SIMULATION OF DESIGN-UNBIASED POINT-TO-PARTICLE SAMPLING AND ALTERNATIVES ON PLANTATION ROWS

Thomas B. Lynch
Dept. Natural Resource Ecology and Management
Oklahoma State University

David Hamlin
Campbell Global

Mark J. Ducey
Dept. Natural Resources and the Environment
University of New Hampshire, Durham

Joint SOMENS & NEMO Meeting , Nov. 2-4, 2014
Raleigh, NC

Plantation row sampling and design unbiased fixed 2*k* (particle sample size) sampling on lines

- Mark Ducey NEMO 2012 – Design unbiased point to particle sampling on lines.
- Ducey - Given in the context of a generalized forest, not plantation. Based on distance along line between trees in strips.
- Bruce Borders SOMENS 2012 – Plantation row sampling. Row lengths with remote sensing, ground sampling for sample trees
- Borders - Fixed length plots with variable tree numbers or fixed tree numbers with variable length plot. Latter not design unbiased. Ratio estimation.
- Combine. Ducey's estimator can be adapted to plantation rows. Row lengths determined by remote sensing.

Ducey's (2012) Estimator

$$\hat{Y}_D = \sum_{i=1}^{N} [w_i^{+k}\delta^{+k}(z) + w_i^{-k}\delta^{-k}(z)]y_i$$

