

Animal Model and INLA

Ingelin Steinsland & Anna Marie Holand

May 15, 2009

Outline



- ▶ Introduction animal model
 - ▶ Quantitative genetics
 - ▶ Questions we want to answer
 - ▶ Animal model
- ▶ The animal model is a GMRF model
- ▶ Approximative inference animal model with gaussian traits
 - ▶ Study of morphological traits for Norwegian House Sparrows
 - ▶ Extended model, and study of Swiss Barn Owls.
- ▶ Approximative inference animal model with non-gaussian traits
 - ▶ Study of Norwegian House Sparrows
- ▶ What next, opportunities and (possible) limitations

Quantitative Genetics

- ▶ *Quantitative genetics is the study of continuous traits (such as height or weight) and its underlying mechanisms. The combined effect of the many underlying genes and different environmental effects results in a continuous distribution of trait*
- ▶ Trait = Genetic + environmental effects.
- ▶ In this study:
 - ▶ Wild life populations

Biological motivation

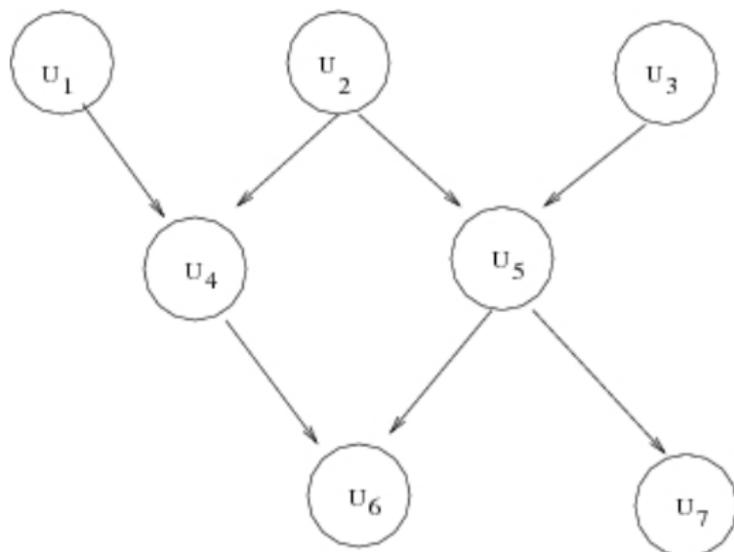
- ▶ An important quantitative genetic parameter is the *heritability* of a trait
- ▶ Speed and direction of evolution.
- ▶ Important for the ability to evolve and adapt to environmental changes, e.g. climate changes.
- ▶ Much used in animal and plant breeding.

Animal Model

Breeding value:

genetic part of the trait

- ▶ Animal model estimate the breeding values for each individual
- ▶ The breeding values are assumed Gaussian with a dependency given by the family structure, i.e. by the corresponding pedigree.



Model, one bird

For bird i ($i = 1, \dots, 3373$):

$$y_i = \sum_{f=1}^{n_f} \beta_{f,l(i)} + u_i + \epsilon_i,$$

- ▶ y_i : observations (i.e. wing length)
- ▶ β_i : “fixed effects” / “group effect” (sex, hatch year and island),
 $\beta \sim N(0, \sigma_\beta^2)$
- ▶ u_i : genetic effects, $u_i \sim N(0, \sigma_u)$
- ▶ ϵ : environmental effects: $\epsilon_i \sim N(0, \sigma_\epsilon)$

Animal model, for the population

$$y = B\beta + Xu + \epsilon$$

- ▶ $y = (y_1, y_2, \dots, y_{mdata})^T$
- ▶ B and X : incidence matrices.
- ▶ $\beta \sim N(0, \sigma_\beta^2 I)$
- ▶ u : genetic effect, $u = (u_1, u_2, \dots, u_{nind})$. $u_i \sim N(0, \sigma_u A)$
 - ▶ A : relationship matrix
- ▶ ϵ : environmental effect. $\epsilon_j \sim N(0, \sigma_\epsilon I)$
- ▶ Constraints: $\sum_{l=1}^L \beta_l = 0$ (for all but one fixed effect) and $\sum_{i=1}^n u_{it} = 0$

Animal model, for the population

- ▶ $\beta \sim N(0, \sigma_{\beta}^2 I)$
- ▶ $u \sim N(0, \sigma_u^2 A)$
- ▶ $\epsilon \sim N(0, \sigma_{\epsilon}^2 I)$
- ▶ The animal model is a (general) linear mixed model

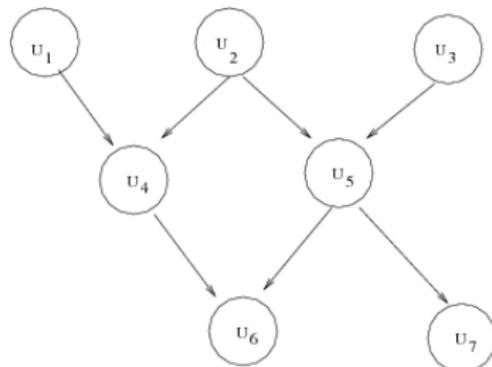
In combination with a complex multi-generational pedigrees these models allow us to estimate causal components of variance, to estimate heritabilities

Relationship matrix A

- ▶ $A_{ij} = 2 \times$ coefficient of coancestry.
- ▶ Coefficient of coancestry: Probability that allele picked at random identical by descent.
- ▶ A is nearly a full matrix.

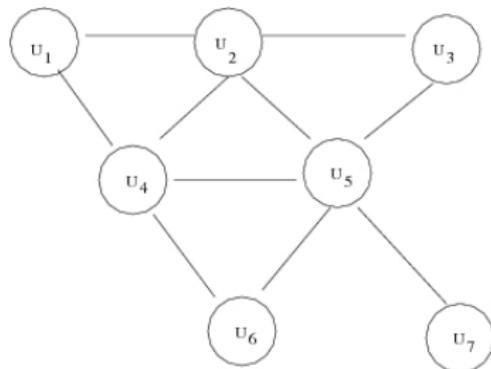
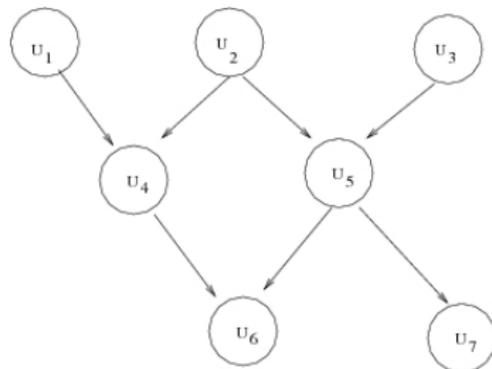
Relationship matrix A

- ▶ $A_{ij} = 2 \times$ coefficient of coancestry.
- ▶ Coefficient of coancestry: Probability that allele picked at random identical by descent.
- ▶ A is nearly a full matrix.
- ▶ Pedigree = DAG (Directed Acyclic Graph)



Relationship matrix A

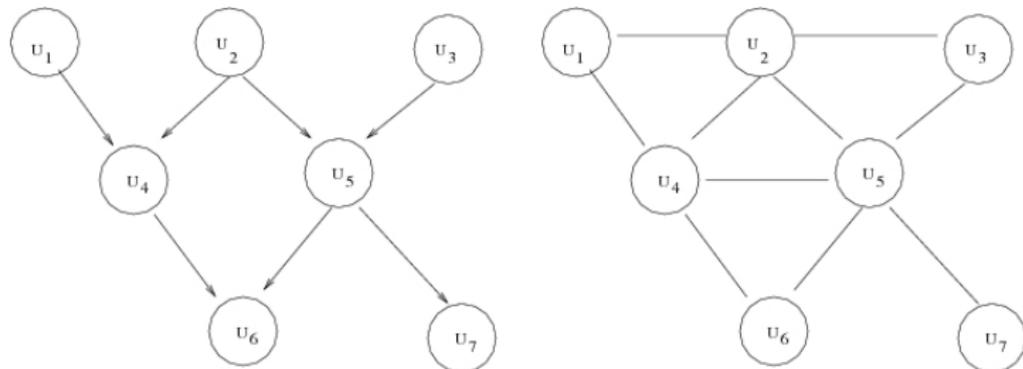
- ▶ $A_{ij} = 2 \times$ coefficient of coancestry.
- ▶ Coefficient of coancestry: Probability that allele picked at random identical by descent.
- ▶ A is nearly a full matrix.
- ▶ Pedigree = DAG (Directed Acyclic Graph)



- ▶ Structure of A^{-1} from moralising the pedigree $\Rightarrow A^{-1}$ sparse.

