Getting milk yields right for genetic (genomic) evaluations: The forgotten pillar

Xiao-Lin Nick Wu (Council on Dairy Cattle Breeding, USA)

Acknowledgement

CDCB	•George R. Wiggans, H. Duane Norman, Javier Burchard, and João Dürr
USDA-AGIL	•Asha M. Miles, Curtis P. Van Tassell, Ransom L. Baldwin VI,
National DHIA & Iyotah Solution	•Steve Sievert, Chip Donatone, Jay Mattison

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

WHO ARE WE?

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

CDCB

COUNCIL ON DAIRY CATTLE BREEDING

CEO: Dr. João Dürr

- Purebred Dairy Cattle Association
- National Association of Animal Breeders
- Dairy Records Processing Centers
- Dairy Records Providers

USDA

Our vision and mission

CDCB VISION

"To be the leading source of genetic information for dairy improvement."

CDCB MISSION

"To drive global dairy cattle improvement by using a collaborative data base to deliver state-of-the-art genetic merit and performance assessments for herd decision making."

Carrillo, J. and Tokushia, K. The US has recorded 5 million genotypes. Hoard's dairyman. March 18, 2021.

GETTING MILK YIELDS RIGHT: WHY?

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

Background

Milk recording is essential for herd management and genetic improvement in dairy cattle.

Lactation (305d) yields are not measured directly but are estimated from test-day milk yields, and the latter were computed from partial daily milk yields.

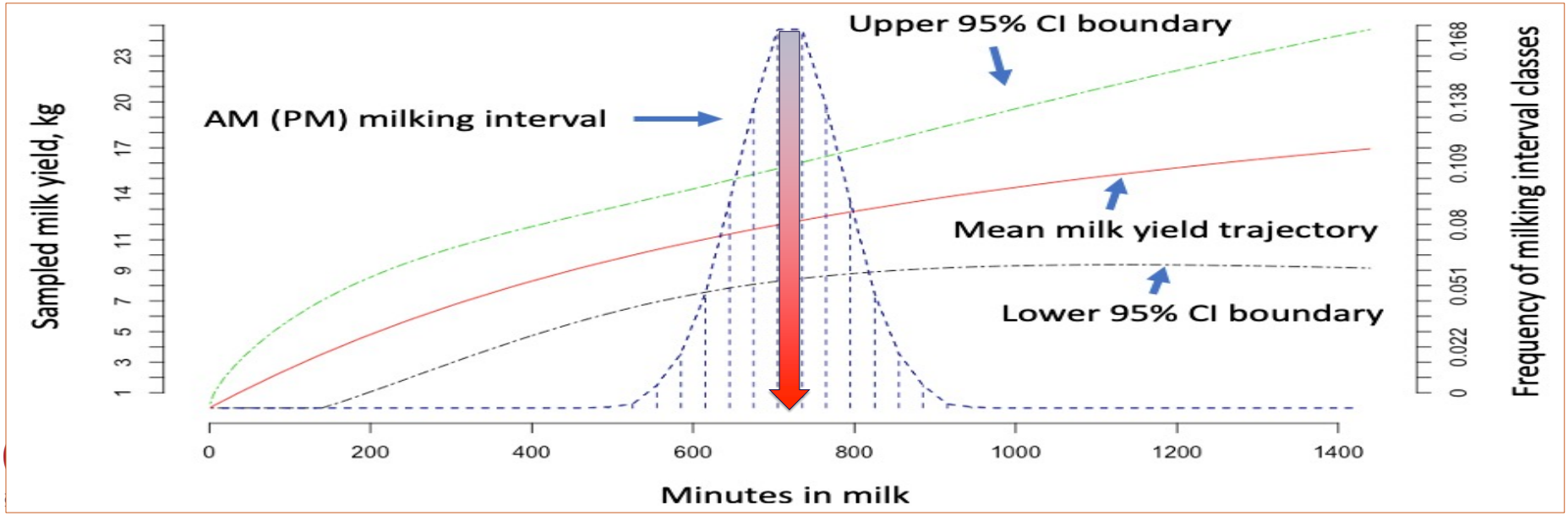
Cows are milked twice or three times or more times daily since 1960s, but not all these milkings are weighed and sampled.

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

AM and PM milking plans: “The ideal city”

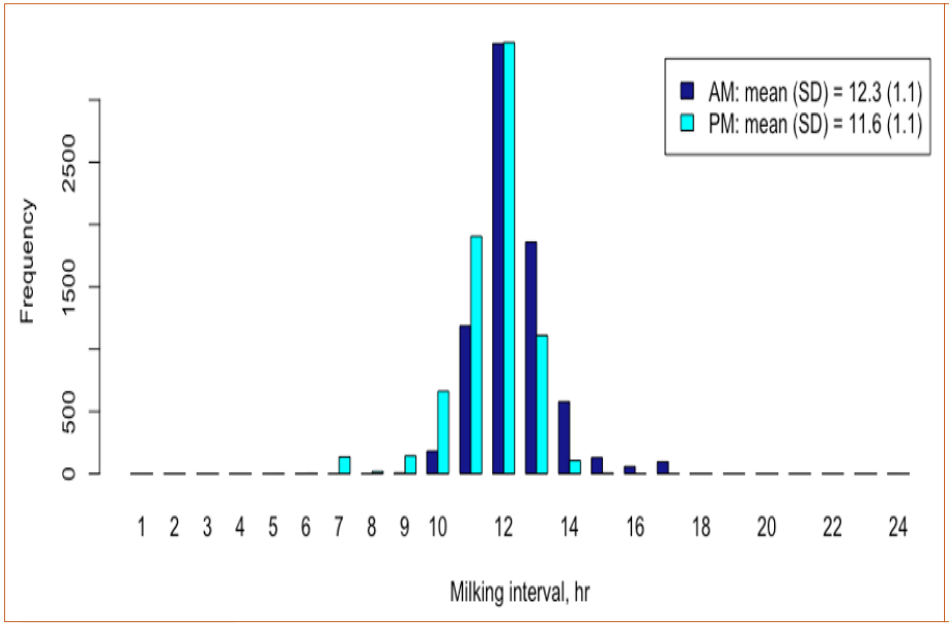
- Assuming equal AM and PM milking times, a test-day yield (y) is taken to be twice the partial (AM or PM) yield on that day (x),

