Genomic predictions of growth curves in Holstein dairy cattle based on parameter estimates from nonlinear models combined with different kernel functions

T. Yin and S. König*

Institute of Animal Breeding and Genetics, Justus-Liebig-University Gießen, 35390 Gießen, Germany

ABSTRACT

Availability of longitudinal body weight (BW) records allows the application of nonlinear models (NLINM) to predict phenotypic and genomic growth curves in dairy cattle. In this regard, we considered a data set including 31,722 BW records from 4,952 female Holstein cattle, during the period from birth (mo 0) to approximately age at first calving (mo 24). Parameters of the growth curves were estimated using 3 NLINM: the logistic (LOG), the Gompertz (GOM), and the Richards (RICH) functions. Residuals for the growth curve parameters from the NLINM applications were used as pseudo-phenotypes in the ongoing genomic analyses with different similarity matrices, including 2 genomic relationship matrices (**G1** and **G2**), a combined pedigree and genomic relationship matrix (**H**), and 3 kernel matrices. The kernels were a weighted “alike by state” kernel function (**K1**), an exponential dissimilarity kernel (**K2**), and a Gaussian kernel (**K3**). On the basis of **G1** and **G2** matrices, genomic heritabilities for the growth curve parameters birth weight (W_0), mature weight (W_m), and growth rate (k), and the shape parameter (m ; only available from RICH) were moderate to large, in the range from 0.29 (m from RICH) to 0.46 (k from RICH). Fitting the similarity matrices based on kernel functions contributed to an increase of the ratio of the variance explained by the similarity matrix in relation to the total variance (compared with the heritability when modeling **G1** or **G2**). Genetic correlations between W_0 , W_m , and k were always positive (>0.30), especially for the same growth curve parameters estimated from different NLINM (>0.90). The shape parameter m from RICH was negatively correlated with other growth curve parameters, from -0.29 to -0.95 . In a next step, estimated genomic breeding

values for growth curve parameters were input data for the respective NLINM, aiming to construct genomic growth curves. Prediction accuracies were correlations between genomic growth curves and genomic breeding values from random regression models for sires and female cattle. Considering all genotyped female cattle with pseudo-phenotypes, prediction accuracies were larger from RICH than from LOG and GOM. However, differences in prediction accuracies from the NLINM $\times$ similarity matrix combinations were quite small. Accordingly, in 5-fold cross-validations using heifer groups with masked phenotypes, very similar prediction accuracies across modeling approaches were identified. Especially for specific age months, genomic growth curve predictions were more accurate for sires than for female cattle, indicating that the relationships between animals in training and validation sets are more important than the selection of specific NLINM $\times$ similarity matrix combinations.

Key words: longitudinal body weight, nonlinear models, growth curve parameters, genomic predictions

INTRODUCTION

Body weight is a trait of increasing importance in dairy cattle breeding programs. Classically, BW was used to determine feed requirements for maintenance and production demands (Visscher et al., 1994). Recently, Pryce et al. (2015) suggested controlling of maintenance costs and improvements for feed efficiency through the optimal combination of BW and production traits in selection index applications. Similarly, Connor (2015) mentioned the possibilities of predicting maintenance costs when simultaneously considering residual feed intake and mature live weight. Consequently, due to the strong influence of BW on feed efficiency traits, dairy cattle breeders have requested inclusion of BW in overall breeding goals, especially in pasture-based production systems (Pryce and Harris, 2006). Furthermore, BW is genetically favorably correlated with production and fertility traits (Berry et al., 2003; Yin and König,

Received December 4, 2019.

Accepted April 6, 2020.

*Corresponding author: sven.koenig@agrar.uni-giessen.de

2018), indicating indirect selection response in a broad pattern of trait categories when incorporating BW in selection indexes.

Automatic weighing systems, as installed in modern dairy cattle farming systems, allow generation of a dense longitudinal BW data structure along the growth trajectory. For genetic analyses of repeated BW measurements, Yin and König (2018) applied linear repeatability and linear random regression models (**RRM**) with continuous time-dependent covariates. For the prediction of BW at specific time points and the modeling of individual growth curves, specific nonlinear functions have also been considered. For instance, Koskan and Ozkaya (2014) modeled growth curves for female Holstein cattle by applying the Brody (Fitzhugh, 1976), Gompertz (France et al., 1996), Logistic (France et al., 1996), Richards (France et al., 1996), and von Bertalanffy functions (von Bertalanffy, 1957). A comprehensive evaluation of the 5 nonlinear models (**NLINM**) was carried out via meta-analysis (Teleken et al., 2017), considering growth data from 14 different publications. Teleken et al. (2017) indicated that estimated parameters from NLINM reflect biological characteristics such as mature weight and maturing rate. However, irrespective of their flexibility, the complexity of NLINM hampered their applications for phenotypic and genetic BW predictions at specific points in time.

Availability of high-density SNP marker panels and sequencing data from genotyped cows allows consideration of alternative approaches regarding genetic-statistical modeling and genomic predictions via NLINM. An obvious problem for estimations of SNP effects with linear regression models, especially for novel functional traits, is the substantially smaller number of genotyped cows compared with the number of SNP. Alternatively, in nonparametric regressions (Gianola and van Kaam, 2008), all or a subset of the markers (**X**) can be used to build the variance-covariance structure between individuals through the definition of a nonparametric function $g(\mathbf{X})$. The function $g(\mathbf{X})$ can be specified by reproducing kernel Hilbert space regression (Gianola and van Kaam, 2008). According to the Representer theorem (Kimeldorf and Wahba, 1971), the solution of $g(\mathbf{X}) = \mathbf{K}\boldsymbol{\alpha}$, where **K** is a n -by- n kernel matrix (n = number of genotyped animals) and $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)$ is a vector of unknown parameters. Element (i, j) in **K** is $K(\mathbf{x}_i, \mathbf{x}_j)$, which is calculated by the kernel function, based on the genotypes of animal i and animal j . Hence, the input marker information is linked to the traits only through the kernel matrix **K** as produced from the kernel function. Schaid (2010a,b) reviewed various kernel functions, including the weight linear kernel (Wu et al., 2011), the polynomial kernel (Zien et al., 2000), and

the Gaussian kernel (Mallick et al., 2005). The general characteristics of all kernels are identified similarities or dissimilarities between individuals, based on, for instance, genotypes, amino acids, or gene expressions. Additionally, when a pedigree-based (**A**) matrix or a marker-based (**G**) relationship with almost 50,000 SNP is chosen as a kernel, a classical additive infinitesimal model in quantitative genetics is depicted (de los Campos et al., 2010a). However, in contrast with **A** and **G** matrices, the variance-covariance structure built on specific kernel functions, such as the polynomial kernel or the Hadamard product of 2 kernel matrices, captures additive genetic as well as nonlinear effects, such as SNP marker interactions (Wang et al., 2015).

To the best of our knowledge, no studies are available that combine parameters from NLINM with genomic information in specific kernel functions for the prediction of growth curves. Therefore, the objectives of the present study were as follows: (i) to use longitudinal BW data from Holstein heifers for the modeling of phenotypic growth curves, considering 3 nonlinear growth curve functions (logistic, Gompertz, and Richards models); (ii) to estimate genomic breeding values (**GBV**) of the growth curve parameters from the 3 nonlinear functions based on 2 **G** matrices and 3 similarity kernels (weighted “alike by state” kernel, exponential dissimilarity kernel and Gaussian kernel); and (iii) to assess the prediction accuracies of genomic growth curves.

MATERIALS AND METHODS

Phenotype Data

Data editing for BW followed the protocol by Yin and König (2018). In this regard, we excluded levels of herd-year-month effects with fewer than 10 BW records and BW outliers based on studentized residuals and corresponding Bonferroni P -values < 0.05 or > 0.95 . Body weights of 4,952 female Holstein cattle were recorded repeatedly during a period from birth (mo 0) to 24 mo (approximately age at first calving). The total number of BW records was 31,722, implying on average 6.41 records per animal (range: 4 to 19 repeated measurements). The cattle were kept in 10 large-scale herds, located in the German federal states of Mecklenburg-West Pomerania and Berlin-Brandenburg, and they were born between 2005 and 2014. The mean, minimum, maximum, and standard deviation for BW were 252.75 kg, 20 kg, 742 kg, and 168.63 kg, respectively. Pedigree after pruning included 20,176 animals, born between 1955 and 2014. On average, every dam had 1.22 phenotyped daughters within a range of 1 to 4. Number of daughter BW records

Figure 1. Number of observations per growth month for all cattle with phenotype BW records (Pheno, 4,952 cattle) and for genotyped cattle with BW records (Geno, 620 cattle).

per dam ranged from 4 to 45, and the average was 7.82. The 4,952 Holstein heifers were daughters from 780 sires, 1 to 173 phenotyped daughter per sire. The number of daughter records per sire ranged from 4 to 1,073 (average: 40.67 daughter records per sire). The distribution of BW records by age is given in Figure 1 for the heifers with phenotypes and for the heifers with phenotypes and genotypes.

