

## Legendre Polynomials versus Linear Splines in the Canadian Test-Day Model

J. Bohmanova, J. Jamrozik, F. Miglior,  
I. Misztal, P.G. Sullivan



## Canadian Test-day Model

- 12 traits  $\Rightarrow$ 
    - 3 lactations
    - 4 traits (milk, fat, protein yield and SCS)
  - Random and fixed regression – Legendre polynomial of order four
  - Only test-days with DIM  $\leq 305$  considered
  - 40% of cows have lactation length  $\geq 305$  DIM
- ↓
- interest to use TD records up to 365 DIM

## Objectives

- Identify the best function for modeling random and fixed regression using test day records measured between 5 and 365 DIM
- Compare two types of functions :
  - Legendre polynomials
  - Linear splines (4, 5 and 6 knots)

## Why Splines?

- Limited sensitivity to the data
  - Records influence only parts of functions in their closeness
- Direct interpretation of variance components
- Easy implementation

## Data for Estimation of Variance Components

- Random sample of 50 herds from the Canadian Holstein database
- 6,094 cows
- 96,756 TD records
- TD from 5 to 365 DIM
- 4 traits: milk, fat, protein, SCS
- 3 lactations
- Pedigree file with 18,178 animals

## Models

$$y_{ijkmt} = HTD_{jkn} + \sum_{l=1}^q \alpha_{jlmn} z_l(d) + \sum_{l=1}^q \beta_{ijln} z_l(d) + \sum_{l=1}^q \gamma_{ijln} z_l(d) + e_{ijkmt}$$

- HTD = herd-test-date effect
- $\alpha$  = fixed regression on DIM x age-season-region of calving class
- $\beta$  = random regressions on DIM for additive genetic effect
- $\gamma$  = random regressions on DIM for permanent environmental effect



## Models

Model	Function	Number of coefficients/knots	Position of knots
LEG	Legendre	5	-
SPL4	Splines	4	[5 65 245 365]
SPL5	Splines	5	[5 65 125 245 365]
SPL6	Splines	6	[5 65 125 245 305 365]

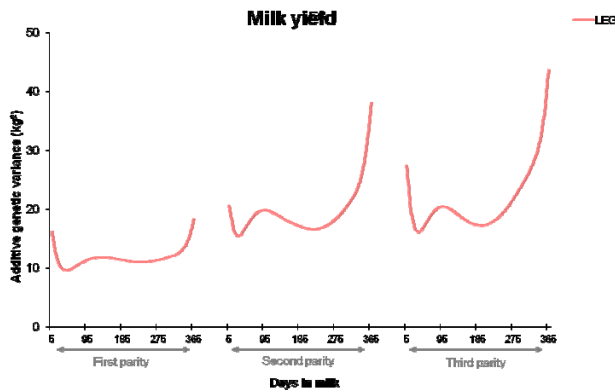


## Estimation of Variance Components

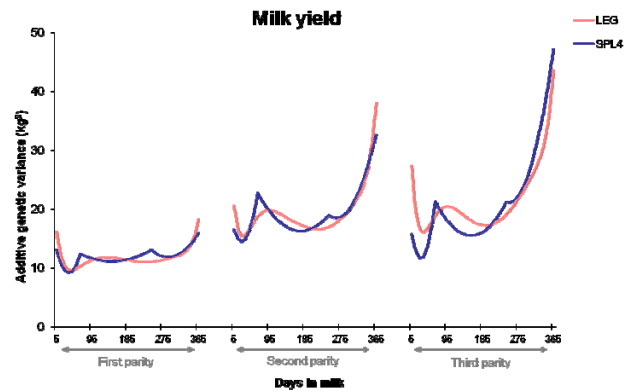
- Bayesian Methods via Gibbs sampler
- A single chain of 100,000 samples
- 20,000 samples discarded as a burn-in
- 80,000 remaining samples used to calculate posterior means of model parameters



