THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Spatial modelling in genetic evaluation of South African Holstein cattle population

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

OPEN ACCESS

Spatial modelling in genetic evaluation of South African Holstein cattle population

I. Houaga[1,2]*, T.P. Oliveira[2], I. Pocrnic[2], E. Lavrenčič[3], C.B. Banga[4], G. Gorjanc[2]

[1] Centre for Tropical Livestock Genetics and Health, Roslin Institute, University of Edinburgh, EH25 9RG, UK;[2] Roslin Institute, University of Edinburgh, EH25 9RG, UK; [3] Biotechnical Faculty, University of Ljubljana, Kongresni trg 12, 1000 Ljubljana, Slovenia; [4] Agricultural Research Council (ARC), Old Olifantsfontein Road, Pretoria, 0002, South Africa, *ihouaga@ed.ac.uk

Abstract
African dairy production systems are characterized by small herd size and low genetic connectedness between herds. This situation makes it difficult to accurately estimate environmental and genetic effects. We evaluated how accounting for spatial relationship between neighbouring herds impacts genetic evaluation of 305-days milk yield in South African Holstein cattle population. We analysed 305-days milk yield and pedigree data for 98,632 Holstein cows collected between 1957 and 2014 from 1,145 herds. Herd and herd-year effects were modelled as independent, whereas spatial effects were modelled as correlated between regions using the Besag model. The results show a regional variation of estimated breeding values and separation of spatial effect from herd effect. The spatial modelling of herd as random effect showed similar accuracy compared to modelling herd-year. Further studies integrating the genotype-by-environment interactions will be needed to better explore the benefits from spatial modelling in African smallholder dairy farming systems.

Introduction
The significant increase of milk yield in the dairy cattle production sector in developed countries has been associated with selective breeding and herd management strategies (Cole et al., 2020). However, in many low to middle-income countries genetic improvement have led to low or no genetic progress, especially in smallholder dairy production systems, due to a lack of appropriate performance records (Ojango et al., 2019). One of the reasons for this is that African dairy production systems are characterized by small herd size and low genetic connectedness between herds. Such a situation is challenging accurate estimation of environmental and genetic effects (Selle et al., 2020). Genetic evaluation partitions genetic and environmental effects on phenotype, which is critical because similarity between relatives may also be caused by shared environmental conditions. Breeders use statistical models to separate these effects by modelling the genetic effect as random with pedigree or genomic relationships between animals and modelling a herd or herd-season effects as fixed or random contemporary group (Mrode, 2014). Traditionally, the herd effect is modelled by assuming the effect of different farms as independent (Cuyabano et al., 2021). However, nearby farms are likely to have similar climate, soils, management, and even social aspects of production and addressing these aspects with spatial models can improve genetic evaluation (Sæbø & Frigessi, 2004; Cuyabano et al., 2021; Selle et al., 2020).
The aim of this study was to evaluate the impact of modelling regions as correlated effects on genetic evaluation of 305 days milk yield in Holstein dairy cattle population in South Africa.

Materials and methods
The 1,331,015 305-days milk yield (MY305) records from 377,921 South African Holstein cows in 7,121 herds was provided by South African Agricultural Research Council. After the quality control we retained 248,597 MY305 records and a pedigree for 196,778 animals, including 98,632 cows from 1,145 herds. The average number of animals per herd across the

whole period was approximately 91 cows per herd (8-2,230). We used postcodes of the herd's location to assign them to administrative regions. The herds were spread over 46 of the 52 South African regions.

Statistical modelling

The baseline model (G_1) for MY305 was accounting for the overall mean, a polynomial function of age within parity, random effect of herd (h), random additive genetic effect of the animal (a), random permanent environmental (p) and random residual (e): $y = Xb + Z_1h + Z_2a + Z_3p + e$; $h \sim N(0, I\sigma_h^2), a \sim N(0, A\sigma_a^2), p \sim N(0, I\sigma_p^2)$, and $e \sim N(0, I\sigma_e^2)$, where X is the design matrix associated with the vector of fixed effects b, Z_1, Z_2, and Z_3 are design matrices respectively associated with herd (h), additive genetic effect (a), and permanent environmental effects (p), and A is the pedigree relationship matrix. We also tested a second model (G_2), where the herd effect was replaced by the herd-year effect (hy), $hy \sim N(0, I\sigma_{hy}^2)$. We extended models G_1 and G_2 using an intrinsic conditional auto-regressive (ICAR) model component (Besag, 1974) to account for spatial variation associated with herd's region: i) $GR_i = G_i + R$, where R is the fixed effect of herd's region and i is the index of the models (i=1,2); ii) $GRU_i = G_i + RU$, where RU is an independent random effect of herd's region, and iii) $GRC_i = G_i + RC$, where RC is a correlated random effect of herd's region via neighbourhood-based covariance matrix (Besag, 1974). We defined scaled (GRC_i) and unscaled ($GRCNS_i$) inverse of the neighbourhood-based covariance matrix using R-INLA package (Rue et al., 2009). The models were fitted using the BLUPf90 suite of programs (Misztal et al., 2018). The AIC criteria was used to select the most parsimonious model. Moreover, accuracies of estimated breeding values (EBV) were calculated as sqrt(1-PEV/σ_g^2) where PEV is the prediction error variance of EBVs and σ_g^2 the additive genetic variance.