Genotype Data

Among the female cattle with phenotypic data, 620 were genotyped using the Illumina Bovine 50K SNP BeadChip V2 (161 animals), or the Illumina Bovine Eurogenomics 10K low-density chip (459 animals; Illumina, San Diego, CA). Additionally, 1,101 sires of 4,952 calves and heifers were also genotyped with the Illumina Bovine 50K SNP BeadChip V2 (180 sires) or the Illumina Bovine Eurogenomics 10K low-density chip (921 sires). The 10K SNP genotypes were imputed to the 50K SNP panel by project partner vit (Verden, Germany), as done for national routine genetic evaluations (Segelke et al., 2012). Genotype quality controls were performed using PLINK software, version 1.9 (Chang et al., 2015; Purcell and Chang, 2019). Filtering criteria for markers were as follows: consideration only of SNP located on *Bos taurus* autosomes; minor allele frequency larger than 0.01; minimum call rate of 0.95; and no significant deviation from Hardy-Weinberg equilibrium ($P > 0.0001$). Finally, 1,721 genotyped

animals with 44,314 SNP were included in genomic analyses.

Nonlinear Growth Functions

Three frequently used nonlinear growth functions (Table 1), including two 3-parameter functions and one 4-parameter function, were employed to model the growth curves of the phenotyped female cattle. The 3-parameter functions were the logistic (**LOG**) and the Gompertz (**GOM**) models, both with the characteristics of a sigmoidal shape and a fixed inflection point. The fixed inflection points of the 2 models are located at about 50% (LOG) and 37% (GOM) of the mature weight (Teleken et al., 2017). The curve parameters with biological meanings of LOG and GOM models are the birth weight (W_0), the mature body weight (W_m), and a constant growth rate (k) reflecting an average velocity from birth to mature weight. The 4-parameter Richards (**RICH**) function embraces all the parameters from LOG and GOM. The auxiliary parameter is a shape parameter (m), indicating associations between proportions of mature weight with the inflection point (Gbangboche et al., 2008).

The “saemix” function from the R package “saemix” (Comets et al., 2017), with an implemented stochastic approximation expectation maximization algorithm, was used to estimate parameters of the 3 NLINM. The equation for the models with homogeneous residuals was as follows:

$$y_{ij} = f(t_{ij}, \Psi_i) + e_{ij}, \quad [1a]$$

and when assuming an exponential error, the model was as follows:

$$\log(y_{ij}) = \log[f(t_{ij}, \Psi_i)] + e_{ij}, \quad [1b]$$

where y_{ij} = the j th BW observation for animal i ; $f(t_{ij}, \Psi_i)$ = one of the nonlinear functions; t_{ij} = age of animal i when the j th observation was recorded; Ψ_i = individual parameters for the nonlinear functions with

$\Psi_i = \Psi_{\text{pop}} + \sum_{l=1}^9 h_{il}\beta_l + \eta_i$, where Ψ_{pop} = one of the popu-

lation growth curve parameters (overall mean), h_{il} = dummy variables for the herd effect l (only the herd effect was included, due to convergence problems), β_l = estimates for a single herd l compared with the reference herd (herd no. 10), η_i = random effects for each animal, $N(\mathbf{0}, \mathbf{\Omega})$; e_{ij} = random residual effects with $e_{ij} \sim N(0, \sigma_e^2)$. The subscripts varied from 1 to 4,952 (number of animals) for i and from 4 or 19 (no. of repeated measurements per animal) for j . Ψ_{pop} was a vector with d elements, and $\mathbf{\Omega}$ was a $d \times d$ variance-covariance matrix. For the 3-parameter functions (LOG and GOM), d was 3, and for RICH, d was 4. Model evaluation criteria were the -2 times maximum log-likelihood ($-2LL$), the Akaike information criterion and the Bayesian information criterion, which are all output from the applied “saemix” function. Low values for $-2LL$, Akaike information criterion, and Bayesian information criterion indicate model superiority.

In the first step, population parameters (Ψ_{pop}), variance-covariance matrices ($\mathbf{\Omega}$), residual variances (σ_e^2), and coefficients for the herd effect were estimated.

Afterward, the individual growth curve parameters (Ψ_i) were calculated based on 100 samples from the conditional distribution $p(\Psi_i | y_i; \widehat{\Psi}_{\text{pop}}; \hat{\Omega}; \hat{\sigma}_e^2)$. Due to the fact

that Ψ_i as estimated from the R package “saemix” comprised the herd effect again, the conditional means of each individual were further adjusted for herd and birth year. Thus, we applied a weighted least squares model, which incorporated the respective conditional variances. The statistic model in scalar notation was

$$y_{ijk} = h_i + b_j + e_{ijk}, \quad [2]$$

where y_{ijk} = conditional means of individual parameters for animal k within herd i and birth year j ; h_i = herd i ; b_j = birth year j ; and e_{ijk} = random residual effects with $e_{ijk} \sim N(0, W_k \sigma_e^2)$, where W_k = reciprocals of conditional variance for animal k . Residuals for the growth curve parameters from the weighted least squares analyses were used as pseudo-phenotypes in the ongoing genomic analyses.

Genetic Parameter Estimations

The residuals of 10 curve parameters (3 for LOG, 3 for GOM, and 4 for RICH) from 620 genotyped female cattle were used as pseudo-phenotypes in animal models to estimate genomic curve coefficients for all animals. We applied the REML algorithm as implemented in the DMU v. 6 software package (Madsen and Jensen, 2013). The statistical model was

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{g} + \mathbf{e}, \quad [3]$$

where $\mathbf{y}$ = a vector of pseudo-phenotypes; μ = the overall mean; $\mathbf{g}$ = a vector of random polygenic effects

Table 1. Applied growth functions for the modeling of growth curves

Model	Equation ¹	Reference
Logistic	$W(t) = \frac{W_m}{1 + \left[\left(\frac{W_m}{W_0} \right) - 1 \right] e^{-kt}}$	France et al., 1996
Gompertz	$W(t) = W_m e^{\ln\left(\frac{W_0}{W_m}\right) e^{-kt}}$	Teleken et al., 2017
Richards	$W(t) = \frac{W_m W_0}{\left[W_0^m + (W_m^m - W_0^m) e^{-kt} \right]^{\frac{1}{m}}}$	France et al., 1996

¹ $W(t)$ = BW at age t ; W_0 = birth weight; W_m = mature BW; k = constant growth rate; m = shape parameter; t = age in months, and e = a mathematical constant.

with a variance-covariance structure of $\mathbf{g} \sim N(0, \mathbf{K}\sigma_g^2)$, where $\mathbf{K}$ = the similarity matrix between individuals, σ_g^2 = the polygenic variance, and $\mathbf{e}$ = a vector of random residual effects with $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$.

Using the genomic marker data, 5 similarity matrices were constructed. First, we created $\mathbf{K}$ according to the genomic relationship matrix as introduced by VanRaden (2008). Therefore, $\mathbf{G}_1 = \frac{\mathbf{Z}\mathbf{Z}'}{\sum 2p_k(1-p_k)}$, where

$\mathbf{Z} = \mathbf{M} - 2\mathbf{P}$ with $\mathbf{M}$ = genotypes (coded as 0, 1, or 2) of the animals; $\mathbf{P}$ = a matrix with allele frequency for marker k in column k ; p_k = allele frequency for marker k in the genotyped population (k varied from 1 to 44,314). The second similarity matrix followed the same rules, but the centralized $\mathbf{Z}$ matrix was standardized with $\sqrt{2p_k(1-p_k)}$, implying $\mathbf{G}_2 = \frac{\mathbf{Z}\mathbf{D}\mathbf{Z}'}{p}$, where p = the number of markers after filtering (44,314) and $\mathbf{D}$ = a p -by- p diagonal matrix with $\frac{1}{2p_k(1-p_k)}$ on the diagonal (Amin et al., 2007). To avoid singularities of $\mathbf{G}_1$ and $\mathbf{G}_2$, a small value of 0.02 was added to the diagonal of both matrices.

