Joint species distribution modelling: how to make more out of community data?

Prof. Otso Ovaskainen

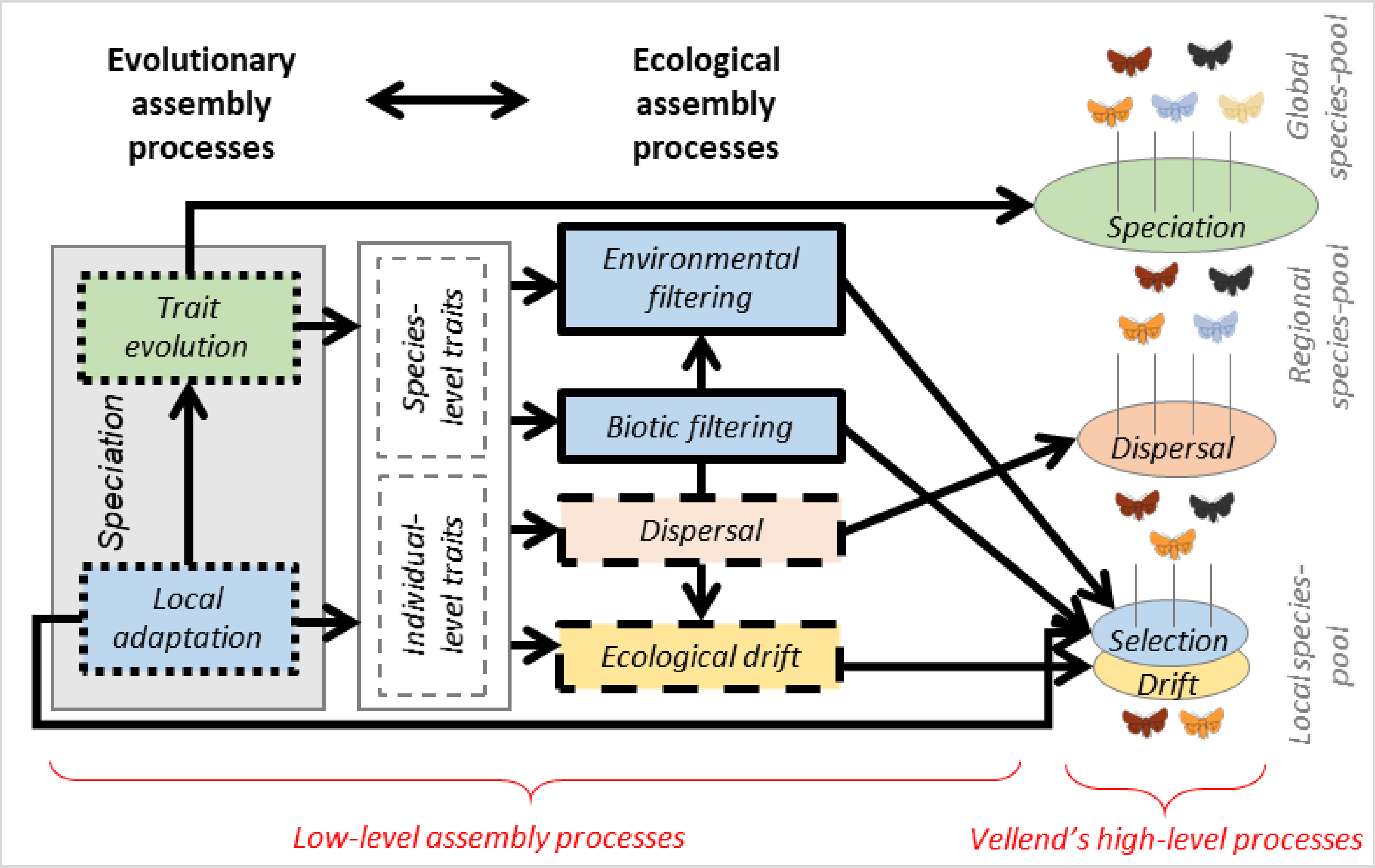
University of Jyväskylä, Finland
University of Helsinki, Finland
NTNU Trondheim, Norway

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

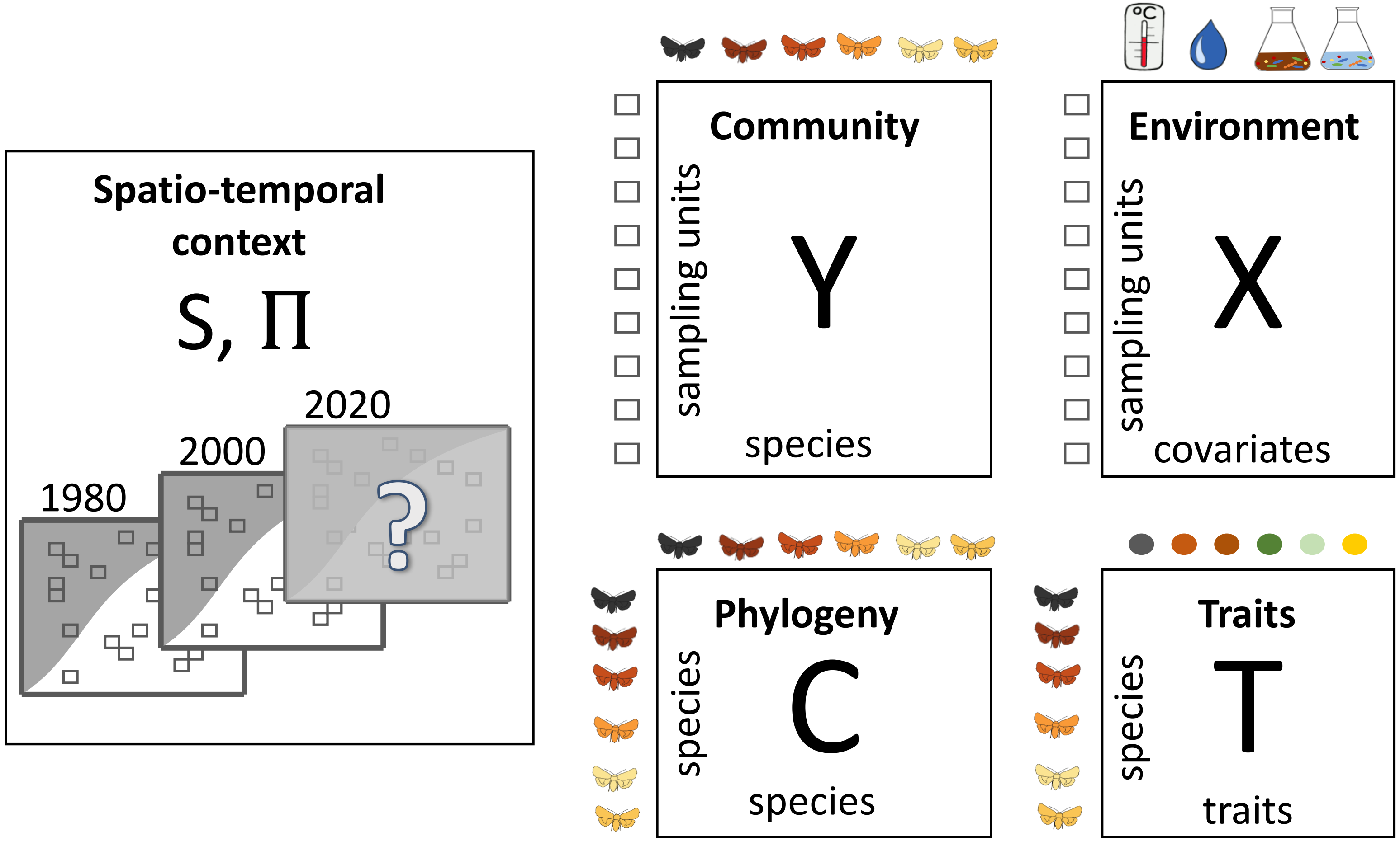
Community ecology:
in the capture of assembly processes

A central goal in community ecology is to understand what processes drive the distributions and dynamics of species

