# Stochastic or deterministic single-tree models: is there any difference in growth predictions?

M. Fortin, L. Langevin

Annals of Forest Science (2012) 69:271–282
DOI 10.1007/s13595-011-0112-0

ORIGINAL PAPER

# Stochastic or deterministic single-tree models: is there any difference in growth predictions?

**Mathieu Fortin · Luc Langevin**

Received: 28 March 2011 / Accepted: 24 June 2011 / Published online: 21 July 2011


**Abstract**
• ***Introduction*** Deterministic single-tree models are commonly used in forestry. However, there is evidence that stochastic events may interact with the nonlinear mechanisms that underlie forest growth. As a consequence, stochastic and deterministic simulations could yield different results for the same single-tree model and the same initial conditions. This hypothesis was tested in this study.
• ***Material and methods*** We used a single-tree growth model that can be implemented either stochastically or deterministically. Two data sets of 186 and 342 plots each were used for the comparisons. For each plot, the simulations were run on a 100-year period using 10-year growth steps. Three different response variables were compared.
• ***Results*** The results showed that there were differences between the predictions from stochastic and deterministic simulations for some response variables and that randomness alone could not explain these differences. In the case of deterministic simulations, the fact that predictions are reinserted into the model at each growth step is a concern. These predictions are actually random variables and their transformations may result in biased quantities. Forest growth modellers should be aware that deterministic simulations may not correspond to the mathematical expectation of the natural dynamics.

**Keywords** Growth modelling · Monte Carlo simulation · Bias · Single-tree models · Stochastic · Deterministic

**Handling Editor**: Daniel Auclair

M. Fortin
INRA, AgroParisTech, UMR1092,
Laboratoire d'Etude des Ressources Forêt-Bois (LERFoB),
Centre INRA de Nancy,
54280 Champenoux, France

M. Fortin (✉) · L. Langevin
Direction de la recherche forestière,
Ministère des Ressources naturelles et de la Faune du Québec,
2700 Einstein,
Québec, QC, Canada G1P 3W8
e-mail: mathieu.fortin@nancy.inra.fr

L. Langevin
e-mail: luc.langevin@mrnf.gouv.qc.ca

## 1 Introduction

In forest management, growth and yield models are considered as essential tools because they provide insights into future forest conditions. Like most statistical models, growth and yield models can be used either deterministically or stochastically. Whether it is at the tree or the stand level, a deterministic model provides the expected growth, i.e., the mathematical expectation of the growth for a particular stand, whereas a stochastic model attempts to illustrate the natural variability of the growth by including random components (Vanclay 1994, p.7). Although forest managers might be more interested in the natural variability of the growth than the mean response alone, stochastic growth and yield models are still rarely used. Most models are used in a deterministic fashion or include only a few stochastic components (Stage 1973; Solomon et al. 1986; Mäkelä et al. 1997; Pretzsch et al. 2002).

The current situation can be explained by the difficulties related to stochastic implementation. Although alternative methods exist (Mowrer and Frayer 1986; Mowrer 1991), stochastic implementation usually relies on Monte Carlo methods. Random numbers are generated in order to provide realisations for the variable of interest. Generating a large number of realisations provides an assessment of the

variability of the response. Monte Carlo methods have been widely used in resampling methods in statistics, e.g., bootstrap methods (Efron and Tibshirani 1993).

Because numerous random numbers are needed to compute the different Monte Carlo realisations, stochastic simulations usually require greater computing power and are often time consuming. Moreover, the correlations between the error terms and the coefficients of the different submodels may be unknown, and assumptions therefore have to be made (Kangas 1999).

On the other hand, deterministic simulations quickly provide forest managers with predictions since they produce an expected mean growth in a single run. Nevertheless, useful information such as prediction uncertainty may not be available because the error propagation is complex and cannot generally be analytically evaluated.

Beyond the advantages and disadvantages of the deterministic and stochastic approaches, there seems to be more basic statistical concerns related to the choice of one or the other. For instance, Zhou and Buongiorno (2004) reported that the nonlinearity in a matrix growth model, i.e., a model expressed as a matrix that contains movement ratios from one diameter class to the other (see Vanclay 1994, p. 43), interacted with the simulation approach (deterministic or stochastic). As a result, deterministic predictions were different from stochastic ones, even though the model and the initial conditions were the same.

Although there is no evidence of such a phenomenon for single-tree growth models, it can be reasonably expected that they exhibit similar behaviour, especially when they are based on differential equations. As a matter of fact, most single-tree models are often iterative, i.e., predictions are reinserted into the model in order to obtain long-term growth projections. As a consequence, in deterministic simulations, some tree- and stand-level predictors are no longer observed after the first growth step. They are instead predicted and, thereby, become random variables. Even if these predictions are assumed to be reliable, their transformation might be biased. Unbiased quantities usually do not transform into unbiased quantities (Duan 1983). These biases may, therefore, cause some distortion in deterministic simulations.

Considering the popularity of single-tree growth models in forest science, determining whether or not these models are sensitive to the simulation approach is a major issue. In this paper, we tested the null hypothesis that the differences between stochastic and deterministic simulations are negligible. For the sake of comparison, we used ARTEMIS-2009, a general single-tree growth model for the province of Quebec, Canada, which implements both approaches (Fortin and Langevin 2010). Two sets of sample plots served as benchmarks. The growth of these sample plots was simulated over 100 years using each approach, and stochastic and deterministic predictions were then compared for several plot variables.

Our intent was not to thoroughly evaluate the model but rather to evaluate the effect of the approach over the predictions. The model has already been evaluated in Fortin and Langevin (2010). Consequently, the discussion focuses on statistical and practical issues related to the use of the stochastic and the deterministic approaches.

## 2 Methods

### 2.1 Model description

In 2007, a research project was undertaken in the province of Quebec, Canada, with the aim of providing forest managers with a general, single-tree model for the 25 most important potential forest types of the commercial forest. Potential forest type refers to the forest composition that is expected in late successional stages (Saucier et al. 1998). Stands belonging to the same potential forest type are assumed to have similar forest dynamics.

The model, which is called ARTEMIS-2009, considers all trees of commercial and non-commercial species with diameter at breast height (dbh at 1.3 m) equal to or greater than 9.1 cm. It is designed for 400-m$^2$ sample plots, which is the standard plot area in the provincial inventory. The model encompasses four dynamic submodels (Fig. 1) that predict mortality, dbh increment, recruitment and recruit dbh, respectively, given that recruitment occurred for a particular growth step, which is set to 10 years. These four submodels were all designed to predict a response variable (mortality, increment, recruitment and recruit dbh) at the end of a particular growth step from predictors that are evaluated at the beginning of the same growth step. Growth forecasts for longer periods are obtained by reinserting predictions into the model. Tree heights and volumes are estimated using a model of the height/diameter relationship and a general volume model that can be found in Fortin et al. (2007, 2009a). The model was implemented in the CAPSIS platform (http://capsis.cirad.fr/).

