Bayesian Inference of Multiple-Traits in a House Sparrow Population using the Animal Model

Ingelin Steinsland & Henrik Jensen

Norwegian University of Science and Technology, Trondheim, Norway

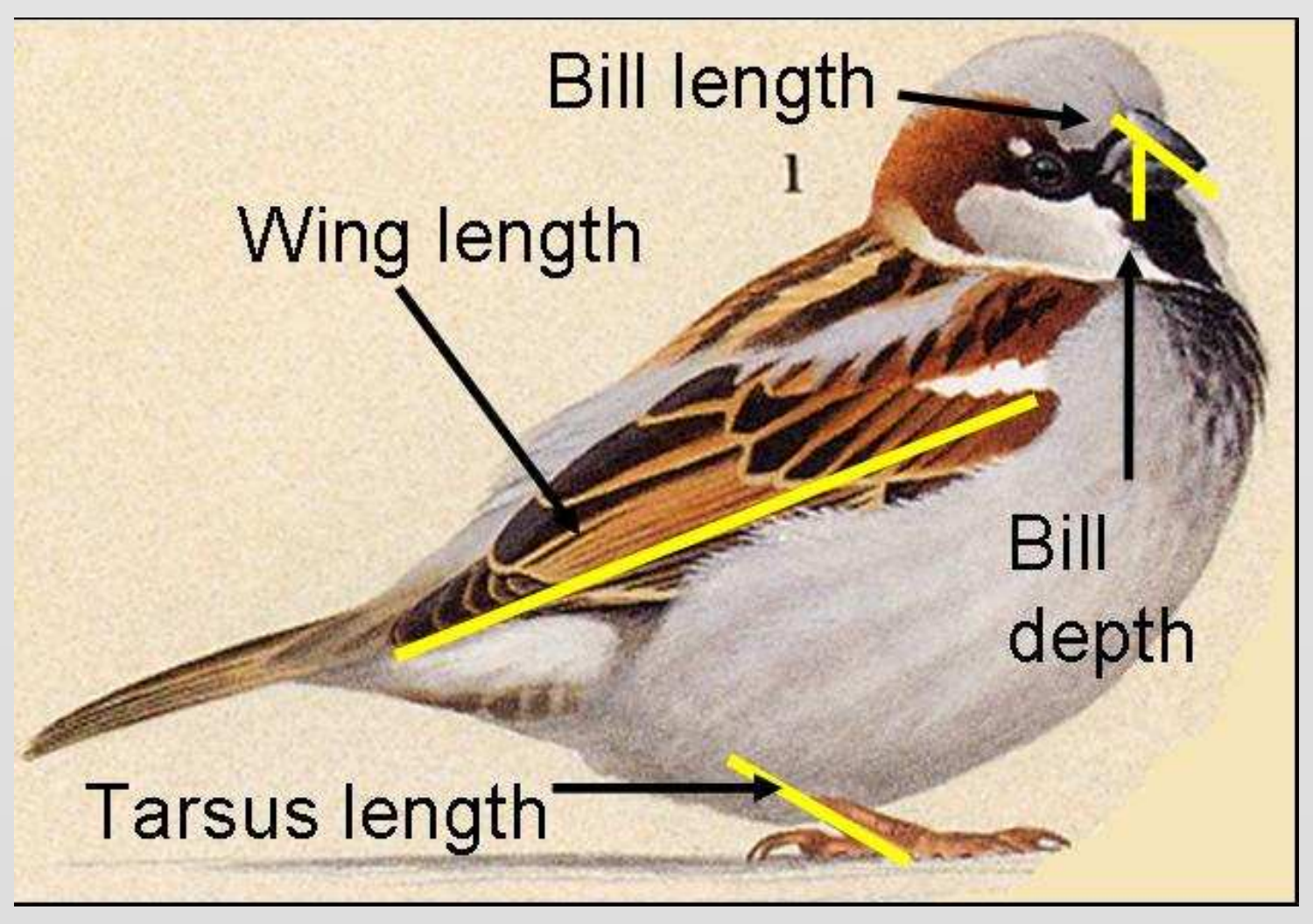
Why Bayesian?

- Can easily access uncertainty for all quantities of interest, e.g. for heritability and generic correlation estimates as well as response to selection.
- Uncertainty of estimated parameters are considered in the further analysis.

Hestmannøy (one of the study islands)

Data

- Most house sparrows of six islands off the coast of Northern Norway are ringed and measuered since 1993.
- Blood samples taken ⇒ Genotyping of 9 microsatalites ⇒ Parenthood analysis ⇒ Pedigree
- Traits measured on adult birds; *tarsus length*, *wing length*, *bill depth*, *bill length*, *body mass*, *total badge* and *visual badge*.