Relationship matrix A

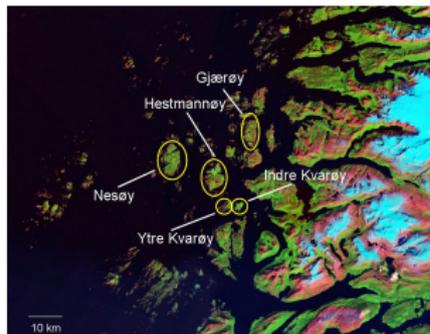
- ▶ $A_{ij} = 2 \times$ coefficient of coancestry.
- ▶ Coefficient of coancestry: Probability that allele picked at random identical by descent.
- ▶ A is nearly a full matrix.
- ▶ Pedigree = DAG (Directed Acyclic Graph)



- ▶ The conditional independence structure of the pedigree \implies animal model is a latent GMRF model.
- ▶ Calculate non-zero elements of A^{-1} as in Quaas (1976)

Data, islands

- ▶ We have studied house sparrow populations on five islands off the coast of Helgeland, Northern Norway, since 1993.
- ▶ Topography and habitat on the islands may result in different environmental conditions \implies different selection pressures in the house sparrow populations.



Hestmannøy

Data, islands

- ▶ The house sparrow (*Passer domesticus*) is a small passerine bird
 - ▶ Sexually dimorphic
 - ▶ Widely spread around the whole globe
- ▶ On these islands house sparrows live near human settlements, and mostly nest inside barns of dairy farms.
- ▶ Relatively easy to capture
- ▶ Low migration rates



Data Collection

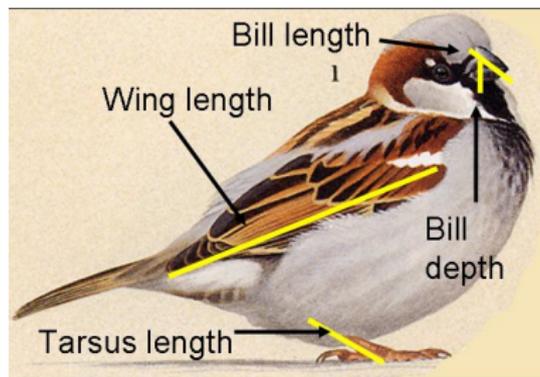
- ▶ Almost all fledglings are marked and blood sample collected in the nest.
 - ▶ Adult and fledged juveniles (birds born the same summer) were captured with mist nets during the summer, collected blood samples and morphological traits.
- ▶ >70% of the adult birds present on each island a given year were marked.



Morphological data

Traits measured for both sexes:

- ▶ Tarsus length
- ▶ Wing length
- ▶ Bill length
- ▶ Bill length
- ▶ Body mass



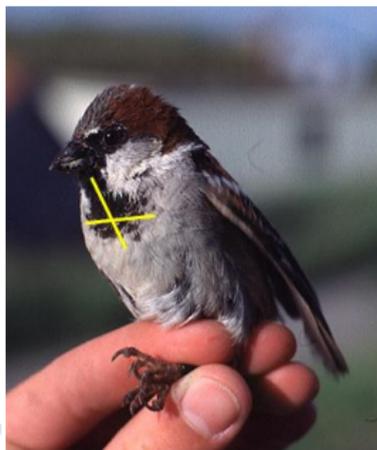
Morphological data

Traits measured for both sexes:

- ▶ Tarsus length
- ▶ Wing length
- ▶ Bill depth
- ▶ Bill length
- ▶ Body mass

For males only:

- ▶ Total badge size, missing for all females
- ▶ Visual badge size, missing for all females



Morphological data

Traits measured for both sexes:

- ▶ Tarsus length
- ▶ Wing length
- ▶ Bill depth
- ▶ Bill length
- ▶ Body mass

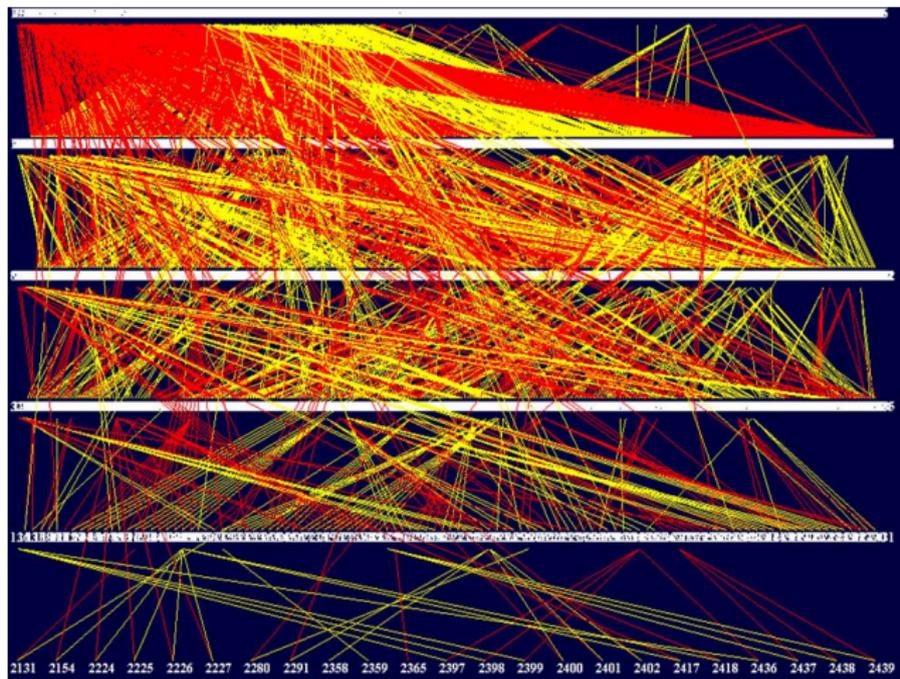
For males only:

- ▶ Total badge size, **missing for all females**
- ▶ Visual badge size, **missing for all females**

Have sex, hatch year and hatch island for all birds.