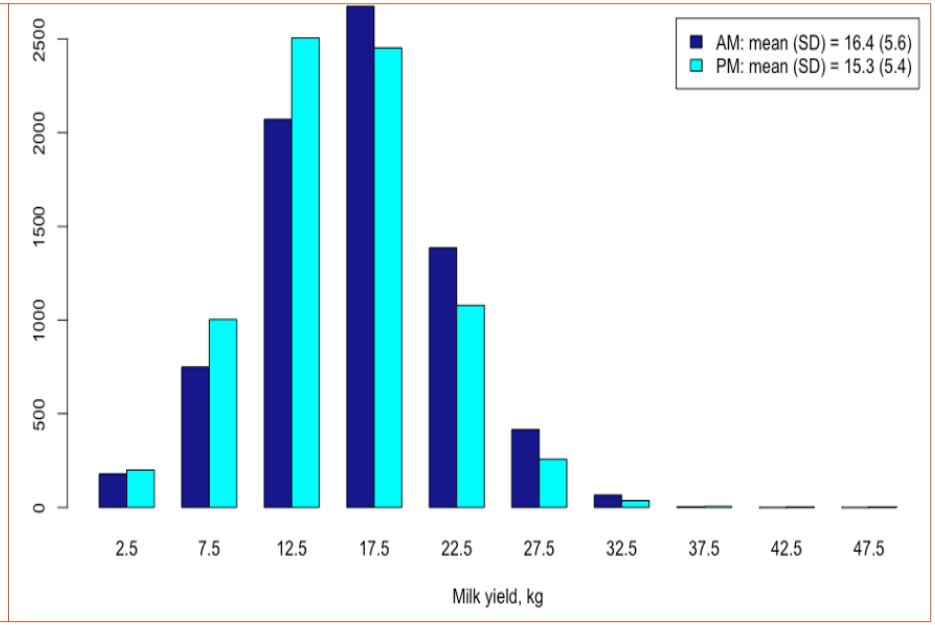
$$y = 2x$$



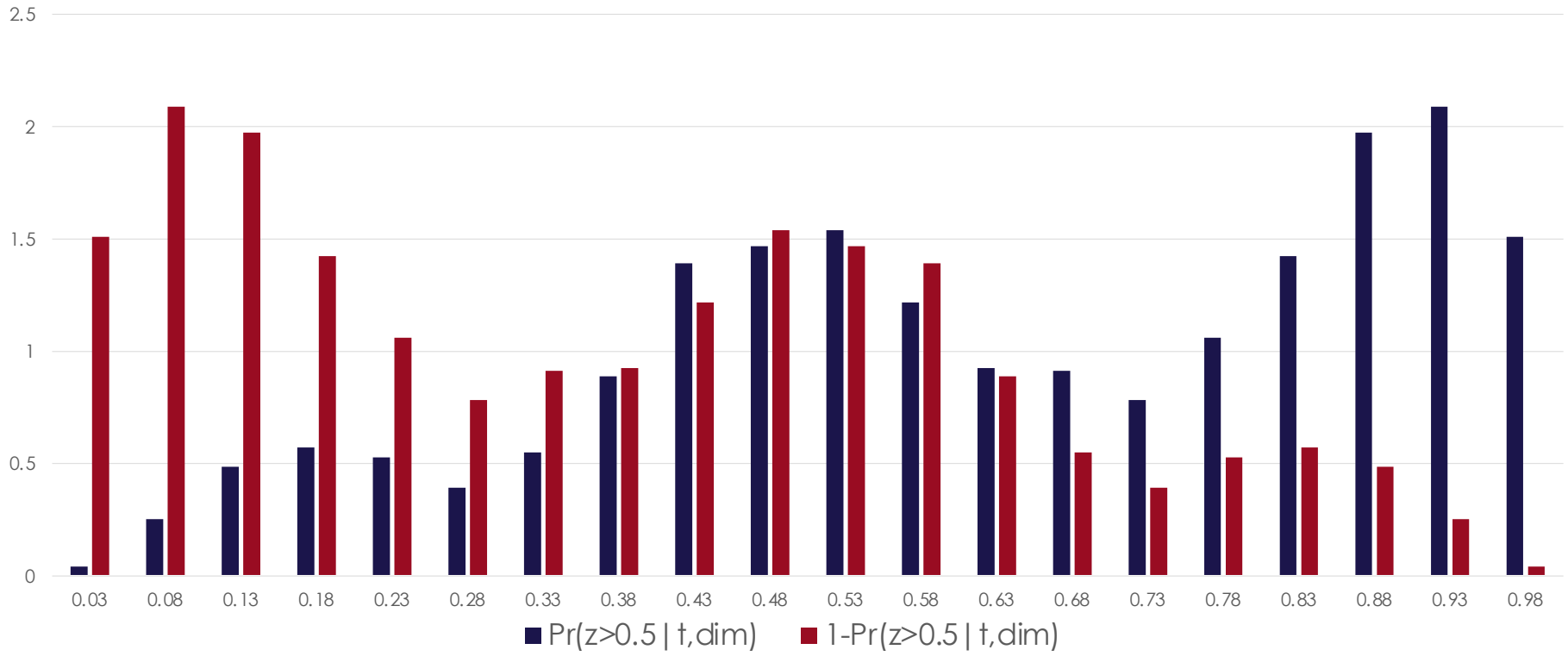
AM and PM milking plans: The real situation

- Morning milking time tends to be longer than evening milking time, and Milk secretion rates may vary between day and night.



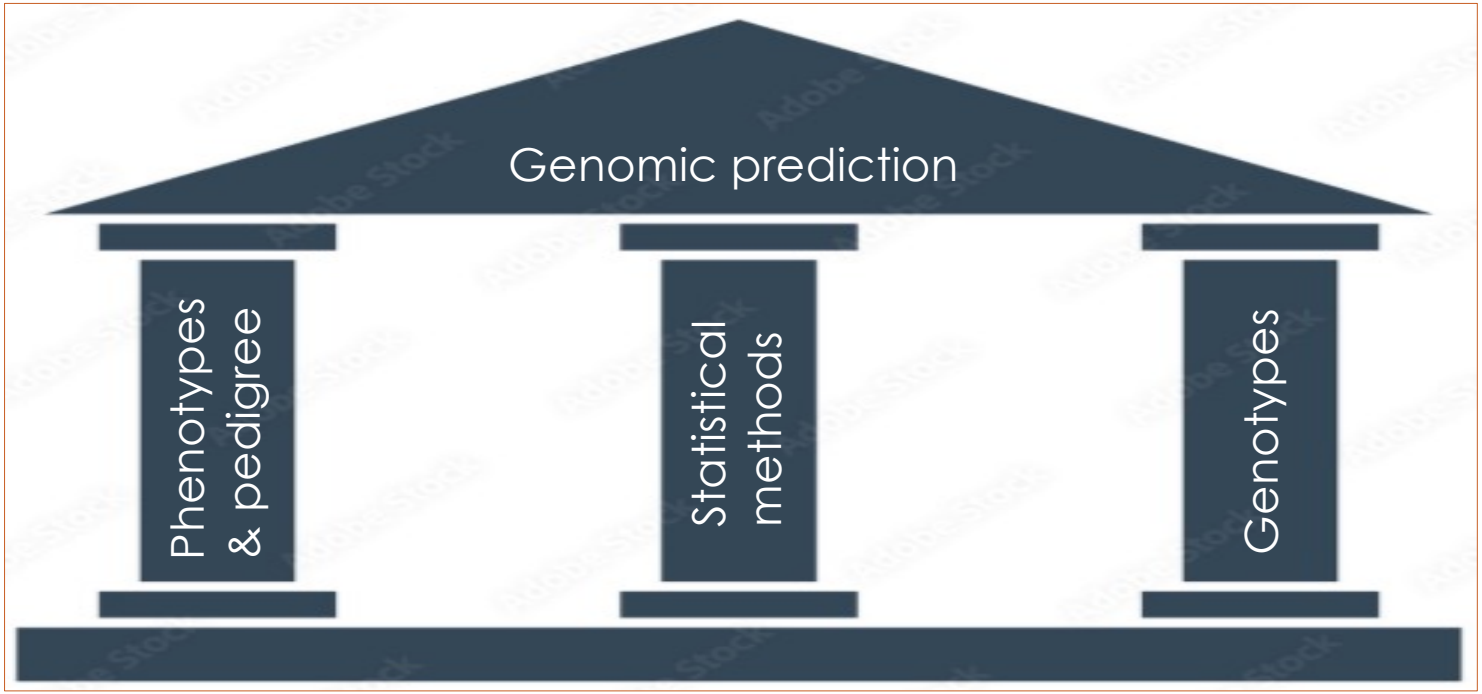


Probability (AM yield > PM yield) = 0.63 (0.25)



CDCB
COUNCIL ON DAIRY CATTLE BREEDING

Tree pillars for genomic prediction



In the past two or three decades, efforts have been be given almost exclusively to the studies on genotypes and statistical methods that can make the best use of these genotypes to make predictions, and the phenotype pillar has faded away.

A technical review on the statistical methods

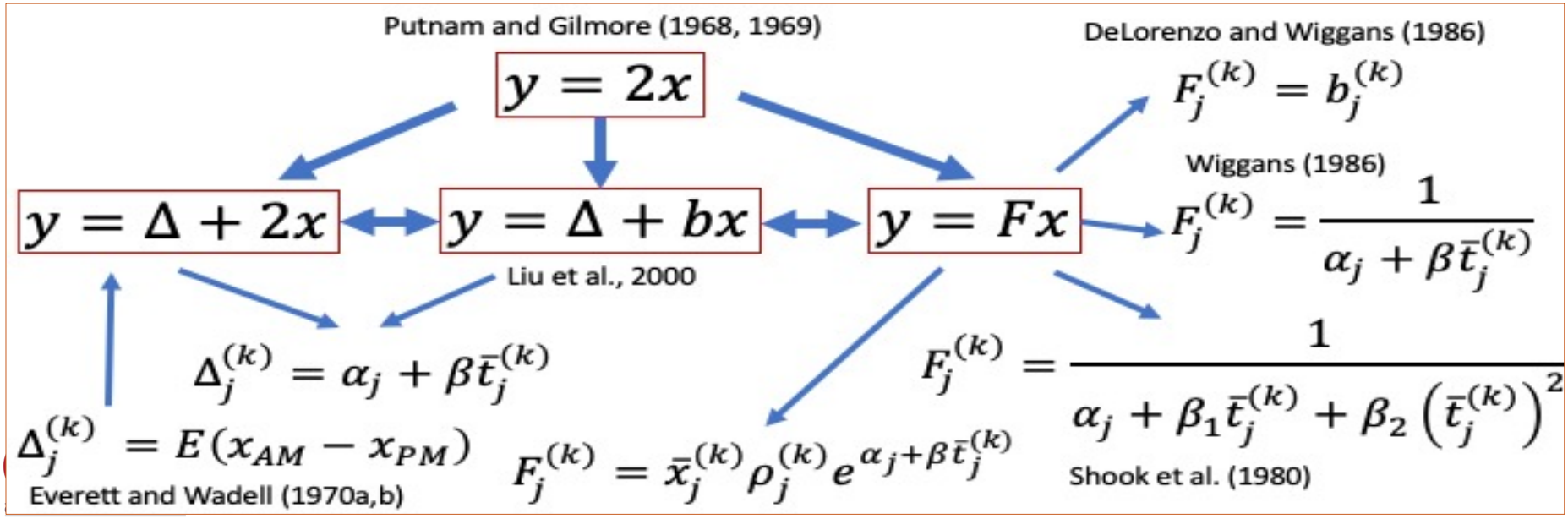
GETTING MILK YIELDS RIGHT: HOW?

Wu, X.-L., Wiggans, G. Norman, H. D., Miles, A. M., Van Tassell, C.P., Baldwin, R.L., Burchard, J. Durr, J., 2023. Daily milk yield correction factors: what are they? JDS Communications. 4:1-6.

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

A historical land map

- Various correction methods were proposed mainly in 1980s, centering on yield factors in two broad categories: additive (ACF) versus multiplicative (MCF) correction factors.



Wu et al., 2022, 2023

What are additive correction factors (ACF) ?

- AM yield (A) known: $y = 2A + \hat{\Delta}_{AM}$
- PM yield (P) known: $y = 2P + \hat{\Delta}_{PM}$

Everett & Wadell, 1970. J. Dairy Sci. 53: 548; 53: 1424.

Where:

t = Milking interval time

d = Days in milk

x = AM or PM milking yield

$$y = \alpha + \beta \mathrm{t} + \gamma(\mathrm{d} - d_0) + 2\mathrm{x} + \epsilon$$

Wu et al., 2023, JDS Communications, 4:1-6.

$$y = \alpha + \beta \mathrm{t} + \gamma(\mathrm{d} - d_0) + \mathrm{bx} + \epsilon$$

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

What is ACF?

- For $y = \Delta + 2x$

$$\Delta_{AM}^{(k)} + \Delta_{PM}^{(k)} = 0$$

- For $y = \Delta + bx$

$$\Delta_{AM}^{(k)} + \Delta_{PM}^{(k)} = (2 - b)\bar{y}^{(k)}$$

where k = the k=th milking interval bin.

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

Wu et al., 2023, JDS Communications, 4:1-6.

