

## MODELS OF BIODIVERSITY

**Jerome Chave and Christophe Thebaud**

*Laboratoire Evolution et Diversité Biologique, CNRS/UPS, Toulouse, France*

**Keywords:** Biogeography, community drift, complex systems, diversity, equilibrium, food web, model, niche, process, scale, species-area curve, species richness, tradeoff

### Contents

1. Introduction
2. Description of the biological diversity
  - 2.1. Access to diversity data
  - 2.2. Measures of biological richness
    - 2.2.1 Indices of diversity
    - 2.2.2 Species-area curves
    - 2.2.3 Abundance models
  - 2.3. Measures of ecosystem complexity
3. Dynamic models of diversity
  - 3.1. Generation of diversity
    - 3.1.1. Speciation
    - 3.1.2. Extinction
  - 3.2. Competitive exclusion and niches
    - 3.2.1 Lotka-Volterra model
    - 3.2.2. Niche models
    - 3.2.3 Tradeoff models
    - 3.2.4 Models of density dependence
  - 3.3. Stochasticity and dispersal
    - 3.3.1 Colonization-extinction dynamics
    - 3.3.2 Metacommunity dynamics and community drift
    - 3.3.3 Disturbances
  - 3.4. Relevance of space – Simulations
4. Synthesis and conclusion
- Glossary
- Bibliography
- Biographical Sketches

### Summary

The goal of this chapter is to review mathematical and theoretical models developed for explaining the emergence and the maintenance of biological diversity.

First, we define the problem and the need for a fundamental research program aimed at disentangling the manifold mechanisms contributing to explain patterns of biodiversity. We define and compare several measures of biological diversity, based on species abundances and on scaling properties.

We then turn to mechanisms commonly invoked for explaining patterns of diversity at

large scales, namely speciation and extinction. These large scale processes are crucial to understand smaller-scale patterns of species coexistence. We review additional processes that operate at the community scale, such as niches, density dependence, tradeoffs, and disturbances. We also emphasize the role of stochasticity in the dynamics of natural communities.

Finally, we review models of species coexistence, and ways of exploring emergent features in these models. We provide examples of spatially-extended and stochastic models that can be used both for testing ecological hypotheses and to extract biologically pertinent parameters out of complex datasets.

## 1. Introduction

The causes of the biological diversity have become a major paradigm in life sciences. Why there are so many species in some ecosystems, while only a few can persist in others is a source of almost inextinguishable wonderment. During the development of the grand inventory of the life, regularities in the distribution and abundance of organisms have been observed that called for a formal explanation, a model, be it mathematical or only verbal. Models of biodiversity seek to describe, then to explain why a vast variety of functional and taxonomic life forms are encountered on the Earth. It would be naïve to think that a single process explains this diversity, and there is indeed a vast diversity of putative causal factors for the origination and the maintenance of biodiversity.

The term biodiversity encompasses not only the number of groups of organisms with distinct morphologic traits (phenotypes) within an ecosystem, but also their genetic diversity. A given organism can harbor a great deal of genetic variation within and among populations: for example not all humans have the same blood group, and this difference signals a difference at the genetic level. It includes also the functional diversity, the value and range of species traits. The debate over the level of description at which the term biodiversity is best defined is still ongoing and is a deep scientific question. DNA-based information allows systematians to reorganize entire families of species, transforming the way of thinking about the organization of biodiversity. Yet, for most applied questions, biodiversity is most conveniently defined in its taxonomic sense: the number of morphologically distinct entities. In practice, many groups of organisms are still poorly known, and their diversity cannot be reliably assessed using morphology (e.g. bacteria, fungi). Recent advances have been made possible by the use of molecular markers for these groups.

The decision-making process of whether a natural area should be protected as a hotspot of diversity primarily refers to the taxonomic diversity that experts have been able to inventory in this area or habitat. This process is necessarily inexact and incomplete, as organisms could be missed, for example if the area is undersampled, organisms have been misidentified, certain scales of descriptions have been overlooked, or simply if the systematian who would be able to correctly identify the organism is unavailable or dead. Such problems are particularly acute in the tropical zone, where the species richness is vastly greater and less studied than in the temperate zone. These problems arising during the systematic description of the Earth's biodiversity have prompted an

effort to understand patterns of diversity from a theoretical perspective. Statistical techniques have been developed to relate partial information on species richness to its community-scale value, to extrapolate local information to larger areas.

