Effect of Cell Size in Test Day Models

E. Grünhagen and H.H. Swalve
Institut für Tierzucht und Haustiergenetik der Universität Göttingen,
Albrecht-Thaer-Weg 3, 37075 Göttingen, Germany

Abstract

Connectedness indices γ (Foulley et al., 1990) and reliabilities of EBVs were estimated for three data sets differing in minimum size of herd test day (HTD). Each data set contained test day (TD) records of 3712 cows from northern Germany with at least 8 subsequent TD records during the first lactation. Under a test day model as suggested by Ptak and Schaeffer (1993) mean connectedness indices γ as well as mean reliabilities of EBVs were almost at the same level across data sets. The high value of mean γ_{HTD}, ~.97, indicates a small contribution of the cell size of HTD to the connectedness of the data set. If there are HTDs with only one observation, the degree of connectedness slightly decreases due to a larger number of γ_{HTD} lower than .95. The effect of age*season at calving (AS) adds only a small amount to connectedness since the value of mean γ_{AS} is almost 1.

Introduction

When using test day models proposed by Ptak and Schaeffer (1993) for genetic evaluation in dairy cattle, cows are compared within herd test day (HTD) and age*season of calving. To improve estimation of breeding values, Ptak and Schaeffer (1993) even suggested to use management groups within a herd as contemporary groups. If cell size becomes smaller there might arise a problem in comparing EBVs. In disconnected data sets some EBVs are not directly comparable leading to mistakes in selection decisions although complete disconnection between random effects can never occur in mixed models. All contrasts between random effects remain estimable with the prior information of the expected mean and variance of those contrasts (Foulley et al., 1984; cit. Boichard et al., 1996).

There are different approaches to measure the connectedness of a data set with respect to a given model. Kennedy and Trus (1993) and Boichard et al. (1996) give a review of measures of connectedness. Kennedy and Trus (1993) favour the average prediction error variance (PEV) of differences in EBVs between animals in different management units since increased connectedness reduces the PEV of comparisons of animals in different management units. The ratio of PEV of contrasts of EBVs with and without management groups (Foulley et al., 1990) provides a qualitative measure of including an additional factor in a model. These two measurements of connectedness are highly correlated (Kennedy and Trus, 1993). Laloë et al. (1996) point out that the method of Foulley et al. (1990) does not detect complete disconnectedness in some situations. Therefore they prefer the generalized coefficient of determination (Laloë, 1993).

The aim of the present study was to analyse whether different sizes of HTD affect the comparability of EBVs.

Material and methods

Data was supplied by VIT, Verden, and covered a region from northern Germany, where herdsizes are representative for the western part of the country. All calvings were from years 1990 to 1995. Table 1 displays the structure of the full and reduced data sets.

Three data sets with at least 1, 4 and 10 cows in the first lactation within HTD were created. Each

cow had to have at least 8 subsequent TD records during this lactation. The reduced data sets were created by eliminating herds and animals randomly from the full data sets to obtain the same number of cows with records in each data set, since the degree of connectedness due to random effects (additive-genetic and permanent enviromental effects) largely depends on the number of animals with records. For the data sets with at least 1 or 4 cows within HTD two different samples were drawn. Table 2 shows that the distribution of cell size across HTD effects for the two samples A and B was similar.

Table 1. Structure of the full and reduced data sets of cows with at least 8 subsequent test day records during the first lactation.

at least ... cows within HTD	1			4			10
	full	reduced		full	reduced		full
		A	B		A	B	
Cows with records, no.	64105	3712	3712	38590	3712	3712	3712
TD records, no.	684948	39691	39592	409213	39597	39539	38675
Herds, no.	2041	106	116	1124	97	97	95
HTD, no.	103133	5764	6144	39635	3752	3677	2157
Cows per HTD, mean no.	6.6	6.9	6.4	10.3	10.6	10.8	17.9
Maximum cows per HTD, no.	46	38	36	45	45	41	44
TD records per cow, mean no.	10.7	10.7	10.7	10.6	10.6	10.7	10.4
Animals in pedigreefile, no.		10306	10354		11033	10759	12136

