Evaluating Model Fit and Selecting among Multiple Models

DO NOT try all possible model combinations and apply "blindly" model selection tools!

9.1 Preselection of Candidate Models

While we present several methods that can be used to select the 'best model' based on quantitative criteria, we do not recommend applying the blind strategy of running all possible model variants through an automated model selection pipeline. Instead, we emphasise that an important task for the ecologist applying HMSC – or any statistical modelling approach – is to use their prior knowledge of the system to restrict the model structure and the sets of candidate variables to be included. The first reason for this is to ensure that the models make sense

- You will risk being distracted from addressing your study questions in a focused way and entering a huge fishing expedition. Do you need all those candidate models to address your study questions?

- "Trying out all models and comparing them" is usually not feasible.

Why it is difficult to compare all possible models?

Example: HMSC model with 100 species, 5 candidate environmental predictors, 3 traits with phylogenetic and spatial data.

- **Environmental covariates.** Linear and quadratic effects: 243 candidate models. Also interactions among predictors: 59,049 candidate models.
- **Traits.** Linear and quadratic effects: 27 candidate models.
- **Phylogeny.** Included or not, 2 candidate models.
- **Spatial random effect.** Included or not, 2 candidate models.

In total: 59,049 x 27 x 2 x 2 = 6.4 million models.

Why it is difficult to compare all possible models?

We can do variable selection for environmental variables separately for each of the 100 species!

In total: (59,049)^100 x 27 x 2 x 2 models =

176185820017003594424751726751564709415710304986914103344960595228717178927946460504785680813156761744471122088410946590680780296391113801605719974769975615015694477435795223423074575191510412214844324166129753116441375723896031363794821398113343803292167786164387095322036662226909778251900406279847022820986019807161684410644975153284363186771260083520294710564518869103428122084286628829643618674752307429430028138561312071954544047652787826683269041811590433387440350300672558298862582783801233400420513205293554732394003685497183693494034407867714486707717332594535663118149101086049304275557976603646302163764831724865414740847042893523016073395680812179755659185673342916646992242620419139353605276891981277428148880138319232537201125217681511972139758215341850515592100288033747644168505482088786116735596303234501435261058213413780905856531592805279062905253858004878975217245072663474507195209396534928474557199531030405101984097009674210577912258417591478393940369612794637804512146040164881008309079727852860716326271587583263743636273331474645656530597382696063955839545491910802573783649704229194706201311136818352315011941037313442759266881333870768759898742601711279843805669019429560889525394695308226793162722165660059053198354699553445512906832540988139997490661667356183178561623357698181150240900897751871034427997126526186984654059383771423353646415505507857850189898072741794194277268627391174768228263134679681882167883166912459521607226897578915933651122543283767840130363934030849415966567759052563372847694721417174921306581316048284535442258285315107535170497350485347625650838438219785261642156131360556317958083911530130154372799366688252344320652718656728363163864170126561795975016032560842044365358749268705278739306027015857241743709556141074151750120277808085920464806934510772875872673679461950424445979434582440133470324827573791936450437396559997179266880948261095793222978086955426974661721691861102208849520357470801285878503484392108897755682306834416263776202316659435272353128435088333934215706597062868436979096164727020733394374272528732723844543899473428972577292657674585840880693434881396409751953181863878254184872178371935676697041817725176373641573484408063946054486842774933275147907473292812153468394992813258073229866003742458064508053752406952142279058225999371938109635873233000316297638326574827289445992376726171271034216223319631613011430135075520096881106295069278728659891755465501441730039014004362567180561927721433411503855542714710282955872304202760793744040622010442305620366421183890111155862541551726699327540366277696996712839371151020547106086018237064172533381174027314171649947184805314631013569700788127698144349435101847728808026183415568056142994075130361556746394608488645023915026024574036918674376157178199128269125494179050478378617616328326703480876862261682749856957080044305560669184764843449835371495624254466734685844158821913160359538809256326631322953347227381292528351972230550031095927159840140001875600543518140194480267666248458213190952713278614679121706755907149571815115728195261024577678885532729752679162071591322431495674223408700921144055121171184295071574285910067144645540350807810686679168820205938831789225058898877765639975017716131699794500776337066257847875805521634677805225348828607876237893702300080837700431920800800575369962041927801497530945224374338014394674655536085217047744610603672646916259627374126169476300138691951506354598599222112777131771447779080220483196330109128004806982730435767328958244743904127294386411836148428537179508862979733431240354864439180442966548078388218834182971309803572411996891788653896495568197072590694089619314653538655228940524400894261390512420785198681997998526946697564743495626364016997405217567274094429812305551965932621615535687229745451492926189784480297727702031473160459173333815448641633640808319422857431600353797504144356474527506142912986332731190153391343711778922756878213797158326774590608431292099391617804190774340265563489268494623195939873887226410879291517792321750312742473984599371799877442369797768601149257428319137346177880246598243990493657126066830381120554900364945918401126808549275276268035460195985640928658848613836574638386269700914807242321675105850970263883083980114265121348754085555879762577365869431426749190367069975892233454152758813016136172881051306725444379004173085343622404249744909842458624459718164471798692741655673630325771051123461419068247015949185172898953368742547890036821925317358315481136767889899484534147989109619594032719605020254701613261832862122223532902755508318046573836812851105772839737440286492966615708444774056873723264837696358267318549102197335001500046131494854757286176642728160237624199544407167783170480558844351963007214755546208179536230763760010 8 models