A more general program has attempted to go beyond this descriptive method, since the early XX<sup>th</sup> century and the works of the Italian mathematician Vito Volterra. Ecologists have developed mathematical models to generate patterns of biodiversity that could be expected from a small number of well-understood ecological processes. Ever since the early development of theoretical ecology, concepts borrowed from dynamical systems theory have come to the forefront of this discipline. The relationship between the diversity of an ecological community and the ‘complexity’ of this community – measured, for example, by the nature and the strength of interactions among species – have been explored. These models have been used to investigate the stability of ecosystems against perturbations such as the invasion of alien species, or the extinction of native species. Other emergent relationships such as the relation between diversity and the productivity of an ecosystem have also received a great deal of attention.

This paper attempts at providing an overview of the statistical and mechanistic modeling frameworks that have been developed to analyze patterns of diversity and to infer processes from these patterns. Throughout this contribution, it is implicit that the tools developed in population ecology are most of the time also applicable in other fields. For example indices of diversity have been developed in community ecology, but have been applied to fields as diverse as genetics, economy, finance or sociology. When reading these definitions, maybe for the first time, the reader should keep in mind that the concept of diversity is applicable to all levels of integration in biology (genes, cells, individuals, social groups, etc...).

## 2. Description of the biological diversity

### 2.1. Access to diversity data

Classifying the diversity is the first, and essential, step toward an understanding of the underlying patterns. Hierarchical classifications using morphological similarities among organisms are the most efficient method for sorting the Earth’s biodiversity. Of foremost importance in this approach are reproductive characters of the organisms (for example the flowers in flowering plants). The tree-like structure of this classification informs on the similarity between two groups of closely related organisms. At the finest level of description, these groups are called *species*. A group of species is called *genus* (one genus, two genera). Each species is described by a genus name and a species name (e.g. for modern humans, the genus name is *Homo*, and the species name, *sapiens*). Each genus can have one or several species, and it must belong to a single *family* (Hominidae for humans). This structure resembles a tree, where each group contains one or more sub-groups, and belongs to exactly one higher-level hierarchical unit of classification. The list of accepted species is the result of a consensus among specialists, and not a fundamental truth. This list is far from being complete. Today, slightly less than 2,000,000 species have been described, most of them arthropods (around 1,000,000), plants (250,000) and animals (45,000). This may represent only a small fraction of the Earth’s species richness, as has been suggested by the entomologist Terry

Erwin in the early 1980s. For example, the 1970 estimate of 250,000 species of flowering plants, one of the best-studied group, has been revised to over 320,000 species by specialists of the Royal Botanic Garden at Kew, UK, the leading authorities in plant taxonomy.

Trees can also be analyzed in a historical context, that of the theory of evolution. Two species in the same genus have a more recent ancestor than either one of these species with a species of another genus. This classification by the degree of historical relatedness is called a *phylogeny*. The reconstruction of these phylogenies using molecular-based information (ie. the degree of variability in DNA molecules, or in proteins among individuals) is the topic of intense ongoing activity. For example, most families of flowering plants have been recently classified, using a set of molecular markers, and the resulting phylogeny, while being consistent with previous classifications, has still revealed surprises. For example, a basic taxonomic distinction in flowering plants is between plants that produce embryos with one or two cotyledons (seed leaves). The former (monocots) encompasses many herbs and palms, while the latter (dicots) comprises most woody species. The molecular phylogeny has best represented two main groups, monocots and eudicots (comprising most dicots) both nested into a third group called Magnoliid dicotyledons that comprises only 3% of the living species. In other words, the distinction monocots/dicots is not natural.

This hierarchical organization greatly simplifies the description of new species and the organization of the current knowledge, especially since the development of computer databases of the world's diversity has made possible a rapid storage and retrieval of this information (e.g. <http://www.sp2000.org/>). Indeed, it can be shown that the optimal way of storing  $n$  pieces of information is in a tree structure, where the search time is generally on the order of  $\ln(n)$ . Optimized storage-retrieval devices, often called data-models, are becoming crucial in the management of the Earth's biodiversity. Ideally, a description of the physiology and behavior, geographical range, natural history, genetic properties (population genetic structure, chromosome number, ...) should be obtained for every known species. This information would occupy at least several megabytes per species. Therefore, complete databases should be more than  $10^9$  bytes (10 terabytes). Consequently, data models will require sophisticated algorithms to quickly retrieve the information in these enormous databases.

