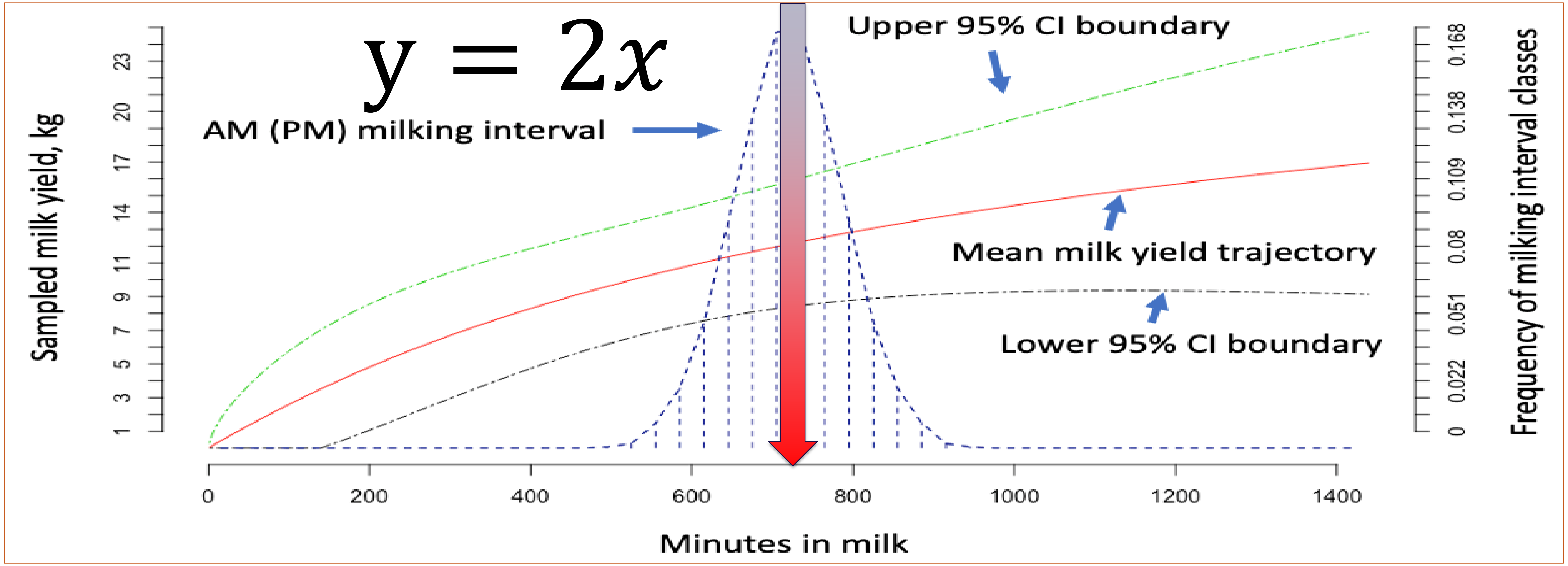
Breaking the bin: Why does modeling continuous milking interval time matter for getting test-day milk yields right?

X.-L. Wu,[1,2] G. R. Wiggans,[1] H. D. Norman,[1] H. A. Enzenaue,[1] A. M. Miles,[3] C. P. Van Tassell,[3] R. L. Baldwin VI,[3] J. Burchard,[1] and J. Dürr [1]

1 Council on Dairy Cattle Breeding, Bowie, MD, USA.
2 Department of Animal and Dairy Sciences, University of Wisconsin, Madison, WI, USA.
3 USDA, ARS, Animal Genomics and Improvement Laboratory, Beltsville, MD, USA.

