

Population dynamics of species-rich ecosystems: the mixture of matrix population models approach

Frédéric Mortier^{1*}, Vivien Rossi², Gilles Guillot³,
Sylvie Gourlet-Fleury¹ and Nicolas Picard¹

Summary

Matrix population models are widely used to predict population dynamics but, when applied to species rich ecosystems with many rare species, the small population sample sizes hinder a good fit of species-specific models. This issue can be overcome by assigning species to groups to increase the size of the calibration data sets. However, the species classification is often disconnected from the models and from the parameter estimation, thus bringing species groups that may not be optimal with respect to the predicted community dynamics. We proposed here a method that jointly classified species into groups and fit the matrix models in an integrated way. The model was a special case of mixture with unknown number of components and was cast in a Bayesian framework. An MCMC algorithm was developed to infer the unknown parameters: the number of groups, the group of each species and the dynamics parameters. We applied the method to a data set from a tropical rain forest in French Guiana.

Key-words: Bayesian, clustering, mixture models, reversible jump Markov chain Monte Carlo.

1 Introduction

The conservation of animal and plant species and their biological control require models to understand and predict population dynamics (Demyanov et al. 2006). Among population dynamics models, projection matrix models have been widely used to investigate the dynamics of age-, stage- or size-structured populations. They provide a simple way of integrating vital rate information such as recruitment, birth, growth or ageing, and mortality.

In species-rich ecosystems high diversity implies that the sample size for most species is limited. The small sample size hinders a good fit of species-specific dynamics models. To address this problem, modellers usually cluster species into groups. To cluster the species while ensuring optimality for predicting community dynamics we need to rely on the mixture model framework.

*frederic.mortier@cirad.fr. CIRAD, UPR Bsef, 34398 Montpellier, France

This study aims at extending mixture modelling to matrix population models (Mortier et al. 2012). The mixture of matrix population models consists in a probabilistic model-based clustering method expressed in terms of matrix population mixture models with an unknown number of components (Marin et al. 2005). The number of groups and the parameters of the matrix population models associated with each group are the unknown quantities. We propose to use a Bayesian framework to infer these unknown quantities.

The mixture of matrix models is defined in the next section. An inference method is then outlined. The mixture matrix model is finally applied to a data set from the Paracou tropical rain forest in French Guiana. The tree species groups thus obtained had consistent ecological behaviours with contrasted functional traits, and compared favourably to other groups obtained by a standard classification technique.

2 Models and inference

2.1 Usher model for a species

The Usher (1969) model is a matrix population model for size-structured populations. It is based on the description of the change of the population by a vector, \mathbf{N}_t , of the numbers of individuals in L ordered state classes at discrete time t . The components of vector \mathbf{N}_t , $N_{l,t}$ for $l = 1, \dots, L$, are the numbers of individuals in each class l . The relationship between \mathbf{N}_t and \mathbf{N}_{t+1} is described by a $L \times L$ transition matrix U , called the Usher matrix:

$$\mathbf{N}_{t+1} = U\mathbf{N}_t \quad (1)$$

where U is equal to:

$$U = \begin{pmatrix} p_1 + f & f & \dots & f \\ q_1 & p_2 & & 0 \\ & \ddots & \ddots & \\ 0 & & q_{L-1} & p_L \end{pmatrix}.$$

p_l is for an individual the probability to stay in class l , q_l the probability to move up from class l to $l + 1$ and f the average fecundity. q_l and p_l take values in $[0, 1]$, whereas f takes values in \mathbb{R}^+ . The probability of dying for an individual in class l is given by $m_l = 1 - p_l - q_l$. Let $\mathbf{d} = (d_1, \dots, d_L)$ be the class-distribution of the population such as $\sum_{l=1}^L d_l = 1$. d_l denotes the probability that an individual of the population chosen at random belongs to the class l . As individuals were assumed to be independent (see Usher 1969 assumptions), the model can be described at the species level. Let $N_{t-1} = \sum_{l=1}^L N_{l,t-1}$ be the total number of individuals of a given species at time $t - 1$. Let $N_{l,l,t}$ denote the number of individuals of this species staying in class l between $t - 1$ and t , $N_{l,l+1,t}$ the number of individuals moving up from class l to $l + 1$ between $t - 1$ and t , and $N_{l,\dagger,t}$ the number of individuals dying in class l between $t - 1$ and t . Let R_t be the number of recruits between $t - 1$ and t , assumed to be a Poisson random variable with parameter fN_{t-1} . Finally, let $\mathbf{N} = (N_{1,l,t}, \dots, N_{L,\dagger,t}, \mathbf{N}_{t-1}, R_t)$

describe the vector of the observations at times $t - 1$ and t . Then, the likelihood is equal to