What are multiplicative correction factors (MCF)? – An empirical interpretation

- Factors for the AM-PM sampling plan – An empirical interpretation (Shook et al., 1980):

Morning-yield factor: $AMF = \frac{A+P}{A}$; $\Rightarrow$ $P_{ik(j=AM)} = A_{ik(AM)} \times AMF_k$

Evening-yield factor: $PMF = \frac{A+P}{P}$ $\Rightarrow$ $P_{ik(j=PM)} = P_{ik(PM)} \times PMF_k$

where: A = (bulk) yield from morning milkings;

P = (bulk) yield from evening milkings.

$$\left(AMF_k^{-1} + PMF_k^{-1}\right) = 1$$

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

Shook, G., Jensen, E.L., Dickson, F. N. 1980. Dairy Herd Improvement Letter, 56:25-30.

D-W yield factors – "a two-faced creature"

$$y_{ijk} = b_{jk}x_{ijk} + e_{ijk} \Rightarrow b_j = \frac{E(y_{ijk})}{E(x_{ijk})} = \frac{\frac{1}{n}\sum_{i=1}^{n} y_{ijk}}{\frac{1}{n}\sum_{i=1}^{n} x_{ijk}} = \frac{\sum_{i=1}^{n} y_{ijk}}{\sum_{i=1}^{n} x_{ijk}}$$

* This form agrees with Shook et al. (1980)

$$b_j = \frac{\sum_{i=1}^{n}(y_{ijk} - \bar{y}_{jk})(x_{ijk} - \bar{x}_{jk})}{\sum_{i=1}^{n}(x_{ijk} - \bar{x}_{jk})^2}$$

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

DeLorenzo & Wiggans, 1980. J. Dairy Sci. 69: 2386-2394.

D-W yield factors accounting for DIM

$$y_{ijk} = b_j x_{ijk} + \gamma\left(d_{ijk} - d_0\right) + \epsilon_{ijk}$$

$$b_j = \frac{E\left(y_{ijk} - \gamma\left(d_{ijk} - d_0\right)\right)}{E\left(x_{ijk}\right)}$$

$$= \frac{E\left(y_{ijk}\right)}{E\left(x_{ijk}\right)} - \frac{\gamma}{E\left(x_{ijk}\right)} \times E\left(d_{ijk} - d_0\right)$$

The regression coefficient corresponds to a MCF described by the Shook et al. (1980) when $E\left(d_{ijk} - d\right)$=0. Otherwise, it is recognized as MCF adjusted for the DIM effects.

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

The Wiggans (1986) model – The *de facto* MCF model

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

$$F_j^{(k)} = E\left(\frac{y_{ij}^{(k)}}{x_{ij}^{(k)}}\right)$$

Apply first-order Taylor approximation

$$E\left(\frac{y_{ij}^{(k)}}{x_{ij}^{(k)}}\right) \approx \frac{E\left(y_{ij}^{(k)}\right)}{E\left(x_{ij}^{(k)}\right)}$$

$$F_{AM}^{(k)} = 1/\left(\hat{\alpha}_1 + \hat{\beta}\bar{t}_1^{(k)}\right)$$

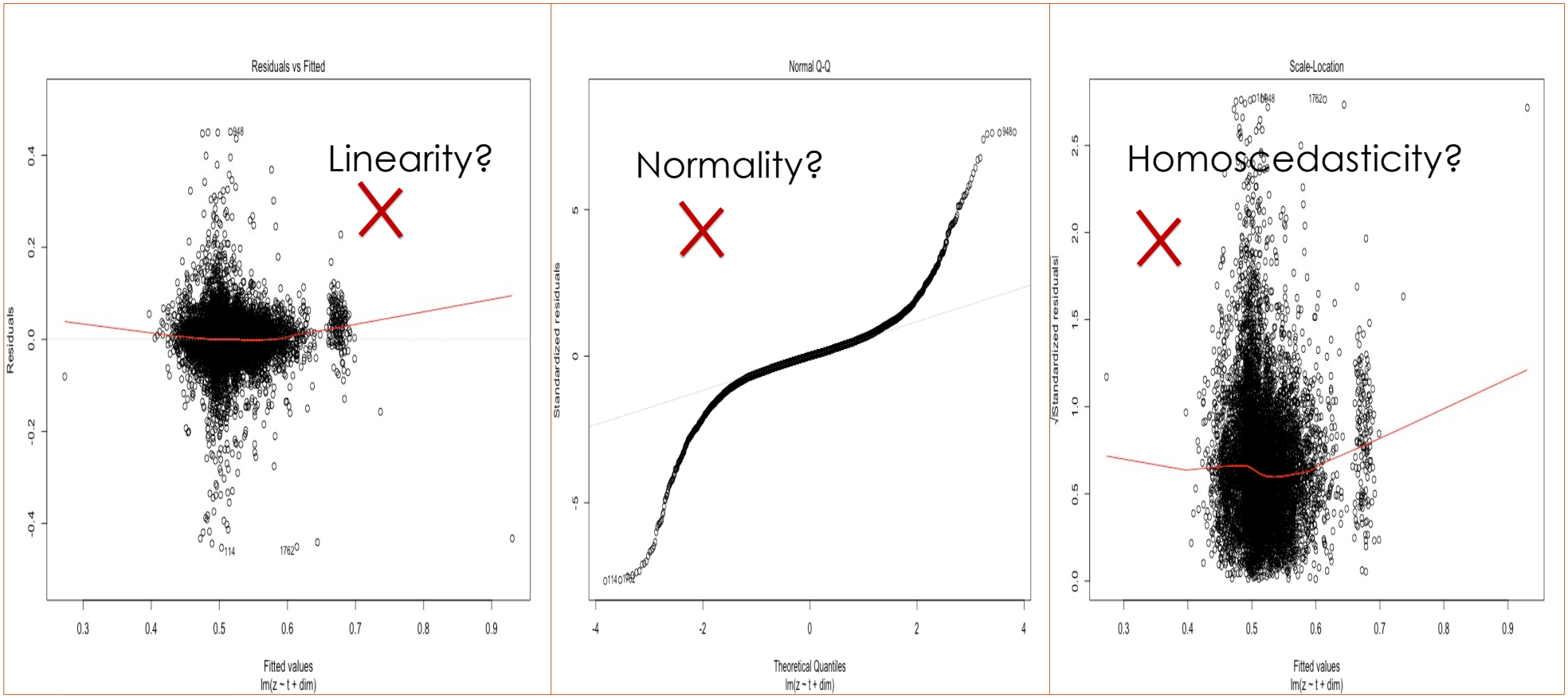
$$F_{PM}^{(k)} = 1/\left(\hat{\alpha}_1 + \hat{\beta}\bar{t}_2^{(k)}\right)$$

$$\hat{y}_{ij}^{(k)} = x_{ij}^{(k)} \times F_j^{(k)} + \hat{\gamma} \times \left(d_{ij}^{(k)} - d_0\right)$$

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

Wiggans, J. Dairy Sci. 1986. 69: 2935-2940.

Extending the Wiggans (1986) model:
– Test of model assumptions



Going beyond linearity: non-linear models

- Polynomial regression
- Step functions
- Regression splines (piecewise polynomials; linear splines; cubic splines; natural splines)
- smoothing splines (Loss + penalty)
- Locally weighted regression (LOESS)
- Generalized additive models (GAM)

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

Wu et al., 2023, a draft in preparation.

Basis functions

- General form:

$$z_i = \beta_0 + \beta_1 b_1(t_i) + \beta_2 b_2(t_i) + \cdots + \beta_K b_k(t_i) + \epsilon_i$$

- Specific forms:
  - Polynomial regression: $b_j(t_i) = x_i^j$
  - Step functions: $b_j(t_i) = I\left(c_j \leq t_i < c_{j+1}\right)$

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

Regression splines vs. Smoothing splines

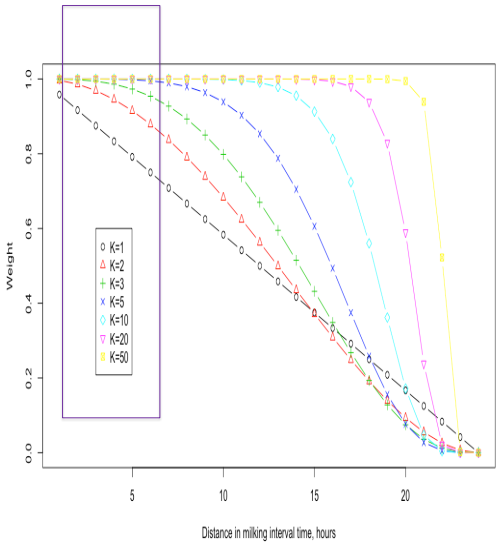
$$\sum_{i=1}^{n} \left(z_i - \hat{f}(t_i)\right)^2 + \lambda \int \hat{f}''(t)^2 dt$$