AM and PM milking plans: “The ideal situation”

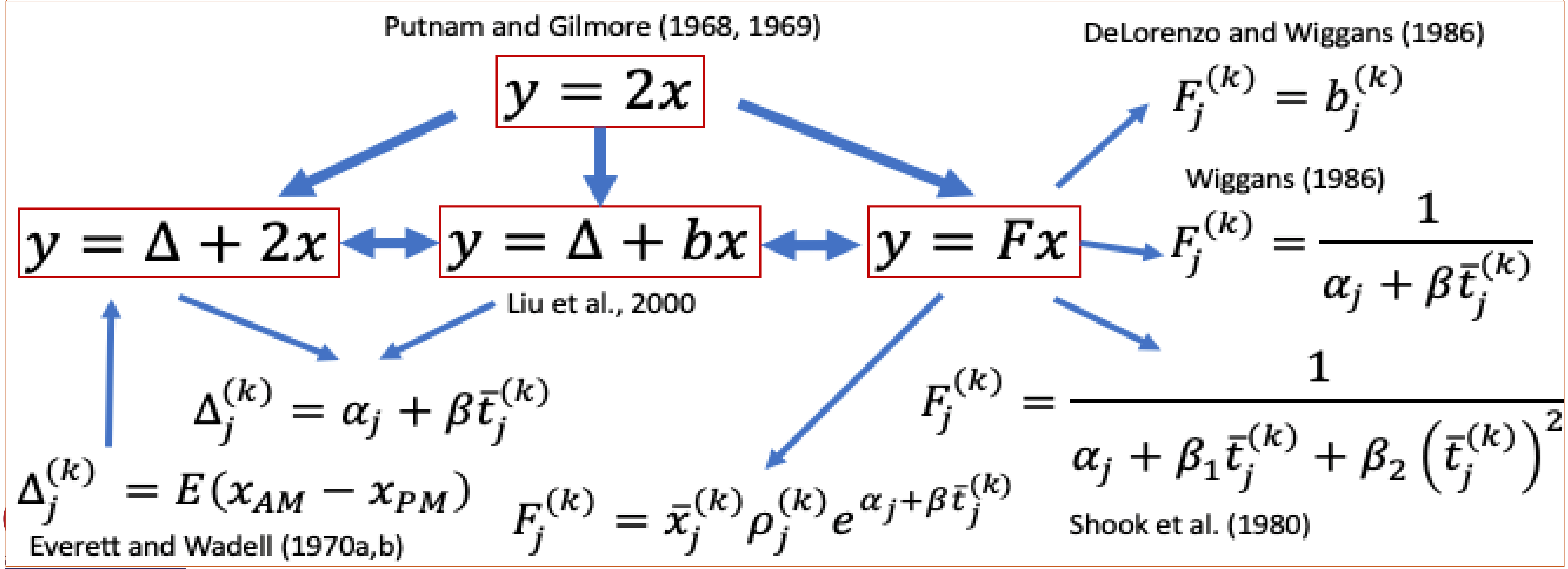
- Either morning or evening milking is sampled alternatively on a test day throughout the lactation period.
- A test-day yield (y) is taken to be twice the partial (AM or PM) yield on that day (x), assuming equal AM and PM milking interval times.
- In reality, however, AM and PM milking intervals are often not equal, and neither are AM and PM milk yields.



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) and Multiplicative (MCF) correction factors.



Wu et al., 2022, 2023

Multiplicative correction factors (MCF)

- MCF, also referred to as ratio factors, are ratio of total daily yield to partial yields.
- Two models: linear regression of daily milk yield on a partial yield without intercept, and linear regression of AM or PM proportional milk yield on milking interval.

$$y = bx + e \Rightarrow \begin{cases} b = \frac{E(y)}{E(x)} = \frac{\frac{1}{n}\sum y}{\frac{1}{n}\sum x} = \frac{\sum y}{\sum x} \\ b = \frac{\sum(y-\bar{y})(x-\bar{x})}{\sum(x-\bar{x})^2} \end{cases}$$

$$E\left(\frac{y}{x}\right) \approx \frac{E(y)}{E(x)}$$

$$\frac{x}{y} = \alpha + \beta t + e \Rightarrow E\left(\frac{y}{x}\right) = E\left(\frac{x}{y}\right)^{-1} = \frac{1}{\alpha + \beta E(t)}$$