## 2.2. Measures of biological richness

### 2.2.1 Indices of diversity

Biodiversity is the irreducible complexity of all life, including not only the great variety of forms among organisms, but also their varying behavior and interactions. In addition, biodiversity is sometimes seen as a currency of value by conservationists. In that context, quantifying the biological diversity is equivalent to an economical valuation process. However, no single objective measure of this diversity is possible, and synthetic measures should be manipulated with caution. For example, one could be interested in the measure of the number of species living in an ecosystem and that are restricted to a specified region or locality. Highly ranked ecosystems, i.e. those containing a large proportion of species with a small distribution range would be

classified as ‘hotspots’, and should be prioritized by protection programs.

The same approaches are shared by ecologists, who seek to quantify species richness in a community, and by geneticists, who are interested in the genetic variability in a population. Measurement of diversity generally concerns a *finite* population of  $N$  organisms, grouped into  $S$  classes. By class, we mean a group of biological objects, for example taxa, species, genera, or others. This definition is straightforwardly generalized to the subsampling of  $N$  individuals in an *infinite* population. The abundance of organisms in a class  $i$  is noted  $N_i$ , and the *relative* abundance is  $p_i = N_i/N$ . The number of classes  $S$  is the most basic measure of biodiversity in ecology. It says nothing, however, about the relative abundance and the commonness of classes within the sample.

The relative abundance  $p_i$  can be interpreted as the probability distribution that an individual chosen at random belongs to taxon  $i$ . By definition of a probability distribution, one has:

$$\sum_{i=1}^S p_i = 1 \quad (1)$$

A simple measure of diversity, is the probability that two randomly chosen organisms belong to different classes, a quantity called *Simpson’s index* of diversity, or Nei’s index of heterozygosity (genetic diversity):

$$D_2 = 1 - \sum_{i=1}^S p_i^2 \quad (2)$$

More precisely, an unbiased estimator of Simpson’s index is

$$D_2' = 1 - \sum_{i=1}^S \frac{N_i(N_i - 1)}{N(N - 1)} \quad (3)$$

This formula compares two organisms taken at random in the system and under the assumption that they are different (they are drawn at random without replacement). Notice that equation (3) is equivalent to equation (2) if, for all  $i$ ,  $N_i$  is much larger than 1. This condition is naturally realized in an infinite population. One can easily generalize this definition, by looking at the probability that  $k$  randomly chosen organisms all belong to different categories:

$$D_k = 1 - \sum_{i=1}^S p_i^k \quad (4)$$

In the language of probability theory, this quantity closely related to the  $k$ th moment of the probability distribution.

Another widely used measure of diversity is related to information theory, and has been interpreted as ‘Shannon’s entropy’, or ‘complexity’ of the system:

$$H = -\sum_{i=1}^S p_i \ln(p_i) \quad (5)$$

This measure is simply related to the first derivative of the index  $D_k$  around  $k=1$ :

$$\left. \frac{dD_k}{dk} \right|_{k=1} = \lim_{k \rightarrow 1} \frac{D_k - D_1}{k - 1} = -\sum_{i=1}^S p_i \left( \lim_{k \rightarrow 1} \frac{p_i^{k-1} - 1}{k - 1} \right) = -\sum_{i=1}^S p_i \ln p_i = H \quad (6)$$

The probabilistic basis for the definition of  $H$  as an index of diversity is weak, and one should always prefer easily understandable indices, such as the species richness  $S$  or the Simpson's index. Many other diversity indices can be defined but they lack a clear statistical and ecological interpretation.

### 2.2.2 Species-area curves

How does the number of species increase with the size of the sampled area? Since it has been first noted by the Swiss botanist de Candolle in 1855, the relation between sampled area and species richness has been discussed and quantified. This question has a vast range of practical implications. When one wants to protect “hotspots” of diversity (again both genetic and specific diversity), the protected area should have a size such that a maximal amount of diversity is protected at a time.

When islands at the same distance from the mainland are compared, it turns out that the species richness increases with the island size. Experimental work has consistently shown that the average number of species  $S$  increases with the area  $A$ , raised to a certain power, that is

$$S = bA^z \quad (7)$$

with  $b$  and  $z$ , two parameters,  $z$  being generally found between 0.1 and 0.4. As early as 1910, this model has been suggested as a good empirical relationship between species richness and area by the Suede botanist Arrhenius. Notice that in saturated communities, the area is proportional to the number of individuals  $N = \rho A$ , and the relationship deduced from the log-normal relative abundance curve implies exactly a power-law relationship, with the condition  $z = 0.25$ , which falls in the range of observed values.

