

Including environmental variables in genomic models for carcass traits in Hanwoo beef cattle

B. O. Makanjuola^{1*}, G. Rovere¹, B.C.D. Cuyabano², S.H. Lee³ and C. Gondro¹

¹ Department of Animal Science, Michigan State University, 474 S Shaw Ln, East Lansing, MI 48824, USA; ² French National Institute for Agriculture, Food and Environment (INRAE), Génétique Animale et Biologie Intégrative (GABI), 78350 Jouy-en-Josas, France; ³ Division of Animal and Dairy Science, Chungnam National University, Daejeon 305764, Korea; *makanjuo@msu.edu

Abstract

The objective of this study was to account for environmental variables to model non-genetic random effects for carcass traits in Hanwoo beef cattle. This model was compared to the traditional model that uses farms as an independent component to model non-genetic random effects. Heritability estimates for all carcass traits in our study ranged from 0.33 to 0.39 when using independent farm effects, whereas with the use of environmental variables, estimated heritabilities dropped and ranged from 0.11 to 0.31. In addition, prediction accuracy increased from a range of 0.39 to 0.45 to a range of 0.50 to 0.58 for all traits when modelling farm effects with environmental variables. This result suggests that environmental variables explain a considerable amount of the observed phenotypic variance, and that not accounting for such environmental variation could upwardly bias heritability estimates. Furthermore, an improvement in the model prediction accuracy was observed when accounting for environmental variables.

Introduction

The impact of climate change resulting from increasing fluctuations in environmental variables and rising temperatures is a major concern for livestock production. Adverse effects of rising temperatures are observed in the form of reduced productivity, performance, growth, and development of animals exposed to such challenging environments (Collier and Gebremedhin, 2015). Traditionally, environmental effects are accounted for by fitting the herd as an effect in genetic evaluation models. With this model, it is traditionally assumed that different herds are completely independent from each other with no similarities. However, herds in closer proximity could be more similar in terms of climatic and geographic factors when compared to more geographically distant herds. Therefore, accounting for relationships of environmental variables between production sites in genetic evaluation models could improve the estimates of genetic components in economically important traits. Improvement in the genetic evaluation of smallholder breeding programs using spatial relationships between herds has already been reported (Selle *et al.*, 2020). In addition, Cuyabano *et al.* (2021) recently showed that using GPS coordinates to create relationships between herds led to an increase in the reliability of the predicted genomic breeding values for carcass traits in beef cattle, when compared to fitting herd effects as independent. Nowadays, public weather stations regularly collect environmental variables and the data is easily accessible for general use. This source of information has been shown to be more accurate than on-farm weather information (Freitas *et al.*, 2006) and can be used to develop a relationship structure between herds based on their weather conditions. Furthermore, temperature-humidity index (THI) as environmental descriptor has been successful in modelling genotype-by-environment interactions with reaction norms (Chen *et al.*, 2021). This present study aimed to: (1) use environmental variables obtained from public weather stations to create a climate relationship

matrix between herds; (2) evaluate the impact on estimated variance components by modelling herd effects through the climate relationship structure and (3) assess prediction accuracy for the predicted genomic breeding values.

Materials & Methods

Animal and phenotypic data. Data from a total of 4,168 Korean Hanwoo beef cattle was used in this study. There were 3,810 steers and 358 bulls born in 2014 and 2015, all slaughtered in 2017. For this study, four carcass traits were analysed: carcass weight (CWT), backfat thickness (BFT), eye muscle area (EMA) and marbling score (MS). Measurements were taken 24h after refrigeration following slaughter. For BFT and EMA, records were measured at the 12-13th rib junction in millimeters and squared centimeters, respectively. MS was measured visually by a trained evaluator on an ordinal scale that ranged for 1 to 9. The 4,168 animals were raised on a total of 124 finishing farms, which were spread across South Korea.

Genotypes and weather data. All animals used in this study were genotyped with the 50K Illumina (San Diego, CA, USA) Bovine SNP50V2BeadChip array. Quality control measures were performed to retain only SNP present on the autosomal chromosomes with a call rate > 90% and minor allele frequency > 1%. After quality control, a total of 43,749 SNP genotypes were retained for further analyses. Weather information was available from 78 public weather stations that were close to the farms. To each farm, weather information from the nearest weather station was assigned. The distances between the farms and the weather stations ranged from 0 to 30 km with approximately 45% of the farms sharing the same location as the weather station. The environmental variables used for this study were monthly averages of the maximum temperature (T_{max}) and of the relative humidity (RH_{min}), spanning from 2013 to 2018. Temperature and relative humidity were then combined into THI using the following formula (NRC, 1971):

$$THI = (1.8T_{max} + 32) - [(0.55 - 0.0055RH_{min}) \times (1.8T_{max} - 26)] \quad (1)$$

THI was considered a suitable environmental variable to use because a previous study already reported that THI influences the modulation of the genetic components of these traits in Hanwoo (Chung *et al.*, 2020).