CDCB COUNCIL ON DAIRY CATTLE BREEDING

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

Discrete MCFs (Wiggans, 1986)

- MCF have been derived for discrete milking interval classes or bins, say every 15 or 30 minutes, assuming that each MCF is a constant within each class.
- If the bin size is too small, there won't be sufficient data for all bins. But, if the bin size is too big, it can lead to a loss of accuracy due to systematic errors.

$$\frac{x}{y} = \alpha + \beta t + e \Rightarrow E\left(\frac{y}{x}\right) = E\left(\frac{x}{y}\right)^{-1} = \frac{1}{\alpha + \beta E(t)}$$

$$F_{t_{(k)}} = E\left(\frac{y_{t_{(k)}}}{x_{t_{(k)}}}\right) = \frac{1}{\hat{\alpha} + \hat{\beta} t_{(k)}} = \frac{1}{\hat{\alpha} + \hat{\beta}\left(\bar{t}_{(k)} + \left(t_{(k)} - \bar{t}_{(k)}\right)\right)}$$

$$F_{t_{(k)}} = \frac{1}{\left\{\hat{\alpha} + \hat{\beta}\bar{t}_{(k)}\right\} + \left\{\hat{\beta}\left(t_{(k)} - \bar{t}_{(k)}\right)\right\}}$$

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

Continuous MCFs (Wu et al., JDS, 2023)

- We proposed deriving MCF for every possible milking interval time unit, denoted as t^*, defined on moving windows.
- Assuming the data are sufficient and well-balanced:

$$F_{t^*} = \frac{1}{E\left(z|t=t^*,\widehat{\theta}\right)} = \frac{1}{\hat{z}_{t^*}}$$

