

# Theory-Driven Analysis of Ecological Data

## Day 3: Spatial data

Vincent Calcagno [vincent.calcagno@inrae.fr](mailto:vincent.calcagno@inrae.fr)

Maxime Dubart [dubartmaxime@gmail.com](mailto:dubartmaxime@gmail.com)



# Outline of the day

## Morning:

- Introduction: typical data? typical questions?
- Part 1: Descriptive approaches (aka statistical models)
- Part 2: Process-based approaches (aka mechanistic models)
- Conclusion: what's best, if anything?

## Afternoon:

- Practical 1: on inferring species interactions from co-occurrences
- **Practical 2: fitting a patch occupancy model to detect competition**

# Outline of the day

## Morning:

- Introduction: typical data? typical questions?
- Part 1: Descriptive approaches (aka statistical models)
- Part 2: Process-based approaches (aka mechanistic models)
- Conclusion: what's best, if anything?

## Afternoon:

- Practical 1: on inferring species interactions from co-occurrences
- **Practical 2: fitting a patch occupancy model to detect competitio**

**Evening speaker:** Davide Martinetti, INRAE, Avignon

“Predicting the risk of invasion of the Japanese beetle *Popillia japonica* in Europe”.



# What are ecological communities ?

A set of species occurring at the same place, at the same time

# What are ecological communities ?

A set of species occurring at the same place, at the same time

Spatial boundaries can be arbitrary (study plots)

**Plants on forest**

**Invertebrates in grassland**

**Corals on reefs**

# What are ecological communities ?

A set of species occurring at the same place, at the same time

**Snails in ponds**

**Ants on islands**

**Flies on fruits**

# What are ecological communities ?

A set of species occurring at the same place, at the same time

Not restricted to one trophic level or resource use

**Plants & herbivores on forest**

**Invertebrates & plants in grassland**

**Corals & fishes on reefs**

# The typical spatial dataset

The core thing

Species (S)

Sampling units (N)

**Community matrix**

Presence/Absence  
(Detection/Non Detection)

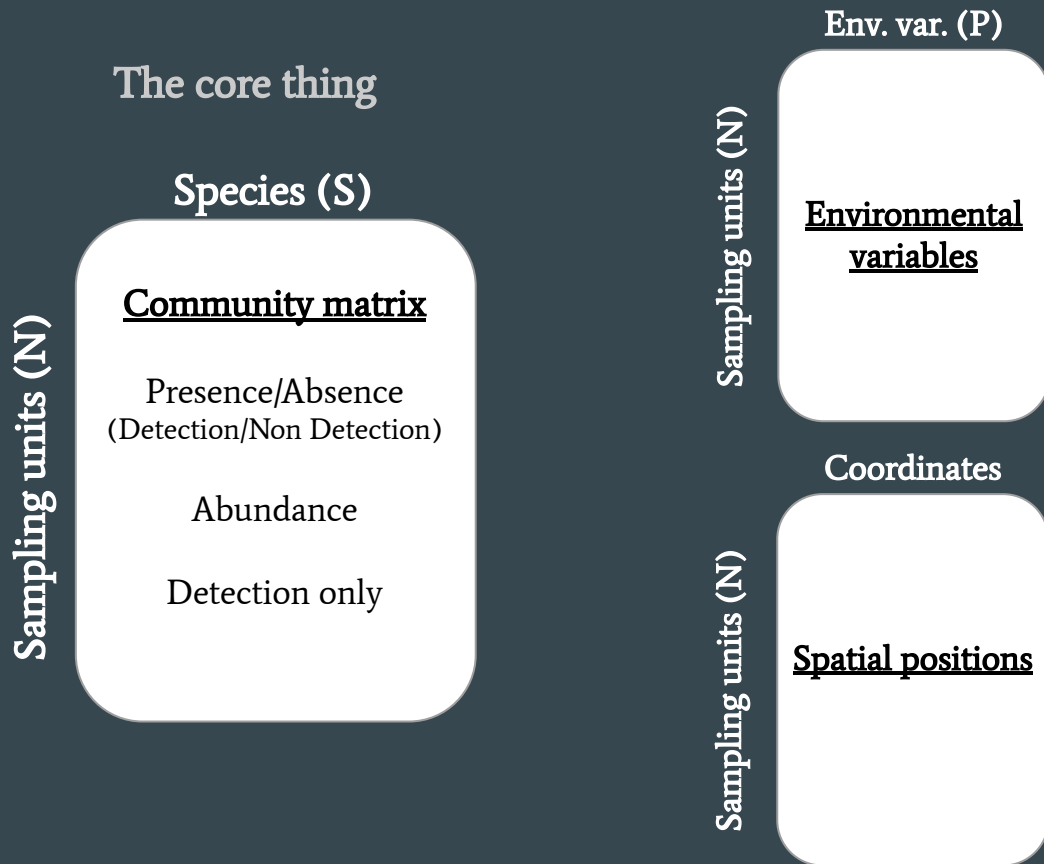
Abundance

Detection only

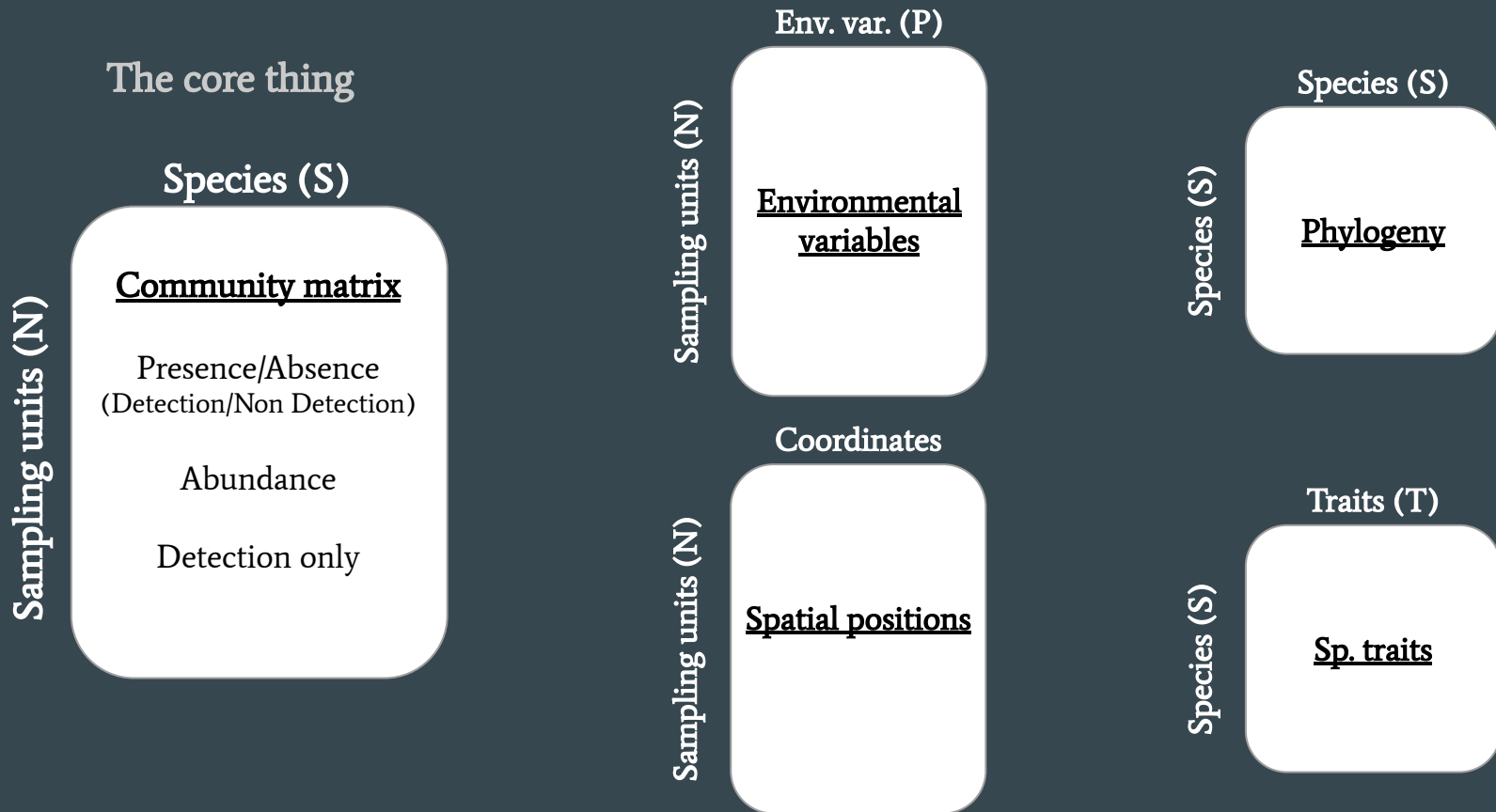


# The typical spatial dataset

Additional data

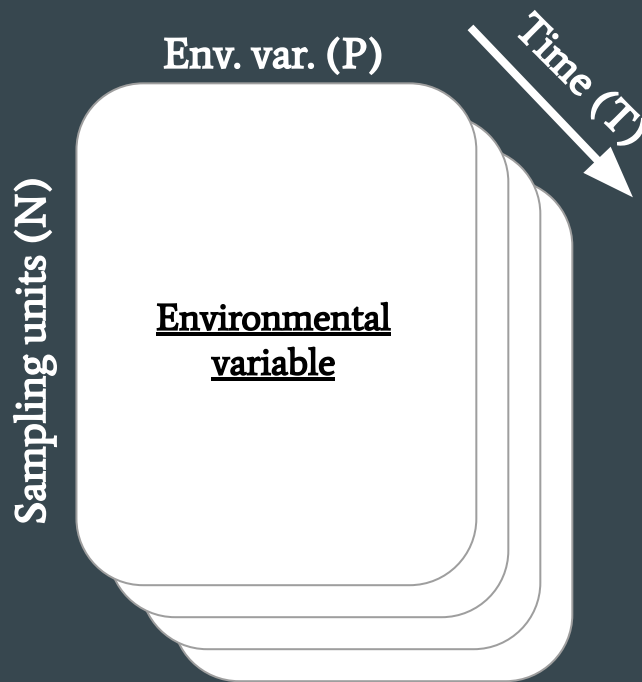
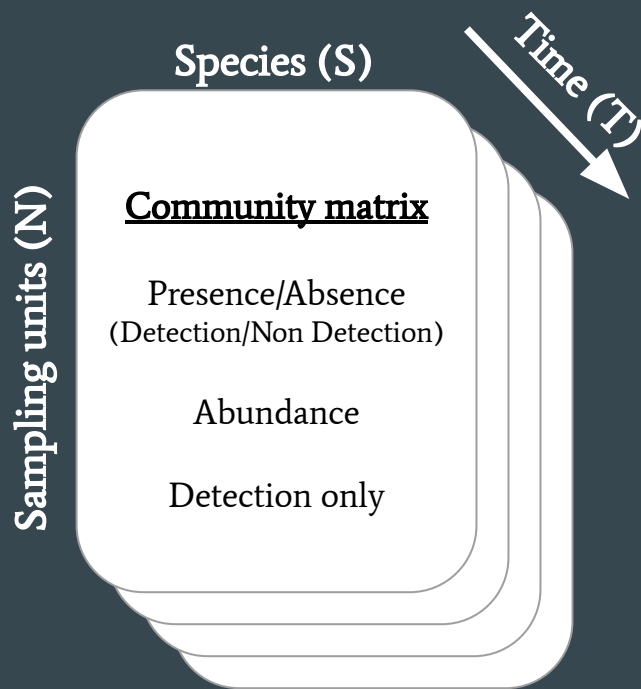


# The typical spatial dataset



# The typical spatial dataset

And surveys can be repeated in time



# The typical spatial dataset

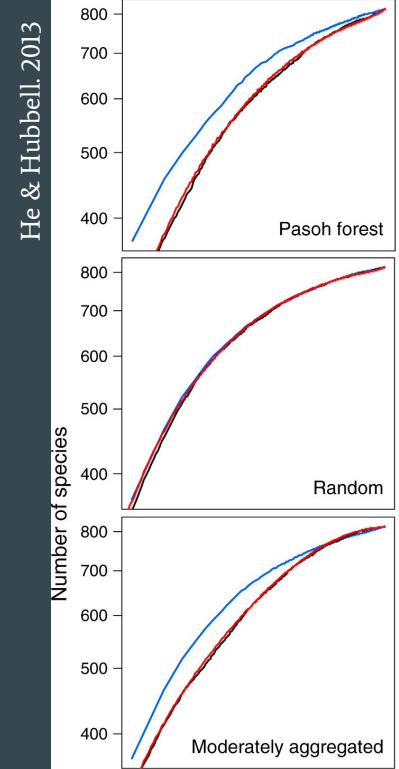
And surveys can be repeated in time

**Note:** having temporally-repeated can greatly improve things

- we can avoid assumptions about system equilibrium (species dyn. and/or envt/species relationships)
- it can help identify interactions among species
- we can more properly account for species detection/non detection (short-term repeats)

# The typical questions addressed

- Patterns of ecological diversity
  - Species-area relationships (SAR curves)
  - Distribution of species richness ( $\alpha$ ,  $\gamma$ )
  - How similar / dissimilar communities are? ( $\beta$ )



Pasoh Nature Reserve of Malaysia



0.2 0.5 2.0 5.0 20.0 50.0  
Area (ha)

# The typical questions addressed

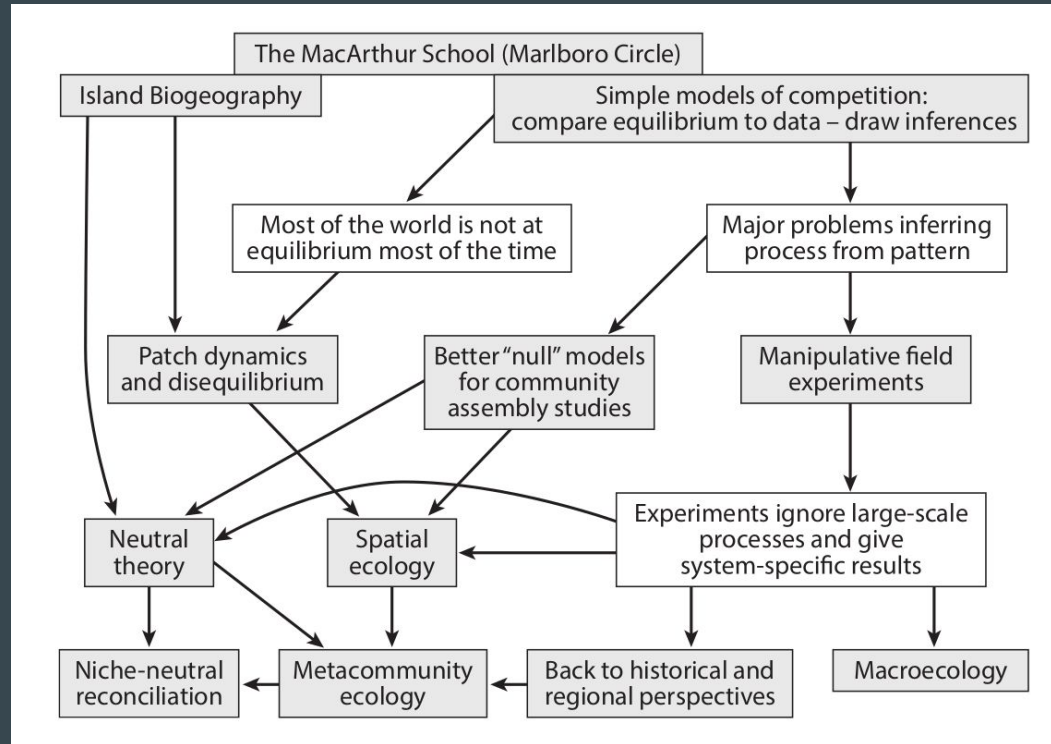
- Patterns of ecological diversity
  - Species-area relationships (SAR curves)
  - Distribution of species richnesses ( $\alpha$ ,  $\gamma$ )
  - How similar / dissimilar communities are? ( $\beta$ )
- Species distributions
  - Habitat preferences / specialization / generalism?
  - Are species interacting, and if so, how?
  - How important is dispersal versus habitat or competition?
  - How contingent is community assembly?
  - ... many more

# The typical questions addressed

- Patterns of ecological diversity
  - Species-area relationships (SAR curves)
  - Distribution of species richnesses ( $\alpha$ ,  $\gamma$ )
  - How similar / dissimilar communities are? ( $\beta$ )
- Species distributions
  - Habitat preferences / specialization / generalism?
  - Are species interacting, and if so, how?
  - How important is dispersal versus habitat or competition?
  - How contingent is community assembly?
  - ... many more
- **Theoretical frameworks to guide analyses?**

# Theoretical frameworks

- We'll make a long story short...



