

Application of a Canonical Transformation with Missing Values to Multitrait Analysis of U.S. Jersey Type Data

N. Gengler,¹ G.R. Wiggans,² J.R. Wright,² H.D. Norman,² and C.W. Wolfe³

¹Chargé de Recherches, Fonds National Belge de la Recherche Scientifique, Unité de Zootechnie, Faculté Universitaire des Sciences Agronomiques, B-5030 Gembloux, Belgium

²Animal Improvement Programs Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705-2350, USA

³American Jersey Cattle Association, Reynoldsburg, OH 43068-2362, USA

Introduction

The American Jersey Cattle Association (AJCA) currently scores 15 linear type traits (stature, strength, dairy form, foot angle, rear legs-side view, body depth, rump angle, thurl width, fore udder attachment, rear udder height, rear udder width, udder depth, udder cleft, front teat placement, and teat length) and assigns a calculated final score for all scored cattle. Breeders are also allowed to request that a final score be assigned by the appraiser for all cows being scored in their herd. Genetic evaluations for type traits have used single-trait sire models. Because the linear scoring system used by AJCA allows for repeated scoring of cattle, a repeatability model has been used. The current evaluation system uses all scores but gives a low weight (.1) to scores after second lactation because they are optional. Predicted transmitting abilities (PTA's) for cows are derived using results from this model.

Recent advances in genetic evaluations for type have included multitrait analysis and use of animal models. A large reduction in computing requirements can be achieved with canonical transformation, which transforms the initial correlated traits to uncorrelated canonical traits. Recent advances have been multiple diagonalization (generalization of canonical transformation to several random effects rather than only additive genetic effect) and development of an expectation-maximization algorithm that permits use of

this approach even if some traits are missing. Inbreeding coefficients now can be calculated and used in genetic evaluations for large populations. The objective of this study was to use these advanced methods to analyze U.S. Jersey type traits.

Data

Scores for type traits were provided by AJCA. Linear type scoring began in January 1, 1980, and only data since then were included. Final score assigned by the appraiser was used instead of calculated final score. At least one score during first lactation was required. After excluding scores after second lactation, 381,511 records for 225,632 cows remained. Records were divided into 14,812 herd-scoring date groups. Not all 16 traits were observed for every record. Scoring started in 1987 for body depth and in 1988 for teat length. Final score assigned by the appraiser became optional in 1987 and is missing in about 10-15% of records since then.

Type scores were adjusted for age and stage of lactation with multiplicative adjustment factors provided by the U.S. Department of Agriculture (USDA). Linear type traits are expressed on a scale between 1 and 50, and means for adjusted scores ranged from 22.0 for teat length to 30.9 for rear udder height. Standard deviations (SD's) also were similar among

traits (from 6.2 for teat length and thurl width to 7.9 for fore udder attachment).

Pedigrees were extracted from the USDA national data base of lactation records. After elimination of animals without a type score and not related to at least two other animals with type scores, 463,787 animals remained.

Model

The following model was applied for all traits:

$$y = Xh + Hc + Ss + Zp + Z^*u + e$$

where

- y** is a vector of type records
- h** is a vector of fixed herd-scoring date effects
- c** is a vector of fixed scoring year-parity-age effects
- s** is a vector of random herd×sire interaction effects
- p** is a vector of random permanent environmental effects
- u** is a vector of random additive genetic effects of animals and genetic groups ($u = a + Qg$, where **a** is a vector of random additive genetic effects of animals expressed as deviations from group means, **g** is a vector of fixed effects of genetic groups, and **Q** is an incidence matrix linking **g** with **u**)
- X**, **H**, **S**, **Z**, and **Z*** are incidence matrices associating **h**, **c**, **s**, **p**, and **u** with **y**; and
- e** is a vector of random residual effects.

The scoring year-parity-age effect (**c**) was included in the model because of concern in recent years on the accuracy of estimates of genetic trend related to problems with age adjustment and changes in maturity rate. Records were divided into 24 groups by parity (first or second lactation), age within parity (<27 mo, 27-31 mo, and >31 mo for first lactation; <39 mo, 39-43 mo, and >43 mo for second

lactation), and scoring year (<1983, 1983-87, 1988-92, and >1992). A herd×sire interaction effect was included to limit the effect of a single herd on a bull's evaluation; 70,281 herd×sire groups were created. To account for differences in average genetic merit of unknown ancestors by birth year, 18 genetic groups were defined based on birth year (<1961, 1961-62, ..., 1991-92, >1992).