- Missing data: Females do not have a badge + some other missing trait values.
- There are 3572 birds in the pedigree, of these 1004 have (some) adult traits measurements.
- For (almost) all birds with data *sex*, *hatch year* and *hatch island* are known.

The Animal Model

For bird i $(i = 1, \dots, 1004)$:

observed traits	=fixed	+genetic	+environmental
y_i	$=\beta_i$	$+u_i$	$+\epsilon_i$

- y_i: observations (traits), $y_i = (y_{tarsus}, y_{wing}, \dots)_i$
- β_i: "*fixed effects*" (sex, hatch year and hatch island), $\beta \sim N(0, \sigma_\beta^2 I)$
- u_i: genetic effects, $u_i \sim N(0, \Sigma_u)$
- ϵ_i: Environmental effects: $\epsilon_i \sim N(0, \Sigma_\epsilon)$
- Conjugate prior for the genetic and environmental covariance matrices Σ_u and Σ_ϵ (inverted Wishart).

For the population:

$$y = X\beta + Zu + \epsilon = W\begin{pmatrix} u \\ \beta \end{pmatrix} + \epsilon$$

- X and Z: incidence matrices, $W = (Z, X)^T$.
- u: genetic effect $u \sim N(0, \Sigma_u \otimes A)$
- A: genetic covariance matrix
- Structure of A^{-1} from moralising the pedigree $\Rightarrow A^{-1}$ sparse.
- ϵ: environmental effect $\epsilon \sim N(0, \Sigma_\epsilon \otimes I)$

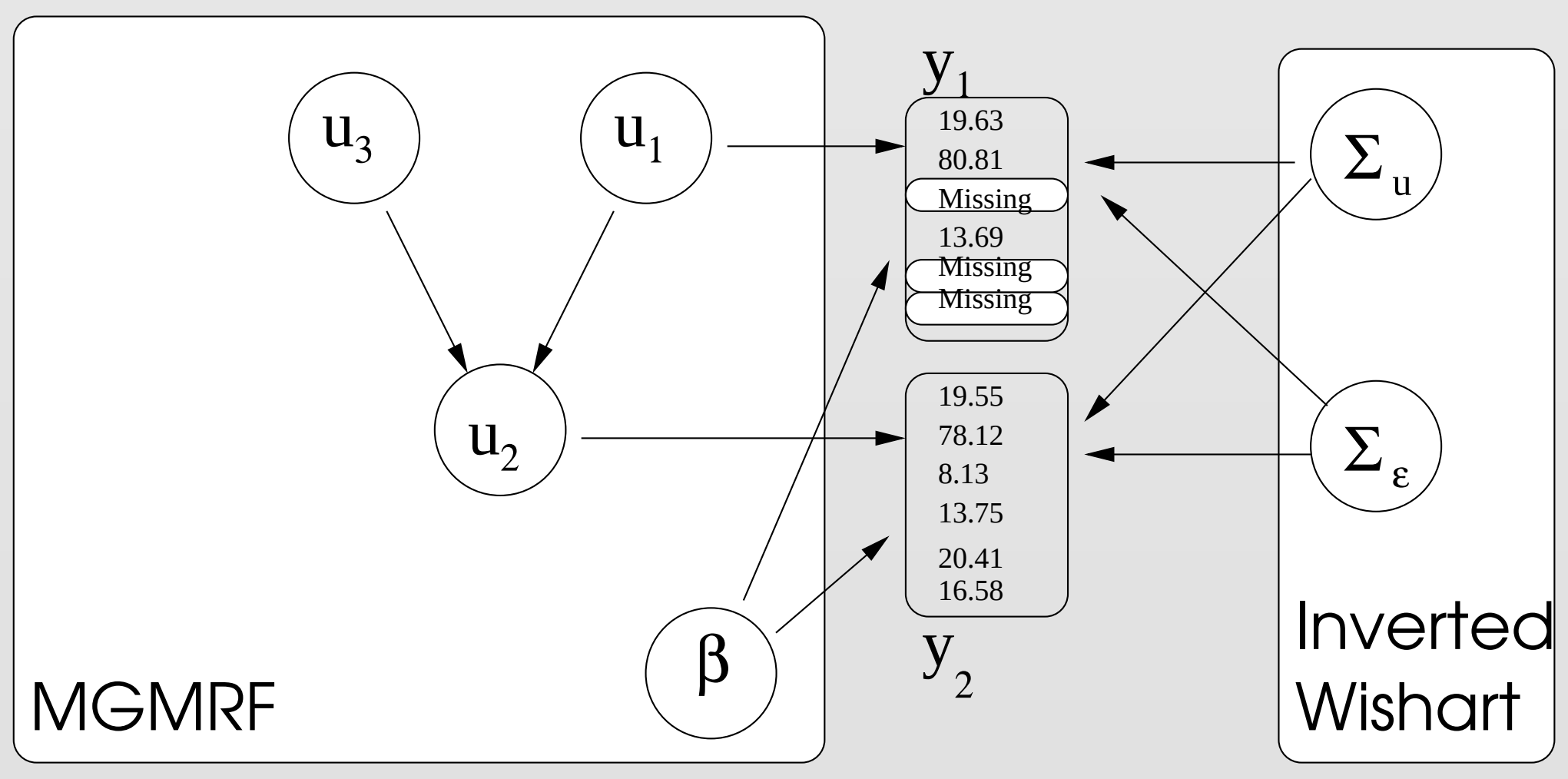
Constraints:

- For all but one fixed effect (f =hatch year and sex), for all traits t:

$$\sum_{l=1}^{Nlevel} \beta_{flt} = 0$$

- For the breeding values, for all traits:

$$\sum_{i=1}^{N} u_{it} = 0$$



Of interest:

- Genetic and environmental covariance matrices (Σ_u and Σ_ϵ), fixed effects (β) and/or individual breeding values (u_i).
- Heritability; $h_j^2 = \frac{\sigma_{uj}^2}{\sigma_{uj}^2 + \sigma_{\epsilon j}^2}$.

Full conditional distributions:

- $\beta, u, y_{miss} | y_{obs}, \Sigma_u, \Sigma_\epsilon \sim$ MGMRF, a multivariate Gaussian Markov Random Field with constraints.
- $\Sigma_u, \Sigma_\epsilon | y, \beta, u \sim$ Inverted Wishart (if no constraints)

Gibbs sampler

Use a Gibbs sampler with two blocks for making inference.

Algorithm

- For each iteration
  1. Sample (β, u, y_{miss}) of dimension 25000!
  2. Sample $(\Sigma_u, \Sigma_\epsilon)$ (Metropolis-Hastings step because of constraints)

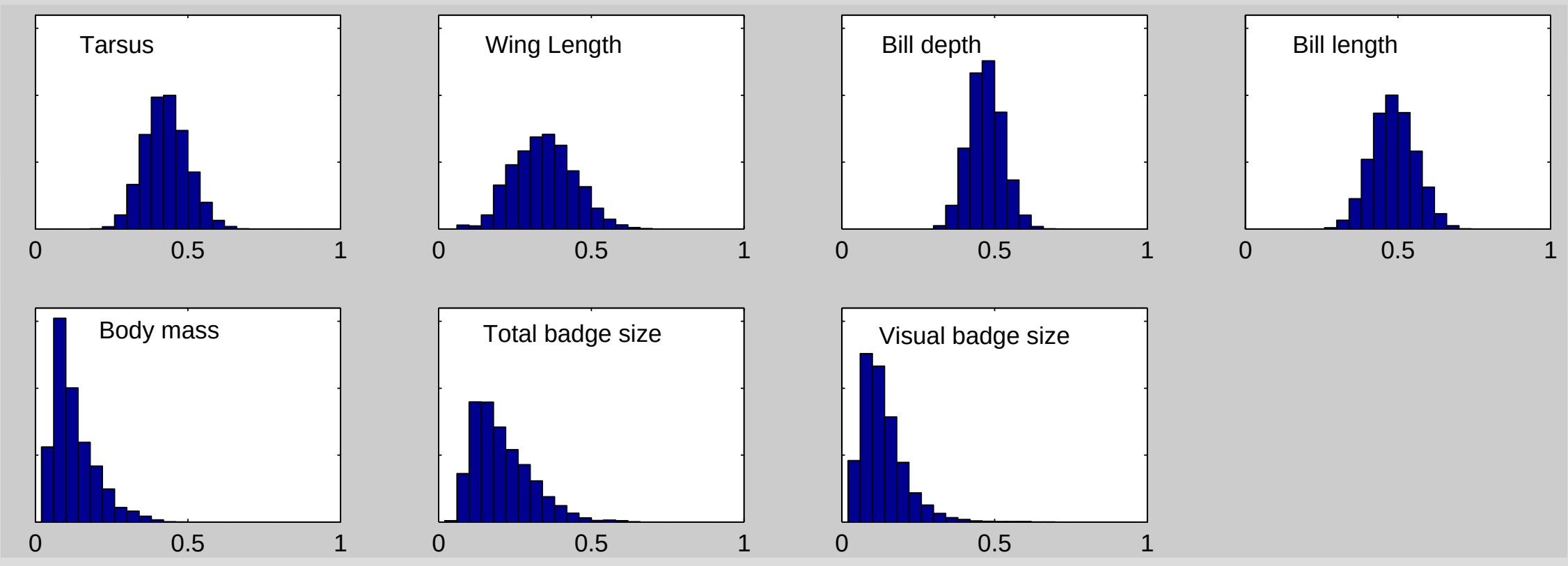
Step 1 is the challenge:

- Sparse $A^{-1} \Rightarrow$ cheaper calculations.
- Complexity $< \mathcal{O}(n^{1.5})$, use GMRFLib.