All submodels belong to the linear or generalised linear type category. Special care was taken during the fitting process to make sure that the distributional assumptions were valid. Using the provincial network of permanent sample plots, the submodels were independently fitted to the potential forest types, thereby providing 25 different versions of the model. The consistency of the whole model was assessed at the plot level using the all-species basal area as the reference variable (Fortin and Langevin 2010). A cross-validation was carried out and did not reveal any major lack of fit. In this case study, we focused on two

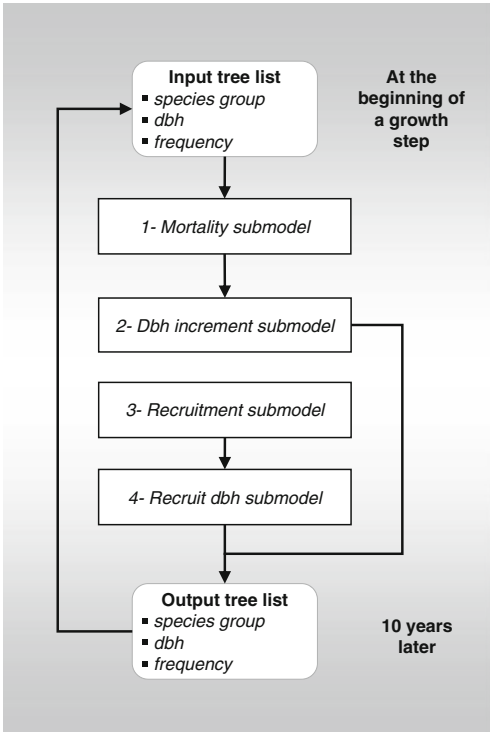


**Fig. 1** Flowchart of ARTEMIS-2009 model

major potential forest types for the sake of comparison: the sugar maple–yellow birch and the yellow birch–balsam fir.

The different parts of the model are briefly described in the following sections. The reader can refer to Fortin and Langevin (2010) for further details about the model. For a better understanding, we will define $i$, $j$, $k$ and $s$ as the plot, tree, growth step and species group indices, respectively.

## 2.2 Mortality submodel

The mortality submodel is of the generalised linear type with the assumption of a Bernoulli distributed response variable. Let us define $m_{\mathrm{ijk}}$ as a binary variable whose value is 1 when the tree $j$ in plot $i$ died during growth step $k$ or 0 otherwise. The model uses a complementary log–log link function in order to express the predictors in a linear fashion and to facilitate the mathematical tractability of the model (Fortin et al. 2008). Basically, the submodel expresses the probability of mortality $p_{ijk}$ for tree $j$ in plot $i$ during growth step $k$ as

$$m_{ijk} \sim \mathrm{Bernoulli}(p_{ijk}) \tag{1a}$$

$$p_{ijk} = 1 - e^{-e^{\mathbf{x}_{ijk}\boldsymbol{\beta}+\ln(\Delta t_{ik})}} \tag{1b}$$

where $\mathbf{x}_{ijk}$ is a row vector of predictors and class variables, $\boldsymbol{\beta}$ is a column vector of fixed-effect parameters, and $\ln(\Delta t_{ik})$ is the natural logarithm of growth step duration (year), which is set to 10 years. The predictors and class variables included in $\mathbf{x}_{ijk}$ are the species group, dbh, the natural logarithm of dbh, the basal area of trees with dbh larger than tree $j$, the occurrence of spruce budworm outbreak, the 1971–2000 mean annual precipitation, the occurrence of partial cutting, as well as some interactions between these variables.

## 2.3 Dbh increment submodel

This submodel is a linear mixed-effects model including some random effects and a covariance structure. Let $\Delta d_{ijk}$ be the dbh increment (cm) for tree $j$ in plot $i$ during growth step $k$. A logarithmic transformation of the response variable was required to ensure the validity of the assumptions of normality and homogeneous variances. The submodel can be expressed as follows

$$\ln(\Delta d_{ijk} + 1) = \mathbf{x}_{ijk}\boldsymbol{\beta} + u_i + u_{ik} + \varepsilon_{ijk} \tag{2a}$$

$$u_i \sim N(0, \sigma^2_{\mathrm{plot}}) \tag{2b}$$

$$u_{ik} \sim N(0, \sigma^2_{\mathrm{step}}) \tag{2c}$$

$$\varepsilon_{ij} = (\varepsilon_{ij1}, \varepsilon_{ij2}, \varepsilon_{ij3}, \ldots)^T \sim \mathrm{MVN}(\mathbf{0}, \mathbf{R}_{ij}) \tag{2d}$$

where $\mathbf{x}_{ijk}$ is a vector of predictors and class variables, $\boldsymbol{\beta}$ is a vector of parameters, $u_i$ is a plot random effect that is normally distributed with mean 0 and variance $\sigma^2_{\mathrm{plot}}$, $u_{ik}$ is a growth step random effect that is also normally distributed with mean 0 and variance $\sigma^2_{\mathrm{step}}$, and $\varepsilon_{ijk}$ is a residual error term. The vector of within-tree residual error terms ($\boldsymbol{\varepsilon}_{ij}$) is assumed to follow a multivariate normal (MVN) distribution with mean $\mathbf{0}$ and variance–covariance $\mathbf{R}_{ij}$. Matrix $\mathbf{R}_{ij}$ is actually the covariance matrix of the within-tree error terms and can be re-expressed in turn as the product of a within-tree variance ($\sigma^2_{\mathrm{tree}}$) and a correlation matrix ($\boldsymbol{\Psi}_{ij}$), i.e., $\mathbf{R}_{ij}=\sigma_{\mathrm{tree}}{}^2\boldsymbol{\Psi}_{ij}$. A correlation structure can be modelled in matrix $\boldsymbol{\Psi}_{ij}$. For this submodel, a linear log correlation structure yielded the best fit. For further details about this correlation structure, the reader can refer to Pinheiro and Bates (2000, p. 226) and Littell et al. (2006, p. 440). The predictors and class variables included in $\mathbf{x}_{ijk}$ are the species group, the dbh, the square of the dbh, the natural logarithm of growth step duration, the plot basal area, the occurrence of spruce budworm outbreaks, the occurrence of partial cutting, and the 1971–2000 mean annual precipitation, as well as some interactions between these variables.

### 2.4 Recruitment submodel

The recruitment submodel is a two-part conditional model. The first part predicts the occurrence of recruitment, whereas the second part aims at predicting the number of recruits conditional on the occurrence of recruitment. Such models are useful for modelling data with excessive zero counts (Cunningham and Lindenmayer 2005).

