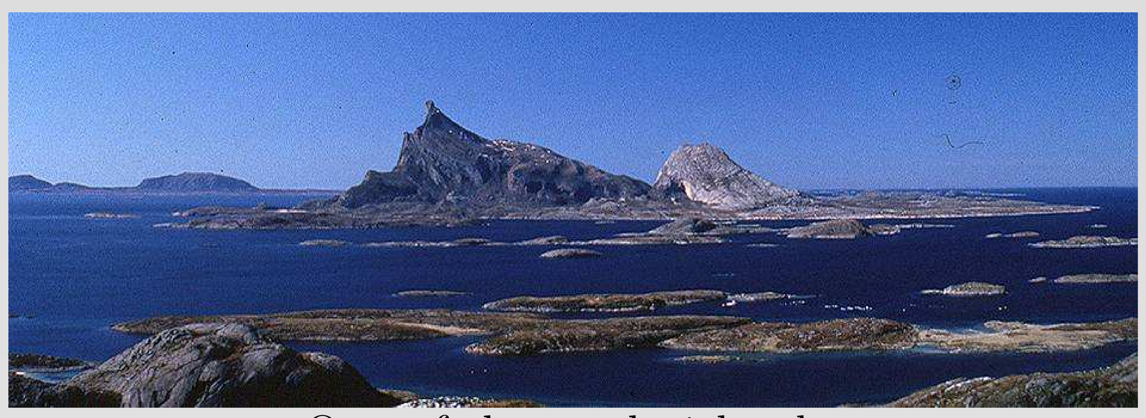
Quantitative Genetics, House Sparrows and a Multivariate Gaussian Markov Random Field Model

Ingelin Steinsland & Henrik Jensen
Norwegian University of Science and Technology, Trondheim, Norway

Of biological interest

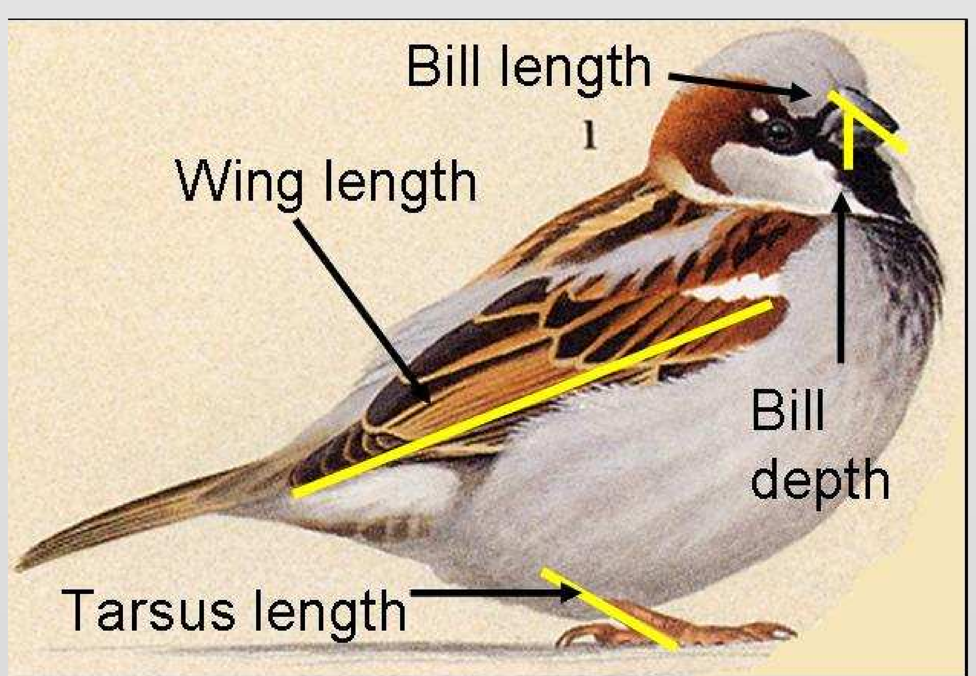
- Quantitative genetics. How much does genetics influence traits?
- Trait = Genetic + environmental effects
- Speed and direction of evolution.
- Important for response to environmental changes, e.g. climate changes.



One of the study islands.

Data

- Most house sparrows of five islands off the coast of Northern Norway are registered since 1993.
- Blood samples taken ⇒ DNA ⇒ Pedigree
- Traits are measured on adult birds, here *tarsus length* (t), *wing length* (w), *bill depth* (bd) and *bill length* (bl) are used:



- There are 2563 birds in the pedigree, of these 746 have (some) adult traits measurements.