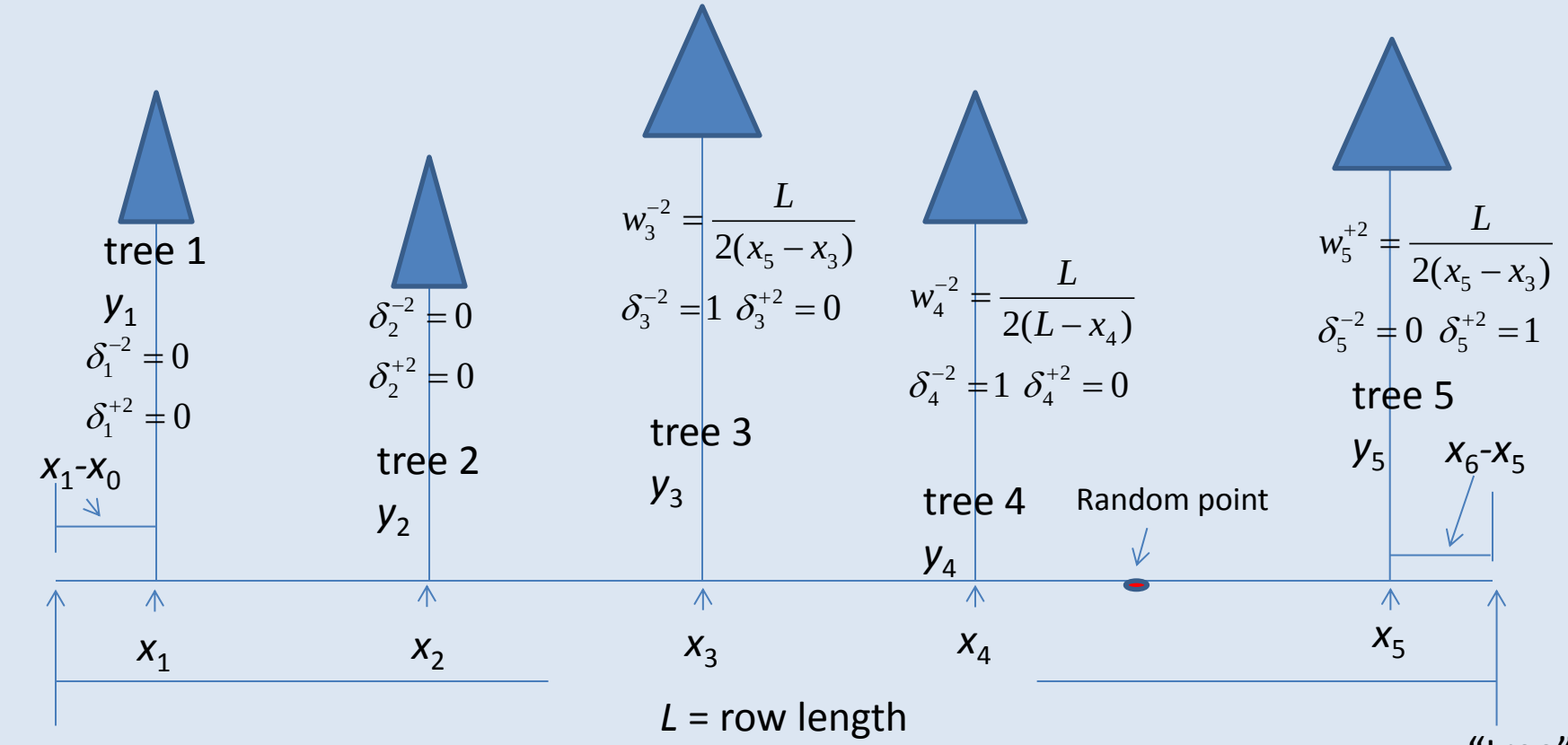
where

$$w_i^{+k} = \begin{cases} \dfrac{L}{2(x_i - x_{i-k})} \text{ if } x_i < L \\ \dfrac{L}{(L - x_{i-k})} \text{ if } x_i = L \end{cases} \qquad w_i^{-k} = \begin{cases} \dfrac{L}{2(x_{i+k} - x_i)} \text{ if } x_i > 0 \\ \dfrac{L}{x_{i+k}} \text{ if } x_i = 0 \end{cases}$$

$$\delta_i^{+k} = \begin{cases} 1 \text{ if } x_{i-k} \le z < x_i \\ 0 \text{ otherwise} \end{cases} \qquad \delta_i^{-k} = \begin{cases} 1 \text{ if } x_i \le z < x_{i+k} \\ 0 \text{ otherwise} \end{cases}$$

L = row length, y_i = attribute of tree i, x_i = distance along row to tree i, z = random point location on row.

Ducey's Estimator for $k = 2$ trees 3, 4 and 5 are in this sample and "tree" 6 is a "false particle"

