A Bayesian hierarchical model to integrate a mechanistic growth model in genomic prediction

H. Yu[1], J. van Milgen[2], E. Knol[3], R. Fernando[1], and J. Dekkers[1*]

[1]Iowa State University, Ames, Iowa 50011, USA; [2]PEGASE, INRAE, 35590 Saint-Gilles, France; [3]Topigs Norsvin, 6640 AA Beuningen, The Netherlands; [*]jdekkers@iastate.edu

Abstract

Genomic prediction can improve the accuracy of estimated breeding values for traits driven by additive genetic effects within common settings but prediction of traits affected by non-additive genetic effects and GxE remains a challenge. Mechanistic growth models express growth performances in terms of nonlinear functional interactions between underlying latent traits and nutritional environmental effects. Assuming the latent traits are less affected by non-additive genetic effects and GxE, these models can capture certain non-additive genetic effects and GxE at the phenotype level and allow prediction at unobserved ages for longitudinal data, e.g. mature weight and mature feed intake. In this study, we developed a Bayesian hierarchical model to integrate a Gompertz model for body weight and feed intake into genomic prediction models for pigs. By predicting breeding values for biologically relevant underlying latent traits, these models have the potential to advance genetic improvement across populations and environments.

Introduction

Genomic prediction (GP) has facilitated genetic improvement by enabling more accurate estimated breeding values at an early age (Meuwissen et al., 2001). While GP is effective in predicting traits dominated by additive genetic effects within common settings, challenges of predicting non-additive genetic effects and extrapolating to other conditions, such as to later ages or to other nutritional environments (GxE) remain. Previous attempts to address these challenges have mainly focused on statistical models, while augmentation of statistical models with biological information has received less attention. Developed from a nutrition perspective, mechanistic growth models (MGM) (e.g. van Milgen et al., 2008) describe growth performance as nonlinear functional interactions between underlying latent traits and nutritional environmental factors and can capture GxE and non-additive genetic effects that are created by the non-linear relationships embedded in the MGM. Following the pioneering ideas of Varona et al. (1997) and Bourdon (1998), several studies have sought to accommodate such MGM in genetic prediction (Doeschl-Wilson et al., 2007; Cai et al., 2012). Their successes were limited either by lack of appropriate statistical and computational methods or by the absence of genomic information. In plant breeding, crop growth models have recently been integrated into GP (Technow et al., 2015; Messina et al., 2018; Campbell et al., 2020), leading to increases in prediction accuracies for performance in other environments compared to standard GP methods. Against this background, the objective of this study was to develop a Bayesian hierarchical model to integrate an MGM into GP for pigs.

Materials and Methods

Bayesian Hierarchical Model for Body Weight and Feed Intake. The proposed model to integrate a MGM for body weight (BW) and daily feed intake (DFI) into GP is presented in Figure 1 and was implemented using the programming language Julia (Bezanson, Jeff, et al., 2017). The MGM of the InraPorc model (van Milgen et al., 2008), with a reparameterized Gompertz model for BW and a gamma function for DFI as a function of BW, was used to connect five latent traits to

observed BW and DFI data. The five latent traits (i.e. DFI50, DFI100, Age115, Shape, and BW65, as defined in Figure 1, were chosen based on biological relevance, and they were modeled as a linear combination of an intercept ($\boldsymbol{\mu}$), contemporary group effects ($\boldsymbol{CG}$), and additive SNP ($\boldsymbol{\alpha}$) and residual effects ($\boldsymbol{e}$). Observed traits of BW and DFI were then modeled as a function of the latent variables through the MGM. While latent traits were modeled as additive effects without GxE, the MGM allows for certain non-additive genetic effects and GxE for BW and DFI by internally creating non-linearity and interactions among genetic and environmental effects. We implemented MCMC to sample the unknown parameters of the MGM using a multi-trait Bayesian regression model. Specifically, the $\boldsymbol{\mu}$, $\boldsymbol{CG}$, and $\boldsymbol{\alpha}$ were sampled conditional on the latent traits using Gibbs sampling, while the five latent traits were jointly sampled using the Metropolis-Hastings algorithm, given the observed BW and DFI data. Variance components were obtained by Gibbs sampling.