Model

For bird i ($i = 1, \dots, 746$):

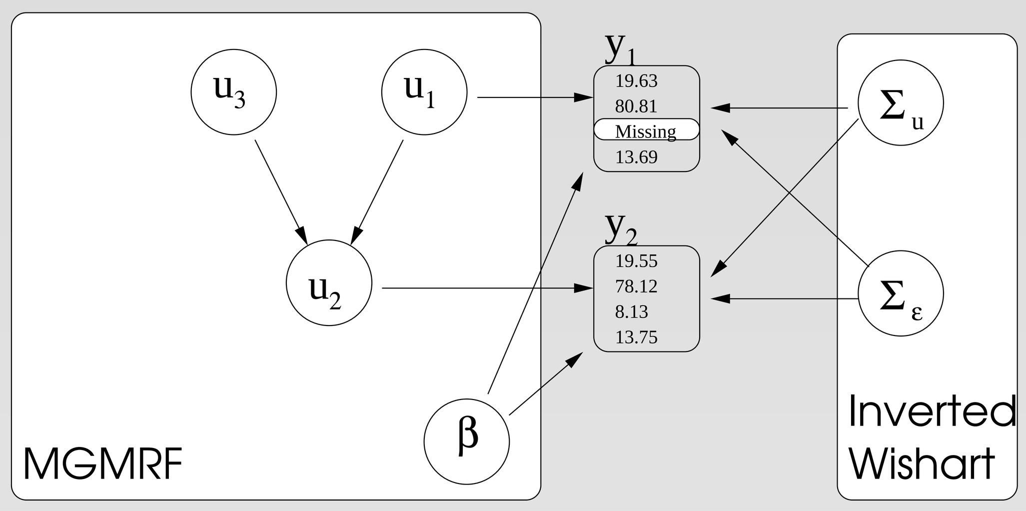
observed traits	=fixed	+genetic	+environmental
y_i	$=\beta_i$	$+u_i$	$+\epsilon_i$

- y_i: observations (traits), $y_i = (y_l, y_w, y_{bd}, y_{bl})_i$
- β_i: "*fixed effects*" (sex and 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 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 ⇒ A^{-1} sparse.
- ϵ: environmental effect $\epsilon \sim N(0, \Sigma_\epsilon \otimes I)$



Of interest:

- u, β, Σ_u and Σ_ϵ
- Heritability; $h_j = \frac{\sigma_{uj}^2}{\sigma_{uj}^2 + \sigma_{\epsilon j}^2}$, $j \in \{t, w, bd, bl\}$

Posterior distributions:

- $\beta, u | y, \Sigma_u, \Sigma_\epsilon \sim$ MGMRF, a multivariate Gaussian Markov Random Field.
- $\Sigma_u, \Sigma_\epsilon | y, \beta, u \sim$ Inverted Wishart

Gibbs sampler

Use a Gibbs sampler with three blocks for making inference.

Algorithm

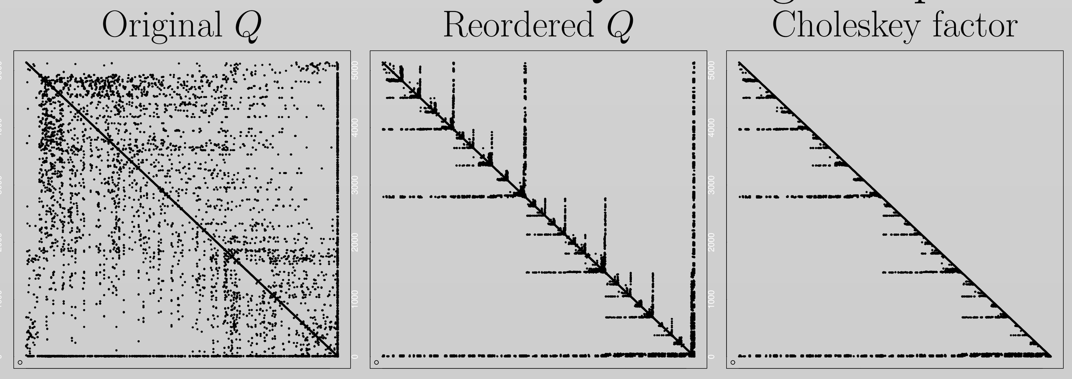
- For each iteration
  1. Sample *Missing data*
  2. Sample (β, u) of dimension 10280!
  3. Sample $(\Sigma_u, \Sigma_\epsilon)$

Step 2 is the challenge:

Fast sampling of MGMRF

- Want to sample $x \sim N(0, Q^{-1})$.
- Find Choleskey factor; $Q = LL^T$, and sample $z \sim N(0, I)$. Solve $L^T x = z$.
- Complexity for general multivariate problem $\mathcal{O}(n^3)$.
- Sparse Q ⇒ cheaper calculations.