Pre-selection of candidate variables (both environmental and species traits)

- Include only those that make ecological sense, not all just because they were measured.

- Remove highly correlated variables by choosing only one of them, or using summary variables such as PCAs or means (but see for RRR later).

- You may run a preliminary analysis e.g. by RDA or species richness analyses to see which variables seem to explained a largest part of the variation.

Model/variable selection strategies

- You may consider also alternative model structures, not only alternative sets of predictors (e.g. models with or without spatial random effects).

- You may decide to keep some of the candidate variables in any case (whether supported or not), e.g. those that control for variation in sampling effort.

- You may select variables hierarchically, e.g. first asking whether “some of the five climatic variables” should be included in the model (models with all and without any climatic variables), and if yes, then examining which of those five explain most variation to select only a subset.

Specific techniques implemented in HMSC for model comparison / variable selection

- Cross-validation (which model has the highest predictive power)?
- WAIC (which model has the lowest WAIC?
- Variable selection by spike and slab prior
- Reduced rank regression

Variable selection by spike and slab prior

"Usual" prior:

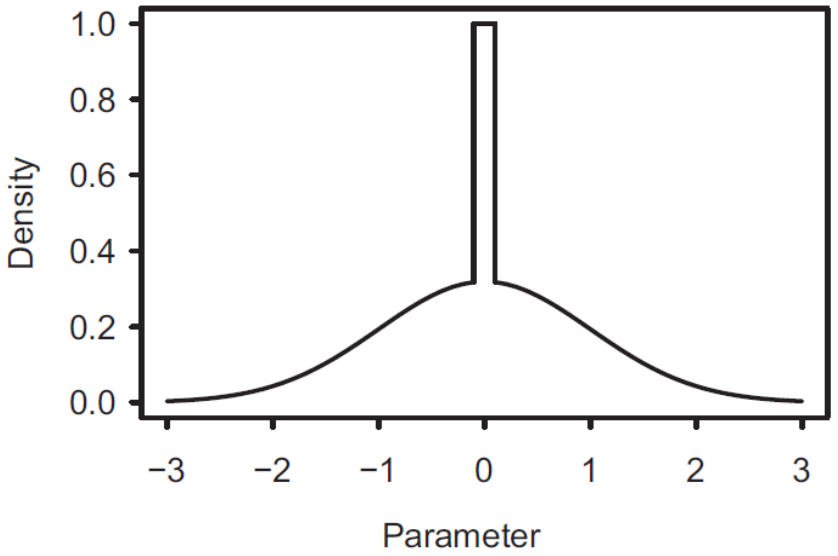
$$\beta \sim N(0, \sigma^2)$$

Spike and slab prior:

$$\beta = p\hat{\beta}$$

$$p \sim \text{Bernoulli}(q)$$

$$\hat{\beta} \sim N(0, \sigma^2)$$



HMSC implements spike and slab prior for the β-parameters

9.4.2 Simulated Case Study with HMSC

```
Hmsc(Y=...,
  XData=...,
  XFormula=...,
  XSelect = ...,
  ...)
```

```
qq = 0.1 #prior probability for a covariate to be included

#1: No variable selection
XSelect.FULL = NULL

#2: Variable selection jointly for all species
XSelect.JOINT = list()
for (k in 2:nc){
 covGroup = k
 spGroup = rep(1, ns)
 q = rep(qq, max(spGroup))
 XSelect.JOINT[[k-1]] = list(covGroup = covGroup,
  spGroup = spGroup, q = q)
}

#3: Variable selection separately for each species
XSelect.SEPARATE = list()
for (k in 2:nc){
 covGroup = k
 spGroup = 1:ns
 q = rep(qq, max(spGroup))
 XSelect.SEPARATE[[k-1]] = list(covGroup = covGroup,
  spGroup = spGroup,q = q)
}
XSelect = list(XSelect.FULL, XSelect.JOINT, XSelect.SEPARATE)
```

