

Joint species modelling

– BEAUTIFUL IN THEORY, TRICKY IN PRACTICE

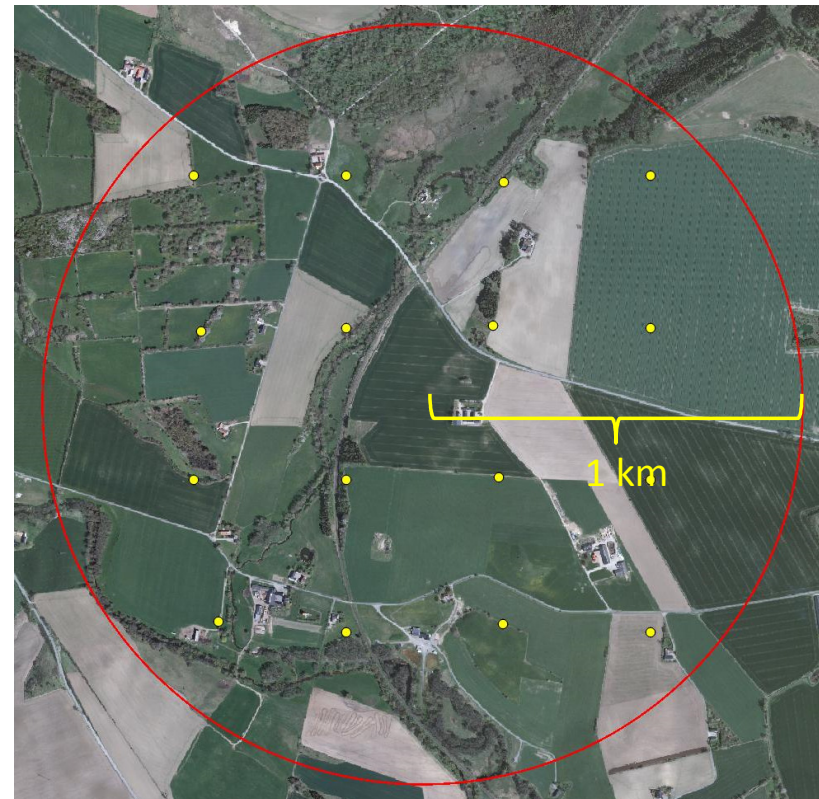
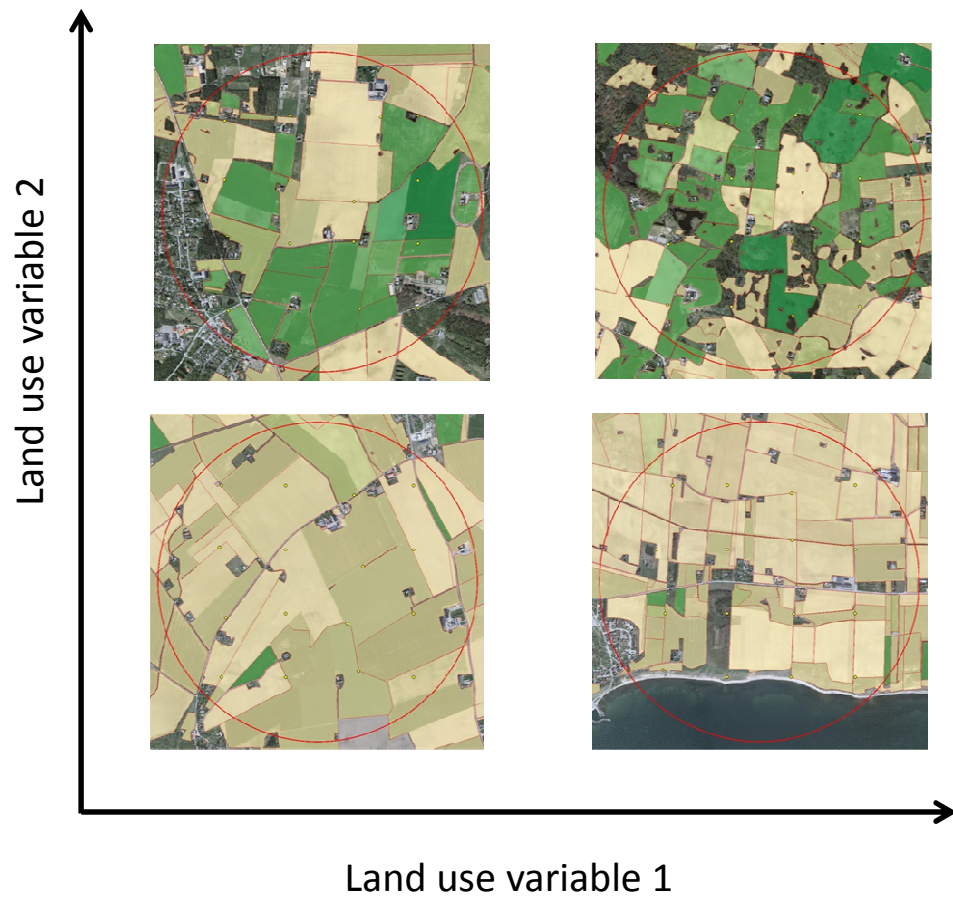
Martin Stjernman
Biodiversity, Lund University



Background

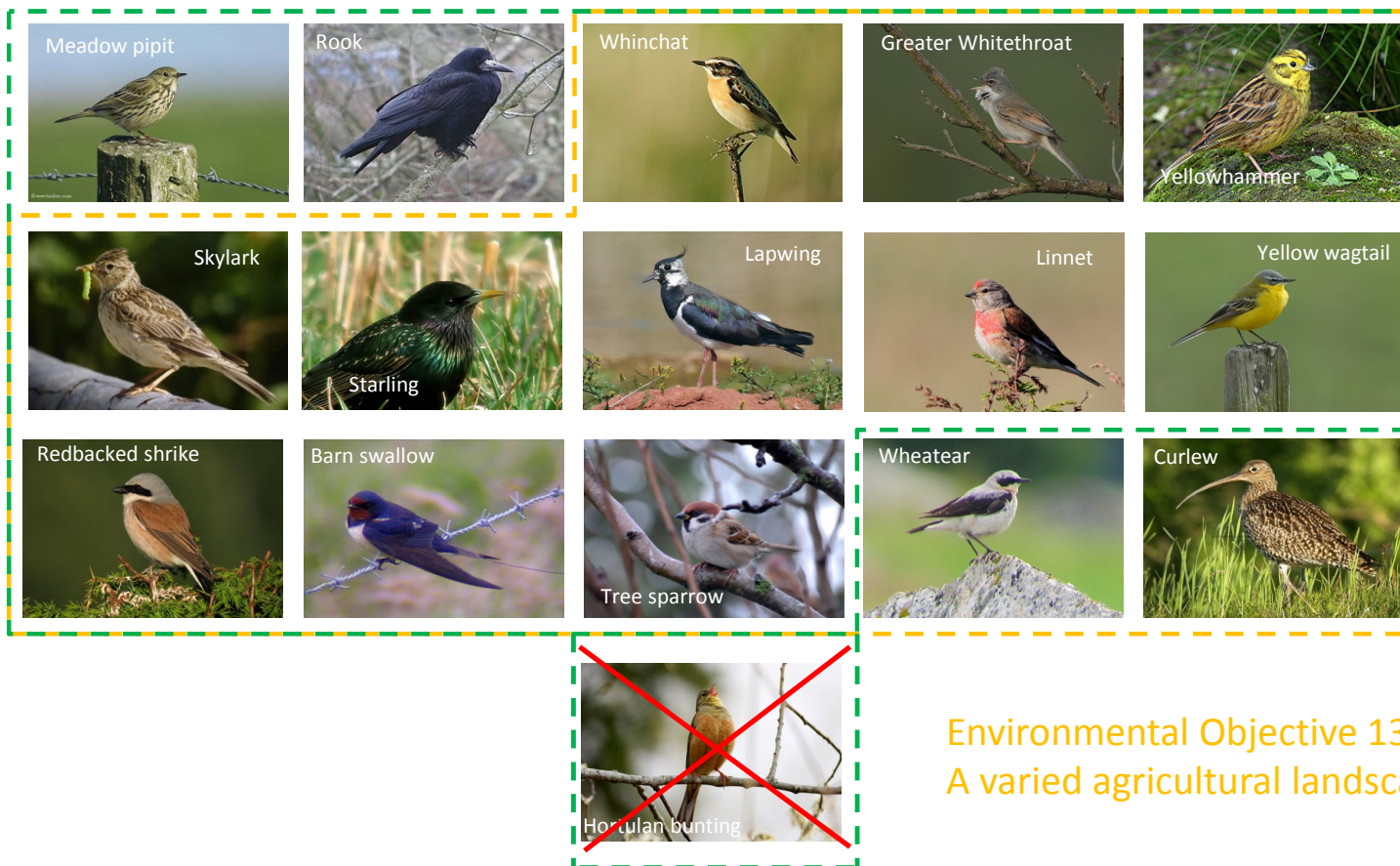
- We sometimes get asked by the Swedish Board of Agriculture –
"What will happen to farmland birds if we change agricultural land use?"
- Approach:
 - Build a model associating bird abundance to land use
 - Combine the model with available data and estimate the parameters/coefficients/variables of the model
 - Apply the model on new land use forecasts according to scenario