In this case study, the probability of observing a given number of recruits ($\mathrm{rec}_{iks}$) for species group $s$ in plot $i$ at the end of growth step $k$ is given by:

$$\Pr(\mathrm{rec}_{iks}) = \begin{cases} 1 - \pi_{iks} & \mathrm{rec}_{iks} = 0 \\ \pi_{iks} \cdot \varphi(\mathrm{rec}_{iks} - 1 | \mu_{iks}, \omega) & \mathrm{rec}_{iks} > 0 \end{cases} \quad (3a)$$

$$\pi_{iks} = \frac{e^{\mathbf{x}_{iks}\boldsymbol{\beta}}}{1 + e^{\mathbf{x}_{iks}\boldsymbol{\beta}}} \quad (3b)$$

$$\mu_{iks} = e^{\mathbf{g}_{iks}\boldsymbol{\gamma}} \quad (3c)$$

where $\pi_{iks}$ is the probability of occurrence of recruitment for species group $s$ in plot $i$ at the end of growth step $k$, $\varphi$ is the probability mass function of a negative binomial distribution with mean $\mu_{iks}$ and dispersion $\omega$, $\mathbf{x}_{iks}$ and $\mathbf{g}_{iks}$ are two vectors of predictors and class variables, and $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are two vectors of parameters. The predictors and class variables included in $\mathbf{x}_{iks}$ and $\mathbf{g}_{iks}$ are the species group, the plot basal area, the stem density of the species groups, the 1971–2000 mean annual precipitation, the 1971–2000 mean annual temperature, and the natural logarithm of growth step duration, as well as some interactions between these variables.

### 2.5 Recruit dbh submodel

Once recruitment occurs, the recruits are not necessarily integrated into the tree list with the minimum dbh, which is set to 9.1 cm. Consequently, recruit dbh had to be modelled. The recruit dbh submodel is a generalised linear model that assumes a Gamma distribution for the response variable. Let $\mathrm{dr}_{ijk}$ be this response variable for recruit $j$ in plot $i$ at the end of step $k$, i.e., the millimetres exceeding the 9.1 cm threshold required to be considered as a recruit. The submodel can be expressed as:

$$\mathrm{dr}_{ijk} \sim \mathrm{Gamma}(\mu_{ijk}, \omega) \quad (4a)$$

$$\mu_{ijk} = e^{\mathbf{x}_{ijk}\boldsymbol{\beta}} \quad (4b)$$

where $\mu_{ijk}$ and $\omega$ are the mean and dispersion parameters of the Gamma distribution, respectively, $\mathbf{x}_{ijk}$ is a row vector of predictors and class variables, and $\boldsymbol{\beta}$ is a column vector of parameters. The predictors and class variables included in $\mathbf{x}_{ijk}$ are the species group, the plot basal area, the occurrence of partial cutting, as well as some interactions between these variables.

### 2.6 Simulation approaches

#### *2.6.1 Stochastic approach*

The stochastic simulations in ARTEMIS-2009 rely on Monte Carlo methods. More specifically, random numbers are generated to account for the variability inherent in each submodel. Three sources of error are taken into account: (1) errors associated with parameter estimates; (2) random effects; and (3) residual errors. The distributions of these errors depend on the distributional assumptions behind each submodel. The estimators that were used to fit the different submodels provided the estimates for the parameters of the assumed distributions.

For each Monte Carlo realisation, random deviates were drawn from the assumed distributions. In the case of MVN distributions, a Cholesky factorisation was used to generate random vectors of deviates from these. An example is given in the Appendix. For example, before simulating the growth, a random normal deviate was drawn to account for the plot random effect in the dbh increment submodel (see Eq. 2b), and random vectors of deviates were also generated to take the errors in the parameter estimates into account.

Then, at the beginning of each growth step, a random normal deviate was drawn to act as a growth step random effect in the dbh increment submodel (see Eq. 2c). At this point, random deviates were drawn to account for the residual errors in each submodel. For example, in the mortality submodel, a random number uniformly distributed over the interval [0,1] was generated for each tree in the tree list. If the number was lower than the predicted probability of mortality $p_{ijk}$ (see Eq. 1a), the tree was considered to be dead. For survivor trees, normally distributed numbers were generated to account for the residual error in the dbh increment model.

When all trees in the tree list had been processed, a random number uniformly distributed over the interval [0,1] was drawn again for each species group in order to determine if some recruitment had occurred. If the random number was lower than the probability $\pi_{iks}$ (see Eq. 3a), it was assumed that recruitment had occurred for species group $s$. Conditional on the occurrence of recruitment, some random numbers were drawn from negative binomial distributions (see Eq. 3a) in order to predict the number of recruits for each species group. A random Gamma-distributed number was then generated for each recruit in order to obtain its dbh (see Eq. 4a). The results of this

growth step were then reinserted into the model until the desired projected length was reached.

The uncertainty of the stochastic simulations was tested by comparing nominal and real confidence interval coverages as proposed in Fortin et al. (2009b). The 95% confidence intervals were found to have a 90–91% real coverage, which indicates that the uncertainty predicted by the stochastic simulations was close to the observed uncertainty (unpublished result).

### 2.6.2 Deterministic approach

Deterministic simulations were run without generating any random deviate. To account for mortality, the expansion factor of a given tree record, which is the number of stems represented by this record, was reduced by the proportion predicted by the mortality submodel. Likewise, the expansion factor of the recruits was determined by the recruitment submodel. As for the dbh increments and the recruit dbh, they were set to the mathematical expectation of the two submodels.

For the dbh increment submodel, however, the mathematical expectation of the increment involves a back transformation because the submodel predicts the log of the dbh increment and not the dbh increment (see Eq. 2a). To correct for the bias induced by the logarithmic transformation, a naive correction inspired from Flewelling and Pienaar (1981) was applied to the predictions as follows:

$$E[\Delta d_{ijk}] = e^{\mathbf{x}_{ijk}\boldsymbol{\beta}+\left(\sigma^2_{\text{plot}}+\sigma^2_{\text{step}}+\sigma^2_{\text{tree}}\right)/2} - 1 \tag{5}$$

where $E[\cdot]$ is the mathematical expectation and $\left(\sigma^2_{\text{plot}} + \sigma^2_{\text{step}} + \sigma^2_{\text{tree}}\right)/2$ is the correction.

For the recruit dbh submodel, the mathematical expectation (see Eq. 4a) corresponds to parameter $\mu_{ijk}$. As in the stochastic approach, the results are reinserted into the model for additional growth steps until the end of the projection is reached.

### 2.6.3 Data sets and comparison

We used two subsets of sample plots located in the forest management unit (FMU) 6152 for the comparison. The FMU 6152 is located in central Quebec, more specifically in the *Laurentides* region (46°51′ N, 74°26′ W). The unit covers an area of 149,992 ha. Among all the sample plots available in the FMU 6152, we selected those that belonged to the sugar maple–yellow birch and the yellow birch–balsam fir potential forest types for a total of 186 and 342 sample plots, respectively.

The plots were established following the standards of the provincial inventory. Within these fixed-radius 400-m$^2$ plots, all trees with dbh greater than or equal to 9.1 cm had their species identified and their dbh measured according to 2-cm diameter classes. For the sake of comparison, we assumed that dbh measurements and species identifications were error free. A summary of the data set is shown in Table 1.

Three variables, the all-species stem density, the all-species basal area and the all-species merchantable volume, were selected as reference variables for the comparison between the predictions of the stochastic and deterministic simulations. For each individual plot, stochastic and deterministic simulations were run using the plot measurements as initial tree lists. In order to compare the different methods on the long term, we ran 100-year simulations starting from 2011 using 10-year growth steps. During the stochastic simulations, 1,000 Monte Carlo realisations were carried out to illustrate the variability of the three response variables for each individual plot to the greatest extent possible. During preliminary trials, this number of 1,000 Monte Carlo realisations had been found to be large enough to ensure the stability of the results. The predictions were estimated as the mean of these 1,000 Monte Carlo realisations. Deterministic simulations were run as described in the previous section. They produced a single simulated tree list from which the predictions of the three response variables were calculated. Once all the simulations were over, the predictions of the different approaches were compared.