$$\begin{aligned} \mathcal{L}(\mathbf{N}|\theta) &= \prod_{l=1}^{L-1} \text{Mult}(N_{l,l,t}, N_{l,l+1,t}, N_{l,\dagger,t} | p_l, q_l, m_l, N_{l,t-1}) \\ &\quad \times \text{Mult}(N_{L,L,t}, N_{L,\dagger,t} | p_L, m_L, N_{L,t-1}) \\ &\quad \times \text{Mult}(N_{1,t-1}, \dots, N_{L,t-1} | d_1, \dots, d_L, N_{t-1}) \\ &\quad \times \mathcal{P}(R_t | f N_{t-1}) \end{aligned} \quad (2)$$

where Mult denotes the multinomial distribution, \mathcal{P} the Poisson distribution and, $\theta = (\mathbf{p}, \mathbf{q}, \mathbf{m}, f, \mathbf{d}) \in [0, 1]^{3L-1} \times \mathbb{R}^+ \times [0, 1]^L$ is the parameters vector where $\mathbf{p} = (p_1, \dots, p_L)$, $\mathbf{q} = (q_1, \dots, q_{L-1})$, $\mathbf{m} = (m_1, \dots, m_L)$.

2.2 Mixture of matrix population models with unknown number of components

In this paper, we focus on the unsupervised classification of S species into different groups (G_1, \dots, G_K) with an unknown number K of groups. Let $\mathbf{C} = (C_1, \dots, C_S)$ be the latent class vector which describes the group of species s . For example, the species 1 is classified in the third group: $C_1 = 3$. The dynamics of each group is modelled by a matrix model, so that the dynamics of the forest stand is a mixture of K matrix models. Let $\theta = (\theta_1, \dots, \theta_K)$ be the parameters associated with the K matrix models. Using the same notation as above but with an additional superscript s , $\mathbf{N}^s = (N_{1,l,t}^s, \dots, N_{L,\dagger,t}^s, \mathbf{N}_{t-1}^s, R_t^s)$, gives the vector of the observations for species $s = 1, \dots, S$ and $\mathbf{N} = (\mathbf{N}^1, \dots, \mathbf{N}^S)$ is the vector of the observations for all species. Then, in terms of the mixture model, posterior density distribution can be expressed as follows:

$$\pi_{\mathbf{C}, \theta, K}^{\mathbf{N}}(\mathbf{C}, \theta, K | \mathbf{N}) \propto \prod_{s=1}^S \mathcal{L}(\mathbf{N}^s | \theta_{C_s}) \pi_{\mathbf{C}|\theta, K}^0(\mathbf{C} | \theta, K) \pi_{\theta|K}^0(\theta | K) \pi_K^0(K) \quad (3)$$

where $\pi_{\mathbf{C}|\theta, K}^0$, $\pi_{\theta|K}^0$ and π_K^0 are the prior densities distributions associated with the class latent random variables, the parameters of each matrix models and the number of groups, respectively.

2.3 Prior distributions

For full Bayesian inference of the model, we set the followings priors on the unknown quantities \mathbf{C} , θ and K . Two prior distributions for the number K of groups were tested: (i) distributed as a uniform distribution: $\pi_K^0(K) \equiv \mathcal{U}(K_{\min}, \dots, K_{\max})$. The only subjective inputs for this prior are K_{\min} , (ii) K is distributed as a Poisson random variable with mean one, truncated to strictly positive values: $\pi_K^0(K) \equiv \mathcal{P}(1) \setminus \{0\}$

$$\pi_{\theta|K}^0(\theta | K) = \prod_{k=1}^K \left\{ \prod_{l=1}^{L-1} \pi_{p,q,m|l,k}^0(p_{lk}, q_{lk}, m_{lk}) \right\} \pi_{p,m|k}^0(p_{Lk}, m_{Lk}) \pi_{f|k}^0(f_k) \pi_{\mathbf{d}|k}^0(\mathbf{d}_k)$$

For all diameter classes l and all groups k , we chose the conjugate Dirichlet prior distribution: $\pi_{p,q,m|l,k}^0 \equiv \mathcal{D}(1, 1, 1)$ for $l < L$ and $\pi_{p,m|k}^0 \equiv \mathcal{D}(1, 1)$. As the number of recruits for each group k was a Poisson random variable with parameter $f_k N_{t-1}^k$, we chose, for all groups k , the conjugate gamma prior distribution \mathcal{G} for the fecundity parameter f_k . Moreover, as the number of births in a group is roughly equal to 1% of the number of individuals in this group: $\pi_{f|k}^0 \equiv \mathcal{G}(0.01, 1)$. This prior distribution reflects the expert's knowledge. Finally, we assumed, for all groups k , that the parameters of the class-population distribution $\mathbf{d}_k = (d_{1k}, \dots, d_{Lk})$ were distributed as a Dirichlet distribution: $\pi_{\mathbf{d}|k}^0 \equiv \mathcal{D}(1, \dots, 1)$.

The prior for the class vector \mathbf{C} assumed that, given the number of groups, each species could equally and independently of the other species be in any group: $\pi_{\mathbf{C}|\theta,K}^0(\mathbf{C}|\theta, K) = \prod_{s=1}^S \pi_{\mathbf{C}|K}^0(C_s|K)$ where $\pi_{\mathbf{C}|K}^0(C_s|K)$ is a uniform distribution on the number of groups: $\mathcal{U}(1, \dots, K)$.

