Mathematical Biosciences 219 (2009) 23–31

Contents lists available at ScienceDirect

Mathematical Biosciences

journal homepage: www.elsevier.com/locate/mbs

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

Nicolas Picard [a,*], Pierrette Chagneau [a], Frédéric Mortier [a], Avner Bar-Hen [b]

[a] CIRAD, Campus international de Baillarguet, 34398 Montpellier Cedex 5, France
[b] Université Paris Descartes, MAP 5, 45 rue des Saint Pères, 75270 Paris Cedex 06, France

ARTICLE INFO

Article history:
Received 11 July 2008
Received in revised form 11 February 2009
Accepted 13 February 2009
Available online 26 February 2009

Keywords:
Asymptotic properties
Matrix model
Population dynamics
Tropical forests
Usher model

ABSTRACT

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:

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

As t tends to infinity, the proportions of individuals in the different categories will grow at an exponential rate, λ, which is the dominant eigenvalue of $\mathbf{A}$ [1]. The population growth rate λ is a model's prediction that is often use 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 Lefkovitch models [8,9]. The Lefkovitch 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).

0025-5564/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.mbs.2009.02.002

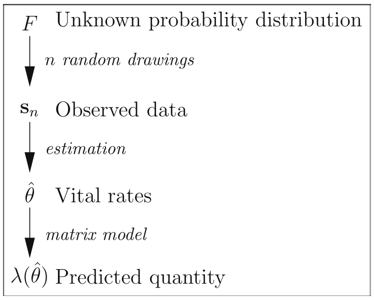

Fig. 1. Schematic diagram of the uncertainty chain showing how any prediction of the matrix model is a random variable whose distribution depends on the distribution of observations.

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, Caesalpiniaceæ), pink gonfolo (*Qualea rosea* Aublet, Vochysiaceæ) and grey gonfolo (*Ruizterania albiflora* (Warming) Marcano-Berti, Vochysiaceæ). 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 $\geqslant 60$ 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 $\geqslant 60$ 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 its 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 ith element of the diagonal of $\mathbf{G}$ is $1 - p_i$ and the ith 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 ith 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 ith 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