An interesting point was to compare the simulations with respect to the initial conditions. Consequently, we defined three classes of initial conditions for each variable tested. For stem density, the classes of initial stem density ($dens_0$) used were: $dens_0<400$, $400\leq dens_0\leq 600$ and $dens_0>600$ stems ha$^{-1}$. For basal area, the classes of initial basal area ($BA_0$) were defined as $BA_0<15$, $15\leq BA_0\leq 25$, and $BA_0>25$ m$^2$ ha$^{-1}$. Finally, for volume, the classes of initial volume ($V_0$) were $V_0<100$, $100\leq V_0\leq 200$ and $V_0>200$ m$^3$ ha$^{-1}$. The results of the stochastic approach were considered as the reference for the comparison. Student's *t* tests were performed on the differences between the stochastic and deterministic predictions under the null hypothesis that there was no difference.

## 3 Results

The results of the 100-year stochastic and deterministic simulations are shown in Table 2 with respect to the different initial conditions. For any response variable and potential forest type, the average 100-year predictions were convergent. In other words, the differences across the initial condition groups were much lower at the end of the projections than they were at the beginning.

**Table 1** Summary of the two data sets used in the comparison of stochastic and deterministic simulations

| Potential forest types and variables | Minimum | Mean | Maximum |
|---|---|---|---|
| Sugar maple–yellow birch ($n$=186) | | | |
| Stem density (stem ha$^{-1}$) | | | |
| All species | 75.0 | 524.6 | 1,075.0 |
| Sugar maple | 0.0 | 298.7 | 975.0 |
| Yellow birch | 0.0 | 62.2 | 400.0 |
| Basal area (m$^2$ ha$^{-1}$) | | | |
| All species | 1.5 | 22.2 | 41.1 |
| Sugar maple | 0.0 | 12.6 | 34.8 |
| Yellow birch | 0.0 | 3.9 | 24.4 |
| Merchantable volume (m$^3$ ha$^{-1}$) | | | |
| All species | 7.7 | 164.1 | 338.2 |
| Sugar maple | 0.0 | 94.3 | 300.6 |
| Yellow birch | 0.0 | 30.1 | 228.8 |
| Yellow birch–Balsam fir ($n$=342)[a] | | | |
| Stem density (stem ha$^{-1}$) | | | |
| All species | 25.0 | 724.0 | 1,975.0 |
| Yellow birch | 0.0 | 79.9 | 550.0 |
| Balsam fir | 0.0 | 258.4 | 1,350.0 |
| Basal area (m$^2$ ha$^{-1}$) | | | |
| All species | 0.6 | 20.2 | 44.8 |
| Yellow birch | 0.0 | 4.7 | 24.5 |
| Balsam fir | 0.0 | 5.3 | 29.3 |
| Merchantable volume (m$^3$ ha$^{-1}$) | | | |
| All species | 1.9 | 130.7 | 354.5 |
| Yellow birch | 0.0 | 35.3 | 207.8 |
| Balsam fir | 0.0 | 29.8 | 194.0 |

$n$ number of sample plots, *merchantable volume* the volume from 15 cm in height up to a minimum diameter under bark of 9.0 cm

For the sugar maple–yellow birch potential forest type, the basal area and the volume predictions converged towards 25–27 and 212–225 m$^3$ ha$^{-1}$. The stem density predictions exhibited more variability, but there was no clear trend between the initial conditions and the predicted values. In fact, the initial stem density between 400 and 600 stems ha$^{-1}$ was the group for which the predicted stem density was the lowest when compared with the other two.

Regarding the differences between stochastic and deterministic predictions, the deterministic approach yielded lower stem densities but higher basal areas and volumes. Taking the stochastic approach as a reference, the relative differences ranged from −5.1% to −2.2% for stem densities and from 2.5% to 4.8% for basal area and volume. The differences decreased with increases of initial density, basal area or volume. According to the results of the $t$ tests, these differences were all highly significant.

For the yellow birch–balsam fir potential forest type, some trends could be distinguished. On average, lower initial densities yielded higher stem density predictions over the 100-year horizon. On the other hand, lower basal area and lower volume predictions could be associated with lower initial basal area and volume, respectively. With respect to the previous potential forest type, predicted stem densities were higher, whereas predicted basal areas and volumes were smaller.

The differences between the stochastic and deterministic predictions revealed a different pattern. For basal area and volume, the differences between the approaches were very small and non-significant, except for the plots that had the largest initial merchantable volume. For the stem density, however, the deterministic approach yielded predictions that were an average of 5% lower than the stochastic approach, and all of these differences were highly significant.

The differences between the simulation approaches over time are shown in Figs. 2 and 3 with the stochastic approach taken as the reference. For the sugar maple–yellow birch potential forest type (Fig. 2), divergences could be observed after the first growth step for most of the initial conditions except for higher stem densities, basal areas and volumes (Fig. 2c, f and i). For stem densities, the divergences increased over time (Fig. 2a, b). Volume and basal area deterministic predictions were initially lower than their stochastic counterparts during the first half of the simulations, but were higher at the end of the simulations (Fig. 2d, e, g and h).

**Table 2** Average 100-year stochastic and deterministic simulations of the stem density, basal area and merchantable volume as a function of the initial conditions and the potential forest types

| Potential forest types, variables and initial conditions | Number of sample plots | Stochastic | Deterministic | Difference |
|---|---|---|---|---|
| Sugar maple–yellow birch | | | | |
| Stem density (stems ha$^{-1}$) | | | | |
| $dens_0$<400 | 41 | 444.0 | 421.5 | −22.5[b] (−5.1%) |
| 400≤$dens_0$≤600 | 96 | 387.2 | 368.5 | −18.6[b] (−4.8%) |
| $dens_0$>600 | 49 | 414.9 | 405.7 | −9.2[a] (−2.2%) |
| Basal area (m$^2$ ha$^{-1}$) | | | | |
| $BA_0$<15 | 41 | 25.7 | 26.8 | 1.1[b] (4.3%) |
| 15≤$BA_0$≤25 | 73 | 25.2 | 26.1 | 0.8[b] (3.4%) |
| $BA_0$>25 | 72 | 25.9 | 26.6 | 0.7[b] (2.8%) |
| Merchantable volume (m$^3$ ha$^{-1}$) | | | | |
| $V_0$<100 | 43 | 214.7 | 224.9 | 10.2[b] (4.8%) |
| 100≤$V_0$≤200 | 89 | 213.9 | 220.8 | 7.0[b] (3.3%) |
| $V_0$>200 | 54 | 212.2 | 217.5 | 5.3[b] (2.5%) |
| Yellow birch–Balsam fir | | | | |
| Stem density (stems ha$^{-1}$) | | | | |
| $dens_0$<400 | 63 | 610.7 | 581.0 | −29.7[b] (−4.9%) |
| 400≤$dens_0$≤600 | 83 | 569.3 | 541.8 | −27.5[b] (−4.8%) |
| $dens_0$>600 | 196 | 564.2 | 538.5 | −25.7[b] (−4.5%) |
| Basal area (m$^2$ ha$^{-1}$) | | | | |
| $BA_0$<15 | 85 | 20.8 | 20.8 | 0.1 (0.3%) |
| 15≤$BA_0$≤25 | 157 | 21.6 | 21.7 | 0.0 (0.2%) |
| $BA_0$>25 | 100 | 22.6 | 22.6 | 0.0 (0.2%) |
| Merchantable volume (m$^3$ ha$^{-1}$) | | | | |
| $V_0$<100 | 108 | 145.5 | 145.8 | 0.3 (0.2%) |
| 100≤$V_0$≤200 | 196 | 157.1 | 157.1 | 0.1 (0.0%) |
| $V_0$>200 | 38 | 165.1 | 162.9 | −2.2[a] (−1.3%) |