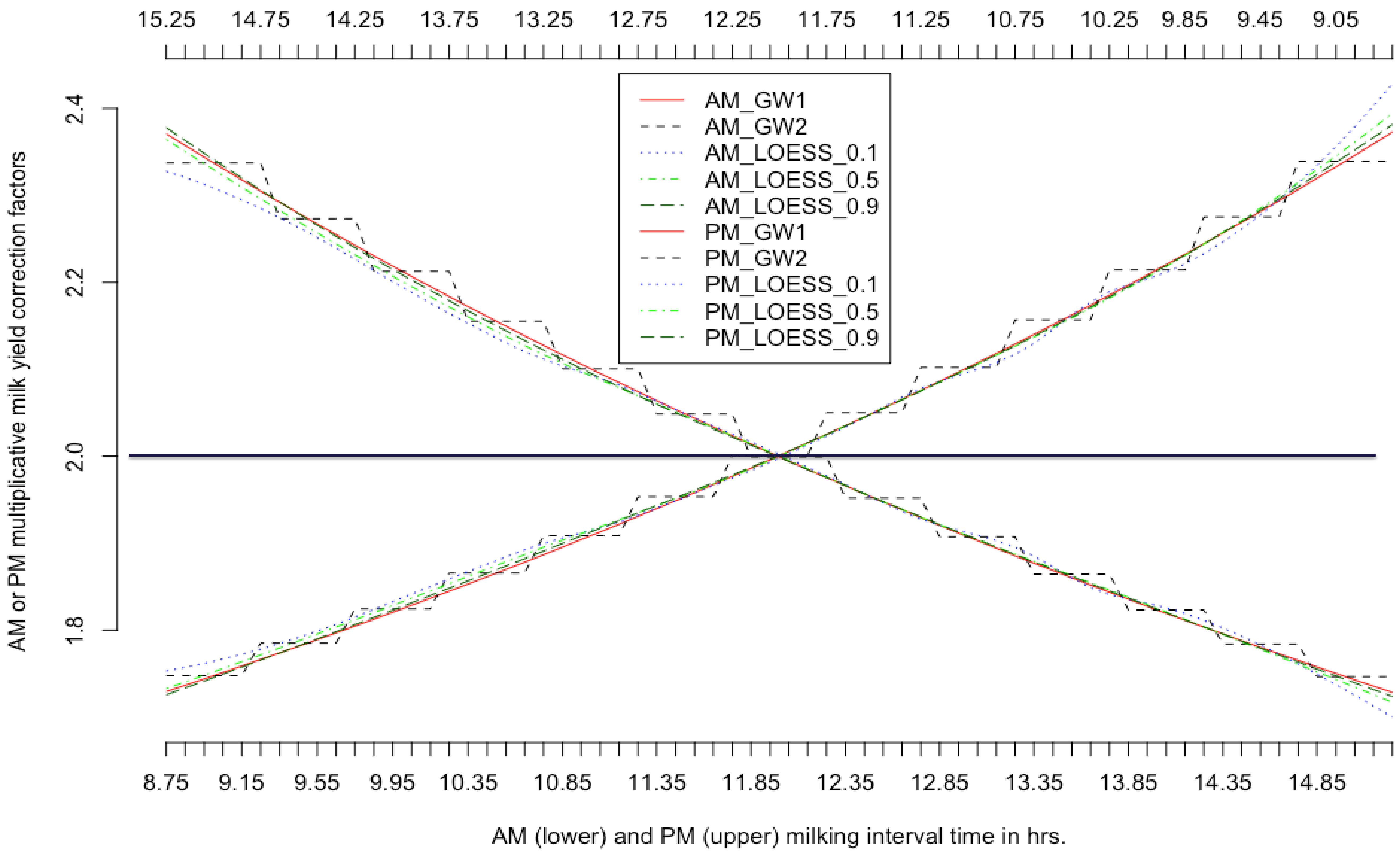
- Otherwise, apply local regression

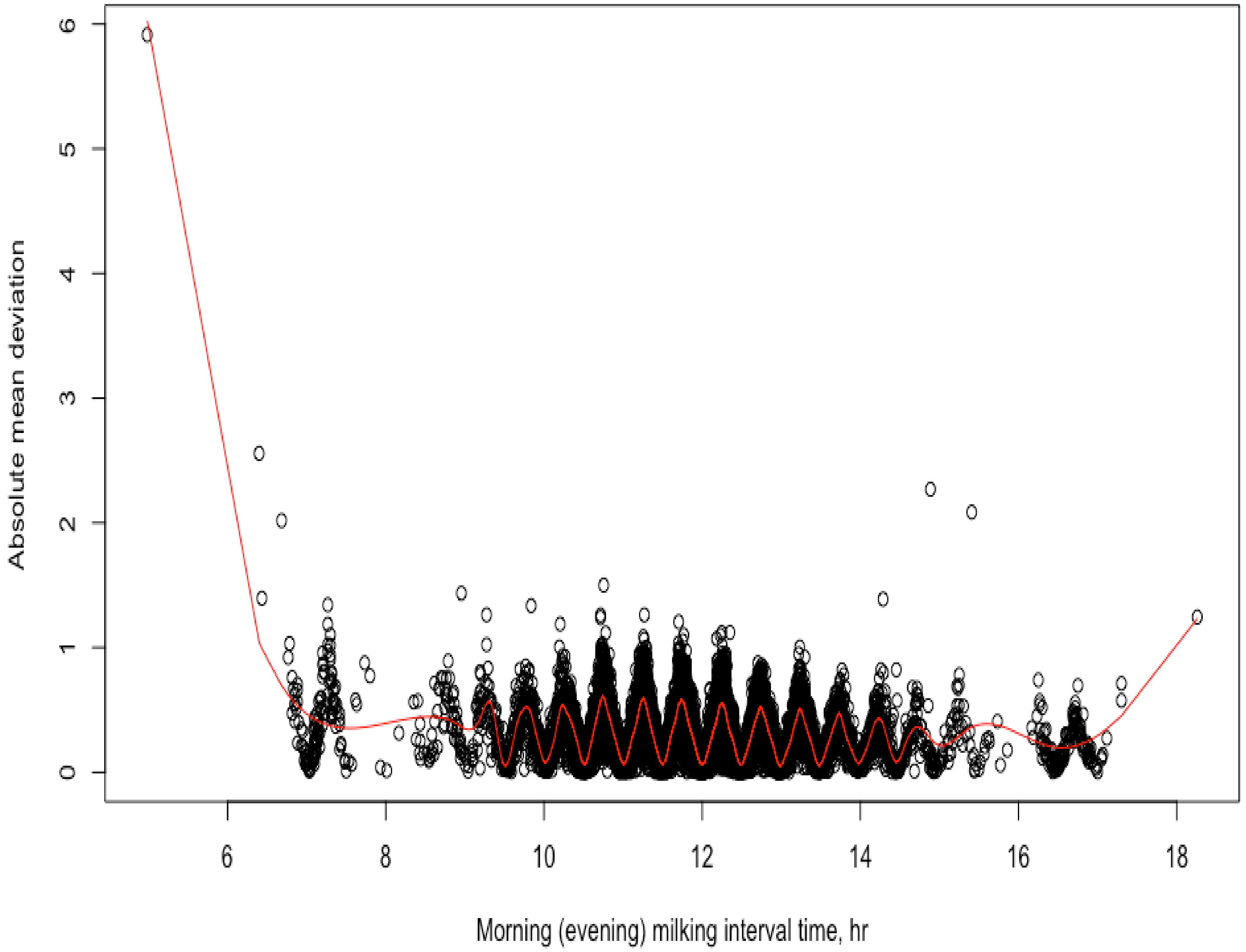
$$F_{t^*} = \frac{1}{E\left(f\left(\mathbf{w}'\mathbf{z}|\boldsymbol{\theta},t\in N(t^*)\right)\right)}$$

CDCB
COUNCIL ON DAIRY CATTLE BREEDING

Discrete vs. continuous MCFs (Simulation)