What we are interested about: assembly processes



What we mostly have data about: patterns that results from assembly processes



Spatio-temporal context

S, Π

Community

Y

sampling units

species

Environment

X

sampling units

covariates

Phylogeny

C

species

species

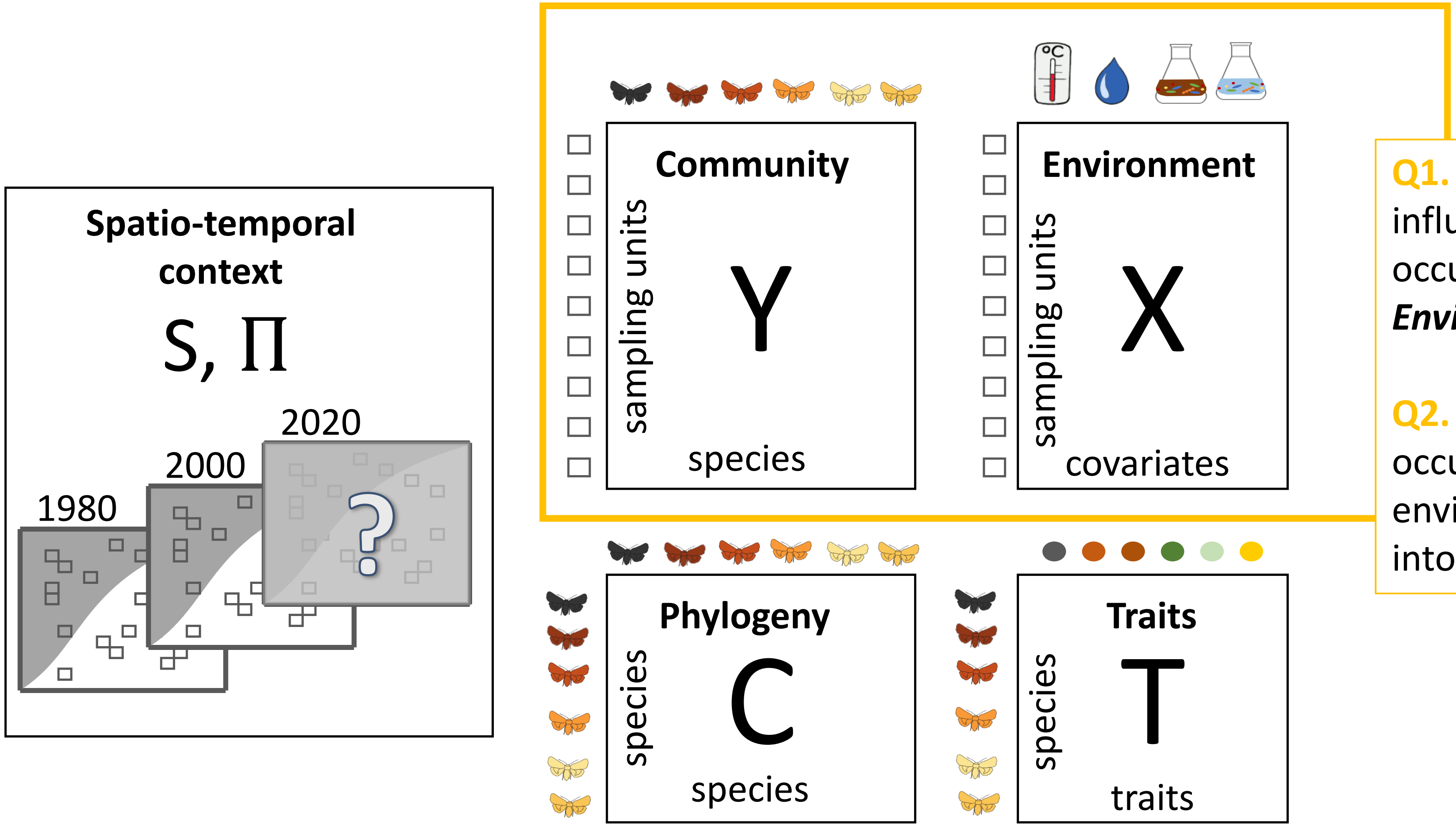
Traits

T

species

traits

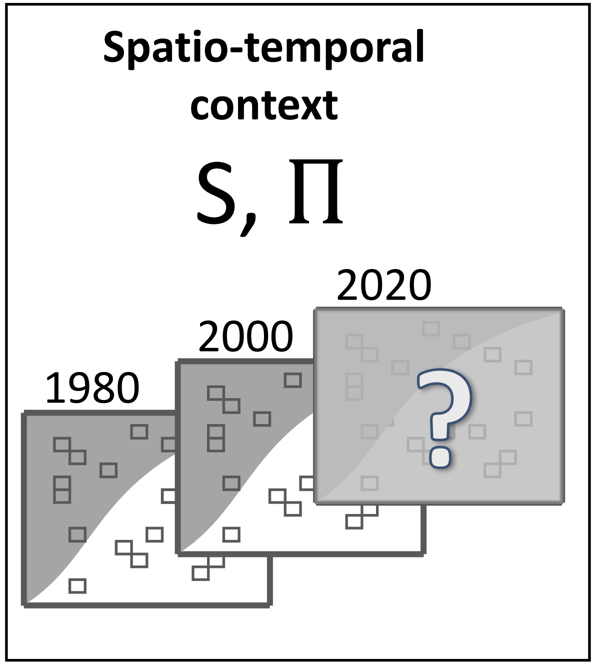
Typical questions asked in community ecology

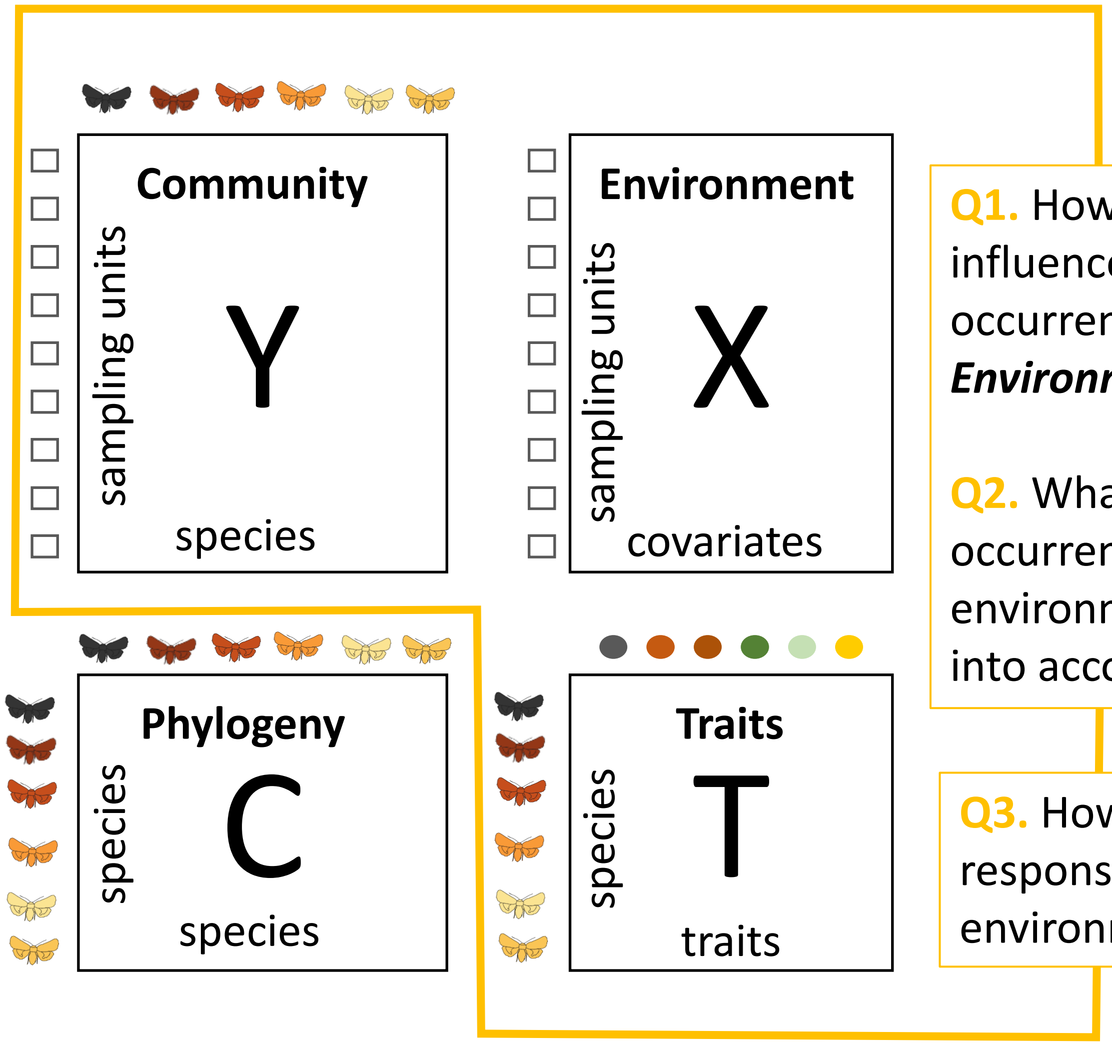


Q1. How do environmental conditions influence species occurrences/community composition? ***Environmental filtering***

Q2. What are the species’ co-occurrence patterns after the effects of environmental conditions are taken into account? ***Biotic filtering***