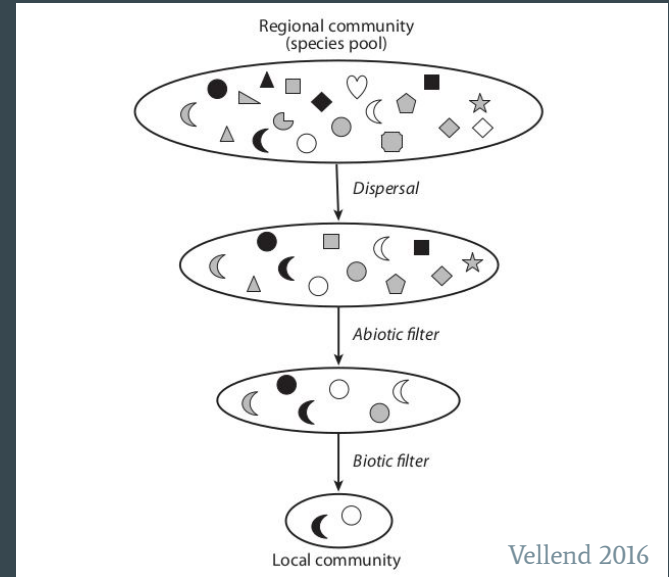
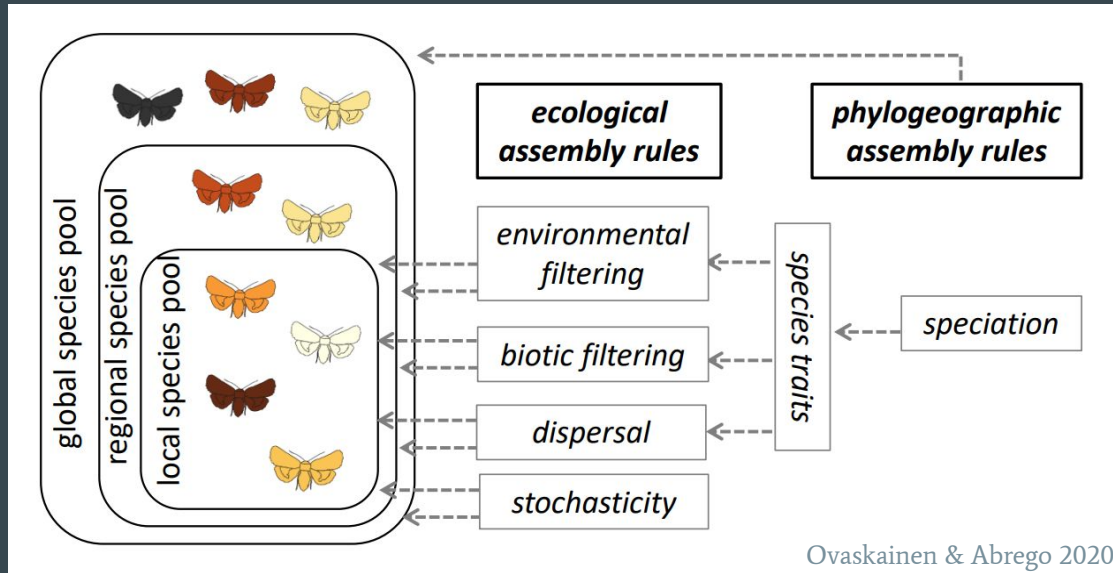


# Theoretical frameworks

A. The 'assembly rules' framework

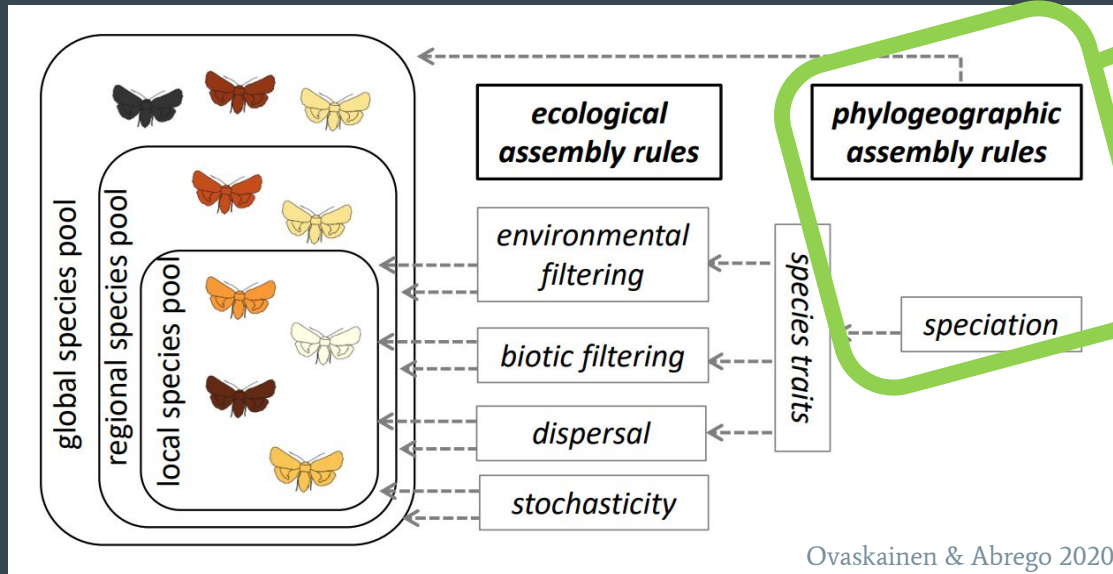
# Theoretical frameworks

## A. The 'assembly rules' framework



# Theoretical frameworks

## A. The 'assembly rules' framework



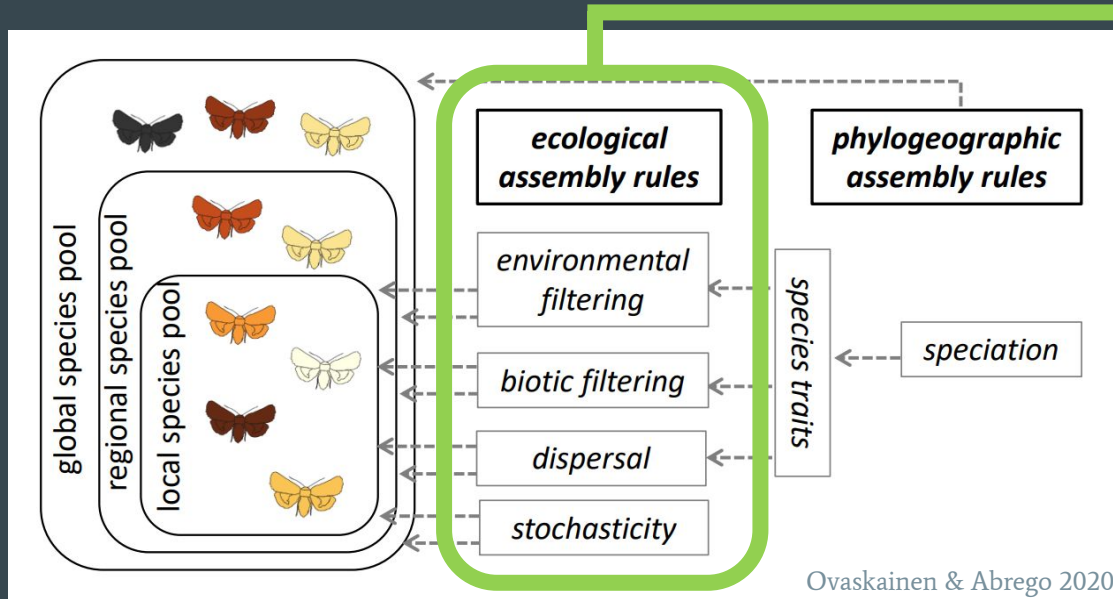
**Historical contingencies**  
(large spatio-temporal scales)

Large-scale migrations  
Geological events  
Speciation  
...

e.g. Great American Biotic Interchange  
e.g. endemic/invasive species

# Theoretical frameworks

## A. The 'assembly rules' framework



**Ecological filters**  
(smaller scales)

Environmental filtering  
Biotic filtering

Niche

Dispersal  
Stochasticity

# Theoretical frameworks

## A. The 'assembly rules' framework

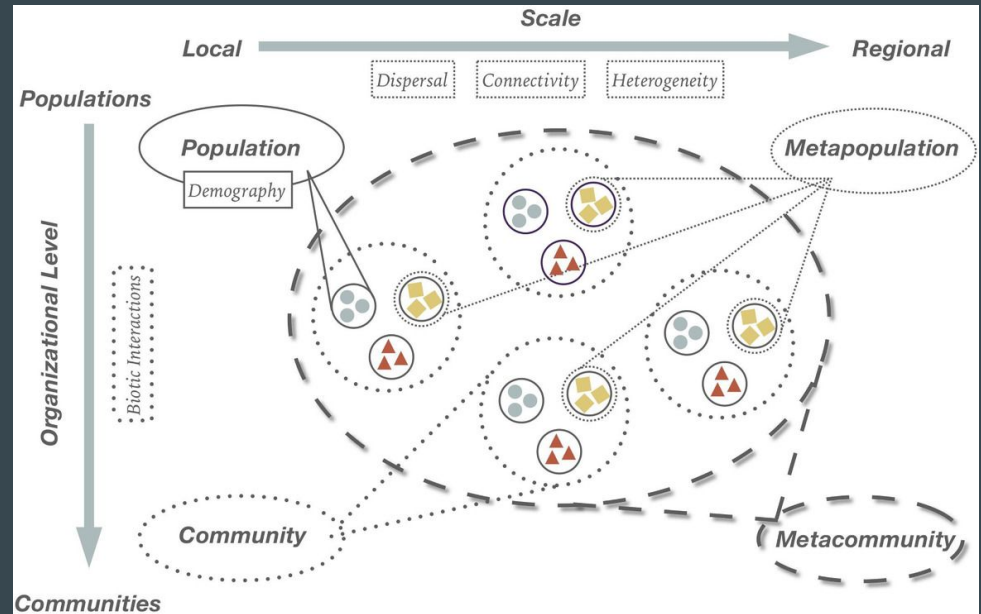
- Strengths
  - Very intuitive and broadly applicable
- Limitations
  - Mostly a static description, with a hierarchical set of filters
  - Focus on each individual site, not on the coupling between them

# Theoretical frameworks

## B. The metacommunity framework

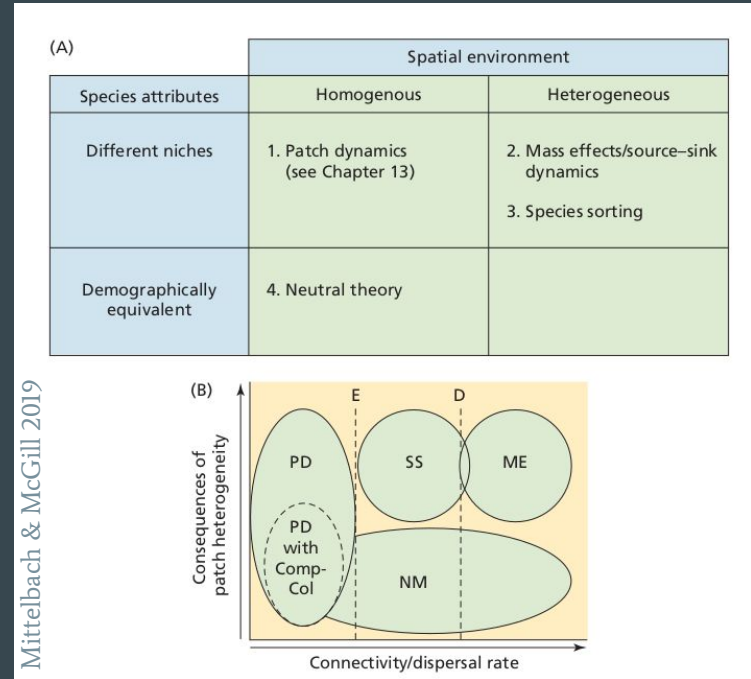
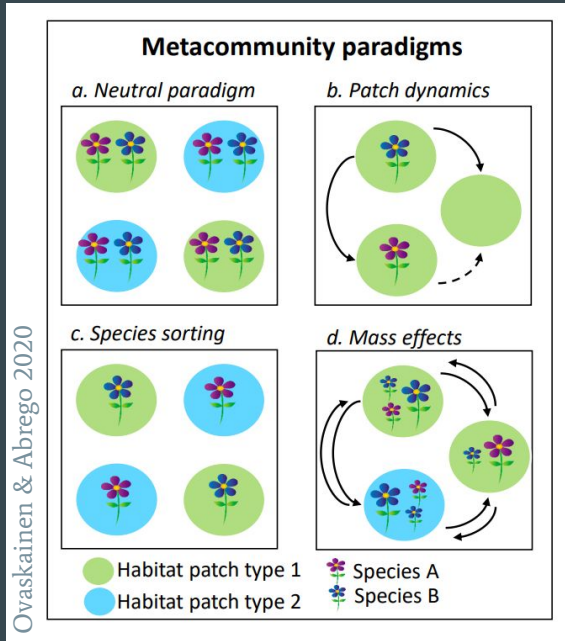
- A metacommunity is set of interconnected local communities
- Directly follows from the metapopulation paradigm