What we have – a nice design



What we have – counts of farmland birds

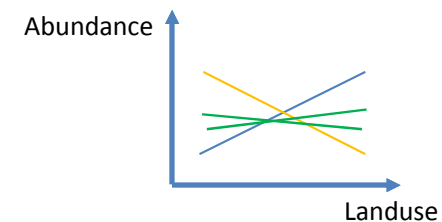
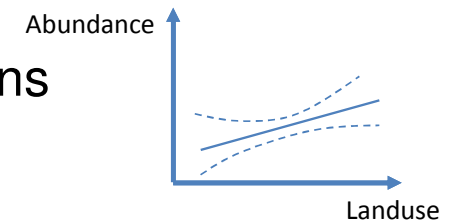
Farmland Bird Index



Environmental Objective 13:
"A varied agricultural landscape"

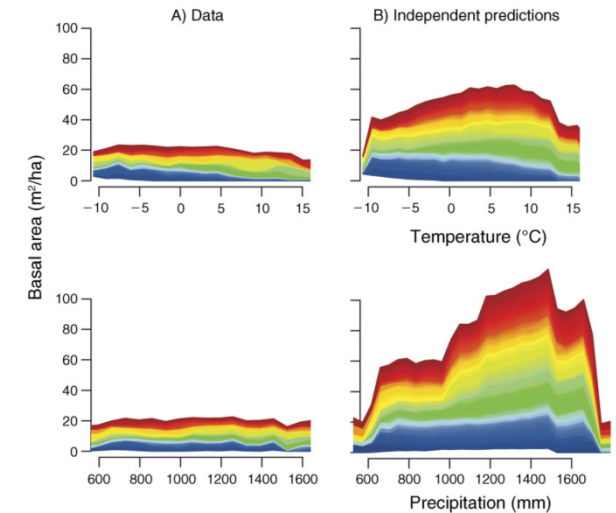
Requirements/Complications

- We need a statistical model (a function) $y = X\beta + Zb$
 - estimated coefficients/parameters which can be applied to new environmental data – make predictions!
 - handle hierarchical design structure (e.g. repeated sampling, different observers)
- We want to be able to describe the uncertainty around predictions
- There are 15 farmland bird species of interest
 - Species may respond differently to changes in land use
 - Species may interact (competition or facilitation)

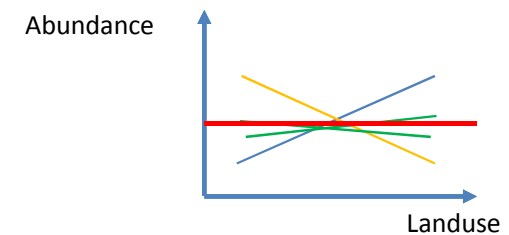


Alternatives (in "the old days")

- Separate models for each species
 - Many models and some rare species may not be possible to analyse
 - Combining results from independent models may yield unrealistic predictions
- Combine species into one response variable (e.g. calculate a diversity measure) and use one model
 - Differential responses to land use among species get "hidden"
- Apply multivariate statistics (ordination etc.)
 - Not model-based (parameters?, non-normal data?, hierarchical designs?)
 - Difficult to use for predictions
 - My ignorance



Clark, J.S. et al, 2014. *Ecological Applications* 24, 990-999.



Joint (Species Distribution) Models – the beautiful solution?

Ecology, 91(9), 2010, pp. 2514–2521
© 2010 by the Ecological Society of America

Modeling species co-occurrence by multivariate logistic regression
generates new hypotheses on fungal interactions

OTSO OVASKAINEN,^{1,3} JENNI HOTTOLA,^{1,2} AND JUHA SIITONEN²

¹Department of Biosciences, University of Helsinki, Viikinkaari 1, FI-00014 Helsinki, Finland

²Finnish Forest Research Institute, Vantaa Research Unit, P.O. Box 18, FI-01301 Vantaa, Finland

Ecological Applications, 24(5), 2014, pp. 990–999
© 2014 by the Ecological Society of America

More than the sum of the parts: forest climate response from joint
species distribution models

JAMES S. CLARK,^{1,2,4} ALAN E. GELFAND,² CHRISTOPHER W. WOODALL,³ AND KAI ZHU¹

¹Nicholas School of the Environment, Duke University, Durham, North Carolina 27708 USA

²Department of Statistical Science, Duke University, Durham, North Carolina 27708 USA

³USDA Forest Service, Northern Research Station, St. Paul, Minnesota 55108 USA

Methods in Ecology and Evolution



Methods in Ecology and Evolution 2014, 5, 397–406

doi: 10.1111/2041-210X.12180

Understanding co-occurrence by modelling species
simultaneously with a Joint Species Distribution Model
(JSMD)

Laura J. Pollock^{1†}, Reid Tingley^{1†}, William K. Morris¹, Nick Golding², Robert B. O'Hara³,
Kirsten M. Parris¹, Peter A. Vesk¹ and Michael A. McCarthy^{1*}

Trends in Ecology & Evolution

CellPress

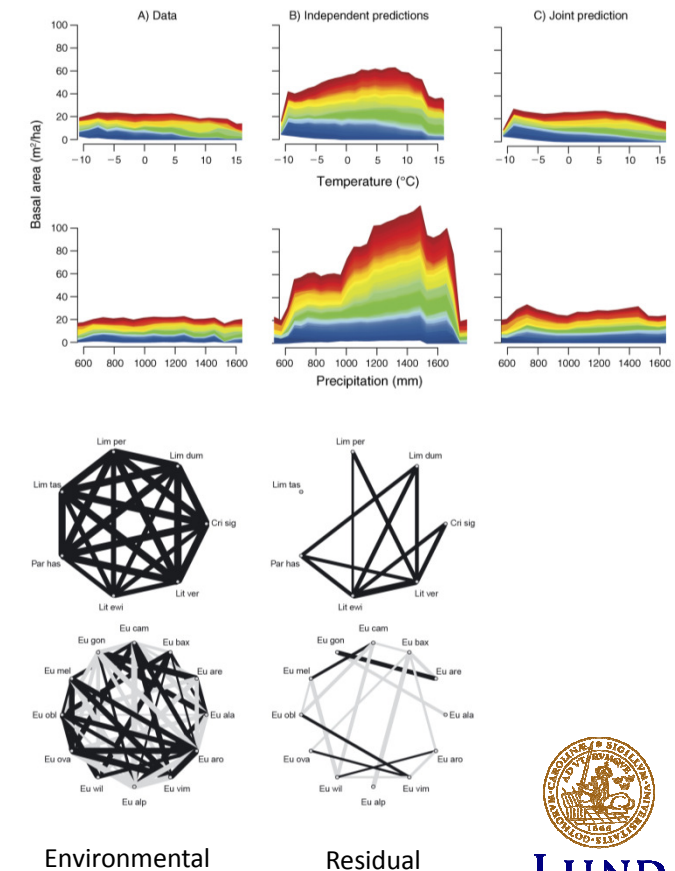
Review

So Many Variables: Joint
Modeling in Community
Ecology

David I. Warton,^{1,*} F. Guillaume Blanchet,² Robert B. O'Hara,³
Otso Ovaskainen,^{4,5} Sara Taskinen,⁶ Steven C. Walker,² and
Francis K.C. Hui⁷

Joint (Species Distribution) Model –the beautiful solution?