(Co)variance components

(Co)variance components were obtained from a subset of the data with recent scores and no missing values by multiple diagonalization using the same multitrait model except that a herd×sire interaction was excluded. Of the estimated (co)variance matrix for permanent environment from that analysis, 40% was assigned to herd×sire interaction and the remainder to permanent environmental (co)variances, which generated a herd×sire variance somewhat less than the 10% of total variance used with the previous sire model. Heritabilities and relative permanent environmental and herd×sire variances used are in Table 1.

Computation of solutions and reliability

Solutions using a multitrait model and data with missing values were obtained with a canonical transformation adapted to include several random effects. Multiple diagonalization was used to simultaneously diagonalize the four (co)variance matrices. Using this transformation, the 16 correlated traits were transformed to 16 uncorrelated traits. Missing values were estimated during iteration on the data without back transformation. When updating missing values for a transformed trait, the newest available solutions for other traits were used. Within trait, second-order Jacobi

iteration was used as it does not need several sorted files as does Gauss-Seidel iteration. Inbreeding was introduced using the method for U.S. yield evaluations. If known, prior solutions were used as starting values to speed convergence. Squared relative change across the 16 transformed traits was the convergence criterion. Solutions on the original scale were obtained by back transformation.

To account for missing values when computing reliability, single-trait prediction error variances were computed for transformed traits using the reduction of information for missing values in original traits. Multitrait reliabilities were obtained through back transformation.

Genetic base

Solutions for additive genetic effects were expressed as PTA's (half the estimates for *u*) relative to the mean PTA for all cows born in 1990. This base (PTA90) was chosen to be similar to the base used for yield traits in the United States and several other countries.

Computing environment

An IBM (Armonk, NY) RISC System 6000, Model 560, workstation with 512 Mbytes of memory was used for computations. One round of iteration ($16 \times 774,554$ estimations) required approximately 23 min. The exact number of rounds to obtain convergence was not available because solutions from preliminary analysis were used as starting values. Theoretically, slightly over 50 rounds of iteration would have been needed to obtain a relative squared difference of approximately 10^{-8} . Using prior solutions reduced the number of rounds of iteration required by about 15. The low relative off-diagonals after diagonalization of 0.36% for herd \times sire interaction and permanent environment and 0.12% for genetic covariance matrices

indicated the success of the multiple diagonalization.

Scoring year-parity-age effect

Solutions for scoring year-parity-age effect were expressed as deviations from mean parity-age solutions within time period (Table 2). Range of solutions was >1.0 for final score, stature, dairy form, rump angle, fore udder attachment, udder depth, and front teat placement for scores after 1992; for rump angle, fore udder attachment, and front teat placement for 1988-92 scores; and for thurl width, udder depth, and teat length for 1983-87 scores. The most extreme ranges were observed for udder depth (5.4 for 1988-92 scores and 4.8 for scores after 1992). These differences reflect lack of fit of current age adjustments. Importance of this effect could be reduced somewhat by changing the herd-scoring date effect to be a herd-parity-scoring date effect.

PTA's

For 363,145 cows born since 1971, final score had a mean PTA of $-.93$ (SD of 1.19) and a mean reliability of .46 (SD of .15). For linear type traits, mean PTA's ranged from -1.37 for dairy form to .37 for udder depth (SD from 0.63 for foot angle to 1.49 for dairy form); mean reliabilities ranged from .36 for rear legs to .51 for stature (SD from .11 for rear legs to .17 for stature). For 11,982 bulls born since 1971, mean PTA and reliability and their SD's were similar for final score to those for cows. For linear type traits, bull mean PTA's ranged from -1.65 for dairy form to .68 for udder depth; mean reliabilities ranged from .39 for rear legs to .48 for stature. For bulls, SD's for both PTA and reliability were somewhat lower than those for cows.

Correlations between PTA with and without inbreeding considered were high ($>.99$). However, when inbreeding was

considered, some PTA's changed by up to .2, which caused some rerankings.

Comparison of multitrait animal and sire models

Correlations of multitrait animal model and sire model PTA's were calculated by birth year (Table 3). Correlations increased over time and were higher for bulls with >40 progeny than for cows born during the same years. Traits with missing values generally had lower correlations than did other traits; correlations for final score decreased slightly for bulls born after 1984, and correlations for body depth and teat length were lower for bulls born before 1980. Differences between solutions from multitrait animal and sire models could have been caused by differences in 1) data included, 2) use of information from relatives, 3) effect of one trait on correlated traits, 4) genetic parameters, 5) use of unknown-parent groups, or 6) consideration of inbreeding.

