Visualizing Relationships between Conformation Traits over Time

M. Misztal[1], S. Tsuruta[1], I. Misztal[1], J. Graham[2], S. Comstock [2], J. Bewley[2], and D. A. L. Lourenco[1]

[1] *Department of Animal and Dairy Science, University of Georgia, Athens, GA 30602*

[2] *Holstein Association USA Inc., Brattleboro, VT 05302*

Corresponding author: Maciej@uga.edu

Abstract

Genetic evaluations often include many correlated traits, which are necessary in selection indices. The genetic parameters underlying their estimation have the potential to change at an accelerated pace under genomic selection. Standard methods for their evaluation are computationally more expensive with large numbers of animals over several years. This study aims to present a methodology to visualize relationships between conformation traits over time. Genetic Parameters via Predictivity is a method that allows this study to graph correlations among 18 linear conformation traits over yearly times slices from 2010 through 2020 using all available genomic, pedigree and phenotypic information For the initial time slice of 2010, the validation population included cows born in 2009 through 2011 while the reference population included cows born from 2000 through 2008. Each subsequent time slice increases the year by one. Using the most recent time slice, with a validation subset of cows born between 2019 through 2021, each trait was regressed on every combination of 2, 3, and 4 remaining traits, choosing the combination with the highest explained variance (R^2). The chosen combinations were then used to compute and graph the predictor traits for all time slice going back to 2010. No more than 4 predictor traits were needed for each target trait, as evidenced by the low 4^{th} eigenvalue in 2010 and 2020, where principal component analysis was used to graph biplots.

Key words: genetic correlation structures, genetic parameters via predictivity, genomic selection, selection index

Introduction

For use in breeding programs, 18 linear conformation traits (also known as type traits) are recorded by the Holstein Association USA Inc. (Brattleboro, VT). These are combined with functional and production traits to the Total Production Index (Holstein Association USA Inc., 2017). Weigel at al. (1998) used conformation traits for the prediction of productive life, defined as the length of time between first calving and culling. Higher reliability was obtained with predictions that used correlated traits, including conformation traits. Tsuruta et al. (2004) examined slow changes in genetic parameters when calculating genetic correlations over 14 years. Cole (2024) discussed the need for periodic recalculation of correlation and covariance matrices in population under selection. Misztal et al. (2021) discussed how such estimation may no longer be feasible with larger genomic datasets. A random regression model over multiple time slices would add computation costs (Tsuruta et al., 2021). Misztal and Gowane (2025) developed a method that addresses these constraints by calculating productivity within and across traits over time slices using Genetic Parameters via Predictivity. Given the high correlation among the traits, visualization of changes over time can be difficult and is addressed using the method in this paper.

Materials and Methods

Data

Tsuruta et al. (2024) used the method outlined in Misztal and Gowane (2025) to compute yearly genetic parameters over the 18-time traits used in this paper. The method combines the deterministic formula for predicting the accuracy of genomic breeding values (Daetwyler et al., 2008) combined with the within- and across-trait predictivity formula developed by Legarra et al. (2008). The correlation between traits i and j is

$$Corr_{ij} = \frac{Corr\left(\mathbf{y}_i - \mathbf{Xb}_i, \hat{\mathbf{u}}_j\right)}{h_i acc_j}$$

where $\mathbf{y}_i$ is the phenotypic vector for trait i, $\mathbf{Xb}_i$ corrects for fixed effects, $\hat{\mathbf{u}}_j$ is the genomic EBV (GEBV) for trait j, h_i is the square root of trait i heritability, and acc_j is the GEBV accuracy for trait j.

Validation and reference datasets were defined annually using a sliding window. For 2010, the validation set comprised cows born 2009–2011 and the reference set comprised cows born 2001–2008. The evaluation was advanced forward by one year through 2020, ending with a validation set born 2019–2021 and a reference set born 2001–2018 (Tsuruta et al., 2024). Genomic estimated breeding values were computed using adjusted phenotypes from validation and reference populations. The reference population ranged from 2,725,000 records for 367,000 genotyped animals born from 2001-2008 to 4,661,000 records from 625,000 genotyped animals for animals born from 2001 through 2018. The validation data were composed of 515,000 animals and 33,000 genotyped animals with a total of 685,000 records for cows born from 2009 through 2011 to 203,000 animals and 67,000 genotyped animals with a total of 191,000 records for cows born from 2019 through 2021.

Methods

Dimensionalities were calculated using the eigenvalues of correlation matrices of the 18 conformation traits in each time slice. Negative or small eigenvalues were set to 10^{-6}. Regressions were then run for each of the 18 target traits using every combination of 2, 3, 4 of the remaining 17 predictor traits.