Results

Variance components

Variance components and heritability estimates are presented in Table 1. Heritabilities in G_1 models were higher than those from G_2 models (Table 1). The additive genetic, permanent environmental, and residual variances were similar between G_1 models. The same trend was observed between G_2 models. However, the additive genetic and residual variance were higher in G_1 models compared to G_2 models due to large difference between herd and herd-year variance estimates. On the other hand, the permanent environment effect was higher in G_2 models than in G_1 models. The herd effect variance was smaller in GR_1 than in model G_1 since the herd effect in model G_1 captured the regional effects. GRU models had a similar estimate of regional variance as the GRC models and scaling was important to get interpretable estimates (GRC_i vs. $GRCNS_i$).

Table 1: Variance components and heritability[1].

Model	σ_g^2	σ_h^2	σ_{hy}^2	σ_{pe}^2	σ_r^2	σ_p^2	h^2
G_1	836 (23)	1,597 (69)	_	338 (17)	_	4,536	**0.18**
GR_1	836(23)	1,402 (62)	_	338 (17)	_	4,340	**0.19**
GRU_1	836 (23)	1,404 (62)	_	338 (17)	205 (64)	4,550	**0.18**
GRC_1	836 (23)	1,411 (63)	_	338 (17)	217 (80)	4,568	**0.18**
$GRCNS_1$	836 (23)	1,411 (63)	_	338 (17)	464 (172)	4,815	**0.17**
G_2	504 (17)	_	2,580 (28)	345(13)	_	4,741	**0.11**
GR_2	507 (17)	_	2,192 (25)	345 (13)	_	4,365	**0.12**
GRU_2	507 (17)	_	2,191 (25)	345(13)	646 (147)	5,009	**0.10**
GRC_2	508 (17)	_	2,190 (25)	345 (13)	1,010 (228)	5,372	**0.12**
$GRCNS_2$	508 (17)	_	2,191 (25)	345 (13)	2,158 (488)	6,520	**0.08**

Regional effect estimates

Figure 2 presents region estimates using GRC_1 and GRC_2 models. From the GRC_1 model, regions with low estimates (indicated in blue) were observed in the central-western and eastern parts of South Africa (Figure 2, left). Most of the regions with low estimates are neighbouring regions in the semi-arid part of the country with high variability in temperature and precipitations. Regions with medium estimates (indicated in purple) were primarily found in the country's northern, south-central, and south-eastern parts. The few regions with high estimates (indicated in red) were in the country's southwestern and eastern parts. GRC_2 model showed similar observations (Figure 2, right).
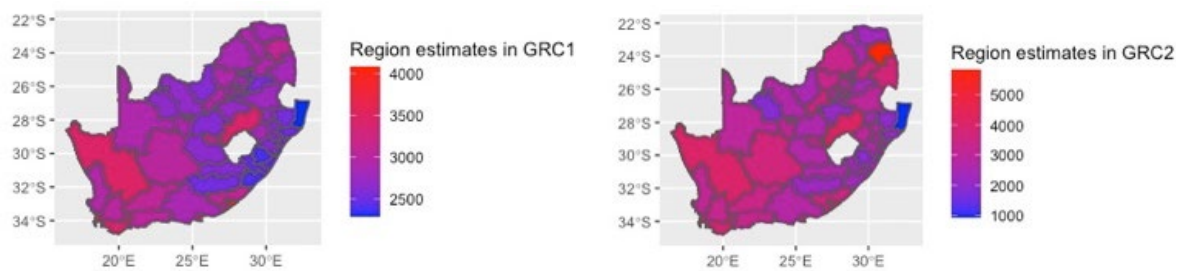

Figure 2. Regional South Africa effect estimates in GRC_1 model (left) and GRC_2 model (right)

