Animal Breeding and Genetics Group
College of Agricultural & Environmental Sciences
UNIVERSITY OF GEORGIA

DHGLMF90: A software tool using double hierarchical generalized linear models for estimating the genetic heterogeneity of residual variance

A. Alvarez Munera, M. Bermann, N. Ibáñez Escriche, C. Casto Rebollo, I. Misztal, D. Lourenco

2024 ASAS Annual Meeting
Calgary, Canada

Motivation

Differences in residual variances among individuals are common

1. Interest in selecting uniform animals expressing consistent responses
2. Modeling heteroskedastic residuals when estimating breeding values
3. Canalisation: Selection to reduce the environmental variability

alejandra.alvarezm@uga.edu

Motivation (cont.)

Hierarchical single-trait model

Mean
Trait performance

\+

Dispersion
Variability

A global criterion
Selection for mean and against the variance

Selection index: $(\bar{y} - y_0)^2 + S_y^2$

alejandra.alvarezm@uga.edu

Problem

Bayesian methods

- Can estimate heteroskedastic residuals
- Metropolis-Hastings algorithm
- Inference with posterior distribution
- **Slow and inefficient**

Double Hierarchical GLM (DHGLM)

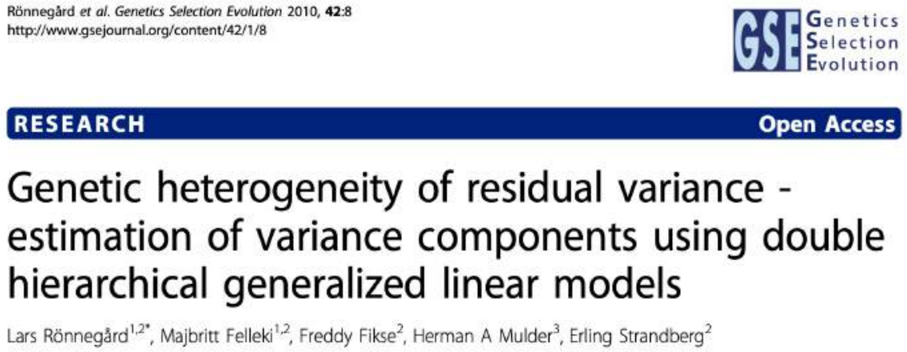
- Lee & Nelder, 2006
- Rönnegård et al., 2010
- No inference with posterior distribution
- **Faster counterpart → Alternative**

Objective

Develop a software named DHGLMF90 to implement DHGLM to estimate heterogeneous residual variances and breeding values

DHGLM

- Similar performance
  - Bayesian methods
  - H-Likelihood
- Bivariate normal-normal model
- Challenges
  - Convergence issues
  - 27/100 replicates failed

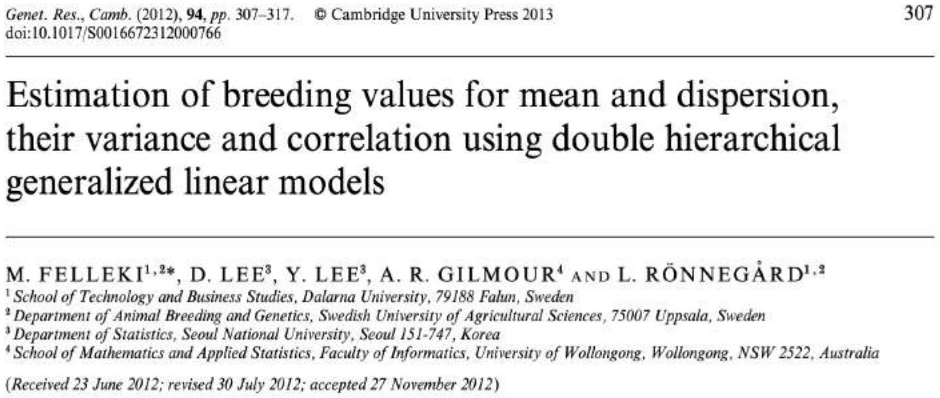


Rönnegård *et al. Genetics Selection Evolution* 2010, **42**:8
http://www.gsejournal.org/content/42/1/8

RESEARCH **Open Access**

Genetic heterogeneity of residual variance - estimation of variance components using double hierarchical generalized linear models

Lars Rönnegård[1,2*], Majbritt Felleki[1,2], Freddy Fikse[2], Herman A Mulder[3], Erling Strandberg[2]



Genet. Res., Camb. (2012), **94**, *pp.* 307–317. © Cambridge University Press 2013 307
doi:10.1017/S0016672312000766