In addition to the 2 genomic relationship matrices $\mathbf{G}_1$ and $\mathbf{G}_2$, 3 kernel matrices were constructed to measure the similarities in terms of genomic marker conformity among all individuals. The weighted “alike by state” kernel function, as introduced by Wu et al. (2011), was defined as the first kernel ($\mathbf{K}_1$). The respective formula is as follows: $\mathbf{K}_1(i, j) = \sum_{k=1}^p w_k (2 - |x_{ik} - x_{jk}|)$, where p was the number of markers; x_{ik} was the genotype of the k th marker for animal i ; and $w_k = \frac{1}{\sqrt{p_k(1-p_k)}}$. Hence,

the allele difference of every SNP was weighted considering the respective allele frequency (Wang et al., 2015). For the exponential kernel $\mathbf{K}_2$ based on a dissimilarity score ($\mathbf{S}$), we used the following algorithm from González-Recio et al. (2008): (1) calculation of frequencies for marker k and genotype s (f_{ks}), implying 1 of the 3 frequencies for every marker corresponding to genotypes 0, 1, and 2 when genotypes for animal i and j at the k th marker are identical; (2) calculation of the dissimilarity score S between animal i and j for the k th marker on chromosome 1

$$\begin{cases} T_k = T_{k-1}f_{ks} \text{ and } S_k = S_{k-1}; \text{ if } x_{ik} = x_{jk} \\ S_k = S_{k-1} + T_{k-1} \text{ and } T_k = 1; \text{ if } x_{ik} \neq x_{jk} \end{cases},$$

where T_k = temporary dissimilarity score for the k th marker, x_{ik} = genotype for animal i and marker k , $T_0 = 1$, and $S_0 = 0$; the loop ran from the first to the last SNP on chromosome 1, resulting in S_{chr1} ; (3) calculation of the dissimilarity score S as described in step 2 for all the other chromosomes, one by one; and (4) calculation of the overall $S = (S_{\text{chr1}} + \dots + S_{\text{chr29}})/p$. The exponential kernel was calculated as $\mathbf{K}_2 = \exp(-S)$.

Another exponential kernel, that is, the Gaussian kernel ($\mathbf{K}_3$), was formed as

$$\mathbf{K}_3(i, j) = \exp\left(-\frac{\|x_i - x_j\|^2}{\delta}\right) = \exp\left(-\frac{(x_i - x_j)'(x_i - x_j)}{\delta}\right),$$

where δ was a bandwidth parameter. The optimal bandwidth parameter was chosen among a grid of values ranging from $0.1 \times \max\{\|x_i - x_j\|^2\}$ to $1.0 \times \max\{\|x_i - x_j\|^2\}$ with an interval of 0.1. Evaluation of the bandwidth parameters was carried out using the R package “rrBLUP” (Endelman, 2011) with respect to the log-likelihood of the model when fitting different δ to scale the kernel matrix. Generally, for all the growth parameters, the log-likelihood was maximized when δ was $0.8 \times \max\{\|x_i - x_j\|^2\}$. Therefore, the optimal bandwidth for $\mathbf{K}_3$ in this study was always set to $0.8 \times \max\{\|x_i - x_j\|^2\}$.

Additionally, genomic growth curve parameters for all female cattle (4,952) were estimated via genomic animal models, applying REML as implemented in the software package DMU v. 6 (Madsen and Jensen, 2013):

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Z}\mathbf{u} + \mathbf{e}, \quad [4]$$

where $\mathbf{y}$, μ , and $\mathbf{e}$ were the same as defined in model 3; $\mathbf{u}$ = a vector of estimated GBV with $\mathbf{u} \sim N(0, \mathbf{H}\sigma_u^2)$;

σ_u^2 = the genetic variance; and $\mathbf{Z}$ = incidence matrix for $\mathbf{u}$. The combined $\mathbf{H}$ matrix was computed by blending $\mathbf{A}$ and the weighted genomic relationship matrix ($\mathbf{G}_w$; Legarra et al., 2009). $\mathbf{G}_w$ was calculated as follows: $\mathbf{G}_w = (0.95 \times \mathbf{G}_1 + 0.05 \times \mathbf{A}_{22})$, where $\mathbf{A}_{22}$ is the submatrix of the pedigree-based relationship matrix for genotyped animals. Heritabilities of the growth curve parameters were estimated from univariate models. Genetic and phenotypic correlations among all growth

Table 2. Descriptions of the models as applied in the present study¹

Model	Model no.	Dependent variable	Similarity matrix	No. of animals	No. of observations
$y_{ij} = f(t_{ij}, \Psi_i) + e_{ij}$	[1a]	Longitudinal BW	—	4,952	31,722
$\log(y_{ij}) = \log[f(t_{ij}, \Psi_i)] + e_{ij}$	[1b]				
$y_{ijk} = h_i + by_j + e_{ijk}$	[2]	Individual growth curve parameters from model 1b	—	4,952	4,952
$\mathbf{y} = \mathbf{1}\mu + \mathbf{g} + \mathbf{e}$	[3]	Residuals from model 2	G1, G2, K1, K2, and K3	620	620
$\mathbf{y} = \mathbf{1}\mu + \mathbf{g} + \mathbf{e}$	[4]	Residuals from model 2	H	4,952	4,952
$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Q}\mathbf{d} + \mathbf{W}\mathbf{m} + \mathbf{Z}\mathbf{p} + \mathbf{S}\mathbf{p}_m + \mathbf{e}$	[5]	Longitudinal BW	H	4,952	31,722

¹**G1** = genomic relationship matrix according to VanRaden, 2008; **G2** = genomic relationship matrix according to Amin et al., 2007; **K1** = kernel matrix according to Wu et al., 2011; **K2** = kernel matrix according to González-Recio et al., 2008; **K3** = Gaussian kernel; **H** = combined matrix blending **A** and the weighted genomic relationship matrix according to Legarra et al., 2009; σ_g^2 = variance explained by the similarity matrix.

curve parameters were estimated with model 3, implying series of bivariate runs for all trait combinations. Standard errors for phenotypic and genetic correlations were calculated based on the estimated (co)variance component matrix, using the “deltamethod” function from the R package “msm” (Jackson, 2019).

Genomic Random Regression Model

Data basis was 31,722 BW records from 4,952 heifers. A Bayesian approach, as implemented in the DMU v. 6 package (Madsen and Jensen, 2013), for genomic RRM was applied to estimate genomic breeding values (**GBV-RRM**) for BW at specific age months. The RRM was defined according to Yin and König (2018):

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Q}\mathbf{d} + \mathbf{W}\mathbf{m} + \mathbf{Z}\mathbf{p} + \mathbf{S}\mathbf{p}_m + \mathbf{e}, \quad [5]$$

where $\mathbf{y}$ = vector of observations for longitudinal BW; $\mathbf{b}$ = vector of fixed effects including herd-year-month, and regressions on age in month using cubic Legendre polynomials nested within birth year; $\mathbf{d}$ = vector of direct additive genetic effects for random regression coefficients, which were modeled with Legendre polynomials of order 3; $\mathbf{p}$ = vector of permanent environmental effects for random regression coefficients, which were modeled with Legendre polynomials of order 3; $\mathbf{m}$ = vector of maternal genetic effects; $\mathbf{p}_m$ = vector of maternal permanent environmental effects; $\mathbf{e}$ = vector of random residual effects; **S** = incidence matrices for $\mathbf{b}$ and $\mathbf{p}_m$, respectively; and **X**, **Q**, **W**, and **Z** = coefficient matrices for $\mathbf{d}$, $\mathbf{m}$, and $\mathbf{p}$, respectively. Heterogeneous residual variances were assumed across the age scale for the following time intervals: 0, 1 to 4, 5 to 8, 9 to 12, 13 to 16, 17 to 20, and 21 to 24 mo. Again, the genetic relationships were constructed using the **H** matrix (Legarra et al., 2009) as explained above.

In the Bayesian framework, we considered a chain length of 300,000 iterations, a burn-in period of 60,000 rounds, and a thinning interval of 10. Effective sample sizes of the estimates ranged from 23.1 for the maternal permanent environmental effect to 21,650.6 for the residual component in mo 5 to 8. Estimated random regression coefficients for direct genetic and permanent environmental effects were averaged, considering the 24,000 Gibbs samples. Afterward, GBV-RRM were calculated by multiplying the final estimated random regression coefficients for direct genetic effects (matrix **A_d** of dimension $20,176 \times 4$) with Legendre polynomials coefficients

$$\mathbf{L} = \left[\sqrt{\frac{1}{2}}, \sqrt{\frac{3}{2}}t_x, \sqrt{\frac{5}{2}}\left(\frac{3}{2}t_x^2 - \frac{1}{2}\right), \sqrt{\frac{7}{2}}\left(\frac{5}{2}t_x^3 - \frac{3}{2}t_x\right) \right],$$

$$t_x = -1 + 2 \frac{(t_l - 0)}{24 - 0},$$

and t_l from 0 to 24 mo, implying the calculation of GBV-RRM via **A_dL'**. Hence, GBV-RRM were available for all animals as included in the pedigree file, with 25 monthly GBV from age 0 to 24 mo.

A detailed overview addressing all models 1 to 5 is given in Table 2.