Regional variation of estimated breeding values (EBV) and accuracy

Figure 3 presents the regional variation of EBV measured by the standard deviation (SD, left) and prediction error variance (PEV, right). The region in the northern part with a very low SD of EBVs had a very high PEV. Similarly, the region in the western part with a low SD of EBVs had a very high PEV. However, most of the regions in the country with medium to high SD of EBV had low to medium PEV. The same trend was seen for GRC_2 models (results not shown). The few regions with low SD of EBV and high PEV suggest that there was not much information on animals in those regions and the difference in distributions is more due to the amount of information. Considering the EBVs of all animals in the pedigree, similar accuracies were observed across the G_1 and G_2 models (results not shown).
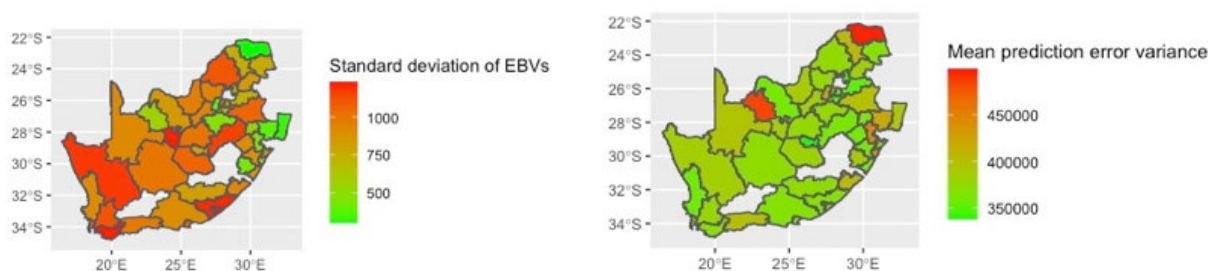

Figure 3. Standard deviation of EBVs (left) and mean of prediction error variance of EBVs (right) in GRC_1 model within south Africa

Discussion

The study evaluated the impacts of modelling spatial variation via neighbouring herd's region on genetic evaluation of MY305 compared to treating herds as entirely independent units. The results show that spatial modelling separated spatial effect from herd effect when treated as correlated and did not negatively impact prediction accuracy. Regional distribution of EBVs

was observed. A recent study showed that modelling proximity between herds better separated genetic effects from environmental effects and increased prediction accuracy in smallholder breeding programmes (Selle et al., 2020). The spatial modelling increases accuracy because it induces environmental connectedness and hence accurately separates a proportion of herd effect that other models may assign to breeding values. Our initial analysis showed that modelling herds using a spatial correlation structure between herd's regions using the Besag model (Besag, 1974) does not increase the prediction accuracy. A recent study on modelling proximity between herds showed increased reliabilities of genomic EBV, but no increase in prediction accuracy of phenotypes (Cuyabano et al., 2021). However, a deeper understanding of data structure, genetic connectedness between herds, the frequent movement of cows between herds, typical to the African livestock production system, is the next important step in this study.

Significant region effects were observed in the present study. These effects are likely due to differences in climatic conditions across regions. To account for large scale environmental effects, Sæbø & Frigessi (2004) modelled veterinary districts as correlated via neighbourhood in genetic evaluation of mastitis resistance in Norwegian Cattle and discovered regional effects in the south-eastern parts of Norway. Despite the benefits from regional modelling in previous studies, coordinate-based models would account more precisely for spatial relationships between farms (Selle et al., 2020; Cuyabano et al., 2021). Furthermore, the current study ignored the genotype-by-environment interaction effect, which are likely significant given the wide range of environmental differences in South Africa.

Conclusion

This study evaluated incorporating region-based spatial effects in genetic evaluation of 305-days milk yield in South African Holstein. Results showed large differences between regions, in some cases more than double, indicating the importance of modelling environmental variation. Further work is required to model environmental variation at a herd level by using a coordinate-based model and possibly expand that model with genotype-by-environment interaction.

References

Besag J. (1974) Spatial Interaction and the Statistical Analysis of Lattice Systems. *JSTOR*, *36*(2), 192–236. http://www.jstor.org/stable/2984812

Cole J. B., Eaglen S.A.E., Maltecca C., Mulder H.A., and Pryce J.E. (2020) Anim. Front. 10(2):37–44. https://doi.org/10.1093/af/vfaa007

Cuyabano B.C.D., Rovere G., Lim D., Kim T.H., Lee H.K., Lee S.H. and Gondro C. (2021) Animals 11(7):2050. https://doi.org/10.3390/ani11072050

Misztal I., Lourenco D., Aguilar I., Legarra A., and Vitezica Z. (2018) Manual for BLUPF90 family of programs. http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all7.pdf

Mrode R. (2014) Linear models for the prediction of animal breeding values (3rd ed.). CAB International

Ojango J.M.K., Mrode R., Rege J.E.O., Mujibi D., Strucken E.M., et al. (2019) JDS 102(6):5266–5278. https://doi.org/10.3168/jds.2018-15807

Rue H., Martino S., and Chopin N. (2009) J R Stat Soc Series B. 71(2):319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x

Sæbø S., and Frigessi A. (2004) GSE 36(5):527–542. https://doi.org/10.1186/1297-9686-36-5-527

Selle M. L., Steinsland I., Powell O., Hickey J. M., and Gorjanc G. (2020) GSE 52(1):69. https://doi.org/10.1186/s12711-020-00588-w