Typical questions asked in community ecology



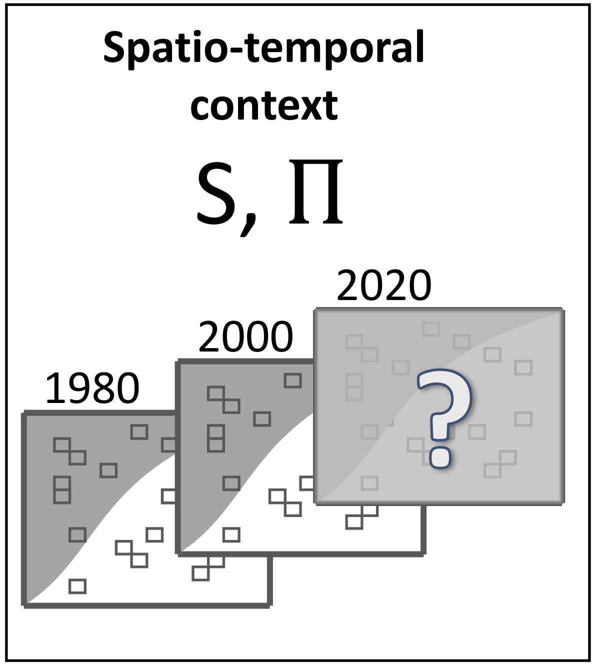


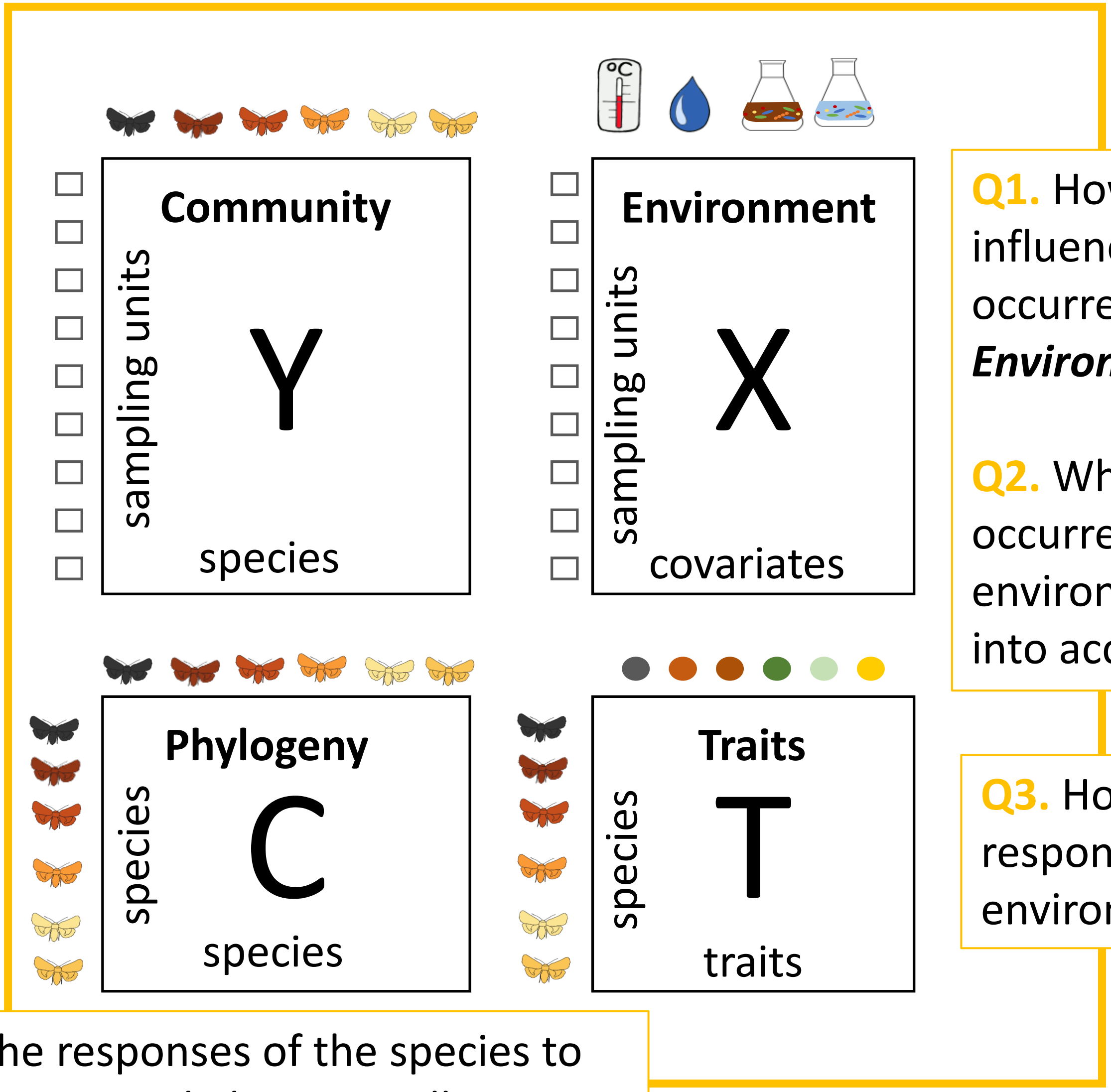
Q1. How do environmental conditions influence species occurrences/community composition? ***Environmental filtering***

Q2. What are the species' co-occurrence patterns after the effects of environmental conditions are taken into account? ***Biotic filtering***

Q3. How do traits influence the responses of the species to the environment?

Typical questions asked in community ecology





Q1. How do environmental conditions influence species occurrences/community composition? ***Environmental filtering***

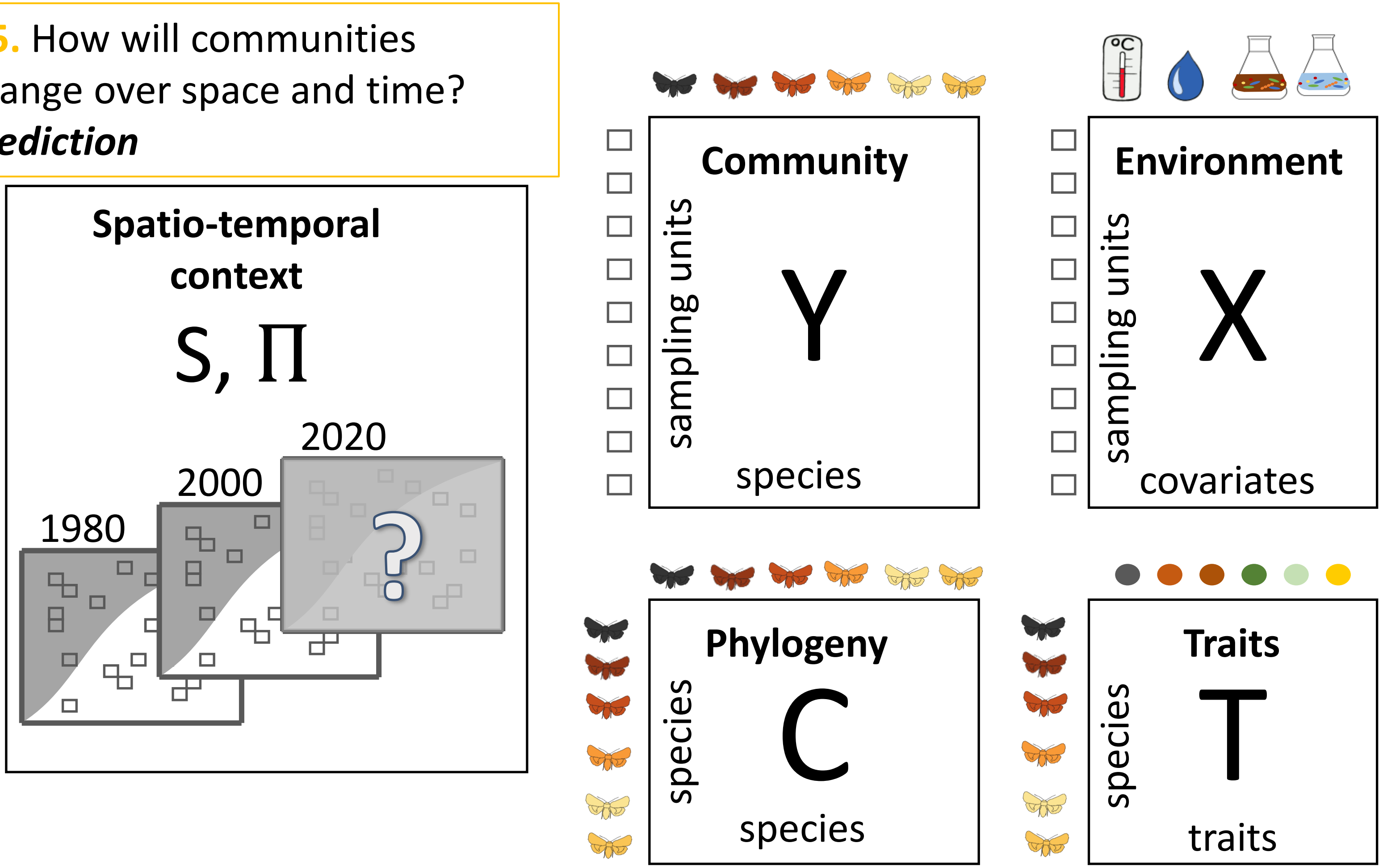
Q2. What are the species' co-occurrence patterns after the effects of environmental conditions are taken into account? ***Biotic filtering***

Q3. How do traits influence the responses of the species to the environment?