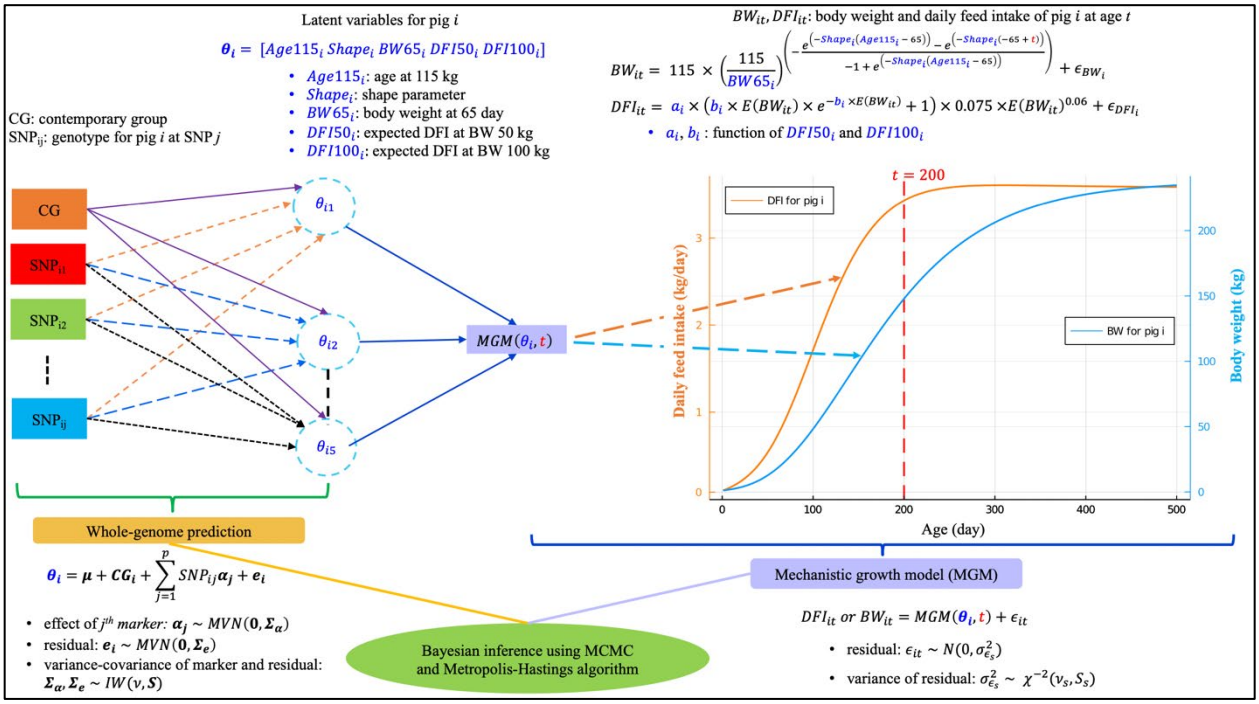

Figure 1. Structure of the Bayesian hierarchical model to integrate a mechanistic growth model (MGM) with five latent variables (blue font) for BW and DFI into genomic prediction.

Application. The model was applied to daily BW and DFI data from ~83 to ~186 days of age on 3,934 purebred boars in 97 contemporary groups, that were genotyped for 24K SNPs. Phenotypes of the 2,000 earlier born pigs were used as training and the remaining pigs as validation. Prediction bias and accuracy were estimated by forward cross-validation (CV) based on regression and correlation (divided by the square root of heritability) of observed BW and DFI adjusted for $\boldsymbol{CG}$ on predictions for the validation data. Because such CV does not allow evaluation of predictions for latent variables and for BW and DFI outside the observed age range, predictions were also evaluated using the linear regression (LR) method of Legarra and Reverter (2018), using predictions based on partial and whole data, which we applied to the non-linear models.

