An Exponential Regression model for estimating daily milk yields

Xiao-Lin Wu,[1,2] George Wiggans,[1] H. Duane Norman,[1] Asha M. Miles,[3] Curtis Van Tassell,[3] Ransom L. Baldwin VI,[3] Javier Burchard,[1] and Joao Durr[1]

[1] Council on Dairy Cattle Breeding, Bowie, MD 20716, USA.

[2] Department of Animal and Dairy Sciences, University of Wisconsin, Madison, WI 53706, USA.

[3] USDA, Agricultural Research Service, Animal Genomics and Improvement Laboratory, Beltsville, MD 20705-2350, USA.

CDCB

COUNCIL ON DAIRY CATTLE BREEDING

ICAR

ICAR 2022

MONTRÉAL CANADA

May 30 - June 3

INTERBULL

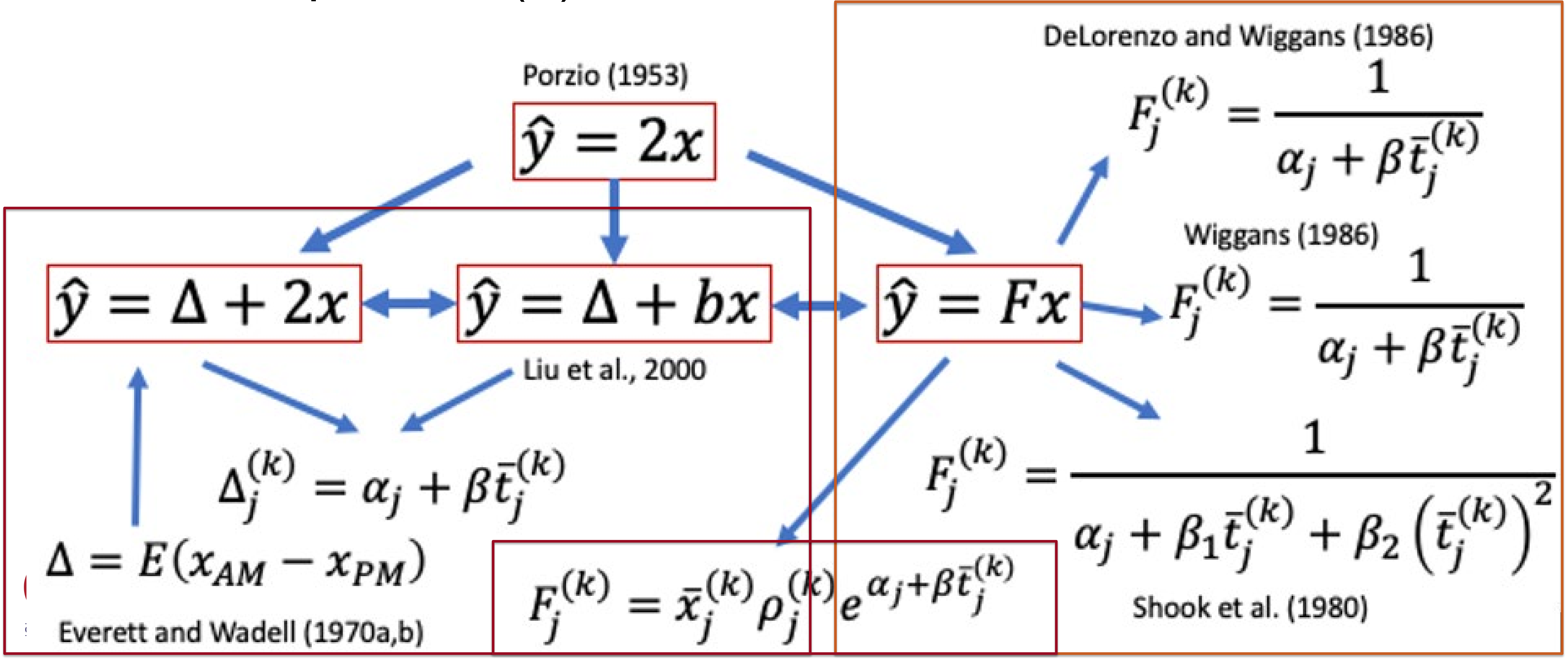
Backgrounds

- A cow is typically milked two or more times each day during her lactation, and not all those milkings are weighed or sampled.
- The initial morning (AM)-evening (PM) milking plan alternately sampled **AM** or **PM** milking on test day throughout the lactation. Daily yield (milk, fat, and protein) was estimated by two times sampled AM or PM yield on each test day, assuming equal AM and PM milking intervals (Porzio, 1953).
- However, AM and PM milking intervals can vary, and milk secretion rates may be different between day and night (Everrett and Wadell,1970a,b).