LOSS Penalty

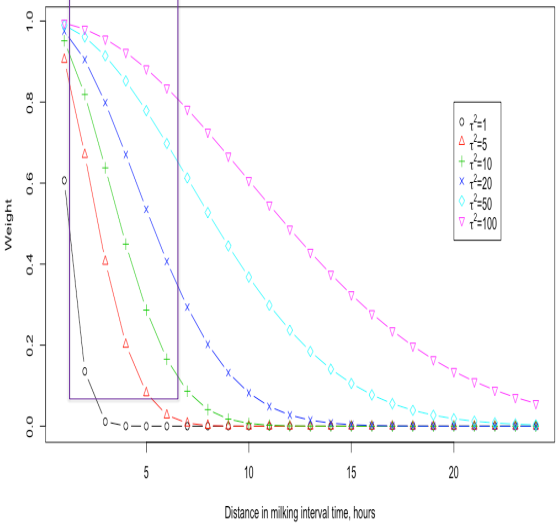
Cubic splines

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

Locally weighted regression (LOESS)



$$\sum_{i=1}^{n} w_i (z_i - \beta_0 - \beta_1 t_i)^2$$



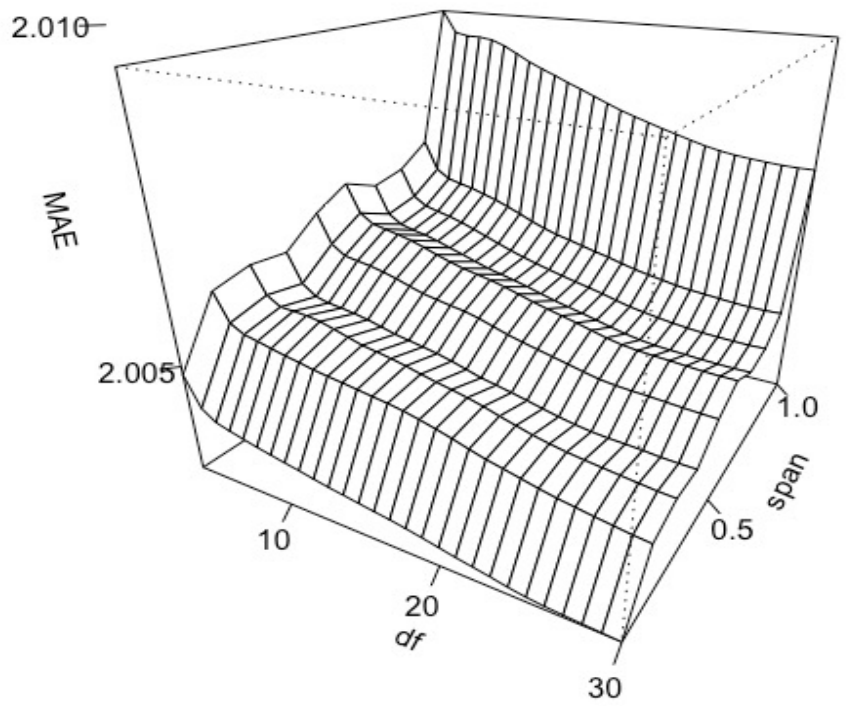
$$w = \left(1 - \left(\frac{dist}{maxdist}\right)^K\right)^K$$

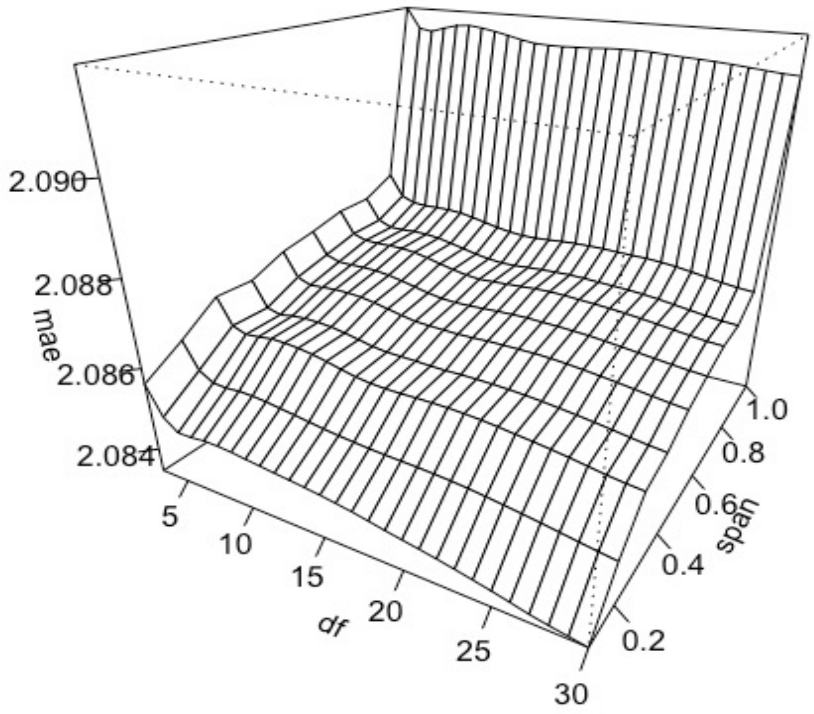
$$w = exp\left(\left(-\frac{(t_i - t^*)^2}{2\tau^2}\right)\right)$$

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

Generalized Additive Models (GAM)

$$z_i = \beta_0 + f_1(t_i) + f_2(t_i) + \cdots + f_k(t_i) + \epsilon_i$$





Exponential regression model: An alternative

$$log\left(\frac{y_{ij}}{x_{ij}}\right) = \alpha_j + \beta t_j + \epsilon_{ij}$$

$b \equiv 1$

$$log(y_{ij}) = \alpha_j + \beta t_j + b log(x_{ij}) + \epsilon_{ij}$$

$Relax\ b \equiv 1$

$$y_{ij} = x_{ij}^{b} e^{(\alpha_j + \beta t_j + \epsilon_{ij})}$$

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

Wu, X.-L., Wiggans, G. Norman, H. D., Miles, A. M., Van Tassell, C.P., Baldwin, R.L., Burchard, J. Durr, J.,. An exponential regression model to estimate daily milk yields. Interbull Bulletin. O. 57. Montréal, Canada, May 30 –June 3, 2022.

Exponential regression: The interpretation

- Exponential growth curve

$$y = a(1 + r)^t$$

- Daily milk yield curve (note: $e \approx 2.718$)

$$y \approx x^b(1 + 1.718)^{\alpha+\beta t+\varepsilon}$$

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

Exponential regression versus existing methods

HOW WELL DO THEY WORK?

Wu, X.-L., Wiggans, G. Norman, H. D., Miles, A. M., Van Tassell, C.P., Baldwin, R.L., Burchard, J. Durr, J., 2022. Statistical methods revisited for estimating daily milk yields: How well do they work? Front. Genet. 13:943706.

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

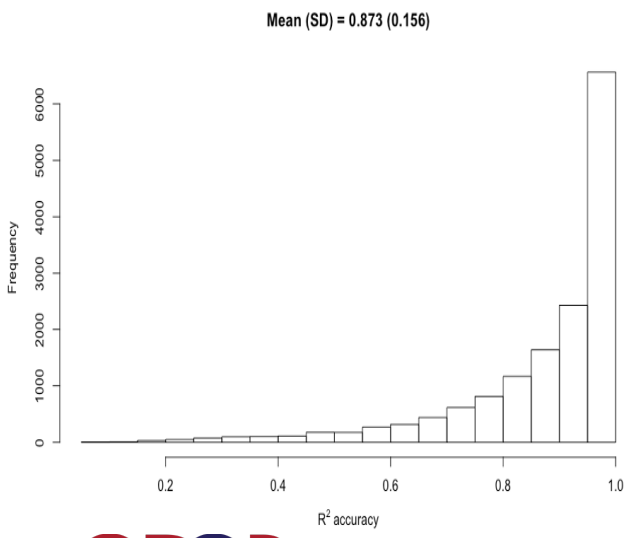
Table 3. Decomposed mean squared error, R^2 accuracy, and correlation between estimated and actual daily milk yield obtained from 10-fold cross-validation [1,2,3,4]

Method	Holstein					Jersey				
	Varb	Bias^2	MSE	Acc	Cor	Varb	Bias^2	MSE	Acc	Cor
M0	0	22.8	22.8	0.821 (0)	0.927 (0)	0.000	14.54	14.54	0.798 (0)	0.948 (0)
M1	0.003	11.3	11.3	0.902 (<0.001)	0.951 (<0.001)	0.012	6.718	6.730	0.895 (<0.001)	0.952 (0.001)
M2A	<0.001	11.3	11.3	0.902 (<0.001)	0.951 (<0.001)	0.002	6.910	6.912	0.892 (<0.001)	0.952 (<0.001)
M2B	<0.001	11.4	11.4	0.902 (<0.001)	0.951 (<0.001)	0.002	6.746	6.748	0.895 (<0.001)	0.952 (<0.001)
M3A	<0.001	10.3	10.3	0.910 (<0.001)	0.951 (<0.001)	0.002	6.078	6.080	0.904 (<0.001)	0.953 (<0.001)
M3B	<0.001	10.3	10.3	0.910 (<0.001)	0.951 (<0.001)	0.003	6.226	6.229	0.902 (<0.001)	0.952 (<0.001)
M4	<0.001	10.2	10.2	0.911 (<0.001)	0.952 (<0.001)	0.025	6.280	6.305	0.901 (<0.001)	0.953 (<0.001)
M5	0.002	11.0	11.0	0.905 (<0.001)	0.951 (<0.001)	0.029	6.707	6.736	0.895 (<0.001)	0.954 (<0.001)
M6	0.001	11.0	11.0	0.904 (<0.001)	0.952 (<0.001)	0.008	6.517	6.525	0.898 (<0.001)	0.953 (<0.001)
M7A	<0.001	10.9	10.9	0.905 (<0.001)	0.952 (<0.001)	0.002	6.570	6.572	0.897 (<0.001)	0.954 (<0.001)
M7B	<0.001	11.0	11.0	0.904 (<0.001)	0.951 (<0.001)	0.004	6.910	6.914	0.892 (<0.001)	0.943 (<0.001)
M8A	0.001	10.1	10.1	0.912 (<0.001)	0.952 (<0.001)	0.003	6.072	6.075	0.905 (<0.001)	0.954 (<0.001)
M8B	0.001	11.0	11.0	0.910 (<0.001)	0.952 (<0.001)	0.010	6.088	6.098	0.903 (<0.001)	0.953 (<0.001)

