



**HAL**  
open science

# Finding confidence limits on population growth rates: Bootstrap and analytic methods

Nicolas Picard, Pierrette Chagneau, Frédéric Mortier, Avner Bar-Hen

## ► To cite this version:

Nicolas Picard, Pierrette Chagneau, Frédéric Mortier, Avner Bar-Hen. Finding confidence limits on population growth rates: Bootstrap and analytic methods. *Mathematical Biosciences*, 2009, 219 (1), pp.23-31. 10.1016/j.mbs.2009.02.002 . hal-01197524

**HAL Id: hal-01197524**

**<https://hal.science/hal-01197524v1>**

Submitted on 3 Aug 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Finding confidence limits on population growth rates: Bootstrap and analytic methods

Nicolas Picard <sup>a,\*</sup>, Pierrette Chagneau <sup>a</sup>, Frédéric Mortier <sup>a</sup>, Avner Bar-Hen <sup>b</sup>

<sup>a</sup> CIRAD, Campus international de Baillarguet, 34398 Montpellier Cedex 5, France

<sup>b</sup> Université Paris Descartes, MAP 5, 45 rue des Saint Pères, 75270 Paris Cedex 06, France

When predicting population dynamics, the value of the prediction is not enough and should be accompanied by a confidence interval that integrates the whole chain of errors, from observations to predictions via the estimates of the parameters of the model. Matrix models are often used to predict the dynamics of age- or size-structured populations. Their parameters are vital rates. This study aims (1) at assessing the impact of the variability of observations on vital rates, and then on model's predictions, and (2) at comparing three methods for computing confidence intervals for values predicted from the models. The first method is the bootstrap. The second method is analytic and approximates the standard error of predictions by their asymptotic variance as the sample size tends to infinity. The third method combines use of the bootstrap to estimate the standard errors of vital rates with the analytical method to then estimate the errors of predictions from the model. Computations are done for an Usher matrix models that predicts the asymptotic (as time goes to infinity) stock recovery rate for three timber species in French Guiana. Little difference is found between the hybrid and the analytic method. Their estimates of bias and standard error converge towards the bootstrap estimates when the error on vital rates becomes small enough, which corresponds in the present case to a number of observations greater than 5000 trees.

## 1. Introduction

Matrix models are population dynamics models for structured populations [1]. They have been widely used to address the conservation and management of animal or plant species [2–6]. Structured populations are described by the number of individuals in each category of the structuring variable (denoted by vector  $\mathbf{N}(t)$ ). Transition rates between categories are gathered into a transition matrix  $\mathbf{A}$ . Time is discrete, and the state of the population at time  $t$  is related to that at time  $t + 1$  by the recurrence relationship:  $\mathbf{N}(t + 1) = \mathbf{A}\mathbf{N}(t)$ . Given an initial composition of the population,  $\mathbf{N}(0)$ , the composition any time later can be predicted as:

$$\mathbf{N}(t) = \mathbf{A}^t \mathbf{N}(0) \quad (1)$$

As  $t$  tends to infinity, the proportions of individuals in the different categories will grow at an exponential rate,  $\lambda$ , which is the dominant eigenvalue of  $\mathbf{A}$  [1]. The population growth rate  $\lambda$  is a model's prediction that is often used to assess if the population is decaying ( $\lambda < 1$ ) or increasing ( $\lambda > 1$ ).

Two types of matrix models are most often distinguished by biologists depending on whether the structuring variable is age or ontogenetic stage [7]. The former corresponds to Leslie models whereas the latter corresponds to Lefkovich models [8,9]. The Lefkovich matrix is the most general. The Leslie matrix has non null elements on its first row and on its sub-diagonal. A third type of matrix model that is familiar to foresters is the Usher model, for size-structured populations [10,11]. The Usher model has non null elements on its first row, its main diagonal and its sub-diagonal.

The parameters of the transition matrix  $\mathbf{A}$  are composed of individual growth, survival or fecundity rates, and are thus known as vital rates. As vital rates are estimated from experimental or census data, their estimates are subject to uncertainties. Then any prediction of the matrix model is also subject to uncertainty. This sampling uncertainty has often been disregarded [12] in preference to other sources of uncertainty such as environmental or demographic variability [3,5,13]. Yet assessing sampling uncertainty is essential to the statistical reasoning, including statistical tests and confidence intervals.

From a statistical point of view, the estimates of the vital rates are realized values of random variables whose distribution depends on the distribution of experimental or census data. The prediction of the matrix model is a function of the vital rates, and its estimator is obtained by plugging the estimator of the vital rates into this function (Fig. 1). Difficulties arise because the function

\* Corresponding author. Present addresses. CIRAD, BP 4035, Libreville, Gabon. Tel.: +241 07 92 12 87; fax: +241 44 47 90.

E-mail addresses: nicolas.picard@cirad.fr (N. Picard), pierrette.chagneau@cirad.fr (P. Chagneau), frederic.mortier@cirad.fr (F. Mortier), avner@math-info.univ-paris5.fr (A. Bar-Hen).

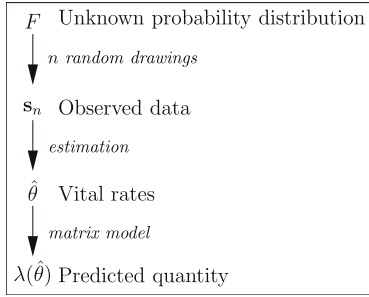


Fig. 1. Schematic diagram of the uncertainty chain showing how any prediction of the matrix model is a random variable whose distribution depends on the distribution of observations.

is often non-linear, thus leading to complicated relationships between the distribution of the vital rates and the distribution of the predicted quantity.

Alvarez and Slatkin [14] have reviewed three methods for computing sampling uncertainty on matrix models predictions: the analytic method, the Monte Carlo method, and resampling methods (including bootstrap, jackknife and related methods). The analytic method relies on a second order Taylor expansion of the predicted quantity with respect to the vital rates, and is approximate. Furthermore it requires that the moments of the vital rates are known. Monte Carlo and resampling methods substitute intensive computer simulations to approximation. The Monte Carlo method is based on the assumption that the distribution of vital rates is known, whereas it is supposed for resampling methods that a reference dataset is available.

Comparisons between the Monte Carlo method and the analytic method have been performed [15–17] and it was concluded that the analytic method is reliable if errors on vital rates remain small. These studies focused on the last arrow of the diagram shown in Fig. 1: they supposed that the uncertainty on the vital rates was known and assessed how this uncertainty contributed to the model's prediction. No link was made with the sampling variability of observations (the first two arrows of the diagram in Fig. 1). Comparisons between the bootstrap and the jackknife have also been performed (see references in [1,14]) and it was concluded that bootstrap should be preferred. However, as far as we know, no comparison has been made between the bootstrap method and the analytic method, integrating the whole chain of uncertainty from observations to predictions as shown in Fig. 1.

This study aims at comparing the analytic method and bootstrap to infer the bias, standard error and confidence interval of the population growth rate predicted by a matrix model for a size-structured population. A hybrid method, using the bootstrap to infer the uncertainty on vital rates from observations, and then the analytical method to incorporate this uncertainty into the model's prediction, will also be compared. These methods will be used to assess the sustainability of logging scenarios for the three major timber species in French Guiana.

## 2. Material and methods

### 2.1. Study site and focus species

Data for this study comes from the Paracou experimental site (5°18'N, 52°23'W) in French Guyana. Paracou is an experimental site dedicated to studying the effects of logging damage on stock recovery. The site lies in a *terra firme* rain forest on the coastal plain with equatorial climate. A dry season occurs from August to mid-November. From March to April, a short drier period interrupts the rainy season. The physiography of the site shows smooth

slopes incised by minor streams. Part of the site is covered by permanently waterlogged areas. The experimental design of the site consists of three blocks of four 300 × 300 m permanent sample plots with a 25 m inner buffer zone. In each central 250 × 250 m square, all trees over 10 cm dbh (diameter at breast height) were identified and georeferenced. Since 1984, girth at breast height, standing deaths, treefalls and newly recruited trees over 10 cm dbh have been monitored annually [18].

Three species were selected: angelique (*Dicorynia guianensis* Amshoff, Caesalpiniaceae), pink gonfolo (*Qualea rosea* Aublet, Vochysiaceae) and grey gonfolo (*Ruizterania albiflora* (Warming) Marcano-Berti, Vochysiaceae). These three species are the most important timber species in French Guiana. Angelique alone represents 34% of the total wood production whereas the two gonfolos represent 32% of it [19]. Angelique is a large canopy tree species endemic to the Guiana shield. Its ecological characteristics are described in [20]. Angelique trees are usually felled above a diameter at breast height (dbh) of 60 cm and natural populations are exploited with a 40-year felling cycle. Pink and grey gonfolos are also large canopy trees that are found on the Atlantic coast of South America. Their ecological characteristics are described in [21,22].