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

Yield correction factors

- Various methods have been proposed focusing on additive (Δ) and multiplicative (F) correction factors



Goals

- Re-evaluate the performance of existing statistical models, compared to the recently proposed exponential regression model, using cross-validation.
- Characterize ACF and MCF obtained from various methods

COUNCIL ON DAIRY CATTLE BREEDING

Additive correction factor (ACF) models

- Regression with categorical variables (Everett and Wadell, 1970)

$$x_{AM} - x_{PM} = f(\theta) + \epsilon$$

$$\hat{y} = f(\hat{\theta}) + 2x_{PM}$$

- A general form:

$$y = f(\theta) + bx + \epsilon$$

where $f(\theta)$ is a function with categorical or/and continuous dependent variables.

ACF evaluated for each MIC, say k

- Regression with categorical variables, say MIC

$$\Delta_j^{(k)} = E\left(f(\theta|j, MIC = k, else)\right)$$

where else = all categorical variables where applicable

- Regression model on discretized milking interval

$$\Delta_j^{(k)} = \hat{\alpha}_j + \hat{\beta}\bar{t}_j^{(k)}$$

COUNCIL ON DAIRY CATTLE BREEDING

Multiplicative correction factor (MCF) models

- Shook et al. (1980)

$$AMF = \frac{AMP+PMP}{AMP};\ PMF = \frac{AMP+PMP}{PMP}$$

- DeLorenzo and Wiggans (1980)

$$y_{ijk} = F_{jk}x_{ijk} + \gamma_{jk}\left(d_{ijk} - \bar{d}_{jk}\right) + \varepsilon_{ijk}$$

- Wiggans (1986)

$$\frac{x_{ij}}{y_{ij}} = \alpha_j + \beta t_{ij} + \gamma\left(d_{ij} - \bar{d}_j\right) + \epsilon_{ij}$$

✓ Assuming $\mathrm{E}\left(d_{ijk} - \bar{d}_{jk}\right) = 0$ and $E\left(\varepsilon_{ijk}\right) = 0$

$$F_{ijk} = \frac{E\left(y_{ijk}\right)}{E\left(x_{ijk}\right)} = \frac{\sum_{i=1}^{n} y_{ijk}}{\sum_{i=1}^{n} x_{ijk}}$$

✓ Applying first-order Taylor approximation

$$\mathrm{E}\left(\frac{y_{ijk}}{x_{ijk}}\right) = \frac{E\left(y_{ijk}\right)}{E\left(x_{ijk}\right)}$$

✓ MCF computed for MIC k, assuming $\mathrm{E}\left(d_{ijk} - \bar{d}_{jk}\right) = 0$ and $E\left(\varepsilon_{ijk}\right) = 0$

$$F_j^{(k)} = \frac{1}{\hat{\alpha}_j + \hat{\beta}\bar{t}_j^{(k)}}$$

Challenges: ACF models

- For regression with discrete MIC, the number of MIC coupled with other categorical variables increased substantially as more variables were considered.
  - ✓ For example, consider twenty MIC, four herd location regions, four years, four seasons, and two parities. Then, there would be $20 \times 4 \times 4 \times 4 \times 2 = 2{,}560$ specific classes for which ACF would need to be estimated while considering all these categorical variables at the same time.
- For regression with continuous milk interval, systematic biases arose from discretizing milking interval.

$$\Delta_j^{(k)} = E\left(\alpha_j + \beta t_{ij}^{(k)}\right) = \alpha_j + \beta E\left(\bar{t}_j^{(k)} + \left(t_{ij}^{(k)} - \bar{t}_j^{(k)}\right)\right) = \alpha_j + \beta \bar{t}_j^{(k)} + \beta E\left(t_{ij}^{(k)} - \bar{t}_j^{(k)}\right)$$

$$= \alpha_j + \beta \bar{t}_j^{(k)}$$

The above holds assuming $E\left(t_{ij}^{(k)} - \bar{t}_j^{(k)}\right) = E\left(t_{ij}^{(k)}\right) - \bar{t}_j^{(k)} = 0.$