Q4. Are the responses of the species to the environment phylogenetically structured? ***Phylogeographic processes***

Typical questions asked in community ecology

Q5. How will communities change over space and time? ***Prediction***



Q1. How do environmental conditions influence species occurrences/community composition? ***Environmental filtering***

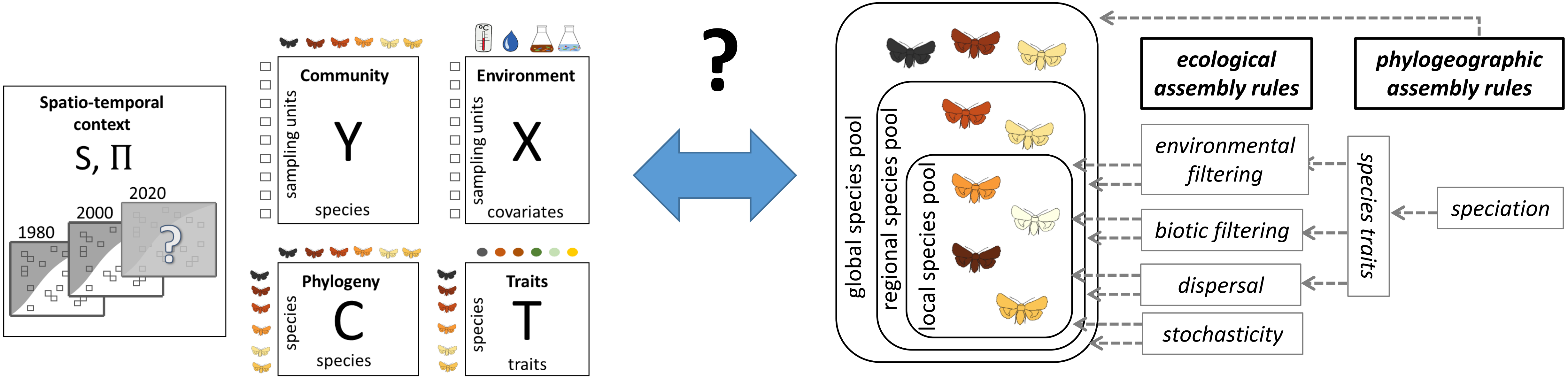
Q2. What are the species' co-occurrence patterns after the effects of environmental conditions are taken into account? ***Biotic filtering***

Q3. How do traits influence the responses of the species to the environment?

Q4. Are the responses of the species to the environment phylogenetically structured? ***Phylogeographic processes***

The BIG QUESTION in *statistical* community ecology:

How to best use the data to address the research questions?



Overview of (typical) methods in community ecology: which data they utilize?

	Input data (as a whole, without subsetting or pre-treatment)				
Method	Community data	Environmental data	Trait data	Phylogenetic data	Spatio-temporal structure
Ordinations	✓	~	~	~	~
Co-occurrence analyses	✓	✗	✗	✗	✗
Network analyses	✓	✗	✗	✗	✗
Diversity metrics	✓	~	✗	✗	~
Joint species distribution modelling	✓	✓	✓	✓	✓

✓Yes, ~ Some variants, ✗No

Overview of (typical) methods in community ecology: which kinds of questions they answer?

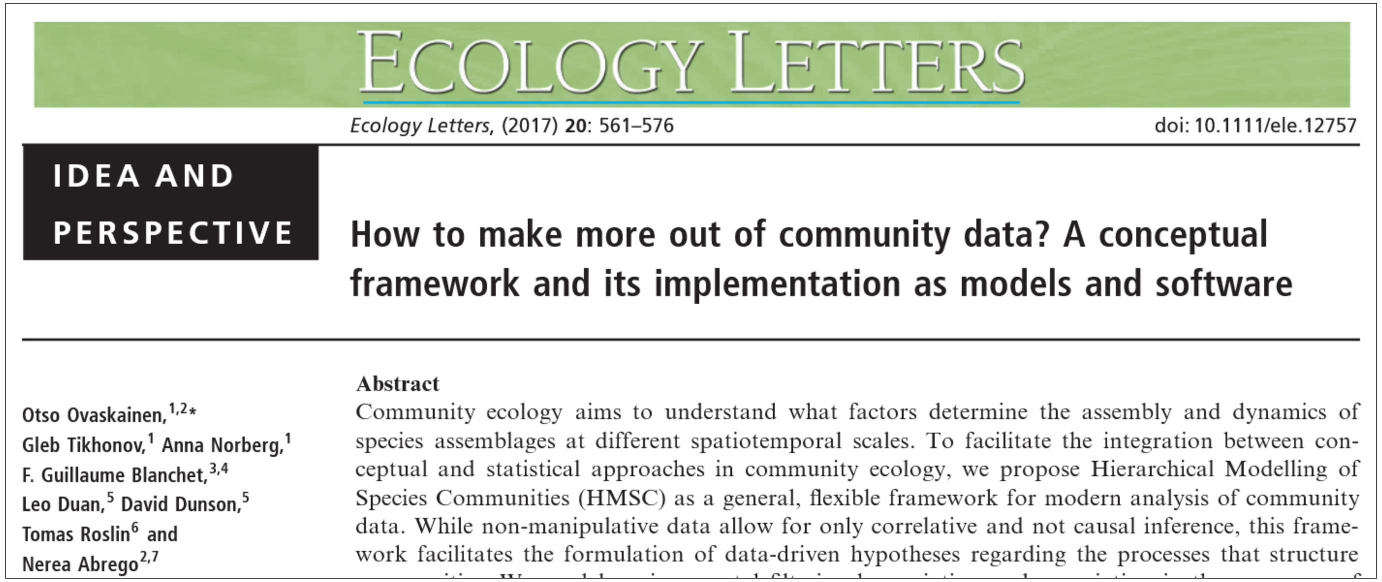
	Outputs (without post-hoc treatment)			
Method	**Environmental filtering**	**Biotic filtering**	**Effects of traits and phylogeny**	**Spatiotemporal predictions**
Ordinations	✓	~	~	✗
Co-occurrence analyses	✗	✓	✗	✗
Network analyses	✓	✓	✗	✗
Diversity metrics	✓	✗	✗	~
Joint species distribution modelling	✓	✓	✓	✓

✓Yes, ~ Some variants, ✗No

Joint species distribution modelling

A fast emerging field in statistical community ecology

2017



ECOLOGY LETTERS

Ecology Letters, (2017) **20**: 561–576 doi: 10.1111/ele.12757

IDEA AND PERSPECTIVE

How to make more out of community data? A conceptual framework and its implementation as models and software

Otso Ovaskainen,[1,2]*
Gleb Tikhonov,[1] Anna Norberg,[1]
F. Guillaume Blanchet,[3,4]
Leo Duan,[5] David Dunson,[5]
Tomas Roslin[6] and
Nerea Abrego[2,7]

Abstract

Community ecology aims to understand what factors determine the assembly and dynamics of species assemblages at different spatiotemporal scales. To facilitate the integration between conceptual and statistical approaches in community ecology, we propose Hierarchical Modelling of Species Communities (HMSC) as a general, flexible framework for modern analysis of community data. While non-manipulative data allow for only correlative and not causal inference, this framework facilitates the formulation of data-driven hypotheses regarding the processes that structure