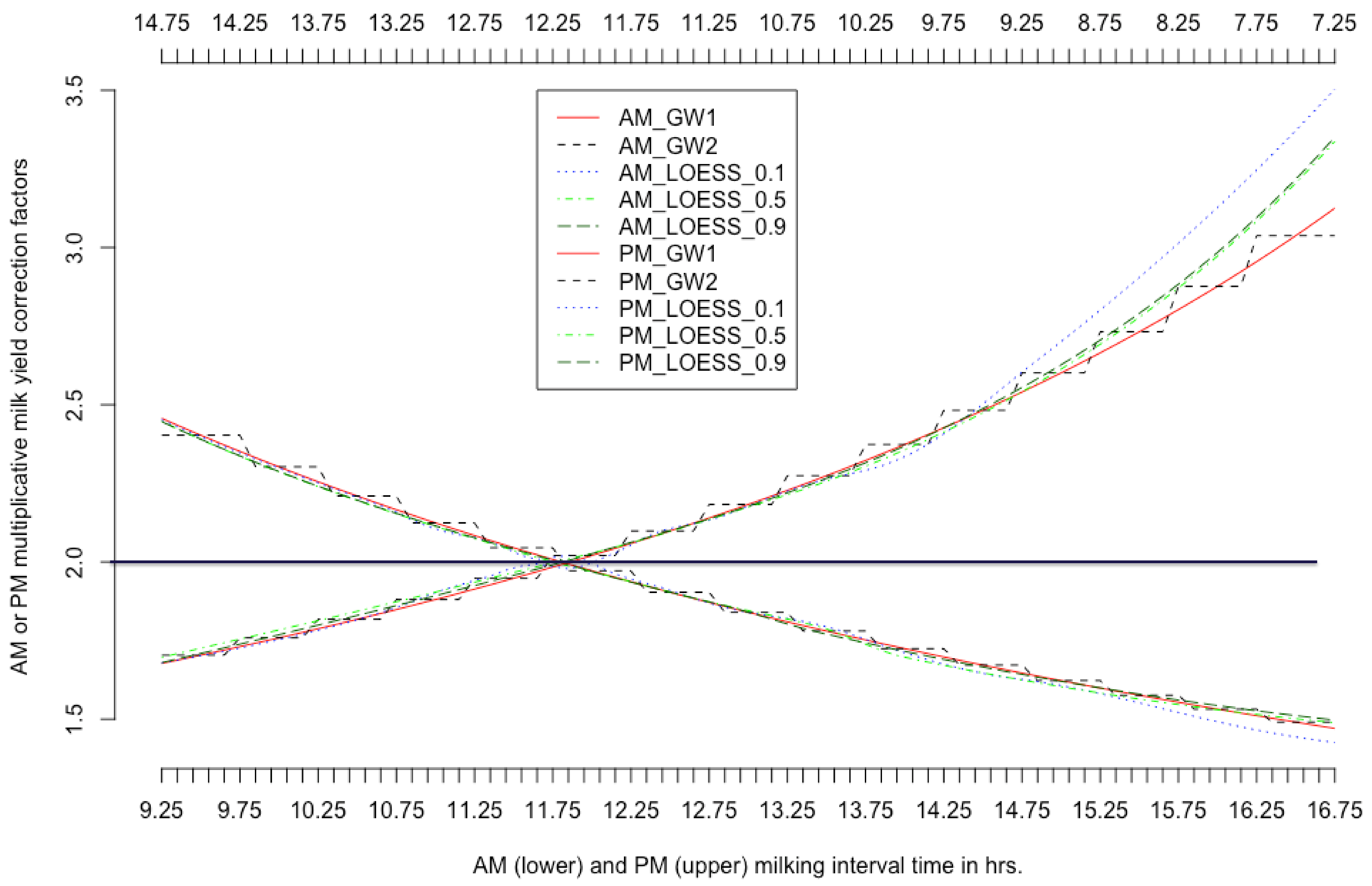
- The distribution of traditional MCF is discrete in the plots whereas the distribution of general forms of MCFs, derived for 0.1-hour intervals, are visually continuous.
- LOESS implements local regression with a span parameter defining the "effective neighborhood", that is, the portion of data points used. Weights vary according to their distance from each local central time point in milking interval time