COUNCIL ON DAIRY CATTLE BREEDING

Challenges: MCF models

- An MCF model is challenged by the “ratio” problem because each has a ratio dependent variable in the data density or smoothing function.
- Shook et al. (1980)

$$\frac{PMP_{2k}}{AMP_{2k}+PMP_{2k}} = \alpha + \beta_1 t_k + \beta_2 t_k^2 + \epsilon_k;\ AMF_{2k} = \frac{1}{\hat{\alpha}+\hat{\beta}_1\bar{t}_k+\hat{\beta}_2\bar{t}_k^2}$$

- DeLorenzo and Wiggans

$$\frac{1}{F_j^{(k)}} = \alpha_j + \beta_j \bar{t}_{jk} + \epsilon_{jk};\ F_j^{(k)} = \frac{1}{\hat{\alpha}_j+\hat{\beta}\bar{t}_j^{(k)}}$$

- Wiggans (1986)

$$\frac{x_{ij}}{y_{ij}} = \alpha_j + \beta t_{ij} + \gamma\left(d_{ij} - \bar{d}_j\right) + \epsilon_{ij};\ \ F_j^{(k)} = \frac{1}{\hat{\alpha}_j+\hat{\beta}\bar{t}_j^{(k)}};$$

$$\hat{y}_{ij} = x_{ij} * F_j^{(k)} + \hat{\gamma}\left(d_{ij} - \bar{d}_j\right)$$

Challenges: The “ratio” problem

- Bias due to dropping (or missing) main effect variables

$$\frac{y}{x} = \alpha + \beta t + \epsilon$$

$$\mathrm{y} = \alpha x + \beta t x + \epsilon x$$

- Bias due to additional measurement errors (δ)

$$\frac{y}{x+\delta} = \alpha + \beta t + \epsilon$$

$$\frac{y}{x} = (\alpha + \beta t + \epsilon) + \left(\alpha \frac{\delta}{x} + \beta \frac{t\delta}{x} + \epsilon_{ij} \frac{\delta}{x}\right)$$

An exponential regression model

- Again, consider milking interval and days in milk (Wu et al., 2022)

$$y_{ij} = x_{ij}^{b} e^{(\alpha_j + \beta t_j + \gamma(d_{ij} - \bar{d}_j) + \epsilon_{ij})}$$

- Logarithm transformation

$$log(y_{ij}) = \alpha_j + \beta t_j + \gamma(d_{ij} - \bar{d}_j) + blog(x_{ij}) + \epsilon_{ij}$$

- Analogous to exponential growth function

$$y_{ij} \approx x_{ij}^{b}(1 + 1.718)^{\alpha_j + \beta t_j + \gamma(d_{ij} - \bar{d}_j) + \epsilon_{ij}}$$

Estimating daily milk yield

- Direct approach

$$\hat{y}_{ij} = x_{ij}^{\hat{b}} e^{\left(\hat{\alpha}_j + \hat{\beta} t_j + \hat{\gamma}\left(d_{ij} - \bar{d}_j\right)\right)}$$

- Indirect approach (via MCF)

