr (2011b) 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. Can J For Res 41:2077–2089

McCarthy MA (2007) Bayesian methods for ecology. Cambridge Univ. Press, Cambridge

Ohmann JL, Gregory MJ, Spies TA (2007) Influence of environment, disturbance, and ownership on forest vegetation of Coastal Oregon. Ecol Appl 17:18–33

Rehfeldt GE (2006) A spline model of climate for the western United States, USDA For. Serv. Gen. Tech. Rep. RMRS-165

Russell MB, Kenefic LS, Weiskittel AR, Puhlick JJ, Brissette JC (2012) Assessing and modeling standing deadwood attributes under alternative silvicultural regimes in the Acadian Forest region of Maine, USA. Can J For Res 42:1873–1883

SAS Institute Inc. (2011) SAS/STAT(R) 9.3 user's guide. SAS Institute Inc, Cary, NC

Snijders TAB, Bosker RJ (2012) Multilevel analysis: an introduction to basic and advanced multilevel modeling, 2nd edn. Sage Publishers, London

Spiegelhalter DJ, Best NG, Carlin BP, Van der Linde A (2002) Bayesian measures of model complexity and fit (with discussion). J R Stat Soc Ser B 64:583–639

Stokland JN, Siitonen J, Jonsson BG (2012) Biodiversity in dead wood. Cambridge University Press, Cambridge

Sustainable Forestry Initiative (2004) Sustainable Forestry Initiative Standard, 2005–2009. The Sustainable Forestry Initiative Inc.
Temesgen H, Barrett TM, Latta GM (2008a) Estimating cavity tree abundance using nearest neighbor imputation methods for western Oregon and Washington forests. Silva Fenn 42:337–354
Temesgen H, Monleon VJ, Hann DW (2008b) Analysis and comparison of nonlinear tree height prediction strategies for Douglas-fir forests. Can J For Res 38:553–565
USDA Forest Service (2013) Research on forest climate change: potential effects of global warming on forests and plant climate relationships in western North America and Mexico. Rocky Mountain Research Station, Moscow Laboratory. http://forest.moscowfsl.wsu.edu/climate/. Accessed 4 Sept. 2013
van Oijen M, Reyer C, Bohn FJ, Cameron DR, Deckmyn G, Flechsig M, Härkönen S, Hartig F, Huth A, Kiviste A, Lasch P, Mäkelä A, Mette T, Minunno F, Rammer W (2013) Bayesian calibration, comparison and averaging of six forest models, using data from Scots pine stands across Europe. For Ecol Manag 289:255–268
Weiskittel AR, Hann DW, Kerhsaw JA Jr, Vanclay JK (2011) Forest growth and yield modeling. Wiley, London
Welsh AH, Cunningham RB, Donnelly CF, Lindenmayer DB (1996) Modelling the abundance of rare species: Statistical models for counts with extra zeros. Ecol Model 88:297–308
Winkelmann R (2008) Econometric analysis of count data. Springer, Berlin
Woodall CW, Domke GM, MacFarlane DW, Oswalt CO (2012) Comparing field- and model-based standing dead tree carbon stock estimates across forests of the US. Forestry 85:125–133
Woodall CW, Liknes GC (2008a) Climatic regions as an indicator of forest coarse and fine woody debris carbon stocks in the United States. Carbon Balance Manag 3:5
Woodall CW, Liknes GC (2008b) Relationships between forest fine and coarse woody debris carbon stocks across latitudinal gradients in the United States as an indicator of climate change effects. Ecol Indic 8:686–690
Woodall CW, Rondeux J, Verkerk P, Ståhl G (2009) Estimating dead wood during national inventories: a review of inventory methodologies and suggestions for harmonization. Environ Manag 44:624–631
Woudenberg SW, Conkling BL, O'Connell BM, LaPoint EB, Turner JA, Waddell KL (2010) The Forest Inventory and Analysis Database: database description and users manual version 4.0 for phase 2. US Department of Agriculture, Forest Service Gen. Tech. Rep. RMRS-24
Zhang J, He CZ (2013) A zero-inflated Bayesian spatial model with near-locating measurements. Environ Ecol Stat 20:609–631

Matthew B. Russell is an Assistant Professor in the Department of Forest Resources at the University of Minnesota. He obtained his Ph.D. in Forest Biometrics for the University of Maine.