- All species modelled simultaneously in one model
- Species specific parameter estimates
- Co-variance between species explicitly modelled
- Rare species "borrow information" from common ones
- Model based
 - + Hierarchical structures handled "easily"
 - + Non-normality "allowed"
 - + Prediction straight-forward
- Bayesian approach
 - + Uncertainty a natural part
 - + Flexible
 - + Example code available



The joint species distribution model (JSDM)

$$\Pr(Y_{ij} = 1) = \Pr(Z_{ij} > 0), \text{ for } i = 1, \dots, n; j = 1, \dots, J$$

$$Z_i \sim N_j(X_i B^*, \Sigma)$$

$$B_{jk}^* \sim N(\mu_k, \sigma_k), \text{ for } k = 1, \dots, K,$$

eqn 1

J = number of species

n = number of sites

K = number of predictors

$$\text{probit}(\Pr(Y_{ij} = 1)) = B_{j1} + B_{j2}X_{i2} + B_{j3}X_{i3} + \dots + B_{jK}X_{iK} \quad \text{eqn 2}$$

N_j is a J-dimensional multivariate normal with mean

$$M_i = X_i \mathbf{B}$$

and (co)variance

$$\Sigma = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1j} \\ \vdots & \ddots & \vdots \\ \sigma_{j1} & \cdots & \sigma_{jj} \end{bmatrix} \sim IW(\nu, ID)$$



Obstacle 1 – Y is not presence/absence, it is counts

Solution: log-normal Poisson

$$Y_{ij} \sim \text{Poisson}(\lambda_{ij})$$

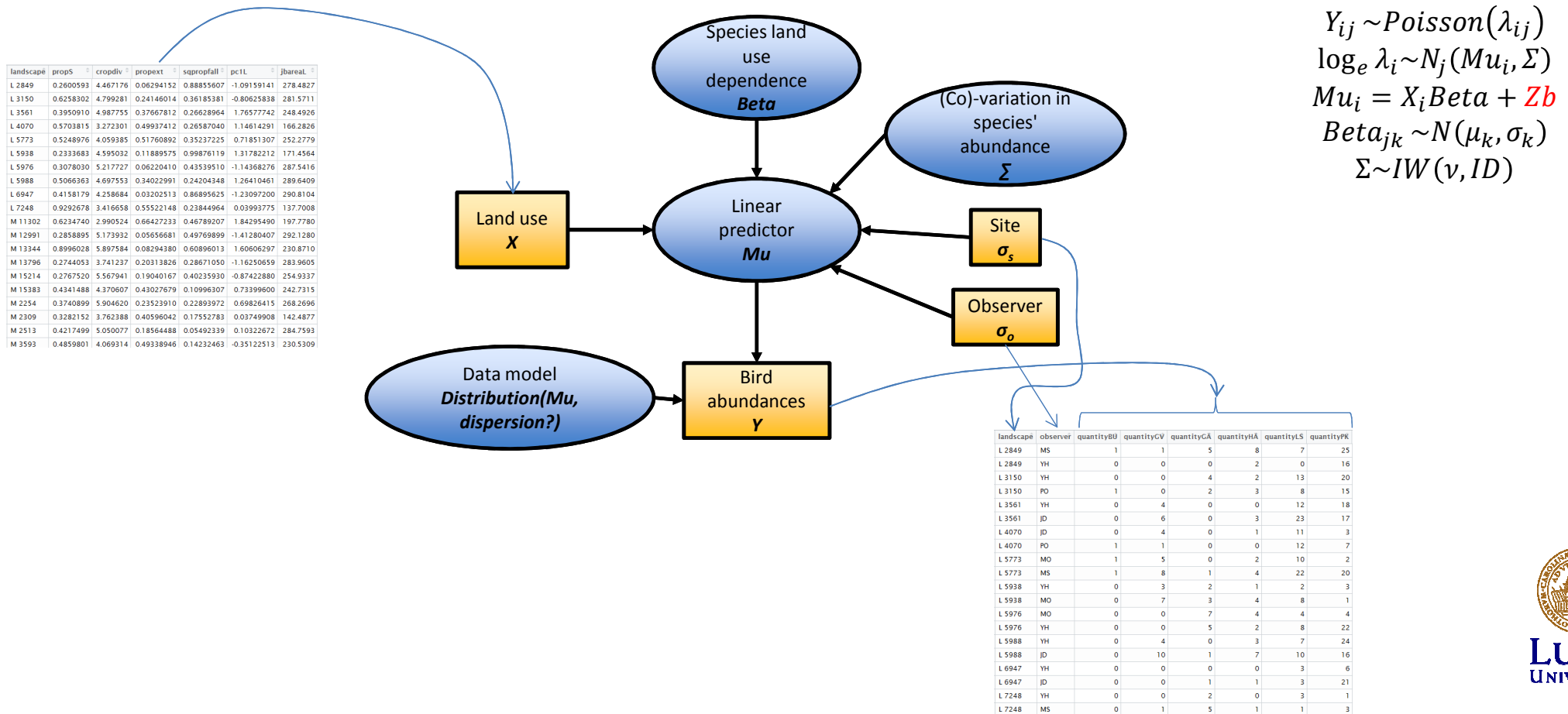
$$\log_e \lambda_i \sim N_J(M_i, \Sigma)$$

$$M_i = X_i B$$

$$B_{jk} \sim N(\mu_k, \sigma_k)$$

$$\Sigma \sim IW(\nu, ID)$$

The joint species abundance model (JSAM)



Obstacle 2 – The prior for Σ

- Commonly used prior for variance-covariance matrix - Inverse-Wishart (IW)

dahtah

about

Search this Blog

Why an inverse-Wishart prior may not be such a good idea

While playing around with Bayesian methods for random effects models, it occurred to me that inverse-Wishart priors can really bite you in the bum. Inverse Wishart-priors are popular priors over covariance functions. People like them priors because they are conjugate to a Gaussian likelihood, i.e, if you have data $\mathbf{y}_1, \dots, \mathbf{y}_n$ with each \mathbf{y}_i :

$$\mathbf{y}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{S})$$

so that the \mathbf{y}_i 's are correlated Gaussian vectors, and you wish to infer the correlation matrix \mathbf{S} , then putting an inverse-Wishart prior on \mathbf{S} is convenient because the posterior distribution is very easy to sample from

From blog by Simon Barthelmé

<https://dahtah.wordpress.com/2012/03/07/why-an-inverse-wishart-prior-may-not-be-such-a-good-idea/>

Search this Blog

Recent Posts

Superpixels in imager
New version of imager package for image processing
vecpack: an R package for packing stuff into vectors
New features in imager 0.30
New features in imager 0.20

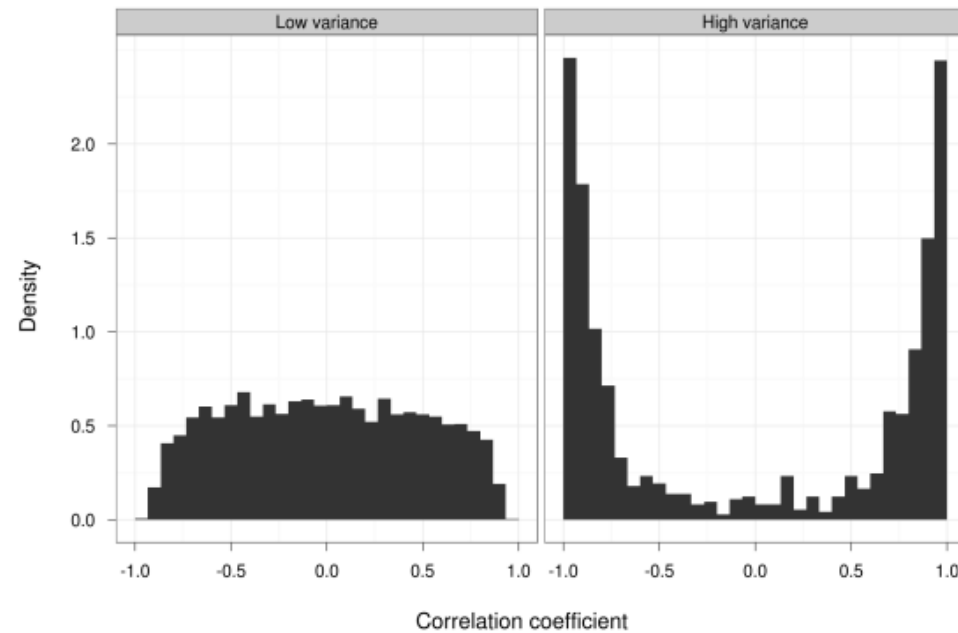
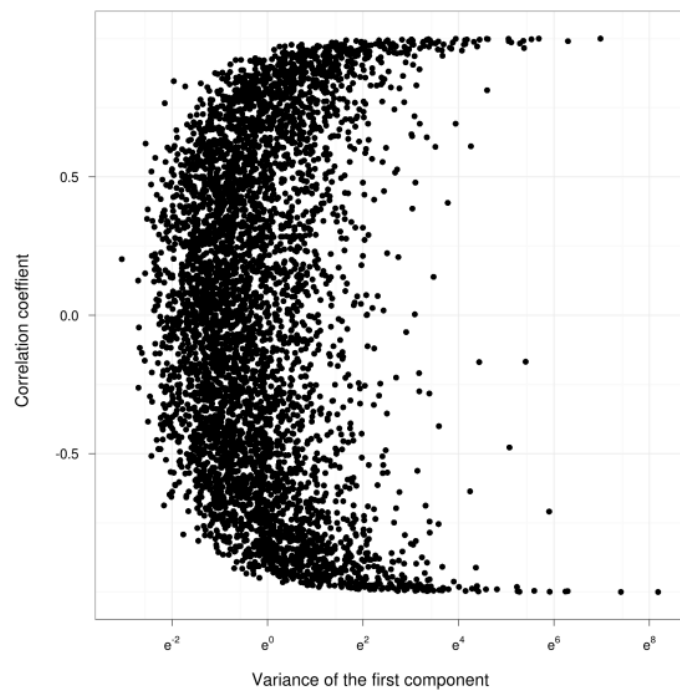
Archives