Accuracies of Genomic Growth Curves

We focused on 3 different evaluation strategies (Figure 2). First, we considered the 620 genotyped female cattle and estimated their GBV (application of model 3) for the pseudo-phenotypes. Afterward, the GBV served as input parameters for the calculation of individual genomic growth curves using the respective NLINM. The genomic growth curves for the 620 cattle were correlated with GBV-RRM from model 5. Alter-

natively, in a second evaluation strategy, GBV for the growth curve parameters (also solutions from model 3) considering the 43 genotyped sires with more than 20 phenotyped daughters, were used for genomic growth curve modeling via NLINM. The correlations between predicted genomic growth curves of the 43 sires and GBV-RRM for the same sires from model 5 were defined as prediction accuracies for sires.

In a third evaluation approach, we focused on a 5-fold cross-validation considering the 620 genotyped female cattle. The females were randomly assigned into 5 groups. For the creation of validation sets, pseudo-phenotypes of cattle in groups 1 to 5 were masked consecutively, and the cattle from the remaining 4 groups represented the training sets. Hence, for the estimation of GBV, pseudo-phenotypes of growth curve parameters from training set animals (496 cattle) were used as input data for model 3. The predicted GBV of the 124 nonphenotyped female cattle in validation sets were input for the respective NLINM to calculate the 3 genomic growth curves. The genomic growth curves for the 124 cattle were correlated with the GBV-RRM of the same animals. The average correlation from 5 validation sets represented accuracies of prediction from 5-fold cross-validations.

The correlations between genomic growth curves and GBV-RRM were calculated as follows: (1) correlations

between aggregated genomic growth curves for age 0 to 24 mo considering GBV of growth curve parameters and aggregated GBV-RRM from the 25 age months (r_{sum}); (2) monthly correlations between genomic growth curves and GBV-RRM for ages 0 to 24 mo (r_{mon}); (3) average of the 25 correlations (birth to 24 mo) from step 2 (r_{mean}). Standard errors (SE) for the correlation coefficients were calculated via a bootstrap approach, using the “boot” function with 1,000 replicates, as implemented in the R package “boot” (Canty and Ripley, 2019).

RESULTS AND DISCUSSION

Nonlinear Models and Growth Curve Parameters

With regard to the growth curve modeling, consideration of the herd effect plus fitting an exponential residual substantially improved the model evaluation criteria (Table 3). This was the case for all 3 functions LOG, GOM, and RICH. Consequently, in the ongoing genomic analyses, pseudo-phenotypes were growth curve parameters estimated from NLINM with herd effect and exponential residual. An across-function comparison for $-2LL$, Akaike information criterion, and Bayesian information criterion indicated superiority of RICH over GOM and LOG. Additionally, the residuals

Figure 2. Strategies to evaluate accuracies of genomic growth curves predicted from genomic breeding values of growth curve parameters.

Table 3. Model evaluations for the nonlinear growth functions (models 1a and 1b) with different covariates and residuals based on -2 times maximized log-likelihood ($-2LL$), Akaike information criterion (AIC), and Bayesian information criterion (BIC)

Growth function	Covariate	Residual	Evaluation criterion		
			$-2LL$	AIC	BIC
Logistic	—	Homogeneous	289,798.7	289,812.7	289,858.2
	—	Exponential	279,817.7	279,831.7	279,877.3
	Herd	Homogeneous	284,216.8	284,284.8	284,506.0
	Herd	Exponential	276,447.4	276,515.4	276,736.7
Gompertz	—	Homogeneous	282,626.9	282,640.9	282,686.5
	—	Exponential	275,008.9	275,022.9	275,068.5
	Herd	Homogeneous	278,917.5	278,985.5	279,206.7
	Herd	Exponential	272,580.3	272,648.3	272,869.5
Richards	—	Homogeneous	283,757.6	283,775.6	283,834.2
	—	Exponential	274,851.8	274,869.8	274,928.4
	Herd	Homogeneous	277,796.9	277,886.9	278,179.8
	Herd	Exponential	270,405.9	270,495.9	270,788.7

for the 3 NLINM ranged from 0.08 (for GOM and RICH) to 0.09 (for LOG; Table 4), again indicating the superiority of GOM and RICH over LOG. The RICH function is a generalized growth function for a variety of other functions (i.e., also for LOG and GOM) due to the possibility of modeling the shape parameter m (Richards, 1959). For m approaching 0, RICH can be transformed to GOM, and for $m = 1$, RICH reflects LOG (Teleken et al., 2017). For our BW data, the population value for m was 0.67 (Table 4), indicating an intermediate growth curve shape between GOM and LOG. Substituting $m \rightarrow 0$, $m = 0.67$, and $m = 1$ into

$$W_{\text{inflection}} = \left(\frac{1}{1+m} \right)^{\frac{1}{m}} W_m \quad (\text{Goshu and Rao, 2013}), \text{ inflec-}$$

tion points for LOG, RICH, and GOM were located at 50%, 46%, and 37% of the mature weight, respectively.

The population parameter estimates for birth weight (W_0) were 40.17 kg (GOM), 40.49 kg (RICH), and 40.72 kg (LOG). The population mature BW (W_m) ranged from 436.45 kg for LOG to 571.19 kg for GOM (Table 4). The population growth rates k ranged from 0.14 (GOM) to 0.31 (LOG) with very small variances.

However, quite large variances were observed for W_m , indicating the substantial BW variations of Holstein heifers close to the first insemination date.

Vázquez et al. (2012) compared the von Bertalanffy, Weibull (Seber and Wild, 1989), modified Hill (López et al., 2000), LOG, and GOM functions, and they identified superiority of the von Bertalanffy function for growth curve modeling of 6 dairy cattle breeds (Ayrshire, Jersey, Guernsey, Holstein, Charolais, and French Friesian). Teleken et al. (2017) applied RICH to modeling growth curves of Holstein bulls and identified an m value of -0.22 (indicator for a von Bertalanffy function). Beltrán et al. (1992) compared Brody with RICH in 2 lines of Angus beef cows, and they suggested the application of specific functions for specific age stages. Koskan and Ozkaya (2014) made similar conclusions for modeling growth curves of Holstein calves. Hence, due to the flexible function parameter m , we generally suggest RICH applications for nonlinear growth curve modeling based on BW data.

The growth curves for 1 cow with 4 observations and 1 animal with 19 observations are plotted in Figure 3A and 3B, respectively. For only a few BW measure-

Table 4. Estimated population growth curve parameters from model 1b (as described in Table 2) with herd effect and exponential residual

Growth function	Growth curve parameters ¹								
	W_0	W_m	k	m	σ_e	$\sigma_{W_0}^2$	$\sigma_{W_m}^2$	$\sigma_k^2 \times 10^4$	σ_m^2
Logistic	40.72	436.45	0.31	—	0.09	10.18	381.21	4.79	—
Gompertz	40.17	571.19	0.14	—	0.08	11.75	437.13	0.71	—
Richards	40.49	452.45	0.25	0.67	0.08	11.30	602.30	0.61	0.01

¹ W_0 = birth weight; W_m = mature BW; k = constant growth rate; m = shape parameter; σ_e = residual standard deviation; $\sigma_{W_0}^2$ = variance of birth weight; $\sigma_{W_m}^2$ = variance of mature BW; σ_k^2 = variance of growth rate; σ_m^2 = variance of shape parameter.

ments during growth, LOG underestimated BW in the period from 18 to 24 mo of age. In contrast to LOG, growth curves predicted from GOM and RICH showed an almost linear positive slope during the same measuring period. The observed differences between GOM and RICH with LOG might be due to the variations in the estimations for k and W_m . More specifically, k was 0.32 for LOG, 0.14 for GOM, and 0.16 for RICH, and W_m

was 397.77 kg for LOG, 525.44 kg for GOM, and 490.03 kg for RICH. For a denser BW structure (Figure 3B) with more BW observations, LOG and RICH generated similar growth curves. Furthermore, only minor differences between GOM curve predictions and real BW observations were observed in the interval from 22 to 24 mo of age.