2020

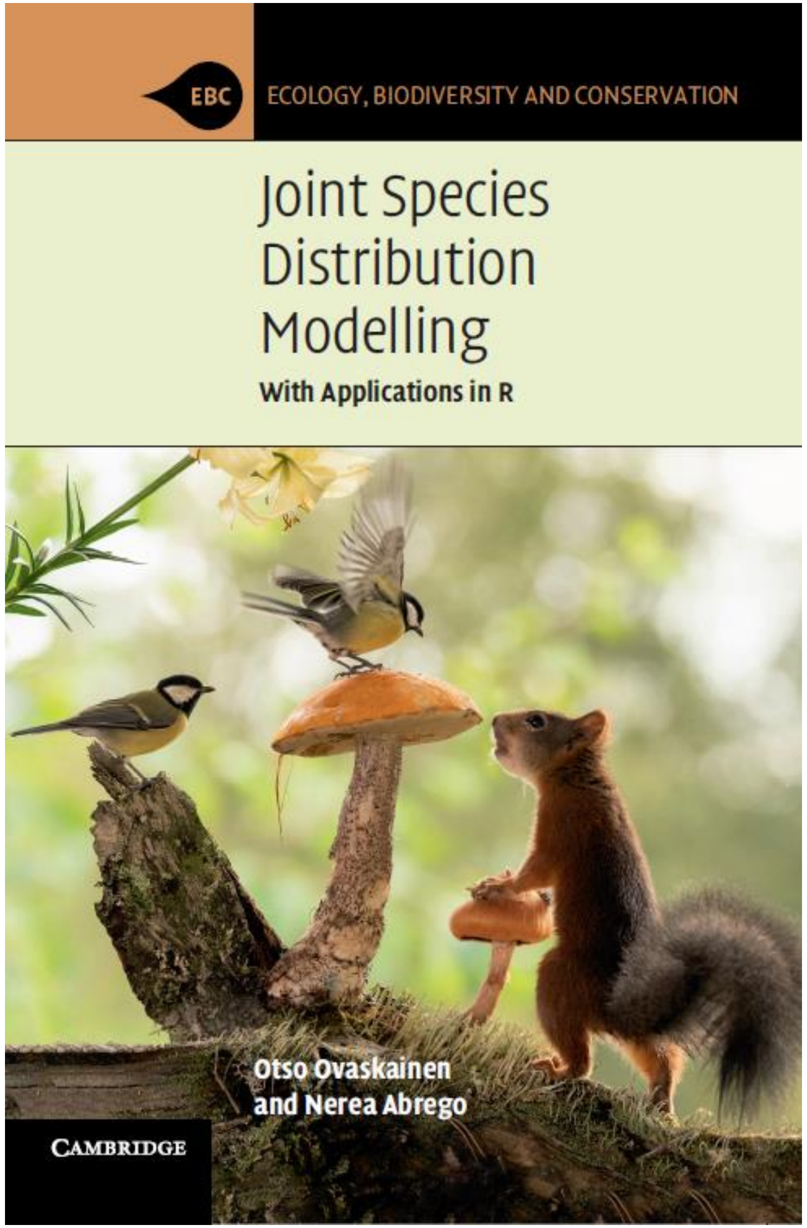
APPLICATION

Methods in Ecology and Evolution BRITISH ECOLOGICAL SOCIETY

Joint species distribution modelling with the R-package HMSC

Gleb Tikhonov[1,2] | Øystein H. Opedal[2,3] | Nerea Abrego[4] | Aleksi Lehikoinen[5] | Melinda M. J. de Jonge[6] | Jari Oksanen[7] | Otso Ovaskainen[2,3]

2020



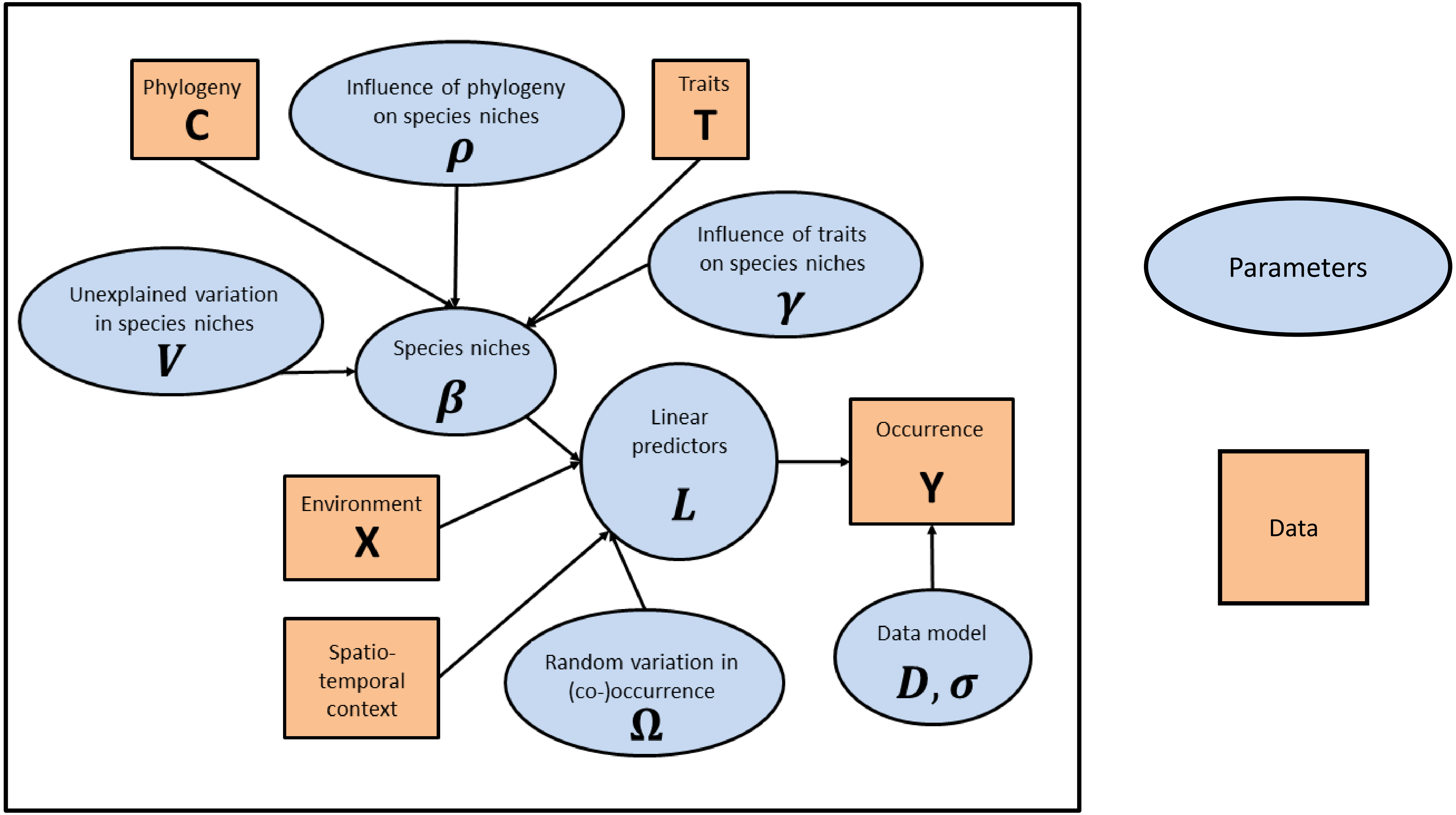
EBC ECOLOGY, BIODIVERSITY AND CONSERVATION

Joint Species
Distribution
Modelling
With Applications in R

Otso Ovaskainen
and Nerea Abrego

CAMBRIDGE

Hierarchical Modelling of Species Communities (HMSC)



Graphical description of the HMSC model as a Directed Acyclic Graph (DAG)

HMSC is a multivariate hierarchical generalized linear mixed model fitted with Bayesian inference

Fixed effects

Random effects

Linear predictor for sampling unit i and species j:

$$L_{ij} = \underbrace{L_{ij}^{F}}_{\text{Fixed effects}} + \underbrace{L_{ij}^{R}}_{\text{Random effects}}$$

Examples of data models:

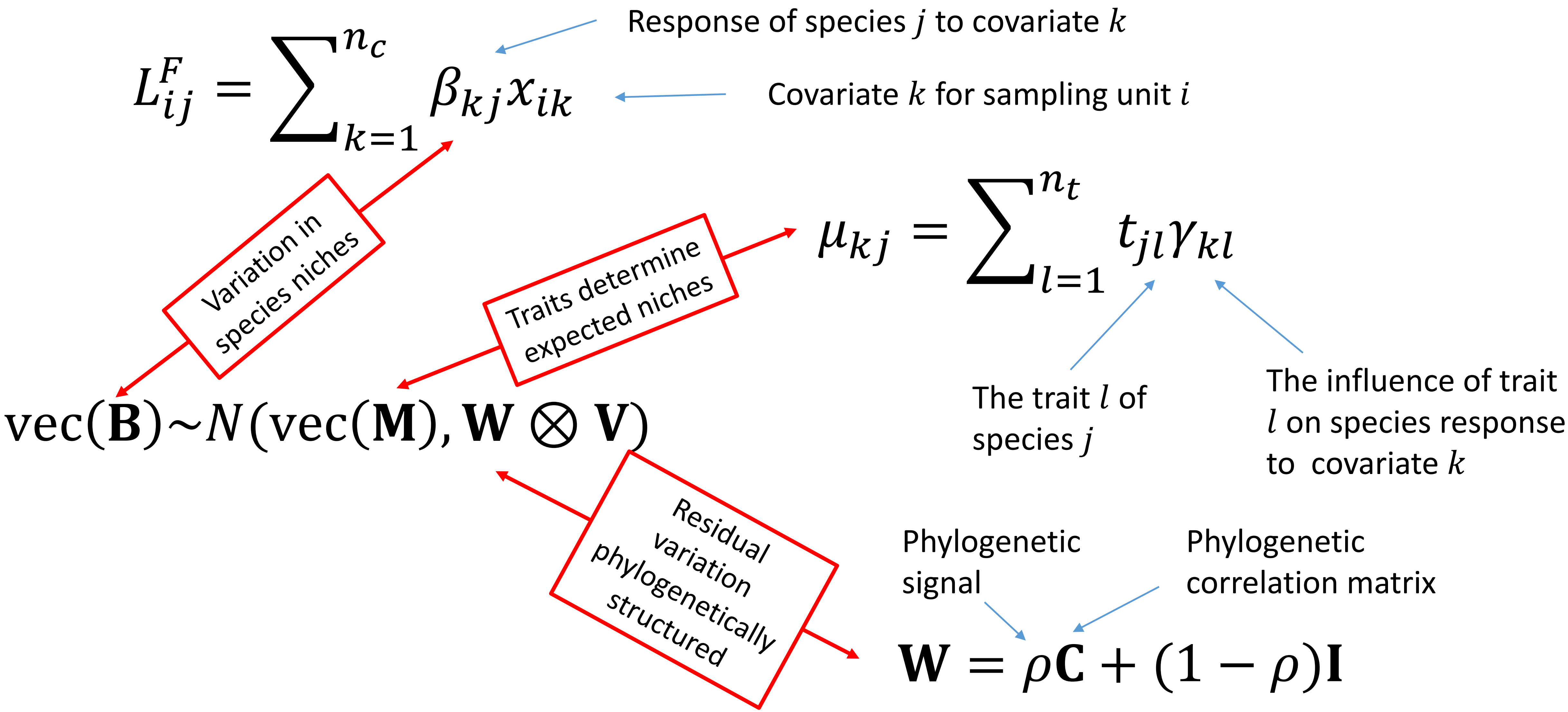
Linear model: $y_{ij} = L_{ij} + \varepsilon_{ij}$