$dens_0$ initial stem density, $BA_0$ initial basal area, $V_0$ initial merchantable volume

[a] Significantly different from 0 at $\alpha$=0.01

[b] Significantly different from 0 at $\alpha$=0.001

For the yellow birch–balsam fir potential forest type (Fig. 3), stochastic and deterministic predictions of basal area and volume were very close throughout the simulations in most cases (Fig. 3d–f, h and i). Simulations with lower initial volume were an exception to this rule since they exhibited divergences during the first 60 years (Fig. 3g). For stem densities, the divergences appeared earlier for lower initial densities but were nearly the same at the end of the simulations, regardless of the initial conditions (Fig. 3a–c).

## 4 Discussion

This study aimed at comparing stochastic and deterministic simulations under the null hypothesis that both approaches yield similar results in a context of single-tree growth modelling. Using ARTEMIS-2009, a growth model that implements both approaches, we compared the simulations in terms of predicted stem densities, basal areas and volumes. In forestry, Monte Carlo methods have been used mostly for error propagation, uncertainty assessment and sensitivity analysis (Gertner and Dzialowy 1984; Mowrer and Frayer 1986; Mowrer 1991; McRoberts et al. 1994), but the comparison between stochastic and deterministic predictions of single-tree models has rarely been addressed. To our knowledge, Ek (1980), Weber et al. (1986) and Vanclay (1991) are the only examples of such a comparison. Whereas they concluded (or assumed) that the differences between stochastic and deterministic predictions were small or negligible, we instead concluded that there may be substantial differences depending on the response variable and the model. In other words, for a given data set and a particular model, stochastic and deterministic predictions may not converge.

Several reasons may explain the fact that we obtained different results. In previous studies, the reference data sets were smaller, the stochastic simulations were based on fewer Monte Carlo realisations, or the forecast periods were shorter. For example, Weber et al. (1986) ran stochastic simulations based on ten Monte Carlo realisations and the comparison was performed on ten plots. Ek (1980) examined 10-year deterministic and stochastic projections for six plots, but only considered the mortality submodel as

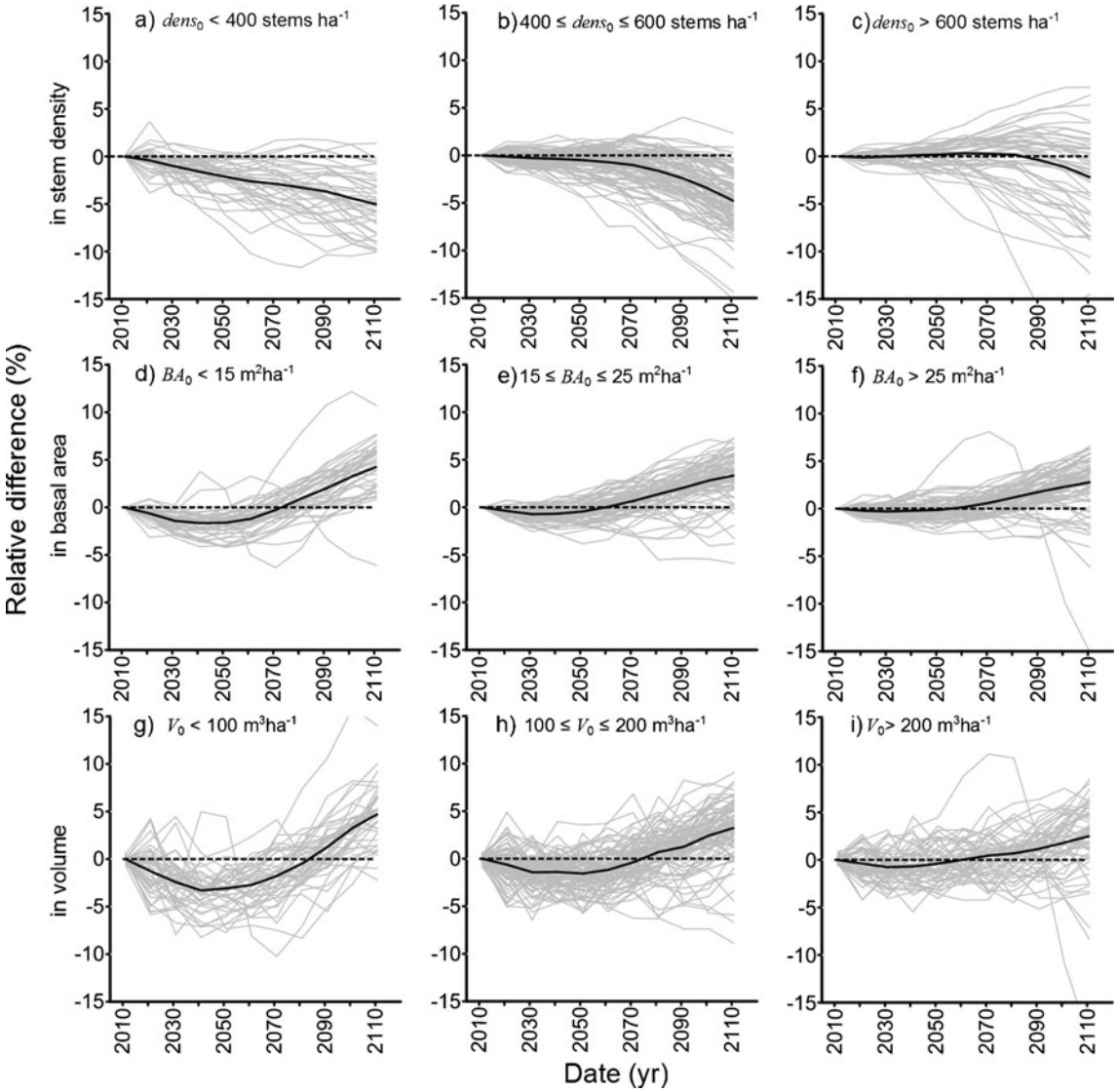


**Fig. 2** Relative differences between deterministic and stochastic simulations for 186 plots of the sugar maple–yellow birch forest type (average difference in *black*; $dens_0$ initial stem density, $BA_0$ initial basal area, $V_0$ initial merchantable volume)

stochastic. Vanclay (1991) showed that stochastic and deterministic simulations could be compatible for single plots in some cases but never performed any extensive comparison. At the time these studies were carried out, computing capacities were much more limited than they are today. With only a few plots and/or short-term projections, the true differences between stochastic and deterministic simulations cannot be distinguished from randomness.