Data from 1992 to 1994 were used for this study. The time interval for subsequent computations thus is 2 years, and the length of the felling cycle corresponds to 20 time intervals. The number of observations was 467 for angelique, 484 for pink gonfolo, and 53 for grey gonfolo. Trees were distributed in diameter classes with equal width and highest bound 60 cm (which is the felling threshold). The number of classes was computed using Sturges' [23] formula. Hence angelique and pink gonfolo trees were distributed in  $m = 8$  diameter classes with breakpoints 10–17.1, 17.1–24.3, 24.3–31.4, 31.4–38.6, 38.6–45.7, 45.7–52.9, 52.9–60 cm and  $\geq 60$  cm, whereas grey gonfolo trees were distributed in  $m = 5$  diameter classes with breakpoints 10–22.5, 22.5–35.0, 35.0–47.5, 47.5–60 and  $\geq 60$  cm.

### 2.2. From observations to population growth rate

In this section, all the elements that compose the diagram shown in Fig. 1 are specified, starting from the matrix model up to the distribution of observations.

#### 2.2.1. Usher model

The Usher model is a matrix model for size-structured populations [10,11]. Size is divided into  $m$  size classes. The special feature of Usher models is Usher's hypothesis, that states that the dynamics of an individual between times  $t$  and  $t + 1$  comprises three, and only three, options: it stays alive in the same class, it stays alive and moves up to the next class, or it dies; moving up by more than one class or moving backwards is not allowed. The Usher hypothesis is appropriate when individual growth is slow and linear on short time intervals, which is the case for trees. It implies that the  $m \times m$  transition matrix,  $\mathbf{U}$ , has non-null elements only on its first row, its main diagonal and its sub-diagonal.

It is convenient to break down  $\mathbf{U}$  into:  $\mathbf{U} = \mathbf{GS} + \mathbf{R}$ , where  $\mathbf{G}$  (the growth matrix) is a stochastic  $m \times m$  matrix with non null elements on its diagonal and its sub-diagonal,  $\mathbf{S}$  (the survival matrix) is a diagonal  $m \times m$  matrix, and  $\mathbf{R}$  (the recruitment matrix) is a  $m \times m$  matrix with non-null elements on its first row. The  $i$ th element of the diagonal of  $\mathbf{G}$  is  $1 - p_i$  and the  $i$ th element of its sub-diagonal is  $p_i$ , where  $p_i$  is the conditional probability that an individual moves from class  $i$  to  $i + 1$  knowing that it stays alive. As no transition can occur beyond the last class,  $p_m = 0$ . The  $i$ th element of the diagonal of  $\mathbf{S}$  is  $1 - d_i$ , where  $d_i$  is the death probability for an individual in class  $i$ . The  $i$ th element of the first row of  $\mathbf{R}$  is  $f_i$ , where  $f_i$  is the fecundity rate of the individuals in class  $i$ . Estimating  $f_i$  requires to know which individual is the mother of any newly

recruited individual. In many situations, and in particular in forestry, this is not possible. Then, only an average fecundity rate  $f$  is used, that is to say:  $f_1 = \dots = f_m = f$ . It represents the ratio of the number of recruited individuals over the number of individuals in the population. Notice that if  $p_i = 1$  for all  $i = 1, \dots, m-1$ , the Usher matrix simplifies to a Leslie matrix. Hence the Leslie model is a special case of the Usher model. In summary, the Usher matrix has  $2m$  vital rates  $(f, d_1, \dots, d_m, p_1, \dots, p_{m-1})$ , that will be denoted in short  $\theta \in (\mathbf{R}^+)^{2m}$ .

### 2.2.2. Model's prediction: stock recovery rate

The model's prediction that motivated this study is the stock recovery rate, that is a forestry concept used to assess the sustainability of forest logging. The management of tropical natural forest is based on a periodic selective harvest of all trees with a diameter greater than a threshold called the "minimum harvest diameter" (MHD). The time interval between two successive logging operations, denoted  $T$ , is the length of the felling cycle. During this time, the forest is left so that the wood stock naturally regenerates. The exploitable wood stock is defined as the number of stems with a diameter greater than a threshold called the "minimum diameter for exploitation" (MDE). The MDE is legally fixed by national forest services whereas the MHD (which is necessarily greater than the MDE) and the felling cycle  $T$  are chosen by forest managers to ensure sustainability.

We assume that logging is instantaneous with respect to forest dynamics. Harvest is defined by a  $m \times m$  diagonal matrix,  $\mathbf{L}$ , such that the  $i$ th element of the diagonal of  $\mathbf{L}$  equals zero if the corresponding diameter class is harvested and one otherwise. Under a periodic felling regime with a felling every  $T$  time steps starting from 0, the composition of the population at the end of the  $k$ th felling cycle is given by:

$$\mathbf{N}(kT) = \mathbf{A}^k \mathbf{N}(0) \quad (2)$$

where  $\mathbf{A} = \mathbf{U}^T \mathbf{L}$ .  $\mathbf{A}$  is a transition matrix that relates the composition of the population at the end of a felling cycle to its composition at the end of the previous felling cycle. Eq. (2) is similar to (1) up to a change of time scale. Then, as  $k$  tends to infinity,  $\mathbf{N}(kT)$  will grow at an exponential rate,  $\lambda$ , that is the dominant eigenvalue of  $\mathbf{A}$ .

By definition, the stock recovery rate at the end of the  $k$ th felling cycle,  $X_k$ , is the ratio of the exploitable stock at time  $kT$  over the exploitable stock at time  $(k-1)T$ . Given the asymptotic behaviour of (2), the asymptotic stock recovery rate, that is the limit of  $X_k$  as  $k$  goes to infinity, identifies with the dominant eigenvalue of  $\mathbf{A}$ :

$$\lim_{k \rightarrow \infty} X_k = \lambda = \max_{\eta} \{\eta : \det(\mathbf{A} - \eta \mathbf{I}) = 0\}$$

In summary, the asymptotic stock recovery rate of the Usher model with transition matrix  $\mathbf{U}$  identifies with the population growth rate of the matrix model with transition matrix  $\mathbf{A} = \mathbf{U}^T \mathbf{L}$ . The population growth rate  $\lambda$  depends on the parameters of  $\mathbf{A}$  but is independent of the initial composition  $\mathbf{N}(0)$ . The asymptotic stock recovery rate thus depends on the  $2m$  vital rates  $\theta$  of  $\mathbf{U}$ . To highlight this dependency, we shall write  $\lambda(\theta)$ , where  $\lambda$  here denotes a function from  $(\mathbf{R}^+)^{2m}$  onto  $\mathbf{R}^+$ .

The stock recovery rate is used by forest managers to assess the sustainability of logging scenarios. A logging scenario is defined by the felling cycle  $T$  and the MHD (that is  $\mathbf{L}$ ). The asymptotic stock recovery rate corresponds to a long-term assessment of sustainability:  $\lambda = 1$  indicates that logging is sustainable;  $\lambda < 1$  indicates that the population is over-logged (then  $T$  or MHD should be increased);  $\lambda > 1$  indicates that logging does not remove the natural ingrowth. However, the estimate of  $\lambda$  is not an end in itself: it is necessary to see if one belongs to the confidence interval of  $\lambda$ .

### 2.2.3. Distribution of observations

In the context of size-structured populations, an observation consists of a pair  $(S_1, S_2)$  giving the size  $S_1$  of an individual at initial measurement time  $t_0$ , and the size  $S_2$  of the same individual at remeasurement time  $t_0 + 1$ . By convention,  $S_1 = 0$  when the individual is newly recruited at  $t_0 + 1$ , and  $S_2 = 0$  when the individual dies between  $t_0$  and  $t_0 + 1$  (but  $(0, 0)$  is not a valid observation). A dataset of size  $n$ , denoted  $\mathbf{s}_n$ , consists of  $n$  observations  $\{(S_{1k}, S_{2k})\}_{k=1, \dots, n}$ .

Size classes are defined by their breakpoints  $u_i$ . The  $i$ th size class is the interval  $[u_i, u_{i+1}[$ , where  $u_{m+1} = \infty$ . An individual is recruited when its size reaches the minimum size for inventory that is  $u_1$ . So  $S_1 = 0$  implies  $S_2 = u_1$ .