Chase et al 2020



# Theoretical frameworks

## B. The metacommunity framework



# Theoretical frameworks

## B. The metacommunity framework

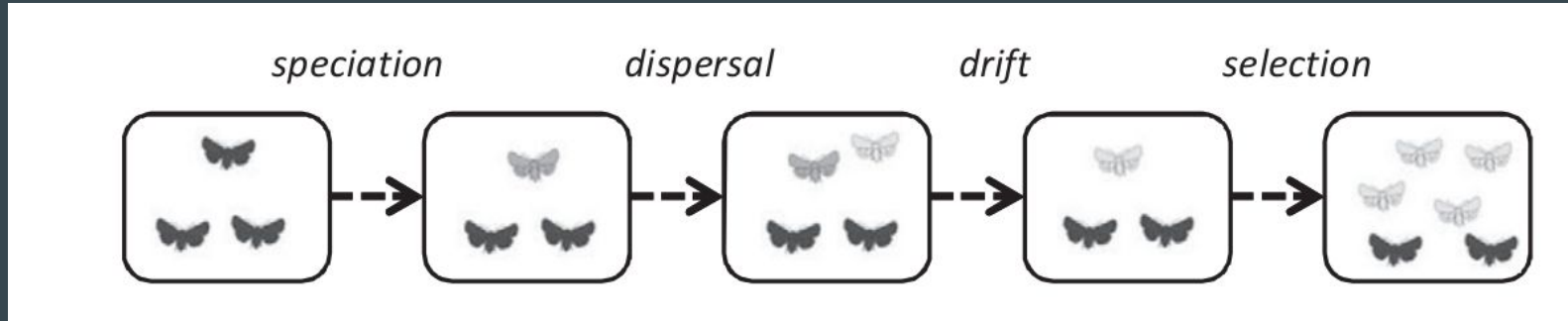
- Strengths
  - Emphasizes dynamical processes and the coupling of communities at different scales
  - Reunified niche, spatial and neutral perspectives
  
- Limitations
  - A collection of models that differ along many axes
  - The four paradigms are not mutually exclusive, and do not map to different processes. They are hard to disentangle



# Theoretical frameworks

## C. Vellends' theory of ecological communities

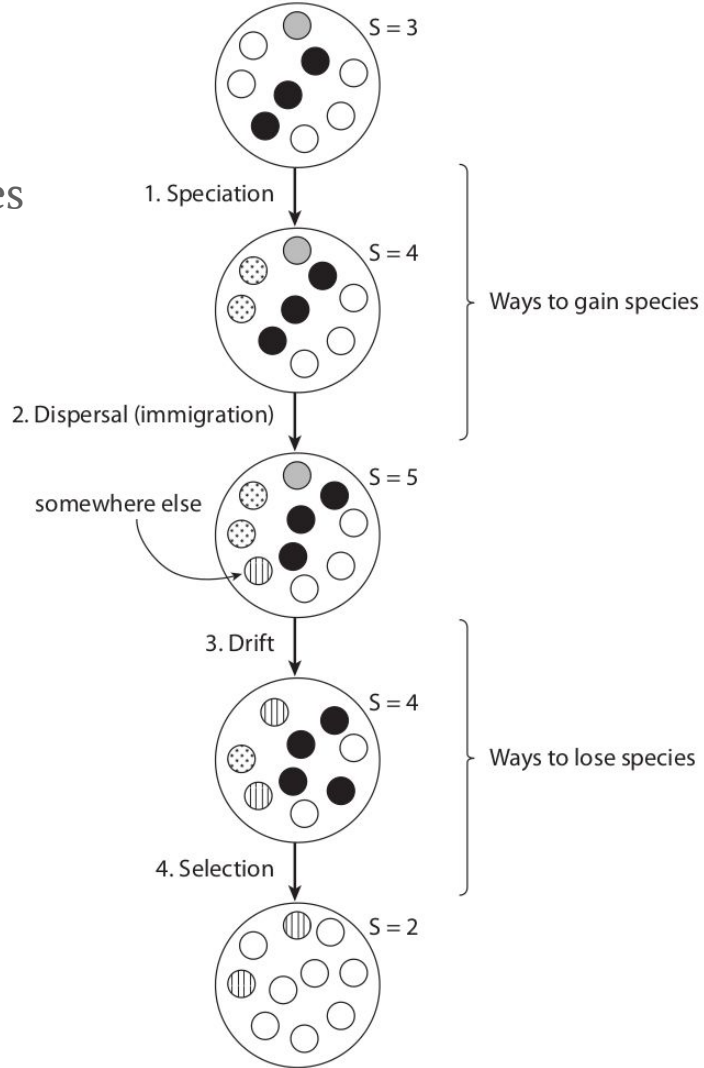
- Vellend's theory recycles the framework of population genetics (and its four evolutionary 'forces'):



Ovaskainen & Abrego 2020

# Theoretical frameworks

## C. Vellends' theory of ecological communities



# Theoretical frameworks

## C. Vellends' theory of ecological communities

- Strengths
  - Emphasizes true and distinct dynamical processes
  - Proximity with population genetics and evolutionary theory
- Limitations
  - Order of the processes is arbitrary.
  - Speciation probably acts on a different timescale (same issue with Neutral theory)
  - The theory is probably too abstract/generic to be very operational.

Go see

<https://reflectionsonpaperspast.wordpress.com/2020/03/15/revisiting-vellend-2010/>

# Theory-driven data analysis

So, in practice: what analyses?

- Part 1: descriptive approaches (more related to assembly rules and metacommunity frameworks)
- Part 2: process-based approaches (more related to metacommunity and Vellend's frameworks)

# Bridge intro / part 1

- questions on Intro



# Part 1. Descriptive approaches

- Statistical approaches (community level) - multivariate analyses (think PCA)

# Part 1. Descriptive approaches

- Statistical approaches (community level) - multivariate analyses (think PCA)
  - (Unconstrained) ordination methods (CA, PCoA, MDMS)
    - multivariate similarity among species / patches,
    - main axis can be (linearly or not) linked to environmental gradients or traits *a posteriori*
  - (Constrained) ordination methods (CCA, RDA, dbRDA, ...)
    - environmental data / spatial positions are incorporated to constrained the ordination
    - variance can be partitioned among multiple explicative tables (e.g. envt vs. space)

# Part 1. Descriptive approaches

- Statistical approaches (community level) - multivariate analyses (think PCA)
  - (Unconstrained) ordination methods (CA, PCoA, MDMS)
    - multivariate similarity among species / patches,
    - main axis can be (linearly or not) linked to environmental gradients or traits *a posteriori*
  - (Constrained) ordination methods (CCA, RDA, dbRDA, ...)
    - environmental data / spatial positions are incorporated to constrained the ordination
    - variance can be partitioned among multiple explicative tables (e.g. envt vs. space)

Based on the notion of a distances among sampling units / species.

Numerous (dis)similarity measures: Euclidean, Bray-Curtis, Chi Square or Euclidean dist. on transformed data (Profiles, Hellinger, ...)

Numerous ways of representing space: spatial neighborhood, spat. weighting matrices (SWM), spatial predictors (orth. polynomials of coordinates, PCNMs, MEMs, etc.)



# Part 1. Descriptive approaches

e.g. Guadeloupe Mollusks

*D. surinamense*



*M. cornuarietis* (1973)



*P. glauca*



*P. columella* (1997)

*M. tuberculata* (1979)

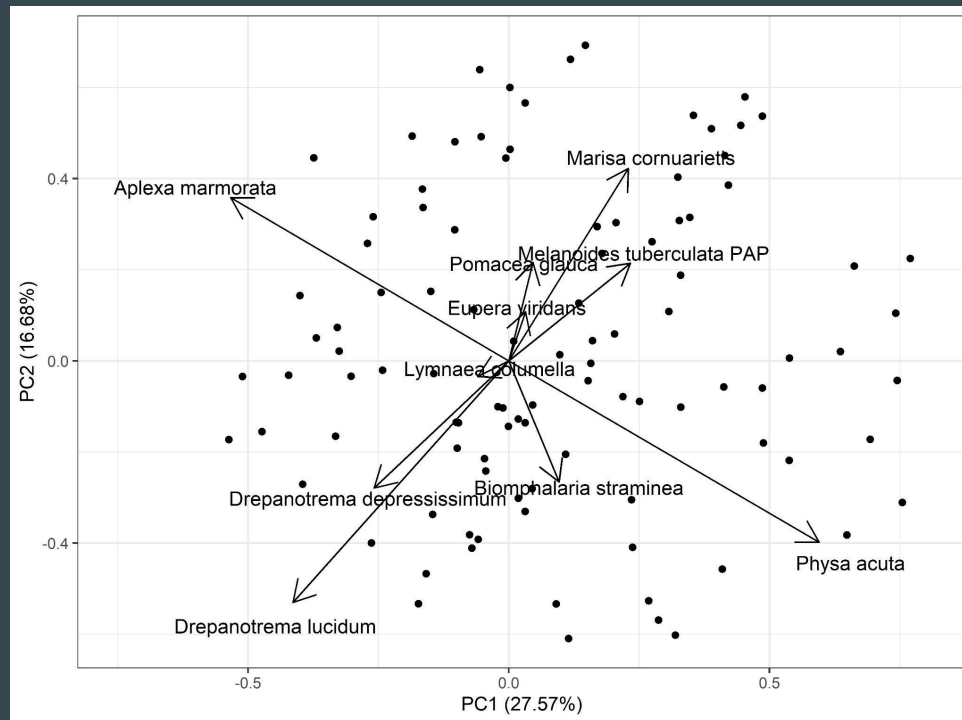


*P. acuta* (1972)



# Part 1. Descriptive approaches

e.g. Guadeloupe Mollusks: unconstrained (PCoA)

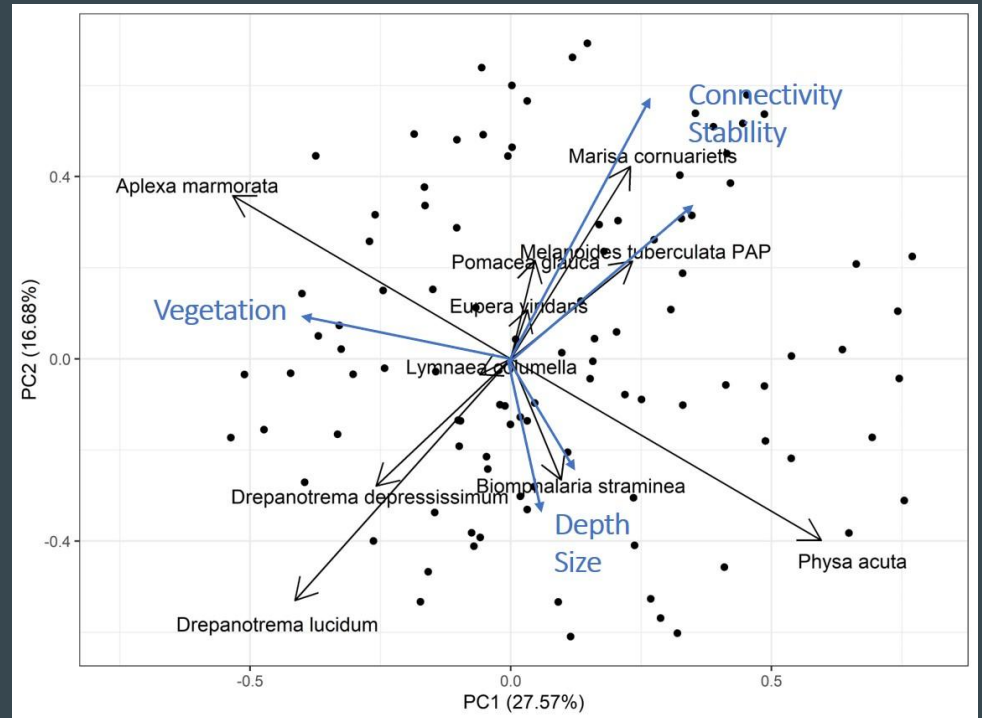


# Part 1. Descriptive approaches

e.g. Guadeloupe Mollusks: unconstrained (PCoA)

*A posteriori* environmental fit

	Dim1	Dim2	r <sup>2</sup>	Pr(>r)	
Vegetation	-0.97466	0.22371	0.1167	0.001	***
Size	0.15422	-0.98804	0.0698	0.001	***
Depth	0.39564	-0.91841	0.0521	0.009	**
Stability	0.71490	0.69923	0.1207	0.001	***
Connectivity	0.44110	0.89746	0.2272	0.001	***

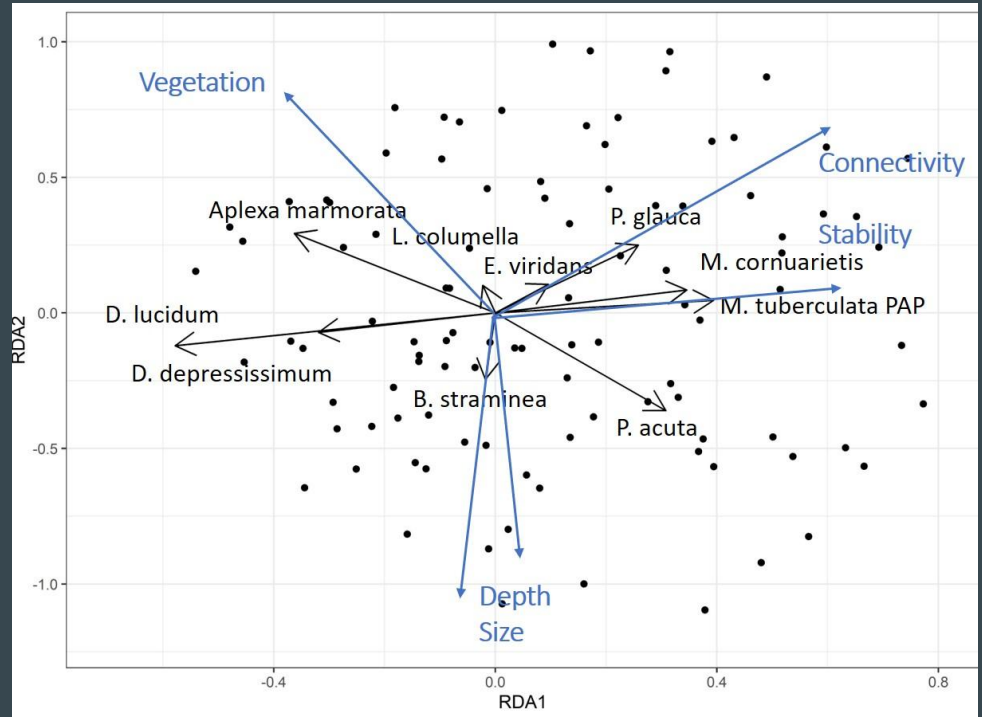


# Part 1. Descriptive approaches

e.g. Guadeloupe Mollusks: RDA (constrained)

	Inertia	Proportion	Rank
Total	0.5950	1.0000	
Constrained	0.0868	0.1459	5
Unconstrained	0.5082	0.8541	10

Inertia is variance



# Part 1. Descriptive approaches

e.g. Guadeloupe Mollusks: partial-RDA & **variance partitioning** with PCNMs

Widely used after Cottenie (2005) publication

the idea, partitioning variance between environmental and spatial predictors

as a way to distinguish whether communities result from niche-based or neutral-based processes...