Genomic evaluation. Genetic parameters were estimated for all traits using both of the following linear mixed models:

$$y = Xb + Zg + Wfarm + e, \quad (2)$$

$$y = Xb + Zg + Wst + e, \quad (3)$$

where y is a vector of phenotypic records for CWT, BFT, EMA and MS; b is a vector of fixed effects that included the effects of sex, age at slaughter, herd size and slaughter date; g is a vector of random additive genetic effects; $farm$ is a vector of non-genetic and independent random effects of farm; st is a vector of non-genetic and correlated random effects of farm; e is a vector of random residuals. X , Z and W are incidence matrices that link respectively the fixed effects, random additive genetic effects and non-genetic random farm effects to the phenotypes. The assumptions posed to the random effects were: $g \sim N(0, G\sigma_g^2)$, $farm \sim N(0, I\sigma_{farm}^2)$ and $st \sim N(0, E\sigma_{st}^2)$. σ_g^2 is the additive genetic variance; σ_{farm}^2 and σ_{st}^2 are the non-genetic variance of the herd effect assuming the farms as independent and correlated, respectively. G is the genomic relationship matrix (VanRaden, 2008), I is an identity matrix, and E is the covariance matrix of the farms using the closest weather station information. E was created using the standardized differences between THI values of the farms. First, a matrix D was calculated using the differences between THI values,

and the matrix was then scaled by dividing the values by the maximum absolute THI difference, to ensure that the elements of the matrix ranged within the $[-1,1]$ interval (i.e. $\mathbf{D}^* = \mathbf{D}/\max\{|D_{ij}|: i, j = 1, \dots, st\}$). Finally, the covariance matrix \mathbf{E} was rescaled to have a diagonal of ones and defined as $E_{ij} = 1 - D_{ij}$ with values ranging from $-1 \leq E_{ij} \leq 1$. With this parameterization, stations that are more similar to each other have values closer to one and the more dissimilar stations will have values closer to minus one. All variance components estimation and genomic predictions were performed using WOMBAT software (Meyer, 2007).

Model predictive accuracy. The predictive accuracy of the different models in predicting the genomic breeding values was performed using a 5-fold cross validation. To evaluate accuracy, the whole data was randomly divided into five subsets of approximately equal size. For each fold of the cross validation, four out of the five subsets (i.e. 80% of data) were assigned as the training set and the remaining subset (i.e. 20% of the data) was used as the testing set. Replication of the 5-fold cross validation procedure was performed 20 times and resulted in a total of 100 estimates. Prediction accuracy in the testing set was calculated as the Pearson correlation between predicted genomic breeding values of the test group and observed phenotypes adjusted for fixed effects ($r = \text{cor}(\hat{g}_{test}, y_{test} - X_{test}\hat{b})$).

Results

Table 1 presents a summary of the phenotypic records, with descriptive statistics for each trait.

Table 1. Descriptive statistics of analysed traits including the number of records (N), mean, standard deviation, minimum and maximum of the observations.

Trait	N	Mean	Standard deviation	Minimum	Maximum
BFT	4168	14.33	5.09	2	47
EMA	4168	96.74	12.05	34	155
CWT	4168	442.02	54.95	159	682
MS	4168	6.44	1.76	1	9

The proportions of variance explained by each random effect using the two different models for each trait are presented in Table 2, along with the prediction accuracy of the predicted genomic breeding values. Heritability estimates for BFT, EMA, CWT and MS were 0.34, 0.35, 0.39 and 0.38, respectively, when fitting the farm as independent effects. These estimates are in line with other studies that used similar models with heritability estimates that ranged from 0.33 to 0.42 (Chung *et al.*, 2020; Cuyabano *et al.*, 2021). Conversely, using the THI to model the farm effects as correlated resulted in a drop in heritability estimates to 0.31, 0.11, 0.29 and 0.18 for BFT, EMA, CWT and MS, respectively. To the best of our knowledge, no study has reported using a covariance matrix created from THI values from public weather stations. Nevertheless, our results are in accordance with those from a study that used GPS coordinates to create a covariance matrix between farms (Cuyabano *et al.*, 2021). These results suggest that environmental variables and geographic location explain a substantial proportion of the phenotypic variation observed. Prediction accuracies obtained were 0.39, 0.40, 0.45 and 0.40 for BFT, EMA, CWT and MS, respectively, when farm was fitted as an independent effect, and increased to 0.50, 0.54, 0.58 and 0.56 for BFT, EMA, CWT and MS, respectively, when fitting farms with a covariance matrix based on environmental variables, described in this study by the THI.