The unknown distribution of observations is denoted  $F$ . It can be broken down into:

$$F = \pi_0 \delta_{(0, u_1)} + \pi_1 F_1 + \pi_2 F_2 \quad (3)$$

where  $\pi_0 = \Pr(S_1 = 0, S_2 = u_1)$  is the probability for an observation to be a newly recruited individual,  $\pi_1 = \Pr(S_2 = 0)$  is the probability for an observation to be a dead remeasured individual,  $\pi_2 = 1 - \pi_0 - \pi_1 = \Pr(S_1 \neq 0, S_2 \neq 0)$  is the probability for an observation to be a live individual at both measurement times,  $\delta_o$  is the Dirac mass at observation  $o$ ,  $F_1(x) = \Pr(S_1 < x | S_2 = 0)$  is the univariate conditional size distribution at  $t_0$  knowing that the individuals are dead at  $t_0 + 1$ , and  $F_2(x, y) = \Pr(S_1 < x, S_2 < y | S_1 \neq 0, S_2 \neq 0)$  is the bivariate conditional distribution of sizes at  $t_0$  and size increments between  $t_0$  and  $t_0 + 1$  knowing that individuals are alive at  $t_0$  and  $t_0 + 1$ .

### 2.2.4. Vital rates estimators

The parameters  $\theta$  of the Usher model are estimated from a dataset  $\mathbf{s}_n$ . The estimate of  $\theta$  is the realized value of its estimator  $\hat{\theta}$ , which is a random variable whose distribution follows from  $F$ . Many estimators of vital rates exist, with contrasted performances in terms of statistical efficiency [24,25]. We selected a biased estimator with a low standard error [26]. The average fecundity rate was estimated using the classical proportion estimator [27]:

$$\hat{f} = \frac{\#\{S_1 = 0\}}{n - \#\{S_1 = 0\}} \quad (4)$$

where  $\#\{S_1 = 0\} = \sum_{k=1}^n \mathbf{1}_{\{S_{1k} = 0\}}$  is the number of newly recruited individuals at remeasurement time. Mortality rates were taken equal in all classes, which introduces a small bias but considerably reduces the standard error. The common mortality rate was estimated by the ratio-of-means estimator [27]:

$$\hat{d}_1 = \dots = \hat{d}_m = \hat{d} = \frac{\#\{S_2 = 0\}}{n - \#\{S_1 = 0\}} \quad (5)$$

Finally, the conditional upgrowth rate from class  $i$  to  $i+1$  was estimated using a revised version of the bias-corrected increment estimator [24–26]:

$$\hat{p}_i = 1 - G_{2i}(u_{i+1} - \hat{a}_i, \hat{\kappa}_i) \quad (6)$$

where  $\hat{a}_i$  is an estimator of the average size increment in class  $i$ , and  $G_{2i}(\cdot, \kappa)$  is a parametric expression depending on parameter  $\kappa$  for the conditional diameter distribution at  $t_0$  knowing that individuals are neither dead nor recruits and that they belong to size class  $i$  at  $t_0$ . The truncated exponential distribution on  $[u_i, u_{i+1}[$  was used for  $G_{2i}$  [28]. Its cumulative distribution function is:

$$G_{2i}(x, \kappa) = \begin{cases} \frac{1 - \exp[-\kappa(x - u_i)/(u_{i+1} - u_i)]}{1 - \exp(-\kappa)} & (\kappa \neq 0) \\ (x - u_i)/(u_{i+1} - u_i) & (\kappa = 0) \end{cases}$$

The average size increment was simply estimated using the empirical mean of size increments in class  $i$ :

$$\hat{a}_i = \frac{\sum_{k=1}^n (S_{2k} - S_{1k}) \mathbf{1}(u_i \leq S_{1k} < u_{i+1}) \mathbf{1}(S_{2k} \neq 0)}{\#\{u_i \leq S_1 < u_{i+1}, S_2 \neq 0\}}$$

whereas the maximum likelihood estimator of the  $\kappa$  parameter was used. Its expression is:  $\hat{\kappa}_i = h^{-1}[(\bar{S}_i - u_{i+1})/(u_{i+1} - u_i)]$ , where

$$\bar{S}_i = \frac{\sum_{k=1}^n S_{1k} \mathbf{1}(u_i \leq S_{1k} < u_{i+1}) \mathbf{1}(S_{2k} \neq 0)}{\#\{u_i \leq S_1 < u_{i+1}, S_2 \neq 0\}}$$

is the empirical mean of sizes at  $t_0$  in class  $i$  excluding dead trees, and  $h$  is a continuous monotone decreasing function (and thus invertible) from  $\mathbf{R}$  onto  $] -1, 0[$ , defined as:  $h(x) = x^{-1} - [1 - \exp(-x)]^{-1}$  if  $x \neq 0$  and  $h(0) = -0.5$ . Estimator (6) for growth rates makes use of the continuous information on size rather than simply counting the number of individuals that grow up from class  $i$  to  $i + 1$  [29–31]. It is a biased estimator but is still more efficient than the classical proportion estimator in terms of quadratic error [25]. Finally, the estimator of the asymptotic stock recovery rate was obtained by plugging the estimator  $\hat{\theta} = (\hat{f}, \hat{d}_1, \dots, \hat{d}_m, \hat{p}_1, \dots, \hat{p}_{m-1})$  of vital rates into  $\lambda$ , that is:  $\hat{\lambda} \equiv \lambda(\hat{\theta})$ .

### 2.3. Computing sampling uncertainty

The estimator  $\hat{\theta}$  of vital rates  $\theta$  is a random vector whose distribution depends on the distribution  $F$  of observations. Then any prediction of the Usher model, and in particular the asymptotic stock recovery rate  $\lambda(\hat{\theta})$ , is also a random variable whose distribution depends on  $F$ . Now that all components of the diagram in Fig. 1 have been specified, we clarify the methods for computing the uncertainty on model's predictions from the uncertainty on observations.

#### 2.3.1. Bootstrap method

Bootstrap methods basically consist in replacing in the diagram of Fig. 1 the unknown distribution  $F$  of observations by the empirical distribution of observations in dataset  $\mathbf{s}_n$ :  $\hat{F}_n = (1/n) \times \sum_{k=1}^n \delta_{(S_{1k}, S_{2k})}$  (where  $\delta_o$ , for any observation  $o$ , is the Dirac mass at  $o$ ). Drawing a sample of  $n$  independent and identically distributed observations according to  $\hat{F}_n$  is then equivalent with drawing randomly  $n$  observations with replacement from  $\mathbf{s}_n$ . Bootstrap algorithms for matrix models have been described in [1] and are a direct application of the algorithms presented by Efron and Tibshirani [32]. The only point to clarify to comply with Efron and Tibshirani's [32] notations is to prove that there exists a functional  $\mathcal{H}$  such that  $\theta = \mathcal{H}(F)$ . The plug-in estimator of  $\theta$  then can be defined as:  $\hat{\theta}_{\text{plug-in}} = \mathcal{H}(\hat{F}_n)$ .

Given a set of  $B$  bootstrap replications  $\lambda(\hat{\theta}_1^*), \dots, \lambda(\hat{\theta}_B^*)$ , the bias of  $\lambda(\hat{\theta})$  was estimated by [32, p. 125]:

$$\widehat{\text{bias}}_{\text{boot}} = \hat{\lambda}_*^* - \lambda(\hat{\theta}_{\text{plug-in}}) \quad (7)$$

where  $\hat{\lambda}_*^*$  is the empirical mean of the  $B$  replications:  $\hat{\lambda}_*^* = \{\sum_{b=1}^B \lambda(\hat{\theta}_b^*)\}/B$ . The standard error of  $\lambda(\hat{\theta})$  was estimated using the usual bootstrap estimator of standard error (see [1, Eq. (12.15), p. 306] or [32, algorithm 6.1 p.47]). The confidence interval of  $\lambda(\hat{\theta})$  was estimated using the percentile method [32, § 13.3, p.170]. All computations were done using  $B = 50,000$  replications. Bootstrap algorithms were implemented in C language interfaced with R software [33]. The code is available at <http://agents.cirad.fr/index.php/Nicolas+Picard>.

#### 2.3.2. Analytic method

Analytic methods approximate the variance of an estimator by  $\sigma^2/n$ , where  $n$  is sample size and  $\sigma^2$  is the asymptotic variance of the estimator (obtained as a limit when  $n$  tends to infinity). Assuming that there exists a functional  $\mathcal{G}$  with suitable properties for differentiability, such that  $\hat{\theta} = \mathcal{G}(\hat{F}_n)$  where  $\hat{F}_n$  is the empirical

distribution of the observations, the asymptotic covariance matrix  $\Sigma$  of the estimator  $\hat{\theta}$  is the  $2m \times 2m$  matrix:

$$\Sigma = \int \text{IC}_{\mathcal{G},F}(x) \text{IC}_{\mathcal{G},F}(x)^\top dF(x) \quad (8)$$

where  $\text{IC}_{\mathcal{G},F}(x)$  is the influence curve of  $\mathcal{G}$  at observation  $x$  relative to distribution  $F$  [34,35], and  $X^\top$  for any matrix or vector denotes the transpose of  $X$ . Given an observation  $x$ , the influence curve is the vector of length  $2m$  given by:

$$\text{IC}_{\mathcal{G},F}(x) = \lim_{\varepsilon \rightarrow 0} \frac{\mathcal{G}[(1-\varepsilon)F + \varepsilon\delta_x] - \mathcal{G}(F)}{\varepsilon} \quad (9)$$

where  $\delta_x$  is a Dirac mass at  $x$ . Assuming that  $\mathcal{G}$  exists, then functional  $\lambda \circ \mathcal{G}$  relates the estimator of the asymptotic stock recovery rate to the empirical distribution of the observations:  $\lambda(\hat{\theta}) = (\lambda \circ \mathcal{G})(\hat{F}_n)$ . The asymptotic variance of  $\lambda(\hat{\theta})$  can then be computed in the same way, using the influence curve.

