



**HAL**  
open science

# Multivariate presence-absence distribution and the neutral theory

F. Laroche

► **To cite this version:**

F. Laroche. Multivariate presence-absence distribution and the neutral theory. GDR TheoMoDive - 2022 annual meeting, Sep 2022, Moulis, France. hal-03799828

**HAL Id: hal-03799828**

**<https://hal.inrae.fr/hal-03799828>**

Submitted on 6 Oct 2022

**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.

# Multivariate presence-absence distribution and the neutral theory

F. Laroche, UMR Dynafor

GDR TheoMoDive, Moulis, September 26th 2022



INRAE

# Ensembles species composition through the prism of metacommunity theory

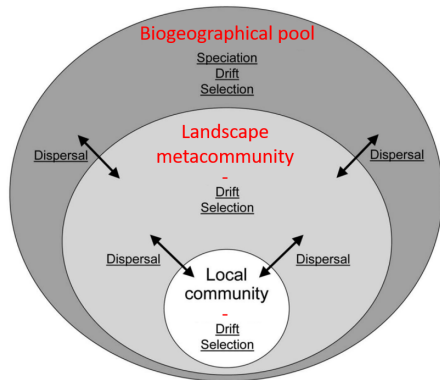
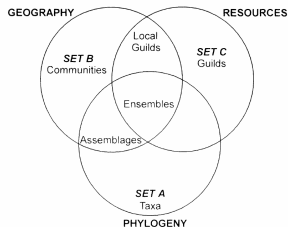


FIGURE 4. THE THEORY OF COMMUNITY ECOLOGY  
Selection, drift, speciation, and dispersal interact to determine community dynamics across spatial scales.

Adapted from Fauth et al. (1996) and Vellend (2010)

# Rephrasing 'selection' in terms of coexistence mechanisms

Classic approach of coexistence among  $J$  species:

$$\left( \frac{dx_j}{dt} = x_j(t)g_j(\mathbf{x}(t)) \right)_{1 \leq j \leq J}$$

where  $x_j(t)$  is the density of species  $j$  at time  $t$ .

- ▶ generalized Lotka-Volterra:  $g_j(\mathbf{x}) = r_j + (A\mathbf{x})_j$ 
  - ▶ differences in  $r_j$ s and  $A_{js}$  drives coexistence, potential equilibria and relative abundances of species

$$\text{Ex : } A = \begin{pmatrix} -\alpha_{11} & -\alpha_{12} \\ -\alpha_{21} & -\alpha_{22} \end{pmatrix}, \text{ coexistence iff } \frac{\alpha_{12}/r_1}{\alpha_{22}/r_2} < 1, \frac{\alpha_{21}/r_2}{\alpha_{11}/r_1} < 1$$

- ▶ to be combined with dispersal and drift

$$\mathbb{E} [x_j(t + dt) - x_j(t) | x_j(t)] = (x_j(t)g_j(\mathbf{x}(t)) + m_j(\mathbf{x}(t))) dt$$

# Assessing the contribution of selection to empirical patterns

Need for a null hypothesis with no selection effect while keeping dispersal and drift.

- ▶ *weak neutral assumption*:  $g_j(\mathbf{x}) = g(\mathbf{x})$  for all species  $j$ ;
  - ▶ generalized Lotka-Volterra framework :  $g(\mathbf{x}) = r + \mathbf{t} \mathbf{a} \mathbf{x}$
- ▶ *strong neutral assumption*:  $g(\mathbf{x}) = g(\sum_{j=1}^J x_j)$ ;
  - ▶ Generalized Lotka-Volterra framework :  $g(\mathbf{x}) = r + a^t \mathbf{1} \mathbf{x}$

In all cases, different from the independent species assumption.

# The neutral model of Hubbell (2001)

One way to combine dispersal and drift with the strong neutral assumption.

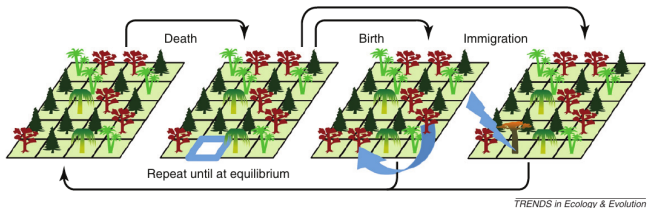


Figure from Rosindell et al. (2011)

Directly adapted from neutral models of population genetics.