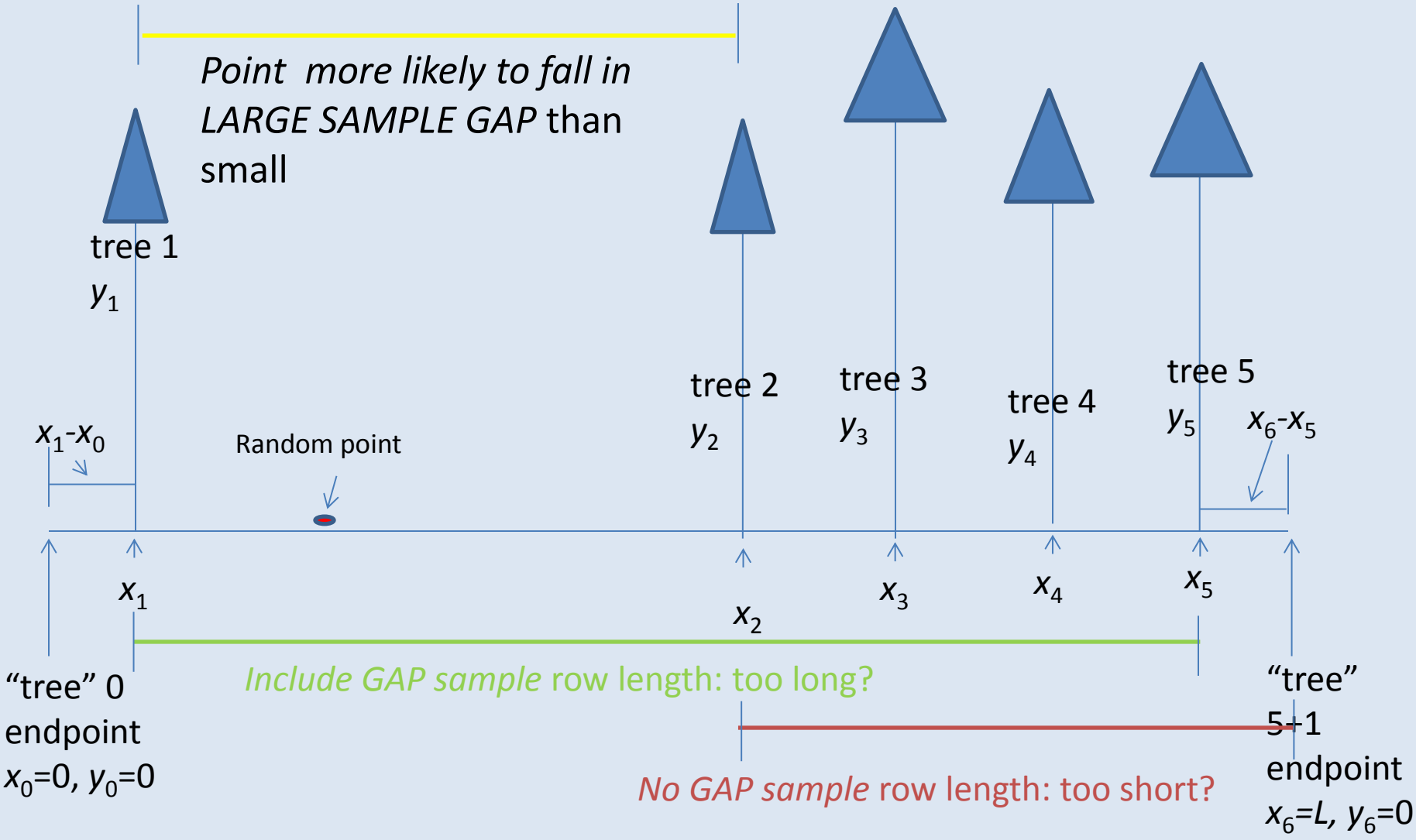


Pr(select trees 4 and 5) = $(x_5\text{-}x_4)/L$

Implementation

- Find plantation row length (remote sensing)
- Plantation -one continuous row, or strata
- Randomly locate sample point *along row*
  NOT nearest tree a 2D random point
- Nearest *k* trees right and left of random point. Measure distances & attributes (e.g., basal area, volume) of sample trees
- Tape Breast Height (BH) midpoint “left -most” tree to “right-most” tree.
- Record distance from left-most tree to each tree BH midpoint along the row. Or laser, vertex.
- Apply Ducey’s estimator. Average *n* values of Ducey’s estimator from *n* randomly sample points
- Estimate sample variance - SRS formula

Alternative Estimators - Measuring one distance including 4 sample trees: to gap or not to gap?



Alternatives: "Mean of Ratios"

$$\hat{Y}_{MEAN\ OF\ RATIOS} = L \frac{1}{\#\text{sample rows}} \sum_{\#\text{sample rows}} \frac{\text{sum attributes on sample row}}{\text{length of sample row}}$$

Including the sample gap:

$$\hat{Y}_{G-MR} = \frac{L}{n} \sum_{h} \frac{\sum_{i=h-1}^{h+2k-1} y_i}{x^*_{h-1} - x^*_{h+2k}}$$

x^*_{h-1} sample from the x_i just left of sample point ground location

y_i = attribute of tree i

n = number of sample points h

$2k$ = fixed number of sample trees

$\hat{Y}_{G-MR}$ mean of ratios estimator of the total that includes the sample gap

Not including the sample gap:

$$\hat{Y}_{NG-MR} = \frac{L}{n} \sum_{h} \frac{\sum_{i=h}^{h+2k} y_i}{x^*_{h} - x^*_{h+2k+1}}$$

x^*_{h} sample from the x_i just right of sample point ground location

$\hat{Y}_{NG-MR}$ mean of ratios estimator of the total that does not include the sample gap

Alternatives: "Ratio of Means"

$$\hat{Y}_{RATIO\ \ OF\ MEANS} = L\frac{\sum_{\#\,sample\,rows}\text{sum of attributes on sample row}}{\sum_{\#\,sample\,rows}\text{length of sample row}}$$

Including the sample gap:

$$\hat{Y}_{G-RM} = L\frac{\frac{1}{n}\sum_{h}\sum_{i=h-1}^{h+2k-1} y_i}{\frac{1}{n}\sum_{h}\left(x^*_{h-1} - x^*_{h+2k}\right)}$$

x^*_{h-1} sample from the x_i just left of sample point ground location

y_i = attribute of tree i

n = number of sample points h

$2k$ = fixed number of sample trees

$\hat{Y}_{G-RM}$ mean of ratios estimator of the total that includes the sample gap

Not including the sample gap (Borders 2012):

$$\hat{Y}_{NG-RM} = L\frac{\frac{1}{n}\sum_{h}\sum_{i=h}^{h+2k} y_i}{\frac{1}{n}\sum_{h}\left(x^*_{h} - x^*_{h+2k+1}\right)}$$

x^*_h sample from the x_i just right of sample point ground location

$\hat{Y}_{NG-RM}$ mean of ratios estimator of the total that includes the sample gap