Variance Components of Growth Curve Parameters

Variance components and genetic parameters for the 10 parameters estimated from the genomic models (model 3) with different similarity matrices are listed in Table 5. For the genomic relationship matrices **G1** and **G2**, σ_g^2 in relation to σ_t^2 are the heritabilities of the growth parameters. Heritabilities for W_0 from LOG and GOM and considering **G1** and **G2** were larger than 0.40. From RICH, the heritabilities for W_0 were 0.35 (**G1**) and 0.37 (**G2**). Heritabilities for W_m in the range from 0.32 to 0.39 were slightly smaller than heritabilities for W_0 . The moderate to high heritabilities for the growth curve parameters W_0 and W_m confirm previous estimates based on “conventional” modeling approaches considering pedigree relationships. For birth weight in Holstein cattle, heritabilities ranged from 0.26 (Johanson et al., 2011) to 0.47 (Yin and König, 2018). For mature weight in Angus beef cattle, Kaps et al. (1999) estimated heritabilities (based on Brody curve predictions) in the range from 0.44 (estimate from a single-trait model) to 0.52 (estimate from a bivariate model). Meyer (1995) used GOM and estimated a heritability for mature weight of 0.49. Our W_m heritabilities of 0.32 (**G1** and RICH) and 0.39 (**G2** and GOM) are in agreement with the heritabilities for mature weight from growth curves using lifetime BW records in 5 cattle breeds (Johnson et al., 1990) and for live weight of lactating Holstein cows (Lassen and Løvendahl, 2016). For the growth rate k from GOM applications in Hereford cattle, Meyer (1995) reported heritabilities ranging from 0.37 to 0.48. Equivalent heritabilities (0.38 to 0.46) were estimated in the present study. However, in the study by Meyer (1995), values for k were very small, with $k < 0.005$. In the present study, k ranged from 0.01 to 0.36. Substantially lower heritabilities (< 0.10) for k were estimated when applying the von Bertalanffy function to modeling growth curves in 2 beef cattle breeds (Oliveira et al., 1994; Forni et al., 2007). Also for the shape parameter m , heritabilities were moderate: 0.29 (**G1**) and 0.31 (**G2**). As far as we know, this is a first study estimating heritabilities for the shape parameter (here based on RICH predictions). However, as an approximation, inflection points also describe the

Figure 3. Growth curves for animals with 4 (A) and 19 (B) repeated BW measurements estimated from model 1b.

Table 5. Variance components and variance ratios for the parameters of the growth functions from model 3 considering different similarity matrices¹

Similarity matrix	Growth function	Parameter	Variance component				SE	h_c^2
			σ_g^2	σ_e^2	σ_t^2	σ_g^2/σ_t^2		
G1	Logistic	W_0	2.60	3.66	6.26	0.42	0.10	0.42
		W_m	38.51	71.74	110.25	0.35	0.09	0.35
		$k \times 10^4$	0.93	1.58	2.51	0.37	0.09	0.37
	Gompertz	W_0	3.06	4.60	7.66	0.40	0.10	0.40
		W_m	23.77	39.76	63.53	0.37	0.09	0.38
		$k \times 10^4$	0.17	0.27	0.44	0.40	0.09	0.40
	Richards	W_0	2.61	4.80	7.41	0.35	0.09	0.35
		W_m	76.46	164.49	240.96	0.32	0.08	0.32
		$k \times 10^4$	0.04	0.04	0.08	0.46	0.05	0.47
		$m \times 10^4$	13.39	32.05	45.44	0.29	0.09	0.30
G2	Logistic	W_0	2.70	3.55	6.25	0.43	0.10	0.44
		W_m	39.32	71.00	110.32	0.36	0.09	0.36
		$k \times 10^4$	1.01	1.51	2.51	0.40	0.10	0.40
	Gompertz	W_0	3.18	4.48	7.65	0.41	0.10	0.42
		W_m	24.70	38.93	63.63	0.39	0.09	0.39
		$k \times 10^4$	0.19	0.25	0.44	0.42	0.10	0.43
	Richards	W_0	2.71	4.69	7.40	0.37	0.10	0.37
		W_m	79.27	161.75	241.02	0.33	0.09	0.33
		$k \times 10^4$	0.02	0.04	0.06	0.38	0.04	0.38
		$m \times 10^4$	14.09	31.36	45.45	0.31	0.10	0.31
K1	Logistic	W_0	13.87	2.95	16.83	0.82	0.07	0.53
		W_m	217.93	58.72	276.65	0.79	0.08	0.47
		$k \times 10^4$	5.35	1.23	6.58	0.81	0.07	0.51
	Gompertz	W_0	16.27	3.78	20.04	0.81	0.08	0.51
		W_m	127.16	33.26	160.42	0.79	0.07	0.48
		$k \times 10^4$	0.97	0.21	1.18	0.82	0.07	0.52
	Richards	W_0	13.49	4.16	17.65	0.76	0.09	0.44
		W_m	421.32	140.74	562.06	0.75	0.08	0.42
		$k \times 10^4$	0.22	0.05	0.27	0.81	0.07	0.50
		$m \times 10^4$	72.81	28.20	101.01	0.72	0.10	0.38
K2	Logistic	W_0	11.35	2.77	14.13	0.80	0.09	0.56
		W_m	194.20	51.26	245.47	0.79	0.08	0.54
		$k \times 10^4$	4.32	1.18	5.49	0.79	0.09	0.53
	Gompertz	W_0	13.60	3.48	17.09	0.80	0.09	0.54
		W_m	106.83	30.81	137.64	0.78	0.08	0.51
		$k \times 10^4$	0.79	0.20	0.99	0.80	0.08	0.55
	Richards	W_0	11.36	3.89	15.25	0.74	0.10	0.47
		W_m	363.38	130.09	493.47	0.74	0.09	0.46
		$k \times 10^4$	0.18	0.05	0.23	0.79	0.08	0.54
		$m \times 10^4$	58.77	27.49	86.25	0.68	0.12	0.40
K3	Logistic	W_0	6.12	1.85	7.96	0.77	0.14	0.70
		W_m	104.25	35.98	140.23	0.74	0.14	0.67
		$k \times 10^4$	2.40	0.78	3.18	0.75	0.14	0.69
	Gompertz	W_0	7.28	2.41	9.69	0.75	0.15	0.68
		W_m	63.01	18.46	81.47	0.77	0.14	0.71
		$k \times 10^4$	0.44	0.12	0.57	0.78	0.14	0.72
	Richards	W_0	6.20	2.91	9.11	0.68	0.15	0.60
		W_m	205.03	94.27	299.30	0.69	0.14	0.61
		$k \times 10^4$	0.10	0.03	0.13	0.76	0.13	0.69
		$m \times 10^4$	31.86	22.61	54.47	0.58	0.16	0.50

¹**G1** = genomic relationship matrix according to VanRaden, 2008; **G2** = genomic relationship matrix according to Amin et al., 2007; **K1** = kernel matrix according to Wu et al., 2011; **K2** = kernel matrix according to González-Recio et al., 2008; **K3** = Gaussian kernel; σ_g^2 = variance explained by the similarity matrix; σ_e^2 = residual variance; σ_t^2 = total variance; σ_g^2/σ_t^2 = variance explained by the similarity matrix in relation to the total variance; SE = standard error of σ_g^2/σ_t^2 ; h_c^2 = corrected heritability according to Legarra (2016); W_0 = birth weight; W_m = mature BW; k = constant growth rate; m = shape parameter.

shape of growth curves. Heritabilities for the inflection points in chicken populations were 0.25 (Manjula et al., 2018) and 0.50 (Grossman and Bohren, 1985).

In comparison with **G1** and **G2**, most of the variance components increased when considering the kernels **K1**, **K2**, or **K3**. The substantially larger variances

Table 6. Heritabilities (bold and diagonal), genetic (above diagonal) and phenotypic correlations (below diagonal) between growth curve parameters (W_0 = birth weight; W_m = mature BW; k = constant growth rate; m = shape parameter) from model 4 using a relationship matrix combining pedigree and genomic information (**H** matrix); SE are given below the parameter estimates

Growth function	Parameter	Logistic			Gompertz			Richards			
		W_0	W_m	k	W_0	W_m	k	W_0	W_m	k	m
Logistic	W_0	0.42	0.33	0.49	0.99	0.35	0.48	0.99	0.39	0.36	−0.39
		0.04	0.08	0.07	0.00	0.08	0.07	0.00	0.08	0.08	0.08
	W_m	0.07	0.30	0.69	0.36	0.95	0.81	0.32	0.96	0.86	−0.56
		0.02	0.04	0.06	0.08	0.01	0.04	0.08	0.01	0.04	0.08
	k	0.27	0.44	0.25	0.47	0.78	0.97	0.41	0.73	0.91	−0.95
Gompertz	W_0	0.01	0.01	0.03	0.07	0.05	0.01	0.07	0.06	0.02	0.01
		0.98	0.09	0.24	0.43	0.36	0.47	1.00	0.40	0.36	−0.35
	W_m	0.00	0.02	0.01	0.04	0.08	0.07	0.00	0.08	0.08	0.08
		0.11	0.89	0.57	0.12	0.31	0.88	0.31	0.95	0.92	−0.67
	k	0.02	0.00	0.01	0.02	0.04	0.03	0.08	0.02	0.02	0.07
Richards	W_0	0.28	0.60	0.96	0.26	0.71	0.25	0.41	0.84	0.94	−0.90
		0.01	0.01	0.00	0.01	0.01	0.02	0.07	0.04	0.01	0.02
	W_m	0.97	0.10	0.19	0.98	0.11	0.22	0.43	0.37	0.31	−0.29
		0.00	0.02	0.02	0.00	0.02	0.02	0.04	0.08	0.08	0.08
	k	0.08	0.91	0.51	0.09	0.83	0.66	0.10	0.28	0.84	−0.60
	m	0.02	0.00	0.01	0.02	0.00	0.01	0.02	0.04	0.04	0.08
		0.18	0.63	0.88	0.17	0.73	0.90	0.13	0.62	0.32	−0.88
		0.02	0.01	0.00	0.02	0.01	0.00	0.02	0.01	0.03	0.03
		−0.30	−0.25	−0.91	−0.26	−0.39	−0.85	−0.20	−0.29	−0.75	0.22
		0.01	0.01	0.00	0.01	0.01	0.00	0.02	0.01	0.01	0.03

linked to the kernel and smaller residuals variances contributed to an increase of the variance ratio σ_g^2/σ_t^2 .