Table 2. Distribution of cell size of HTD effects in the data sets of 3712 cows with records.

at least ... cows within HTD	1		4		10
	A	B	A	B	
cell size=1	11.4 %	11.2 %			
2 # cell size < 4	18.7 %	20.7 %			
4 # cell size < 10	43.2 %	46.1 %	52.2 %	49.6 %	
10 # cell size < 20	24.3 %	20.4 %	40.6 %	43.4 %	65.3 %
cell size $ 20	2.4 %	1.6 %	7.2 %	7.0 %	34.7 %

In this study connectedness was measured by the ratio of PEV of contrasts of EBVs with and without management groups known as connectedness index γ (Foulley et al., 1990) since this criteria provides a measure of including a certain factor in a model:

$$\gamma_{\triangle}(x) = \frac{x' C_R x}{x' C_F x}$$

with C_R and C_F denoting the submatrix of the inverse of the coefficient matrix pertaining to the

animal equations with (C_F) and without (C_R) effect of Δ (HTD or AS) and x indicating a vector of a contrast. The linear model for the analysis of the data was

$$y_{ijki} = HTD_i + AS_j + \sum_{m=1}^{4} b_{jm} X_{jm} + a_k + pe_k + e_{ijkl}$$

where

y_{ijkl} = TD record, milk yield
HTD_i = effect of HTD, fixed
AS_j = effect of age*season of calving, fixed, 15 levels
X_{jm} = fixed covariables nested within AS, DIM=days in milk
X_{j1}=DIM/305, X_{j2}=$(DIM/305)^2$, X_{j3}=ln(305/DIM), X_{j4}=$(ln(305/DIM))^2$
a_k = animal additive genetic effect, random
pe_k = effect of permanent environment of the cow during lactation, random
e_{ijkl} = residual effect, random

The mixed model equations were set up with a modified version of MTDFREML (Boldman et al., 1993). Genetic parameters were taken from the study of Swalve (1995). The submatrices of the inverted coefficient matrices were calculated using FSPAK-Routines (Misztal and Perez-Enciso, 1993).

Results and discussion

Reliability

In Table 3 raw means and standard deviations (SD) for reliabilities of EBVs of cows with records are given along with the corresponding extrema. The mean reliability seems to be almost independent from the smallest possible cell size of HTD. The SD of reliability tends to be larger, the minimum reliability tends to be smaller in the data set with at least one cow within HTD than those in the other data sets.

Table 3. Raw means, standard deviations and extrema of reliabilities of EBVs of cows with records (data sets of 3712 cows with records).

at least ... cows within HTD	1		4		10
	A	B	A	B	
Mean reliability	.45	.45	.47	.46	.46
SD of reliability	.054	.053	.045	.044	.048
Minimum reliability	.13	.01	.25	.21	.27
Maximum reliability	.57	.56	.56	.56	.55

Connectedness

Table 4 and 6 displays the descriptive statistics of the connectedness indices γ_Δ (Foulley et al., 1990). The value of γ_Δ can vary between 0 and 1. A value of 1 indicates that the effect Δ does not affect the accuracy predicting the contrast between two EBVs. If γ_Δ is 0, only prior information is used to predict the contrast of EBVs. Mean γ_Δ describes the degree of connectedness among animals due to the influence of the effects included in the reduced model calculating γ_Δ.

Influence of HTD

Only slight differences in mean γ_{HTD} appear between the three different data sets. The mean value of γ_{HTD} (~.97) indicates a small contribution of cell size of HTD to the connectedness of the data set. Other factors of influence (AS, relationships between animals and the effect of permanent environment) lead to a satisfying level of connectedness.