$$\hat{y}_{ij} = x_{ij}^{(k)} \times F_j^{(k)}$$

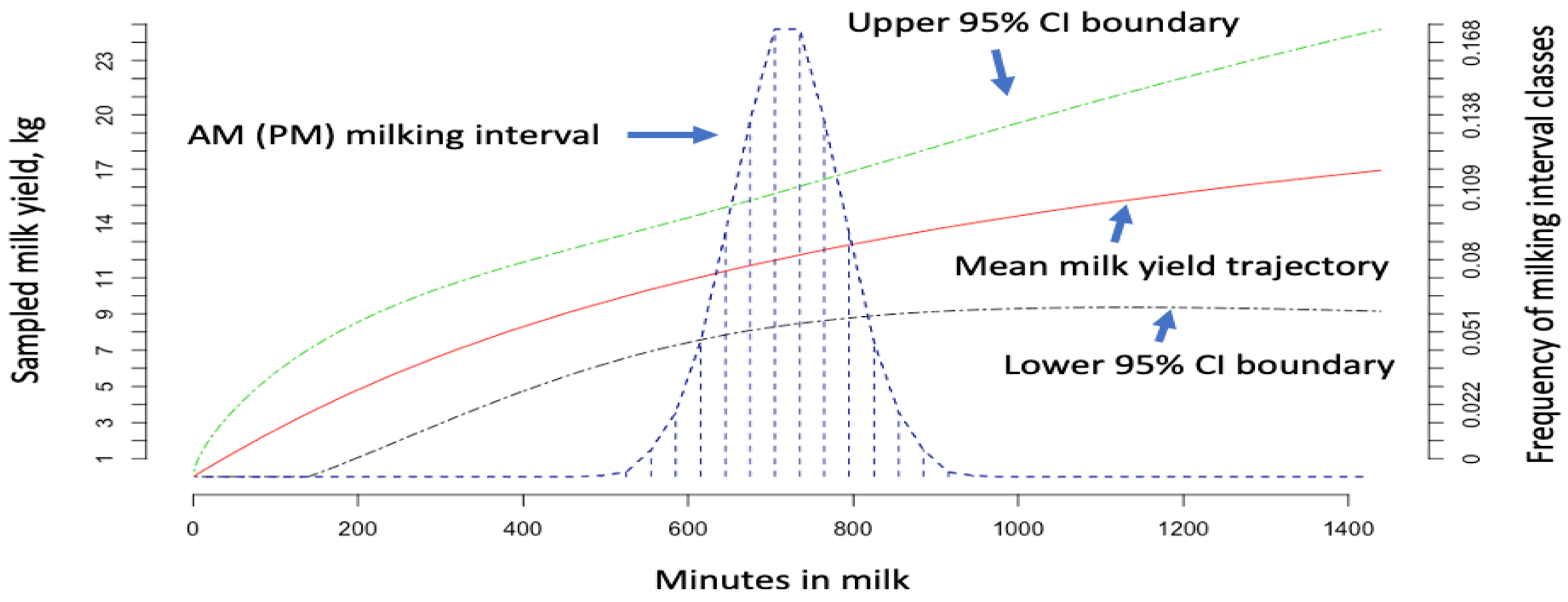
$$F_j^{(k)} = \frac{E\left(y_{ij}^{(k)}\right)}{E\left(x_{ij}^{(k)}\right)} = \rho_j^{(k)} \left(\bar{x}_j^{(k)}\right)^{b-1} e^{\alpha_j + \beta \bar{t}_j^{(k)} + \gamma\left(d_{jk} - \bar{d}_j^{(k)}\right)}$$

where: $\rho_j^{(k)} = e^{\frac{1}{2}\left(V\left(y_{ij}^{(k)}\right)\left(\bar{y}_j^{(k)}\right)^{-2} - bV\left(x_{ij}^{(k)}\right)\left(\bar{x}_j^{(k)}\right)^{-2}\right)}$

A simulation study

- Daily milk yields were simulated based on a modified Michael-Menten function (Klopcic et al., 2012), where $y_{720} \sim TN(12,2)$ and $k \sim TN(0.8,0.1)$.
- Thirty replicates were simulated; Each replicate consists of 3,000 cows; 2/3 for training and 1/3 for testing
- The performance (accuracy and decomposed MSE) for eight methods were evaluated by 10-fold cross-validation.

Mean and 95% confidential intervals of daily milk yield and frequency distribution of AM (PM) milking intervals.