Results and Discussion

Figure 2 shows GP for BW and DFI from 0 to 500 days for the validation data. Note that predictions extend outside the age range with phenotypes, which is facilitated by the MGM. The GP for BW and DFI plateaued at ~400 and ~300 days, respectively. Previous studies used a 2-step method to incorporate growth models into genetic evaluation (Kachman et al., 1998; Vautier et al., 2013), with latent variables estimated on an individual basis in a first step, using only that individual's data, followed by genetic analysis of the estimated latent variables as phenotypes. Our model integrates these two steps and leverages relationships among individuals to estimate the latent traits for all individuals, regardless of the amount of data available for each.

Estimates of accuracy and bias for BW and DFI based on the LR and CV methods are also in Figure 2. Leveraging the LR method, we computed the accuracy and bias of GP of the latent traits (Table 1) and for BW and DFI outside the observed data (Figure 2). The LR method uses the relationship between predictions based on partial and whole data and can, therefore, estimate accuracy and bias of predictions of traits that do not have corresponding observed phenotypes (Legarra and Reverter, 2018). The latent traits have specific biological meaning related to growth processes based on the MGM of van Milgen et al. (2008). The GP for latent trait DFI100 had the highest accuracy (0.51), while the GP for Shape had the lowest accuracy (0.43). As Shape is related to the inflection point and mature weight, its low accuracy is likely because the BW and DFI data are from a very linear part of the growth curve. Prediction bias was found for all latent traits, ranging from 0.61 to 0.91, compared to the expectation of 1 for unbiased predictions.

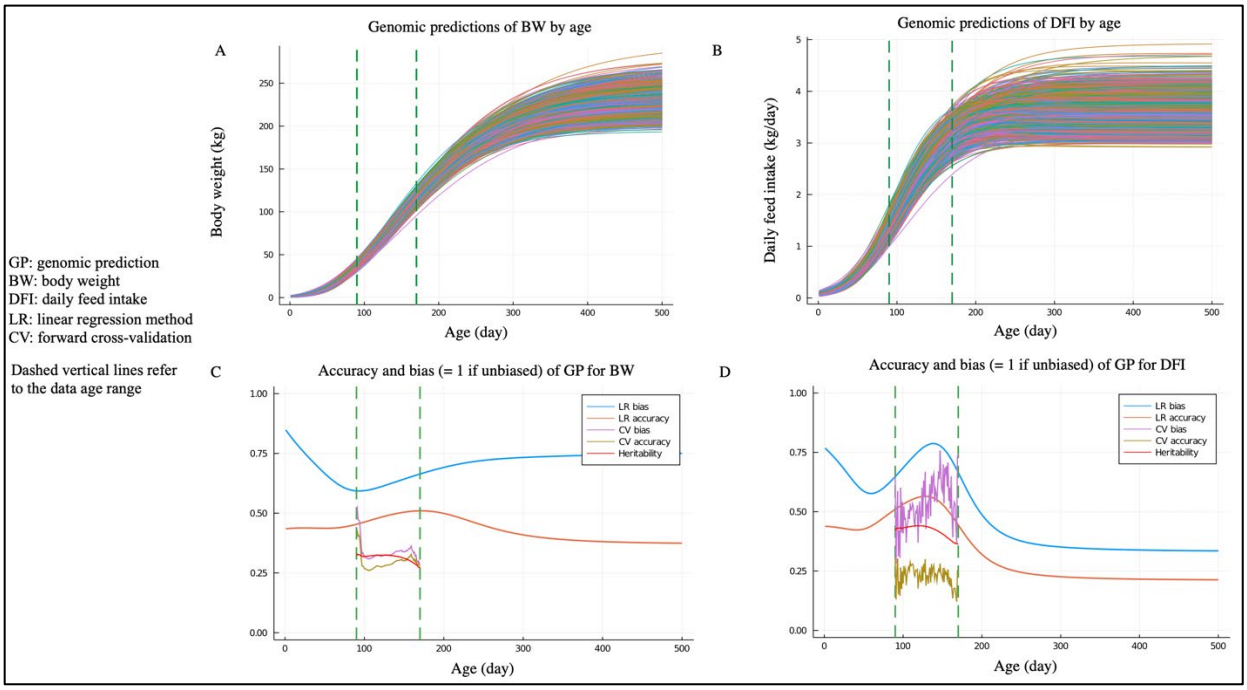

Figure 2. GP for BW (A) and DFI (B) for the validation data and estimates of bias and accuracy of GP of BW (C) and DFI (D) using the LR and CV methods.