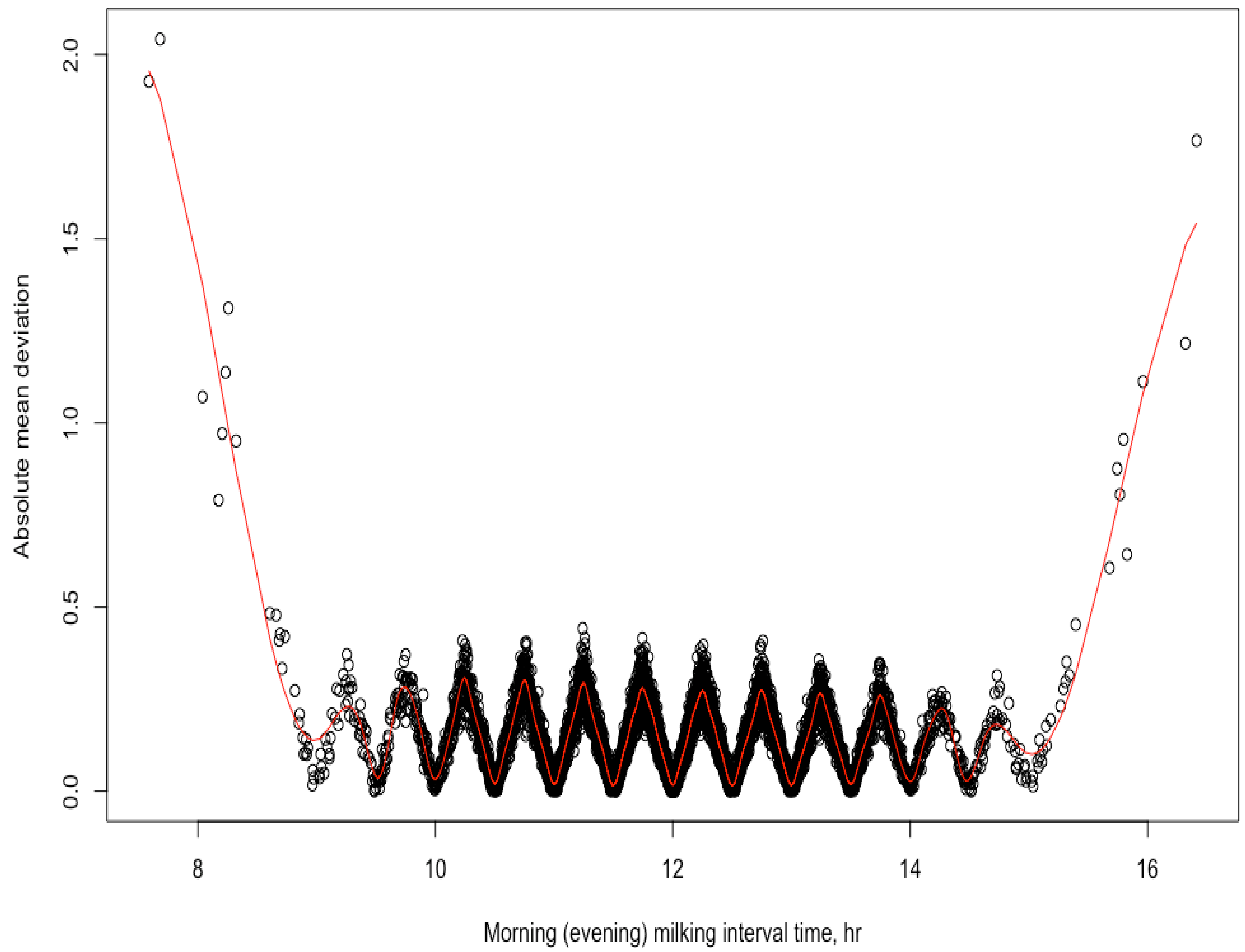




Discrete vs. continuous MCFs (Holstein)

- We obtained similar results and reached similar conclusions when using a US Holstein milking test-day data.

