ESTIMATION OF LACTATION YIELDS FROM TEST DAY YIELDS OF MILK, FAT AND PROTEIN FOR PORTUGUESE HOLSTEIN CATTLE USING AN AUTOREGRESSIVE TEST DAY MODEL

J. Vasconcelos[1], A. Martins[2], M. Petim-Batista[3], A. Silvestre[3], A. Ferreira[2], J. Colaço[3] and J. Carvalheira[1]

[1]Instituto de Ciências Biomédicas Abel Salazar and Centro de Estudos de Ciência Animal, Universidade do Porto, Rua do Monte-Crasto, 4485-661 Vairão, Portugal
[2]Associação para o apoio à Bovinicultura Leiteira do Norte, Portugal
[3]Universidade de Trás-os-Montes e Alto Douro, Portugal

INTRODUCTION

The Portuguese dairy cattle associations estimate cumulative lactation yields for milk and milk components by the Test Interval Method (TIM, Sargent, 1968). The Portuguese National milk data set is relatively small with an average of about 90 000 new lactations (L) records added per year. However, National genetic evaluations use only yields from terminated L with 210 or more days-in-milk (DIM), extended to 305-d if necessary, using pre-calculated factors (Gama *et al.*, 2000). All L with less than 210-d are excluded from the genetic analysis, representing each year 11 % to 14 % of all terminated L (ANABLE, 1999). On top of this number, it is also necessary to add all L still in progress at the moment of data retrieval for analysis. At the end of the year 2000, for example, there were in the first three L, 61,000 cows in milk (30 % of all terminated L in that year) that were also excluded from any analysis. In a relatively small data set as this, any information that may be added to the evaluation would be an important contribution to improve the reliability of the predicted breeding values.

Test day (TD) models are becoming the standard for genetic evaluations of production traits in dairy cattle. This methodology may also be applied to estimate 305-d cumulative yields of milk, fat and protein using the predicted daily yields from the model. This method is being applied in Canada (Schaffer and Jamrozik, 1996) using the Wood equation (Wood, 1967) in a multiple trait approach incorporating the covariances between the three traits. In the US, VanRaden (1997) obtained more accurate 305-d yields using a best prediction TD model to estimate these traits. More recently, Carvalheira *et al.* (1998 ; 2001) indicated that autoregressive TD repeatability models (ARTD) could estimate accurate L curves (on a herd basis) and predict all TD residual yields from all cows present in the analysis. Given the acknowledged potential of this modeling approach to reduce the size of residual variance components, ARTD models may be a viable alternative to derive 305-d L yields from observed TD yields. In this manner, the estimated 305-d L records may reflect the general advantages associated with TD models by better accounting for environmental effects.

The objective of this work was to develop and test an ARTD model to predict missing TD for every cow with records. The cow effect would be partitioned in two independent variables : the first measuring correlations between L, and the second accounting for correlations between TD within L. The covariance matrices for these variables, will assume first-order autoregressive processes. Cumulative 305-d L may then be estimated, summing the appropriate means to the predicted TD residuals.

MATERIAL AND METHODS

Data. The Portuguese National Association of Dairy Cattle Breeders (ANABLE) provided the data for this study, with a total of 6 694 962 TD records, produced between June 1994 and December 2000. The editing process conformed the data for the analysis following the assumptions of the ARTD model. Only records from a.m.-p.m. supervised test plans belonging to the first three L were used. Records missing relevant information, outliers, or with less than 5 or more than 305 DIM were deleted. Criteria on minimum and maximum age limits within L were adopted from ANABLE. The sequence between TD within L for each cow was checked and those with intervals smaller than 15-d or greater than 75-d were not used. Cows with only one record per L were also deleted. Table 1 gives the structure of the data set for milk, fat and protein submitted to the analysis.

Table 1. Number of test day milk, fat and protein yields per lactation (L) after editings (data is from June 1994 to December 2000)

Data	Milk	Fat	Protein
Original	6 694 962	6 694 962	6 694 962
Final	2 101 098	2 093 524	2 103 506
1st L	915 338	913 001	916 500
2nd L	695 628	692 260	696 289
3rd L	490 132	488 263	490 717

The final data consisted of records produced by 141 681, 141 777 and 141 730 cows for milk, fat and protein, respectively.

The model. The ARTD model takes into account environmental effects that are specific to each test date, age at calving and lactation curve within each herd. This partition of the fixed effects permits the estimation of specific means for each herd, increasing the accuracy of the predicted TD. The model used in the present study was a "cow model" assuming no genetic relationships between animals. The ARTD model for milk, fat and protein yields, was as follows :

$$\mathbf{y}_{\mathbf{ijkLmn}} = \mathbf{HTD}_{\mathbf{i}} + \mathbf{Age(H)}_{\mathbf{j}} + \mathbf{DIM(H)}_{\mathbf{k(L)}} + \mathbf{p}_{\mathbf{m(L)}} + \mathbf{t}_{\mathbf{n(mL)}} + \mathbf{e}_{\mathbf{ijkLmn}},$$

Where, **y** is the TD yield, **HTD** is the fixed effect of herd-test-date, **Age(H)** is the fixed effect of age at calving nested within herd, **DIM(H)** is the fixed effect of DIM nested with herd and L, **p** is the random effect of the long term environmental effects accounting for the correlations generated by the cow across L, **t** is the random effect of the short term environmental effects accounting for the correlations due to the cow between TD and within each L, and **e** is the random residual effect. Both, **p** and **t** are fitted with first-order autocorrelation structures.