The computation of asymptotic variances using the influence curve provides a rigorous mathematical justification of the second-order Taylor expansions (sometimes also referred as the  $\delta$ -method) previously used to compute the variances in an analytic way [14–17]. In particular it can be used to prove that the asymptotic variance of  $\lambda(\hat{\theta})$ , denoted  $\sigma_\lambda^2$ , is related to the asymptotic covariance matrix of  $\hat{\theta}$  by:

$$\sigma_\lambda^2 = (D_{\mathcal{G}}\lambda)\Sigma(D_{\mathcal{G}}\lambda)^\top \quad (10)$$

where  $D_{\mathcal{G}}\lambda$  is the Jacobian matrix of  $\lambda$  at  $\theta$ , that is a  $1 \times 2m$  matrix whose  $i$ th element is equal to the partial derivative of  $\lambda$  with respect to the  $i$ th element of  $\theta$ . The influence curve can also be used to state that  $\lambda(\hat{\theta})$  has an asymptotic normal distribution with mean  $\lambda[\mathcal{G}(F)]$ . Then the bias of  $\lambda(\hat{\theta})$  in the analytic approach is  $\lambda[\mathcal{G}(F)] - \lambda[\mathcal{H}(F)]$ , whereas its confidence interval are computed using the percentiles of the normal distribution.

The estimates of the asymptotic moments of  $\lambda(\hat{\theta})$  were obtained by replacing the unknown distribution  $F$  by the empirical distribution  $\hat{F}_n$ . Hence the bias was estimated as:

$$\widehat{\text{bias}}_\infty = \lambda(\hat{\theta}) - \lambda(\hat{\theta}_{\text{plug-in}})$$

and the estimate of  $\sigma_\lambda$  was:  $\hat{\sigma}_\lambda^2 = (D_{\hat{\theta}}\lambda)\widehat{\Sigma}(D_{\hat{\theta}}\lambda)^\top$ , where  $\widehat{\Sigma}$  is obtained by replacing  $F$  by  $\hat{F}_n$  in the expression of  $\Sigma$ .

#### 2.3.3. Hybrid method

The hybrid method makes a compromise between the accuracy of the bootstrap approach (whereas the analytic method is approximate) and the speed of computation of the analytic method (whereas the bootstrap method requires potentially long simulations). It is based on the fact that, once the covariance matrix of the vital rates estimator is known, the variance of any prediction of the matrix model can be computed from them (see last arrow of the diagram in Fig. 1). Hence the hybrid method uses the bootstrap approach to compute the moments of the vital rates estimator with good accuracy, and then uses the analytic method to compute the moments of  $\lambda(\hat{\theta})$ . Notice that with this approach, bootstrap simulations are needed once for all, whatever the prediction of the model to be subsequently studied is. The bias was estimated as:

$$\widehat{\text{bias}}_{\text{hyb}} = \lambda(\hat{\theta}_*^*) - \lambda(\hat{\theta}_{\text{plug-in}})$$

where  $\hat{\theta}_*^*$  is the empirical mean of the  $B$  bootstrap replications:  $\hat{\theta}_*^* = (\sum_{b=1}^B \hat{\theta}_b^*)/B$ . The variance of  $\lambda(\hat{\theta})$  was computed using the same formula as (10) but replacing  $\mathcal{G}(F)$  by  $\hat{\theta}_*^*$ , and replacing the asymptotic covariance matrix  $\Sigma$  of  $\hat{\theta}$  by its bootstrap estimate. The confidence interval was again computed using the percentiles of the normal distribution.

### 3. Results

#### 3.1. Plug-in estimator of vital rates

The relationship between the distribution of observations  $F$  and the vital rates  $\theta$  of the Usher model is given by the following Proposition.

**Proposition 1.** *There exists a functional  $\mathcal{H}$  such that  $\theta = \mathcal{H}(F)$ , where  $\theta$  are the parameters of the Usher model and  $F$  is the distribution of observations given by (3). The functional  $\mathcal{H}$  is defined by:*

- The first element of  $\mathcal{H}(F)$ , that corresponds to the mean fecundity rate, is:

$$[\mathcal{H}(F)]_1 \equiv f = \frac{\pi_0}{1 - \pi_0} \quad (11)$$

- The  $(i + 1)$ th element of  $\mathcal{H}(F)$  for  $i = 1, \dots, m$ , that corresponds to the mortality rate in class  $i$ , is:

$$[\mathcal{H}(F)]_{i+1} \equiv d_i = \pi_1 \frac{F_1(u_{i+1}) - F_1(u_i)}{G_1(u_{i+1}) - G_1(u_i)} \quad (i = 1, \dots, m) \quad (12)$$

where

$$G_1(x) = \pi_0 \delta_0 + \pi_1 F_1(x) + \pi_2 F_2(x, \infty)$$

is the size distribution at measurement time  $t_0$  (including a Dirac mass at 0 for individuals that will be recruited).

- The  $(m + 1 + i)$ th element of  $\mathcal{H}(F)$  for  $i = 1, \dots, m - 1$ , that corresponds to the conditional upgrowth rate from class  $i$  to  $i + 1$  is:

$$[\mathcal{H}(F)]_{m+1+i} \equiv p_i = \frac{F_2(u_{i+1}, u_{i+2}) - F_2(u_i, u_{i+2}) - F_2(u_{i+1}, u_{i+1}) + F_2(u_i, u_{i+1})}{F_2(u_{i+1}, \infty) - F_2(u_i, \infty)} \quad (i = 1, \dots, m - 1) \quad (13)$$

The proof is given in Appendix A. This Proposition then yields:

**Proposition 2.** *The plug-in estimator of  $\theta$  is given by:*

- The plug-in estimator of the mean fecundity rate is:

$$\hat{f}_{\text{plug-in}} = \frac{\#\{S_1 = 0\}}{n - \#\{S_1 = 0\}}$$

- The plug-in estimator of the mortality rate in class  $i$  is:

$$\hat{d}_{i\text{plug-in}} = \frac{\#\{u_i \leq S_1 < u_{i+1}, S_2 = 0\}}{\#\{u_i \leq S_1 < u_{i+1}\}}$$

- The plug-in estimator of the conditional upgrowth rate from class  $i$  to  $i + 1$  is:

$$\hat{p}_{i\text{plug-in}} = \frac{\#\{u_i \leq S_1 < u_{i+1}, u_{i+1} \leq S_2 < u_{i+2}\}}{\#\{u_i \leq S_1 < u_{i+1}, S_2 \neq 0\}}$$

Hence the plug-in estimator of  $\theta$  is the classical proportion estimator of the transition rates [24,27]. The estimator of  $f$  that we used is identical to its plug-in estimator. The estimator of  $d_i$  that we used is a weighted mean of the plug-in estimators of the  $d_i$ s. However the estimator of  $p_i$  that we used is completely different from its plug-in estimator. This means that  $\hat{d}_i$  and  $\hat{p}_i$  are biased. One may also notice that formula (7) for the bootstrap estimate of the bias differs from that given by Caswell [1, Eq.(12.23), p. 318]. The formula given by Caswell is valid only if the estimators of the vital rates are unbiased, and is thus inexact in the present case where  $\hat{d}_i$  and  $\hat{p}_i$  are biased.

#### 3.2. Asymptotic covariance matrix of vital rates

Let  $\mathcal{M}$  be the functional that maps any distribution  $F$  on its expectation:

$$\mathcal{M}(F) = \int x dF(x)$$

The relationship between the empirical distribution of observations  $\hat{F}_n$  and the vital rates estimator  $\hat{\theta}$  is given by the following Proposition.