The variance components were further transformed into comparable scales by scaling the variances linked to the similarity matrices with D_k , where $D_k = \text{diag}(\mathbf{K}) - \bar{\mathbf{K}}$ (Legarra, 2016). In our study, D_k was 1.01 for **G1** and **G2**, 0.24 for **K1**, 0.30 for **K2**, and 0.71 for **K3**. Corrected heritabilities from **K1** and **K2** were larger than from **G1** and **G2**. When setting **K3** as relationship matrix, the corrected heritabilities increased to 0.50 for the shape parameter and to 0.72 for the growth rate. However, SE for the variance ratios from **K3** were quite large (Table 5).

The heritabilities from the single-trait model 4 as well as genetic (r_g) and phenotypic correlations (r_p ; from the series of bivariate runs using model 4) among growth curve parameters are given in Table 6. Heritabilities for growth curve parameters from the **H** matrix approach were identical or slightly smaller compared with the respective parameter heritabilities based on **G1** or **G2** modeling. However, compared with the **H** matrix estimates, heritabilities and variance ratios from **G1** and **G2** approaches had larger SE, probably due to the different numbers of animals with pseudo-phenotypes (4,952 cattle in model 4, and 620 cattle in model 3). Genetic correlations between the same growth curve parameters from different NLINM were larger than 0.90: 0.99 between W_0 from LOG with W_0 from GOM or RICH. The corresponding phenotypic correlations

were larger than 0.82. Birth weight (W_0) was genetically moderately correlated with growth rate (k) and with mature weight (W_m ; from 0.33 to 0.49). Corresponding phenotypic correlations ranged from 0.07 to 0.28. The large correlations between k and W_m (0.69 to 0.88 for r_g , and 0.44 to 0.73 for r_p) indicate the strong influence of growth rate on mature weight phenotypically and genetically. Accordingly, a pronounced favorable genetic correlation between k and W_m (0.82) was reported for predictions from the von Bertalanffy function in Nelore cattle (Forni et al., 2007). The growth rate k was genetically negatively correlated with the shape parameter m (r_g from −0.88 to −0.95). Also for all remaining curve parameters, genetic correlations with m were negative. The negative correlations indicate that increasing W_0 , W_m , and k influences the shape of growth curves from RICH by decreasing the percentages of mature weight at inflection points. Standard errors for correlation estimates were quite small, between 0.01 and 0.08 for r_g and generally lower than 0.02 for r_p .

Accuracies of Genomic Growth Curves for Animals with Phenotypes

The correlations between genomic growth curves based on GBV of growth curve parameters and respective GBV-RRM for all genotyped female cattle are listed in Table 7. For r_{sum} , the accuracies ranged from 0.75 to 0.79 with corresponding SE from 0.01 to 0.02. Accuracies were largest when combining estimated

Figure 4. Correlations between genomic growth curves (model 3) and respective genomic breeding values from the random regression model per age month (model 5). **G1** = genomic relationship matrix according to VanRaden, 2008; **G2** = genomic relationship matrix according to Amin et al., 2007; **K1** = kernel matrix according to Wu et al., 2011; **K2** = kernel matrix according to González-Recio et al., 2008; **K3** = Gaussian kernel.

pseudo-phenotypes from RICH with **G1**, **G2**, **K1**, or **K2**. RICH always performed better than LOG or GOM, and, throughout, the lowest accuracies were identified for LOG. The increase in accuracies for RICH over LOG ranged from 0.02 (for **G2** and **K1**) to 0.03 (for **G1**, **K2** and **K3**), reflecting the phenotypic NLINM evaluations (Table 3). Apart from LOG, accuracies were maximized when modeling the kernel **K1**, but the accuracy differences for different similarity matrices were quite small.

Accuracies according to r_{mean} varied between 0.65 for GOM combined with **K3** and 0.68 for RICH combined with **G1**, **G2**, **K1**, or **K2**. Generally, values for r_{mean} were lower than the respective values for r_{sum} . Nevertheless, applications of RICH combined with **G2** or **K1**

implied the largest accuracies, independent from the evaluation criterion. Standard errors for r_{mean} were close to 0.02.

The detailed correlations in monthly intervals (r_{mon}) explain the differences between r_{sum} and r_{mean} (Figure 4). The correlations between genomic growth curves and GBV-RRM were larger than 0.60 from birth until the age of 20 mo. However, probably due to the low data coverage in the period from 21 to 24 mo, the accuracies substantially decreased from approximately 0.60 to -0.05 at the end of the time scale. Accordingly, the accuracy for r_{mean} was quite large (0.77 for **G1**) when excluding correlations beyond 21 mo. The pattern of accuracies along the growth trajectory was very similar

Table 7. Correlations between genomic growth curves (model 3) and genomic breeding values (GBV) from the random regression model (model 5) considering genotyped female cattle with BW records (SE ranged from 0.01 to 0.02)¹

Similarity matrix	r_{sum}			r_{mean}		
	Logistic	Gompertz	Richards	Logistic	Gompertz	Richards
G1	0.76	0.78	0.79	0.67	0.66	0.68
G2	0.77 ²	0.78	0.79	0.67 ²	0.66	0.68
K1	0.77	0.78 ²	0.79 ²	0.67	0.67 ²	0.68 ²
K2	0.76	0.78	0.79	0.67	0.66	0.68
K3	0.75	0.77	0.78	0.66	0.65	0.67

¹**G1** = genomic relationship matrix according to VanRaden, 2008; **G2** = genomic relationship matrix according to Amin et al., 2007; **K1** = kernel matrix according to Wu et al., 2011; **K2** = kernel matrix according to González-Recio et al., 2008; **K3** = Gaussian kernel; r_{sum} = correlation between the sum of genomic growth curve breeding values and the sum of GBV from the random regression model; r_{mean} = average of 25 monthly correlations between genomic growth curve breeding values and GBV from the random regression model.

²Highest correlation coefficient within each column.

Table 8. Correlations between growth curves calculated from genomic breeding values (GBV) of growth curve parameters (model 3) and GBV from the random regression model (model 5) for genotyped female cattle from a 5-fold cross-validation and for 43 genotyped sires (SE ranged from 0.07 to 0.08 for heifers and from 0.14 to 0.15 for sires)¹

Testing set	Similarity matrix	r_{sum}			r_{mean}		
		Logistic	Gompertz	Richards	Logistic	Gompertz	Richards
Female	G1	0.40	0.40	0.41	0.34 ²	0.33 ²	0.34 ²
	G2	0.40	0.40	0.41	0.34	0.33	0.34
	K1	0.40 ²	0.41 ²	0.42 ²	0.33	0.33	0.33
	K2	0.39	0.40	0.41	0.32	0.32	0.32
	K3	0.40	0.40	0.41	0.33	0.32	0.33
Sire	G1	0.43	0.44	0.41	0.38	0.39	0.37
	G2	0.43	0.44	0.42	0.38	0.39	0.38 ²
	K1	0.42	0.42	0.40	0.36	0.37	0.35
	K2	0.41	0.41	0.39	0.35	0.36	0.35
	K3	0.45 ²	0.45 ²	0.42 ²	0.38 ²	0.39 ²	0.37

¹**G1** = genomic relationship matrix according to VanRaden, 2008; **G2** = genomic relationship matrix according to Amin et al., 2007; **K1** = kernel matrix according to Wu et al., 2011; **K2** = kernel matrix according to González-Recio et al., 2008; **K3** = Gaussian kernel; r_{sum} = correlation between sum of genomic growth curves based on GBV of growth curve parameters and the sum of GBV from the random regression model; r_{mean} = average of 25 monthly correlations between genomic growth curves and GBV from the random regression model; Female = genotyped female cattle with masked pseudo-phenotypes; Sire = genotyped sires with more than 20 phenotyped daughters.

²Highest correlation coefficient for each nonlinear function based on 5-fold cross-validations for sires.

for the 5 different similarity matrices. Standard errors for r_{mon} were smallest at age 6 to 9 mo (0.01), but were larger at birth (0.02) and at age 24 mo (0.04).