Out of the possible 136, 680 and 2,380 combinations of 2, 3 and 4 predictor traits in 2020, the combinations yielding the highest R^2 were kept. These combinations were estimated for the previous 10 time slices for each target trait. These were graphed to illustrate the temporal stability of the traits over time. Biplots of the first two principal components were constructed for 2010 and 2020 to inspect geometric shifts in trait vectors. Dendrograms were created assigning distances of 0 to 2 for traits that were perfectly correlated to perfectly negatively correlated, respectively. All statistical analyses were executed in Julia using GLM.jl, MultivariateStats.jl, and Clustering.jl.

Results and Discussion

Rear udder width as a target trait in the case of three predictor traits is used as an example of applying the best fit. Rear udder height, udder depth and rear teat placement are the combination with the highest % variance explained (R^2) of 84. The next best combination keeps the first two traits but replaces rear teat placement with body depth. R^2 does not decrease significantly. The third best combination replaces body depth with rump width, with R^2 dropping to 0.83. Finally, the worst combination of 3 traits R^2 of 0.08. These traits are the least correlated with rear udder width and have the least exploratory power.

Table 1: Coefficient of regression and R^2 for rear udder width as a target trait for three best and the worst combination of remaining 17 traits

Model	1	2	3	680
Udder Height	0.91	0.82	0.83	
Udder Depth	-0.51	-0.39	-0.43	
R Teat Place	0.31			
Body Depth		0.30		
Rump Width			0.30	
Rump angle				0.01
Foot Angle				0.27
Teat length				0.06
R^2	0.84	0.84	0.83	0.08

The combination chosen would be the first combination (udder height, udder depth, and rear teat placement) which had the highest variance explained. This analysis was performed for 2020 and then these same three traits would be estimated for all 11 time slices and then graphed.

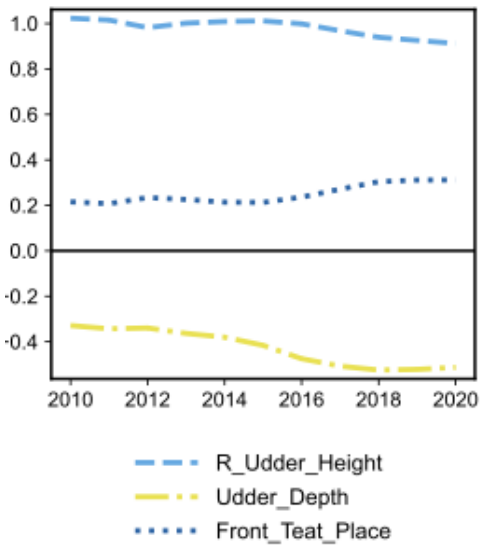


Figure 1. The coefficients for the best-fit model for rear udder width with optimal three predictor traits in 2020 over all 11 time slices

Eigenvalue analysis of the 2020 genetic correlation matrix yielded 18 values ranging from 6.92 down to -0.05. After setting non-positive eigenvalues to 10^{-6}, 8 principal components accounted for 90% of total variance, and 14 components explained 99%. This highlights substantial statistical redundancy across the 18 linear traits.

For best fit combinations of predictor traits, adding extra traits does not increase R^2 considerably (Table 1). Average change from 2 to 3 predictor traits was 0.04 for all traits and 0.03 from 3 to 4 predictor traits. Combined with low 4th eigenvalues, 4 predictor traits suggest possible overfitting and more predictor traits are unnecessary.

Table 2: Highest proportions of variance explained (%) for selected target traits and the corresponding optimal 2 ,3 and 4 trait models

	Proportion of variance (R^2)		
Target Trait	2-trait	3-trait	4-trait
Body Depth	96	97	97
Strength	95	96	97
Stature	92	95	95
Dairy Form	88	92	93
Udder Depth	81	89	95
R Udder Height	79	84	88
R Udder Width	75	84	88
Foot Angle	69	76	78
Rump Width	52	55	57
Teat Length	28	33	33
Rump Angle	9	17	23

Discussion

Highly correlated traits were better explained than poorly correlated traits. Body depth, strength, and stature all have explained variances of above 92 percent even with only two predictor traits. Their best predictor traits were frequently each other. They all measured body capacity traits and the most frequent predictor traits for all 18 target traits. Udder depth, rear udder height and rear udder depth were also highly correlated and were the best predictor traits for each other but were less frequently the optimal predictor traits for other traits.