**Proposition 3.** *There exists a functional  $\mathcal{G}$  such that  $\hat{\theta} = \mathcal{G}(\hat{F}_n)$ , where  $\hat{\theta}$  is the estimator of vital rates that was previously defined and  $\hat{F}_n$  is the empirical distribution of observations. Given a distribution  $F$  that can be written as (3), the functional  $\mathcal{G}$  is defined by:*

- The first element of  $\mathcal{G}(F)$ , that corresponds to the mean fecundity rate, is:

$$[\mathcal{G}(F)]_1 = \frac{\pi_0}{1 - \pi_0}$$

- The  $(i + 1)$ th element of  $\mathcal{G}(F)$  for  $i = 1, \dots, m$ , that corresponds to the mortality rate in class  $i$ , is:

$$[\mathcal{G}(F)]_{i+1} = \frac{\pi_1}{1 - \pi_0} \quad (i = 1, \dots, m)$$

- The  $(m + 1 + i)$ th element of  $\mathcal{G}(F)$  for  $i = 1, \dots, m - 1$ , that corresponds to the conditional upgrowth rate from class  $i$  to  $i + 1$  is:

$$[\mathcal{G}(F)]_{m+1+i} = 1 - G_{2i} \left[ u_{i+1} - \mathcal{M}(F_{22i}), h^{-1} \left( \frac{\mathcal{M}(F_{21i}) - u_{i+1}}{u_{i+1} - u_i} \right) \right] \quad (i = 1, \dots, m - 1)$$

where

$$F_{22i}(x) = \frac{F_2(x, \infty)}{F_2(u_{i+1}, \infty) - F_2(u_i, \infty)} \mathbf{1}(u_i \leq x < u_{i+1})$$

is the conditional size distribution at  $t_0$  knowing that the individuals are alive at both measurement times and belong to class  $i$  at  $t_0$ ,  $\mathbf{1}(p)$  is the indicator function of proposition  $p$ , and

$$F_{22i}(x) = \frac{\int_{u_i}^{u_{i+1}} \left\{ \int_{-\infty}^{a+x} \frac{\partial^2 F_2}{\partial x \partial y}(a, b) db \right\} da}{F_2(u_{i+1}, \infty) - F_2(u_i, \infty)}$$

is the conditional distribution of size increments between  $t_0$  and  $t_0 + 1$  knowing that the individuals are alive at both measurement times and belong to class  $i$  at  $t_0$ .

The proof is similar to that of Proposition 1 and is not given here to save space. Carrying forward this expression of  $\mathcal{G}$  in (9) gives the influence curve  $\text{IC}_{\mathcal{G}, F}(x)$  for any observation  $x$ . Carrying forward this expression of the influence curve in (8) gives the asymptotic covariance matrix of  $\hat{\theta}$ . Computations are long and are not shown here to save space. They give the following expression for the asymptotic covariance matrix  $\Sigma$  of the vital rates estimator  $\hat{\theta}$  (where  $\Sigma_{ij}$  is the element on the  $i$ th row and  $j$ th column of  $\Sigma$ ):

- The asymptotic variance of the estimator of the mean fecundity rate is:

$$\Sigma_{11} = \frac{\pi_0}{(1 - \pi_0)^3}$$

- The asymptotic covariance matrix of the estimators of mortality rates is:

$$\Sigma_{(i+1)(j+1)} = \frac{\pi_1 \pi_2}{(1 - \pi_0)^3} \quad (i, j = 1, \dots, m)$$

- The asymptotic variance of the estimator of the conditional upgrowth rate from class  $i$  to  $i+1$  is:

$$\begin{aligned} \frac{\Sigma}{(m+1+i)(m+1+i)} &= \frac{1}{\pi_2[F_2(u_{i+1}, \infty) - F_2(u_i, \infty)]} \\ &\times \left\{ \left[ \frac{\partial G_{2i}}{\partial X}(u_{i+1} - \mu_{2i}, \alpha_i) \right]^2 \sigma_{2i}^2 \right. \\ &+ \left[ \beta_i \frac{\partial G_{2i}}{\partial \kappa}(u_{i+1} - \mu_{2i}, \alpha_i) \right]^2 \sigma_{1i}^2 \\ &\left. - 2\beta_i \frac{\partial G_{2i}}{\partial X}(u_{i+1} - \mu_{2i}, \alpha_i) \frac{\partial G_{2i}}{\partial \kappa}(u_{i+1} - \mu_{2i}, \alpha_i) c_i \right\} \end{aligned}$$

where

$$\alpha_i = h^{-1} \left( \frac{\mu_{1i} - u_{i+1}}{u_{i+1} - u_i} \right)$$

$$\beta_i = \frac{1}{u_{i+1} - u_i} (h^{-1})' \left( \frac{\mu_{1i} - u_{i+1}}{u_{i+1} - u_i} \right)$$

$\mu_{1i} = \mathcal{M}(F_{21i})$  is the conditional expectation of size in class  $i$ ,  $\mu_{2i} = \mathcal{M}(F_{22i})$  is the conditional expectation of size increment in class  $i$ ,  $\sigma_{2i}^2$  is the conditional variance of size increments in class  $i$ ,  $\sigma_{1i}^2$  is the conditional variance of sizes in class  $i$ , and  $c_i$  is the conditional covariance between size and size increment in class  $i$ .

- All other elements of  $\Sigma$  are null.

To complement the latter expression, we give the expressions for the derivatives of  $h^{-1}$  and  $G_{2i}$ :  $(h^{-1})'(x) = [h'(h^{-1}(x))]^{-1}$ ,

$$h'(x) = \begin{cases} -\frac{1}{x^2} + \frac{\exp(-x)}{(1-\exp(-x))^2} & (x \neq 0) \\ -1/12 & (x = 0) \end{cases}$$

$$\frac{\partial G_{2i}}{\partial X}(x, \kappa) = \begin{cases} \frac{\kappa \exp[-\kappa(x-u_i)/(u_{i+1}-u_i)]}{(u_{i+1}-u_i)(1-\exp(-\kappa))} & (\kappa \neq 0) \\ 1/(u_{i+1}-u_i) & (\kappa = 0) \end{cases}$$

and:

$$\frac{\partial G_{2i}}{\partial \kappa}(x, \kappa) = \begin{cases} \frac{(x-u_i)(u_{i+1}-x)}{2(u_{i+1}-u_i)^2} & (\kappa = 0) \\ \frac{1}{(1-\exp(-\kappa))^2} \left\{ \left( \frac{x-u_i}{u_{i+1}-u_i} \right) \exp \left[ -\kappa \left( \frac{x-u_i}{u_{i+1}-u_i} \right) \right] \right. \\ \left. + \exp(-\kappa) \left[ \left( \frac{u_{i+1}-x}{u_{i+1}-u_i} \right) \exp \left[ -\kappa \left( \frac{x-u_i}{u_{i+1}-u_i} \right) \right] - 1 \right] \right\} & (\kappa \neq 0) \end{cases}$$

### 3.3. Asymptotic variance of the stock recovery rate

Given the expression of the asymptotic covariance matrix of vital rates given in the previous Section and Eq. (10), the expression of the Jacobian matrix  $D_\theta \lambda$  will give the asymptotic variance of the stock recovery rate. Function  $\lambda$  that maps the vital rates  $\theta$  onto the asymptotic stock recovery rate can be broken down into:

$$\lambda(\theta) = (\mathcal{A} \circ \mathcal{B} \circ \mathcal{C} \circ \mathcal{D})(\theta)$$

where  $\mathcal{D}$  is the function that maps the vector of vital rates  $\theta$  onto the Usher matrix,  $\mathcal{C}$  is the function that maps any  $m \times m$  matrix onto its  $T$ th power,  $\mathcal{B}$  is the function that maps any  $m \times m$  matrix onto its right product with  $\mathbf{L}$ , and  $\mathcal{A}$  is the function that maps any  $m \times m$  matrix onto its dominant eigenvalue. Using chain rule for Jacobian [36, p.91],

$$D_\theta \lambda = (D_{\mathcal{B} \circ \mathcal{C} \circ \mathcal{D}}(\theta) \mathcal{A})(D_{\mathcal{C} \circ \mathcal{D}}(\theta) \mathcal{B})(D_{\mathcal{D}}(\theta) \mathcal{C})(D_\theta \mathcal{D})$$

Let  $Z$  be the vector of length  $m$  whose elements are all equal to zero. Let  $U_i$  be the vector of length  $m$  whose elements are all equal to zero except the  $i$ th that is equal to one. Let

$$D_i = -(1-p_i)U_i - p_i U_{i+1} \quad (i = 1, \dots, m)$$

$$E_i = -(1-d_i)U_i + (1-d_i)U_{i+1} \quad (i = 1, \dots, m-1)$$

where, by convention,  $p_m = 0$  and  $U_{m+1} \equiv Z$ . Then the Jacobian of  $\mathcal{D}$  is the  $m^2 \times 2m$  matrix:

$$D_\theta \mathcal{D} = \begin{bmatrix} U_1 & D_1 & Z & & Z & Z & E_1 & Z & & Z \\ \vdots & Z & D_2 & & \vdots & \vdots & Z & E_2 & & \vdots \\ \vdots & \vdots & Z & \ddots & Z & \vdots & \vdots & Z & \ddots & Z \\ \vdots & \vdots & \vdots & & D_{m-1} & Z & \vdots & \vdots & & E_{m-1} \\ U_1 & Z & Z & & Z & D_m & Z & Z & & Z \end{bmatrix}$$