Although this functional relationship between land area and species richness has been often assumed to be a truth, it is worthwhile emphasizing that its use to estimate species numbers at the regional scale from local inventories should be made only with the utmost care. For example, experiments conducted by the American ecologist Gleason on plant diversity in the early 1920s, have challenged Arrhenius power-law relationship between species richness and area. Gleason's works lead to the conclusion that for a saturated community, a good empirical model was

$$S = \alpha \ln \left( 1 + \frac{N}{\alpha} \right) \quad (8)$$

In a saturated community, the number of individuals  $N$  is proportional to the land area  $A$ , if  $\rho = N/A$  is the plant density per unit area. The parameter  $\alpha$  is a free parameter of the model. This apparent paradigm, discussed by Connor and McCoy in 1979, is partly a consequence of the imprecise definition of species-area relationships, and of the differences in the processes involved in the maintenance of species richness in animal or in plant communities. We will see in section 3.3.2 how this empirical model relates to a dynamic model of diversity.

-  
-  
-

TO ACCESS ALL THE 25 PAGES OF THIS CHAPTER,  
Visit: <http://www.eolss.net/Eolss-sampleAllChapter.aspx>

### Bibliography

M. Bramson, J.T. Cox and R. Durrett. A spatial model for the abundance of species. *Annals of Probability* 26, 658–709, 1998. (a difficult but insightful paper on spatially-explicit mathematical ecological models)

D.L. DeAngelis, W.M. Mooij. Individual-based modeling of ecological and evolutionary processes. *Annual Review of Ecology, Evolution and Systematics*. 36, 147-168, 2005. (a review of individual-based models)

S. Engen. *Stochastic Abundance Models*. Monographs on Applied Probability and Statistics, Chapman and Hall, 1978. (a useful introduction to the probabilistic theory of species sampling)

W.J. Ewens. *Mathematical Population Genetics*. Berlin Springer-Verlag. 1979. (a fundamental textbook – recently reedited – in theoretical population genetics).

G.F. Gause. *The Struggle for Existence*. Williams and Wilkins, Baltimore 1934. (a classic of ecological theory).

S.P. Hubbell. *A Unified Neutral Theory of Biodiversity and Biogeography*. Monographs in Population Biology, Princeton University Press, Princeton NJ, 2001. (a landmark publication presenting and discussing the neutral theory of biodiversity)

M.A. Huston. *Biological Diversity: Coexistence of Species on Changing Landscapes*. Cambridge University Press, Cambridge UK, 1994. (a textbook on theories of species coexistence)

R.H. MacArthur and E.O. Wilson. *The Theory of Island Biogeography*. Monographs in Population Biology, n 1, Princeton University Press, Princeton NJ, 1967. (a classic of island biogeography)

R.E. Ricklefs and D. Schuler. *Species Diversity in Ecological Communities. Historical and Geographical Perspectives*. The University of Chicago Press, 1993. (a series of articles by renowned ecologists on species diversity, patterns and processes)

M.L. Rosenzweig. *Species Diversity in Space and Time*. Cambridge University Press, Cambridge UK, 1995. (a thorough review of patterns of biodiversity)

D. Tilman. *Plant Strategies and the Dynamics and Structure of Plant Communities*. Princeton University

press, Princeton NJ, 1988. (a classic book on plant competitive strategies)\

### **Biographical Sketches**

**Jerome Chave**, is researcher with the Centre National pour la Recherche Scientifique, a nation-wide Research Institute. He is working in Toulouse in collaboration with Université Paul Sabatier. He is a theoretical ecologist, and is currently working on understanding the mechanisms by which species can coexist in similar environments. Specifically, Jerome Chave is combining mathematical and computer-based models to explore the different mechanisms proposed to explain the astounding biodiversity found in tropical forests. Recently, this has lead him to investigate the relationships between ecological theories and theory in population genetics. He is also interested in the biogeochemical cycling coupling ecosystems and the atmosphere. And is involved in projects aimed at providing biometric methods for measuring C stock and balance for tropical forests.

**Christophe Thebaud**, is professor of Evolutionary Ecology in Université Paul Sabatier, Toulouse. His main interest is to understand the mechanisms by which biological diversity arises, specifically through the appearance of new species. In recent years he has been working on a model plant species which shows a rapid pattern of morphological differentiation, possibly related to an ongoing process of speciation. He is combining ecological experiments to molecular biology techniques both in the light of evolutionary theory. In the past he has been actively involved in projects of island biogeography, and the study of diversification patterns of both plant and bird natural communities in the Mascarene archipelago.