With 1,000 Monte Carlo realisations per plot and at least 38 plots per group (Table 2), our study clearly shows that differences exist and that they cannot be attributed to randomness alone, considering the results of the $t$ tests. While comparing stochastic and deterministic simulations from a density-dependent transition matrix over 10,000 years, Zhou and Buongiorno (2004) also found large divergences. They concluded that the interaction between nonlinearity and stochasticity could have effects on the prediction of stand dynamics and stand structure.

On the basis of our experience, we concluded that this interaction is actually a matter of random variable transformation. Unbiased quantities on a given scale usually do not transform into unbiased quantities on a transformed scale (Duan 1983). These biases may intervene during any deterministic simulation if (1) some of the explanatory variables are actually derived from the transformation of other variables and/or (2) the effect of the explanatory variables on the response variable is nonlinear. The two different cases are explained in the following paragraphs.

Single-tree growth models usually rely on explanatory variables that are actually transformations of other variables. For example, basal area is derived from the sum of squared diameters. In statistics, it is well known that the

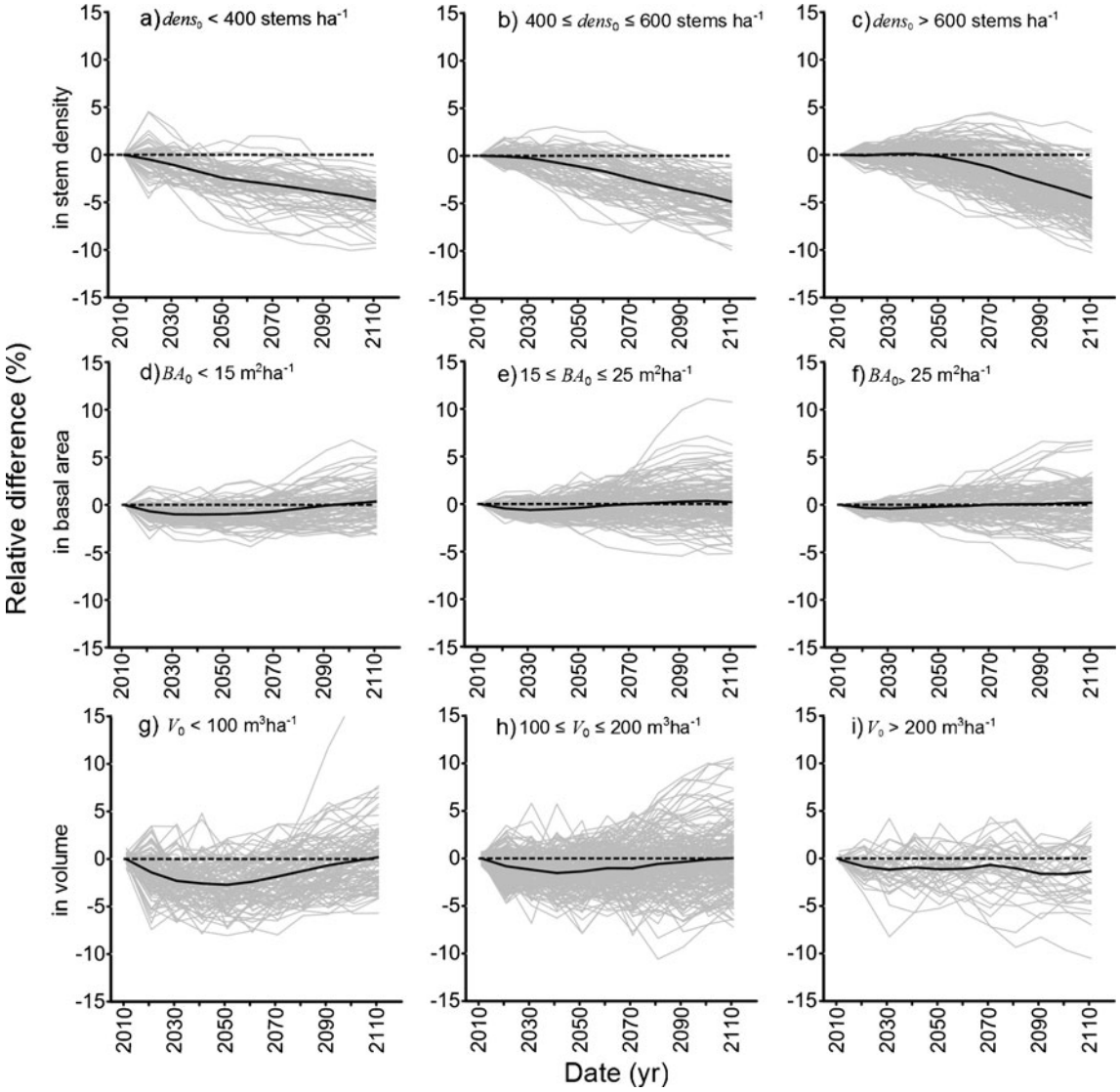


**Fig. 3** Relative differences between deterministic and stochastic simulations for 342 plots of the yellow birch–balsam fir forest type (average difference in *black*; $dens_0$ initial stem density, $BA_0$ initial basal area, $V_0$ initial merchantable volume)

squared mathematical expectation of a random variable underestimates the mathematical expectation of the squared random variable, i.e., $(E[Y])^2 < E[Y^2]$ with $Y$ being any random variable (Gregoire et al. 2008). In deterministic simulations, adding these underestimated quantities together to estimate basal area systematically leads to the underestimation of the true basal area, even after a single growth step. A crude estimator of this bias is the variance of the random variable (Gregoire et al. 2008). For basal area estimation, this bias would be proportional to the sum of the variances in the dbh increment predictions.

For both potential forest types, this basal area underestimation partially explains the differences between the deterministic and stochastic simulations in the first decades (Figs. 2d–f and 3d–f). The fact that the relative differences are larger for plots with lower initial basal areas (Figs. 2d and 3d) is in accordance with the idea of basal area underestimation. Lower initial basal areas in these plots lead to higher dbh increments. Since these dbh increments also exhibit larger variances, the basal area underestimation is relatively greater. Note that the differences between volume predictions exhibit a similar pattern because this variable is also derived from the square of the dbh.