COUNCIL ON DAIRY CATTLE BREEDING

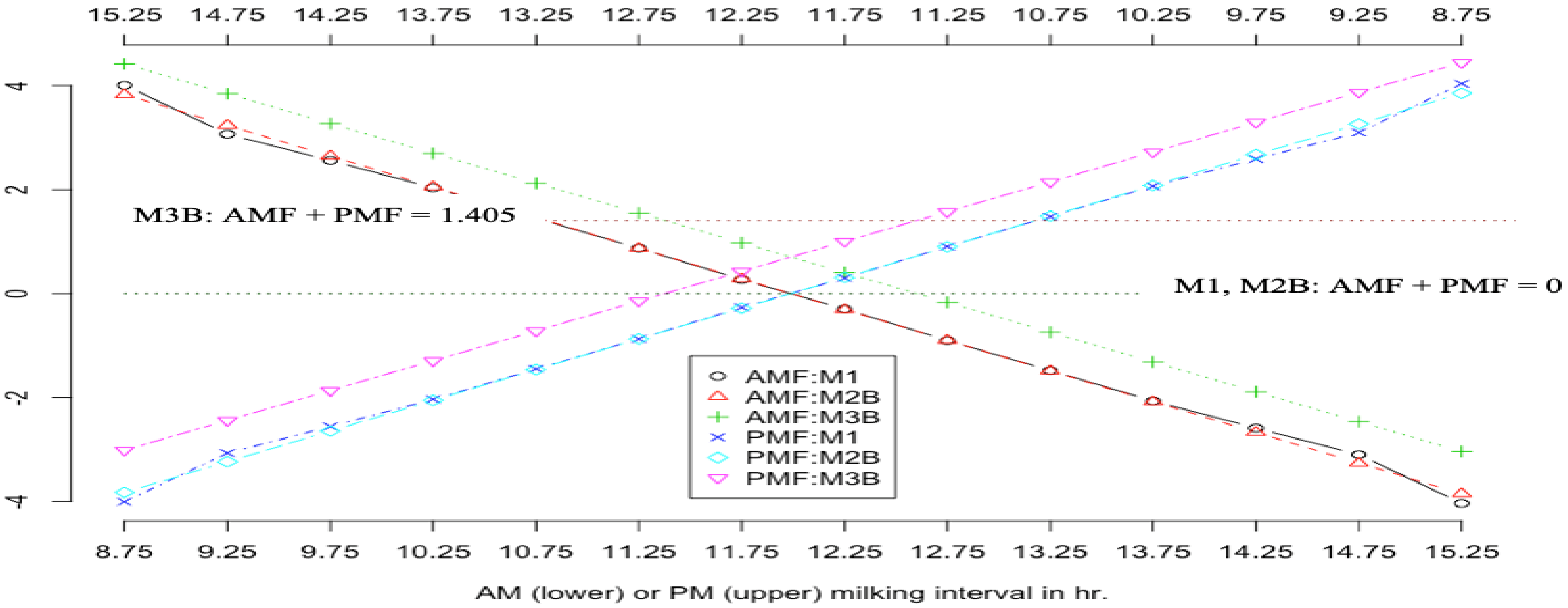
Decomposed MSE and Accuracy

Models	Variance	Bias²	MSE	Accuracy
M1	4.90E-04	0.486	0.486	0.968
M2A	7.40E-05	0.448	0.448	0.971
M2B	7.50E-05	0.480	0.480	0.968
M3A	7.40E-05	0.435	0.435	0.972
M3B	7.40E-05	0.465	0.465	0.970
M4	4.60E-04	0.422	0.422	0.972
M5	4.30E-04	0.420	0.421	0.972
M6A	5.90E-05	0.386	0.386	0.975
M6B	5.90E-05	0.417	0.417	0.973
M7A	5.90E-05	0.376	0.376	0.976
M7B	7.60E-05	0.385	0.385	0.975