Probit model: $y_{ij} \sim \mathrm{Bernoulli}(\Phi(L_{ij}))$

Lognormal Poisson model: $y_{ij} \sim \mathrm{Poisson}\left(\exp\left(L_{ij} + \varepsilon_{ij}\right)\right)$

Fixed effects ~ environmental filtering

Linear predictor for fixed effects for sampling unit i and species j:



$$L_{ij}^{F} = \sum_{k=1}^{n_c} \beta_{kj} x_{ik}$$

Response of species j to covariate k

Covariate k for sampling unit i

Variation in species niches

$$\mu_{kj} = \sum_{l=1}^{n_t} t_{jl} \gamma_{kl}$$

Traits determine expected niches

The trait l of species j

The influence of trait l on species response to covariate k

$$\mathrm{vec}(\mathbf{B}) \sim N(\mathrm{vec}(\mathbf{M}), \mathbf{W} \otimes \mathbf{V})$$

Residual variation phylogenetically structured

Phylogenetic signal

Phylogenetic correlation matrix

$$\mathbf{W} = \rho \mathbf{C} + (1 - \rho)\mathbf{I}$$

Random effects ~ biotic filtering and missing covariates

Number of factors

$$L_{ij}^{R} = \sum_{h=1}^{n_f} \eta_{ih}\lambda_{hj}$$

Site loadings Species loadings

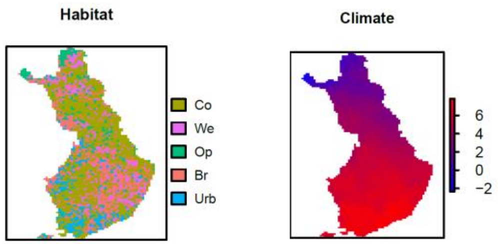
$\lambda_{hj} \sim$ Multiplicative gamma process shrinking prior (Bhattacharya and Dunson 2011)

$$\eta_{ih} \sim N(0,1) \quad \Rightarrow \quad \mathrm{Cov}\left[L_{i_1 j_1}^{R}, L_{i_2 j_2}^{R}\right] = \sum_{h=1}^{n_f} \lambda_{hj_1}\lambda_{hj_2}\delta_{i_1 i_2}$$

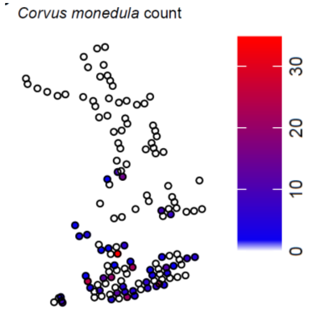
HMSC input and output

EXAMPLE DATA ON 50 MOST COMMON SPECIES OF FINNISH BIRDS

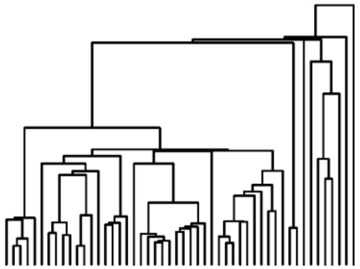
LAND COVER AND CLIMATIC DATA



SPECIES DATA



TRAITS AND PHYLOGENIES

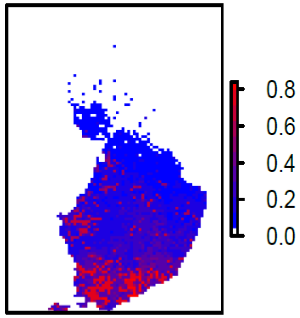


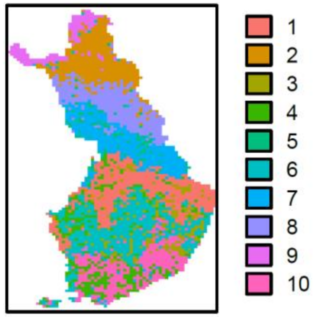
JOINT SPECIES DISTRIBUTION MODELLING BY HMSC

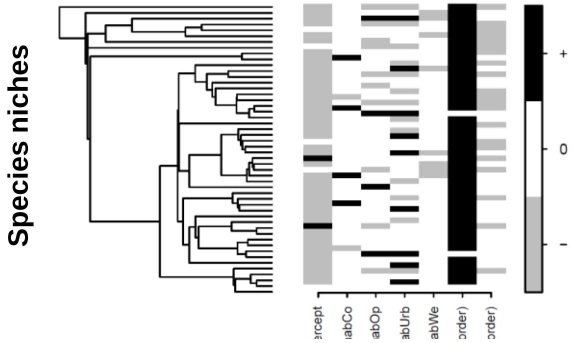
Current distribution of biodiversity

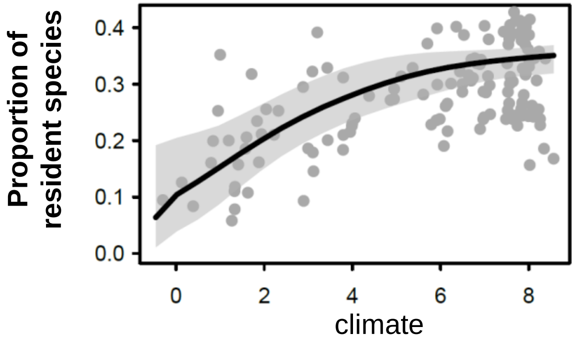
Response to environmental covariates

Response to environmental change


***Corvus monedula* occurrence probability**


Regions of common profile





MODEL FITTING

INFERENCE AND PREDICTION

The overall workflow of HMSC in the R-package Hmsc

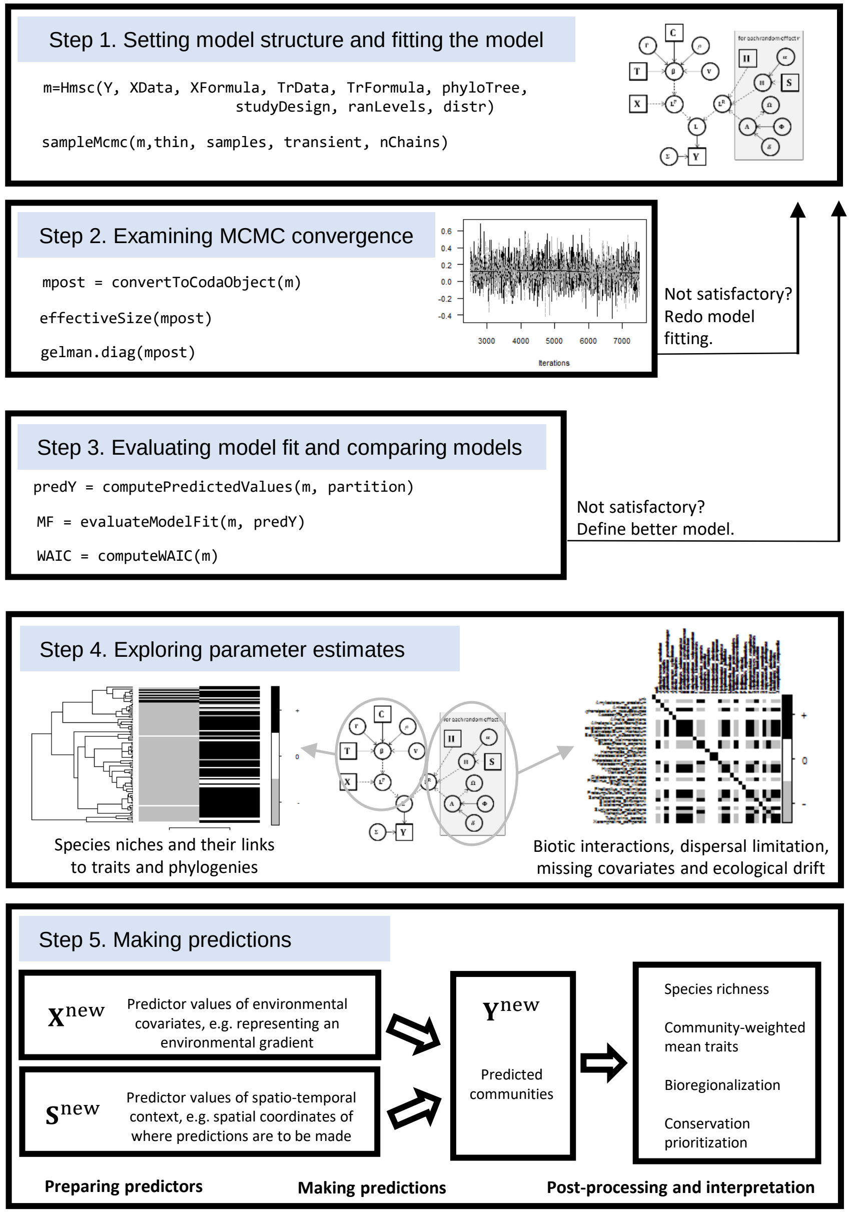
Step 1. Setting model structure and fitting the model

Step 2. Examining MCMC convergence

Step 3. Evaluating model fit and comparing models

Step 4. Exploring parameter estimates

Step 5. Making predictions



Step 1. Setting model structure and fitting the model

```
m=Hmsc(Y, XData, XFormula, TrData, TrFormula, phyloTree,
                        studyDesign, ranLevels, distr)

sampleMcmc(m,thin, samples, transient, nChains)
```

Step 2. Examining MCMC convergence

```
mpost = convertToCodaObject(m)

effectiveSize(mpost)

gelman.diag(mpost)
```

Not satisfactory?
Redo model
fitting.