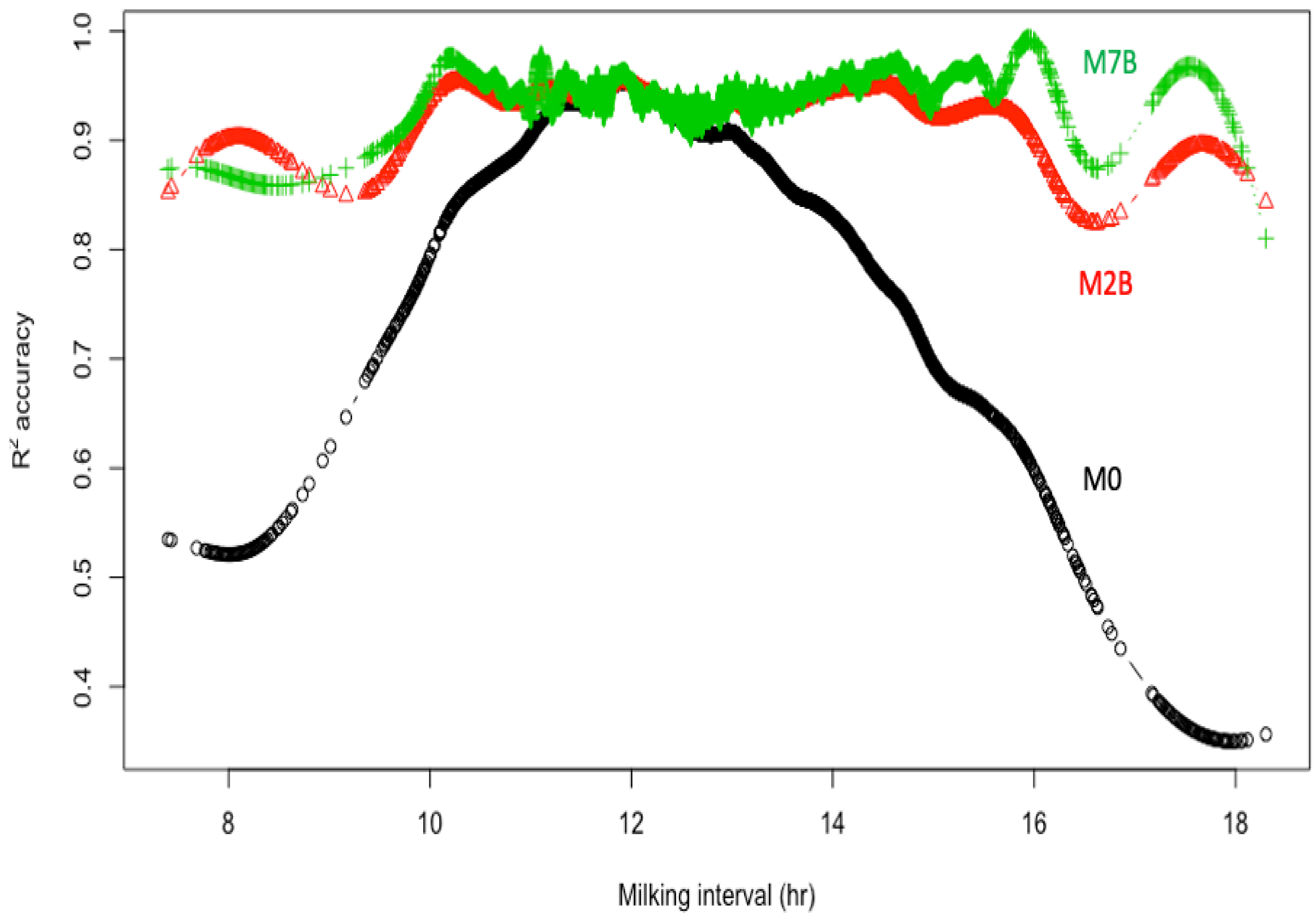




Comparing models and strategies

- All models performed similarly with equal AM/PM milking intervals, but they vary substantially with unequal milking intervals . MCF model is slightly better than ACF models, both surpass the 2X method.
- Biases were lower and R^2 accuracies were higher with general ACF/MCF, compared to classic ACF/MCF for every 30 minutes.

Method	MSE	R2 Accuracy
M0: Baseline (2X)	22.8	0.821
M2A: ACF_G	11.3	0.902
M2B: ACF_D	11.4	0.902
M5: Shook (1980)	11.0	0.905
M6: D-W (1986)	11.0	0.904
M7A: Wiggans_G	10.9	0.905
M7B: Wiggans_D	11.0	0.904
M8A: Wu_G	10.1	0.912
M8B: Wu_D	11.0	0.910



Take-home messages

- Following the traditional methods to estimate test-day milk yields, choosing the right size for milking interval classes or bins is challenging. On one hand, if we go too small with our bin size, we might not have enough data for each category. On the flip side, making the bins too large can lead to systematic errors.
- But there's a promising alternative: by using local regression, we can derive continuous yield correction factors. This approach is flexible and has shown to boost the accuracy of our estimated test-day milk yields.