*Ecology Letters*, (2005) **8**: 1175–1182      doi: 10.1111/j.1461-0248.2005.00820.x

---

**LETTER**

## Integrating environmental and spatial processes in ecological community dynamics

---

**Karl Cottenie**  
*National Center for Ecological Analysis and Synthesis,  
University of California, Santa Barbara, 725 State Street Suite*

**Abstract**  
The processes controlling the abundances of species across multiple sites form the cornerstone of modern ecology. In these metacommunities, the relative importance of local environmental and regional spatial processes is currently hotly debated, especially in terms of the validity of neutral model. I collected 158 published data sets with

# Part 1. Descriptive approaches

e.g. Guadeloupe Mollusks, partial-RDA & **variance partitioning** with PCNMs

Widely used after Cottenie (2005) publication

the idea, partitioning variance between environmental and spatial predictors

**Table 1** Decision tree for relationship between significance structure and metacommunity types

[E]	[S]	[E S]	[S E]	Metacommunity type
(not) sig.	(not) sig.	sig.	not sig.	SS
(not) sig.	(not) sig.	sig.	sig.	SS + ME
(not) sig.	(not) sig.	not sig.	sig.	NM/PD
(not) sig.	(not) sig.	not sig.	not sig.	Undetermined
not sig.	not sig.	not sig.	not sig.	No found

Relationship between significance structure of the four important variation components and associated metacommunity types. The components are environment [E], space [S], environment independent of space [E|S], and space independent of environment [S|E]. sig., variation component explains a significant part of the variation in community structure; not sig., no significant part; (not) sig., either.

Deeply rooted in the metacommunity framework

# Part 1. Descriptive approaches

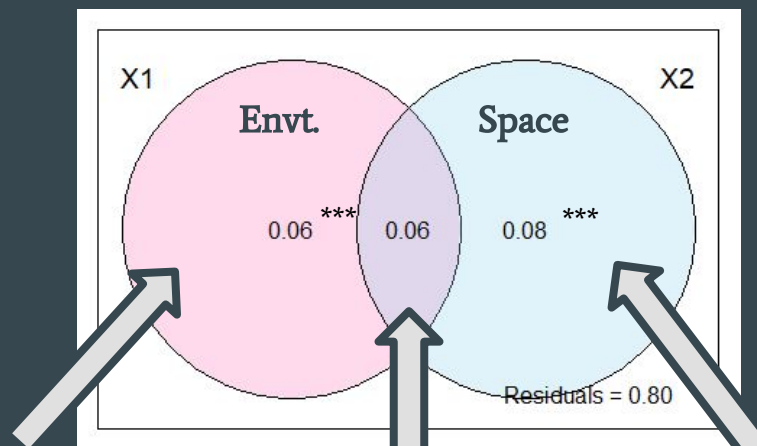
e.g. Guadeloupe Mollusks, partial-RDA & **variance partitioning** with PCNMs

Using Principal Coordinates of Neighbourhood Matrix (PCNM, Borcard & Legendre, 2002) as spatial descriptors

**Table 1** Decision tree for relationship between significance structure and metacommunity types

[E]	[S]	[E S]	[S E]	Metacommunity type
(not) sig.	(not) sig.	sig.	not sig.	SS
(not) sig.	(not) sig.	sig.	sig.	SS + ME
(not) sig.	(not) sig.	not sig.	sig.	INM/PD
(not) sig.	(not) sig.	not sig.	not sig.	Undetermined
not sig.	not sig.	not sig.	not sig.	No found

Relationship between significance structure of the four important variation components and associated metacommunity types. The components are environment [E], space [S], environment independent of space [E|S], and space independent of environment [S|E]. sig., variation component explains a significant part of the variation in community structure; not sig., no significant part; (not) sig., either.



Envt. independent of space

Spatially autocorrelated envt. gradient

Space independent of envt.

# Part 1. Descriptive approaches

- Statistical approaches (community level) - multivariate analyses
  - Among distance matrices regressions
    - Mantel and partial Mantel tests
    - MRM (Lichstein, 2007)

Might be better suited for testing some theory (e.g. Neutral Theory),  
See Tuomisto & Ruokolainen, 2006.

But ... see also Tuomisto, Ruokolainen & Ruokolainen, 2012.





# Part 1. Descriptive approaches

- Statistical approaches (species level)

**Rationale:** species that do not interact and have the same habitat affiliations should be distributed **independently** over sites\*.

\*Think it through

# Part 1. Descriptive approaches

- Statistical approaches (species level)
  - Fitting null distributions, pairwise tests of species independence (Forbes 1907, Veech et al . 2013)

	species 1 present	species 1 absent	
species 2 present	a	c	Number of sites
species 2 absent	b	d	

# Part 1. Descriptive approaches

- Statistical approaches (species level)
  - Fitting null distributions, pairwise tests of species independence (Forbes 1907, Veech et al . 2013)

	species 1 present	species 1 absent
species 2 present	a	c
species 2 absent	b	d

Number of sites

Arita (2016)

$$P(a) = \frac{\binom{a+b}{a} \binom{c+d}{c}}{\binom{N}{a+c}}$$

Positive association

No association

Negative association

Expected number of co-occurrences (if independent)

# Part 1. Descriptive approaches

- Statistical approaches (species level)
  - Fitting null distributions, pairwise tests of species independence (Forbes 1907, Veech et al . 2013)
  - Permutation based (null model) approaches (Diamond, Simberloff, Gotelli, Ulrich)

# Part 1. Descriptive approaches

- Statistical approaches (species level)
  - Fitting null distributions, pairwise tests of species independence (Forbes 1907, Veech et al . 2013)
  - Permutation based (null model) approaches (Diamond, Simberloff, Gotelli, Ulrich)

The C-score:

$$C_{ij} = (r_i - S_{ij})(r_j - S_{ij})$$

Number of sites with  
species i

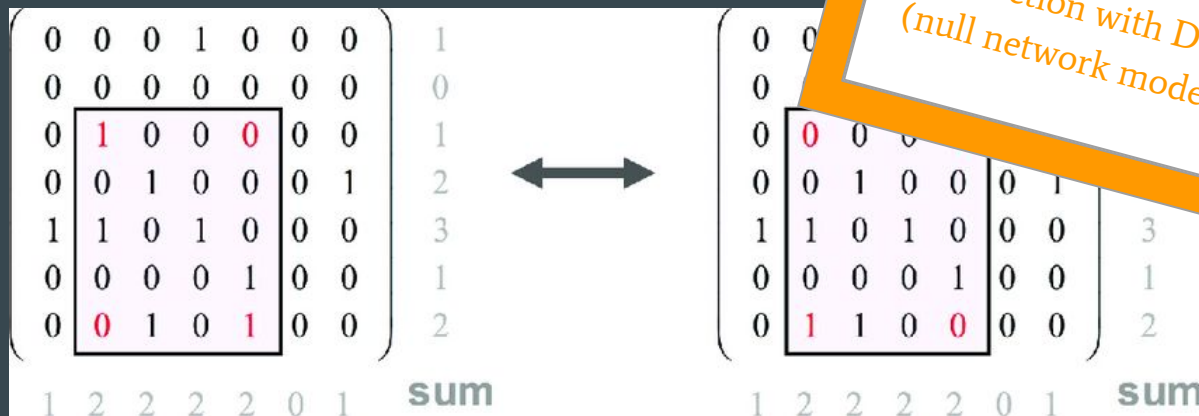
Number of sites with  
species j

Number of sites with  
both species

# Part 1. Descriptive approaches

- Statistical approaches (species level)
  - Fitting null distributions, pairwise tests of species independence (Forbes 1907, Veech et al . 2013)
  - Permutation based (null model) approaches (Diamond, Simberloff, Gotelli, Ulrich)

Reshuffle the matrix many times (fixed-fixed swap algorithm):

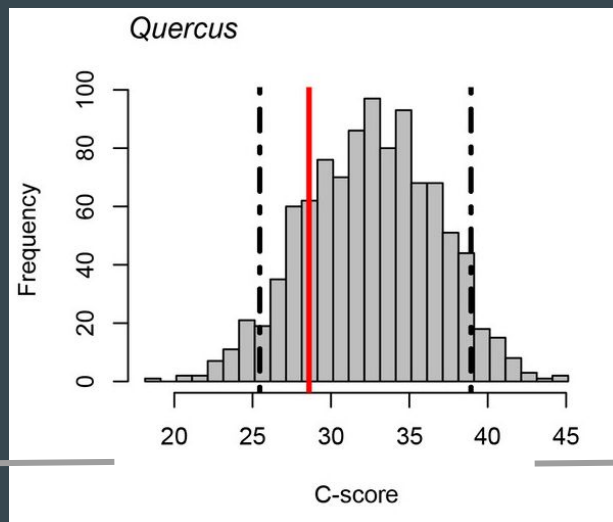


# Part 1. Descriptive approaches

- Statistical approaches (species level)
  - Fitting null distributions, pairwise tests of species independence (Forbes 1907, Veech et al . 2013)
  - Permutation based (null model) approaches (Diamond, Simberloff, Gotelli, Ulrich)

Generate null distribution  
for  $C_{ij}$  and decide  
significance and  
standardized effect size:

Positive  
association ←



→ Negative  
association

# Part 1. Descriptive approaches

- Statistical approaches (species level)
  - Fitting null distributions, pairwise tests of species independence (Forbes 1907, Veech et al . 2013)
  - Permutation based (null model) approaches (Diamond, Simberloff, Gotelli, Ulrich)
  - Extension to constrained null models by Peres Neto et al (2001)

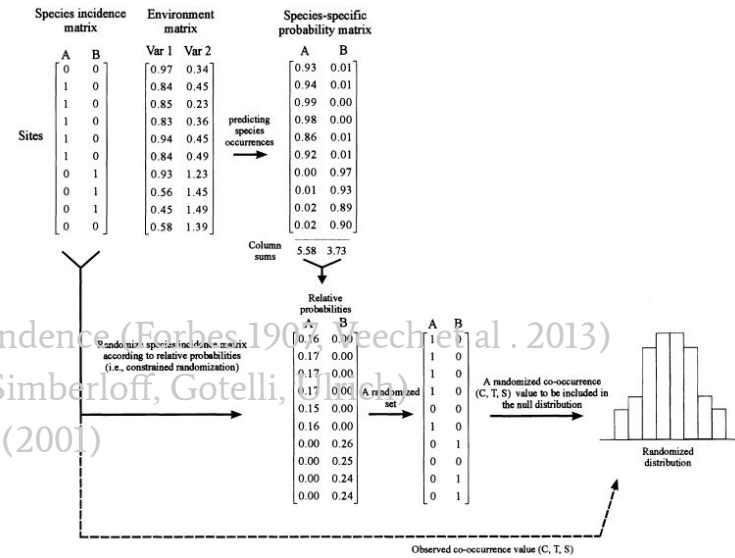


# Part 1. Descriptive approaches

- Statistical approaches (species level)

- Fitting null distributions, pairwise tests of species independence (Forbes 1900, Beecher et al. 2013)
- Permutation based (null model) approaches (Diamond, Simberloff, Gotelli, Hubbell)
- Extension to constrained null models by Peres Neto et al (2001)

Peres Neto et al., 2001





# Part 1. Descriptive approaches

- Statistical approaches (species level)
  - (Joint) Species Distribution Modelling
  - Occupancy models (emphasis on detection process)

# Part 1. Descriptive approaches

- Statistical approaches (species level): Species Distribution Models (SDM)
  - From a theoretical viewpoint : they catch the ‘realized niche’ of one species through regression or a broad range of statistical/classification methods
  - **Main aim : making predictions - on past, contemporary, and future sp. distributions**
  - Have been criticized regarding the lack of consideration for :
    - equilibrium assumption (species are wherever they can) (Araújo & Pearson, 2005)
    - dispersal, dispersal limitation in particular
    - species interactions (but see Anderson, 2017, JoB)
    - but also sampling, methods, model transferability...

# Part 1. Descriptive approaches

- Statistical approaches (species level): Species Distribution Models (SDM)

A lot of methods...

Table 3.1 *Summary of some popular and recently emerged SDM frameworks used to model community data. The SDM frameworks are classified either as single-species distribution models or joint species distribution models.*

<b>Single-species distribution models</b>	<b>Reference</b>
Boosted regression trees (BRT)	Hijmans et al. (2017); Ridgeway (2017)
Generalised additive model (GAM)	Wood (2011)
Generalised linear model (GLM)	R Development Core Team (2019)
Gradient nearest neighbour (GNN)	Crookston & Finley (2008)
Maximum-entropy approach (MaxEnt)	Phillips et al. (2006)
Multivariate adaptive regression spline (MARS-COMM)	Milborrow (2017)
Multivariate regression tree (MRTS)	De'ath et al. (2014)
Random forest (RF)	Liaw & Wiener (2002)
Support vector machine (SVM)	Meyer et al. (2017)
Gradient extreme boosting (XGB)	Chen et al. (2018)

From Ovaskainen & Abrego, 2020

# Part 1. Descriptive approaches

- Statistical approaches (species level): Species Distribution Models (SDM)

With several species:

- One SDM per species, then combine → **Stacked** SDMs
- SDM for all species at the same time → **Joint** SDMs (JSDMs)

Think: several point estimates of means *versus* one ANOVA

# Part 1. Descriptive approaches

- Statistical approaches (species level): Species Distribution Models (SDM)

Even more methods...

<b>Joint species distribution models</b>	
Bayesian community ecology analysis (BC)	Golding & Harris (2015)
Bayesian ordination and regression analysis (BORAL)	Hui (2017)
Generalised joint attribute modelling (GJAM)	Clark et al. (2017)
Hierarchical modelling of species communities (HMSC)	Ovaskainen et al. (2017b)
Multivariate stochastic neural network (MISTN)	Harris (2015)
Species archetype model (SAM)	Hui et al. (2013)

From Ovaskainen & Abrego, 2020

# Part 1. Descriptive approaches

- JSDBMs: one relatively simple and flexible approach
  - **Hierarchical Models of Species Composition (HMSC)** - Ovaskainen et al. (2017)
    - Bayesian framework
    - Several latent factors (random effects)
    - Species occupancies + habitat variables + species traits + Phylogeny + ...