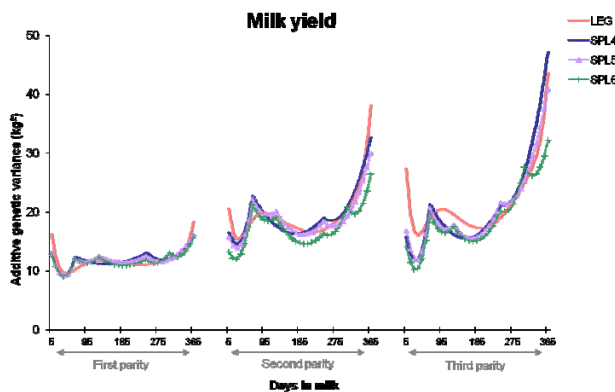
## Additive Genetic Variance



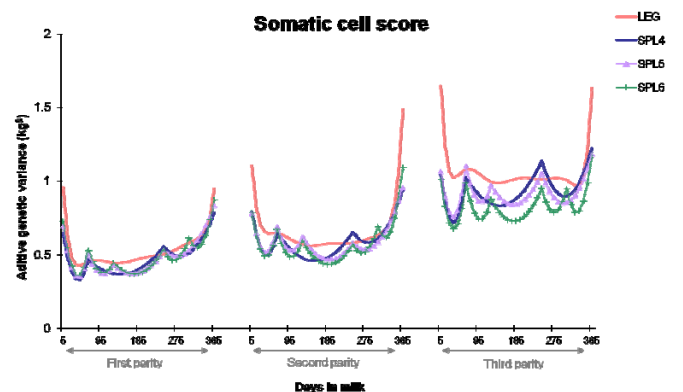
## Additive Genetic Variance



## Additive Genetic Variance

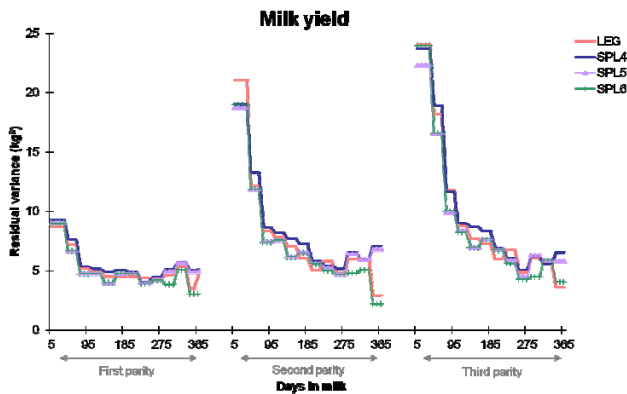


## Additive Genetic Variance





## Residual Variance



## Model Comparison

- Deviance Information Criterion (DIC)
 
$$DIC = \bar{D} + p_D$$
- $\bar{D}$  = the measure of the fit of the model
- $p_D$  = penalty for increasing model complexity

Model	DIC	Rank
LEG	255,808	2
SPL4	274,532	4
SPL5	258,924	3
SPL6	236,646	1



## Genetic Evaluations

- D01: all TD records available until August 2001
  - 27 million TD records, 1.6 million cows
- D06: all TD records available until August 2006
  - 45 million TD records, 2.7 million cows
- Mixed model equations solved by iteration on data with PCG algorithm and block diagonal preconditioner



## Error of Prediction

$$ERP = \sqrt{\frac{\sum_{i=1}^n (ebv06_i - pa01_i)^2}{n}}$$

- *ebv06* - EBV calculated from D06
- *pa01* - parent average predicted from D01
- *n* - number of sires with no daughters in D01 and  $\geq 25$  daughters in D06  
n = 1,929 sires



## ERP – Milk yield

Model	Lactation 1	Lactation 2	Lactation 3
LEG	705	834	782
SPL4	701	814	765
SPL5	709	825	773
SPL6	699	801	759



## ERP – Milk yield

Model	Lactation 1	Lactation 2	Lactation 3
LEG	0%	0%	0%
SPL4	-1%	-2%	-2%
SPL5	0%	-1%	-1%
SPL6	-1%	-4%	-3%



## ERP – SCS

Model	Lactation 1	Lactation 2	Lactation 3
LEG	0.43	0.48	0.58
SPL4	0.42	0.47	0.56
SPL5	0.41	0.47	0.55
SPL6	0.40	0.45	0.52



## ERP – SCS

Model	Lactation 1	Lactation 2	Lactation 3
LEG	0%	0%	0%
SPL4	-3%	-1%	-3%
SPL5	-4%	-2%	-4%
SPL6	-6%	-6%	-10%



## Convergence

Model	Number of iterations	CPU time per iteration	Total CPU time
LEG	667	1,076	8d 7h
SPL4	848	1,020	10d 0h
SPL5	804	1,345	12d 12h
SPL6	911	1,685	17d 18h

2.40 GHz processor, convergence criterion =  $9.9 \times 10^{-13}$

Correlation between coefficients ranged:

RRM with Legendre polynomials  
RRM with splines

$r_g = [-0.5, 0.4]$   
 $r_g = [0.0, 0.8]$



## Conclusions

- Spline model with 6 knots (SPL6) had:
  - ☺ - had the best goodness of fit (DIC)
  - ☺ - had lowest variances at extremes of lactation
  - ☺ - EBV were most stable (ERP)
  - ☹ - had slowest convergence
- SPL6
  - is a superior model to the currently used Legendre polynomials for 5-365 DIM TD records
  - slower convergence → diagonalization has to be applied to make it feasible for a routine genetic evaluation



## Acknowledgments

- DairyGen Council of Canadian Dairy Network 

- NSERC