Estimation of breeding values for mean and dispersion, their variance and correlation using double hierarchical generalized linear models

M. FELLEKI[1,2*], D. LEE[3], Y. LEE[3], A. R. GILMOUR[4] AND L. RÖNNEGÅRD[1,2]
[1] *School of Technology and Business Studies, Dalarna University, 79188 Falun, Sweden*
[2] *Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, 75007 Uppsala, Sweden*
[3] *Department of Statistics, Seoul National University, Seoul 151-747, Korea*
[4] *School of Mathematics and Applied Statistics, Faculty of Informatics, University of Wollongong, Wollongong, NSW 2522, Australia*

(*Received 23 June 2012; revised 30 July 2012; accepted 27 November 2012*)

Algorithm

Iteratively calculates variance components and leverages for a bivariate model

Initialize weights $\Sigma^{-1} = \begin{bmatrix} \boldsymbol{\phi}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathrm{diag}(\frac{1-q}{2}) \end{bmatrix}$ and working variables Z_d

Fit correlation structure $\mathbf{G} \otimes \mathbf{A}$ between a and a_d, but y and Z_d conditional uncorrelated

Update σ^2, e and q, and thereby Z_d and $\mathrm{diag}((1-q)/2)$ in $\boldsymbol{\Sigma}^{-1}$

Repeat step 2

Update $\boldsymbol{\phi}^{-1}$ and update $\boldsymbol{\Sigma}^{-1}$

Reweighted
least squares
Algorithm (IRWLS)

alejandra.alvarezm@uga.edu

Improving the algorithm

$$H = \begin{bmatrix} X^T \boldsymbol{\phi}^{-1} X & X^T \boldsymbol{\phi}^{-1} Z \\ Z^T \boldsymbol{\phi}^{-1} X & Z^T \boldsymbol{\phi}^{-1} Z + \frac{1}{(1-\rho^2)\sigma_a^2} A^{-1} \end{bmatrix}$$