Genetic trend

Genetic trends were computed using estimated breeding values ($EBV=2 \times PTA$)

for 302,219 cows born between 1977 and 1993. Quadratic regression equations were fit for mean EBV by birth year. If the quadratic effect was nonsignificant, a linear equation was used. Difference between mean EBV's of cows born in 1992 and 1993 indicates recent genetic trend (Table 4). Genetic trend was greater for dairy form than for final score or rear udder traits. Almost no trend was evident for feet and leg traits and fore udder attachment; trend for udder depth was negative. The regression equations explained over 60% of total variance except for traits with no trend.

Conclusions

Type evaluations can be computed more accurately and at a reasonable cost using a multitrait animal model with canonical transformation and multiple diagonalization to account for missing values. This methodology should give U.S. Jersey breeders the advanced tools they need to make more accurate genetic selections for final score and linear type traits.

Table 1. Estimated heritabilities and assumed variances for permanent environment and herd×sire interaction relative to total variance for Jersey type traits

Trait	Heritability	Relative variance	
		Permanent environment	Herd×sire interaction
Final score	.293	.112	.075
Stature	.399	.104	.070
Strength	.264	.076	.051
Dairy form	.277	.089	.060
Foot angle	.128	.073	.049
Rear legs (side view)	.133	.070	.047
Body depth	.273	.081	.054
Rump angle	.314	.121	.080
Thurl width	.223	.065	.044
Fore udder attachment	.218	.083	.055
Rear udder height	.275	.110	.073
Rear udder width	.260	.093	.062
Udder depth	.323	.093	.062
Udder cleft	.199	.098	.065
Front teat placement	.286	.106	.071
Teat length	.306	.106	.071

Table 2. Scoring year-parity-age solutions for Jersey type traits expressed as deviations from mean parity-age solutions within time period

Scoring year	Parity	Age (mo)	Trait ¹															
			FS	ST	SR	DF	FA	RL	BD	RA	TW	FU	RH	RW	UD	UC	TP	TL
<1983	1	<27	-.4	-.2	-.3	-.1	.1	.0	-.1	.0	.0	.0	.0	-.2	.3	-.1	.0	.1
		27-31	.0	.1	-.1	.0	.1	.0	.1	.0	.3	.0	.2	.1	.2	.0	.0	.0
		>31	.1	.2	.2	.1	.1	.0	.1	.0	.2	.2	.0	.1	.2	.0	.0	.1
	2	<39	.3	-.3	-.2	.2	-.1	.0	.0	.0	-.1	-.1	.0	.3	-.4	.2	-.2	-.1
		39-43	.1	.2	.0	.0	.0	.0	.0	.0	-.1	-.2	.0	.0	-.1	.0	.1	.0
		>43	.1	.1	.3	-.1	-.1	-.1	-.1	.1	-.4	-.1	.0	-.2	-.1	-.1	.0	.0
1983-87	1	<27	-.3	-.1	.3	-.2	.1	-.4	-.3	.1	-.6	.8	.1	-.2	1.2	-.2	-.3	-2.0
		27-31	-.4	.0	-.2	-.4	.1	.1	-.3	.0	-.5	.6	-.1	-.3	.5	-.1	.1	-1.1
		>31	-.3	.1	.0	-.4	.2	.1	-.1	-.1	-.1	.5	-.1	-.5	.2	.0	.3	-.5
	2	<39	.4	.1	.0	.2	-.1	-.1	.2	.0	.3	-.4	.2	.3	-.4	.2	-.1	1.1
		39-43	.2	-.1	-.3	.3	-.1	.1	.1	.0	.3	-.8	.1	.3	-.8	.0	-.1	1.4
		>43	.3	-.1	.0	.4	.0	.1	.3	-.2	.7	-.7	.0	.2	-.8	.0	.0	1.2
1988-92	1	<27	.3	-.2	-.4	.2	-.2	-.1	-.3	.6	-.4	-.2	.0	-.1	-.5	-.3	-.3	.0
		27-31	-.1	.1	-.1	-.1	.0	.0	-.1	.2	-.1	-.1	-.1	.0	-.5	.0	.0	-.1
		>31	-.1	-.1	.0	.0	.1	.0	.0	.7	.0	-.3	.0	.1	-1.7	-.2	-.3	-.2
	2	<39	-.2	.4	.2	-.4	-.1	.0	.2	-2.0	.2	.9	-.2	-.2	3.7	.6	.9	.3
		39-43	-.1	.0	.2	.0	.1	.1	.1	-.3	.2	.2	.0	.0	.3	.1	.2	.0
		>43	.1	-.3	-.1	.3	.0	.0	-.1	.6	.1	-.5	.2	.3	-1.4	-.1	-.6	-.1
>1992	1	<27	.8	-.4	-.4	.6	.5	-.4	-.2	.7	-.3	.3	-.2	.0	.6	-.3	-.2	-.4
		27-31	.4	-.1	-.2	.5	.2	-.1	-.1	.3	.0	.0	.0	.2	.0	.0	.0	-.2
		>31	.1	-.2	-.2	.3	.0	.0	-.4	.8	-.1	-.5	.2	.3	-1.2	-.2	-.4	-.3
	2	<39	-.6	.7	.5	-1.0	-.4	.0	.6	-2.0	.4	1.0	-.5	-.5	2.9	.5	1.0	.5
		39-43	-.3	.3	.2	-.4	-.2	.2	.1	-.3	.2	.0	.1	-.1	-.2	.1	.2	.1
		>43	-.4	-.1	.1	.1	-.3	.2	.0	.6	.0	-1.0	.5	.3	-1.9	.0	-.5	.2