A detailed description of the model expectations and (co)variance structure for this type of models, may be found in Carvalheira *et al.*, (2001). To help ensure accuracy, at least 3 observations in each contemporary group were required. Farms (herds) failing this requirement were grouped based in cluster analysis using the method of Ward (SAS, 1989), where phenotypic mean and standard deviation in first L defined similarity between herds. It was assumed that farms with similar phenotypic mean and standard deviation had similar management. Analysis were for single trait and (co)variance components were estimated by DFREML-SIMPLEX method (Nelder *et al.*, 1965). Cumulative 305-d L yields were then estimated summing the appropriate within herd means to the predicted TD residuals for each cow.

RESULTS AND DISCUSSION

Presently, only terminated 305-d L yields with more than 210 DIM (approximately the number of records with > 210-d in table 2) are used for the National genetic evaluations. ARTD procedures allowed a substantial increase in the total number of 305-d L records available for those analysis (records with < 210-d in table 2). These records (either L in progress or terminated) were projected to 305-d L yields corresponding to a recuperation of 20.8 %, 20.5% and 20.2 % of the data for milk, fat and protein, respectively, for the first 3 L. This data set now contains 245 466, 254 474 and 254 610 of ARTD 305-d L yields of milk, fat and protein, respectively, and is available for further analysis. In total, the National data set for the first 3 L, was incremented by 51 106, 52 274 and 51 369 records for milk, fat and protein, respectively, which, for a relatively small database, may be a significant improvement.

Table 2. Number of estimated 305-d lactation yields derived from test day yields of milk, fat and protein from the Portuguese Holstein database (from June 1994 to December 2000) using an autoregressive test day model

Lactation		Milk	Fat	Protein
	< 210-d	19 688	19 960	19 577
1st	> 210-d	89 022	88 817	89 194
	Total	108 710	108 777	108 771
	< 210-d	18 113	18 348	18 022
2nd	> 210-d	66 948	66 652	67 070
	Total	85 061	85 000	85 092
	< 210-d	13 305	13 966	13 770
3rd	> 210-d	46 890	46 736	46 977
	Total	60 695	60 702	60 747

Milk, fat and protein phenotypic means and standard deviations for the first 3 ARTD L are in table 3. These means were slightly greater but not significantly different from those estimated by TIM. Trends of ARTD 305-d yields were positive for all traits and L for the period studied (109,4 kg/yr, 115,3 kg/yr e 105,9 kg/yr for milk, 4,1 kg/yr, 4,3 kg/yr e 4,0 kg/yr for fat and 4,3 kg/yr, 4,5 kg/yr, 4,2 kg/yr for protein, for 1st, 2nd and 3rd L, respectively).

Table 3. Phenotypic means ($\bar{x}$) and standard deviations (SD) for yields of milk, fat and protein (kg) for the first 3 lactations (L) estimated at 305-d by the autoregressive test day model

L	Milk		Fat		Protein	
	$\bar{x}$	SD	$\bar{x}$	SD	$\bar{x}$	SD
1st	6 946.5	1 472.6	251.6	57.0	220.3	46.9
2nd	7 563.5	1 749.2	272.9	66.2	241.3	54.3
3rd	7 667.8	1 741.4	277.0	66.5	242.0	53.9

Correlations between ARTD 305-d L yields and corresponding TIM records (only for those L having 210 or more DIM) are in table 4. The high correlations obtained in this study suggest that the ARTD model may be a viable alternative to estimate 305-d L yields, although more studies are necessary to determine differences in accuracy between the two methods.

Table 4. Correlations between the 305-d lactation (L) yields estimated by the autoregressive test day model and by the test interval method (only for lactations of length 210-d or more) for the first 3 lactations of Portuguese Holstein cattle

Trait	1st L	2nd L	3rd L
Milk	0.99	0.98	0.98
Fat	0.98	0.98	0.97
Protein	0.99	0.98	0.98

REFERENCES

ANABLE (1999) « Publicação de Resultados »*,* Editors Verde Milho, Aveiro, Portugal.

Carvalheira J., Blake, R.W., Pollak, E.J., Quaas, R.L. and Duran-Castro, C.V. (1998) *J. Dairy Sci.* **81** : 2738-2751.

Carvalheira, J., Pollak, E.J., Quaas, R.L. and Blake, R.W. (2001) *Proc. 52nd EAAP* : 6.

Gama, L.T., Carolino, I. and Carolino, N. (2000) *Proc. Congresso de Zootecnia* : 104.

Nelder, J.A., and Mead, R. (1965) *Computer J.* **7** : 308-313.

Sargent, F.D., Lytton, V.H. and Wall, O.J. Jr. (1968) *J. Dairy Sci.* **51** : 170.

SAS Institute Inc., (1989) SAS/STAT User's Guide, Version 6, Fourth Edition, Volume 1, Cary, NC., 943 pp.

Schaeffer, L.R. and Jamrozik, J. (1996) *J. Dairy Sci.* **79** : 2044-2055.

Vanraden, P.M. (1997) *J. Dairy Sci.* **80** : 3015-3022.

Wood, P.D.P. (1967) *Nature* **216** : 164-165.