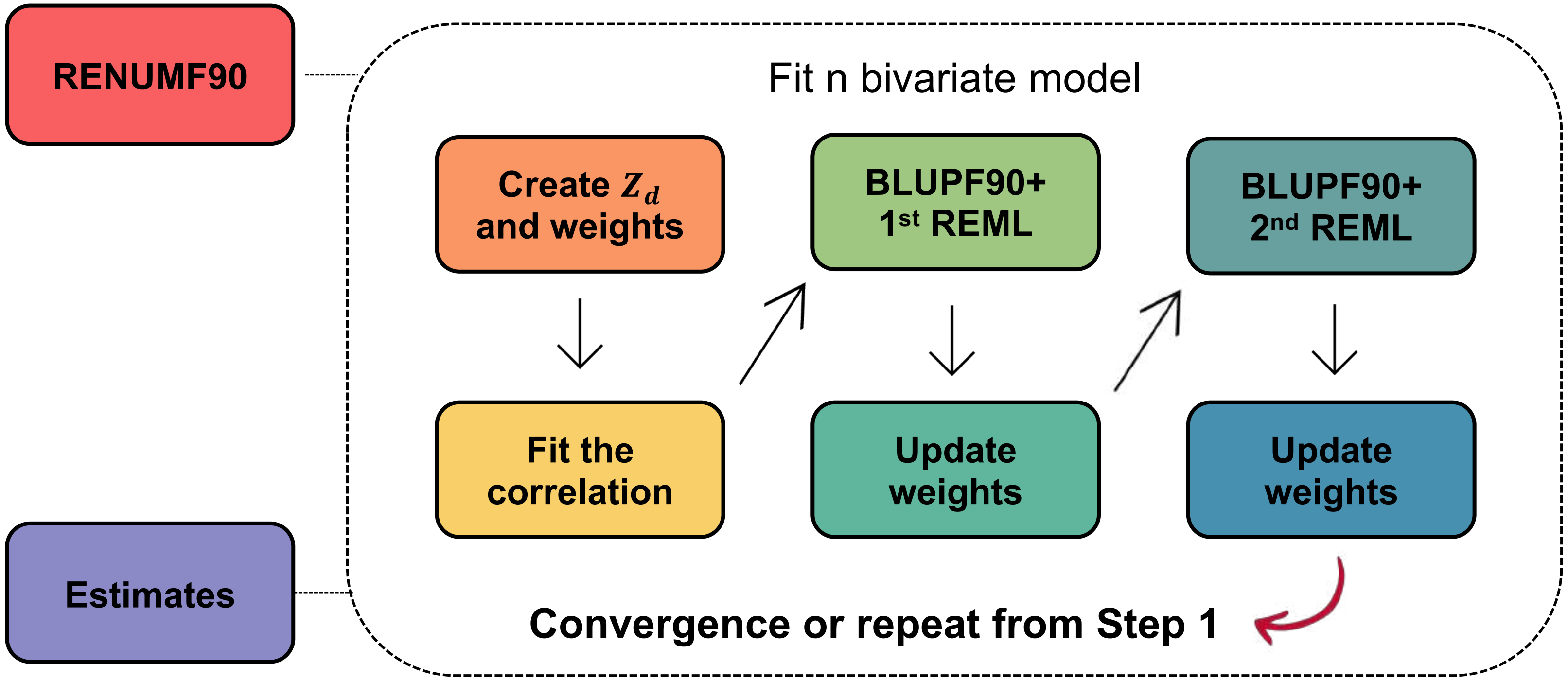
Leverages

Felleki et al., (2012)

$$C = \begin{bmatrix} X^T \boldsymbol{\phi}^{-1} X & 0 & X^T \boldsymbol{\phi}^{-1} Z & 0 \\ 0 & X_d^T \operatorname{diag}(\frac{1-q}{2}) X_d & 0 & X_d^T \operatorname{diag}\left(\frac{1-q}{2}\right) Z \\ Z^T \boldsymbol{\phi}^{-1} X & 0 & Z^T \boldsymbol{\phi}^{-1} Z + \frac{1}{(1-\rho^2)\sigma_a^2} A^{-1} & -\frac{\rho}{(1-\rho^2)\sqrt{\sigma_a^2 \sigma_{a_d}^2}} A^{-1} \\ 0 & Z^T \operatorname{diag}(\frac{1-q}{2}) X_d & -\frac{\rho}{(1-\rho^2)\sqrt{\sigma_a^2 \sigma_{a_d}^2}} A^{-1} & Z^T \operatorname{diag}\left(\frac{1-q}{2}\right) Z + \frac{1}{(1-\rho^2)\sigma_{a_d}^2} A^{-1} \end{bmatrix}$$

DHGLMF90 flowchart

DHGLMF90 uses BLUPF90+ to estimate variance components using REML



RENUMF90

Fit n bivariate model

Create Z_d and weights

Fit the correlation

BLUPF90+ 1st REML

Update weights

BLUPF90+ 2nd REML

Update weights

Convergence or repeat from Step 1

Estimates

Renumber the data

Initialize weights $\boldsymbol{\Sigma}^{-1}$ and Z_d

Fit the correlation $\mathbf{G} \otimes \mathbf{A}$ for a/a_d

REML for the mean part

Update σ_a^2, Zd and $\mathrm{diag}((1-q)/2)$

REML for the dispersion part

Update $\sigma_{a_d}^2$, $\boldsymbol{\phi}^{-1}$ and weights $\boldsymbol{\Sigma}^{-1}$

VCE + EBV + Residuals

alejandra.alvarezm@uga.edu

Data sets

	Data 1	Data 2
Trait	Pigs littler size	Daily feed intake
Animals in the pedigree	51	5460
Records	306	109680
Effects	Additive + PE	Additive + PE