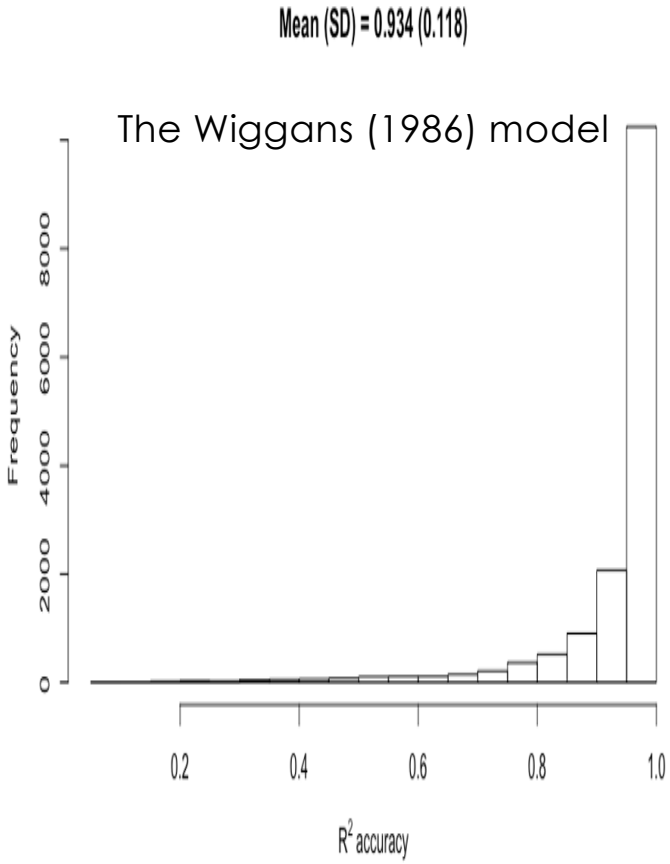
[1] M0 = daily milk yield (DMY) estimated by doubling morning (AM) or evening (PM) milk yield; M1 = additive correction factor (ACF) model with categorical milking interval classes (MIC) and lactation months; M2A = ACF model with continuous variables for milking interval and daily in milk (DIM); M2B = M2A with ACF computed on discretized MIC; M3A = linear regression of daily milk yield on milking interval and DIM; M3B = M3A with ACF computed on discretized MIC; M4 = M3A plus quadratic terms for milking interval and DIM; M5 = multiplicative correction factor (MCF) model according to Shook et al. (1980); M6 = MCF model according to DeLorenzo and Wiggans (1986); M7A = linear regression of AM or PM proportion of DMY on milking interval and DIM (Wiggans, 1986); M7B = M7A with MCF computed for discretized MIC (Wiggans, 1986); M8A = exponential regression model (Wu et al., 2022); M8B = M8A with MCF computed on discretized MIC.