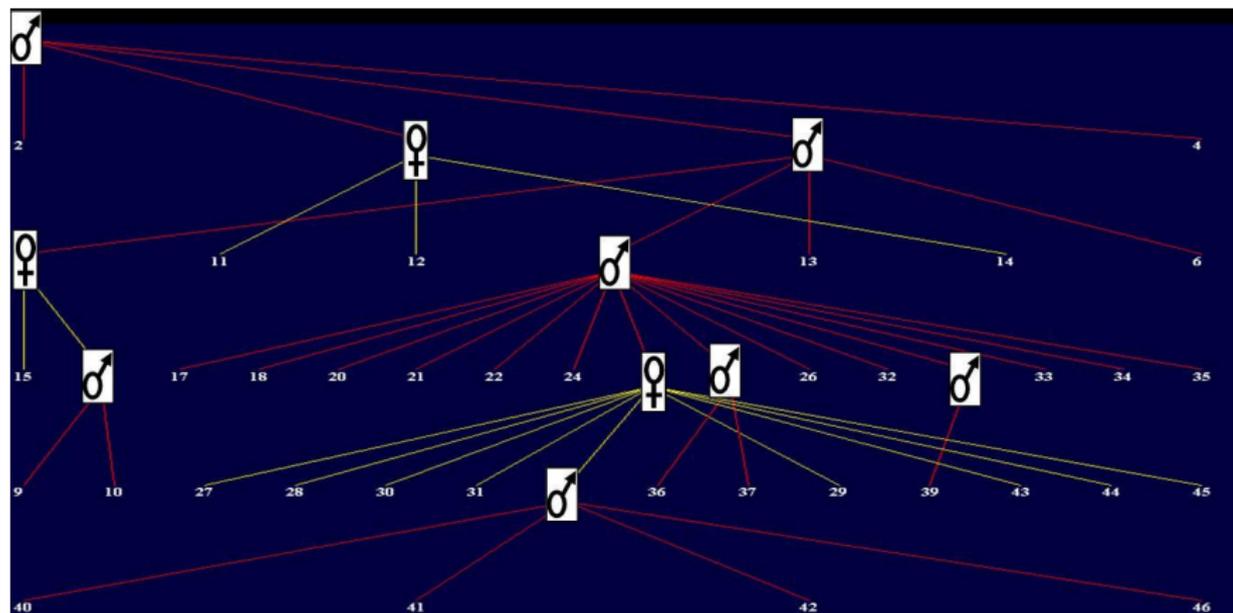
Pedigree data

- ▶ Blood samples \Rightarrow DNA \Rightarrow capture and observational data \Rightarrow Pedigree



Pedigree data

- ▶ Blood samples \Rightarrow DNA \Rightarrow capture and observational data \Rightarrow Pedigree



Summary Data

- ▶ 3572 birds in the pedigree (from 1993 - 2002)
- ▶ 1004 with measured traits as one year olds.
- ▶ Have up to 7 traits for adult birds.
 - ▶ There are missing data, e.g. females do not have badge.
- ▶ Sex, hatch year and island is known for all the birds.



Approximative inference for Gaussian Animal Model

Conditioned on hyper-parameters 'everything' is Gaussian.

Our model formulation:

$$y = B\beta + Xu_A + \epsilon$$

Non-gaussian hyper-parameters:

$$u \sim N(0, \sigma_u^2 A), \epsilon \sim N(0, \sigma_\epsilon^2 I).$$

Approximative inference for Gaussian Animal Model

Conditioned on hyper-parameters 'everything' is Gaussian.

Inla model formulation:

- ▶ Likelihood: $y_i | \eta_i \sim \pi(y_i | \eta_i)$
- ▶ Latent field: $\eta_i = \sum f_k(c_{ki}) + \mathbf{z}_i^T \beta + \epsilon_i$

Our model formulation:

$$y = B\beta + Xu_A + \epsilon$$

Non-gaussian hyper-parameters:

$$u \sim N(0, \sigma_u^2 A), \epsilon \sim N(0, \sigma_\epsilon^2 I).$$

Approximative inference for Gaussian Animal Model

Conditioned on hyper-parameters 'everything' is Gaussian.

Inla model formulation:

- ▶ Likelihood: $y_i | \eta_i \sim \pi(y_i | \eta_i)$
- ▶ Latent field: $\eta_i = \sum f_k(c_{ki}) + z_i^T \beta + \epsilon_i$

Our model formulation:

$$y = B\beta + Xu_A + \epsilon$$

Animal model Inla

- ▶ Likelihood: $y_i | \eta_i \sim N(\eta_i, 1/10)$
- ▶ Latent field: $\eta_i = \sum_{l=1}^3 \beta_{g(l,i)} + u_i + \epsilon_i$

Non-gaussian hyper-parameters:

$$u \sim N(0, \sigma_u^2 A), \epsilon \sim N(0, \sigma_\epsilon^2 I).$$

Two case-studies

Norwegian House Sparrows



with Henrik Jensen,
Department of Biology, NTNU

- ▶ Heritability?
- ▶ Evolution?

Swiss Barn Owls



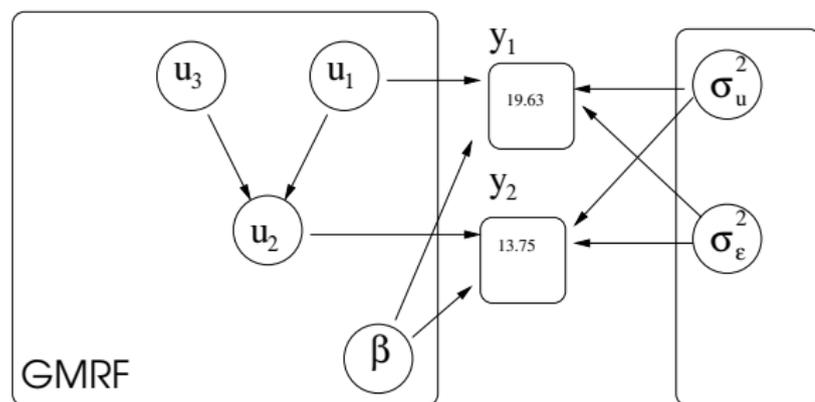
with Alexandre Roulin,
Lausanne, Switzerland,
Henrik Jensen et al

Finding posterior heritability

$$h^2 = \frac{\sigma_u^2}{\sigma_\epsilon^2 + \sigma_u^2}$$

- ▶ A function of two hyper-parameters.
- ▶ Need joint posterior; $\pi(\sigma_u^2, \sigma_\epsilon^2 | y)$
- ▶ Use inla options
 - ▶ `FP_HYPERPARAM = hyper.dat,`
 - ▶ `int_strategy = grid`
 - ▶ `hessian_force_diagonal=1`
- ▶ Need many evaluations \Rightarrow Computationally expensive

Evolution?



Do the breeding values change over time?

$$\sum_{i \in \text{year1}} u_i \neq \sum_{i \in \text{year5}} u_i$$

- Posterior of a linear combination of some latent field variables.

House Sparrow Case Study

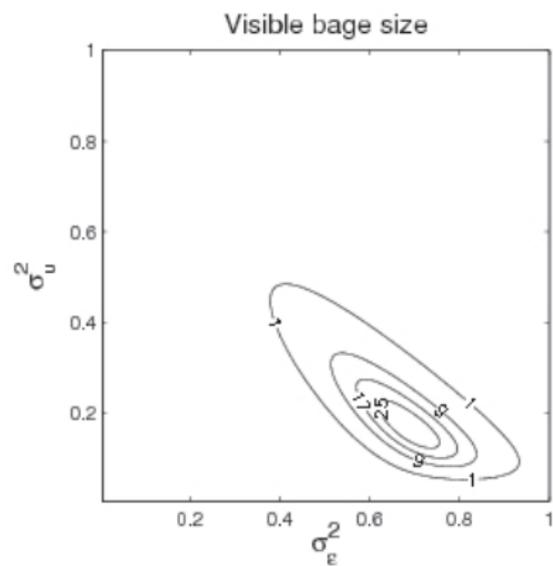
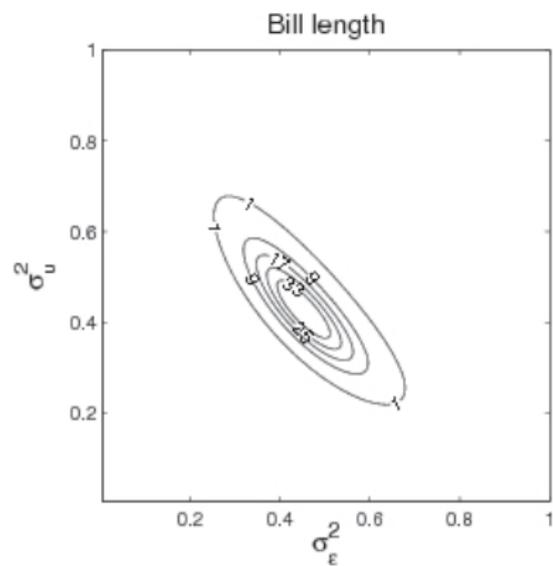
Explore heritability and look for evolution:



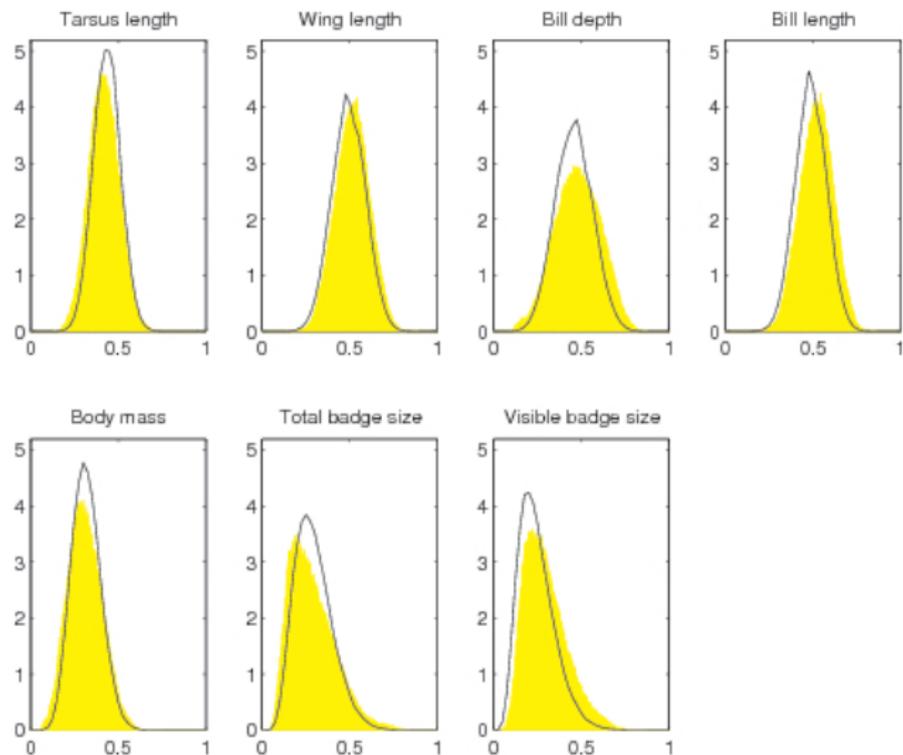
Joint work with Henrik Jensen.

Utilising Gaussian Markov Random Field properties of Bayesian Animal models In revision for Biometrics.

Joint Posterior for $(\sigma_u^2, \sigma_\epsilon^2)$

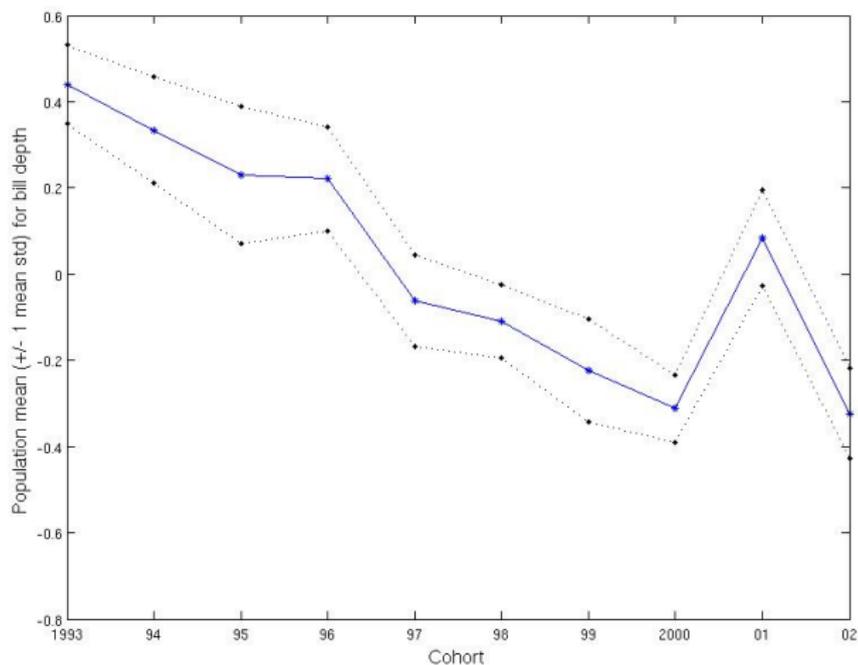


Heritability

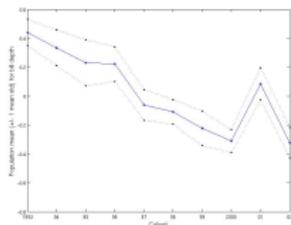


Evolution of bill depth?

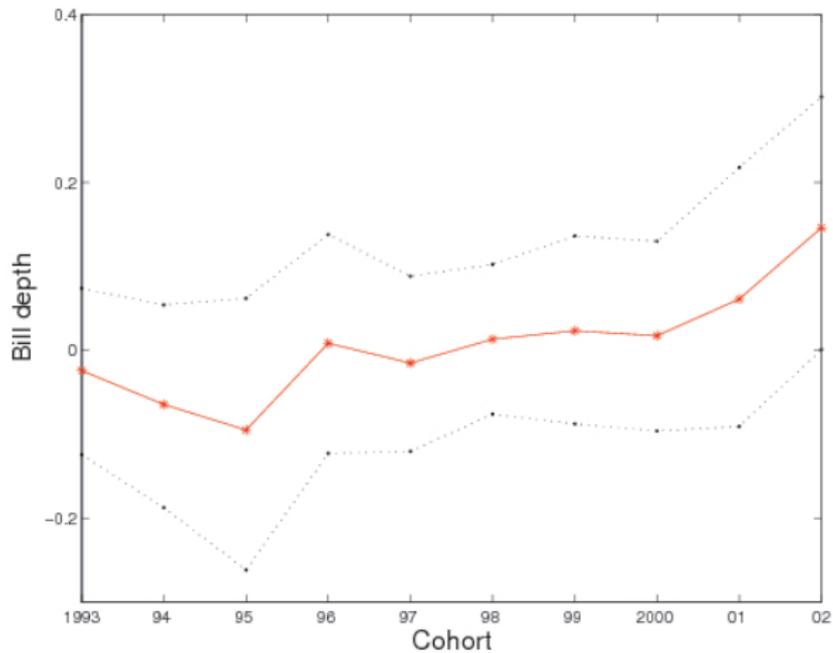
Observed bill depth vs hatch year:



Evolution of bill depth?