The inference of parameters is made through the investigation of the posterior distribution $\pi_{\mathbf{C},\theta,K}^{\mathbf{N}}(\mathbf{C}, \theta, K|\mathbf{N})$ defined by equation 3. As the number of groups is unknown, the posterior distribution is not available in an analytic form. Then, we propose to run a Metropolis within Gibbs Monte Carlo Markov chain (MCMC). The algorithm consists of three moves: increasing the number of groups (birth case); decreasing the number of groups (death case); keeping the same number of groups but potentially changing one species assignment (no jump case). In the first two cases, the number of parameters is not constant, so a Reversible Jump MCMC approach is used, whereas in the third case, a Gibbs step can be used. All moves are equally distributed with probability 1/3. To save place, details and R package can be found in appendix of Mortier et al. (2012).

3 Application

Data were collected at the Paracou experimental site (5°18'N, 52°53'W), French Guiana. The site is located in a undisturbed *terra firme* forest under equatorial climate. Three 250 m × 250 m permanent sample plots (18.75 ha in total) have been established in 1984 and left as control of the undisturbed forest dynamics. All trees greater than 10 cm dbh (diameter at breast height) have been identified and georeferenced. Girth at breast height, standing deaths, treefalls and newly recruited trees greater than 10 cm dbh have been monitored either annually or every two years since 1984 (Gourlet-Fleury et al. 2004). The data set consisted of 93 species collected in 1993 and 1995 on the three control plots.

Based on 50 different chains, and 20000 iterations after a burn-in of 10000 iterations, five groups were obtained. The mixture of Usher matrix models classified species according to both their growth rate and their maximum size (Picard et al. 2012). When plotting species along these two axes, species groups were clearly separated (Figure 1). Because these two axes can be used to order species along a continuum of ecological strategies, this means that the mixture of Usher matrix models was also able to classify species in a way that is consistent with their autecology.

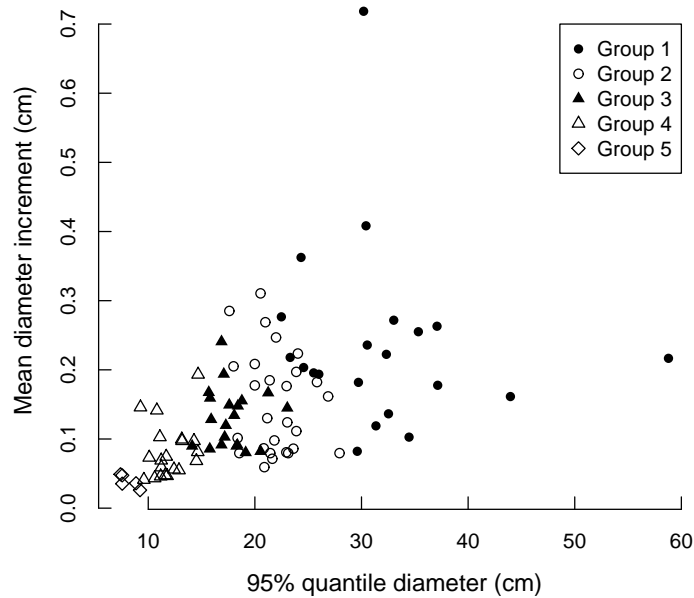


Figure 1: Upper bound of diameters (95% quantile of dbh in 1995, in cm) versus mean diameter increment between 1993 and 1995 (cm) for 93 species at Paracou, French Guiana. The five different symbols correspond to the five groups defined by the mixture matrix model.

References

- Demyanov, V., Wood, S. N., and Kedwards, T. J. (2006). Improving ecological impact assessment by statistical data synthesis using process-based models. *J. Roy. Stat. Soc. C (App. Stat.)*, 55(1):41–62.
- Gourlet-Fleury, S., Guehl, J. M., and Laroussinie, O., editors (2004). *Ecology and Management of a Neotropical Rainforest. Lessons Drawn from Paracou, a Long-Term Experimental Research Site in French Guiana*, Paris. Elsevier.
- Marin, J. M., Mengersen, K., and Robert, C. P. (2005). Bayesian modelling and inference on mixtures of distributions. In Dey, D. and Rao, C. R., editors, *Bayesian Thinking, Modeling and Computation*, number 25 in Handbook of Statistics, pages 459–507, Amsterdam, The Netherlands. Elsevier.
- Mortier, F., Rossi, V., Guillot, G., Gourlet-Fleury, S., and Picard, N. (2012). Population dynamics of species-rich ecosystems: the mixture of matrix population models approach. *Methods in Ecology and Evolution*, In press.
- Picard, N., Köhler, P., Mortier, F., and Gourlet-Fleury, S. (2012). A comparison of five classifications of species into functional groups in tropical forests of French Guiana. *Ecol. Complex.*, 11:75–83.
- Usher, M. B. (1969). A matrix model for forest management. *Biometrics*, 25(2):309–315.