Distribution of R^2 accuracies

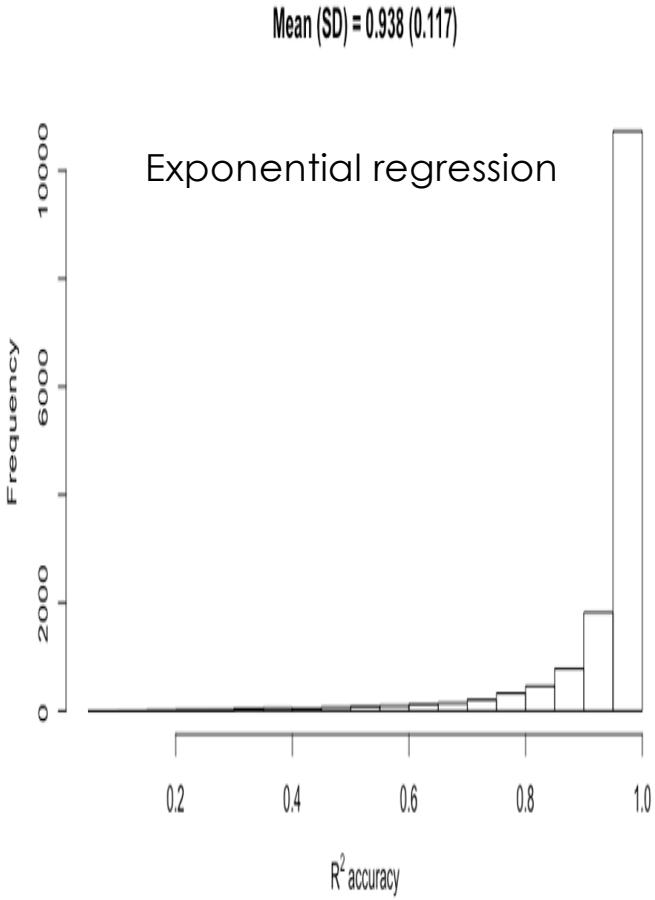
$y = 2x$



The Wiggans (1986) model

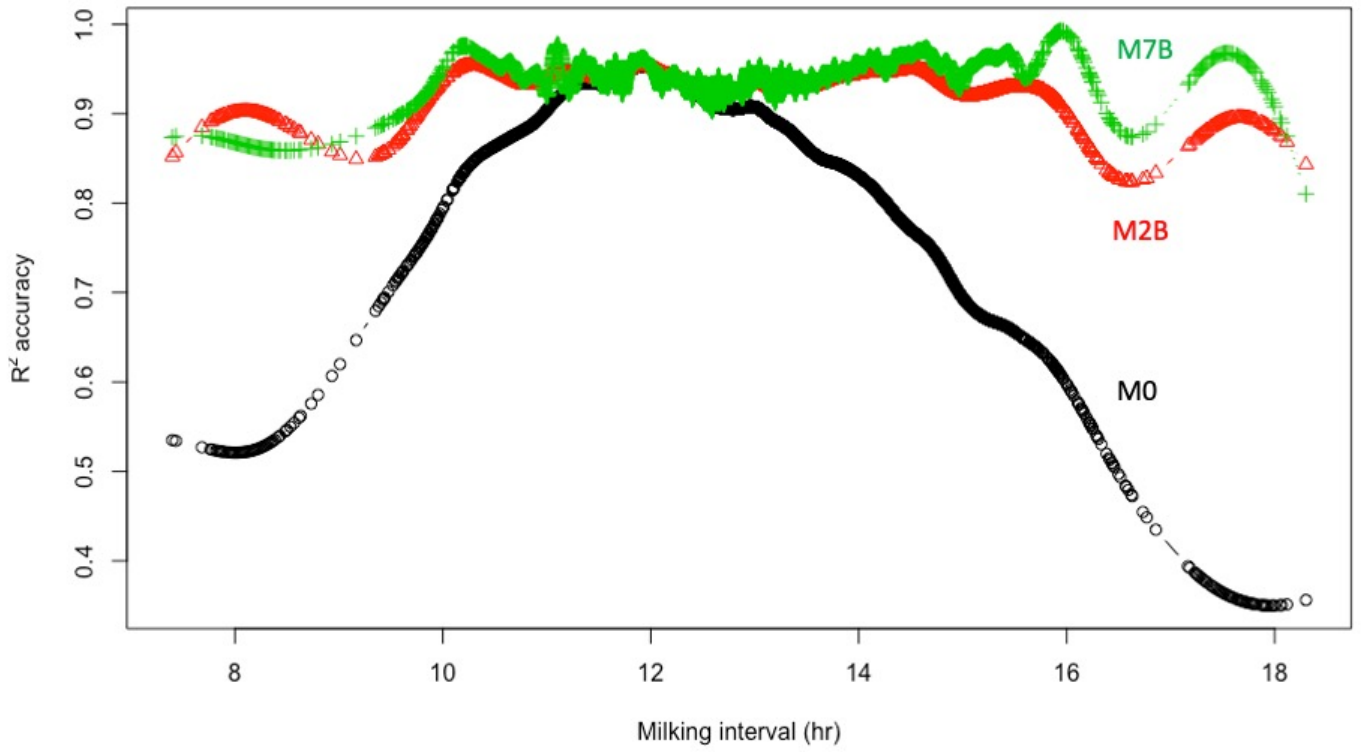


Exponential regression

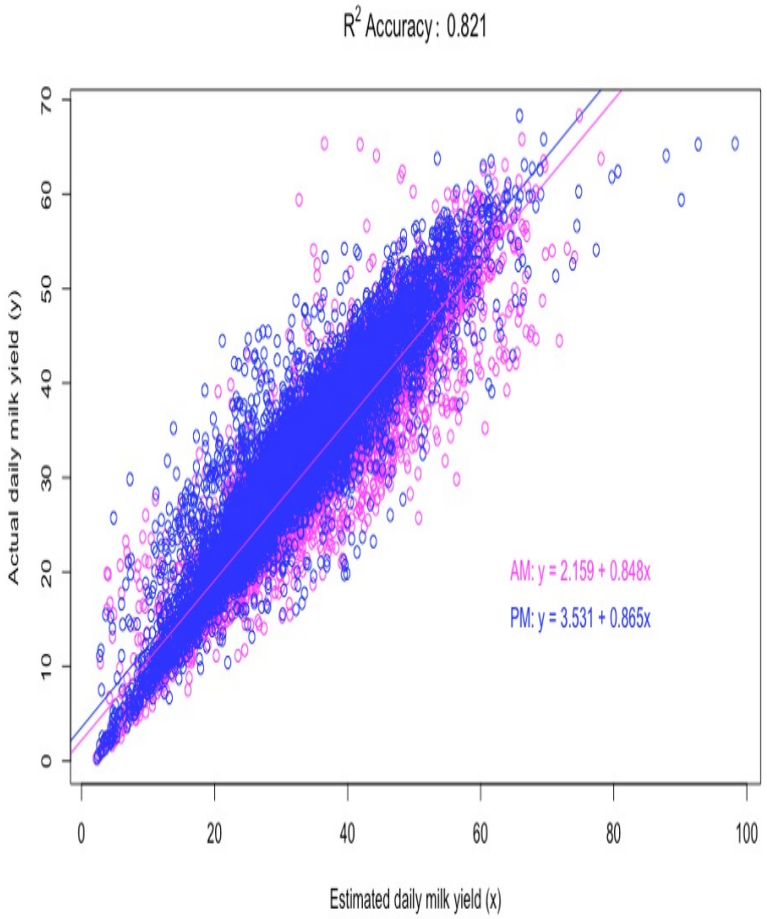


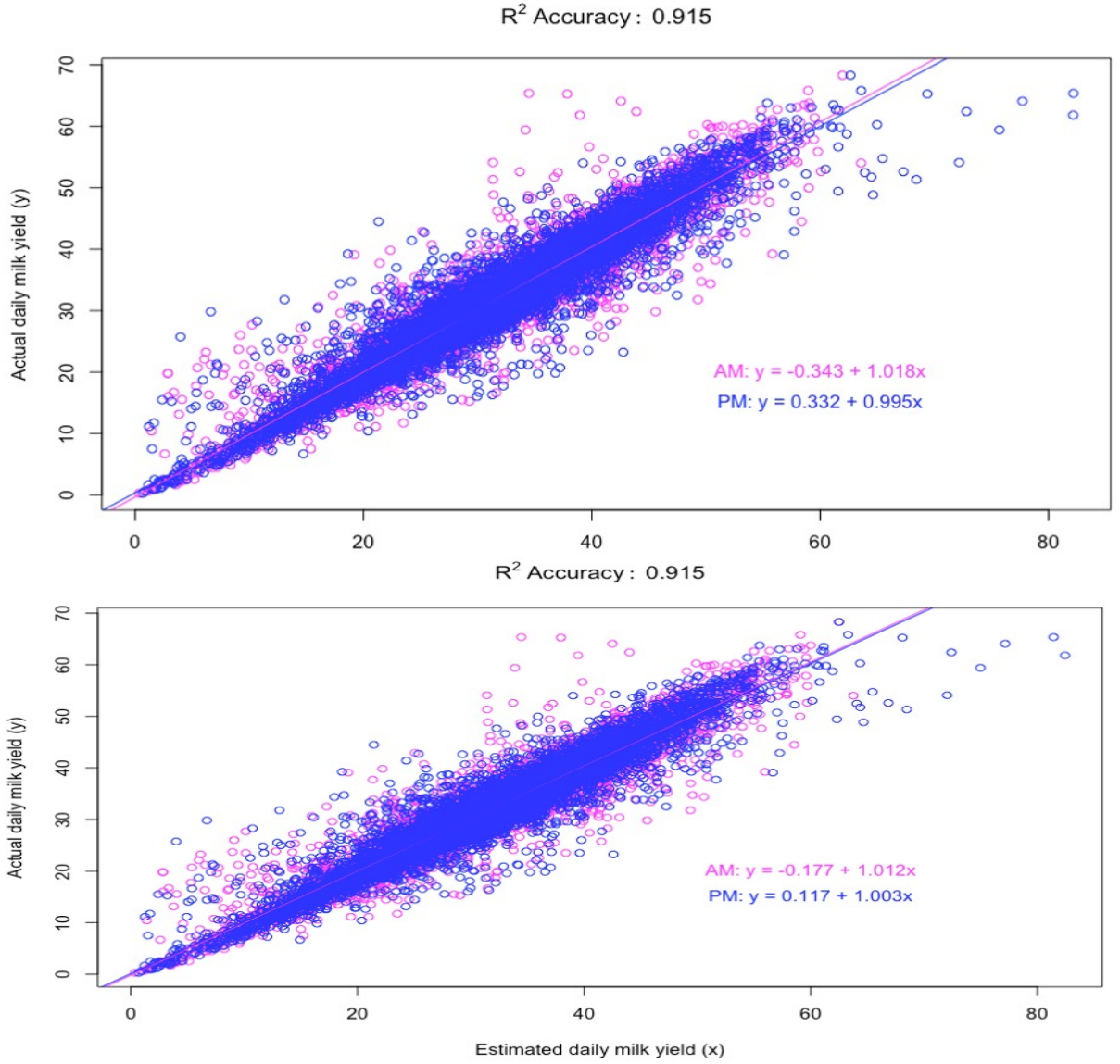
CDCB
COUNCIL ON DAIRY CATTLE BREEDING

Larger errors arose from unequal milking time



CDCB
COUNCIL ON DAIRY CATTLE BREEDING





CDCB
COUNCIL ON DAIRY CATTLE BREEDING

(2) The Wiggans Linear regression versus non-linear models

HOW WELL DO THEY WORK?

Wu et al., 2023. Estimating test-day milk yields by modeling proportional daily milk yields: Going beyond linearity. (A manuscript in preparation)

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

Comparing yield factors

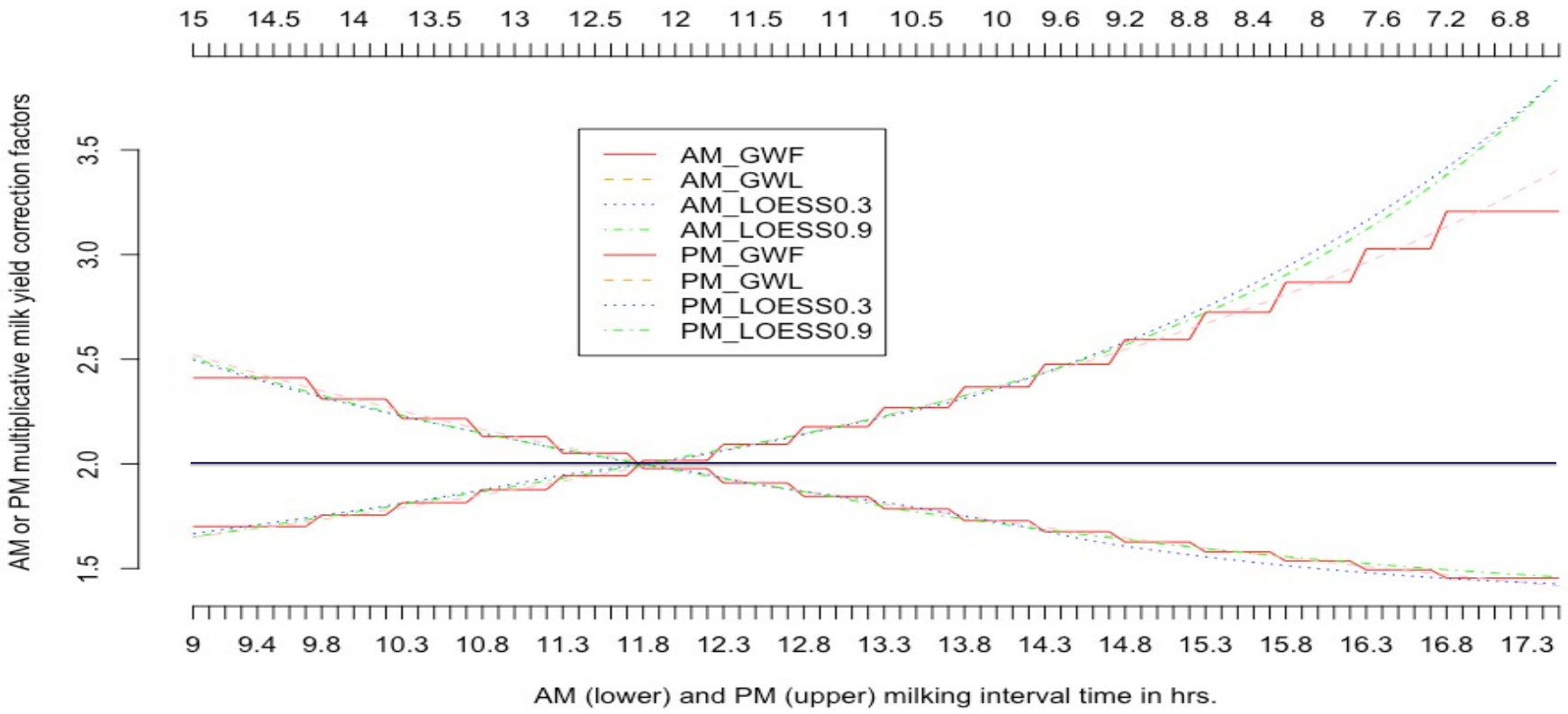


Table 2. Mean absolute errors (MAE), mean squared errors (MSE), R^2 accuracy, and correlation (r) between actual and estimated daily milk yields, obtained from ten-fold cross-validation using different models

Model [1]	AM milking				PM milking			
	MAE	MSE	R2ACC	CORR	MAE	MSE	R2ACC	CORR
GW1	2.069	10.47	0.9085	0.9520	2.069	10.47	0.9085	0.9520
GW2	2.093	10.62	0.9073	0.9514	2.094	10.62	0.9073	0.9514
QR	2.066	10.46	0.9086	0.9521	2.067	10.46	0.9086	0.9521
SF	2.080	10.50	0.9083	0.9521	2.080	10.50	0.9083	0.9521
CS	2.066	10.47	0.9086	0.9522	2.066	10.46	0.9086	0.9522
LOESS	2.062	10.44	0.9088	0.9520	2.062	10.43	0.9089	0.9520
GAM	2.049	10.30	0.9098	0.9527	2.049	10.30	0.9098	0.9527
S-J-D	2.085	10.52	0.9081	0.9517	2.085	10.53	0.9081	0.9517
DW	2.076	10.45	0.9087	0.9518	2.076	10.45	0.9087	0.9518

[1] GW1 = linear regression according to Wiggans (1986); GW2 = multiplicative yield factors according to Wiggans (1986); QR = quadratic regression; SF = step functions; CS = cubic splines; LOESS = locally estimated scatterplot smoothing; GAM = generalized additive model; S-J-D = multiplicative yield factors according to Shook et al. (1986); D-W = multiplicative yield factors according to DeLorenzo and Wiggans (1986).