March 2017



LUND
UNIVERSITY

Dependence between variance and correlation



From blog by Simon Barthelmé
<https://dahtah.wordpress.com/2012/03/07/why-an-inverse-wishart-prior-may-not-be-such-a-good-idea/>

Alternative priors for the covariance matrix

inverse Wishart (IW)

$$\Sigma \sim IW(\nu, ID)$$

Scaled inverse Wishart (SIW)

$$\Sigma = \Delta Q \Delta$$

$$\Delta_{ii} = \xi_i \Delta \text{ is diagonal}$$

$$Q \sim IW(\nu, ID)$$

$$\xi_i \sim N^+(0, 1)$$

Separation strategy with IW (SSIW)

$$\Sigma = \Delta R \Delta$$

$$\Delta_{ii} = \sigma_i \Delta \text{ is diagonal}$$

$$R = \Delta_q Q \Delta_q$$

$$(\Delta_q)_{ii} = \frac{1}{\sqrt{Q_{ii}}} \Delta_q \text{ is diagonal}$$

$$Q \sim IW(\nu, ID)$$

$$\sigma_i \sim \text{Cauchy}^+(0, 5)$$

Separation strategy with LKJ (SSLKJ)

$$\Sigma = \Delta R \Delta$$

$$\Delta_{ii} = \sigma_i \Delta \text{ is diagonal}$$

$$R \sim LKJ(1)$$

$$\sigma_i \sim \text{Cauchy}^+(0, 5)$$

Barnard, J., McCulloch, R., Meng, X.-L., 2000. Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica* 10, 1281-1311.

Lewandowski, D., Kurowicka, D., Joe, H., 2009. Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis* 100, 1989-2001.

Simpson, Matt, 2012, "Prior distributions for covariance matrices: the scaled inverse-Wishart prior", blog @ <http://www.themattsimpson.com/2012/08/20/prior-distributions-for-covariance-matrices-the-scaled-inverse-wishart-prior/>

Alvarez, I., Niemi, J., Simpson, M., 2014. Bayesian inference for a covariance matrix, in: Song, W. (Ed.), Annual conference on applied statistics in agriculture. Kansas State University Libraries, New Prairie Press, Manhattan, Kansas.



LUND
UNIVERSITY

Obstacle 3 – alternative priors difficult to implement in BUGS/JAGS

- Gibbs sampling wants conjugate priors (I'm told...)
- Prior distributions not available
- Solution: Stan!
 - Hamiltonian Monte Carlo - conjugacy a non-issue
 - Has built-in prior for correlation matrix (LKJ)



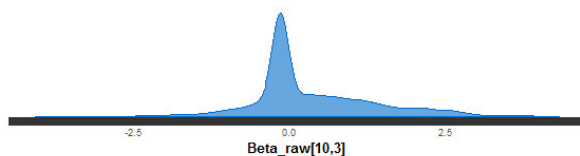
Obstacle 4 – Convergence etc.

- A first try to translate BUGS/JAGS code into Stan resulted in bad looking chains, low effective sample size, divergent transitions, high Rhat etc.

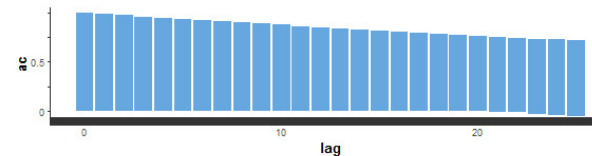
Rhat	n_eff	mean	sd	2.5%	50%	97.5%
1.15	20	0.35	0.95	-1.16	-0.15	2.61

☐ Include warmup

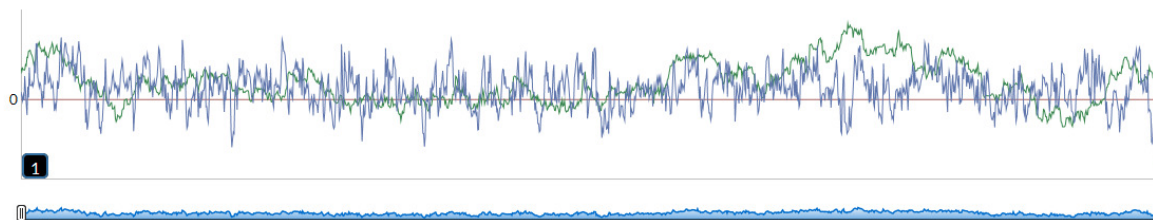
Kernel Density Estimate



Autocorrelation

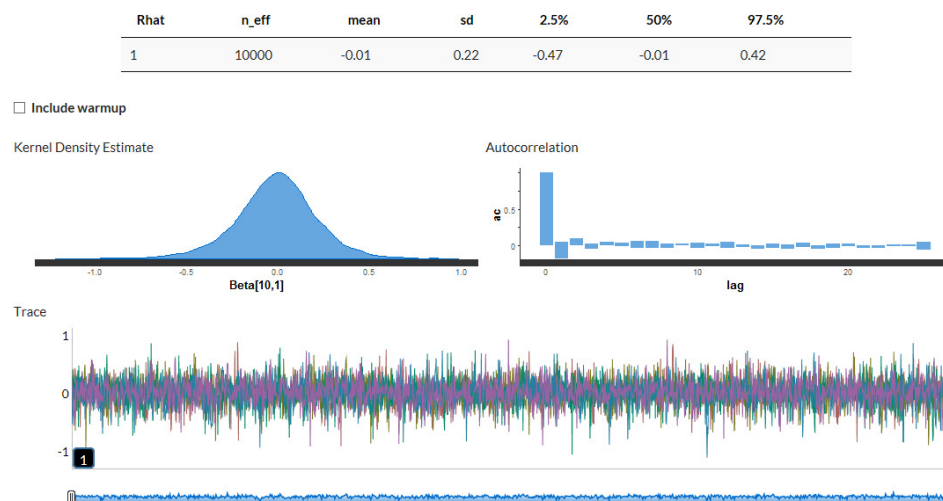


Trace



Obstacle 4 – Convergence etc.

- Solution: the Stan manual!
 - Reparameterization and vectorization
 - Example code