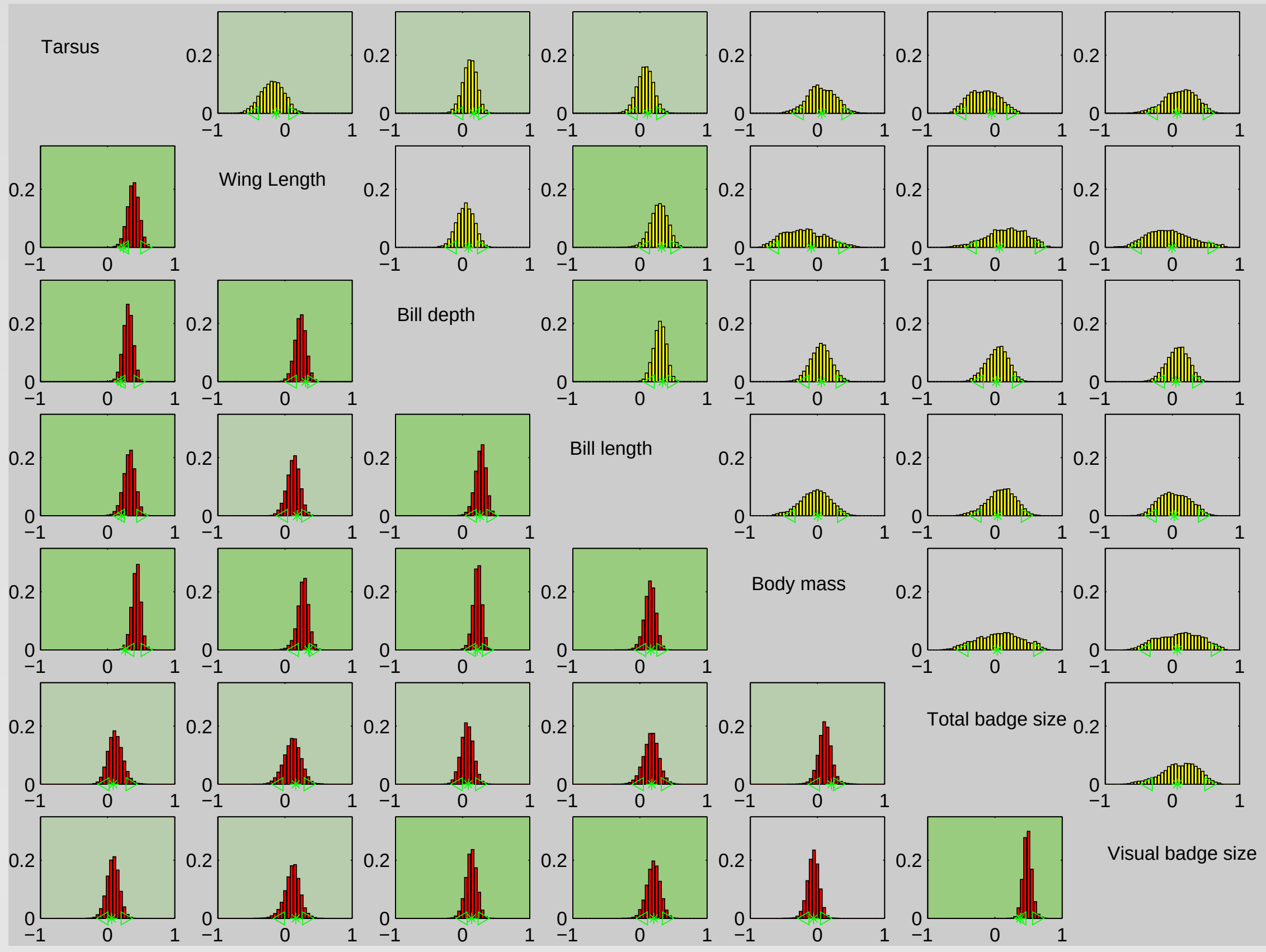
Results

PS: This is ongoing work, the software is still tested

- The Gibbs sampler ran for 55000 iterations



Posterior distributions for the heritabilities



Posterior distributions for genetic correlation (Σ_u, yellow) and environmental correlation (Σ_ϵ, red).

- All environmental correlation estimates are positive.
- Genetic correlation estimates for traits with low heritability (body mass) and/or many missing data (badges) are uncertain.

Summary

- Animal model of multiple traits = MGMRF model.
- Use a Gibbs sampler with only two blocks.
- Able to find uncertainty in heritability and genetic correlation
- Can handle constraints, e.g. in "fixed effects".

Further work

- Extend methodology to make inference about response to selection with uncertainty.