Discretizing milking time leads to accuracy loss

- When daily milking yield is non-linear with milking interval

$$E\left(\mu_{\bar{t}_j^{(k)}}\right) \neq a_j + \beta \bar{t}_j^{(k)}$$

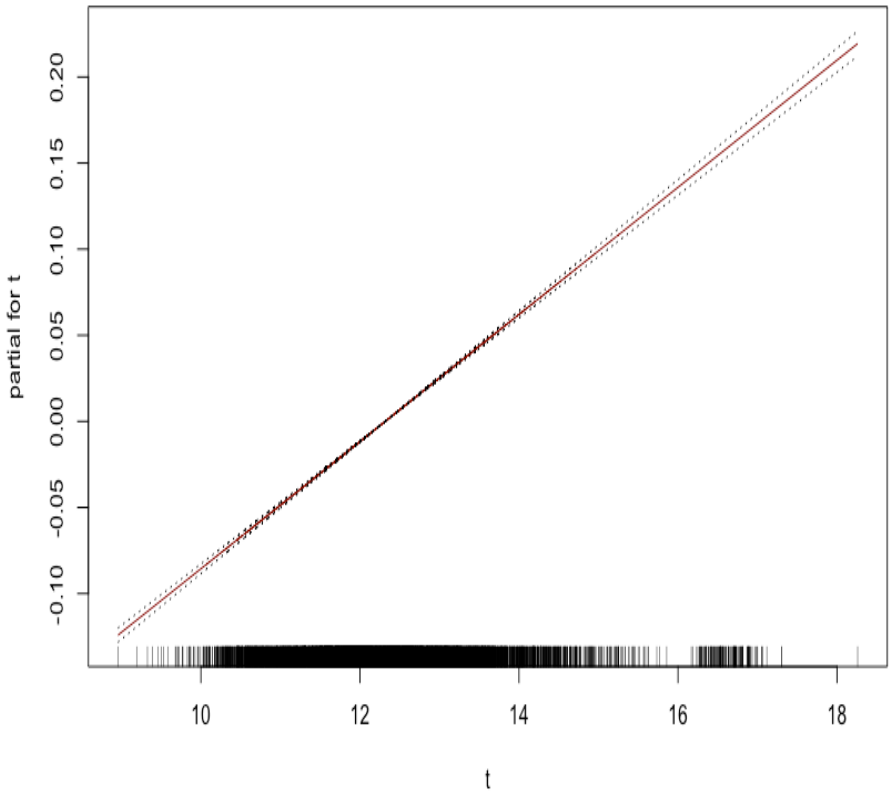
- When daily milking yield is linear with milking interval

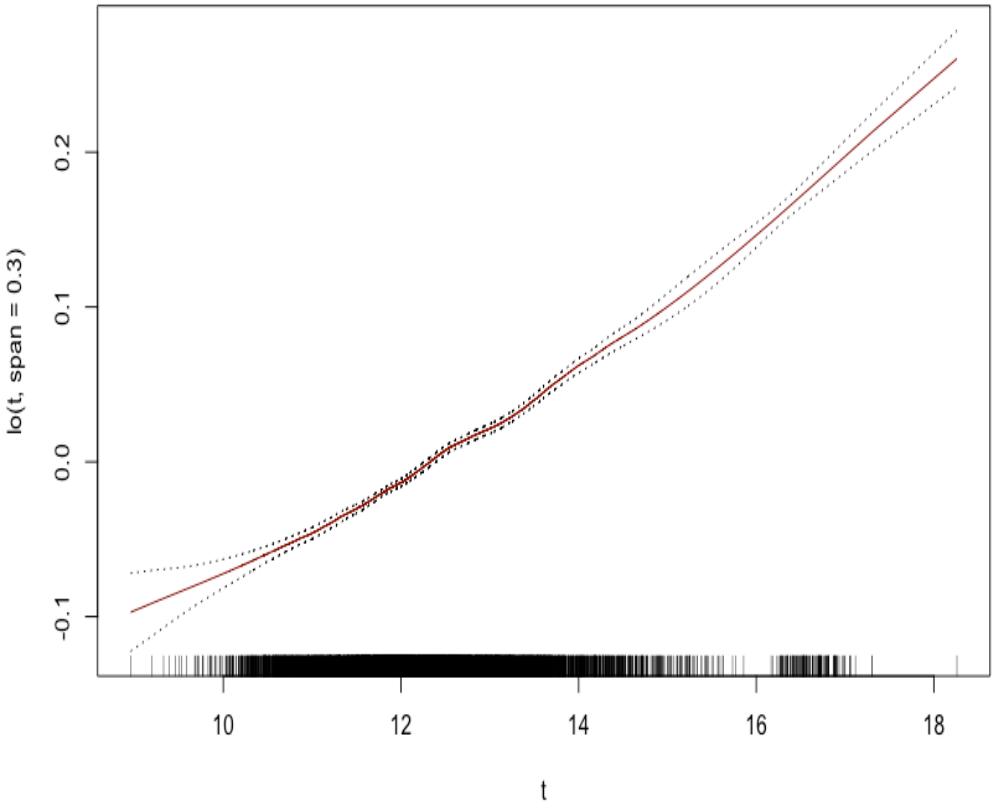
$$E\left(\mu_{\bar{t}_j^{(k)}}\right) = \mathrm{E}\left(\alpha_j + \beta t_{ij}^{(k)}\right) = \alpha_j + \beta \bar{t}_j^{(k)} + \beta E\left(t_{ij}^{(k)} - \bar{t}_j^{(k)}\right)$$

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

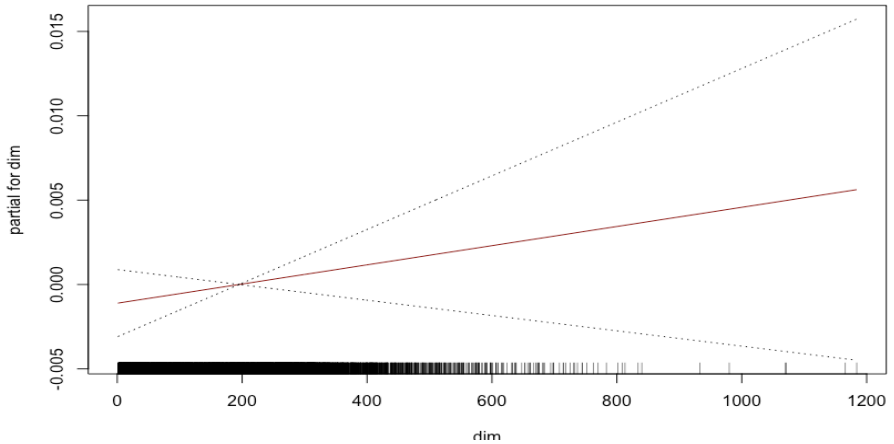
CDCB
COUNCIL ON DAIRY CATTLE BREEDING

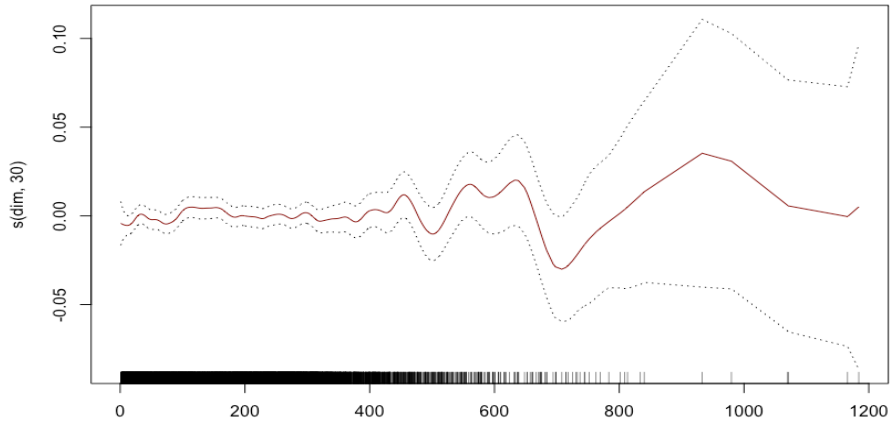
Fitting linear vs. LOESS for milking interval time