HMSC implements spike and slab prior for the β-parameters

9.4.2 Simulated Case Study with HMSC

Explanatory power:

```
##                  Model FULL Model JOINT Model SEPARATE
## Data FULL          0.929       0.822          0.729
## Data JOINT         0.712       0.470          0.603
## Data SEPARATE      0.713       0.286          0.587
```

Predictive power:

```
##                 Model FULL Model JOINT Model SEPARATE
## Data FULL         0.375       0.228          0.263
## Data JOINT        0.154       0.284          0.074
## Data SEPARATE     0.099       0.067          0.106
```

Reduced Rank Regression (RRR)

- PCA (and other such techniques) can in a first step be applied to the original covariates to reduce their dimensionality & collinearity, and then in a second step use a small set of dominating PCAs as predictors in **X**.

- RRR combines these two steps so that instead explaining most variation in **X**, the dimension-reduced covariates explain as much variation as possible of **Y**

Reduced Rank Regression (RRR)

$$y_{i,t} = c_i + \sum_{j=1}^{m} \alpha_{i,j} y_{j,t-1} + e_{i,t},$$

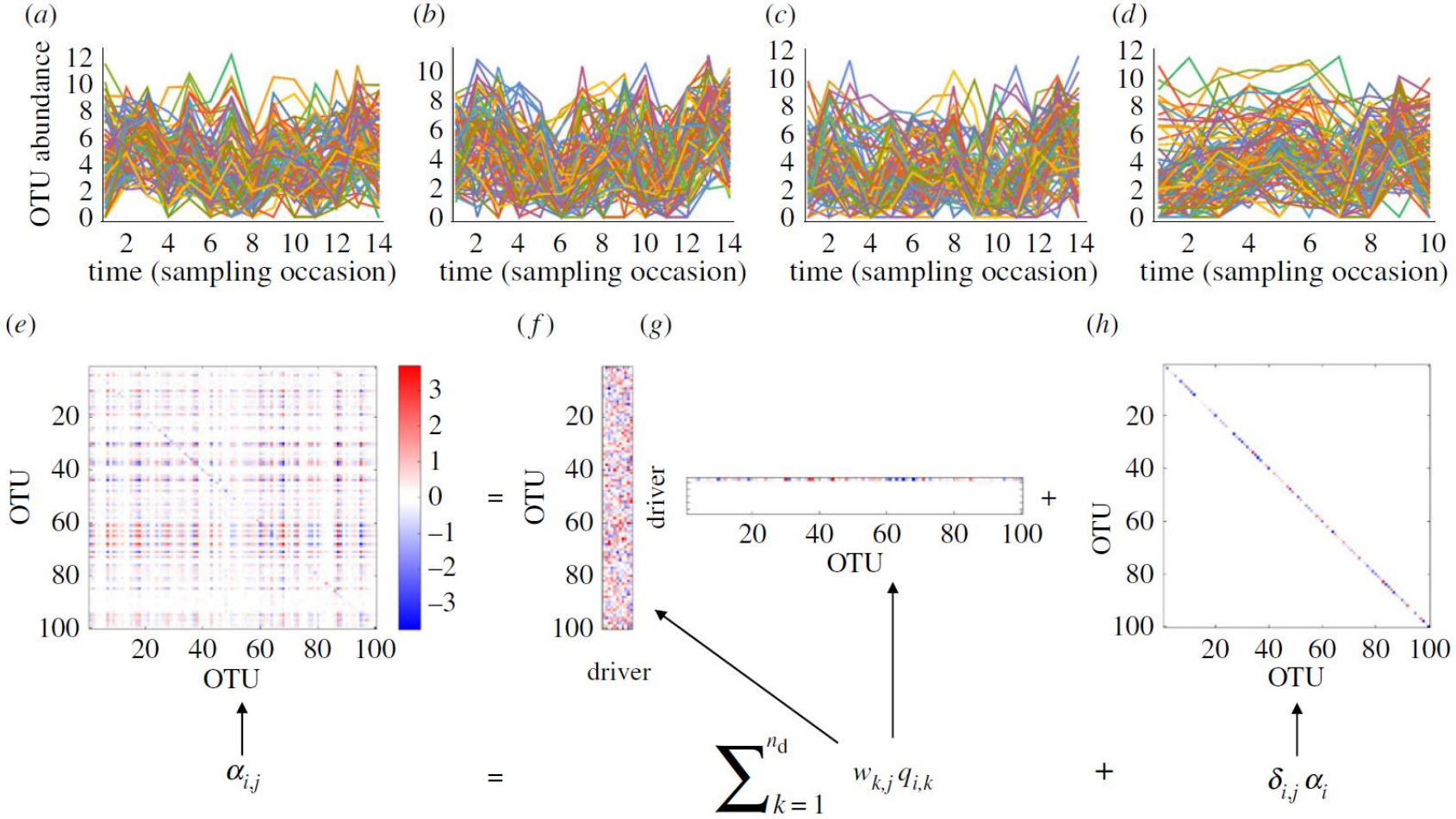
PROCEEDINGS B

rspb.royalsocietypublishing.org

Research

How are species interactions structured in species-rich communities? A new method for analysing time-series data

Otso Ovaskainen[1,2], Gleb Tikhonov[1], David Dunson[3], Vidar Grøtan[2], Steinar Engen[4], Bernt-Erik Sæther[2] and Nerea Abrego[2,5]



Reduced Rank Regression (RRR)

Reduced covariates

Weights (estimated)

Original covariates

$$x_{ik} = \sum_{l=1}^{n_c} w_{kl} \tilde{x}_{il}$$

$$k = 1, \dots, n_c^{RRR}$$

Prior parameters for RRR

$$w_{kl} \mid \phi_{kl}^{RRR}, \delta \sim N\left(0, \left(\phi_{kl}^{RRR}\right)^{-1}\left(\tau_k^{RRR}\right)^{-1}\right), \tau_k^{RRR} = \Pi_{h=1}^{k}\ \delta_h^{RRR} \tag{9.9}$$