Step 3. Evaluating model fit and comparing models

```
predY = computePredictedValues(m, partition)

MF = evaluateModelFit(m, predY)

WAIC = computeWAIC(m)
```

Not satisfactory?
Define better model.

Step 4. Exploring parameter estimates

Species niches and their links
to traits and phylogenies

Biotic interactions, dispersal limitation,
missing covariates and ecological drift

Step 5. Making predictions

$\mathbf{X}^{\text{new}}$ Predictor values of environmental covariates, e.g. representing an environmental gradient

$\mathbf{S}^{\text{new}}$ Predictor values of spatio-temporal context, e.g. spatial coordinates of where predictions are to be made

$\mathbf{Y}^{\text{new}}$

Predicted communities

Species richness

Community-weighted mean traits

Bioregionalization

Conservation prioritization

Preparing predictors

Making predictions

Post-processing and interpretation

HMSC belongs to the class of Joint Species Distribution Models

- **HMSC** by Ovaskainen, Abrego et al: Hierarchical Modelling of Species Communities
- **BORAL** by Hui et al: Bayesian Ordination and Regression Analysis
- **GLLVM** by Hui, Taskinen et al.: Generalized Linear Latent Variable Models
- **GJAM** by Clark et al.: Generalised Joint Attribute Modelling
- **JDSDM** by Thorson et al: Joint Dynamic Species Distribution Models
- **MISTN** by Harris: Multivariate Stochastic Neural Network
- **BC** by Golding and Harris: Bayesian Community Ecology Analysis
- **SSDM** = Stacked Species Distribution Models: ”Pick your favourite single-species model and apply it one by one to each species”

How “joint” is a joint species distribution model?

Niche structure

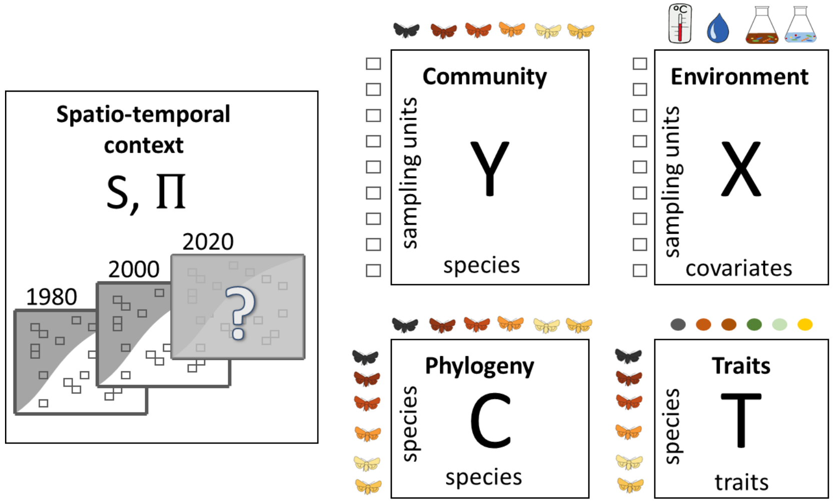
N1: joint modelling of species niches
N2: dependency on traits
N3: dependency on phylogeny

Biotic interactions

B1: Estimation of association networks
B2: dependency on traits
B3: dependency on phylogeny

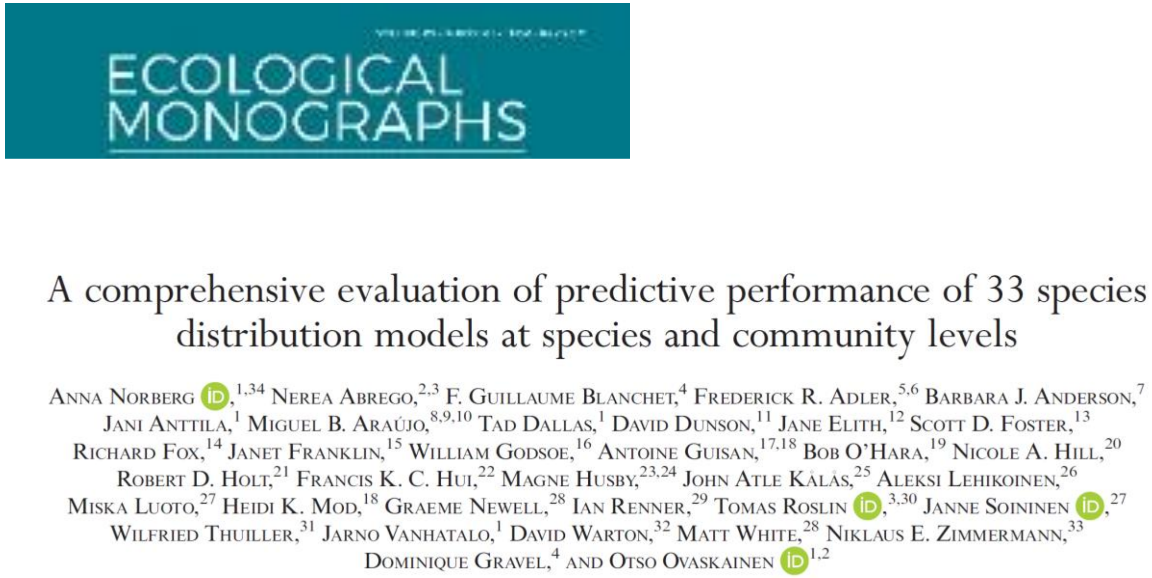
MODEL	N1	N2	N3	B1	B2	B3
HMSC	X	X	X	X		
BORAL	X	X		X		
GJAM				X		
JDSDM	X			X		
MISTN				X		
GLLVM	X	X		X		
BC				X		
SSDM						

Data types utilized by joint species distribution models



MODEL	Y	X	T	C	S
HMSC	X	X	X	X	X
BORAL	X	X	X		
GJAM	X	X	X		
JDSDM	X	X			X
MISTN	X	X			
GLLVM	X	X	X		
BC	X	X			
SSDM	X	X			X