Teat length and rump angle had the lowest variance explained and were the least correlated with other traits and the least likely to be optimal predictor traits for any other target traits. They measure unique and distinct

biological features. In the Trait Performance Index, dairy form, strength and stature are featured in more than one composite while rump angle is the one trait that is not featured in any of the composites.

Comparing all the traits with each other, the trait with the most significant change in relationships between 2010 and 2020 was dairy form. In 2010, it was clustered tightly with rear udder height and rear udder width. In 2020, dairy form was most more closely correlated with strength, body depth, and rump width. This may be related to breeding objectives, as dairy form is correlated negatively with fertility (Dechow et al., 2004), whereas strength is generally regarded as desirable by the industry.

Conclusions

This study provides a framework to convert correlation matrices derived by the Genetic Parameters via Predictivity method into visuals that present complex relationships between related traits over time. The best fit coefficient graphs, biplots, and dendrograms allow an intuitive interpretation for an otherwise overwhelming amount of subtle changes. The 2- and 3- trait models summarize linear conformation traits in Holsteins that exhibit high statistical redundancy. Rump angle and teat length are structurally independent, and desired selection pressure should be applied to them individually. This framework provides an objective approach to adjust to shifts in genetic parameters while including all genomic, pedigree and phenotypic information, which is crucial for trait selection in selection indices.

Acknowledgements

This research was primarily supported by grants from Holstein Association USA (Brattleboro, VT) and the U.S. Department of Agriculture's National Institute of Food and Agriculture (Agriculture and Food Research Initiative competitive grant 2015-67015-22936. This proceeding presents a summary of research accepted for publication in The Journal of Dairy Science

References

Cole, J.B. 2024. Are there too many traits in our selection indices? *J. Dairy Sci.* 108 (11), 12483–12505. https://doi.org/10.3168/jds.2024-25833

Daetwyler, H.D., Villanueva, B. and Woolliams, J.A. 2008. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS ONE*, 3 (10), e3395. https://doi.org/10.1371/journal.pone.0003395

Dechow, C.D., Rogers, G.W., Klei, L., Lawlor, T.J. and VanRaden, P.M. 2004. Body condition scores and dairy form evaluations as indicators of days open in US Holsteins. *J. Dairy Sci.* 87 (10), 3534–3541. https://doi.org/10.3168/jds.S0022-0302(04)73490-6

Holstein Association USA. 2017. Updates to the Total Performance Index (TPI) and type composites. Accessed Mar. 2, 2026. https://www.holsteinusa.com/pdf/Upcoming_Changes_aug17.pdf

Legarra, A., Robert-Granié, C., Manfredi, E. and Elsen, J.M. 2008. Performance of genomic selection in mice. *Genetics*, 180 (1), 611–618. https://doi.org/10.1534/genetics.108.088575

Misztal, I., Aguilar, I., Lourenco, D., Ma, L., Steibel, J. and Toro, M. 2021. Emerging issues in genomic selection. *J. Anim. Sci.* 99 (9), skab092. https://doi.org/10.1093/jas/skab092

Misztal, I. and Gowane, G. 2025. Estimation of heritabilities and genetic correlations by time slices using predictivity in large genomic models. *Genetics*, 230 (2), iyaf066. https://doi.org/10.1093/genetics/iyaf066

Tsuruta, S., Misztal, I. and Lawlor, T.J. 2004. Genetic correlations among production, body size, udder, and productive life traits over time in Holsteins. *J. Dairy Sci.* 87 (5), 1457–1468. https://doi.org/10.3168/jds.S0022-0302(04)73297-X

Tsuruta, S., Lawlor, T.J., Lourenco, D.A.L. and Misztal, I. 2021. Bias in genomic predictions by mating practices for linear type traits in a large-scale genomic evaluation. *J. Dairy Sci.* 104, 662–677. https://doi.org/10.3168/jds.2020-18668

Tsuruta, S., Lourenco, D., Lawlor, T. and Misztal, I. 2024. Over-time genetic correlations based on accuracy in GEBV for 18 linear conformation traits in US Holsteins. Page 785 in Book of Abstracts of the 75th Annual Meeting of the European Federation of Animal Science. Florence, Italy. https://docs.eaap.org/boa/2024_Florence_EAAP_Book_Abstracts.pdf

Weigel, K.A., Lawlor, T.J., VanRaden, P.M. and Wiggans, G.R. 1998. Use of linear type and production data to supplement early predicted transmitting abilities for productive life. *J. Dairy Sci.* 81 (7), 2040–2044. https://doi.org/10.3168/jds.S0022-0302(98)75778-9