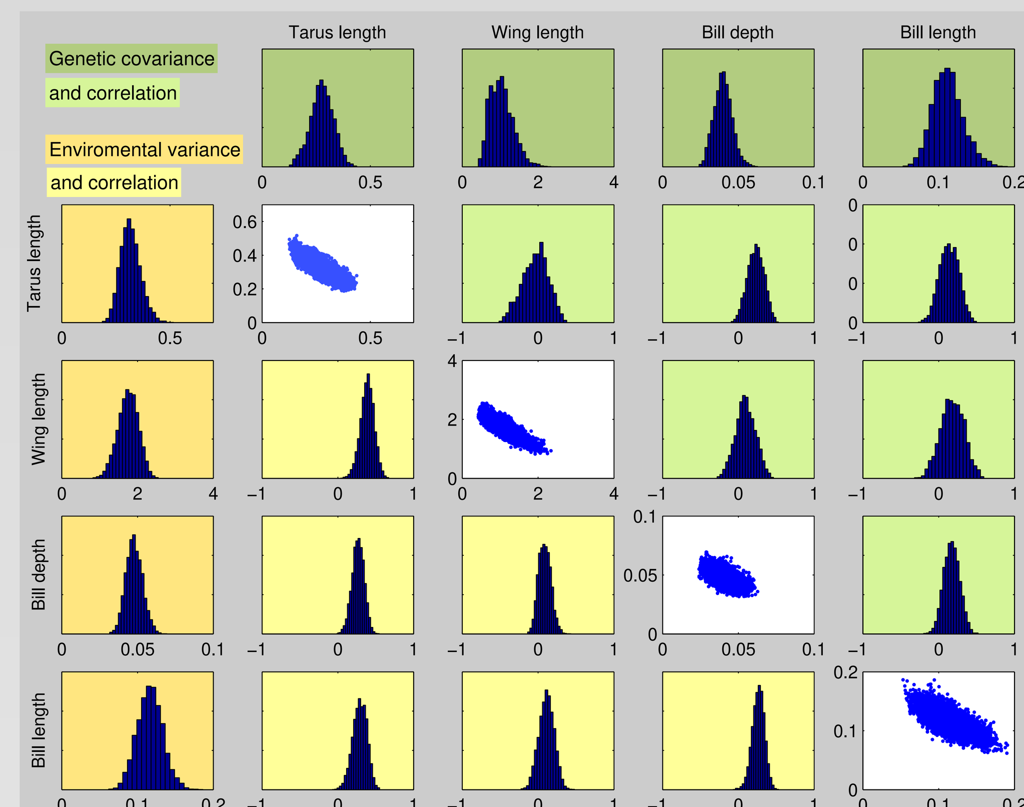
Reorder elements in Q to get sparse L.



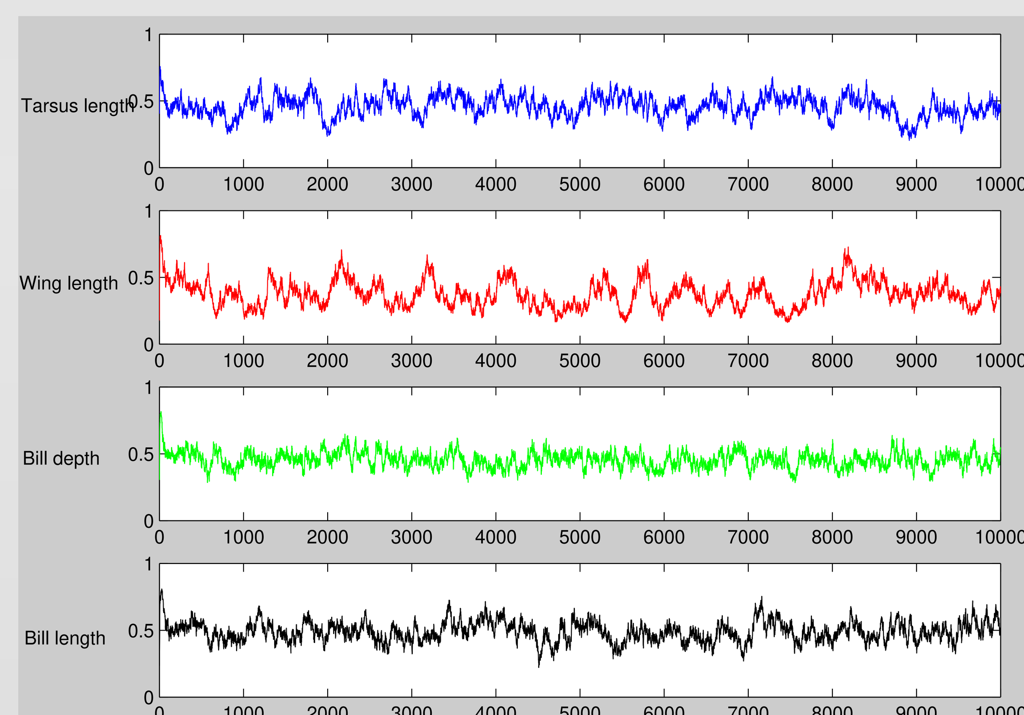
Complexity $< \mathcal{O}(n^{1.5})$!

Results

- The Gibbs sampler ran for 10000 iterations (one hour), use GMRFLib.
- Results agree with results from the classical analysis (REML).
- Mixing seems to be good



Posterior estimates for Σ_u^2 (green) and Σ_ϵ^2(*yellow*).



Trace-plot for the heritabilities

Summary

- Breeding model of multiple traits = MGMRF model.
- Use a Gibbs sampler with three blocks
- Because of the sparse structure the pedigree impose, sampling from a MGMRF of dimension 10000 is fast.

Further work

- Use the breeding values u to explore whether selection is genetic or environmental.
- Response to selection.