Stan Modeling Language User's Guide and Reference Manual

Stan Development Team

Stan Version 2.14.0

Monday 26th December, 2016

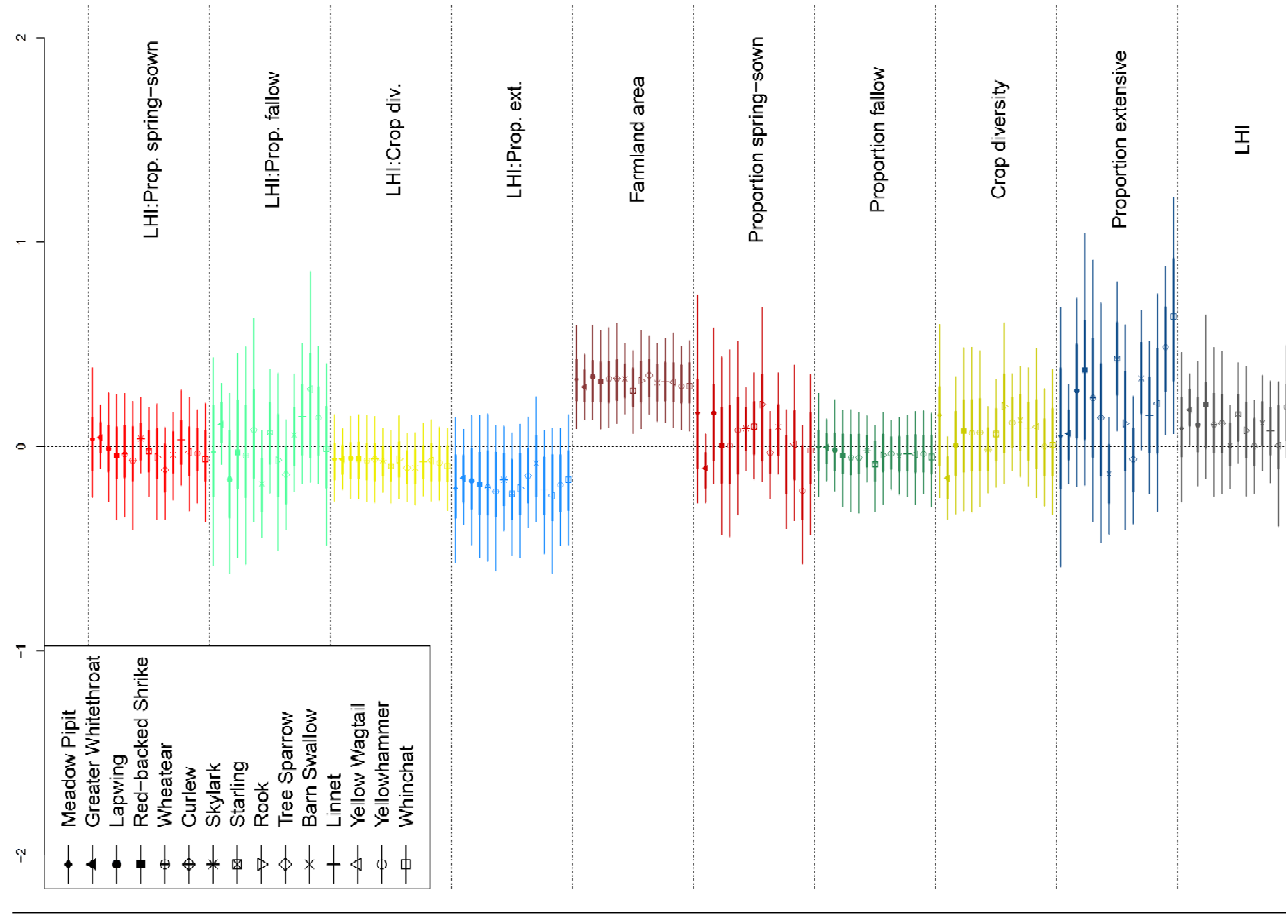


mc-stan.org

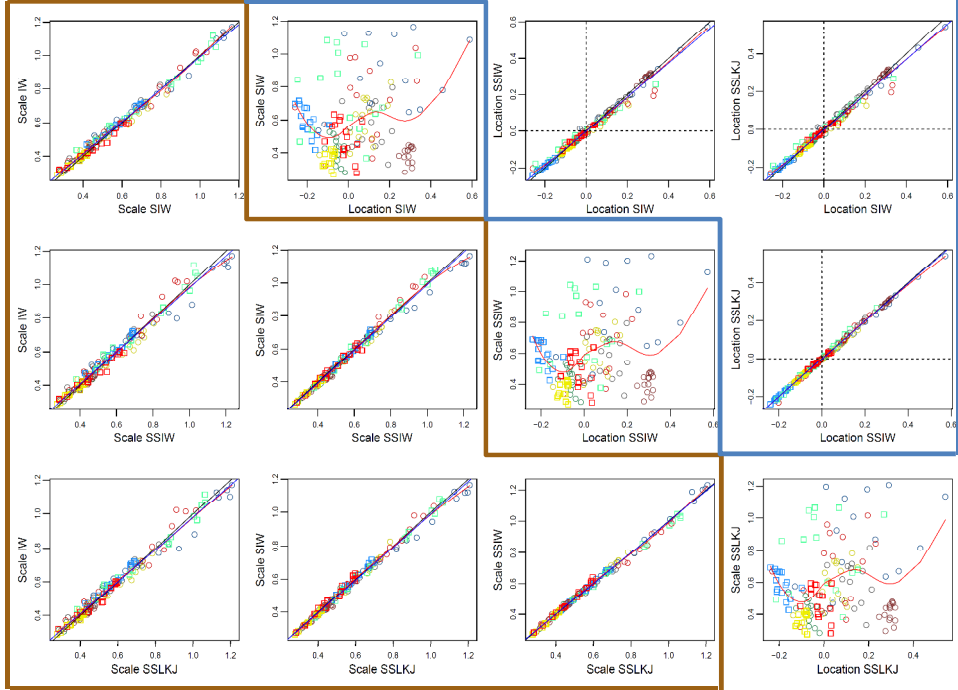


LUND
UNIVERSITY

Finally some results – species specific Beta



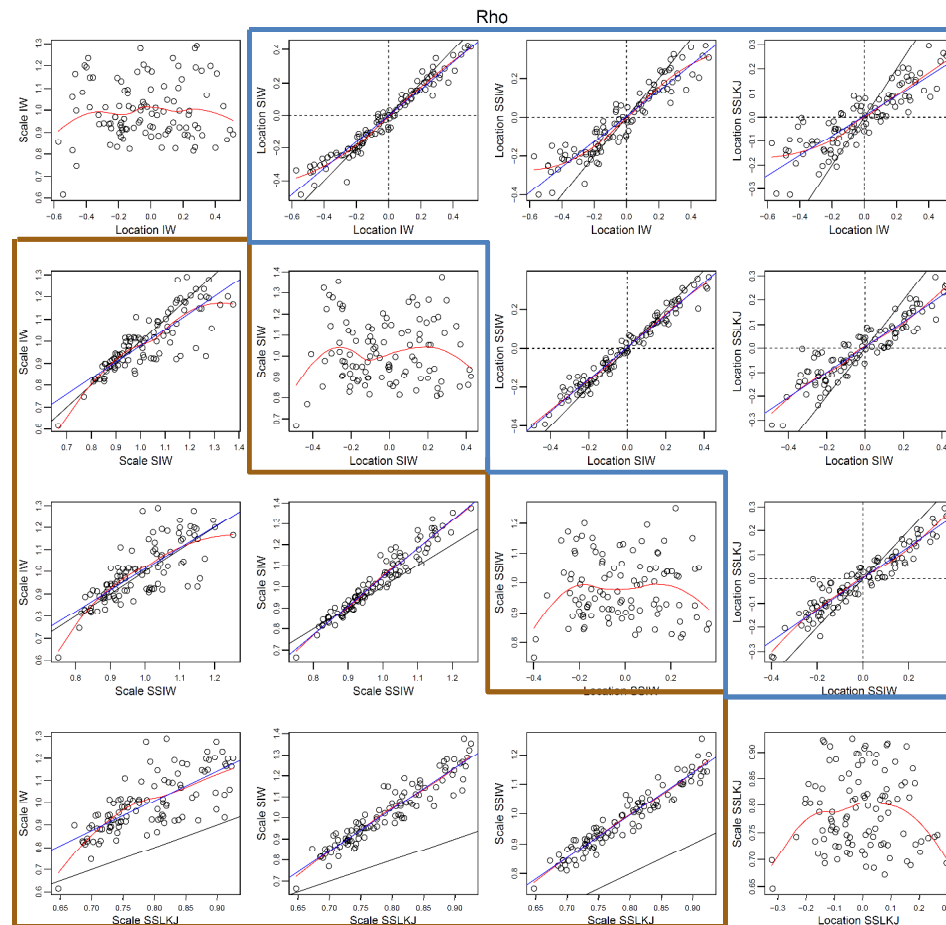
Beta (intercept excluded)