# Part 1. Descriptive approaches

- JSDBMs: one relatively simple and flexible approach
  - **Hierarchical Models of Species Composition (HMSC)** - Ovaskainen et al. (2017)

Nice R package, nice book,

## Methods in Ecology and Evolution

APPLICATION |  Open Access |   

### Joint species distribution modelling with the R-package HMSC

Gleb Tikhonov, Øystein H. Opedal, Nerea Abrego, Aleksí Lehtikoinen, Melinda M. J. de Jonge, Jari Oksanen, Otso Ovaskainen 

First published: 25 December 2019 | <https://doi.org/10.1111/2041-210X.13345> | Citations: 5

New Results

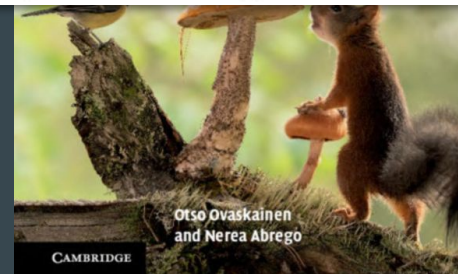
 Follow this preprint

### Accelerating joint species distribution modeling with Hmsc-HPC: A 1000x faster GPU deployment

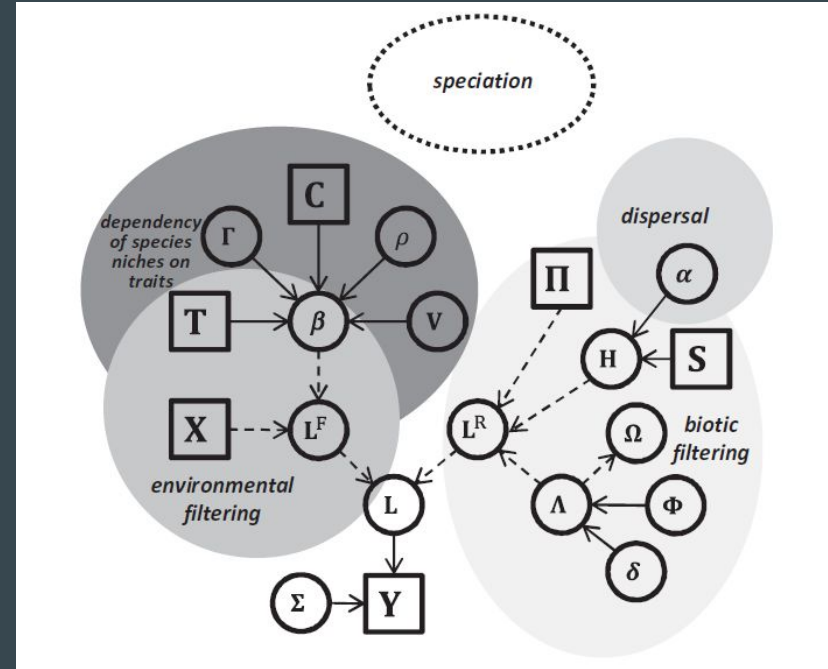
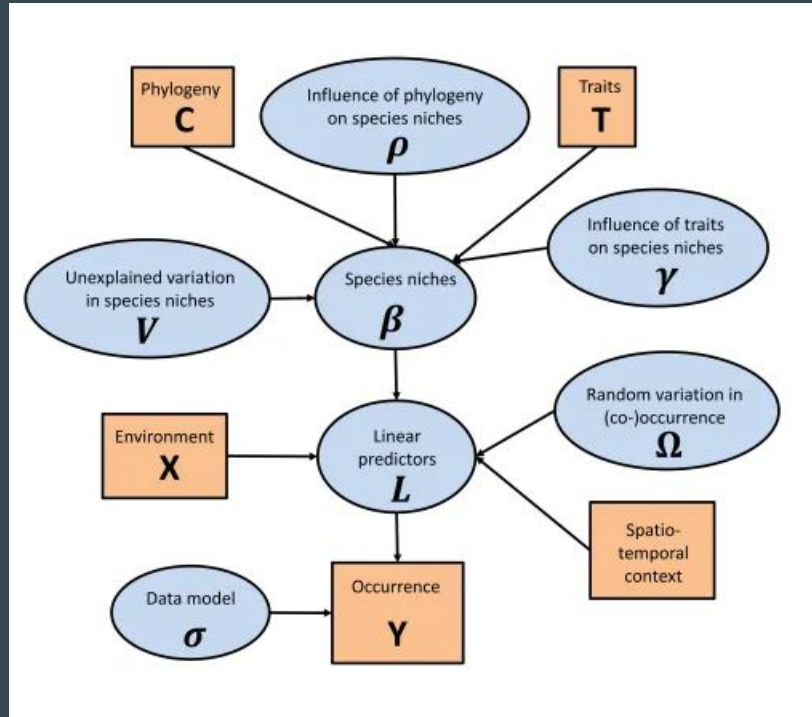
 Anis Ur Rahman, Gleb Tikhonov, Jari Oksanen, Tuomas Rossi, Otso Ovaskainen

doi: <https://doi.org/10.1101/2024.02.13.580046>

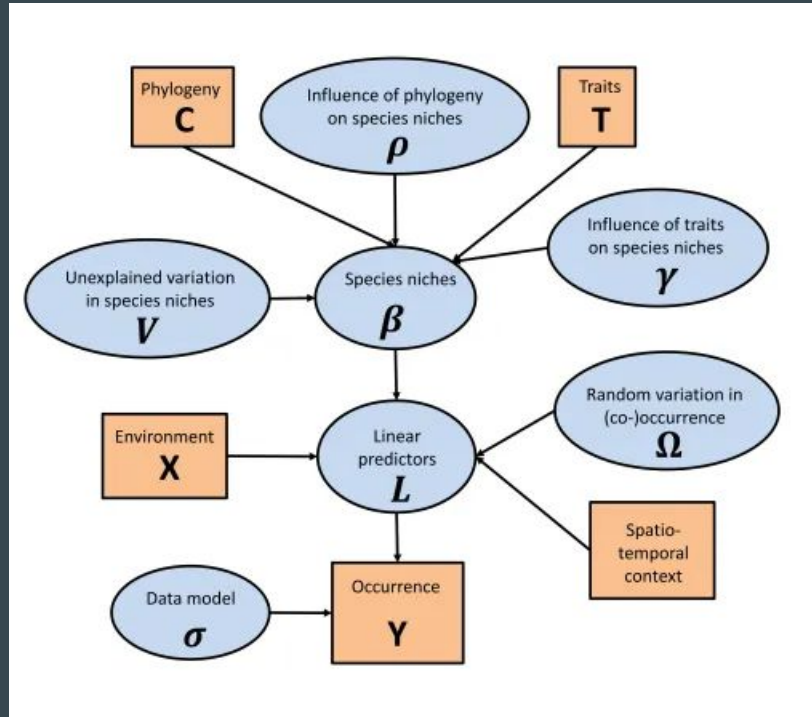
This article is a preprint and has not been certified by peer review [what does this mean?].



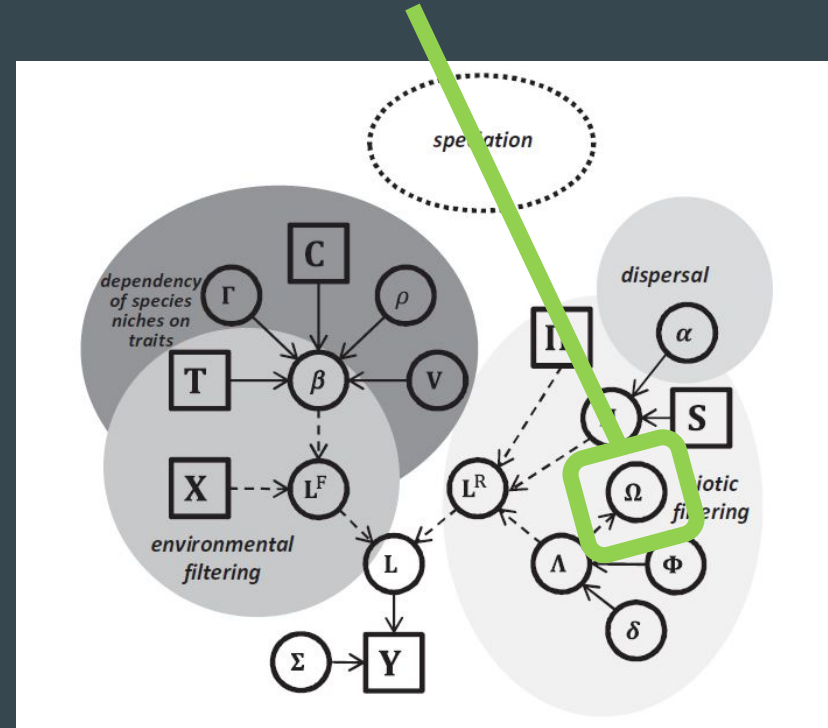
# Part 1. Descriptive approaches



# Part 1. Descriptive approaches



Species interactions must be deduced from **residual covariances**\*



\*Same rationale as for null-model approach earlier

# Part 1. Descriptive approaches

- An example of HMSC application: fungal communities of decaying trees

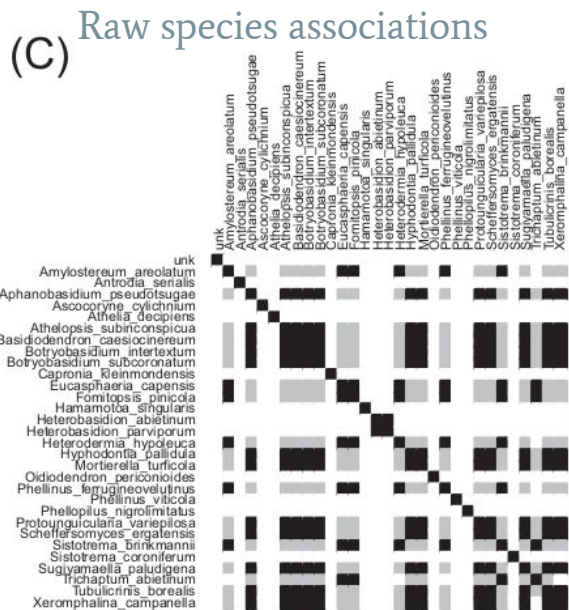


*Botryobasidium subcoronatum* (@Wikipedia)

# Part 1. Descriptive approaches

- An example of HMSC application: fungal communities of decaying trees

From Ovaskainen & Abrego, 2020

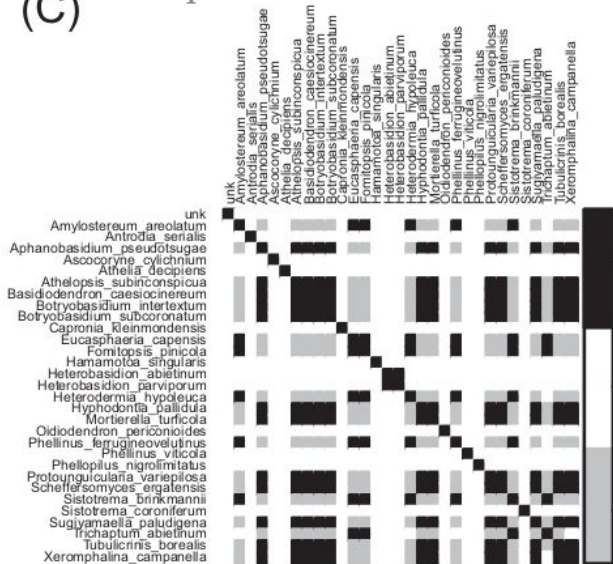


*Botryobasidium subcoronatum* (@Wikipedia)

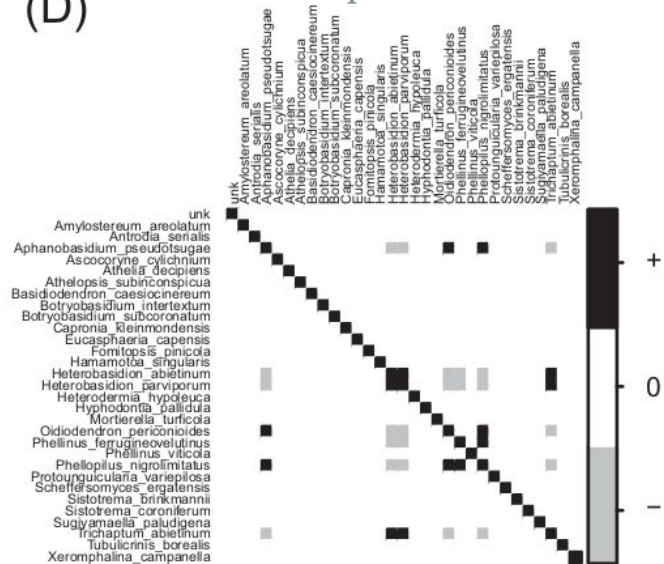
# Part 1. Descriptive approaches

- An example of HMSC application: fungal communities of decaying trees

(C) Raw species associations



(D) Residual species associations



# Part 1. Descriptive approaches

- Statistical approaches (species level): Occupancy models



Ecography 40: 281–295, 2017

doi: 10.1111/ecog.02445

© 2016 The Author. Ecography © 2016 Nordic Society Oikos

Subject Editor: Miguel Araújo. Editor-in-Chief: Miguel Araújo. Accepted 15 June 2016

## Modelling of species distributions, range dynamics and communities under imperfect detection: advances, challenges and opportunities

Gurutzeta Guillera-Arroita

*G. Guillera-Arroita (gurutzeta.guillera@unimelb.edu.au), School of Biosciences, Univ. of Melbourne, Australia.*

Building useful models of species distributions requires attention to several important issues, one being imperfect detection of species. Data sets of species detections are likely to suffer from false absence records. Depending on the type of survey, false positive records can also be a problem. Disregarding these observation errors may lead to important biases in model estimation as well as overconfidence about precision. The severity of the problem depends on the intensity of these errors and how they correlate with environmental characteristics (e.g. where species detectability strongly depends on habitat

# Part 1. Descriptive approaches

- Statistical approaches (species level): Occupancy models

Close to SDMs, but account for imperfect species detection (non-detection  $\neq$  absence)

The idea: distinguish *true* occupancy states from *observed* occupancy states,  
and estimate a **detection probability** which allows to account for false negatives

Use short-term survey repeats (short enough to assume no change in occupancy state)



# Part 1. Descriptive approaches

- Statistical approaches (species level): Occupancy models

Close to SDMs, but account for imperfect species detection (non-detection  $\neq$  absence)

The idea: distinguish *true* occupancy states from *observed* occupancy states, and estimate a **detection probability** which allows to account for false negatives

*Bufo americanus*



Andrew Hoffman, 2007

Initial survey: detected in *ca.* 30% of sites

Considering detection probability:  
probable presence in *ca.* 50% (+44%) of sites

# Part 1. Descriptive approaches

- Statistical approaches (species level): Occupancy models

Large number of (more mechanistic\*) extensions based on occupancy modelling ...

Metapopulation dynamics

Environmental covariates

Multiple species

Explicit dispersal

...