# The mathematical side of Hubbell (2001) model

Community size  $K$  is a finite deterministic parameter. The dynamics of species abundances  $\mathbf{Y}$  is a *jump process* on the  $J$ -dimensional discrete simplex  $\Delta_K$ .

$$\begin{aligned} \dot{\mathbb{P}}(\mathbf{Y}(t) = \mathbf{y}) \\ = Kd \left( \begin{array}{c} \sum_{j_d, j_b=1}^J \mathbb{P}(\mathbf{Y}(t) = \mathbf{y} + \mathbf{e}_{j_d} - \mathbf{e}_{j_b}) \\ \times \frac{y_{j_d} + 1}{K} \left( (1 - m) \frac{y_{j_b} - 1}{K - 1} \mathbb{I}_{y_{j_b} \geq 1} + m\pi_{j_b} \right) \\ - \mathbb{P}(\mathbf{Y}(t) = \mathbf{y}) \end{array} \right) \end{aligned}$$

where  $\mathbf{e}_j$  is the vector with 1 at coordinate  $j$  and 0 elsewhere, and  $\boldsymbol{\pi}$  is the vector of regional relative abundances.

## Stationary distribution in Hubbell (2001) model

Stationary distribution  $\mathbb{P}^*$  follows *detailed balance*:

$$\begin{aligned}\mathbb{P}^*(\mathbf{y}) &\times \frac{y_{j_1}}{K} \left( (1-m) \frac{y_{j_2}}{K-1} \mathbb{I}_{y_{j_2} \geq 1} + m\pi_{j_2} \right) \\ &= \mathbb{P}^*(\mathbf{y} + \mathbf{e}_{j_2} - \mathbf{e}_{j_1}) \times \frac{y_{j_2} + 1}{K} \left( (1-m) \frac{y_{j_1} - 1}{K-1} \mathbb{I}_{y_{j_1} \geq 1} + m\pi_{j_1} \right)\end{aligned}$$

Defining the effective number of immigrants

$I = m(K-1)/(1-m)$  (Etienne and Olf, 2004):

$$\mathbb{P}^*(\mathbf{y}) = \binom{K}{\mathbf{y}} \frac{\prod_{j=1}^J (I\pi_j)^{(y_j;1)}}{I^{(K;1)}} \text{ with } \theta^{(y,c)} := \prod_{k=0}^{y-1} (\theta + ck)$$

which is a Dirichlet-multinomial distribution  $\mathcal{DM}_{\Delta_K}(I\pi)$  (Donnelly et al., 2001; Harris et al., 2017)



## Relaxing zero-sum assumption

Community composition  $\mathbf{y}$  now follows a jump process on  $\mathbb{N}^J$ .

Birth, death and immigration are distinct events.

Strong neutral assumption  $\implies$  rates depend only on the total number of individuals  $|\mathbf{y}|$

$$\begin{aligned}\dot{\mathbb{P}}(\mathbf{Y}(t) = \mathbf{y}) = & \\ & \sum_{j_d=1}^J \mathbb{P}(\mathbf{Y}(t) = \mathbf{y} + \mathbf{e}_{j_d})(y_{j_d} + 1)d(|\mathbf{y}| + 1) \\ & + \sum_{j_b=1}^J \mathbb{P}(\mathbf{Y}(t) = \mathbf{y} - \mathbf{e}_{j_b}) [(y_{j_b} - 1)b(|\mathbf{y}| - 1) + m(|\mathbf{y}| - 1)\pi_{j_b}] \\ & - \mathbb{P}(\mathbf{Y}(t) = \mathbf{y}) [|\mathbf{y}|(b(|\mathbf{y}|) + d(|\mathbf{y}|)) + m(|\mathbf{y}|)]\end{aligned}$$

where:

- ▶  $b(\cdot)$  and  $d(\cdot)$  stand for per capita birth and death rates respectively ;
- ▶  $m(\cdot)$  is the immigration rate.