¹FS = final score;

FA = foot angle;

TW = thurl width;

UD = udder depth;

TL = teat length.

ST = stature;

RL = rear legs (side view);

FU = fore udder attachment;

UC = udder cleft;

SR = strength;

BD = body depth;

RH = rear udder height;

TP = front teat placement

DF = dairy form;

RA = rump angle;

RW = rear udder width;

and

Table 3. Correlations between predicted transmitting abilities from multitrait animal and sire models for Jersey type traits by birth year of animal

Birth year	Number of animals	Trait ¹															
		FS	ST	SR	DF	FA	RL	BD	RA	TW	FU	RH	RW	UD	UC	TP	TL
Cows																	
1975-79	30,739	.56	.67	.62	.59	.65	.6569	.56	.64	.61	.58	.61	.66	.63	...
1980-84	70,923	.76	.86	.81	.79	.82	.8683	.77	.77	.80	.78	.71	.84	.85	...
>1984	147,856	.88	.92	.87	.90	.88	.8991	.84	.83	.91	.90	.83	.88	.91	...
Bulls with >40 progeny																	
1970-74	181	.67	.83	.79	.74	.73	.82	.50	.85	.76	.76	.74	.71	.79	.81	.81	.78
1975-79	188	.82	.90	.83	.83	.81	.81	.63	.91	.84	.86	.86	.84	.83	.87	.87	.80
1980-84	231	.91	.93	.91	.93	.85	.88	.82	.93	.91	.91	.91	.89	.89	.88	.92	.77
>1984	192	.88	.95	.91	.90	.88	.90	.92	.93	.90	.88	.91	.90	.86	.89	.92	.75

¹FS = final score;
FA = foot angle;
TW = thurl width;
UD = udder depth;
TL = teat length.

ST = stature;
RL = rear legs (side view);
FU = fore udder attachment;
UC = udder cleft;

SR = strength;
BD = body depth;
RH = rear udder height;
TP = front teat placement

DF = dairy form;
RA = rump angle;
RW = rear udder width;
and

Table 4. Differences between estimated breeding values (EBV's) of cows born in 1992 and 1993 (trend) from linear or quadratic fit of mean EBV by birth year for 302,219 Jersey cows born from 1977 to 1993

Trait	Trend	Equation	R ²
Final score	.55	Quadratic	99.8
Stature	.18	Quadratic	67.8
Strength	.10	Quadratic	59.7
Dairy form	.77	Quadratic	99.7
Foot angle	.06	Quadratic	75.3
Rear legs (side view)	.02	Quadratic	96.5
Body depth	.21	Quadratic	85.7
Rump angle	.16	Quadratic	74.4
Thurl width	.17	Quadratic	93.7
Fore udder attachment	.05	Quadratic	36.1
Rear udder height	.50	Quadratic	99.1
Rear udder width	.52	Quadratic	99.6
Udder depth	-.13	Linear	99.3
Udder cleft	.12	Quadratic	97.3
Front teat placement	.08	Linear	85.0
Teat length	-.02	Linear ¹	12.9

¹Nonsignificant, $P=.09$.