Simultaneous Estimation of Spatial and Genetic Effects Using Hierarchical Generalized Linear Models

L. Rönnegård,[1,2] M. Felleki,[1,2] M. Alam,[2] and X. Shen[1]
[1]SLU, Uppsala, Sweden, [2]Dalarna University, Sweden

ABSTRACT: Spatial modelling is recognized as an important factor for assessment of breeding values in plant trials. 841 Scots pines were analyzed using version 2.0 of the **hglm** package. Both random additive genetic effects and spatial effects were included in the model. The covariance structures for the genetic and spatial effects were given by the additive relationship matrix and a conditional autoregressive (CAR) model, respectively. The genetic variance decreased by 24% when the spatial effects were included, which confirms the importance of including spatial effects in plant trials. The possibility to fit CAR models in the **hglm** package is expected to facilitate spatial modelling in genetic studies including animal breeding applications.

Key words: breeding values; random spatial effects; spatial correlation.

INTRODUCTION

Spatial modelling has been recognized as an important factor to consider in genetic plant trials (Dutkowski et al. 2002, Zas 2008) but the available software to fit both spatial and genetic effects is limited. ASReml (Gilmour et al. 2006) allows modelling when the data has been collected in a regular spatial matrix structure. There are also Bayesian methods such as WinBUGS, which uses Gibbs sampler, and INLA (Rue et al. 2009) which is a deterministic algorithm combining Laplace approximation and numerical integration for a limited set of models. INLA has recently also been developed to include modelling of additive genetic effects (Steinsland & Jensen 2010, Holand et al. 2013). A reasonably fast and non-Bayesian software on CRAN for fitting both genetic and spatial effects, where the spatial points do not have to be regular, seems to be missing though.

We have recently developed the **hglm** package (Rönnegård et al. 2010) in R available on CRAN for spatial modelling and where the random effects can follow several different distributions. The aim of this paper is to investigate, using the **hglm** package, the importance of including spatial effects in heritability estimation for a Swedish tree breeding trial data set.

MATERIALS AND METHODS

Data. 841 individuals in a Scots pine progeny study conducted in northern Sweden were analyzed and collected by The Forestry Research Institute of Sweden, SkogForsk. Tree height was measured on the 562 trees surviving until 26 years of age (Figure 1). The plants were placed in 8.8 × 22 m blocks each having 40 seedlings on a 2.2 × 2.2 m grid. Our study includes the blocks in the north-western part of the study area. Parental plants were crossed following a partial diallel design of 52 parent trees, which were assumed unrelated and the offspring were placed on the grids unrestricted randomly. See Waldmann and Ericsson (2006) for a detailed description of the data set. A spatial-genetic model was fitted by Finley et al. (2009) using an MCMC method.

Model. The fitted linear mixed model with tree height, y, as outcome was

$$y = \mu + Za + Ws + e \qquad (1)$$

where μ is an intercept term, Z and W are incidence matrices connecting the individual random effects with the observed phenotypes, a is a normal distributed additive genetic effect with (co)variance matrix $A\sigma_a^2$ with A being the additive relationship matrix, s is a random spatial effect from a normal distribution described below, and e is the residual effect $e \sim N(0, \sigma_e^2)$.

The random spatial effects follow a conditional autoregressive (CAR) model. The (co)variance matrix for the spatial effect s is Σ and its inverse is given by

$$\Sigma^{-1} = \frac{1}{\tau}(I - \rho D) \qquad (2)$$

where τ is a spatial variance parameter, ρ is a spatial correlation parameter, I is the identity matrix and D is a neighbourhood matrix (same size as the matrix A) having elements 0's and 1s indicating which seedlings are standing next to each other. From version 2.0 of the *hglm* package this distribution has been implemented and is specified as `rand.family=CAR(D=nbr)` where `nbr` is the neighbourhood matrix given by the user. The model is fitted in a computationally fast way by using an eigen decomposition of D and adding an additional hierarchy in the hierarchical generalized linear model approach (Lee & Nelder 1996, Rönnegård et al. 2010, Lee & Rönnegård 2013).

RESULTS AND DISCUSSION

The estimated variance component without spatial effects included in the model were $\hat{\sigma}_a^2 = 53.3$ and $\hat{\sigma}_e^2 = 113.6$ giving a heritability of $h^2 = 0.32$. When model (1) was used, ie with spatial effects included, the estimated variance components decreased to $\hat{\sigma}_a^2 = 40.43$ and $\hat{\sigma}_e^2 = 75.4$. The estimated spatial effect variance and correlation parameters were $\hat{\tau} = 21.3$ and $\hat{\rho} = 0.126$, respectively. Hence, the spatial effects explained a large portion of the total variance and influenced the genetic variance dramatically. Both spatial and genetic random effects were computed for all trees including those without observed height (Figure 2).

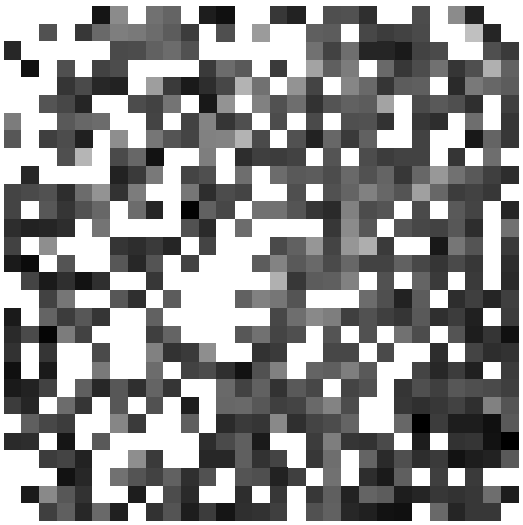

Figure 1: Location of each tree and their height given in grey scale. Darkness increases with height and white indicates missing phenotype

These results confirm the importance of including spatial effects in plant trials. The implementation of spatial modelling in the **hglm** package simplifies the practical application of spatial modelling. Note that the matrix D in (2) could be defined using distances between points rather than neighbouring areas, which facilitates modelling of irregular spatial points as for instance farms in animal breeding.