Decomposing stationary distribution into sum and split:

$$\mathbb{P}^*(\mathbf{y}) = \mathbb{Q}^*(|\mathbf{y}|) \mathbb{P}_{|\mathbf{y}|}^*(\mathbf{y})$$

# The effective number of migrants hypothesis

Stationary distribution does not necessarily follow detailed balance, an ancillary hypothesis is needed:

$$\exists I \in \mathbb{R}^+, \forall y \in \mathbb{N}^*, \frac{m(y)}{b(y)} = I \quad (1)$$

- ▶ constant immigration of juveniles
- ▶ regulation occurs on juveniles
  - ▶ quite suited for trees

Under hypothesis (1), the split is Dirichlet-multinomial (Haegeman and Etienne, 2008):

$$\mathbb{P}_{|\mathbf{y}|}^*(\mathbf{y}) \sim \mathcal{DM}_{\Delta_{|\mathbf{y}|}}(I\boldsymbol{\pi})$$

$I$  is still called 'effective number of migrants'.

# Outline

**Question:** Can we relax of the effective number of migrants hypothesis (1) to reach more general neutral models ?

1. Closure property of neutral stationary distributions
2. Presence-absence distributions through transfer functions
3. Polynomial transfer functions : interpretation and simulation

## Closure property of stationary distribution

Let species indices  $l, m \in \{1, J\}$  with  $l < m$ . For  $\mathbf{x} \in \mathbb{R}^J$ , define  $\tilde{\mathbf{x}} \in \mathbb{R}^{J-1}$  such that:

$$\begin{cases} \tilde{x}_j = x_j & \text{if } j < l \\ \tilde{x}_j = x_l + x_m & \text{if } j = l \\ \tilde{x}_j = x_j & \text{if } l < j < m \\ \tilde{x}_j = x_{j+1} & \text{if } m \leq j \end{cases}$$

Then :

$$\begin{aligned} \dot{\mathbb{P}}(\tilde{\mathbf{Y}}(t) = \tilde{\mathbf{y}}) = & \sum_{j_d=1}^J \mathbb{P}(\tilde{\mathbf{Y}}(t) = \tilde{\mathbf{y}} + \mathbf{e}_{j_d})(\tilde{y}_{j_d} + 1)d(|\tilde{\mathbf{y}}| + 1) \\ & + \sum_{j_b=1}^J \mathbb{P}(\tilde{\mathbf{Y}}(t) = \tilde{\mathbf{y}} - \mathbf{e}_{j_b}) [(\tilde{y}_{j_b} - 1)b(|\tilde{\mathbf{y}}| - 1) + m(|\tilde{\mathbf{y}}| - 1)\tilde{\pi}_{j_b}] \\ & - \mathbb{P}(\mathbf{Y}(t) = \tilde{\mathbf{y}}) [|\tilde{\mathbf{y}}|(b(|\tilde{\mathbf{y}}|) + d(|\tilde{\mathbf{y}}|)) + m(|\tilde{\mathbf{y}}|)] \end{aligned}$$

which implies the closure property of the stationary distribution:

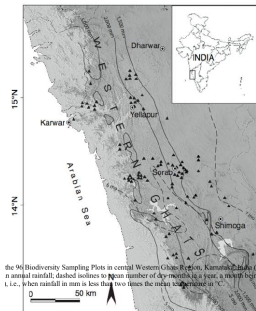
$$\tilde{\mathbb{P}}_{\pi}^*(\tilde{\mathbf{y}}) = \mathbb{P}_{\tilde{\pi}}(\tilde{\mathbf{y}})$$

all other parameters being kept constant.

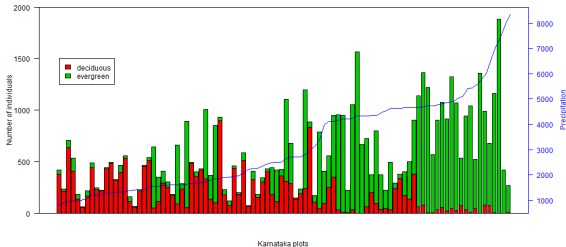
# Testing closure in abundance data with the $DM$ split

$DM$  split implies closure property (Peyhardi et al., 2021).