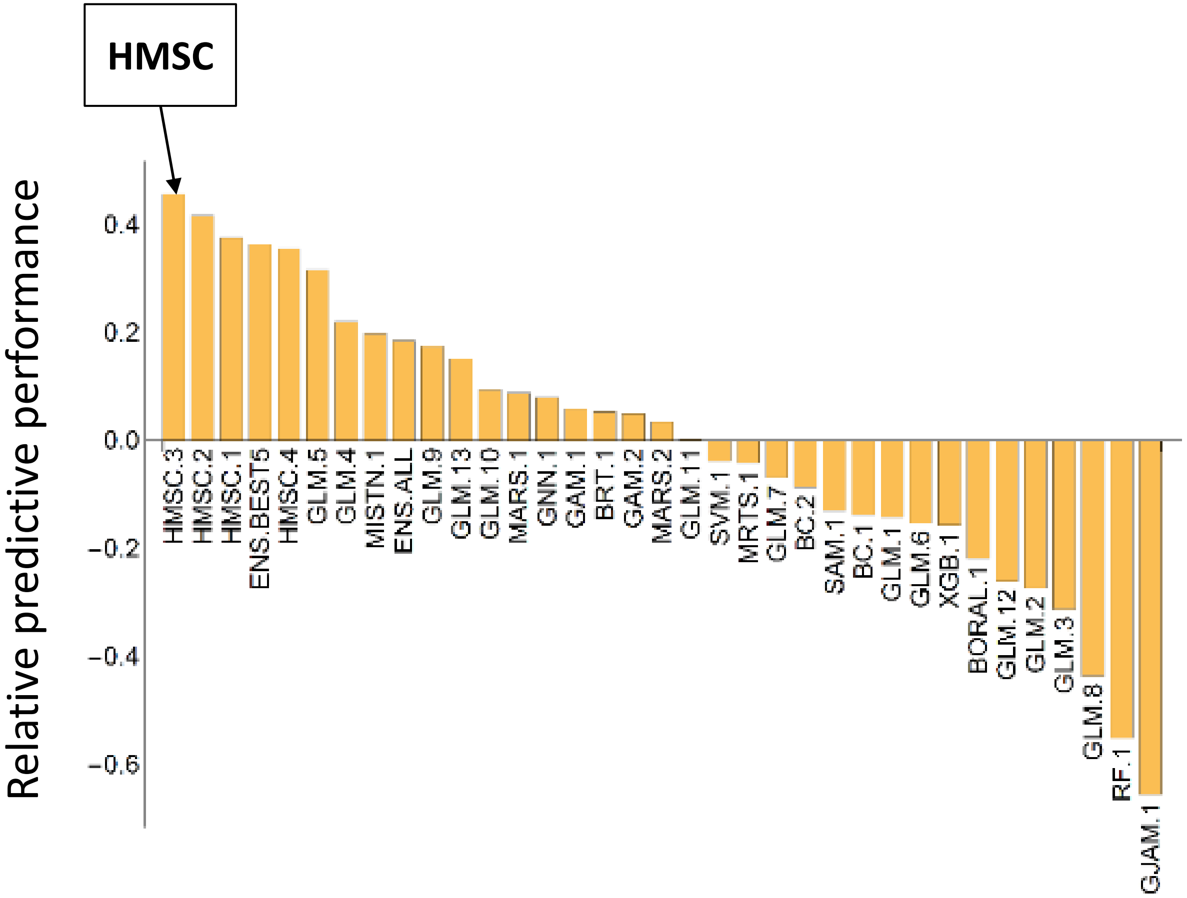
Predictive performance of (joint) species distribution models



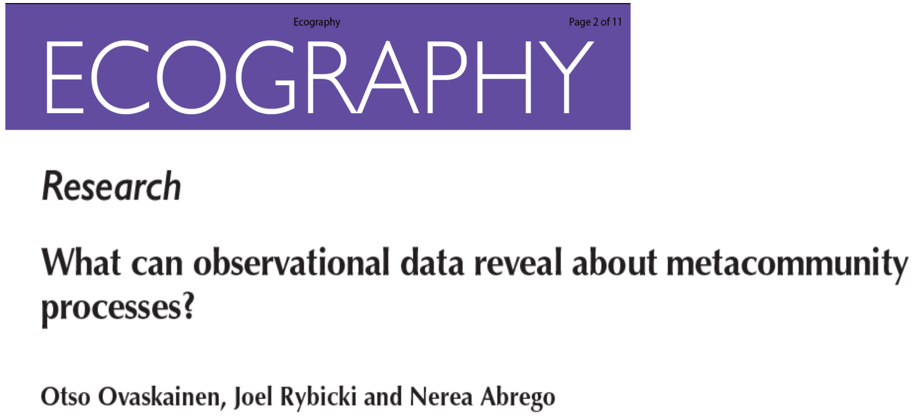
ECOLOGICAL MONOGRAPHS

A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels

Anna Norberg,[1,34] Nerea Abrego,[2,3] F. Guillaume Blanchet,[4] Frederick R. Adler,[5,6] Barbara J. Anderson,[7] Jani Anttila,[1] Miguel B. Araújo,[8,9,10] Tad Dallas,[1] David Dunson,[11] Jane Elith,[12] Scott D. Foster,[13] Richard Fox,[14] Janet Franklin,[15] William Godsoe,[16] Antoine Guisan,[17,18] Bob O'Hara,[19] Nicole A. Hill,[20] Robert D. Holt,[21] Francis K. C. Hui,[22] Magne Husby,[23,24] John Atle Kålås,[25] Aleksi Lehikoinen,[26] Miska Luoto,[27] Heidi K. Mod,[18] Graeme Newell,[28] Ian Renner,[29] Tomas Roslin,[3,30] Janne Soininen,[27] Wilfried Thuiller,[31] Jarno Vanhatalo,[1] David Warton,[32] Matt White,[28] Niklaus E. Zimmermann,[33] Dominique Gravel,[4] and Otso Ovaskainen[1,2]



How well do JSDMs do in terms of inference?



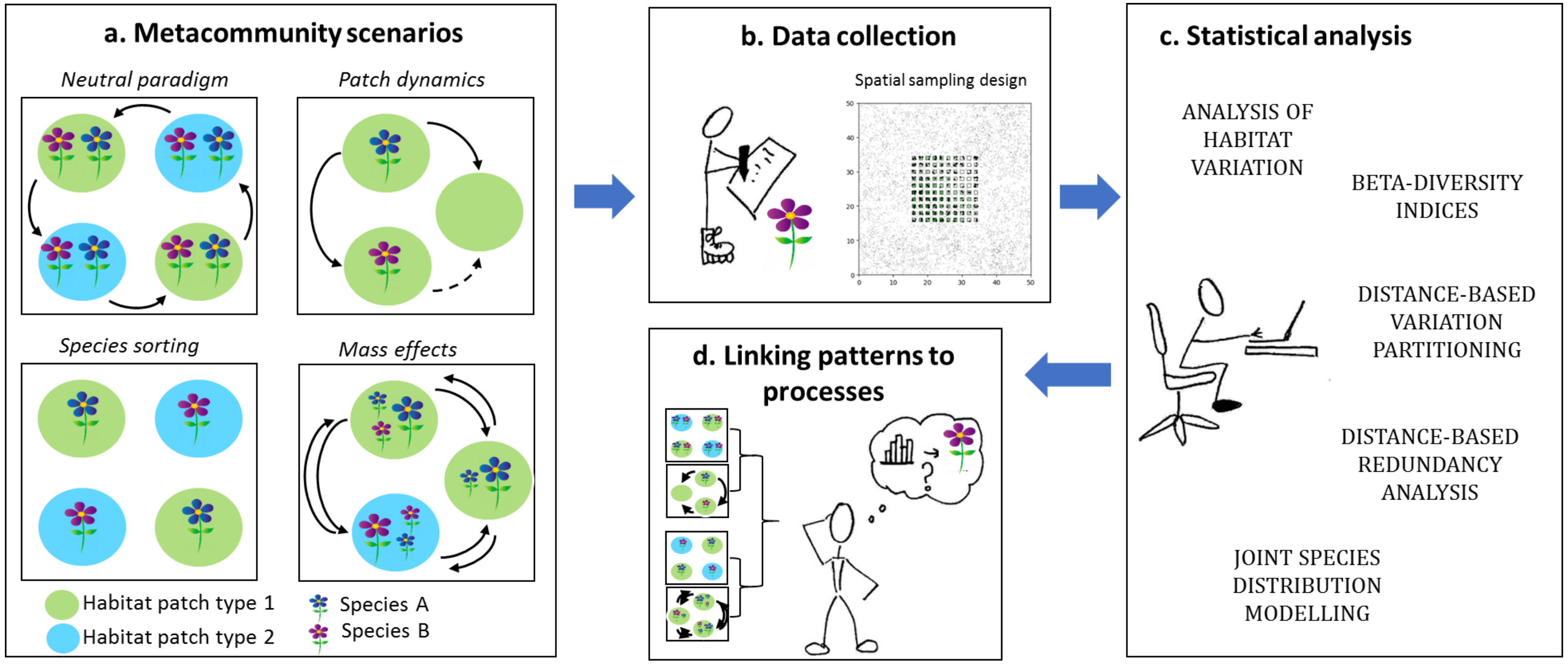
Ecography Page 2 of 11

ECOGRAPHY

Research

What can observational data reveal about metacommunity processes?

Otso Ovaskainen, Joel Rybicki and Nerea Abrego



Statistical approach	HAB	BETA	db-VP	db-RDA	JSDM	ALL
C1. presence versus absence of gradient	0.94		0.88	0.82	0.79	0.97
C2. patchy versus continuous landscape	1	0.99	0.78	0.97	0.92	1
C3. uniform versus varying patch quality						
C4. dynamic versus.static patches					0.78	0.82
C5. slow versus.fast patch turnover					0.8	0.93
C6. generalists versus specialists		0.76			0.9	0.96
C7. partial versus strict specialists						0.81
C8. variation versus no variation in dispersal strategy						
C9. short versus long dispersal		0.87		0.8	0.75	0.93
C10. missing versus not missing covariates					0.87	0.9

JSDM outputs were most informative in disentangling which underlying community assembly mechanisms generated the patterns in the data

Conclusions

- The processes underlying community assembly are complex, and we have typically not data directly on them, merely on patterns generated by the processes.
- We need statistical methods that can relate data on patterns to the underlying processes.
- Joint species distribution models (JSDMs) are one promising approach for doing so. They have been shown to be efficient both for inference and prediction.
- JSDMs include single species distribution models and model-based ordinations as special cases, and thus they are merging two fields that have long developed in isolation from each other.
- Methods development with JSDMs is currently very rapid, with several research groups developing related approaches.
- In this course we will learn how to apply the R-package Hmsc!