Herd effects are commonly modelled as fixed effects in animal models. For small and unbalanced herd sizes the use of random herd effects are recommended (Ugarte et al. 1992, Oikawa & Sato 1997). However, so far spatial information has not been used to fit random herd effects as far as we know. It is reasonable to expect that spatial modelling would improve model fit for animal models applied on agricultural systems where the number of animals per farm is small (eg in developing countries).

There are alternatives to the **hglm** package that can fit both spatial and genetic effects. ASReml has been used in several plant studies (Dutkowski et al. 2002, Piepho et al. 2008, Zas 2008) but is limited to regularly structured spatial points. WinBUGS and other MCMC methods (Finley et al. 2009) can fit more general spatial models but is notoriously slow and convergence is difficult to assess. A more exciting alternative is INLA (Rue et al. 2009) which is less flexible than WinBUGS but combines sparse matrix techniques (similar to those in ASReml) with Laplace approximation resulting in a fast and accurate method producing posterior likelihoods for the model parameters.

A problem not considered in the present study (nor in Finley et al. 2009) is whether the phenotypes are missing at random. This is a topic for the future and has, to our knowledge, rarely been dealt with in animal breeding applications. Furthermore, spatial modeling could potentially also include GxE interactions, which would be an additional development to the analysis.

Apart from spatial modelling, several other new options are also available from version 2.0 of the **hglm** package. Model selection tools using likelihood ratio tests and AIC are available. The user can specify several random effects from different distributions, and it is also possible to add linear predictors for the random effect variance components.

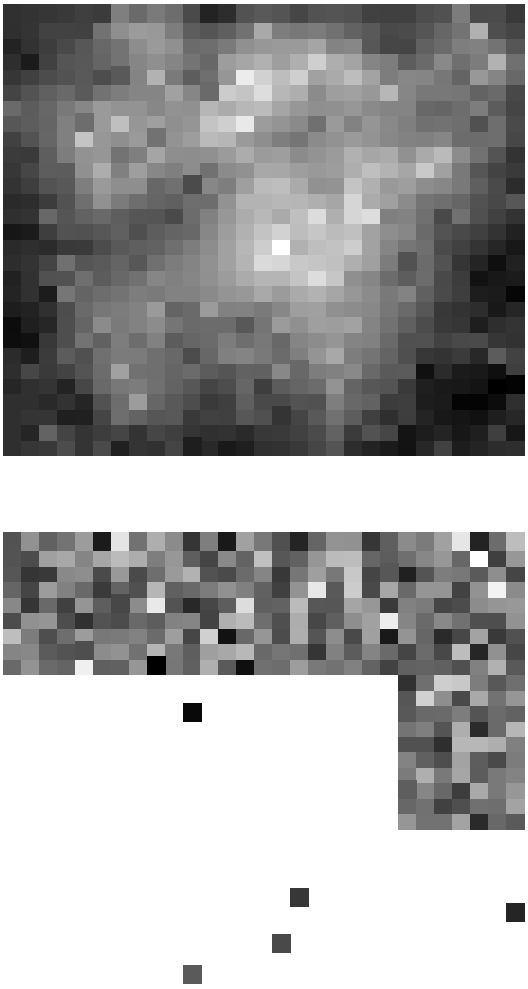

Figure 2: Estimated spatial and genetic random effects for each tree. Darkness increases with higher values.

CONCLUSION

Using version 2.0 of the **hglm** package in R, the importance of including spatial effects in tree trials for selective breeding has been confirmed. This version of the **hglm** package is a substantial development especially for spatial modelling.

LITERATURE CITED

Dutkowski GW, Costa e Silva J, Gilmour AR & Lopez GA (2002) Can J For Res 32: 2201–2214.

Finley AO, Banerjee S, Waldmann P and Ericsson T (2009). Biometrics 65, 441–451.

Gilmour AR, Gogel BJ, Cullis BR, and Thompson R: Asreml user guide release 2.0 VSN International, Hemel Hempstead 2006.

Holand AM, Steinsland I, Martino S and Jensen H (2013) G3 3:1241-51.

Lee Y. and Nelder JA (1996) J R Statist Soc B 58: 619-678.
Lee Y. and Rönnegård L (2013) J Anim Breed Genet 130:415-416.
Oikawa T and Sato K (1997) J Anim Breed Genet 114: 177-183.
Piepho HP, Möhring J, Melchinger AE and Büchse A (2008) Euphytica 161:209–228.
Rönnegård, L, Shen, X and Alam, M. (2010) The R Journal 2:20-28.
Rue H, Martino S and Chopin N (2009) J Royal Stat Soc B 71: 319–392.
Steinsland I and Jensen H (2010) Biometrics 66: 763-771.
Ugarte E, Alenda R and Carabaño MJ (1992) J Dairy Sci 75:269-278.
Waldmann, P. and Ericsson, T. (2006). Theor. Appl. Genet. 112:1441–1451.
Zas R (2008) Can J For Res 38: 114-124.