- ▶ testing functional groups through comparing estimates of  $\hat{I}$  with various levels of species aggregation
- ▶ example in tropical trees ensembles



the 90 Biodiversity Sampling Plots in central Western Ghats, Karnataka, India (KBS n annual rainfall; dashed isolines to mean number of dry months in a year, a month being  $< 1$ , i.e., when rainfall in mm is less than two times the mean temperature in  $^{\circ}C$ )  
recompos a wide range of habitats from hyper-wet evergreen forests (rainfall  $\geq 4000$  mm/yr)



Lumping into deciduous/evergreen yields lower  $\hat{I}$   
Suggests species sorting along rainfall gradient

Ramesh et al. (2010); Laroche et al. (2020)

# The transfer function driving neutral occupancies

What if not working on tropical trees ? Characterizing a general set of distributions with closure property ?

- ▶ start simple: presence-absence (PA) distributions

Let a neutral jump process. Closure property implies that there exists a transfer function  $f : [0, 1] \rightarrow [0, 1]$  such that:

$$\forall J \in \mathbb{N}^*, \forall \boldsymbol{\pi} \in \Delta_J, \mathbb{P}(Y_j = 1) = f(\pi_j)$$

Closure property implies that for all  $\mathbb{S} \in \mathcal{P}(\{1, \dots, J\})$ :

$$\mathbb{P}(\cup_{j \in \mathbb{S}} \{Y_j = 1\}) = f\left(\sum_{j \in \mathbb{S}} \pi_j\right)$$

which is sufficient to determine the multivariate PA distribution (Teugels, 1990).

- ▶ Proposal : forget the underlying process and focus on the transfer function ?

# Necessary and sufficient conditions for valid transfer functions

Not all  $f : [0, 1] \rightarrow [0, 1]$  yields a valid multivariate PA distribution.  
Focusing on smooth functions,  $f$  is a valid transfer function iff :

$$\begin{cases} f(0) = 0 \\ f(1) \leq 1 \\ \forall k \in \mathbb{N}^*, \forall \pi \in [0, 1], (-1)^{k-1} f^{(k)}(\pi) \geq 0 \end{cases}$$

Biologically speaking :

- ▶  $f(0) = 0$  makes sense;
- ▶  $f$  monotonically increasing over  $[0, 1]$  also makes sense;
- ▶  $f(1)$  is the occupancy of the full ensemble.

Example :  $f(\pi) = 1 - e^{-\theta\pi}$ ,  $\theta \in \mathbb{R}^+$  is a valid transfer function that yields to independent marginals

## Split-polynomial transfer functions : biological interpretation

Let  $f$  a polynomial transfer function. Its degree  $K$  can be interpreted as the maximum number of individuals within the community.

Example :  $f(\pi) = \theta\pi, \theta \in [0, 1]$  is a valid transfer function that yields to no co-occurrences.

Further assuming that  $f$  is split on  $\mathbb{R}$ , then  $\exists 1 = \lambda_1 \leq \dots \leq \lambda_K$ :

$$\forall \pi \in [0, 1], f(\pi) = 1 - \prod_{k=1}^K \left(1 - \frac{\pi}{\lambda_k}\right)$$

Taking  $\lambda_k = 1 + \frac{k-1}{I}, I \in \mathbb{R}^{+*}$  yields Hubbell (2001) model

- ▶ other profiles of  $\lambda_k$  can be chosen;
- ▶  $1/\lambda_k$  may be understood as a probability of immigration along a backward process.



## Split-polynomial transfer functions : urn simulation algorithm

The interpretation of  $\lambda_k$  in terms of immigration along a backward process yields a simulation algorithm for split-polynomial  $f$ :

1. initialize  $\mathbf{a} \in (\{1, \dots, J\})^K$  at  $a_i = 0, \forall 1 \leq i \leq K$ ;
2. draw  $a_1 \sim \mathcal{M}(1, \boldsymbol{\pi})$ ;
3. set  $i$  to 2;
4. if  $i > K$  go directly to step 9
5. draw  $B \sim \mathcal{B}(\frac{1}{\lambda_i})$ ;
6. if  $B = 1$ , draw  $a_i \sim \mathcal{M}(1, \boldsymbol{\pi})$  else draw  $a_i$  at random in  $\{a_1, \dots, a_{i-1}\}$ ;
7. set  $i$  to  $i + 1$ ;
8. go back to step 4;
9. let  $\mathbf{x} \in \{0, 1\}^J$  such that  $\forall j \in \mathbb{I}_J, x_j = 0$  if  $\{a_i, a_i = j\} = \emptyset$ ,  $x_j = 1$  otherwise ;
10. return  $\mathbf{x}$ .