The nonlinearity of the submodels may also induce some biases in the deterministic predictions. For generalised linear statistical models such as mortality, recruitment and recruit dbh submodels, the use of link functions results in nonlinear relationships between the predictors and the response variables. Except for the identity link function, a random error on a given predictor has a direct effect on the predictions. In general, the mathematical expectation of the response variable no longer matches the result of the link

function based on the mathematical expectation of the predictor, i.e., $E[\phi(\mathbf{x}\boldsymbol{\beta})] \neq \phi(E[\mathbf{x}\boldsymbol{\beta}])$, where $\phi(\cdot)$ is the link function of the argument, $\mathbf{x}$ is the row vector of predictors with some being random variables, and $\boldsymbol{\beta}$ is a column vector of parameters. These errors in predictors can actually be seen as random effects. McCulloch et al. (2008, p. 190) provide a good example of the aforementioned inequality for generalised linear models. Note that the same development applies to the errors in the parameter estimates.

Estimating the magnitude and the sign of these biases over the response variable is rather complex. They both depend on the distribution of the errors in the predictors as well as the sign and the value of the parameter estimates associated with the predictors. These errors are propagated throughout the model over several growth steps, and the situation becomes so complex that the exact cause of the differences between stochastic and deterministic simulations can no longer be identified. Because the model relies on differential equations, the errors in a particular growth step may have multiple impacts on subsequent growth steps. In deterministic simulations, for example, underestimating basal area for a particular growth step results in higher dbh increments and higher recruitment in the following growth steps. The fact that the differences between the stochastic and deterministic simulations exhibit different patterns across the potential forest types (Figs. 2 and 3), especially for basal area and volume, is thought to be related to this complex error propagation.

We first thought that the errors in the parameter estimates could be responsible for the divergence between the approaches. To test this possibility, we ran simulations under the assumption that the model parameters were known, i.e., no random deviates were simulated for the parameter estimates. The results we obtained were very close to those shown in this paper. The only change we actually observed was that the differences for the first step were slightly smaller. However, for the subsequent growth steps, the differences were of the same magnitude. We explain this small impact of the parameter deviates by the fact that the model was fitted using large data sets. In this case study, 1,015 and 846 permanent plots were used to fit the model in the sugar maple–yellow birch and the yellow birch–balsam fir potential forest types, respectively (Fortin and Langevin 2010). Given that the uncertainty associated with the parameter estimates decreases with the number of observations, it is not surprising that the errors in the parameter estimates have such a small effect on the simulations. For models with less accurate parameter estimates, the effect would probably be greater.

The results of our study showed that differences exist between stochastic and deterministic predictions when using 10-year growth steps. Would these differences also exist for other growth step durations? For example, would the differences be greater or smaller if we had simulated the growth over 100 years using 1-year growth steps instead? Unfortunately, ARTEMIS-2009 cannot be used to test the interaction between the growth step duration and the simulation approach because the model was designed for 10-year growth steps.

The differences we observed are probably due to the interaction between the model uncertainty and the number of growth steps in the projections. The uncertainty is obviously closely related to the growth step duration. The predictions of 10-year growth steps are generally less accurate than the predictions of 1-year growth steps. However, we hypothesize that the differences between stochastic and deterministic predictions are more highly affected by the number of growth steps than by the model uncertainty. For example, deterministic and stochastic predictions over a century would be much closer if the model used a single 100-year growth step. Basically, there would no reinsertion of predicted variables into the model during such deterministic simulations. Moreover, the effect of the random effects and the errors in the parameter estimates would be easily assessable and a correction factor could be easily implemented.

On the other hand, if the model was based on 1-year growth steps, projections over a century would mean a hundred reinsertions of predicted variables during the deterministic simulations. Even if each 1-year prediction was more accurate than in the previous case, the error propagation would probably be out of control after a few growth steps and the risk of larger differences would arise. This remains a hypothesis to be tested in future studies.

We attempted to include correction factors in the deterministic simulations during preliminary trials in order to account for the errors in the explanatory variables. We estimated the variance on dbh predictions as the cumulative sum of the variance on the dbh increment throughout the growth steps. This variance on the dbh prediction served as a correction factor for the calculation of the basal area. Unfortunately, the results of these "corrected" deterministic simulations proved to be unsatisfactory. The correction factor did not, in fact, account for the correlation between log-transformed error terms associated with dbh increment predictions (see Eq. 2d) and, for this reason, it could be considered as being underestimated. Appropriate correction factors for deterministic single-tree growth models remain to be developed. However, considering the complexity of the error propagation in such models, stochastic simulations may be easier to implement and even less time consuming.

A major question remains: which approach is the best one? Unfortunately, long-term data sets are usually unavailable to test each approach with respect to real-world observed data. From a biological point of view, stochastic simulations are thought to produce consistent predictions

because they attempt to illustrate the natural variation we expect from forest stand growth (Vanclay 1994, p. 7). This is true as long as the simulations remain within the limits of the data that served to fit the different submodels. There is no guarantee that the predictions remain biologically consistent outside of the calibration range. In such cases, inconsistent stem densities, basal area or volume might be predicted and these can be partially responsible for the differences between the stochastic and the deterministic approaches.

The sugar maple–yellow birch and the yellow birch–balsam fir versions of ARTEMIS-2009 were fit to 1,015 and 846 permanent plots, respectively, with three to five measurements per plot. In spite of these large data sets, a few extreme inconsistent numbers of recruits would be rarely predicted during stochastic simulations. To prevent this, a biologically consistent limit was set and inconsistent predictions beyond that limit were brought back to the limit with a warning message displayed. During the Monte Carlo simulations we ran in this case study, the limit was reached approximately 15 times for each comparison data set, which means that the probabilities of occurrence are around $8.1\times 10^{-6}$ and $4.4\times10^{-6}$ for the sugar maple–yellow birch and the yellow birch–balsam fir potential forest type, respectively. Considering these low probabilities, inconsistent predictions had a negligible impact on our results and the differences between both approaches appear to be primarily due to the statistical concerns expressed above.

## 5 Conclusions

In population biology, the study of density-dependent deterministic models has been found to be highly problematic since small environmental disturbances can considerably alter the dynamics of deterministic biological mechanisms (Higgins et al. 1997). The deterministic simulation is much simpler but clearly omits the possibility of stochastic events, whereas they are considered to be part of natural stand dynamics. With a transition-matrix model, Zhou and Buongiorno (2004) clearly showed that small stochastic events maintained some scarce species, whereas deterministic simulations led to the conclusion that they would disappear.

Using a deterministic approach with single-tree models oversimplifies the natural dynamics. The iterative reinsertions of predicted explanatory variables into the model for long-term projections are a concern from a statistical standpoint since the complex error propagation due to variable transformation may lead to biases. In this case study, relative differences as large as 5% could be observed between the deterministic and the stochastic approaches for some response variables when forecasting growth over a 100-year horizon (Table 2). Considering these differences and the results of their associated *t* tests, we are forced to reject our null hypothesis and to conclude that deterministic and stochastic simulations may yield different predictions.

On the basis of the results of this study, it cannot be concluded that deterministic simulations are systematically biased due to random variable transformation. Our recommendation is instead that modellers should be aware that deterministic simulations may yield different results from stochastic simulations when using the same data and the same model. Unless the stochastic simulation produces inconsistent predictions, it would be more consistent with stand dynamics. Regardless of the level of reliability associated with the stochastic approach, there is no certainty that a deterministic model truly represents the expectation of the natural dynamics. Deterministic models should instead be considered as one of the possible evolutions that may or may not be close to the mathematical expectation of the natural dynamics.