Table 4. Raw means, standard deviations and extrema of connectedness indices γ_{HTD} (Foulley et al., 1990) for the data sets of 3712 cows with records.

at least ... cows within HTD	1		4		10
	A	B	A	B	
no. of estimable γ_{HTD}	53=101=665	53=597=481	60=858=028	57=872=661	73=635=180
Mean γ_{HTD}	.964	.963	.976	.969	.973
SD of γ_{HTD}	.038	.035	.026	.031	.033
Minimum γ_{HTD}	.233	.369	.375	.304	.299
Maximum γ_{HTD}	1.00	1.00	1.00	1.00	1.00

From the distribution of the connectedness indices γ_{HTD} given in Table 5 it may be concluded, that the percentage of $\gamma_{HTD} < .95$ in the data set with at least one cow within HTD tends to be larger than those in the other data sets. Therefore mean γ_{HTD} is slightly lower and SD of γ_{HTD} is slightly larger than those in the data sets with more than one cow in each HTD.

Table 5. Distribution of connectedness indices γ_{HTD} (Foulley et al., 1990) in the data sets with 3712 recorded cows.

at least ... cows within HTD	1		4		10
	A	B	A	B	
no. of estimable γ_{HTD}	53=101=665	53=597=481	60=858=028	57=872=661	73=635=180
$\gamma_{HTD} < .80$	.82 %	.53 %	.25 %	.29 %	.38 %
$.80 \# \gamma_{HTD} < .95$	19.21 %	21.50 %	8.15 %	15.87 %	13.04 %
γ_{HTD} $ .95	79.97 %	77.97 %	91.60 %	84.84 %	86.58 %

Influence of AS

Since the covariates describing the lactation curve of animals are nested within AS the reduced model used to calculate γ_{AS} contains neither AS effects nor the covariates. As Table 6 shows there seems to be only a very small increase in connectedness if AS effects are added to the model. In each data set mean γ_{AS} is almost 1.

Table 6. Raw means, standard deviations and extrema of connectedness indices γ_{AS} (Foulley et al., 1990) for the data sets of 3712 cows with records.

at least ... cows within HTD	1		4		10
	A	B	A	B	
no. of estimable γ_{AS}	53=101=665	53=597=481	60=858=028	57=872=661	73=635=180
Mean γ_{AS}	.999	.999	.999	.999	.999
SD of γ_{AS}	.002	.002	.002	.002	.002
Minimum γ_{AS}	.901	.912	.909	.948	.912
Maximum γ_{AS}	1.00	1.00	1.00	1.00	1.00

Conclusions

The effect of age*season at calving adds nearly nothing to the degree of connectedness of the data set. Although Swalve (1995) pointed out that the size of contemporary groups in countries with smaller herds is especially crucial, this study shows that the comparibility of EBVs seems to be only slightly affected by a smaller size of HTD. It appears to be not necessary to combine HTD effects to increase cell size.

References

Boichard, D., Foulley, J.L. and Hanocq, E. 1996. Measures of connectedness in animal breeding, Gen.-stat. Ausschuß, 19.-21.3.96, Werder.

Boldman, K.G., Kriese, L.A., Van Vleck, L.D. and Kachman, S.D. 1993. A Manual for use of MTDFREML, USDA-ARS, Lincoln, Nebraska.

Foulley, J.L., Bouix, J., Goffinet, B. and Elsen, J.M. 1984. In `Insémination artificielle et amélioration génétique: bilan et perspectives critiquesA p. 131-176, Editors J.M. Elsen und J.M. Foulley, Les colloques de l'INRA 29.

Foulley, J.L., Bouix, J., Goffinet, B. and Elsen, J.M. 1990. In `Advances in statistical methods for genetic improvement of livestockA p. 277-308, Editors D. Gianola and K. Hammond, Springer Verlag, Heidelberg.

Kennedy, B.W. and Trus, D. 1993. *J. Anim. Sci. 71,* 2341-2352.

Misztal, I. and Perez-Enciso, M. 1993. *J. Dairy Sci. 76,* 1479-1483.

Laloë, D. 1993. *Genet. Sel. Evol. 25,* 557-576.

Laloë, D., Phocas, F. and Ménissier, F. 1996. *Genet. Sel. Evol. 28,* 359-378.

Ptak, E. and Schaeffer, L.R. 1993. *Livest. Prod. Sci. 34,* 23-34.

Swalve, H.H. 1995. *J. Dairy Sci. 78,* 929-938.