Accuracies of Predicted Genomic Growth Curves for Animals Without Phenotypes

Genomic prediction accuracies for the female cattle from the 5-fold cross-validation and for the sires are given in Table 8 for r_{sum} and r_{mean} . Accuracies were lower than the accuracies for female cattle with pseudo-phenotypes, such as approximately 0.40 for r_{sum} (SE = 0.07 – 0.08) and approximately 0.33 for r_{mean} (SE = 0.08). The lower accuracies were expected, because only accuracies for female cattle with masked pseudo-phenotypes were considered. Genomic growth curves predicted from parameters of RICH showed only minor gains in accuracies compared with GOM and LOG. With a focus on the different similarity matrices, values for r_{sum} were again largest when modeling the kernel **K1**. According to r_{mean} , prediction accuracies were largest when modeling **G1** but showed only minor differences compared with the **G2** and **K1** modeling approaches. Standard deviations for the prediction accuracies from the 5-fold cross-validation ranged from 0.06 at birth to 0.14 at age 19 and 20 mo.

Given the fact that only influential sires with more than 20 phenotyped daughters were considered, the prediction accuracies for sires were larger than for the female cattle. The SE for the accuracies ranged from 0.14 to 0.15 for all NLINM and similarity matrix applications. However, the differences in prediction

accuracies between sires and female cattle were quite small, because only 31 sires additionally had daughters with genotypes. For sires, and for both evaluation criteria r_{sum} and r_{mean} , accuracies from GOM were slightly larger than from RICH or LOG. However, when considering the corresponding SE, differences in prediction accuracy with similarity combinations did not differ significantly. de los Campos et al. (2010b) compared the Gaussian kernel average with Bayesian ridge regression and with **G**. For protein content, Bayesian ridge regression and **G** performed better than the Gaussian kernel average, but for daughter fertility, de los Campos et al. (2010b) detected opposite results. Hence, trait specifics should be considered when selecting the most appropriate similarity matrix. Furthermore, de los Campos et al. (2010a) identified almost the same residual variances and mean squared errors from a Gaussian kernel with optimal bandwidth and from the Gaussian kernel average. In the current study, the optimal bandwidth for the Gaussian kernel **K3** was chosen according to log-likelihood values. Hence, for the sire data set evaluation, **K3** modeling was associated with the largest prediction accuracies (apart from r_{mean} and RICH pseudo-phenotypes). In contrast, in genomic predictions for sires without daughter records, González-Recio et al. (2008) indicated a better predictive ability for the kernel **K2** compared with **A**, trinomial kernels, linear regressions, or 24 informative SNP in a BayesA approach.

Correlations across the growth curve in monthly intervals (r_{mon}) for female cattle without phenotypes (SE = 0.07 – 0.09) and for sires (SE = 0.11 – 0.16)

Figure 5. Correlations between predicted genomic growth curves based on the genomic relationship matrix (model 3) and respective genomic breeding values from the random regression model per month (model 5). Female = genotyped female cattle with masked pseudo-phenotypes; sire = genotyped sires with more than 20 phenotyped daughters.

were almost the same for different similarity matrices. Thus, only results from **G1** are plotted in Figure 5. The prediction accuracies for female cattle and sires were very similar at the extreme ends of the age scale. From birth to 5 mo of age, r_{mon} varied from 0.3 and 0.4 for both validation strategies. Afterward, the correlation increased gradually to 0.58 for sires, but was 0.40 for female cattle at age 18 mo. At the end of the growth curve interval, probably due to the smaller number of animals with phenotypic records, r_{mon} in both data sets (sires and heifers) substantially decreased, and was close to 0 at 24 mo. The shape of the prediction accuracy curves (the values for r_{mon} at specific ages) was very similar for pseudo-phenotypes from LOG, GOM, and RICH applications for sires as well as for female cattle.

CONCLUSIONS

For the modeling of phenotypic growth curves based on longitudinal BW records in dairy cattle from birth to approximate age at first calving, RICH applications with herd effects and exponential residuals contributed to superior NLINM over LOG and GOM. Heritabili-

ties for NLINM growth curve parameters based on **G1** and **G2** modeling approaches were the same or slightly larger than the corresponding heritabilities based on the **H** matrix. Genetic correlations among the growth curve parameters W_0 , W_m , and k were always positive (>0.30), especially for the same growth curve parameters estimated from different NLINM (>0.90). The shape parameter m in RICH was genetically negatively associated with other growth curve parameters. Accuracies of the genomic growth curves for the 620 female cattle with pseudo-phenotypes favored RICH applications over LOG and GOM, irrespective of the similarity matrices used. However, due to the large SE, differences in prediction accuracies from the different approaches considering sires with more than 20 daughters with BW records, or the results from the 5-fold cross-validations, were not significant. Generally, when focusing on animals without phenotypes, prediction accuracies were larger for sires than for female cattle, especially for specific ages in the interval from age 6 mo to 18 mo. Hence, increasing the phenotypic BW data set is more important for the increase of genomic growth curve accuracies than are function or similarity matrix specifications. For complex modeling approaches, as used in the present study based on a quite large number of phenotypes, it is also imperative to increase the number of genotyped animals.

ACKNOWLEDGMENTS

The authors gratefully acknowledge funding from the German Federal Ministry of Education and Research (BMBF; Bonn, Germany) and from the Förderverein Bioökonomieforschung e.V. (FBF; Bonn, Germany)/German Holstein Association (DHV; Bonn, Germany) for the collaborative project “KMU-innovativ-10: Kuhl—Cow calibration groups for the implementation of selection strategies based on high-density genotyping in dairy cattle,” grant no. 031A416C. The authors have not stated any conflicts of interest.

REFERENCES

- Amin, N., C. M. van Duijn, and Y. S. Aulchenko. 2007. A genomic background based method for association analysis in related individuals. *PLoS One* 2:e1274. <https://doi.org/10.1371/journal.pone.0001274>.
- Beltrán, J. J., W. T. Butts Jr., T. A. Olson, and M. Koger. 1992. Growth patterns of two lines of Angus cattle selected using predicted growth parameters. *J. Anim. Sci.* 70:734–741. <https://doi.org/10.2527/1992.703734x>.
- Berry, D. P., F. Buckley, P. Dillon, R. D. Evans, M. Rath, and R. F. Veerkamp. 2003. Genetic relationships among body condition score, body weight, milk yield, and fertility in dairy cows. *J. Dairy Sci.* 86:2193–2204. [https://doi.org/10.3168/jds.S0022-0302\(03\)73809-0](https://doi.org/10.3168/jds.S0022-0302(03)73809-0).