Breeding value bill depth vs hatch year:



Animal model for several traits

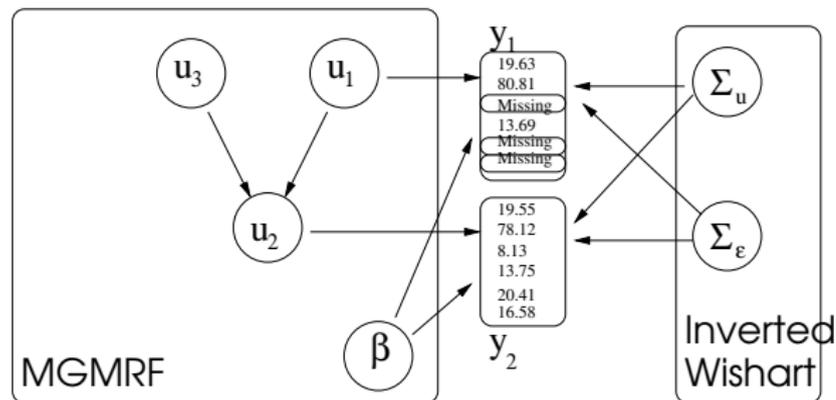
Why?

- ▶ From Kruuk et al (2002), Evolution *Antler size in red deer: Heritability and selection, but no evolution*
- ▶ One hypotheses: *Genetic correlation between a given trait and other traits under selection will constrain the direction and pace of any evolutionary change*

Animal model for several traits

How?

$$y = B\beta + Xu + \epsilon$$

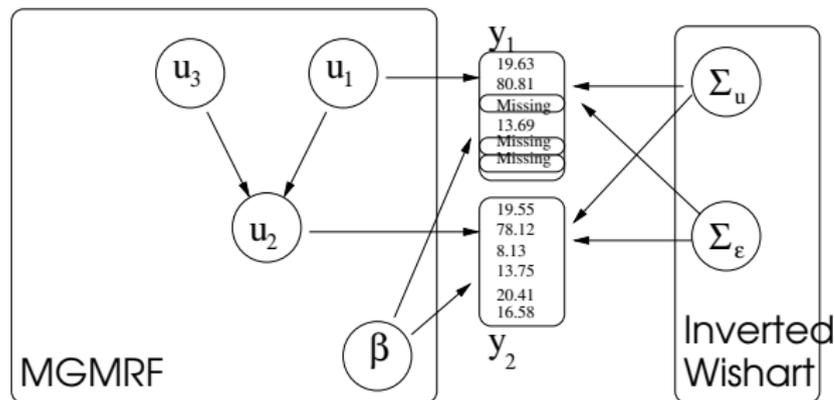


- ▶ $u_i \sim N(0, \Sigma_u)$, $u \sim N(0, \Sigma_u \otimes A)$
- ▶ $\epsilon_i \sim N(0, \Sigma_\epsilon)$, $\epsilon \sim N(0, \Sigma_\epsilon \otimes I)$

Animal model for several traits

How?

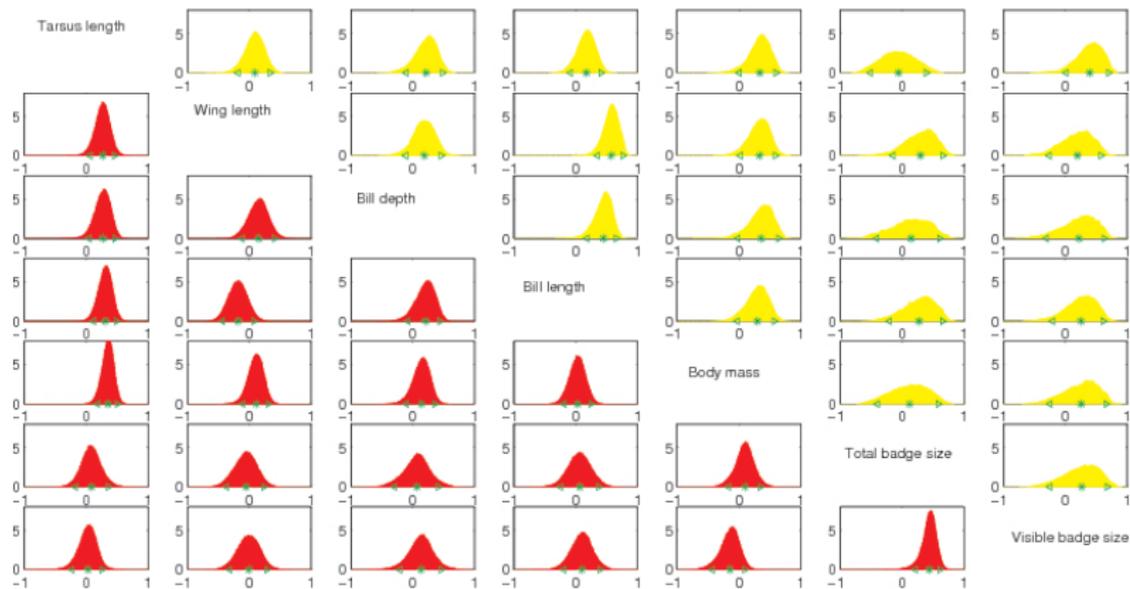
$$y = B\beta + Xu + \epsilon$$



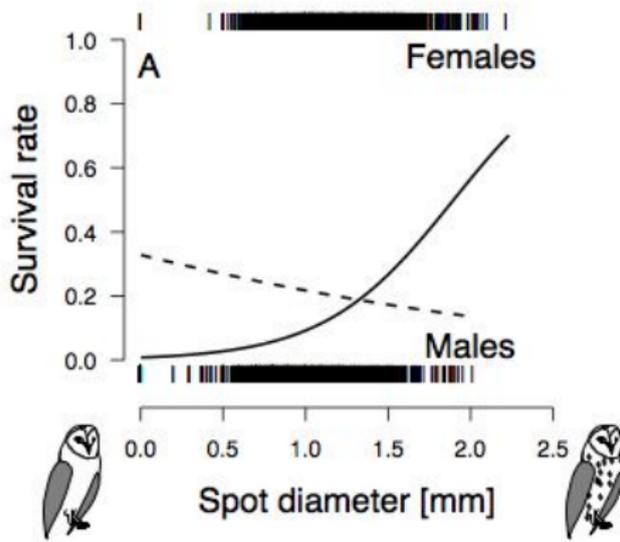
Problem

- ▶ A model with n -traits has $n(n + 1)$ non-Gaussian hyperparameters.
- ▶ Inla can do up to 5 non-Gaussian hyperparameters

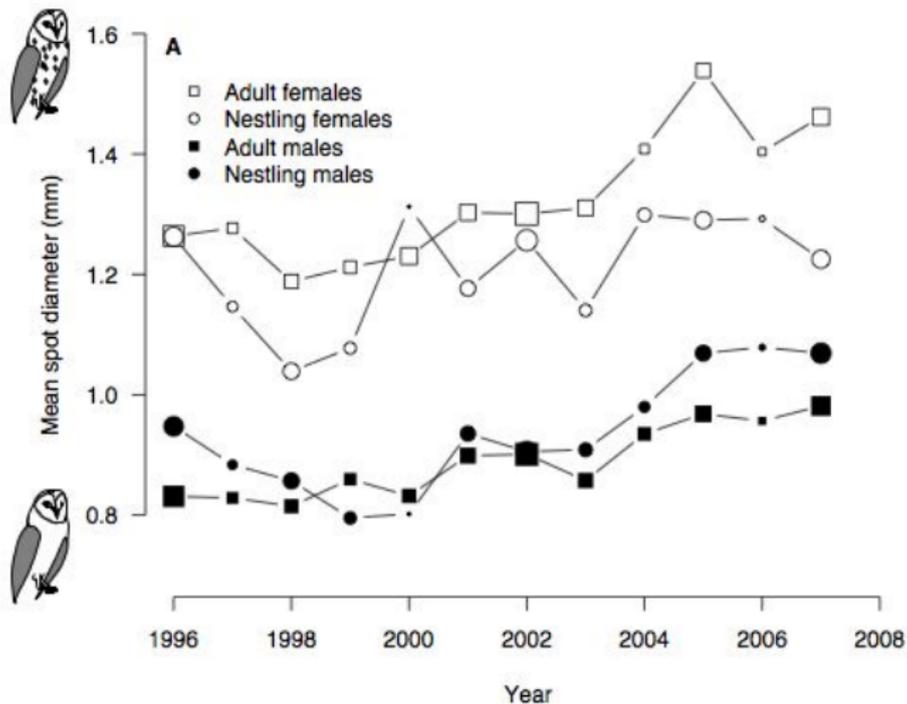
Estimated correlations



Spot diameter and survival



Evolution in spots?



