Unsupervised classification of species groups based on mixture matrix population models

Frédéric Mortier[1], Vivien Rossi[2], Nicolas Picard[2] and Sylvie Gourlet-Fleury[2]

[1] CIRAD, Biological Systems Dpt, Research Unit Genetic Diversity and Breeding of Forest Tree Species, Montpellier, France
[2] CIRAD, Environments and Societies Dpt, Research Unit Dynamics of Natural Forests, Montpellier, France

The high species diversity of mixed tropical forests hinders the development of forest dynamics models, because most species are scarce and their dynamics is poorly known. A solution commonly adopted is to cluster species in groups, yielding enough data to allow model calibration. Many tree species classifications have been proposed for modelling purposes: they either rely on life-history and morphological traits or on a quantification of the dynamics processes using data from permanent sample plots. Most often, groups are defined empirically or result from multivariate analyses (PCA and cluster analyses). They usually evidence the existence of two extreme species strategies, i.e. heavy shade-tolerance of species able to germinate and accomplish their whole life-cycle in the understorey and extreme light-dependency of pioneers that need large gaps to germinate, develop and reproduce. In this paper, we propose a new method to classify species according to their evolution characteristics which are modelled by an Usher matrix population model. The Usher model describes the evolution of a population by the vector, N(t), of the number of individuals in m ordered state classes at discrete time t. The species classification method presented is a MCMC based algorithm built in a hierarchical bayesian framework. We developed a reversible jump MCMC algorithm to estimate simultaneously the number of species groups, the classification of the species into the groups and the parameters of the associated matrix population models. We used data on growth, mortality and recruitment from the Paracou Experimental Site in French Guiana, compared the resulting species groups with groups proposed in previous works and discussed their ecological meaning. This method has two main advantages: 1) it provides a unsupervised classification (no a priori on the number of groups) which avoids the sensitive problem of the choice of a model selection criterion 2) it fully takes into account the modelling perspective of the classification (the resulting classification tends to maximize the likelihood of the observations).