↓

Simulated data

Felleki et al., 2012

Test performance

↓

Real data

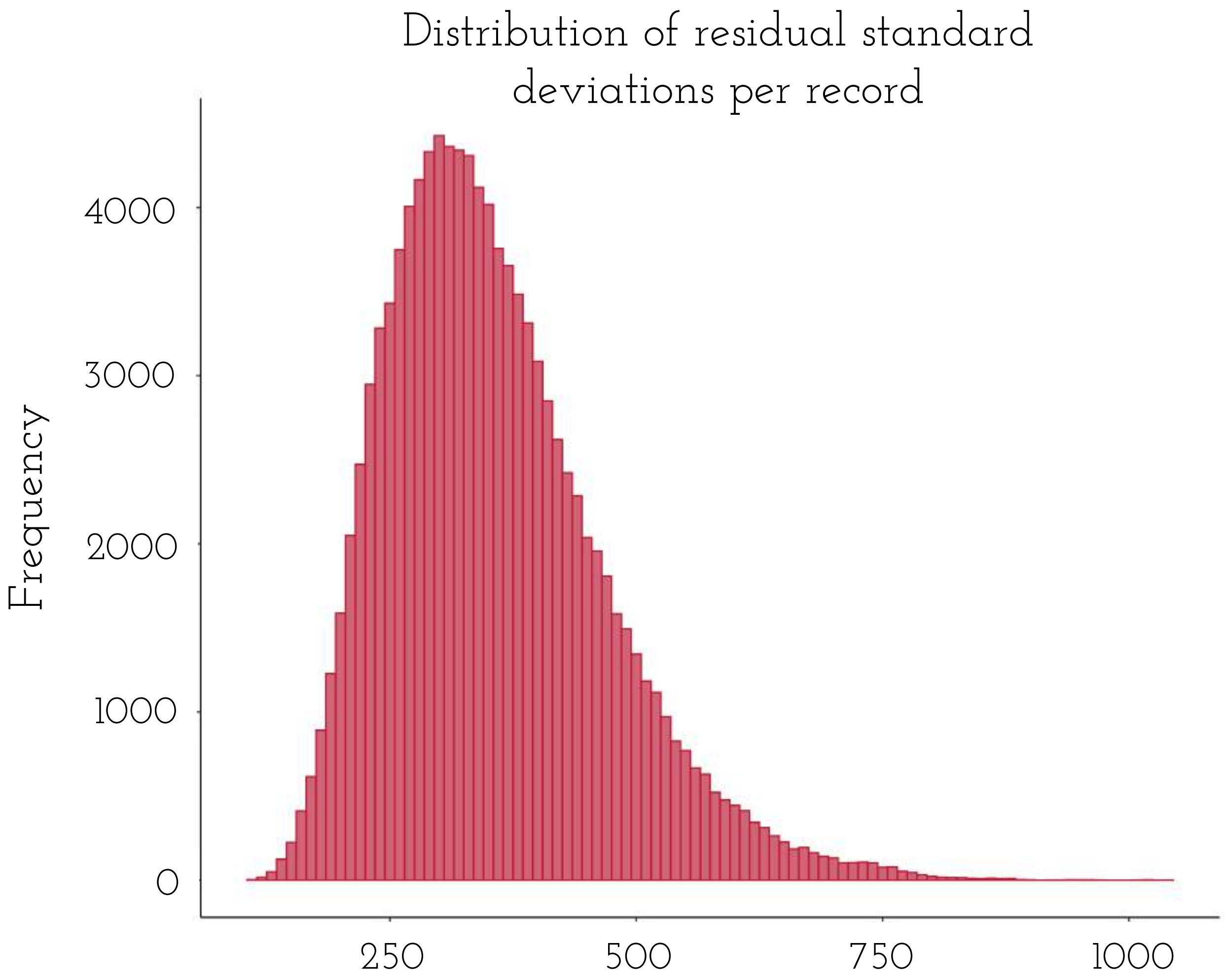
Pietrain pigs

Test method

Data - Results

1. All replicates converged
2. No significant differences

Method	σ_a^2	Convergence
H-Likelihood	1.0	97/100
IRWLS	1.0	74/100
DHGLMF90	1.0	100/100



Computing times

DHGLMF90 converged in order of magnitude faster than Bayesian methods

Method	Data 1	Data 2
DHGLMF90	1 min	10 min
Bayes	15 min	>1 day

X10 computing time **X144 computing time**

Other implementations

- Variance components estimates for Uterine capacity
  - Corr = 0.98
  - Corr = 0.99

Ibáñez-Escriche *et al. (2008; 2024)*

Program	σ_a^2	$\sigma_{a_v}^2$	σ_p^2	$\sigma_{p_v}^2$	r_g	Time
DHGLMF90	0.85 (0.23)	0.23 (0.06)	0.27 (0.12)	0.10 (0.05)	-0.80 (0.08)	~5 min
Bayes	0.81 (0.20)	0.15 (0.05)	0.40 (0.14)	0.03 (0.03)	-0.72 (0.10)	~2.5 h

X30 computing time

- No significant differences

alejandra.alvarezm@uga.edu

Challenges

Further developments for the DHGLMF90

1. Extend to multiple-trait models (in progress)
2. Analyzing genomic datasets
3. How to do inference without a posterior distribution?

alejandra.alvarezm@uga.edu

Conclusion

DHGLMF90 gives a good approximation

1. Comparable performance with other methodologies
2. Faster convergence
3. Fast to estimate breeding values

alejandra.alvarezm@uga.edu

Acknowledgments

UNIVERSITY OF GEORGIA
College of Agricultural & Environmental Sciences
Animal Breeding and Genetics Group

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Animal Breeding and Genetics Group
College of Agricultural & Environmental Sciences
UNIVERSITY OF GEORGIA

Questions?

Alejandra Alvarez Munera
alejandra.alvarezm@uga.edu