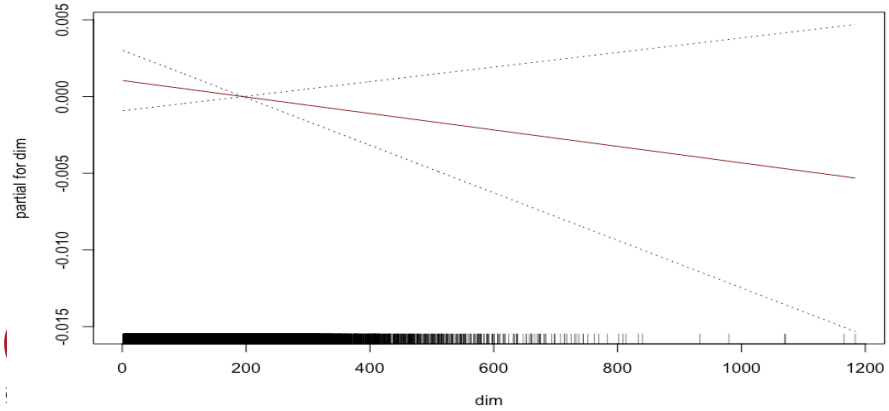


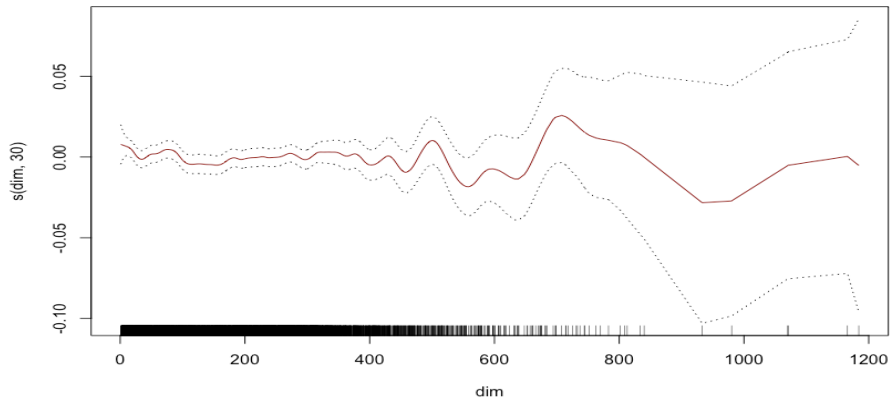


Fitting linear vs. Cubic splines for days in milk









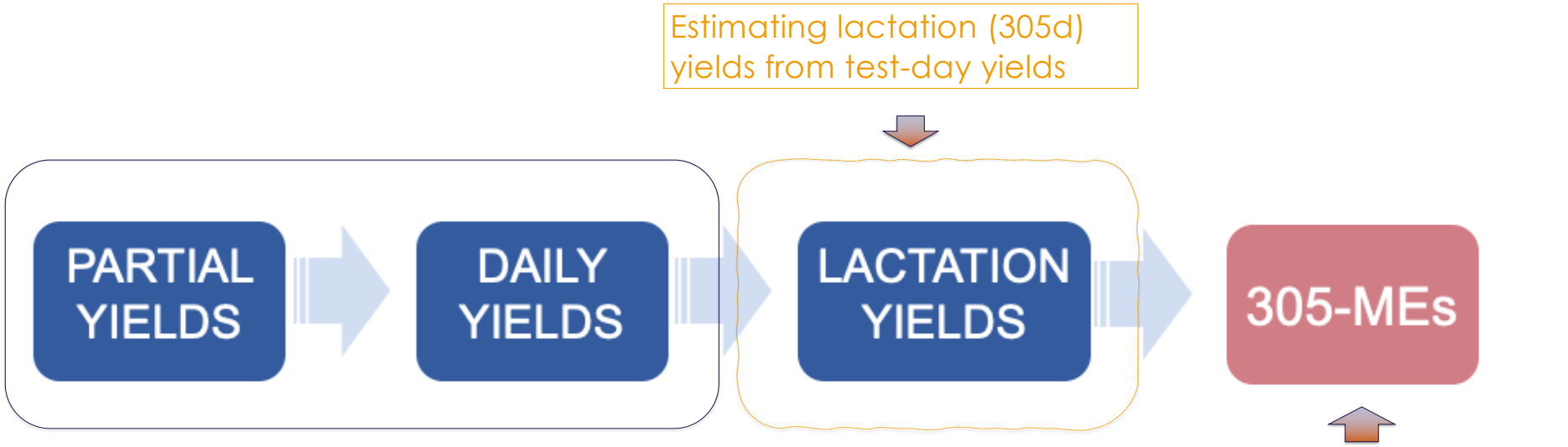
Standardization of lactation records

GOING FURTHER

A project led by Dr. Asha Miles (USDA-AGIL)

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

Milk yield corrections: The full spectrum

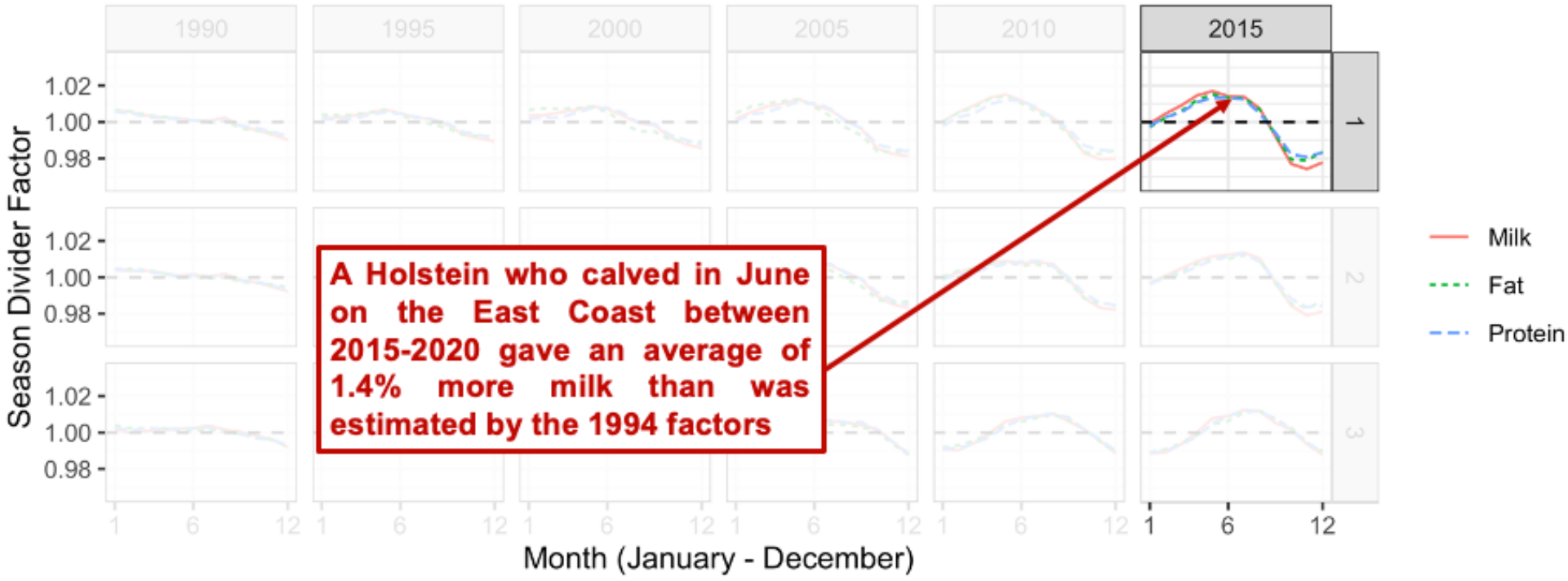


Mature equivalent records are lactation records that have been adjusted for age at freshening, frequency of milking and season of the year at calving.

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

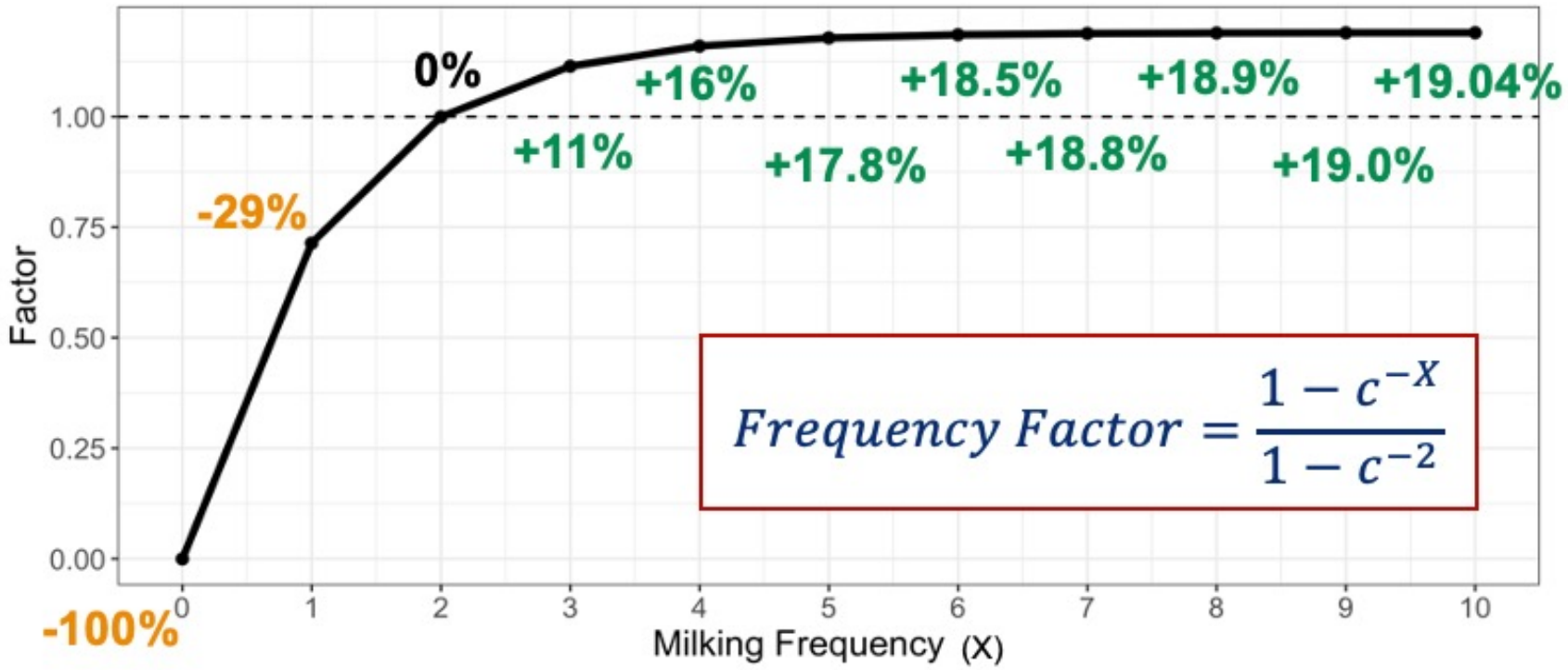
SEASON-REGION CORRECTIONS

Holstein Example



FREQUENCY-INTERVAL CORRECTION

Milk Yield Example



Take-home message

- The current linear systems work but the parameters need to be updated.
- Computing MCF on large discretized milking interval bins can lead to accuracy loss. Optimal bins needs to be small, which otherwise invite nonlinear models.
- Non-linear models can further improve the accuracy of estimated daily milk yields. Promising models include local regression (LOESS), GAM, and exponential regression.
- A joint effort is ongoing between CDCB, USDA-AGIL, and NHDIA to collect data toward updating the current systems.

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

CDCB
COUNCIL ON DAIRY CATTLE BREEDING