- Canty, A., and B. Ripley. 2019. Bootstrap Functions [R package boot version 1.3–24]. Accessed Jul. 30, 2019. <https://cran.r-project.org/web/packages/boot/boot.pdf>.
- Chang, C. C., C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, and J. J. Lee. 2015. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* 4:7. <https://doi.org/10.1186/s13742-015-0047-8>.
- Comets, E., A. Lavenu, and M. Lavielle. 2017. Parameter estimation in nonlinear mixed effect models using saemix, an R implementation of the SAEM algorithm. *J. Stat. Softw.* 80. <https://doi.org/10.18637/jss.v080.i03>.
- Connor, E. E. 2015. Invited review: Improving feed efficiency in dairy production: challenges and possibilities. *Animal* 9:395–408. <https://doi.org/10.1017/S1751731114002997>.
- de los Campos, G., D. Gianola, G. J. M. Rosa, K. A. Weigel, and J. Crossa. 2010a. Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. *Genet. Res. (Camb.)* 92:295–308. <https://doi.org/10.1017/S0016672310000285>.
- de los Campos, G., D. Gianola, G. J. M. Rosa, K. A. Weigel, A. I. Vazquez, and D. B. Allison. 2010b. Semi-parametric marker-enabled prediction of genetic values using reproducing kernel Hilbert spaces regressions. In *Proc. 9th World Congress on Genetics Applied to Livestock Production*, Leipzig, Germany. Accessed Nov. 20, 2020. <http://www.wcgalp.org/proceedings/2010>.
- Endelman, J. B. 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome* 4:250–255. <https://doi.org/10.3835/plantgenome2011.08.0024>.
- Fitzhugh, H. A. Jr. 1976. Analysis of growth curves and strategies for altering their shape. *J. Anim. Sci.* 42:1036–1051. <https://doi.org/10.2527/jas1976.4241036x>.
- Forni, S., M. Piles, A. Blasco, L. Varona, H. N. Oliveira, R. B. Lôbo, and L. G. Albuquerque. 2007. Analysis of beef cattle longitudinal data applying a nonlinear model. *J. Anim. Sci.* 85:3189–3197. <https://doi.org/10.2527/jas.2006-677>.
- France, J., J. Dijkstra, and M. S. Dhanoa. 1996. Growth functions and their application in animal science. *Anim. Res.* 45(Suppl. 1):165–174. <https://doi.org/10.1051/animres:19960637>.
- Gbangboche, A. B., R. Glele-Kakai, S. Salifou, L. G. Albuquerque, and P. L. Leroy. 2008. Comparison of non-linear growth models to describe the growth curve in West African dwarf sheep. *Animal* 2:1003–1012. <https://doi.org/10.1017/S1751731108002206>.
- Gianola, D., and J. B. C. H. M. van Kaam. 2008. Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. *Genetics* 178:2289–2303. <https://doi.org/10.1534/genetics.107.084285>.
- González-Recio, O., D. Gianola, N. Long, K. A. Weigel, G. J. M. Rosa, and S. Avendaño. 2008. Nonparametric methods for incorporating genomic information into genetic evaluations: An application to mortality in broilers. *Genetics* 178:2305–2313. <https://doi.org/10.1534/genetics.107.084293>.
- Goshu, A. T., and K. P. R. Rao. 2013. Derivation of inflection points of nonlinear regression curves—Implications to statistics. *Am. J. Theor. Appl. Stat.* 2:268–272. <https://doi.org/10.11648/j.ajtas.20130206.25>.
- Grossman, M., and B. B. Bohren. 1985. Logistic growth curve of chickens: Heritability of parameters. *J. Hered.* 76:459–462. <https://doi.org/10.1093/oxfordjournals.jhered.a110145>.
- Jackson, C. 2019. Multi-state modelling with R: themsmpackage. Accessed Oct. 16, 2019. <https://cran.r-project.org/web/packages/msm/vignettes/msm-manual.pdf>.
- Johanson, J. M., P. J. Berger, S. Tsuruta, and I. Misztal. 2011. A Bayesian threshold-linear model evaluation of perinatal mortality, dystocia, birth weight, and gestation length in a Holstein herd. *J. Dairy Sci.* 94:450–460. <https://doi.org/10.3168/jds.2009-2992>.
- Johnson, Z. B., C. J. Brown, and Brown, A. H., Jr. 1990. Evaluation of growth patterns of beef cows. *Ark. Agric. Exp. Stn. Bull.* 923.
- Kaps, M., W. O. Herring, and W. R. Lamberson. 1999. Genetic and environmental parameters for mature weight in Angus cattle. *J. Anim. Sci.* 77:569–574. <https://doi.org/10.2527/1999.773569x>.
- Kimeldorf, G., and G. Wahba. 1971. Some results on Tchebycheffian spline functions. *J. Math. Anal. Appl.* 33:82–95. [https://doi.org/10.1016/0022-247X\(71\)90184-3](https://doi.org/10.1016/0022-247X(71)90184-3).
- Koskan, O., and S. Ozkaya. 2014. Determination of growth curves of female Holstein calves using five non-linear models. *Pak. J. Agric. Sci.* 51:225–228.
- Lassen, J., and P. Løvendahl. 2016. Heritability estimates for enteric methane emissions from Holstein cattle measured using noninvasive methods. *J. Dairy Sci.* 99:1959–1967. <https://doi.org/10.3168/jds.2015-10012>.
- Legarra, A. 2016. Comparing estimates of genetic variance across different relationship models. *Theor. Popul. Biol.* 107:26–30. <https://doi.org/10.1016/j.tpb.2015.08.005>.
- Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. *J. Dairy Sci.* 92:4656–4663. <https://doi.org/10.3168/jds.2009-2061>.
- López, S., J. France, W. J. Gerrits, M. S. Dhanoa, D. J. Humphries, and J. Dijkstra. 2000. A generalized Michaelis-Menten equation for the analysis of growth. *J. Anim. Sci.* 78:1816–1828. <https://doi.org/10.2527/2000.7871816x>.
- Madsen, P., and J. Jensen. 2013. A user's guide to DMU: A package for analysing multivariate mixed models. Accessed Sep. 5, 2019. https://dmu.ghpc.au.dk/DMU/Doc/Current/dmuv6_guide.5.2.pdf.
- Mallick, B. K., D. Ghosh, and M. Ghosh. 2005. Bayesian classification of tumours by using gene expression data. *J. R. Stat. Soc. B* 67:219–234. <https://doi.org/10.1111/j.1467-9868.2005.00498.x>.
- Manjula, P., H.-B. Park, D. Seo, N. Choi, S. Jin, S. J. Ahn, K. N. Heo, B. S. Kang, and J.-H. Lee. 2018. Estimation of heritability and genetic correlation of body weight gain and growth curve parameters in Korean native chicken. *Asian-Australas. J. Anim. Sci.* 31:26–31. <https://doi.org/10.5713/ajas.17.0179>.
- Meyer, K. 1995. Estimates of genetic parameters for mature weight of Australian beef cows and its relationship to early growth and skeletal measures. *Livest. Prod. Sci.* 44:125–137. [https://doi.org/10.1016/0301-6226\(95\)00067-4](https://doi.org/10.1016/0301-6226(95)00067-4).
- Oliveira, H. N., R. B. Lôbo, and C. S. Pereira. 1994. Relationships among growth curve parameters, weights and reproductive traits in Guzera beef cows. In *Proc. 5th World Congress on Genetics Applied to Livestock Production*, Guelph, Canada.
- Pryce, J. E., O. Gonzalez-Recio, G. Nieuwhof, W. J. Wales, M. P. Coffey, B. J. Hayes, and M. E. Goddard. 2015. Hot topic: Definition and implementation of a breeding value for feed efficiency in dairy cows. *J. Dairy Sci.* 98:7340–7350. <https://doi.org/10.3168/jds.2015-9621>.
- Pryce, J. E., and B. L. Harris. 2006. Genetics of body condition score in New Zealand Dairy Cows. *J. Dairy Sci.* 89:4424–4432. [https://doi.org/10.3168/jds.S0022-0302\(06\)72490-0](https://doi.org/10.3168/jds.S0022-0302(06)72490-0).
- Purcell, S., and C. Chang. 2019. PLINK 1.9. Accessed Aug. 9, 2019. www.cog-genomics.org/plink/1.9/.
- Richards, F. J. 1959. A flexible growth function for empirical use. *J. Exp. Bot.* 10:290–301. <https://doi.org/10.1093/jxb/10.2.290>.
- Schaid, D. J. 2010a. Genomic similarity and kernel methods I: advancements by building on mathematical and statistical foundations. *Hum. Hered.* 70:109–131. <https://doi.org/10.1159/000312641>.
- Schaid, D. J. 2010b. Genomic similarity and kernel methods II: Methods for genomic information. *Hum. Hered.* 70:132–140. <https://doi.org/10.1159/000312643>.
- Seber, G. A. F., and C. J. Wild. 1989. *Nonlinear Regression*. John Wiley and Sons, New York, NY.
- Segelke, D., J. Chen, Z. Liu, F. Reinhardt, G. Thaller, and R. Reents. 2012. Reliability of genomic prediction for German Holsteins using imputed genotypes from low-density chips. *J. Dairy Sci.* 95:5403–5411. <https://doi.org/10.3168/jds.2012-5466>.
- Teleken, J. T., A. C. Galvão, and W. S. Robazza. 2017. Comparing non-linear mathematical models to describe growth of different animals. *Acta Sci. Anim. Sci.* 39:73–81. <https://doi.org/10.4025/actascianimsci.v39i1.31366>.
- VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91:4414–4423. <https://doi.org/10.3168/jds.2007-0980>.

- Vázquez, J. A., J. M. Lorenzo, P. Fuciños, and D. Franco. 2012. Evaluation of non-linear equations to model different animal growths with mono and bisigmoid profiles. *J. Theor. Biol.* 314:95–105. <https://doi.org/10.1016/j.jtbi.2012.08.027>.
- Visscher, P. M., P. J. Bowman, and M. E. Goddard. 1994. Breeding objectives for pasture based dairy production systems. *Livest. Prod. Sci.* 40:123–137. [https://doi.org/10.1016/0301-6226\(94\)90042-6](https://doi.org/10.1016/0301-6226(94)90042-6).
- von Bertalanffy, L. 1957. Quantitative laws in metabolism and growth. *Q. Rev. Biol.* 32:217–231. <https://doi.org/10.1086/401873>.
- Wang, X., E. P. Xing, and D. J. Schaid. 2015. Kernel methods for large-scale genomic data analysis. *Brief. Bioinform.* 16:183–192. <https://doi.org/10.1093/bib/bbu024>.
- Wu, M. C., S. Lee, T. Cai, Y. Li, M. Boehnke, and X. Lin. 2011. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet.* 89:82–93. <https://doi.org/10.1016/j.ajhg.2011.05.029>.
- Yin, T., and S. König. 2018. Genetic parameters for body weight from birth to calving and associations between weights with test-day, health, and female fertility traits. *J. Dairy Sci.* 101:2158–2170. <https://doi.org/10.3168/jds.2017-13835>.
- Zien, A., G. Ratsch, S. Mika, B. Scholkopf, T. Lengauer, and K. R. Muller. 2000. Engineering support vector machine kernels that recognize translation initiation sites. *Bioinformatics* 16:799–807. <https://doi.org/10.1093/bioinformatics/16.9.799>.

ORCIDS

S. König <https://orcid.org/0000-0002-1716-2511>