**\*see Part 2 “Patch occupancy models”.**

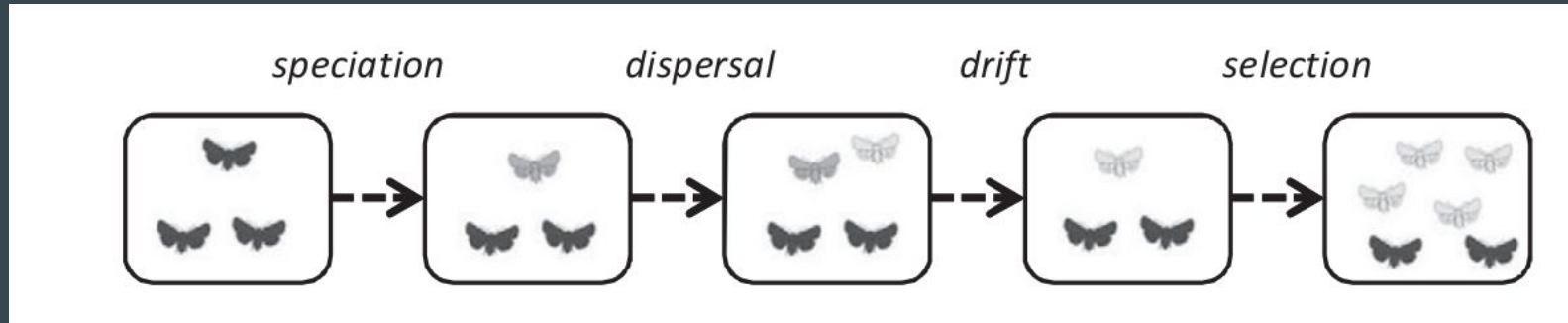
# Bridge part 1 / part 2

- Coffee break / questions on Part 1



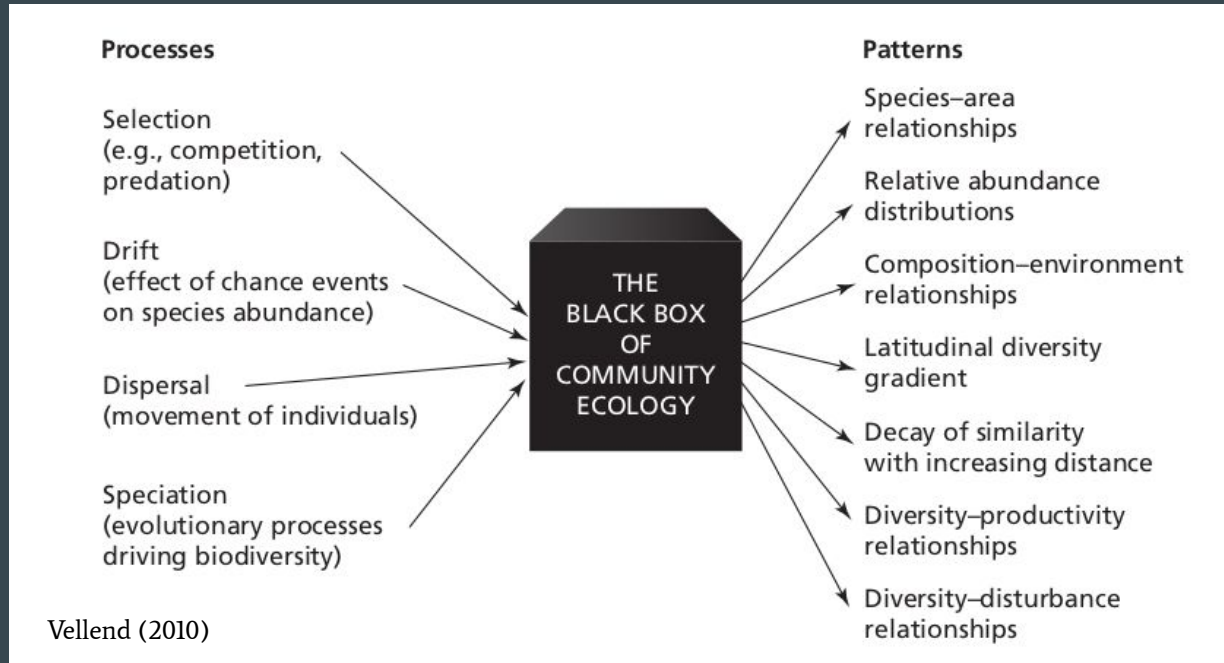
## Part 2. Process-based approaches

- Describing the processes and how they operate in space and time



# Part 2. Process-based approaches

- Describing the processes and how they operate in space and time



# Part 2. Process-based approaches

- Describing the processes and how they operate in space and time

Some mathematical  
(dynamical) model

## Processes

Selection  
(e.g., competition,  
predation)

Drift  
(effect of chance events  
on species abundance)

Dispersal  
(movement of individuals)

Speciation  
(evolutionary processes  
driving biodiversity)



## Patterns

Species–area  
relationships

Relative abundance  
distributions

Composition–environment  
relationships

Latitudinal diversity  
gradient

Decay of similarity  
with increasing distance

Diversity–productivity  
relationships

Diversity–disturbance  
relationships

Vellend (2010)

# Part 2. Process-based approaches

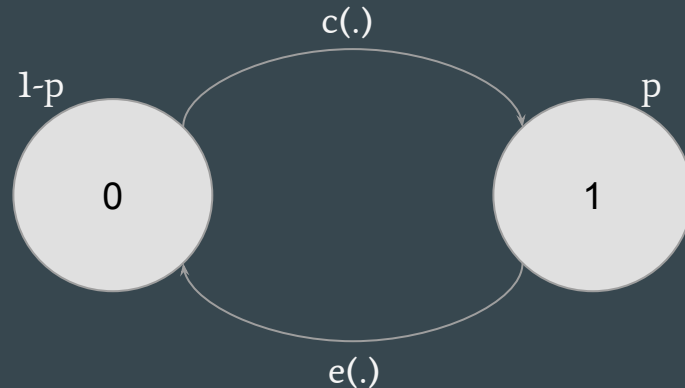
- A. Patch occupancy models

In one site (patch), a species is either present ( $p$ ) or absent ( $1-p$ )

Local dynamics can be neglected.

The probability of presence is a dynamic equilibrium between:

- The rate of colonization ( $c$ )
- The rate of extinction ( $e$ )



# Part 2. Process-based approaches

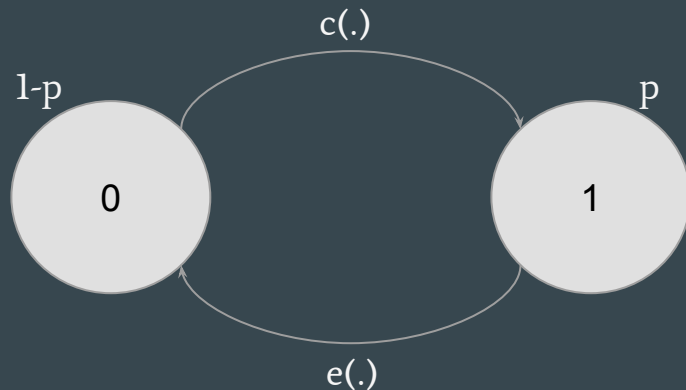
- A. Patch occupancy models

In one site (patch), a species is either present ( $p$ ) or absent ( $1-p$ )

Local dynamics can be neglected.

The probability of presence is a dynamic equilibrium between:

- The rate of colonization ( $c$ )
- The rate of extinction ( $e$ )



$$\frac{dp}{dt} = c(.) (1 - p) - e(.) p$$



# Part 2. Process-based approaches

- A. Patch occupancy models

$$\frac{dp}{dt} = c(\cdot)(1 - p) - e(\cdot)p$$

Now if you assume that:

- The extinction rate  $e(\cdot)$  is just a constant:  $e(\cdot) = e$
- The colonization rate is proportional to the overall occupancy, which, if all patches are homogeneous, is just  $p$ :  $c(\cdot) = c p$

you get

$$\frac{dp}{dt} = cp(1 - p) - ep$$

Levin's (1969) metapopulation model

$$p^* = 1 - \frac{e}{c}$$

Equilibrium metapopulation occupancy

# Part 2. Process-based approaches

- A. Patch occupancy models

$$\frac{dp}{dt} = c(\cdot)(1 - p) - e(\cdot)p$$

Now if you assume that:

- The extinction rate  $e(\cdot)$  is just a constant:  $e(\cdot) = e$
- The colonization rate is proportional to the overall occupancy, which, if all patches are homogeneous, is just  $p$ :  $c(\cdot) = cp$

you get

$$\frac{dp}{dt} = cp(1 - p) - ep$$

Levin's (1969) metapopulation model

$$p^* = 1 - \frac{e}{c}$$

Equilibrium metapopulation occupancy

\*Note: you have seen this model yesterday. Noticed?

# Part 2. Process-based approaches

- A. Patch occupancy models

$$\frac{dp}{dt} = c(\cdot)(1 - p) - e(\cdot)p$$

Now if you assume that:

- The extinction rate  $e(\cdot)$  is just a constant:  $e(\cdot) = e$
- The colonization rate is also just a constant:  $c(\cdot) = c$

you get

$$\frac{dp}{dt} = c(1 - p) - ep$$

McArthur & Wilson's (1967) continent-island model

$$p^* = \frac{c}{c + e}$$

Equilibrium island occupancy

# Part 2. Process-based approaches

- A. Patch occupancy models

Levins' (1969) formed the basis of metapopulation theory

MW's (1967) model formed the basis of the theory of island biogeography (TIB)

In both cases:

**Colonization** should depend on isolation (**distance to mainland** or **patch isolation**):  
nearer islands/more connected patches are more likely to receive migrants

**Extinction** should depend on island or patch **size** (area):  
larger islands/patches, having larger population sizes, have lower extinction risk

# Part 2. Process-based approaches

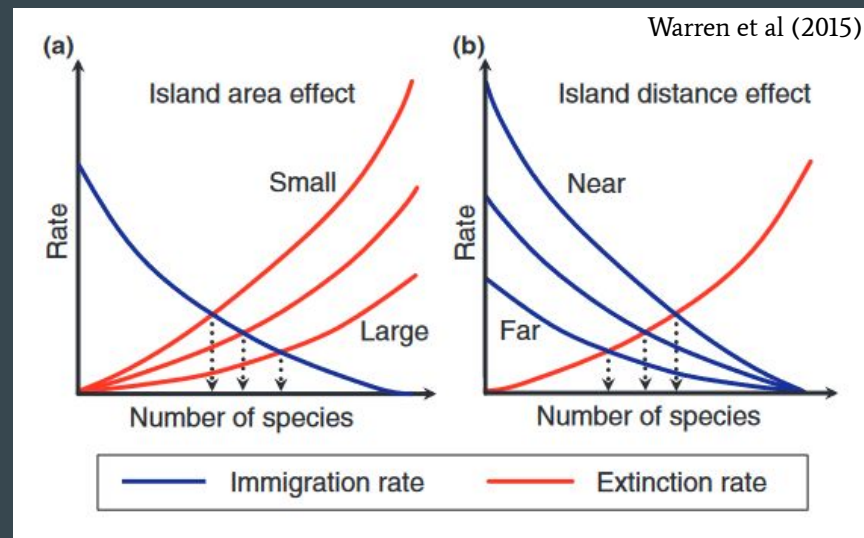
- A. Patch occupancy models

With several species, we have one (independent) equation per species, and we must sum these equations to get the dynamics of the **number** of species.

Take the MW's model.

If all species were similar and non interacting, we'd get **linear** changes in the rates of extinction and colonizations with species richness

**In practice**, we rather expect\* **convex** functions:  
– colonization decreases slower than linearly  
– extinction increases faster than linearly



\*Why?

## Part 2. Process-based approaches

- A. Patch occupancy models

We can **adjust** the MW model to time series of colonizations and extinctions events on islands to test for this prediction: the example of Manne et al. (1998) JAE

The data: presence/absence of breeding birds on 13 British islands, followed over consecutive years (as many as 20 years)

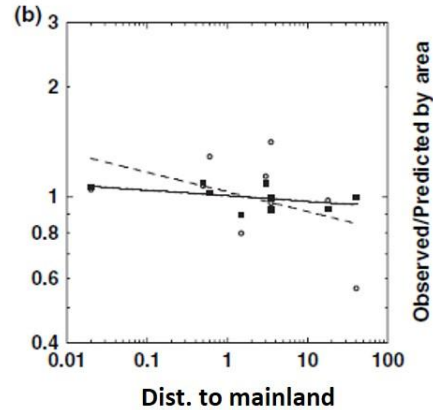
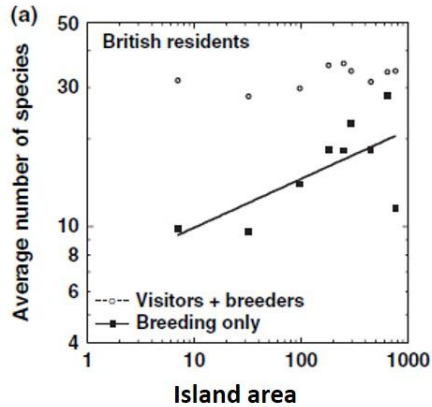
# Manne et al. 1998



Oystercatcher pair @wikipedia



The location of the nine islands in this study. Map modified from *J.* (2006).



# Manne et al. 1998

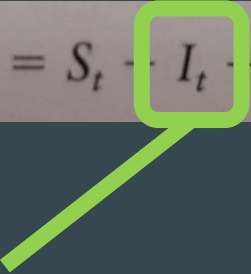
The number of species on a particular island at time  $t$  ( $S_t$ ) is known exactly, and modelled as

$$S_{t+1} = S_t + I_t - E_t$$



# Manne et al. 1998

The number of species on a particular island at time  $t$  ( $S_t$ ) is known exactly, and modelled as

$$S_{t+1} = S_t + I_t - E_t$$


Number of immigration events.  
Probability:

$$\gamma_t = \beta_1(1 - S_t/P)^{\beta_2}$$

# Manne et al. 1998

The number of species on a particular island at time  $t$  ( $S_t$ ) is known exactly, and modelled as

$$S_{t+1} = S_t + I_t - E_t$$

Number of immigration events.  
Probability:

$$\gamma_t = \beta_1(1 - S_t/P)^{\beta_2}$$

Number of extinction events.  
Probability:

$$\theta_t = \beta_3(S_t/P)^{\beta_4}$$

# Manne et al. 1998

The number of species on a particular island at time  $t$  ( $S_t$ ) is known exactly, and modelled as

$$S_{t+1} = S_t + I_t - E_t$$

Number of immigration events.  
Probability:

$$\gamma_t = \beta_1(1 - S_t/P)^{\beta_2}$$

Number of extinction events.  
Probability:

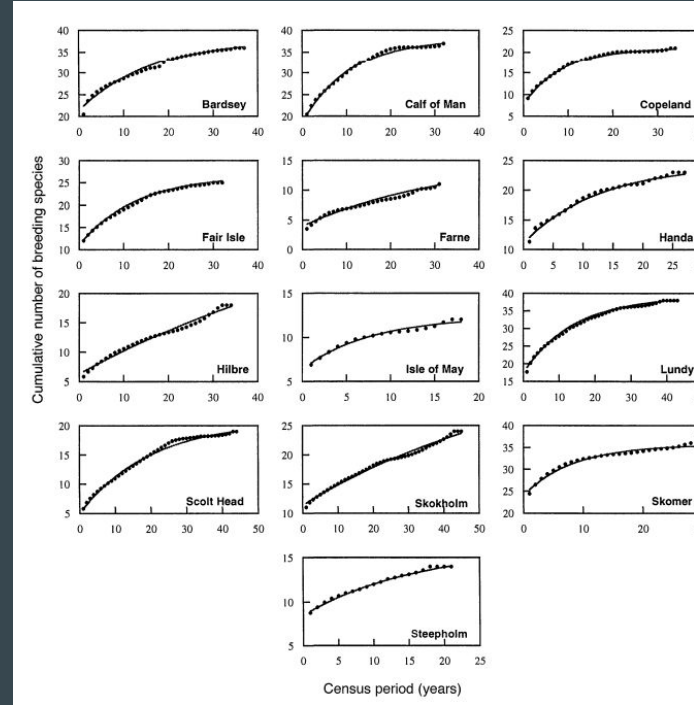
$$\theta_t = \beta_3(S_t/P)^{\beta_4}$$

Number of species  
in regional pool

# Manne et al. 1998

From there, we need to:

1. Estimate the number of species in the regional pool(  $P$  )

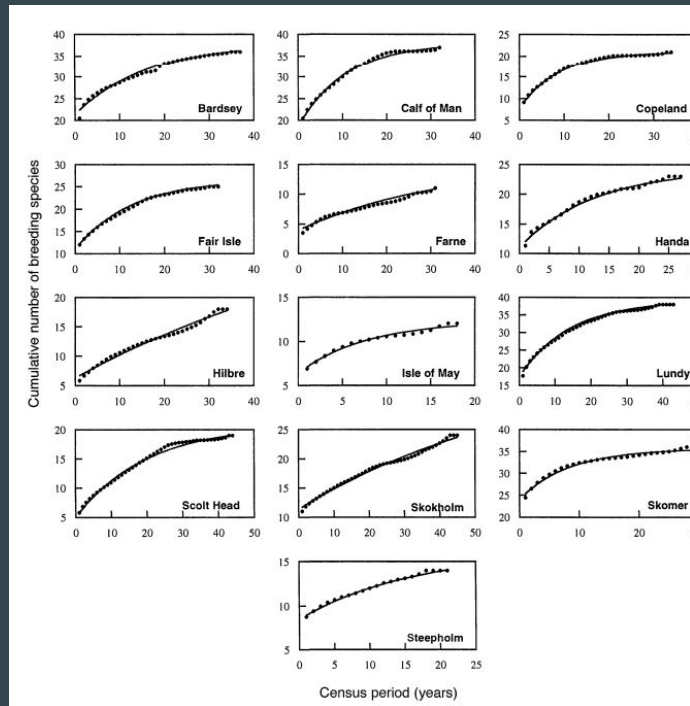


# Manne et al. 1998

From there, we need to:

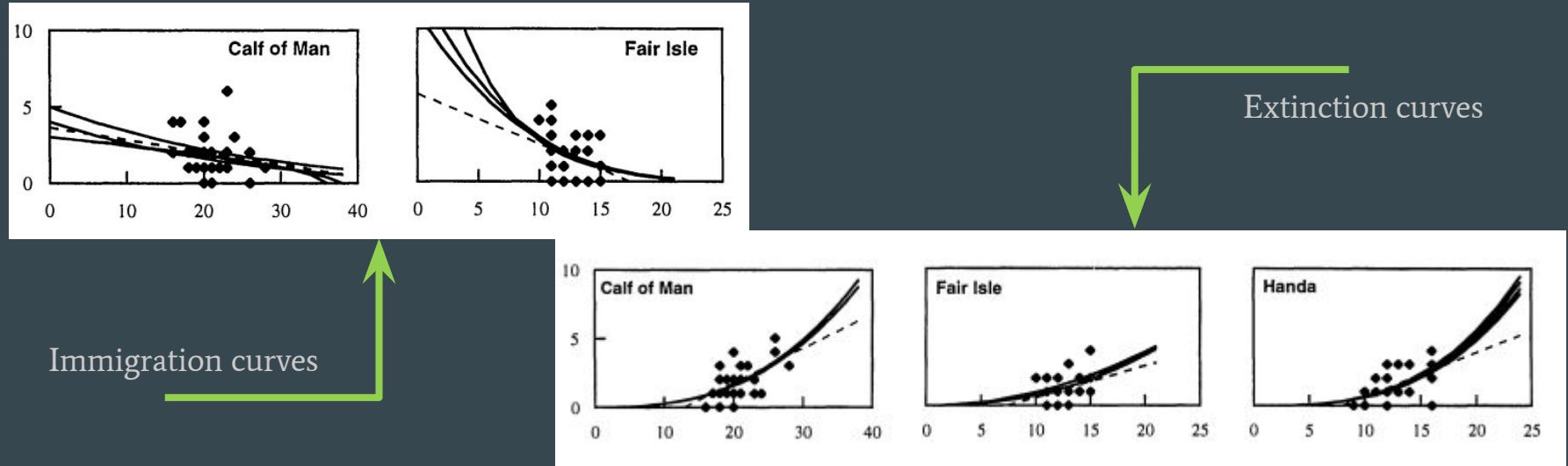
1. Estimate the number of species in the regional pool( P)
2. Express and maximize the likelihood

$$\begin{aligned}
 L(\mathbf{I}, \mathbf{E}; \beta) &= \prod_{t=1}^T \text{Bin}(I_t | P - S_t, \gamma_t) \prod_{t=1}^T \text{Bin}(E_t | S_t, \theta_t) \\
 &= \prod_{t=1}^T \gamma_t^{I_t} (1 - \gamma_t)^{P - S_t - I_t} \theta_t^{E_t} (1 - \theta_t)^{S_t - E_t}
 \end{aligned}$$



# Manne et al. 1998

For a majority of islands, Manne et al. (1998) did find **non-linear curves**, as expected from theory\*.



# Part 2. Process-based approaches

- A. Patch occupancy models

The idea that colonization rate is a function of distance and patch extinction rate varies with patch size can also be incorporated in Levins' metapopulation model (**spatially realistic models**): These are called **incidence function models** (IFM), starting from Hanski (1991)

# Part 2. Process-based approaches

- A. Patch occupancy models

The idea that colonization rate is a function of distance and patch extinction rate varies with patch size can also be incorporated in Levins' metapopulation model (**spatially realistic models**): These are called **incidence function models** (IFM), starting from Hanski (1991)

$$c_i = m_0 \exp(-\beta_1 D_i)$$

$$e_i = m_0 \exp(-\beta_2 A_i)$$

Decay rate of colonization with distance ( $D_i$ )

Decay rate of extinction with patch area ( $A_i$ )



# Part 2. Process-based approaches

- A. Patch occupancy models

IFM models can be fitted to (snapshot) occupancy data:

- Some parameters have to be estimated separately from other data (e.g., decay rates)
- We must assume dynamic equilibrium for fit, but can be used for projection

American Pika (Moilanen et al. 1998)



*Melitaea cinxia* (Wahlberg et al. 1996)



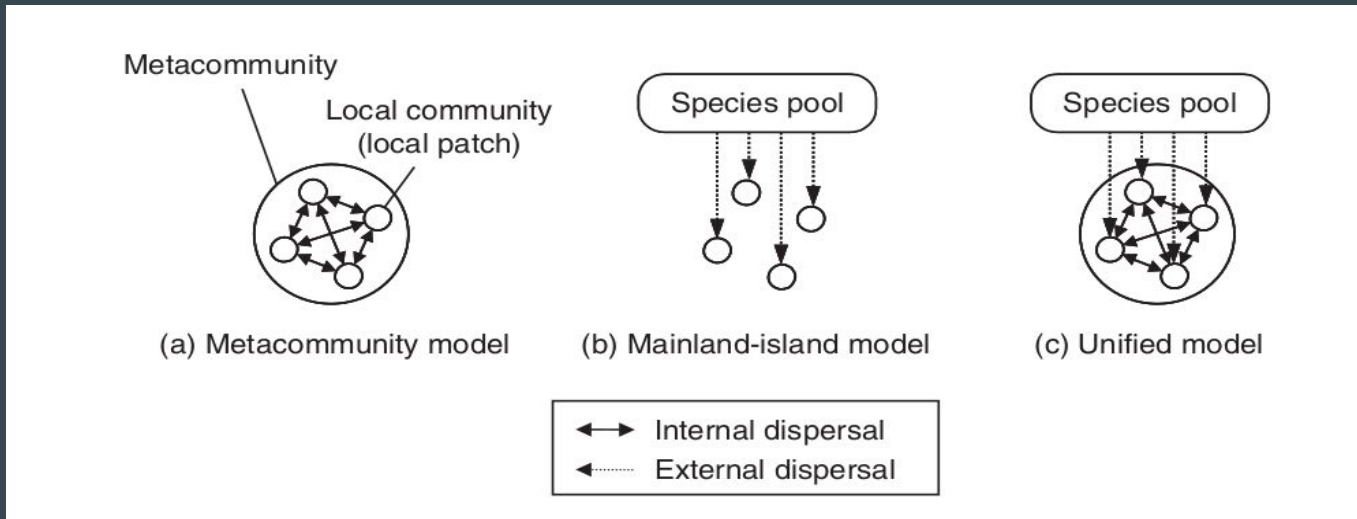
*Hyla arborea* (Vos et al. 2000)



# Part 2. Process-based approaches

- A. Patch occupancy models

Of course, you can mix the two assumptions for colonization and obtain a “unified” model\*:



Modified from Fukami (2005)

\*We let you do the maths as an exercise.

# Part 2. Process-based approaches

- A. Patch occupancy models

Patch occupancy models can be extended in many directions:

- multiple species with interactions (e.g. competition colonization trade-offs)
- different classes of patches (different habitats)
- spatially explicit models (contact networks)
- food webs (trophic chains or networks)
- ...

Of course, the more you complicate them, the more difficult analysis is, and also, the more difficult it is to parametrize them and fit them to data.

# Part 2. Process-based approaches

- A. Patch occupancy models

Even complex patch occupancy models can be fitted to temporal data (time series) - using the temporal turnover (i.e. apparent colonization and extinctions)

This is often called **dynamic occupancy modelling** (MacKenzie, 2003; Bailey et al. 2014)

They usually include a detection layer (i.e. a statistical modelling of the observation process)

# Part 2. Process-based approaches

- A. Patch occupancy models

The idea: modelling patch states (e.g.,  $\mathbf{X} = \{0, 1\}$ ) and transition probabilities (matrix  $\mathbf{T}$ ) among states, such that  $\mathbf{X}_{i,t+1} = \mathbf{X}_{i,t} \mathbf{T}$

Often based on Levins' metapopulation model (in discrete time)

equivalent to  $C(\cdot)$

Colonization prob.

$$\mathbf{T} = \begin{bmatrix} 1 - \gamma & \gamma \\ \varepsilon & 1 - \varepsilon \end{bmatrix}$$

Extinction prob.

$$P(X_{t+1} = 0) = P(X_t = 0) (1 - \gamma) + P(X_t = 1) \varepsilon$$

$$P(X_{t+1} = 1) = P(X_t = 0) \gamma + P(X_t = 1) (1 - \varepsilon)$$

equivalent to  $E(\cdot)$

# Part 2. Process-based approaches

- A. Patch occupancy models

The idea: modelling patch states (e.g.,  $\mathbf{X} = \{0, 1\}$ ) and transition probabilities (matrix  $\mathbf{T}$ ) among states, such that  $\mathbf{X}_{i,t+1} = \mathbf{X}_{i,t} \mathbf{T}$

Often based on Levins' metapopulation model (in discrete time)

equivalent to C(.)

Colonization prob.

$$\mathbf{T} = \begin{bmatrix} 1 - \gamma & \gamma \\ \varepsilon & 1 - \varepsilon \end{bmatrix}$$

Extinction prob.

$$P(X_{t+1} = 0) = P(X_t = 0) (1 - \gamma) + P(X_t = 1) \varepsilon$$

$$P(X_{t+1} = 1) = P(X_t = 0) \gamma + P(X_t = 1) (1 - \varepsilon)$$

Colonization and extinction can be patch and/or time specific, and then can include patch environmental characteristics, other species effects as well as distance-based dispersal

equivalent to E(.)

# Part 2. Process-based approaches

One example: metapopulation dynamics of *D. depressissimum* in Guadeloupe  
(Lamy et al. 2013)

Transition matrix

$$P_{WW} = \begin{matrix} & 0 & 1 \\ \begin{matrix} 0 \\ 1 \end{matrix} & \begin{bmatrix} 1-\gamma_W & \gamma_W \\ (1-\Phi_W)*(1-\gamma_W) & \Phi_W + (1-\Phi_W)*\gamma_W \end{bmatrix} \end{matrix}$$

With  $\Phi = 1 - \varepsilon$  ; i.e. the persistence probability



© Jean-Pierre POINTIER/EPHE/USR 3278/CNRS Photothèque

# Part 2. Process-based approaches

One example: metapopulation dynamics of *D. depressissimum* in Guadeloupe (Lamy et al. 2013)

Table 4: Posterior means of intercept and covariate effects in the fit their 95% credible intervals (CIs)

Parameter	Coefficient	Mean (95% CI)	Predicted mean
$\Phi_{W,t}$			
Intercept	$a_1$	1.31 (.883, 1.758)	.787
Vegetation	$\beta_2$	.961 (.650, 1.297)	
Little rainy season	$\beta_4$	-.631 (-.998, -.298)	
$\gamma_{W,t}$			
Intercept	$a_3$	-.843 (-1.100, -.599)	.288
Size	$\beta_9$	.382 (.166, .604)	
Stability	$\beta_{11}$	-.839 (-1.180, -.538)	
Connectivity	$\beta_{12}$	.353 (.142, .575)	
$p_W$ :			
Detectability		.777 (.742, .811)	

Note: Each coefficient is given in logit scale. Consequently, for each parameter ( $\Phi_w$ ,  $\Phi_D$ ,  $\gamma_w$ , and  $\gamma_D$ ), we computed the predicted mean by setting all covariates to their conditional means in wet sites (for  $\Phi_w$  and  $\gamma_w$ ) or dry sites (for  $\Phi_D$  and  $\gamma_D$ ) in the prediction formula (eqq. [4]–[7]) and using the inverse-logit function to back-transform into the natural scale.

Model	$q_{i,t}$	Parameter	Estimates in wet sites (CI)		
			$p_W$	$\Phi_W$	$\gamma_W$
Model 1	+	$\Phi_W + \gamma_W$	.79 (.75, .82)	.76 (.69, .82)	.32 (.28, .36)

Detection prob.

Mean persistence prob.

Mean colonization prob.