**Acknowledgements** The authors are thankful to Mrs. Isabelle Auger of the Natural Resources and Wildlife Ministry of Quebec-Forest Research Branch (*Ministère des Ressources Naturelles et de la Faune du Québec-Direction de la Recherche Forestière*) and to two anonymous reviewers for their helpful comments on a preliminary version of this paper.

## Appendix

Random vector drawn from a multivariate normal distribution

Let us consider the vector $\varepsilon$ as a random vector that follows a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance $\mathbf{V}$, i.e., $\varepsilon \sim MVN(\boldsymbol{\mu},\mathbf{V})$. The Cholesky decomposition provides the lower triangular matrix $\mathbf{A}$ that satisfies the condition $\mathbf{V}=\mathbf{A}\mathbf{A}^T$. If $\mathbf{y}$ is a vector of independent standard normal random variates (i.e., all independently and normally distributed with mean 0 and variance 1), the sum $\boldsymbol{\mu}+\mathbf{A}\mathbf{y}$ yields a random vector from the desired multivariate distribution.

For example, let us consider the bivariate case where:

$$\boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ and } \mathbf{V} = \begin{bmatrix} 0.700 & 0.500 \\ 0.500 & 2.000 \end{bmatrix}$$

The Cholesky decomposition of $\mathbf{V}$ yields:

$$\mathbf{A} = \begin{bmatrix} 0.837 & 0.000 \\ 0.598 & 1.282 \end{bmatrix}$$

Now, let us draw a random vector of independent standard normal variates where:

$$\mathbf{y} = (0.246,\ -1.976)^{\mathrm{T}}$$

The sum of $\boldsymbol{\mu}+\mathbf{Ay}$, i.e.,

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0.837 & 0.000 \\ 0.598 & 1.282 \end{bmatrix} \times \begin{bmatrix} 0.246 \\ -1.976 \end{bmatrix} = \begin{bmatrix} 0.206 \\ -2.386 \end{bmatrix}$$

yields a vector that follows a MVN distribution with mean $\boldsymbol{\mu}$ and covariance $\mathbf{V}$.

## References


Cunningham RB, Lindenmayer DB (2005) Modeling count data of rare species: some statistical issues. Ecology 86:1135–1142

Duan N (1983) Smearing estimate: a nonparametric retransformation method. J Am Stat Assoc 78:605–610

Efron B, Tibshirani RJ (1993) An Introduction to the Bootstrap. Chapman & Hall, Boca Raton

Ek A.R. (1980) A preliminary trial of alternative methods for treating mortality in the multipurpose forest projection system (MFPS) model. Univ. Minn., Coll. For., Dep. For. Resour., Staff Pap. Ser. No. 8.

Flewelling JW, Pienaar LV (1981) Multiplicative regression with lognormal errors. For Sci 27:281–289

Fortin M., and Langevin L. (2010) ARTEMIS-2009: un modèle de croissance basé sur une approche par tiges individuelles pour les forêts du Québec. Gouvernement du Québec. Ministère des Ressources naturelles et de la Faune. Direction de la recherche forestière. Mémoire de recherche n°156. 48 p.

Fortin M, DeBlois J, Bernier S, Blais G (2007) Mise au point de tarif de cubage général pour les forêts québécoises: une approche pour mieux évaluer l'incertitude associée aux prévisions. For Chron 83:754–765

Fortin M, DeBlois J, Bédard S, Meunier S (2008) Modeling mortality probabilities of individual trees for different species groups in hardwood stands under uneven-aged management in southern Québec. Ann For Sci 65:205

Fortin M., Bernier S., Saucier J.-P., and Labbé F. (2009a) Une relation hauteur-diamètre tenant compte de l'influence de la station et du climat pour 20 espèces commerciales du Québec. Gouvernement du Québec. Ministère des Ressources naturelles et de la Faune. Direction de la recherche forestière. Mémoire de recherche n° 153. 22 p.

Fortin M, Bédard S, DeBlois J, Meunier S (2009b) Assessing and testing prediction uncertainty for single tree-based models: a case study applied to northern hardwood stands in southern Québec. Canada Ecol Modell 220:2770–2781

Gertner GZ, Dzialowy PJ (1984) Effects of measurement errors on an individual tree-based growth projection system. Can J For Res 14:311–316

Gregoire TG, Lin QF, Boudreau J, Nelson R (2008) Regression estimation following the square-root transformation of the response. For Sci 54:597–606

Higgins K, Hasting A, Sarvela JN, Botsford LW (1997) Stochastic dynamics and deterministic skeletons: population behaviour of Dungeness crab. Science 276:1431–1435

Kangas AS (1999) Methods for assessing uncertainty of growth and yield predictions. Can J For Res 29:1357–1364

Littell R.C., Milliken G.A., Stroup W.W., Wolfinger R.D., and Schabenberger O., 2006. SAS system for mixed models. 2nd edition. SAS Institute Inc., Cary. 814 p

Mäkelä A, Vanninen P, Ikonen V-P (1997) An application of process-based modelling to the development of branchiness in Scots pine. Silva Fennica 31:369–380

McCulloch CE, Searle SR, Neuhaus JM (2008) Generalized, linear, and mixed models. Second edition. Wiley, Hoboken, p 384

McRoberts RE, Hahn JT, Hefty GJ, Van Cleve JR (1994) Variation in forest inventory field measurement. Can J For Res 24:1766–1770

Mowrer HT (1991) Estimating components of propagated variance in growth simulation model projections. Can J For Res 21:379–386

Mowrer HT, Frayer WE (1986) Variance propagation in growth and yield projections. Can J For Res 16:1196–1200

Pinheiro JC, Bates DM (2000) Mixed-effects models in S and S-PLUS. Springer, New York, 528 p

Pretzsch H, Biber P, Ďurský J (2002) The single tree-based stand simulator SILVA: construction, application and evaluation. For Ecol Manage 162:3–21

Saucier J.-P., Bergeron J.-F., Grondin P., and Robitaille A. (1998) Les régions écologiques du Québec méridional: un des éléments du système hiérarchique de classification écologique du territoire mis au point par le ministère des Ressources naturelles. L'Aubelle, février-mars 1998, 8 p.

Stage A.R. (1973) Prognosis model for stand development. USDA For. Serv. Res. Pap. INT-137. 32 p.

Solomon DS, Hosmer RA, Hayslett HT Jr (1986) A two-stage matrix model for predicting growth of forest stands in the Northeast. Can J For Res 16:521–528

Vanclay JK (1991) Compatible deterministic and stochastic predictions by probabilistic modeling of individual trees. For Sci 37:1656–1663

Vanclay JK (1994) Modelling forest growth and yield, applications to mixed tropical forests. CAB International, Wallingford, 312 p

Weber LA, Ek AR, Droessler TD (1986) Comparison between stochastic and deterministic mortality estimation in an individual tree based stand growth model. Can J For Res 16:1139–1141

Zhou M, Buongiorno J (2004) Nonlinearity and noise interaction in a model of forest growth. Ecol Modell 180:291–304