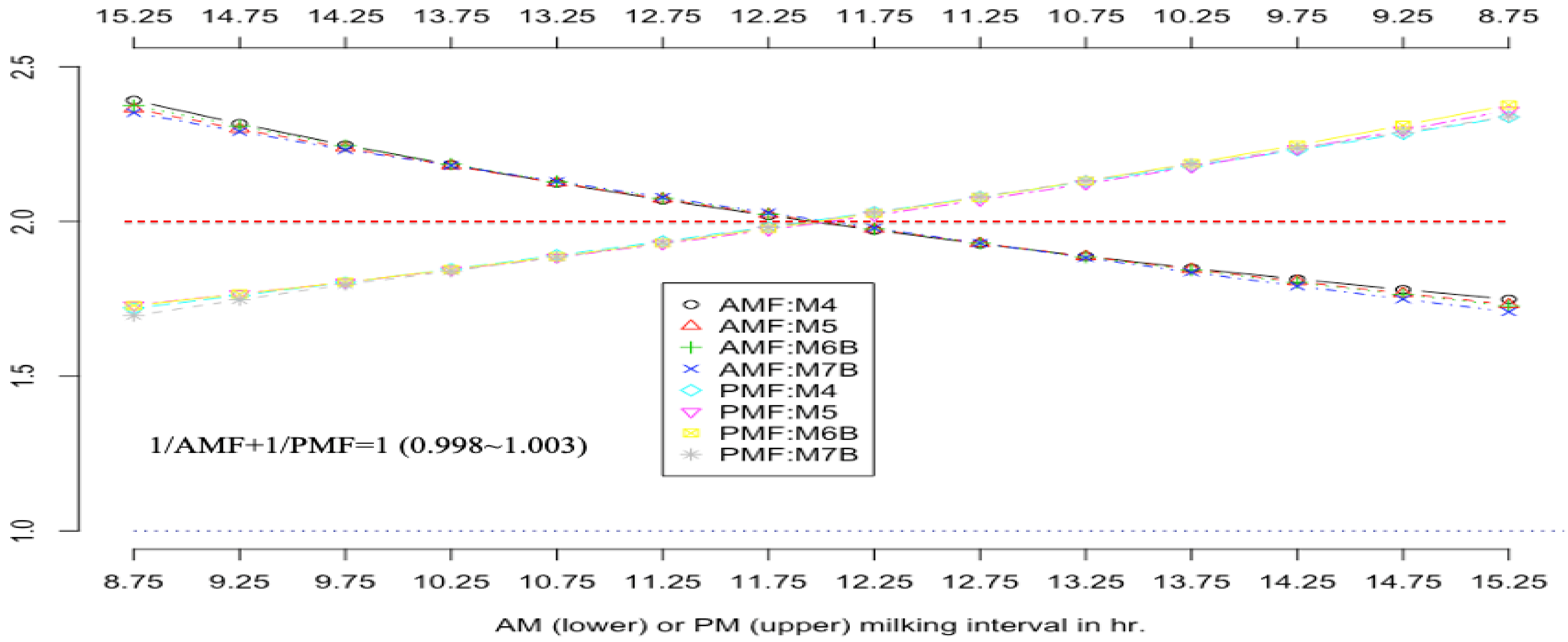
Comparing model parameters

Statistical models	Model parameters				Linear regression fit	Correlation
	α_{AM}	α_{PM}	β	b		
M2A	14.17	14.19	-1.182	2.0	AM: $y = 0.822 + 0.966\ \hat{y}$	0.985
	(0.095)	(0.095)	(0.008)	---	PM: $y = 0.580 + 0.976\ \hat{y}$	0.985
M3A	14.46	14.48	-1.147	1.942	AM: $y = 0.123 + 0.995\ \hat{y}$	0.986
	(0.096)	(0.096)	(0.008)	(0.004)	PM: $y = -0.126 + 1.005\ \hat{y}$	0.985
M6A	0.208	0.208	0.024	---	AM: $y = 0.684 + 0.972\ \hat{y}$	0.986
	(0.002)	(0.002)	(<0.001)	---	PM: $y = 0.577 + 0.976\ \hat{y}$	0.986
M7A	1.324	1.324	-0.048	0.977	AM: $y = 0.102 + 0.996\ \hat{y}$	0.988
	(0.005)	(0.005)	(<0.001)	(0.002)	PM: $y = -0.009 + 1.001\ \hat{y}$	0.987

Additive correction factors



Multiplicative correction factors



CDCB
COUNCIL ON DAIRY CATTLE BREEDING

Conclusions

- Doubling AM (PM) milk yields was approximately taken with equal AM and PM milking intervals, but it was subject to large errors with unequal milking intervals. In contrast, the use of ACF and MCF has effectively reduced the biases due to unequal milking intervals.
- All the methods had high precision but they differ considerably in the accuracy of estimates. Comparably speaking, MCF and linear regression models had smaller MSE and higher accuracy than ACF.
- The exponential regression models had the smallest MSE and the greatest accuracies of all the model evaluated, thus providing a promising alternative tool for estimating DMY.
- The methods were presented with the simulated AM and PM milking data, yet the principles are generally applicable to cows milked more than two times a day, and they apply to the estimation of daily fat and protein yields.

Comments and questions?