The Jacobian of  $\mathcal{C}$  is the  $m^2 \times m^2$  matrix:

$$D_U \mathcal{C} = \sum_{i=1}^T (\mathbf{U}^\top)^{T-i} \otimes \mathbf{U}^{i-1}$$

where  $\otimes$  denotes the Kronecker product [36]. Let  $c$  be the last size class that is not harvested, that is to say:  $\mathbf{L}$  is the diagonal matrix whose  $i$ th element on the diagonal equals one if  $i \leq c$  and zero if  $i > c$ . Then the Jacobian of  $\mathcal{B}$  is the  $m^2 \times m^2$  diagonal matrix whose  $i$ th element on the diagonal equals one if  $i \leq mc$  and zero if  $i > mc$ . Finally, the Jacobian of  $\mathcal{A}$  is given in [1, Eq. (9.14), p. 210].

### 3.4. Comparison between the three methods

Table 1 shows the estimate of the asymptotic (as time tends to infinity) stock recovery rate using the bootstrap, the analytic or the hybrid method, for the three species at Paracou. For all three species, one is included in the 95% confidence interval. However the confidence interval is centred on a value higher than one for angelique and grey gonfola, whereas it is centred on one for pink gonfola. Moreover this estimate is an overestimation for angelique and pink gonfola whereas it is an underestimation for grey gonfola. Estimates for angelique and pink gonfola rely on a similar number of observations ( $n = 467$  for angelique and  $n = 484$  for pink gonfola), whereas sample size ( $n = 53$ ) is lower for grey gonfola. The standard error is correspondingly higher for grey gonfola than for angelique or pink gonfola.

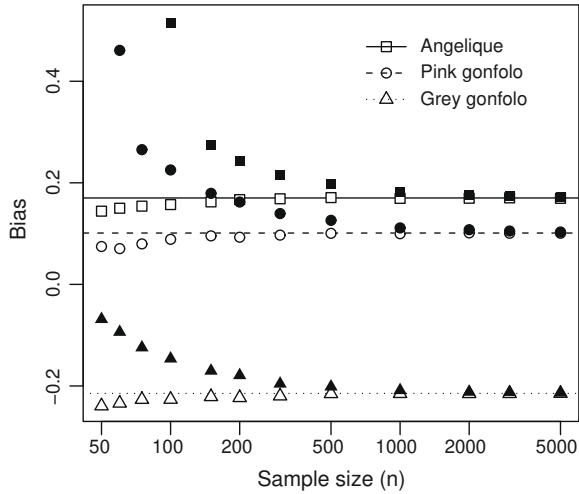
Reference values in Table 1 are those given by the bootstrap method since this method does not make any approximation. On the contrary the analytic and the hybrid methods approximate the true values by their asymptotic values as the sample size tends to infinity. The analytic method underestimates the asymptotic stock recovery rate and its standard error. It either underestimates (for angelique and pink gonfola) or overestimates (for grey gonfola) the bias. The hybrid method slightly outperforms the analytic method (angelique, grey gonfola) or does as well (pink gonfola) in estimating the standard error. However the hybrid method does as well (angelique) or worse (pink and grey gonfola) than the analytic method in estimating the bias.

Table 1 is complemented by Figs. 2 and 3 that show the speed of convergence of the bootstrap and hybrid estimates of the bias and

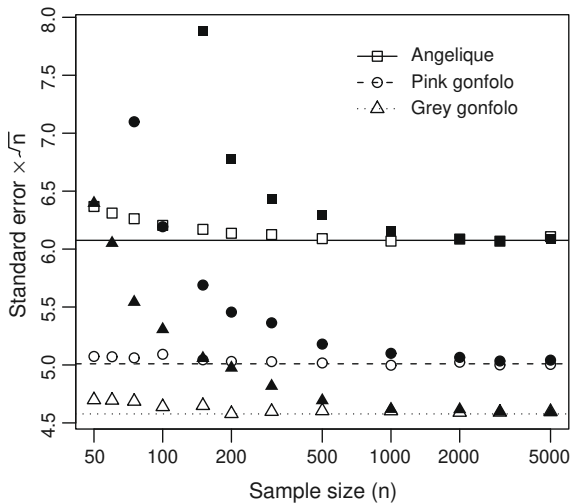
**Table 1**

Estimate of the asymptotic stock recovery rate, and of the bias, standard error and 95% confidence interval of its estimator using the bootstrap, analytic or hybrid method, for the three species angelique (*Dicorynia guianensis*), pink gonfola (*Qualea rosea*) and grey gonfola (*Ruizterania albiflora*) at Paracou.

| Species      | Method    | Estimate | Bias   | Standard error | Lower bound | Upper bound |
|--------------|-----------|----------|--------|----------------|-------------|-------------|
| Angelique    | Bootstrap | 1.450    | 0.199  | 0.291          | 0.975       | 2.090       |
| Angelique    | Analytic  | 1.422    | 0.170  | 0.281          | 0.882       | 1.973       |
| Angelique    | Hybrid    | 1.422    | 0.170  | 0.285          | 0.875       | 1.980       |
| Pink gonfola | Bootstrap | 1.003    | 0.125  | 0.236          | 0.618       | 1.534       |
| Pink gonfola | Analytic  | 0.979    | 0.101  | 0.228          | 0.542       | 1.425       |
| Pink gonfola | Hybrid    | 0.979    | 0.098  | 0.228          | 0.539       | 1.423       |
| Grey gonfola | Bootstrap | 1.390    | -0.073 | 0.864          | 0.466       | 3.187       |
| Grey gonfola | Analytic  | 1.249    | -0.215 | 0.629          | 0.042       | 2.481       |
| Grey gonfola | Hybrid    | 1.224    | -0.239 | 0.644          | -0.011      | 2.486       |



**Fig. 2.** Convergence speed of the bootstrap and hybrid estimates of bias toward its asymptotic (as sample size tends to infinity) value, for the three species at Paracou: solid symbols represent the bootstrap estimate; open symbols represent the hybrid estimate; lines represent the asymptotic value; squares or solid line are for angelique (*Dicorynia guianensis*); circles or dashed line are for pink gonfolo (*Qualea rosea*); triangles or dotted line are for grey gonfolo (*Ruizterania albiflora*).



**Fig. 3.** Convergence speed of the bootstrap and hybrid estimates of standard error toward its asymptotic (as sample size tends to infinity) value, for the three species at Paracou: solid symbols represent the bootstrap estimate; open symbols represent the hybrid estimate; lines represent the asymptotic value; squares or solid line are for angelique (*Dicorynia guianensis*); circles or dashed line are for pink gonfolo (*Qualea rosea*); triangles or dotted line are for grey gonfolo (*Ruizterania albiflora*).

standard error of the stock recovery rate estimator as sample size tends to infinity. For all three species, convergence is reached for sample sizes greater than 5000 observations. Again, the reference values are those given by the bootstrap method. Both the analytic and the hybrid method underestimate the bias and the standard error of the stock recovery rate estimator. The hybrid method does slightly better than the analytic method in estimating the standard error, whereas the analytic method does better than the hybrid method in estimating the bias.

#### 4. Discussion

With the bootstrap method we can estimate the bias, standard error and confidence interval of the stock recovery rate estimator

without any approximation. This method requires potentially long computing time. The analytic method, which relies on the asymptotic values of these quantities as the sample size tends to infinity, is comparatively very fast but is approximate if the sample size is small. The convergence between the bootstrap and the analytic method was obtained for sample size greater than 5000 for the three species at Paracou. Although the hybrid method makes a compromise between the bootstrap and the analytic method by using the former for the second arrow of Fig. 1 and the latter for the last arrow of Fig. 1, its estimates are very close to those of the analytic method, with slightly better results for standard error and slightly worse results for bias. This means that most of the discrepancy between the bootstrap and the analytic method for small samples is due to the last arrow of Fig. 1. As a first conclusion, the hybrid method has little interest as compared to the analytic method. As a second conclusion, the analytic method is unreliable unless sample size is large enough. Bootstrap then should be preferred. The threshold for the analytic method to be reliable (about 5000 observations on the basis of the three species at Paracou) is beyond most sample sizes commonly found in tropical rainforests. The value of this threshold depends on the type of estimator used for vital rates and on the prediction of interest. For instance, Zetlaoui et al. [37] and Chagneau [38] found a threshold of about 10000 observations to predict the stock recovery rate at the end of the first felling cycle or the population growth rate, using the proportion estimator. Due to the high species richness of tropical rainforests and the corresponding low densities of their species, capturing 5000 trees (*a fortiori* 10000) most often require several hundreds (if not thousands) of hectares.