## Take-home messages

- ▶ The essence of neutral hypothesis is the *closure property* (CP) of associated multivariate count distributions → we propose to build null distributions from CP directly, rather than considering specific dynamical neutral models.
- ▶ Multivariate presence-absence distributions with CP can be parsimoniously described through a single *transfer function* that relates regional relative abundance of any species to its local occupancy.
- ▶ Split-polynomial transfer functions are biologically interpretable as a *backward process*, are easy to simulate through an *urn algorithm*, and extend classic neutral models.

## Next steps

- ▶ clarifying the link between  $\lambda_k$  and assumptions on immigration in split-polynomial  $f$
- ▶ determining whether transfer functions can be non-split polynomial (on  $\mathbb{R}$ );
- ▶ implementing statistical routines to fit split-polynomial transfer functions;
- ▶ adapting the approach of Laroche et al. (2020) to detect distinct neutral clusters in presence-absence datasets;
- ▶ apply this framework to saproxylic beetles communities within tree-related microhabitats.



# Acknowledgements

**Collaborators** — F. Mortier and J. Peyhardi on the study of closure properties in multivariate count distributions; F. Munoz, A. Taudiere and C. Violle on the study of tropical tree communities in Karnataka.

**Fundings** — This work is supported by the BloBiForM ANR project (ANR-19-CE32-0002-01) → #BloBiForM

# References I

- Donnelly, P., Nordborg, M., and Joyce, P. (2001). Likelihoods and simulation methods for a class of nonneutral population genetics models. *Genetics*, 159(2):853.
- Etienne, R. S. and Olf, H. (2004). A novel genealogical approach to neutral biodiversity theory. *Ecology Letters*, 7(3):170–175. WOS:000189232400002.
- Fauth, J. E., Bernardo, J., Camara, M., Resetarits, W. J., Van Buskirk, J., and McCollum, S. A. (1996). Simplifying the jargon of community ecology: a conceptual approach. *The American Naturalist*, 147(2):282–286.
- Haegeman, B. and Etienne, R. S. (2008). Relaxing the zero-sum assumption in neutral biodiversity theory. *Journal of Theoretical Biology*, 252(2):288–294.
- Harris, K., Parsons, T. L., Ijaz, U. Z., Lahti, L., Holmes, I., and Quince, C. (2017). Linking statistical and ecological theory: Hubbell's unified neutral theory of biodiversity as a hierarchical dirichlet process. *Proceedings of the IEEE*, 105(3):516–529.

## References II

- Hubbell, S. P. (2001). *The unified neutral theory of biodiversity and biogeography*. Princeton University Press, Princeton, NJ.
- Laroche, F., Violle, C., Taudiere, A., and Munoz, F. (2020). Analyzing snapshot diversity patterns with the Neutral Theory can show functional groups effects on community assembly. *Ecology*, 101(4):e02977. Publisher: John Wiley & Sons, Ltd.
- Peyhardi, J., Fernique, P., and Durand, J.-B. (2021). Splitting models for multivariate count data. *Journal of Multivariate Analysis*, 181:104677.
- Ramesh, B. R., Swaminath, M. H., Patil, S. V., Dasappa, Plissier, R., Venugopal, P. D., Aravajy, S., Elouard, C., and Ramalingam, S. (2010). Forest stand structure and composition in 96 sites along environmental gradients in the central western ghats of india. *Ecology*, 91(10):3118–3118.
- Rosindell, J., Hubbell, S. P., and Etienne, R. S. (2011). The unified neutral theory of biodiversity and biogeography at age ten. *Trends in Ecology & Evolution*, 26(7):340–348.

## References III

- Teugels, J. L. (1990). Some representations of the multivariate bernoulli and binomial distributions. *Journal of Multivariate Analysis*, 32(2):256–268.
- Vellend, M. (2010). Conceptual synthesis in community ecology. *Quarterly review of biology*, 85(2):183–206.