Table 2. Estimates of variance proportions explained by additive genetics (h^2), non-genetic effects of farm or weather station ($(\text{farm/st})^2$) and residual (e^2) and predictive accuracy (r) with their respective standard errors (se).

Trait	Model	$h^2 \pm \text{se}$	$(\text{farm/st})^2 \pm \text{se}$	$e^2 \pm \text{se}$	r
BFT (mm)	g + farm	0.34 ± 0.03	0.03 ± 0.01	0.63 ± 0.03	0.39
	g + st	0.31 ± 0.04	0.13 ± 0.08	0.57 ± 0.06	0.50
EMA (cm ²)	g + farm	0.35 ± 0.03	0.07 ± 0.02	0.58 ± 0.03	0.40
	g + st	0.11 ± 0.03	0.69 ± 0.08	0.19 ± 0.05	0.54
CWT (kg)	g + farm	0.39 ± 0.03	0.04 ± 0.01	0.58 ± 0.03	0.45
	g + st	0.29 ± 0.05	0.29 ± 0.12	0.43 ± 0.07	0.58
MS	g + farm	0.38 ± 0.03	0.05 ± 0.01	0.57 ± 0.03	0.40
	g + st	0.18 ± 0.02	0.54 ± 0.01	0.27 ± 0.01	0.56

Discussion

Our study compared modelling herd effects considering farms as correlated or independent. A similar study has already been performed for Hanwoo beef cattle, using GPS coordinates to build correlation between farms (Cuyabano *et al.*, 2021). We innovate with our study by using environmental information from public weather stations, in the form of THI, to build this correlation between farms. Our results showed that a considerable proportion of the phenotypic variation can be explained by correlated environment, in agreement to what was reported by Cuyabano *et al.* (2021), and that this leads to a reduction in the additive genetic variance, and consequently to lower heritability estimates. This finding may suggest that not accounting for environmental variation could be upwardly biasing heritability estimates. Additionally, the highest prediction accuracy for all traits was obtained when using the model that accounted for the THI covariances between farms. Such improvement was not observed by Cuyabano *et al.* (2021), suggesting that actual weather information may be more accurate to estimate the random effects than GPS coordinates. Further research is warranted to demonstrate that environmental variables can be used to improve models in another independent population.

References

- Chen S.Y., Freitas P.H.F., Oliveira H.R., Lázaro S.F., Huang Y. J., *et al.* (2021) *Genetics Selection Evolution* 53(1):1-18. <https://doi.org/10.1186/s12711-021-00645-y>
- Chung Y., Lee S. H., Lee H. K., Lim D., van der Werf J., *et al.* (2020) *Frontiers in Genetics*, 11:1632. <https://doi.org/10.3389/fgene.2020.576377>
- Collier R. J., and Gebremedhin K. G. (2015) *Annu. Rev. Anim. Biosci.* 3:513-532. <https://doi.org/10.1146/annurev-animal-022114-110659>
- Cuyabano B. C. D., Rovere G., Lim D., Kim T. H., Lee H. K., *et al.* (2021) *Animals* 11(7):2050. <https://doi.org/10.3390/ani11072050>
- Freitas M. S., Misztal I., Bohmanova J., and West J. (2006) *Livestock Science* 105(1): 223-228. <https://doi.org/10.1016/j.livsci.2006.06.011>
- Meyer K. (2007) *J Zhejiang Univ. Sci. B.* 8:815-821. <http://doi.org/10.1631/jzus.2007.B0815>
- NRC. (1971) *A guide to environmental research on animals*. National Research Council, Washington, DC: National Academies, USA.
- Selle M. L., Steinsland I., Powell O., Hickey J. M., and Gorjanc G. (2020) *Genetics Selection Evolution* 52(1): 1-17. <https://doi.org/10.1186/s12711-020-00588-w>
- VanRaden P. M. (2008) *J Dairy Sci* 91(11): 4414-4423. <https://doi.org/10.3168/jds.2007-0980>