Applications of these methods to assess sampling variability are numerous [39–45]. In the present case they allow us to test whether the asymptotic stock recovery rate differs from one. Although the three species at Paracou have contrasted expected values for  $\lambda$ , none of them is significantly different from one at level 5%. More observations would be required to estimate  $\lambda$  with greater accuracy, and to be able to decide whether it is significantly different from one or not. The accuracy (*sensu* [46]) of the estimate of  $\lambda$  presently is 38% for angelique, 46% for pink gonfolo, and 98% for grey gonfolo at confidence level 95%. Taking  $\hat{F}_n$  for the theoretical distribution  $F$  of observations (see first arrow of Fig. 1), it would require 6870 observations for angelique (to compare to 467 presently available), 9860 observations for pink gonfolo (to compare to 484), and 5060 observations for grey gonfolo (to compare to 53) to get an accuracy of 10% at confidence level 95%, which is generally considered as an acceptable accuracy. For these sample sizes, and assuming again that the theoretical distribution of observations is  $\hat{F}_n$ ,  $\lambda$  is significantly greater than one at level 5% for angelique ( $p$ -value  $< 10^{-4}$ ) and grey gonfolo ( $p$ -value =  $10^{-4}$ ), and is not significantly different from one at level 5% for pink gonfolo ( $p$ -value = 0.67).

Another application would consist in comparing sampling variability with other sources of variability, in particular environmental variability. For instance we used data from 1992 to 1994, but it would be interesting to assess how the variability generated by the use of another time step compares to sampling variability. Depending on this comparison, it will be possible or not to detect significant differences in vital rates between years. This is left for future work.

The approximate expression for bias corresponds to a Taylor expansion to order zero with respect to  $(1/\sqrt{n})$  (where  $n$  is sample size), whereas the approximate expression (10) for standard error corresponds to a Taylor expansion up to order one with respect to  $(1/\sqrt{n})$ . The threshold at which the bootstrap method converges toward the analytic expression could be lowered by making Taylor expansion to higher order. For instance bias could be approximated by:



$$\left\{ \lambda[\mathcal{G}(F)] + \frac{1}{2}(\text{vec}\Sigma)^\top \text{vec}(H_{\mathcal{G}(F)}\lambda) \right\} - \lambda[\mathcal{H}(F)]$$

where  $H_{\theta}\lambda$  is the Hessian matrix of  $\lambda$  at  $\theta$ , that is a  $2m \times 2m$  matrix whose term on the  $i$ th row and  $j$ th column is the second derivative of  $\lambda$  with respect to the  $i$ th and  $j$ th elements of  $\theta$ , and  $\text{vec}$  is the vec operator, that is the operator that vectorizes a matrix by stacking its columns. In this equation, the term between braces corresponds to a Taylor expansion up to order  $(1/\sqrt{n})^2$ . Similarly, the approximate expression for standard error could be expanded up to higher order with respect to  $(1/\sqrt{n})$ , which would entail computing the successive derivatives of  $\lambda$  with respect to  $\theta$ . These computations are not complicated in the principle, but are tedious in practice. Motivation for computation speed is then needed to prefer the analytic method to bootstrap.

The results of the present study are consistent with those of previous studies [15–17] in the sense that the analytic method is reliable only if errors on vital rates remain small. However these previous studies focused on the last arrow of Fig. 1 and ignored sampling variability. The present study relates sample size to the acceptable errors on vital rates. As sample size increases, the errors on vital rates vanished and the conclusions of the previous studies are found.

The present study also extends previous results that were obtained for other estimators of vital rates and other predicted quantities. Zetlaoui et al. [37,47] and Chagneau [38] used for the statistical model for observations (i.e. the first arrow of Fig. 1) a multinomial distribution with the vital rates as parameters. The corresponding maximum likelihood estimator of vital rates was the proportion estimator of Michie and Buongiorno [27], which also corresponds to the plug-in estimator of the present study. The asymptotic distribution for the maximum likelihood estimator can be computed using the  $\delta$ -method. The expression for the asymptotic variance is basically the same as the one given by Eq. (10), but  $\Sigma$  then designates the Fisher information matrix. Computations are noticeably more complicated with any estimator as the one used in the present study than with the maximum likelihood estimator. Zetlaoui et al. [37,47] focused on the population growth rate (i.e. the dominant eigenvalue of the Usher matrix) and the corresponding stable diameter distribution, while Chagneau [38] focused on the stock recovery rate at the end of the first felling cycle. The present study shows that these results can easily be extended to other quantities predicted by the matrix model, as long as these quantities are differentiable functions of the vital rates. This simply entails replacing  $D\lambda$  in Eq. (10) by the appropriate Jacobian matrix. Notice that the expression of the Jacobian  $D\lambda$  can often be found as part of a sensitivity analysis [1, Chapter 9]. Nevertheless, the bootstrap method is comparatively easier to extend to other vital rate estimators and other predictions of the matrix model than the analytic method. It can also deal with an explicit distribution for the observations (first arrow of Fig. 1), by performing parametric bootstrap [32,48].

### Appendix A. Plug-in estimator of vital rates

We here give the proof of Propositions 1 and 2. By definition of the mean fecundity rate,

$$f = \frac{E(\#\{S_1 = 0\})}{E(\#\{S_1 \neq 0\})} = \frac{n\Pr(S_1 = 0)}{n - n\Pr(S_1 = 0)} = \frac{\pi_0}{1 - \pi_0}$$

By definition of the mortality rate in class  $i$ ,

$$d_i = \Pr(S_2 = 0 | u_i \leq S_1 < u_{i+1})$$

Using Bayes formula, this becomes:

$$d_i = \frac{\Pr(S_2 = 0)\Pr(u_i \leq S_1 < u_{i+1} | S_2 = 0)}{\Pr(u_i \leq S_1 < u_{i+1})} = \pi_1 \frac{F_1(u_{i+1}) - F_1(u_i)}{G_1(u_{i+1}) - G_1(u_i)}$$

By definition of the conditional upgrowth rate from class  $i$  to  $i + 1$ ,

$$p_i = \Pr(u_{i+1} \leq S_2 < u_{i+2} | u_i \leq S_1 < u_{i+1}, S_2 \neq 0)$$

This can be written as:

$$p_i = \frac{\Pr(u_i \leq S_1 < u_{i+1}, u_{i+1} \leq S_2 < u_{i+2})}{\Pr(u_i \leq S_1 < u_{i+1}, S_2 \neq 0)}$$

The numerator can be written as:

$$\begin{aligned} & \Pr(u_i \leq S_1 < u_{i+1}, u_{i+1} \leq S_2 < u_{i+2} | S_1 \neq 0, S_2 \neq 0) \times \Pr(S_1 \neq 0, S_2 \neq 0) \\ & = \pi_2 \{F_2(u_{i+1}, u_{i+2}) - F_2(u_i, u_{i+2}) - F_2(u_{i+1}, u_{i+1}) + F_2(u_i, u_{i+1})\} \end{aligned}$$

whereas the denominator can be written as:

$$\begin{aligned} & \Pr(u_i \leq S_1 < u_{i+1} | S_1 \neq 0, S_2 \neq 0) \times \Pr(S_1 \neq 0, S_2 \neq 0) \\ & = \pi_2 \{F_2(u_{i+1}, \infty) - F_2(u_i, \infty)\} \end{aligned}$$

which gives the expression (13) of  $p_i$ .

By replacing  $F$  by the empirical distribution  $\hat{F}_n$  of the observations, plug-in estimators are obtained. Thus, the plug-in estimators of  $\pi_0$ ,  $\pi_1$ ,  $F_1$ ,  $G_1$  and  $F_2$  are:

$$\hat{\pi}_{0\text{plug-in}} = \#\{S_1 = 0\}/n$$

$$\hat{\pi}_{1\text{plug-in}} = \#\{S_2 = 0\}/n$$

$$\hat{F}_{1\text{plug-in}}(x) = \#\{S_1 < x, S_2 = 0\}/\#\{S_2 = 0\}$$

$$\hat{G}_{1\text{plug-in}}(x) = \#\{S_1 < x\}/n$$

$$\hat{F}_{2\text{plug-in}}(x, y) = \#\{0 < S_1 < x, 0 < S_2 < y\}/\#\{S_1 \neq 0, S_2 \neq 0\}$$

Carrying forward these expressions in (11)–(13), the plug-in estimator of  $\theta$  is obtained.

### Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.mbs.2009.02.002.

### References

- [1] H. Caswell, Matrix Population Models: Construction, Analysis and Interpretation, second ed., Sinauer Associates, Inc. Publishers, Sunderland, Massachusetts, 2001.
- [2] W.G. Doubleday, Harvesting in matrix population models, *Biometrics* 31 (1) (1975) 189.
- [3] M.S. Boyce, Population viability analysis, *Annu. Rev. Ecol. Syst.* 23 (1992) 481.
- [4] E.S. Menges, Population viability analyses in plants: challenges and opportunities, *Trends Ecol. Evol.* 15 (2) (2000) 51.
- [5] J. Fieberg, S.P. Ellner, Stochastic matrix models for conservation and management: a comparative review of methods, *Ecol. Lett.* 4 (3) (2001) 244.
- [6] C.E. Hauser, E.G. Cooch, J.D. Lebreton, Control of structured populations by harvest, *Ecol. Model.* 196 (3–4) (2006) 462.
- [7] J. van Groenendaal, H. de Kroon, H. Caswell, Projection matrices in population biology, *Trends Ecol. Evol.* 3 (10) (1988) 264.
- [8] P.H. Leslie, In the use of matrices in certain population mathematics, *Biometrika* 33 (1945) 183.
- [9] L.P. Lefkovich, The study of population growth in organisms grouped by stages, *Biometrics* 21 (1965) 1.
- [10] M.B. Usher, A matrix approach to the management of renewable resources, with special reference to the selection forests, *J. Appl. Ecol.* 3 (1966) 355.
- [11] M.B. Usher, A matrix model for forest management, *Biometrics* 25 (2) (1969) 309.
- [12] E.R. Alvarez-Buylla, R. García-Barrios, C. Lara-Moreno, M. Martínez-Ramos, Demographic and genetic models in conservation biology: applications and perspectives for tropical rain forest tree species, *Annu. Rev. Ecol. Syst.* 27 (1996) 387.
- [13] T.N. Kaye, D.A. Pyke, The effect of stochastic technique on estimates of population viability from transition matrix models, *Ecology* 84 (6) (2003) 1464.

- [14] E.R. Alvarez-Buylla, M. Slatkin, Finding confidence limits on population growth rates, *Trends Ecol. Evol.* 6 (7) (1991) 221.
- [15] F. Houllier, J.D. Lebreton, D. Pontier, Sampling properties of the asymptotic behavior of age- or stage-grouped population models, *Math. Biosci.* 95 (2) (1989) 161.
- [16] E.R. Alvarez-Buylla, M. Slatkin, Finding confidence limits on population growth rates: Monte Carlo test of a simple analytic method, *Oikos* 68 (1993) 273.
- [17] E.R. Alvarez-Buylla, M. Slatkin, Finding confidence limits on population growth rates: three real examples revised, *Ecology* 75 (1) (1994) 255.
- [18] S. Gourlet-Fleury, J.M. Guehl, O. Laroussinie (Eds.), *Ecology and Management of a Neotropical Rainforest. Lessons Drawn from Paracou a Long-Term Experimental Research Site in French Guiana*, Elsevier, Paris, 2004.
- [19] J. Barret (Ed.), *Atlas illustré de la Guyane*, IRD, Paris, France, 2001.
- [20] S. Gourlet-Fleury, G. Cornu, S. Jéssel, H. Dessard, J.G. Jourget, L. Blanc, N. Picard, Using models for predicting recovery and assessing tree species vulnerability in logged tropical forests: a case study from French Guiana, *For. Ecol. Manag.* 209 (1-2) (2005) 69.
- [21] GERDAT-CTFT, *Fiches techniques sur des bois guyanais: gonfalo*, *Bois For. Trop.* 202 (1983) 69-72.
- [22] J. Thiel, *Fiches de reconnaissance pratique des arbres sur pied de la forêt guyanaise: gonfalo gris*, *Bois For. Trop.* 201 (1983) 46.
- [23] H.A. Sturges, The choice of a class interval, *J. Am. Stat. Assoc.* 21 (1926) 65.
- [24] N. Picard, A. Bar-Hen, S. Gourlet-Fleury, Estimator of upgrowth transition parameters for size-classified matrix from small samples, *Ecol. Model.* 204 (1-2) (2007) 59.
- [25] N. Picard, F. Mortier, P. Chagneau, Influence of estimators of the vital rates in the stock recovery rate when using matrix models for tropical rainforests, *Ecol. Model.* 214 (2-4) (2008) 349.
- [26] N. Picard, Y. Yalibanda, S. Namkossereana, F. Baya, Estimating the stock recovery rate using matrix models, *For. Ecol. Manag.* 255 (10) (2008) 3597.
- [27] B.R. Michie, J. Buongiorno, Estimation of a matrix model of forest growth from re-measured permanent plots, *For. Ecol. Manag.* 8 (1984) 127.
- [28] D.J. Olive, *Applied Robust Statistics*, Southern Illinois University, Department of Mathematics, Carbondale, Illinois, 2007, book published on-line. Available from: <<http://www.math.siu.edu/olive/run.pdf>>.
- [29] R. Ralston, J. Buongiorno, B. Schulte, J. Fried, Non-linear matrix modelling of forest growth with permanent plot data: The case of uneven aged Douglas-fir stands, *Int. Trans. Oper. Res.* 10 (5) (2003) 461.
- [30] J. Liang, J. Buongiorno, R.A. Monserud, Growth and yield of all-aged douglas-fir – western hemlock forest stands: a matrix model with stand diversity effects, *Can. J. For. Res.* 35 (10) (2005) 2368.
- [31] J. Liang, J. Buongiorno, R.A. Monserud, Estimation and application of a growth and yield model for uneven-aged mixed conifer stands in California, *Int. For. Rev.* 7 (2) (2005) 101.
- [32] B. Efron, R.J. Tibshirani, *An Introduction to the Bootstrap*, No. 57 in *Monographs on Statistics and Applied Probability*, Chapman & Hall, New York, 1993.
- [33] R Development Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2005.
- [34] F.R. Hampel, The influence curve and its role in robust estimation, *J. Am. Stat. Assoc.* 69 (1974) 383.
- [35] A. Cuevas, J. Romo, On the estimation of the influence curve, *Can. J. Stat.* 23 (1) (1995) 1.
- [36] J.R. Magnus, H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley Series in Probability and Statistics, second ed., John Wiley & Sons, Chichester, 1999.
- [37] M. Zetlaoui, N. Picard, A. Bar-Hen, Asymptotic distribution of stage-grouped population models, *Math. Biosci.* 200 (1) (2006) 76.
- [38] P. Chagneau, *Optimisation sous contraintes spatiales; application à la mise en place de parcelles permanentes de suivi des forêts tropicales humides*, MSc. Thesis, University of Montpellier 2, Montpellier, France, 2006.
- [39] L.L. Eberhardt, A.K. Majorowicz, J.A. Wilcox, Apparent rates of increase for two feral horse herds, *J. Wildl. Manag.* 46 (2) (1982) 367.
- [40] A. Taberner, P. Castañera, E. Silvestre, J. Dopazo, Estimation of the intrinsic rate of natural increase and its error by both algebraic and resampling approaches, *Bioinformatics* 9 (5) (1993) 535.
- [41] H. Caswell, S. Brault, A.J. Read, T.D. Smith, Harbor porpoise and fisheries: an uncertainty analysis of incidental mortality, *Ecol. Appl.* 8 (4) (1998) 1226.
- [42] R.D. Mace, J.S. Waller, Demography and population trend of grizzly bears in the swan mountains, Montana, *Conserv. Biol.* 12 (5) (1998) 1005.
- [43] R.M. Sibly, F.T. Hansen, V.E. Forbes, Confidence intervals for population growth rate of organisms with two-stage life histories, *Oikos* 88 (2) (2000) 335.
- [44] R.A. Hinrichsen, The accuracy of alternative stochastic growth rate estimates for salmon populations, *Can. J. Fish. Aquat. Sci.* 59 (6) (2002) 1014.
- [45] J.R. Skalski, J.J. Millsbaugh, K.E. Ryding, Effects of asymptotic and maximum age estimates on calculated rates of population change, *Ecol. Model.* 212 (3-4) (2008) 528.
- [46] B. Zeide, Plot size optimization, *For. Sci.* 26 (2) (1980) 251.
- [47] M. Zetlaoui, N. Picard, A. Bar-Hen, Asymptotic distribution of density-dependent stage-grouped population dynamics models, *Acta Biotheor.* 56 (1-2) (2008) 137.
- [48] E.R. Alvarez-Buylla, Density dependence and patch dynamics in tropical rain forests: matrix models and applications to a tree species, *Am. Nat.* 143 (1994) 155.