Envt. effects on persistence

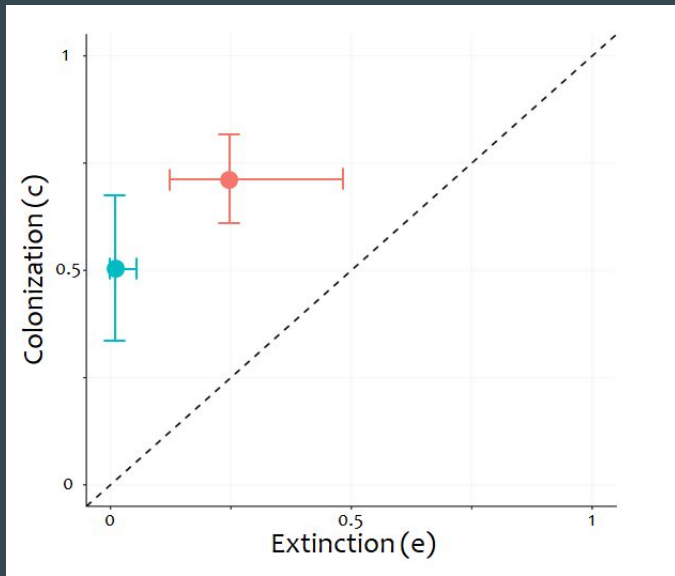
Envt. effects on colonization



# Part 2. Process-based approaches

With multiple species: *P. acuta* & *A. marmorata* - Guadeloupe

Rates in **absence** of the other species



*Aplexa marmorata*

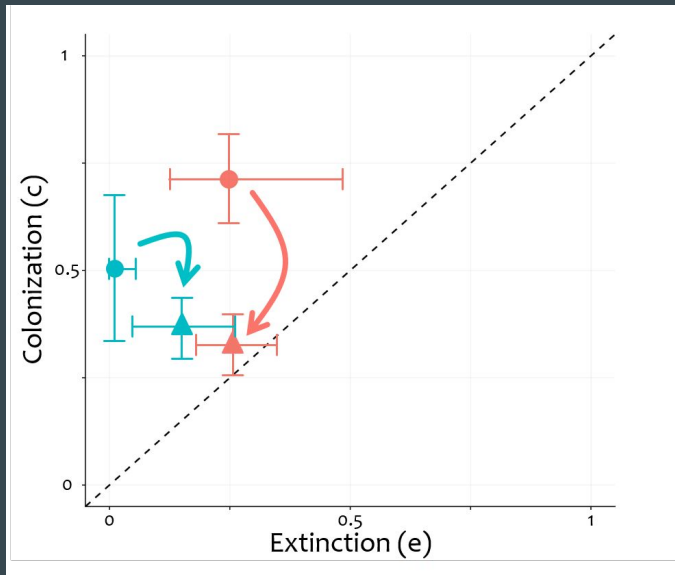


*Physa acuta*

# Part 2. Process-based approaches

With multiple species: *P. acuta* & *A. marmorata* - Guadeloupe

Rates in **presence** of the other species



*Aplexa marmorata*



*Physa acuta*

*A. marmorata* increases *P. acuta* extinction rate

*P. acuta* reduces *A. marmorata* colonization rate

# Part 2. Process-based approaches

With multiple species: *P. acuta* & *A. marmorata* - Guadeloupe

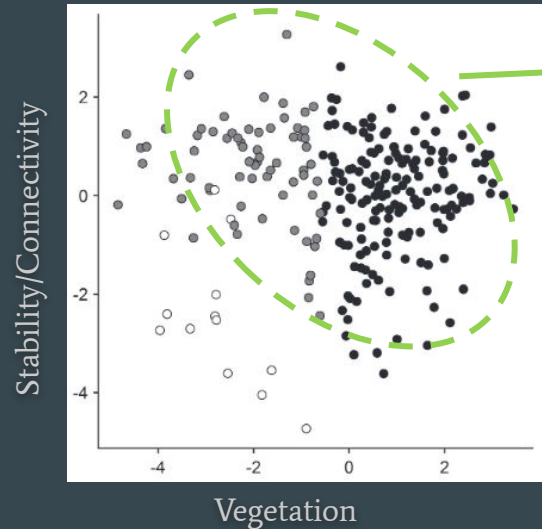
These interaction effects depends on environmental conditions



*Aplexa marmorata*



*Physa acuta*



*A. marmorata* fundamental niche

# Part 2. Process-based approaches

With multiple species: *P. acuta* & *A. marmorata* - Guadeloupe

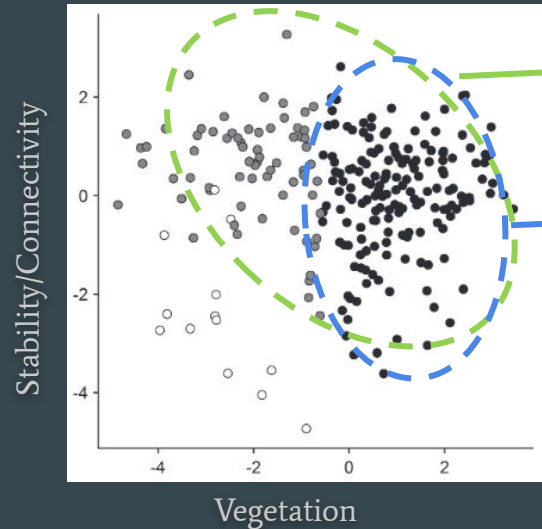
These interaction effects depends on environmental conditions



*Aplexa marmorata*



*Physa acuta*

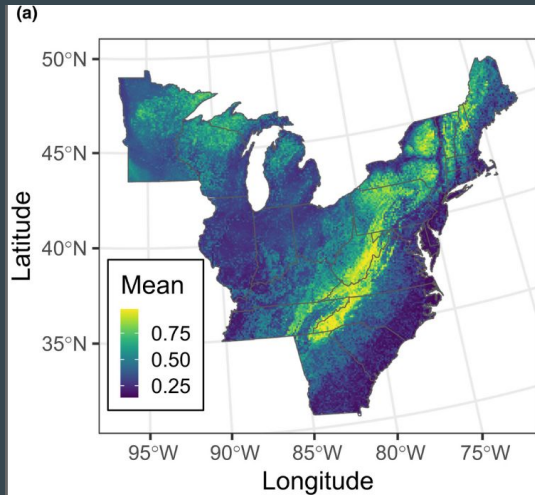


*A. marmorata* fundamental niche

*A. marmorata* realized niche


# Part 2. Process-based approaches

There are increasingly flexible and easy-to-use solutions to implement occupancy models



Methods in Ecology and Evolution

BRITISH  
ECOLOGICAL  
SOCIETY

APPLICATION |  Open Access |   

**spOccupancy: An R package for single-species, multi-species,  
and integrated spatial occupancy models**

Jeffrey W. Doser  Andrew O. Finley, Marc Kéry, Elise F. Zipkin

First published: 15 May 2022 | <https://doi.org/10.1111/2041-210X.13897> | Citations: 19

## Part 2. Process-based approaches

- B. Models with local population dynamics

All previous mechanistic models describe presence/absence only (i.e. neglected local dynamics)

It may be more realistic to describe within patch population dynamics, and more data could be used (when abundance data is available)

## Part 2. Process-based approaches

- B. Models with local population dynamics

There are, again, many models that attempt to do this.

- > The simplest approach possible is to take a patch occupancy model (e.g. Levins' model), and add a dynamical variable to describe local population density (e.g. logistic growth within each patch). This approach was taken for instance by Gyllenberg & Hanski (1992)
- > Alternatively, one can simply distinguish “small” populations (i.e. recently colonized) from “large” populations (small populations that have grown). This adds a second equation to Levins' model, but remains quite simple (Hanski 1985)

# Part 2. Process-based approaches

- B. Models with local population dynamics

Such structured metapopulation models can predict non linear changes of extinction and colonization rates with the total fraction of occupied patches.

This can cause **rescue effects** and **alternative stable states**  
(similar to an Allee effect at metapopulation level)

It has implications for the response to perturbations (Eriksson et al. (2014))



## Part 2. Process-based approaches

- B. Models with local population dynamics

There are, again, many models that attempt to do this.

> Another example is Hubbell's **neutral model of biodiversity** over several connected communities.

Owing to the fact that all species are assumed identical, the dynamics of local abundances (assuming some constant total abundance) is relatively easy to model.

Etienne (2005)

$$P[D|\theta, m, J] = \frac{J!}{\prod_{i=1}^S n_i \prod_{j=1}^J \Phi_j!} \frac{\theta^S}{(\theta)_J} \\ \times \sum_{A=S}^J \left( K(D, A) \frac{(\theta)_J}{(\theta)_A} \frac{I^A}{(I)_J} \right)$$

# Part 2. Process-based approaches

- B. Models with local population dynamics

There are, again, many models that attempt to do this.

> **Reaction-diffusion models** are yet another framework to model the spatial spread of species, combining local dispersal (diffusion) and local population dynamics (reaction)

# Part 2. Process-based approaches

- B. Models with local population dynamics

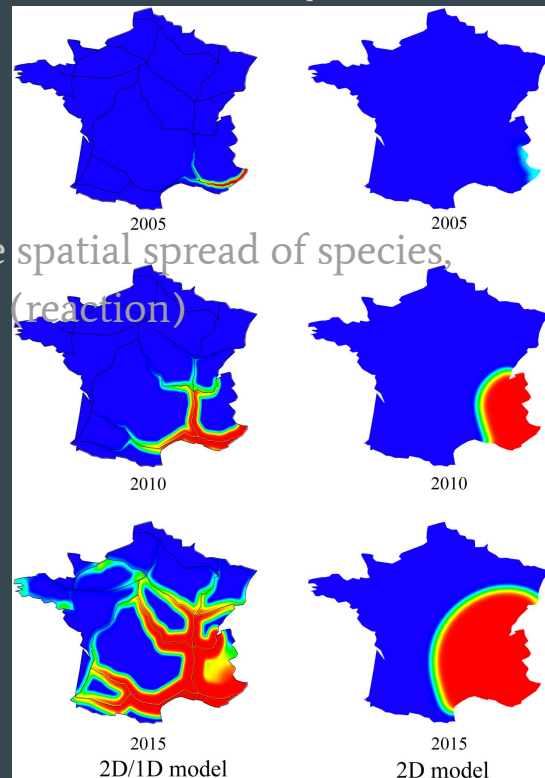
There are, again, many models that attempt to do this.

> **Reaction-diffusion models** are yet another framework to model the spatial spread of species, combining local dispersal (diffusion) and local population dynamics (reaction)

*Aedes albopictus* @ Wikipedia



Roques & Bonnefon 2016



## Part 2. Process-based approaches

- B. Models with local population dynamics

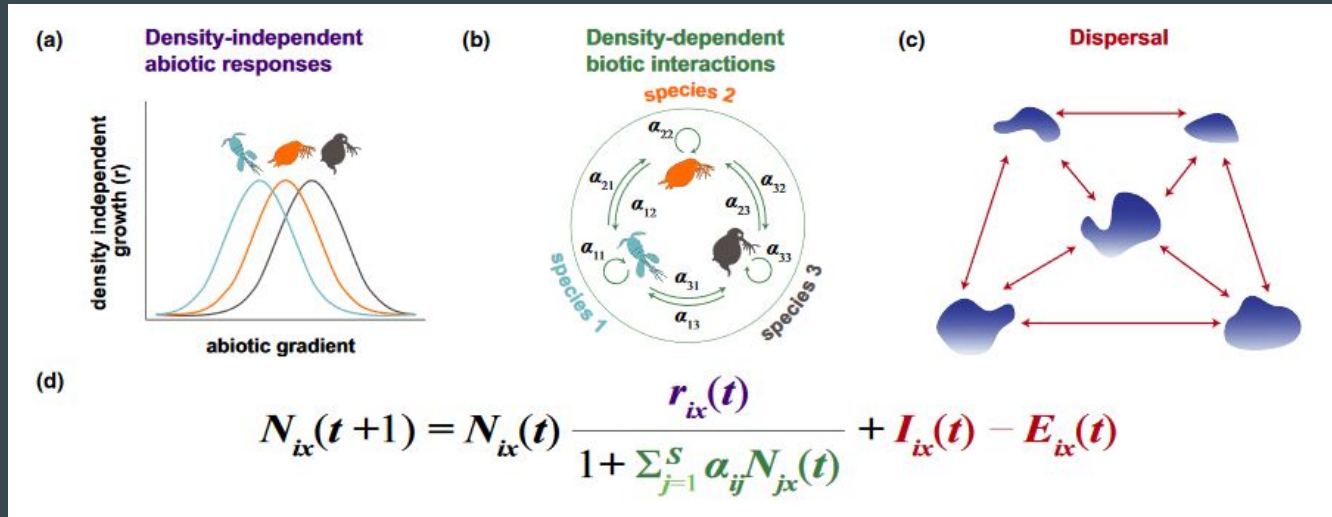
Describing the local population dynamics with several interacting species quickly becomes **too complicated** to be usefully modelled with simple mathematical models.

In this case, we usually turn to **stochastic simulation models**.

# Part 2. Process-based approaches

- C. Stochastic simulation models

One relatively general metacommunity model has been proposed by Thompson et al. (2020)



Stochasticity is modelled as the random draw of the values of  $E(t)$  and  $N(t)$

# Conclusion

## Ordination methods

### Pros:

- Very large literature
- Widely used/applicable
- Able to work on (very) large datasets

### Cons:

- Links with theory unclear
- So many methods
- Sometimes controversial

## JSDMs

### Pros:

- Sophisticated models with will for linking components to processes
- Very flexible
- Can assess relationship between species traits/phylo. and 'niche'

### Cons:

- Still SDMs with residual correlation as criterion
- Results interpretation not always clear

## Process-based models

### Pros:

- Explicit processes and link to theory
- Can be customized to specific system based on knowledge
- Can address 'what if' questions

### Cons:

- Greedy in data & comp. resources
- Limited in the number of species they can handle
- Highly parametric

# Conclusion

- So what?

# Conclusion

- So what?

So, no good method!

*“In conclusion, we find no reason why a proponent of either of the two extremes of correlative and process-based species distribution modelling should hold the moral high ground. ‘Correlationists’ should be humble: their model’s success maybe due to spurious correlations. ‘Mechanists’ should be unassertive about their approach, because they will only find effects of processes that they included. Either approach must comply with nature, statistically or mechanistically, and be aware of the kinds of questions they are best suited to answer.” Dormann et al. 2012*



# To go further

Main reference textbooks:

Legendre & Legendre. **2012**. *Numerical Ecology*.

Vellend. **2016**. *The Theory of Ecological Communities*.

Leibold & Chase. **2018**. *Metacommunity Ecology*.

Mittelbach & McGill. **2019**. *Community Ecology*.

Ovaskainen & Abrego. **2020**. *Joint Species Distribution Modelling*

# Ready for practicals?

- questions on morning session



# Afternoon: practicals

## 1. Simulate data/test methods

Simulate with the following code:

<https://github.com/mxdub/TMDSpatial>

Try different analyses

→Does it work?

To install : JAGS

R packages : vegan, R2jags, Hmsc

And

```
devtools::install_github('mxdub/TMDSpatial')
```

## 2. The effects of patch disturbance

Article walkthrough:

Functional Ecology



RESEARCH ARTICLE | Open Access |

**Metacommunity dynamics and the detection of species associations in co-occurrence analyses: Why patch disturbance matters**

Vincent Calcagno, Nik J. Cunniffe , Frédéric M. Hamelin

First published: 04 April 2022 | <https://doi.org/10.1111/1365-2435.14047>

Go fetch the R markdown at:

<https://github.com/nikcunniffe/MetacommunityDynamics>

# Practical 2: the effects of patch disturbance

1. What is patch disturbance?

2. How do you include it in Levins' metapopulation model?

(write down the equations)

(walk through model formulation: patch age, equilibrium...)

3. What consequences for species co-occurrence patterns?

4. Test the expectations/predictions