Table of Contents

Organisation

- Break-out groups Wednesday and Thursday for Europe
- The course covers more or less the whole book

Introduction to community ecology & Typical data collected in community ecology

Community ecology: Assemblage and their interactions

John H. Lawton (1999): Are there general laws in ecology?

Although conceptual frameworks have been developed in the last decades

HMSC follows the definition from *Fauth et al. (2016)*: Simplifying the Jargon of Community Ecology: A conceptual approach

Prevailing theories: - Assembly rules framework:

$\rightarrow$ restricted species combinations to which competitive interactions can lead (see *Diamond (1975)*)

$\rightarrow$ nowadays assembly processes: filters act at various spatial scales

$\rightarrow$ ecological assembly rules: environmental filtering (β), biotic filtering (Ω), dispersal (α), stochasticity (R^2)

$\rightarrow$ species traits (γ) influence Ω, α and β

$\rightarrow$ phylogeographic assembly rules

- Metacommunity theory
- Vellends theory of ecological communities

$\rightarrow$ HMSC components can be related to ecological theory

Four archetypes of metacommunity theory: - Neutral paradigm - Patch dynamics - species sorting - mass effects

Read: *Ovaskainen et al: What can observational data reveal about metacommunity processes?*

Typical data collected in community ecology?

Lab experiments & field experiments (manipulative) & field data (non-manipulative)

$\rightarrow$ non-manipulative field data are shaped by full complexity of assembly processes

$\rightarrow$ Abundances/rel. Abundances, presence-absence, spatial-temporal design, environmental covariates, traits, phylogeny

Overview of the structure of HMSC

Theoretical

Multivariate hierarchical generalized linear mixed model fitted with Bayesian inference

HMSC is a correlative model!

- HMSC is a joined distribution modelling approach, not stacked:

  $\rightarrow$ There is random variation in pairs of species (co-occurrences, species associations)

  $\rightarrow$ Share information across species while modelling

- **Q:** Species associations are modelled as random effect:

  $\rightarrow$ Dimensionality problem (cannot put too many as predictors/as fixed effects)

  $\rightarrow$ Species occurrences are the response variable, HMSC creators refrain from putting species to explanatory variable part

Species associations are modelled using a latent approach, otherwise to computational intensive

Practical

Finding relevant covariates if many covariates are present. Approach Mirkka Jones:

In fitting pilot models, where people did not suggest 2-3 X covariates themselves, I did the following as a quick route to selecting from among numerous X-predictors: I calculated the correlations between the X-predictors and the first couple of axes of an ordination of community composition based on the full Y-matrix. I included those X variables in the hmsc model that had a high positive or negative correlations with the first couple of ordination axes (but excluding any X covariates that covaried strongly).

1) Setting model structure and fit the model

   $\rightarrow$ call *Hmsc()*

   $\rightarrow$ model is defined but not fitted! (unlike *lm()* for example)

   $\rightarrow$ prior: default prior applies too most datasets (according to Ovaskainen)

   $\rightarrow$ call: *sampleMcmc()* for model fitting. Start always with a small amount of thinning!

2) Examining MCMC convergence (how good is the sample of the posterior distribution

   $\rightarrow$ represent the posterior distribution)

   $\rightarrow$ Use *library(coda)* utilities (loaded when loading HMSC package)

   $\rightarrow$ Effective Size (should be close to actual sample size), Gelman diagnostics, Trace plots

   $\rightarrow$ Run the whole pipeline also with bad convergence to obtain preliminary results

3) Evaluating model fit and comparing model

   $\rightarrow$ Bigger issue than variable selection: Should one include traits, phylogeny, spatial structure and how?

   $\rightarrow$ WAIC (widely applicable AIC)

   $\rightarrow$ Predictive power (how well does the model predict/explain the data?). Gives insight in how models differ (prediction difference, bias etc.)

   $\rightarrow$ CV used for predictive power (the more variables one includes the less good predictions will be with this approach?), without CV for all data than model fitted for all data for explanatory purposes $\rightarrow$ Output: $TjurR^2$, AUC

   $\rightarrow$ CV can be done across sampling units and also across **species**! (conditional predictions)

$\rightarrow$ Predictive values can be checked for variance in residuals

4) Exploring parameter estimates

$\rightarrow$ One joint posterior distribution for all parameters

$\rightarrow$ Could use tools of the coda package, or Hmsc functions, e.g.: *plotBeta(m, post, supportLevel)*

positive, zero or negative slope of beta parameters (species niches); supportLevel: how much posterior probability?

$\rightarrow$ Association plots: *corrplot(computeAssociations(m))*

$\rightarrow$ Variance partitioning: how much variation is explained by different compartments of the model? (*computeVariancePartitioning(m)*)

$\rightarrow$ How is variance partitioning calculated? See section 5.5. in *Joint Species Distributions Modelling*

5) Making predictions

5.1 prepare predictions: predictor variables of env. covariates + predictor values of spatio-temporal context

$\rightarrow$ e.g.:

Gradient <- constructGradient(m, focalVariable)

pred <- predict(m, Gradient)

plotGradient(m, Gradient, pred, measure, showData)

5.2 Make predictions (Y new) with uncertainty estimate (95 % credible interval):

- General:

  $\rightarrow$ Construct gradient of new data for variable of interest (focalvariable) to make predictions on

  $\rightarrow$ The non.focalvariables are the other variables needed for predictions (e.g. *list(hab = 1)* uses the most common type)

  $\rightarrow$ predict model on gradient with *predict()* function (creates posterior predicted distribution, see MCMC)

  $\rightarrow$ expected = TRUE: predicting probabilities; expected = FALSE: predicting realisations (e.g. Presence-absence) mean over this gives the posterior mean

- Spatial predictions:

  $\rightarrow$ New X Data & new spatial data

  $\rightarrow$ Prediction for each posterior sample

4

5.3 Post-processing and interpretation (species richness, community-weighted mean traits, . . .)

How HMSC models variation in species niches?

- Full HMSC case: multiple species, environ. covariates, traits, phylogeny

- Linear predictor is now a matrix! L_{ij} , $i =$ sampling unit, $j =$ species

- β is also a matrix

- Environ. variables do not depend on species (i.e. x_{ik}, where $k =$ covariate, but it can be changed to be species-specific

- Joint species distribution model: model is fitted simultaneously, so species models are dependent on each other

- How are species niches structured? HMSC implements continuous variation in species niches (in contrast to *Hui et al.* species archetype models)

- Parameters estimated $\rightarrow$ used to estimate Variance-Covariance matrix ?

  $\rightarrow$ Borrowing information from other species

How to utilize traits and phylogeny?

See chapter 6

Traits

- Modelled as regression to species traits

- Expected value μ is specific to the trait expressed by each species (t_{jl}, $t =$ trait value) and can thus be modified by the trait information

- γ_{kl} parameter: How species niches depend on species traits (one parameter for community as response to covariate, e.g. climatic conditions)

Phylogeny

- Phylogenetic tree or taxonomic table

- Correlation between species in phylogenetic trees: That's the fraction of shared evolutionary history. Can be expressed in a correlation matrix.

- Residual variation in species niches (Matrix M, expected species niches) is modelled as a weighted average of correlation matrix and (modified) identity matrix (W)

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

- ρ measures the strength of the phylogenetic signal in species niches