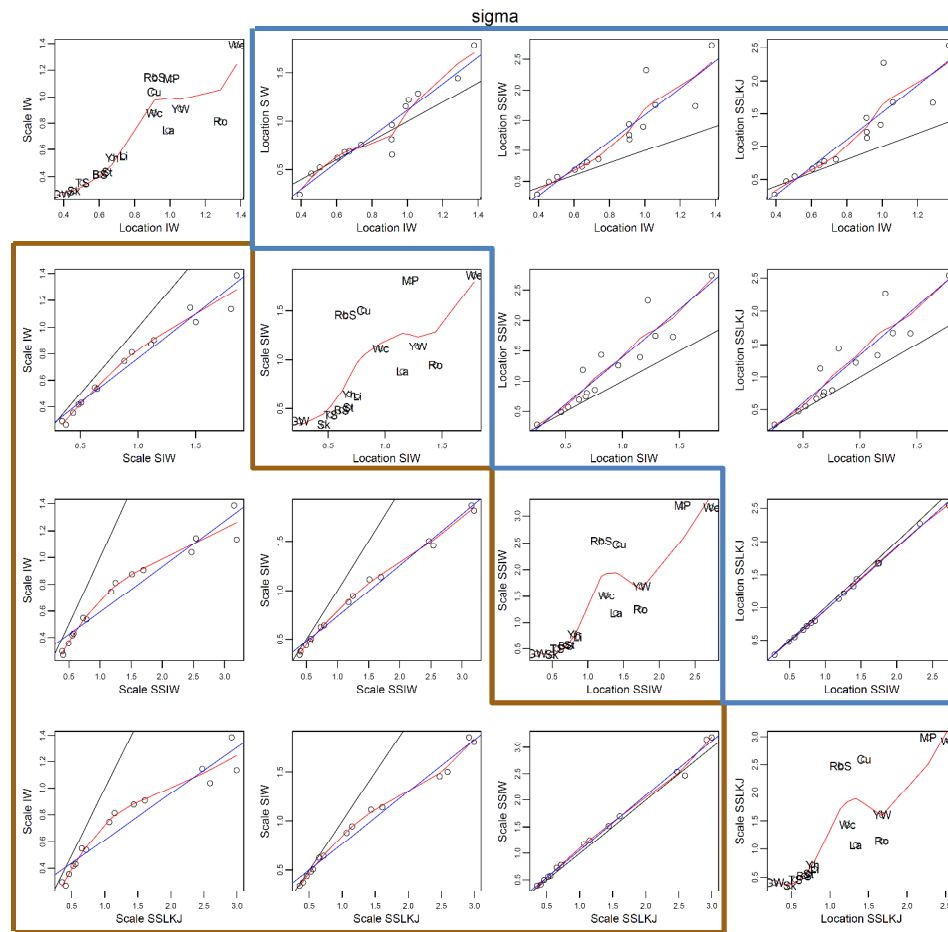
Mean posterior

Width of HPD interval

Finally some results – effect of prior on Σ Rho (correlations between species)



Finally some results – effect of prior on Σ sigma (standard deviation)