Sex specific inheritance

- ▶ Autosomal: 'Normal genes'
- ▶ Sex-linked:
 - Humans: Males: XY , Females XX
 - Birds: Males: ZZ , Females ZW

A different dependence structure for sex-linked genes.

Extended animal model for the population

$$y = B\beta + Xu_A + Xu_Z + \epsilon$$

- ▶ $y = (y_1, y_2, \dots, y_{mdata})^T$
- ▶ B and X : incidence matrices.
- ▶ u_A : genetic effect autosomal genes. $u_A \sim N(0, \sigma_{u_A}^2 A)$
- ▶ A : relationship matrix autosomal genes
- ▶ u_Z : genetic effect Z-linked genes. $u_Z \sim N(0, \sigma_{u_Z}^2 A_Z)$
- ▶ A_Z : relationship matrix z-linked genes
- ▶ ϵ : environmental effect $\epsilon \sim N(0, \sigma_\epsilon^2 I)$

Extended animal model for the population

$$y = B\beta + Xu_A + Xu_Z + \epsilon$$

- ▶ $y = (y_1, y_2, \dots, y_{mdata})^T$
- ▶ B and X : incidence matrices.
- ▶ u_A : genetic effect autosomal genes. $u_A \sim N(0, \sigma_{u_A}^2 A)$
- ▶ A : relationship matrix autosomal genes
- ▶ u_Z : genetic effect Z-linked genes. $u_Z \sim N(0, \sigma_{u_Z}^2 A_Z)$
- ▶ A_Z : relationship matrix z-linked genes
- ▶ ϵ : environmental effect $\epsilon \sim N(0, \sigma_\epsilon^2 I)$

Relationship matrix A_Z

- ▶ $A_Z(i, j)$: $2 \times$ coefficient of coancestry for Z-linked genes.
- ▶ Structure of A_Z^{-1} , sparser than A^{-1} (no link between mothers and daughters).

Extended animal model and INLA

Inla model formulation:

- ▶ Likelihood: $y_i | \eta_i \sim \pi(y_i | \eta_i)$
- ▶ Latent field: $\eta_i = \sum f_k(c_{ki}) + z_i^T \beta + \epsilon_i$

Our model formulation:

$$y = B\beta + Xu_A + Xu_Z + \epsilon$$

Animal model Inla

- ▶ Likelihood: $y_i | \eta_i \sim N(\eta_i, 1/10)$
- ▶ Latent field: $\eta_i = \sum_{l=1}^3 \beta_{g(l,i)} + u_{Ai} + u_{Zi} + \epsilon_i$

Extended animal model and INLA

Inla model formulation:

- ▶ Likelihood: $y_i | \eta_i \sim \pi(y_i | \eta_i)$
- ▶ Latent field: $\eta_i = \sum f_k(c_{ki}) + z_i^T \beta + \epsilon_i$

Our model formulation:

$$y = B\beta + Xu_A + Xu_Z + \epsilon$$

Animal model Inla

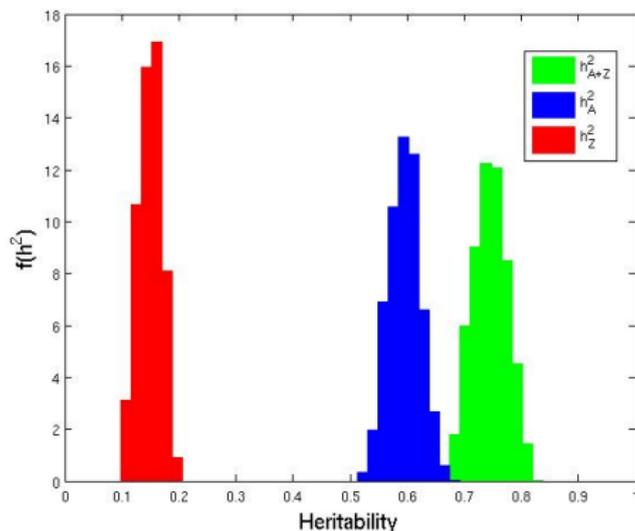
- ▶ Likelihood: $y_i | \eta_i \sim N(\eta_i, 1/10)$
- ▶ Latent field: $\eta_i = \sum_{l=1}^3 \beta_{g(l,i)} + u_{Ai} + u_{Zi} + \epsilon_i$

Hyper-parameters:

$$u_A \sim N(0, \sigma_{ua}^2 A), \quad u_Z \sim N(0, \sigma_{uz}^2 A_z), \quad \epsilon \sim N(0, \sigma_\epsilon^2 I).$$

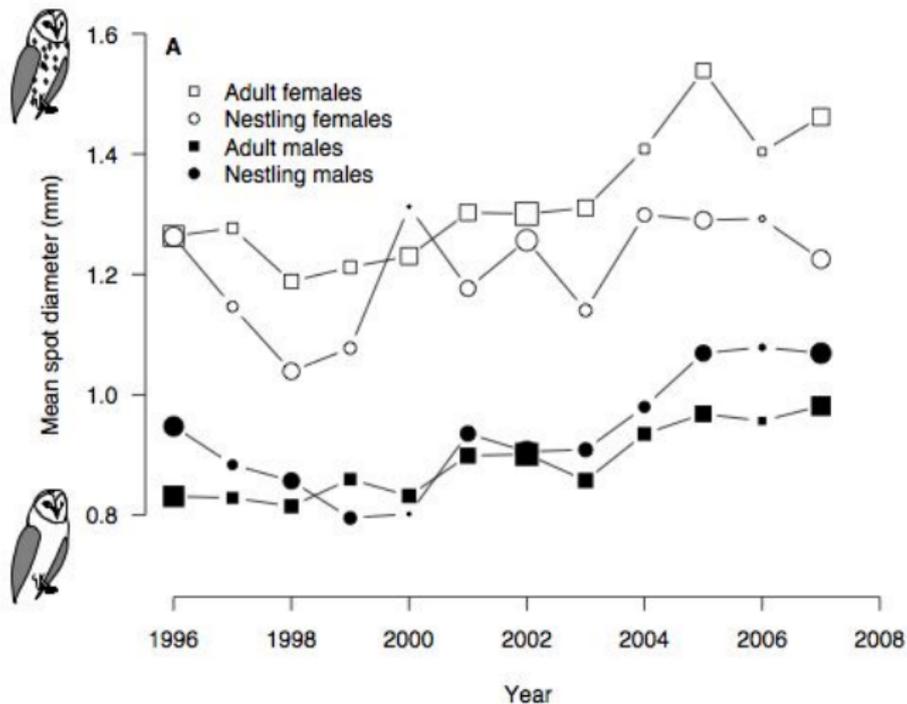
Heritabilities

- ▶ Heritability females: 0.59
- ▶ Heritability males: 0.82

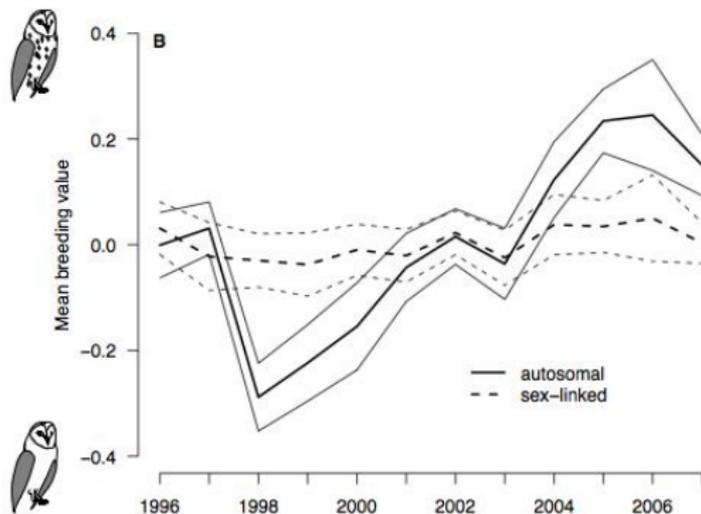
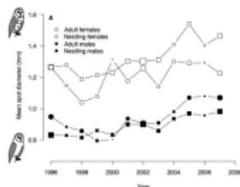


- ▶ Autosomal heritability: 0.59
- ▶ Z-linked heritability: 0.15

Evolution in spots?