Table 1. Bias and accuracy of genomic predictions of latent traits based on the LR method.

	Age115	Shape	BW65	DFI50	DFI100
Bias (unbiased if = 1)	0.66	0.73	0.61	0.83	0.91
Accuracy	0.51	0.43	0.45	0.51	0.55

Compared to the CV method, which only allows evaluation of predictions within the age range with phenotypes, the LR method resulted in much higher estimates of accuracy and lower estimates of bias (i.e. closer to 1) of predictions for BW and DFI for that age rage, although the trends with age were similar (Figure 2). Macedo et al. (2020) pointed out that the LR method can result in false estimates of bias when a seriously flawed model is fitted but that estimates of accuracy are robust. However, in a separate simulation study, we found that estimates of both bias and accuracy from the LR method can be biased when using a flawed model, in contrast to the CV method.
In conclusion, the proposed Bayesian hierarchical model allows GP for biologically relevant latent traits based on a MGM and can, thereby, accommodate certain non-additive genetic effects and GxE and allow prediction of traits outside the age range with data. Selection on GP for latent traits is expected to lead to greater rates of genetic improvement and the ability to better deal with GxE, if the latent traits are less affected by non-additive genetic effects and GxE. Whether they are, however, requires further investigation and potentially reparameterization of the latent variables. However, caution must be exercised when using the LR method to evaluate predictions as it may provide biased estimates of accuracy and bias when the model used is not valid for the data at hand.

Funded by USDA-NIFA grant # 2020-67015-31031.

References

Bezanson J., Edelman A., Karpinski S. and Shah V.B. (2017) SIAM review. 59(1):65-98. https://doi.org/10.1137/141000671
Bourdon R.M. (1998) J. Animal Sci 76(9):2308-2323. https://doi.org/10.2527/1998.7692308x
Cai W., Kaiser M.S., and Dekkers J.C.M. (2012) J. Animal Sci. 90(1):127-141. https://doi.org/10.2527/jas.2011-4293
Campbell M.T., Grondin A., Walia H., and Morota G. (2020) J. Experimental Botany 71(18):5669-5679. https://doi.org/10.1093/jxb/eraa280
Doeschl-Wilson A.B., Knap P.W., Kinghorn B.P., and Van der Steen H.A.M. (2007) Animal 1(4):489-499. https://doi.org/10.1017/S1751731107691848
Kachman S.D., Baker R.L., and Gianola D. (1988) Theor. Appl. Genetics 76(1):148-156. https://doi.org/10.1007/BF00288846
Legarra A., and Reverter A. (2018) Genet. Sel. Evol. 50(1):1-18. https://doi.org/10.1186/s12711-018-0426-6
Macedo F.L., Reverter A., and Legarra A. (2020). J. Dairy Sci., 103(1):529-544. https://doi.org/10.3168/jds.2019-16603
Messina C.D., Technow F., Tang T., Totir R., Gho C., et al. (2018) European J. Agronomy, 100:151-162. https://doi.org/10.1016/j.eja.2018.01.007
Meuwissen T.H.E., Hayes B.J., and Goddard M.E. (2001) Genetics 157(4):1819-1829. https://doi.org/10.1093/genetics/157.4.1819
Technow F., Messina C.D., Totir L.R., and Cooper M. (2015) PloS one 10(6):e0130855. https://doi.org/10.1371/journal.pone.0130855
Varona L., Moreno C., García Cortés L.A., and Altarriba J. (1997) Livest. Prod. Sci. 47(3):201-209. https://doi.org/10.1016/S0301-6226(96)01415-7
van Milgen J., Valancogne A., Dubois S., Dourmad J.Y., Sève B., et al. (2008) Animal Feed Sci. and Technology, 143(1-4):387-405. https://doi.org/10.1016/j.anifeedsci.2007.05.020
Vautier B., Quiniou N., van Milgen J., and Brossard L. (2013) Animal 7(8):1265-1273. https://doi.org/10.1017/S1751731113000554