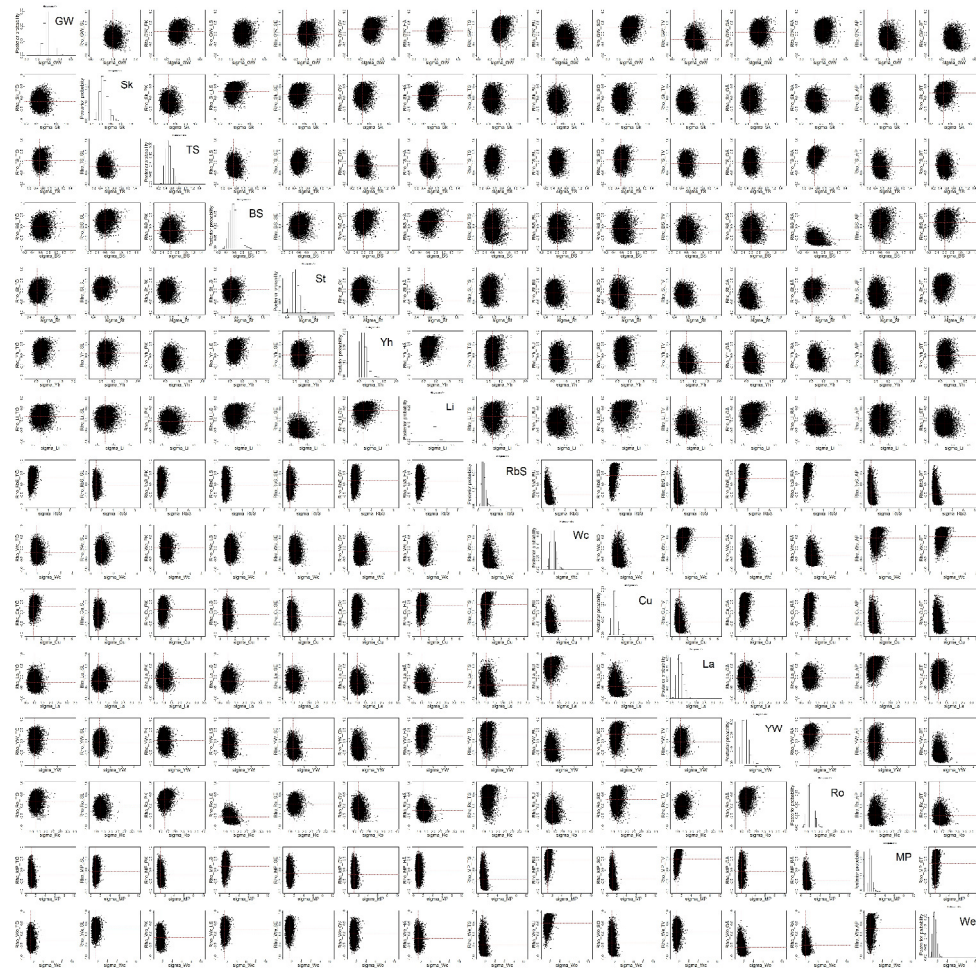


Mean posterior

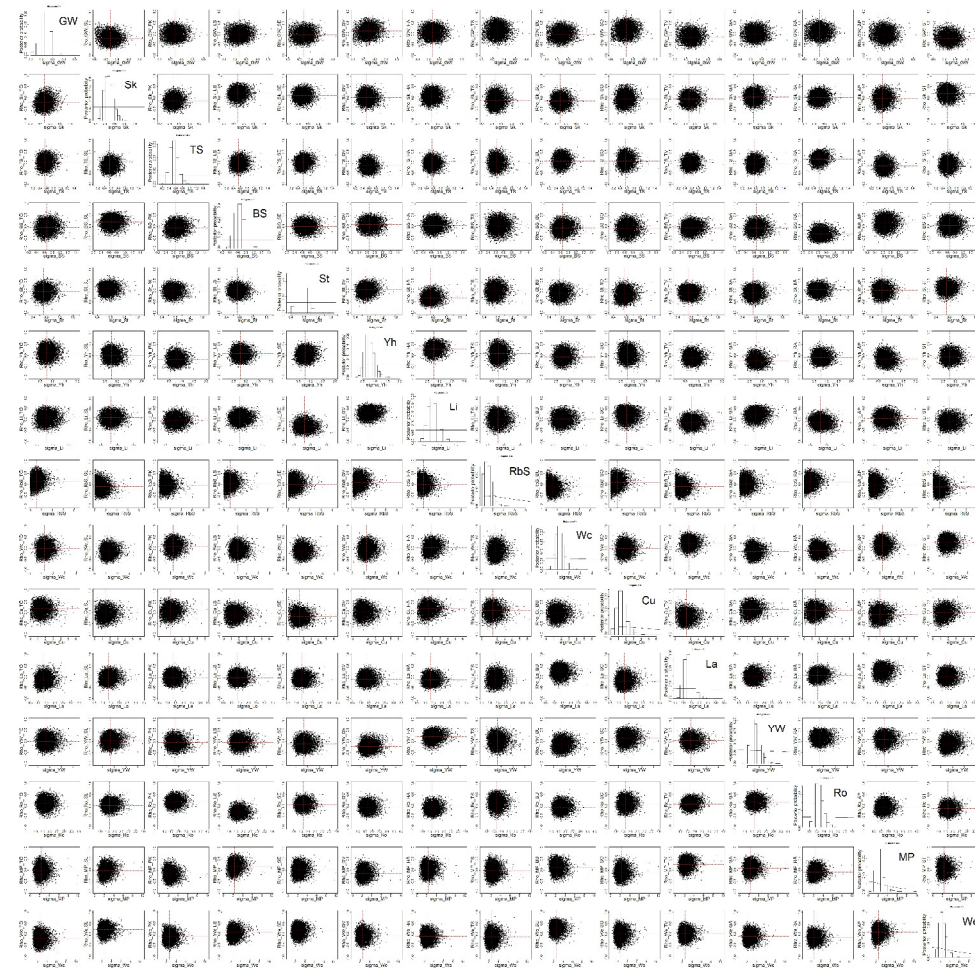
Width of HPD interval

Correlation

IW

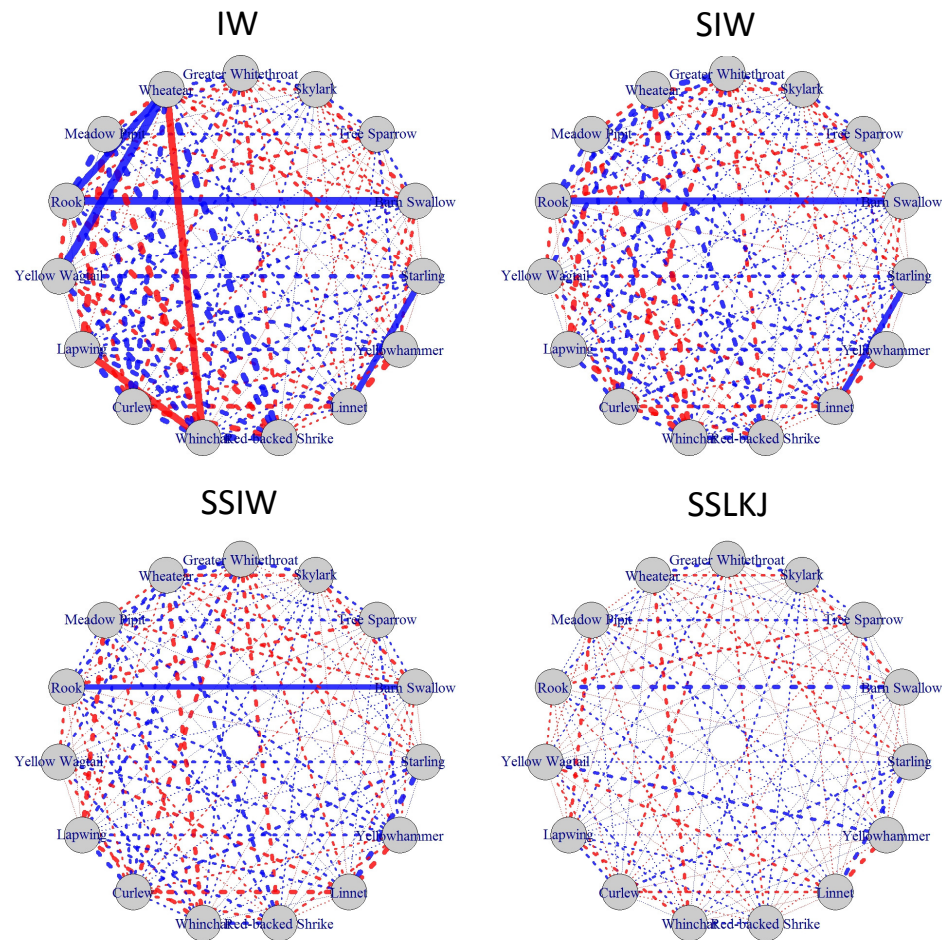


SSLKJ



sigma (standard deviation)

Finally some results – effect of prior on Σ



What did we learn?

- The Bayesian approach is very flexible – we can model anything!
- Flexibility demanding – we need to make active decisions
- Was it useful? – we won't know until we've tried
- Complex output – but we only need to communicate relevant parts to stakeholders
- Bayesian analyses bring things in the open
- Enhanced understanding
- There is lots of help "out there"
- And more...

The end

Thanks for listening!

Thanks also to Ullrika Sahlin and Johan Lindström for encouragement and inputs of wisdom!

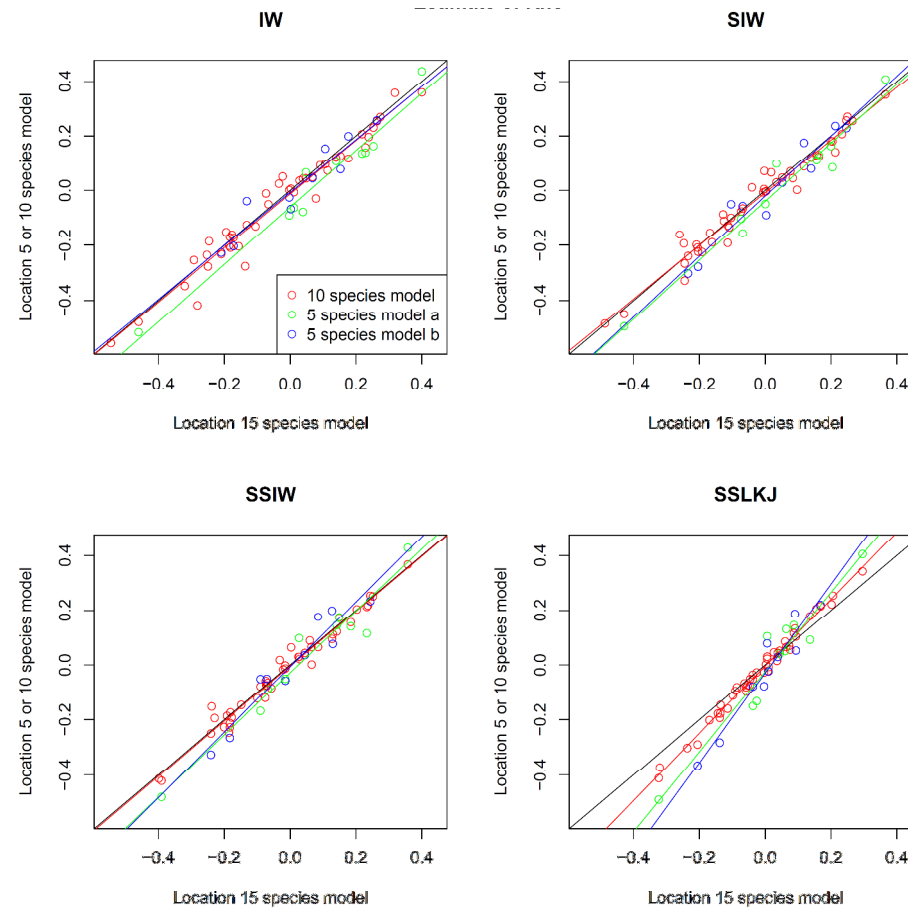


Forskningsrådet Formas

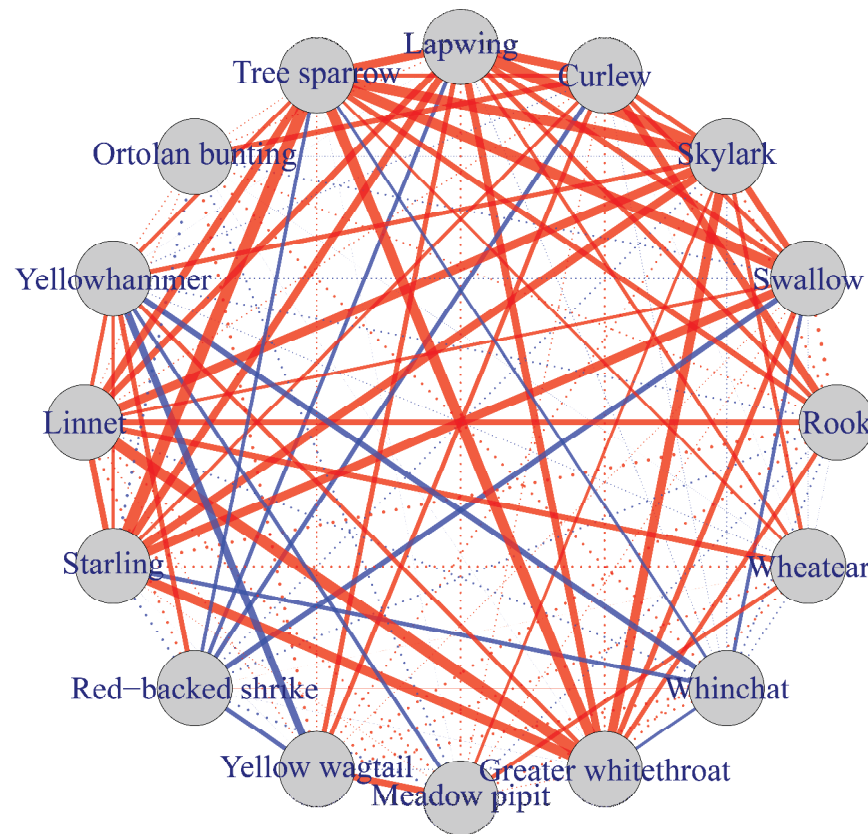
Formas främjar framstående forskning för hållbar utveckling