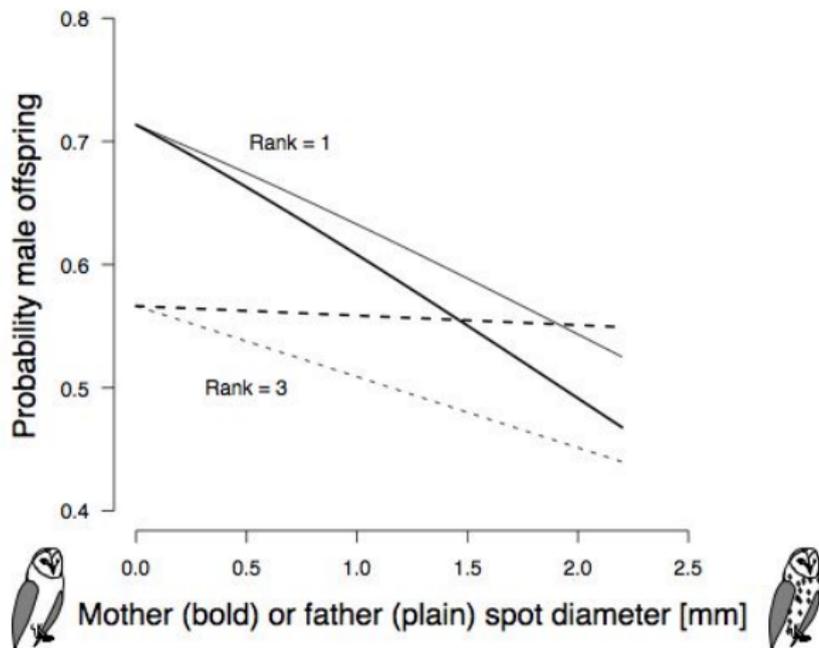


Evolution in spots?



Conclusion

Masculinized females invest in sons and feminized males in daughters. Submitted paper by *A Roulin, R Altwegg, H Jensen, I Steinsland and M Schaub*



TO DO: Z-linked extended animal model

- ▶ Need non-zero values for A_z^{-1} .
- ▶ Have used MENDEL til find A_z and inverted in Matlab

Think it is possible to find an algorithm similar to the one we use for A^{-1} .

Approximative inference for non-Gaussian animal model

- ▶ Important fitness-related traits have a non-Gaussian distribution;
 - ▶ lifespan
 - ▶ reproductive success
 - ▶ dispersal behavior
 - ▶ survival (dead or alive)

Approximative inference for non-Gaussian animal model

- ▶ Important fitness-related traits have a non-Gaussian distribution;
 - ▶ lifespan
 - ▶ reproductive success
 - ▶ dispersal behavior
 - ▶ survival (dead or alive)
- ▶ Estimating the heritability (h^2) of fitness-related non-Gaussian traits.

Is dispersal inheritable?

Data

- ▶ Pedigree
- ▶ Migrated or not (1/0).
- ▶ Covariates (hatch year, hatch island, sex) as group effects.



Model, one bird

For bird i ($i = 1, \dots, 3292$):

$$y_i \sim \text{bin}(p_i, 1)$$

- ▶ p_i : Probability of moving

Logit-link:

$$p_i = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)}$$

η_i =group +genetic +environmental

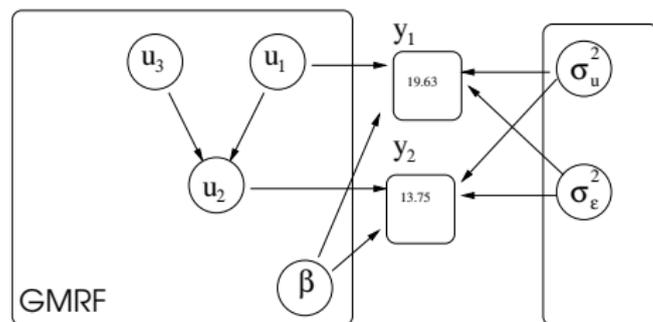
$$\eta_i = \sum_{f=1}^{n_f} \beta_{f,I(i)} + u_i + \epsilon_i$$

Animal model, for the population

$$\eta = B\beta + Xu + \epsilon$$

- ▶ β : group effect, $\beta \sim N(0, \sigma_\beta^2 I)$.
- ▶ ϵ : environmental/individual effect, $\epsilon \sim N(0, \sigma_\epsilon^2 I)$.
- ▶ u : genetic effect, $u \sim N(0, \sigma_u A)$.
- ▶ A : relationship matrix
- ▶ Non-gaussian hyperparameters: σ_u^2 and σ_ϵ^2
- ▶ Constraints: $\sum_{l=1}^L \beta_l = 0$ (for all but one group effect) and $\sum_{i=1}^n u_{it} = 0$

Non-Gaussian animal model



- ▶ u : genetic effect, $u = (u_1, u_2, \dots, u_{nind})$. $u \sim N(0, \sigma_u A)$
- ▶ A : relationship matrix
- ▶ $\epsilon \sim N(0, \sigma_\epsilon^2 I)$

Non-gaussian trait animal model is a general linear mixed model / latent GMRF model.

Inla and non-Gaussian animal model

Inla model formulation:

- ▶ Likelihood: $y_i | \eta_i \sim \pi(y_i | \eta_i)$
- ▶ Latent field: $\eta_i = \sum f_k(c_{ki}) + z_i^T \beta + \epsilon_i$

Our model formulation:

- ▶ $y_i | \eta_i \sim \text{bin}(\text{logit}(\eta_i), 1)$
- ▶ $\eta = B\beta + Xu + \epsilon$

Hyper-parameters:

$$u \sim N(0, \sigma_u^2 A), \quad \epsilon \sim N(0, \sigma_\epsilon^2 I).$$

Inla and non-Gaussian animal model

Inla model formulation:

- ▶ Likelihood: $y_i | \eta_i \sim \pi(y_i | \eta_i)$
- ▶ Latent field: $\eta_i = \sum f_k(c_{ki}) + z_i^T \beta + \epsilon_i$

Non-gaussian animal model Inla

- ▶ Likelihood: $y_i | \eta_i \sim \text{bin}(\text{logit}(\eta_i), 1)$
- ▶ Latent field: $\eta_i = \sum_{l=1}^3 \beta_{g(l,i)} + u_i + \epsilon_i$

Our model formulation:

- ▶ $y_i | \eta_i \sim \text{bin}(\text{logit}(\eta_i), 1)$
- ▶ $\eta = B\beta + Xu + \epsilon$

Hyper-parameters:

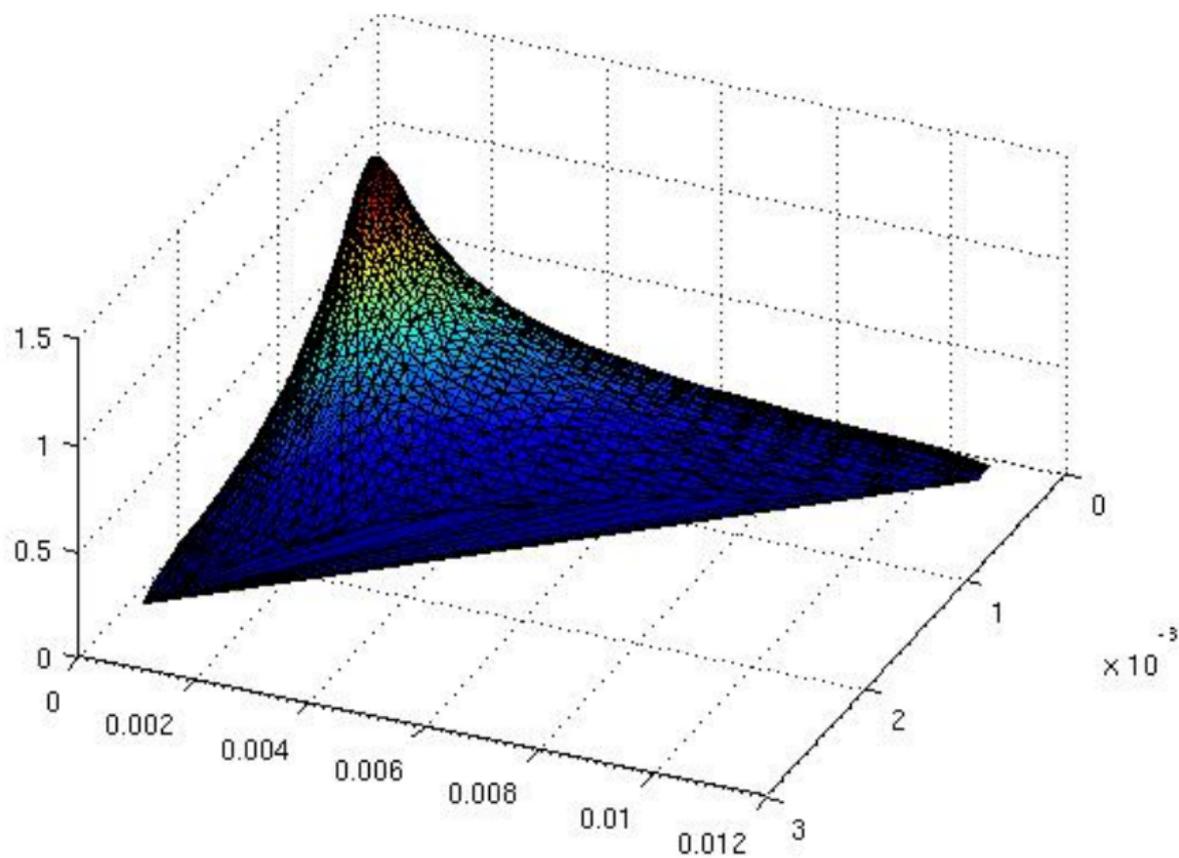
$$u \sim N(0, \sigma_u^2 A), \epsilon \sim N(0, \sigma_\epsilon^2 I).$$

Heritability and INLA

$$h^2 = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\epsilon^2}$$

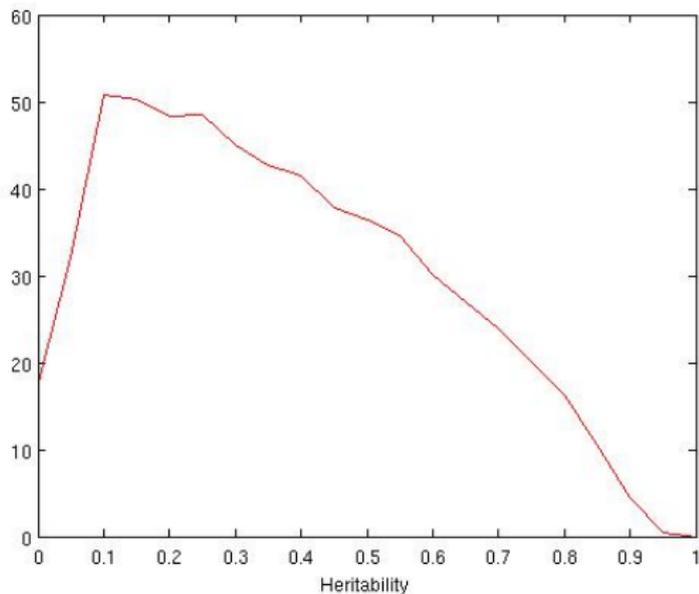
- ▶ Need $\pi(\sigma_u^2, \sigma_\epsilon^2 | y)$
- ▶ Evaluate joint posterior on a grid of variance values.
- ▶ Most specify the Q-matrix ourself, and use the generic option.

Posterior of σ_{μ}^2 and σ_{ϵ}^2

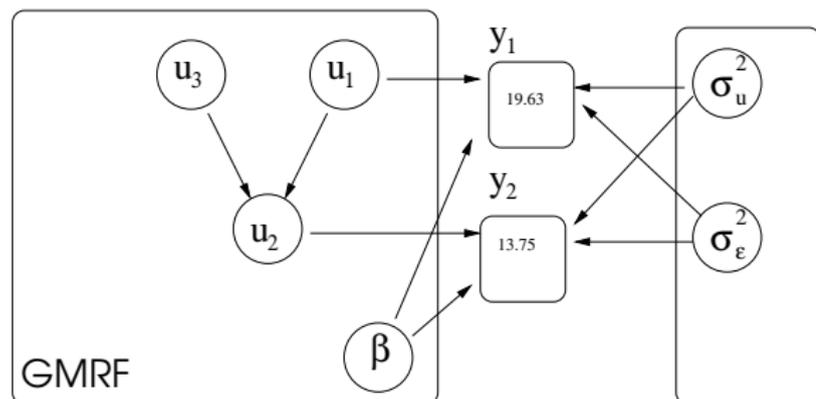


Posterior of heritability

Heritability: $h^2 = \sigma_u^2 / (\sigma_u^2 + \sigma_\epsilon^2)$



Summary INLA and animal model



- ▶ 'Pedigree dependence' gives GMRF structure.
- ▶ Animal model = latent GMRF model.
 - ▶ Gaussian traits OK
 - ▶ Non-gaussian traits OK
- ▶ Constraints
- ▶ Heritability: function of two hyper-parameters.
- ▶ Evolution of breeding values: linear combination of some latent variables.

Experience using INLA

- ▶ Making the input-files takes most of the time.
- ▶ Have made the Q -matrix, and used generic option.
- ▶ Easy to modify /extend a model
- ▶ Joint posterior for hyper-parameters, computationally expensive. (and we could probably do this smarter)
- ▶ Easy to make requests such that computation time is a problem.

Limitations

- ▶ Two or more traits not possible (yet)

Our plans

- ▶ **Animal R-Inla** R-package for animal model
 - ▶ Make Q-matrix from pedigree.
 - ▶ Good default settings for finding $\pi(\sigma_u^2, \sigma_\epsilon^2 | y)$
 - ▶ Calculate $\pi(h^2 | y)$
 - ▶ Calculate $\sum_{i \in \text{year}} u_i$
 - ▶ Easy for biologists and fast (an alternative for REML).
- ▶ Use Animal R-Inla for analysing our house sparrow dataset.
- ▶ Extend the animal model.