$$\phi_{kl} \mid \upsilon \sim \mathrm{Ga}(\upsilon^{RRR}/2, \upsilon^{RRR}/2) \tag{9.10}$$

$$\delta_1 \sim \mathrm{Ga}\left(a_1^{RRR}, b_1^{RRR}\right), \delta_h \sim \mathrm{Ga}\left(a_2^{RRR}, b_2^{RRR}\right) \text{ for } h \geq 2 \tag{9.11}$$

As default values of the prior parameters, Hmsc assumes $\upsilon^{RRR} = 3$, $a^{RRR} = (1, 50)$ and $b^{RRR} = (1,\ 1)$.

HMSC implements RRR for the β-parameters

9.5.1 Simulated Case Study with HMSC

Hmsc(Y=…,
XData=…,
XFormula=…,
XRRRData=…,
XRRRFormula=…,
ncRRR=…,
…)

```
models = list()
for(dataset in 1:3){
    tmp = list()
    for (model in 1:3){
        switch(model,{
            m = Hmsc(Y = Y[[dataset]], XData = XData,
              XFormula = ~., distr = "normal")
        },
        {
            pc = princomp(XData)
            XData.PC = data.frame(pc$scores[,1])
            m = Hmsc(Y = Y[[dataset]], XData = XData.PC,
              XFormula = ~., distr = "normal")
        },
        {
            m = Hmsc(Y = Y[[dataset]], XData = XData, XFormula = ~1,
              XRRRData = XData, XRRRFormula = ~.-1, ncRRR=1,
                distr = "normal")
        }
        )
```

HMSC implements RRR for the β-parameters

9.5.1 Simulated Case Study with HMSC

Explanatory power:

```
##                    Model FULL    Model PC   Model RRR
## Data FULL             0.929       0.108       0.431
## Data PC               0.667       0.374       0.392
## Data RRR              0.797       0.070       0.607
```

Predictive power:

```
##                  Model FULL Model PC Model RRR
## Data FULL             0.491   -0.113     0.093
## Data PC               0.155    0.259     0.228
## Data RRR              0.300   -0.011     0.416
```