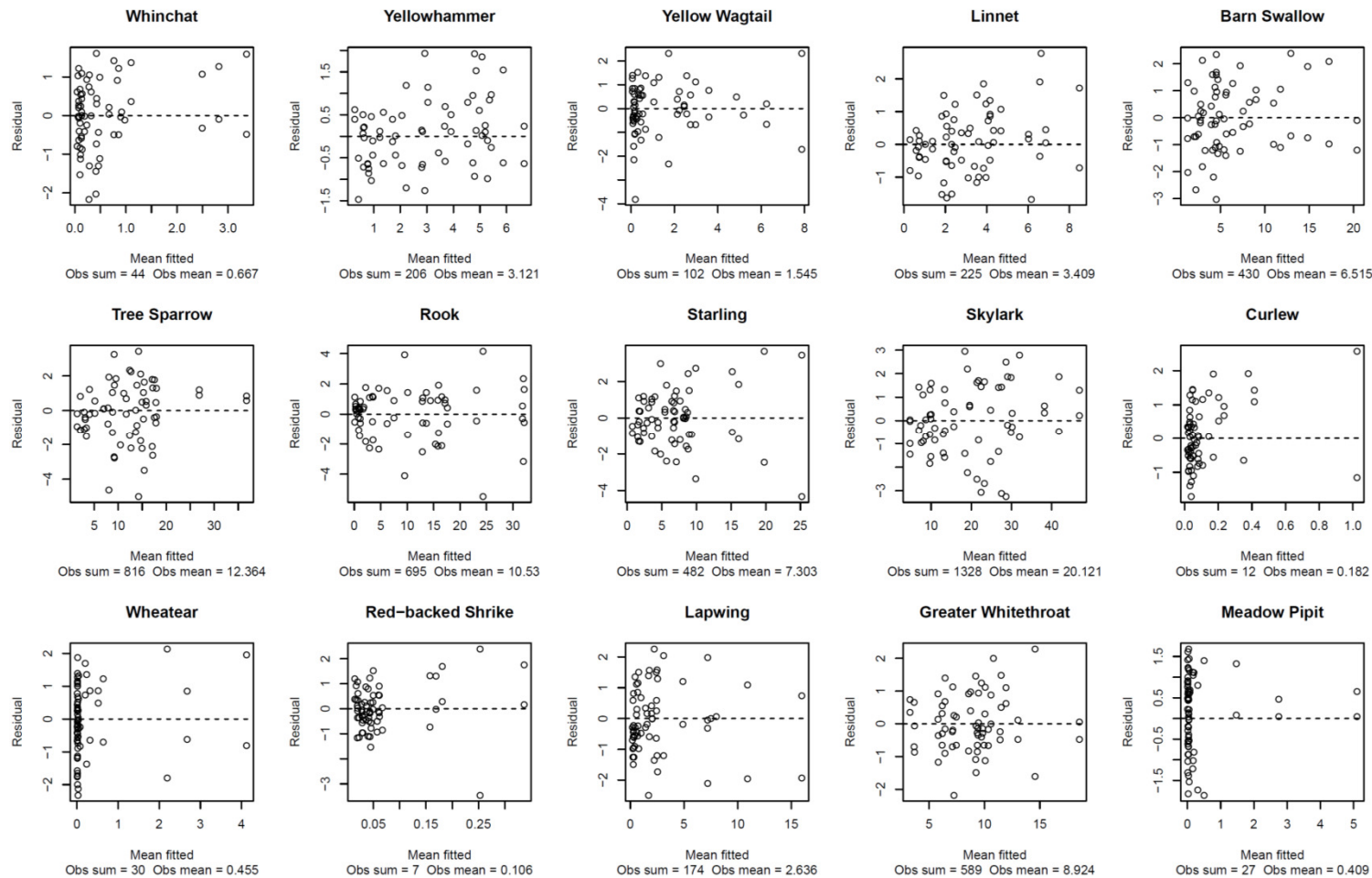
LKJ exhibit less shrinkage with fewer species



Larger dataset shows that LKJ does not necessarily shrink correlations to zero



Slight pattern in residuals



Random effects:

```
Groups Name      Variance Std.Dev.
Eps  (Intercept) 0.62099  0.78803
fZoo  (Intercept) 0.00000  0.00000
Number of obs: 87, groups: Eps, 87; fZoo, 8
```

Fixed effects:

```
Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.86900    0.31221  -9.189  < 2e-16
cLSize      -0.32280    0.12098  -2.668  0.007626
cVisitors   -0.02220    0.11102  -0.200  0.841534
fFeeding2    -0.68762    0.24484  -2.808  0.004979
fOc2         0.96309    0.25101   3.837  0.000125
fOther2      0.27994    0.23899   1.171  0.241460
fEnrichment2 0.13473     0.26567   0.507  0.612073
fGroup2     -0.85641    0.22000  -3.893  9.91e-05
fSex2       0.05347    0.20214   0.265  0.791390
cEnclosure  -0.07352    0.11001  -0.668  0.503951
cVehicle    0.24409    0.12485   1.955  0.050571
cAge       -0.14584    0.10620  -1.373  0.169628
```

The est
and the est
with Zoo
observation
The res
fGroup a

As part of the model validation process, we extract the Pearson residuals and plot them against the fitted values (Figure 7.5), against all covariates in the model, and against all covariates not in the model. The first graph, plotting residuals versus fitted values, shows a pattern for the smaller fitted values. On nearly every occasion that we use the observation-level random intercept in a GLMM, we see a similar pattern if residuals are plotted versus fitted values. If we plot the fitted values against the observed data, we will most likely find that the model gives a perfect fit. This is because the `glmer` function includes the random effects in the fitted values. To demonstrate this, we can calculate the fitted values ourselves. First we calculate the matrix `X` with all the covariates.

```
> X <- model.matrix(~ cLSize + cVisitors +
+ fFeeding + fOc + fOther + fEnrichment +
+ fGroup + fSex + cEnclosure + cVehicle +
+ cAge, data = ZD)
```

The first column of `X` contains ones, the second column is the `cLSize` values, etc. Categorical variables are included as a column with the values 0 and 1. Next we extract the estimated regression parameters:

```
> Betas <- fixef(M2)
```

and the random effects z_i and ϵ_{ij} :

```
> z <- ranef(M2)$fZoo$(Intercept)'
> Eps <- ranef(M2)$Eps$(Intercept)'
```

The final requirement is a vector with values identifying which observations belongs to a particular zoo:

```
> re <- as.numeric(ZD$fZoo)
```

We now have all required components to calculate the predictor function η and probabilities of stereotypic behaviour π .

```
> eta <- X %*% Betas + z[re] + Eps
> pi <- exp(eta) / (1 + exp(eta))
> fitted(M2) - pi
```

The results of the final line are not presented here but include only zeros, indicating that the fitted function applied to a `glmer` gives fitted values that include the random intercepts. This should not be a problem if we are aware that fitted values may contain both the covariate

calculating

not indicate

tain the

Apparently a common problem with "observation-level random effects"

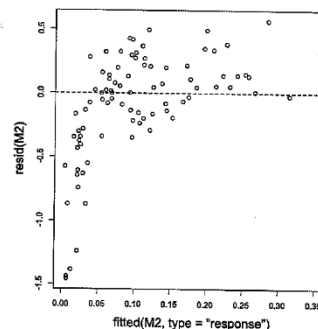
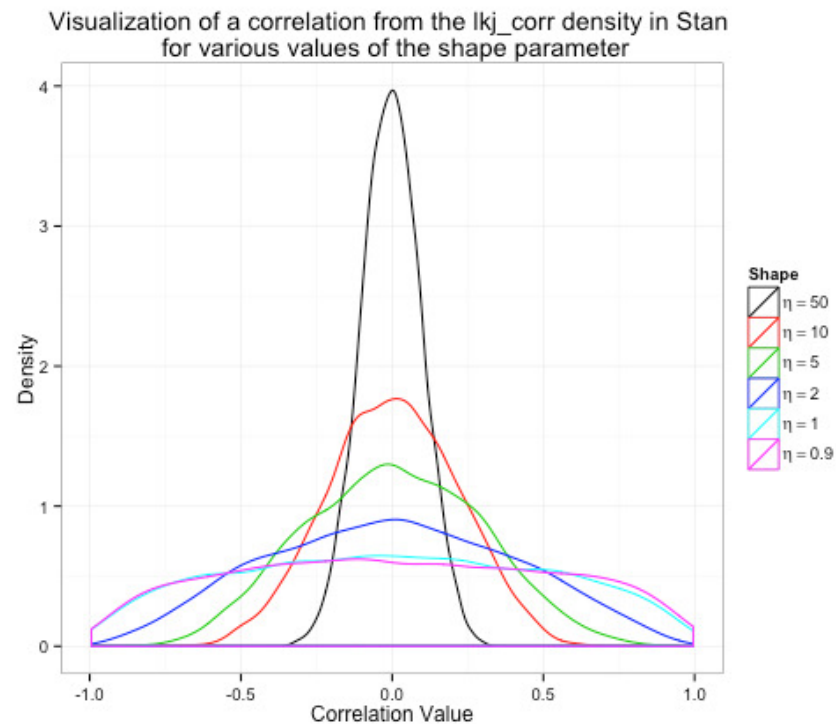


Figure 7.5. Pearson residuals plotted versus fitted values for the binomial GLMM with random intercept Zoo and observation-level random intercept.

From Zuur, A.F., Hilbe, J.M., Ieno, E.N., 2013. A Beginner's Guide to GLM and GLMM with R: A Frequentist and Bayesian Perspective for Ecologists. Highland Statistics.

LKJ prior



3x3 correlation matrix

Taken from Scott Baldwin's site "Psychstatistics"
<http://www.psychstatistics.com/2014/12/27/d-lkj-priors/>