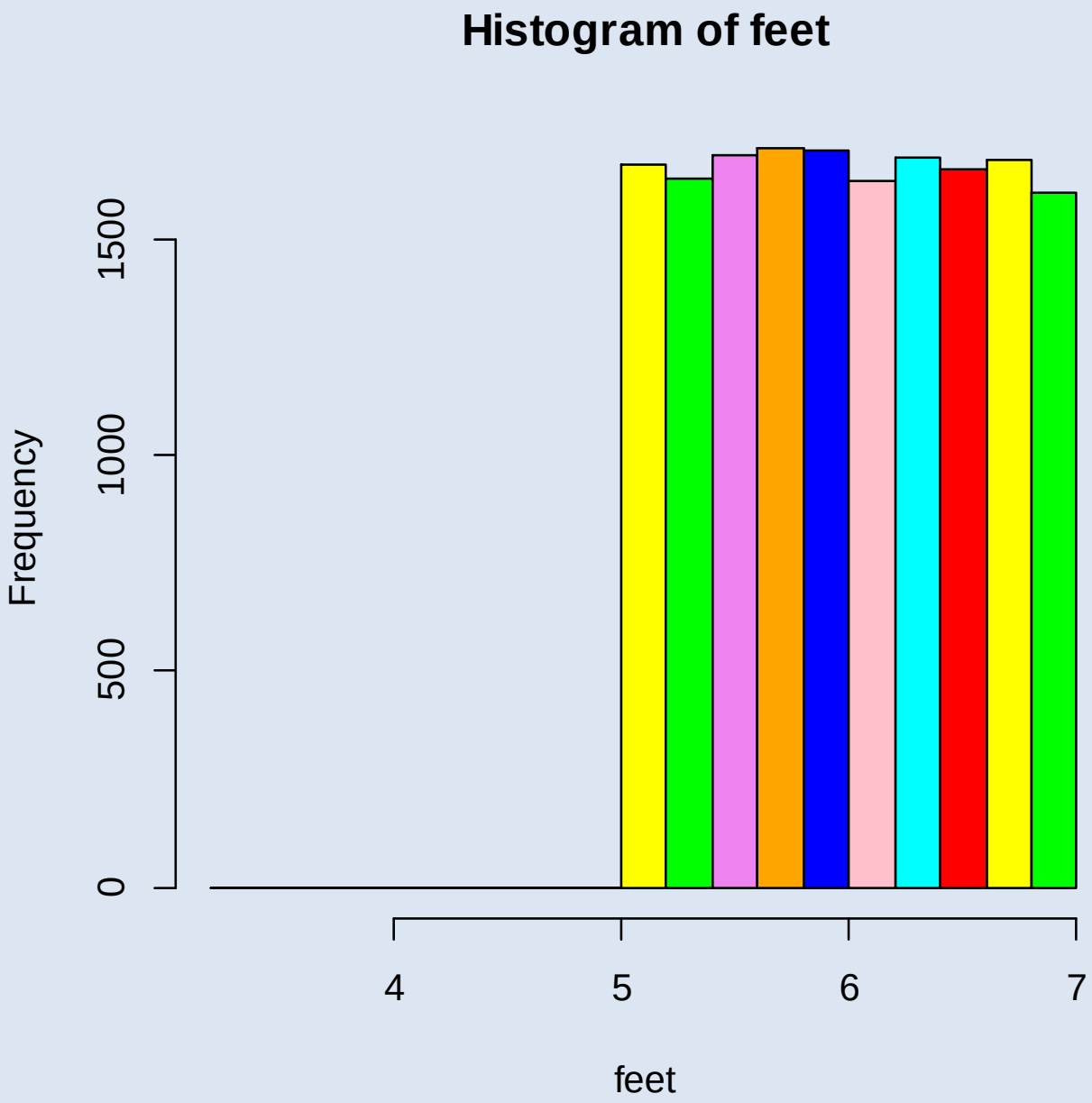
Estimator simulation and variance computation

- Two simulated plantation rows: “Gappy” and “Not Gappy”
- “Not Gappy” tree positions have a small amount of random error
- “Gappy” – large row gaps – thinning/mortality
- Ducey and Mean of Ratios – exact bias/variances of estimators computed on each row
- Ratio of means – R Simulation script: 1 million random row locations used to estimate bias/variance of estimators
- Fixed-length rows – R Simulation script: 1 million random row locations used to estimate variance of estimator

Not Gappy row population

- Target row length: 100,000 feet
- Mean inter-tree distance: 6 feet
- Standard deviation inter-tree distance: 3.5 feet
- Minimum inter-tree distance: 3.6 feet
- Maximum inter-tree distance: 7 feet
- Total number of trees: 16,667

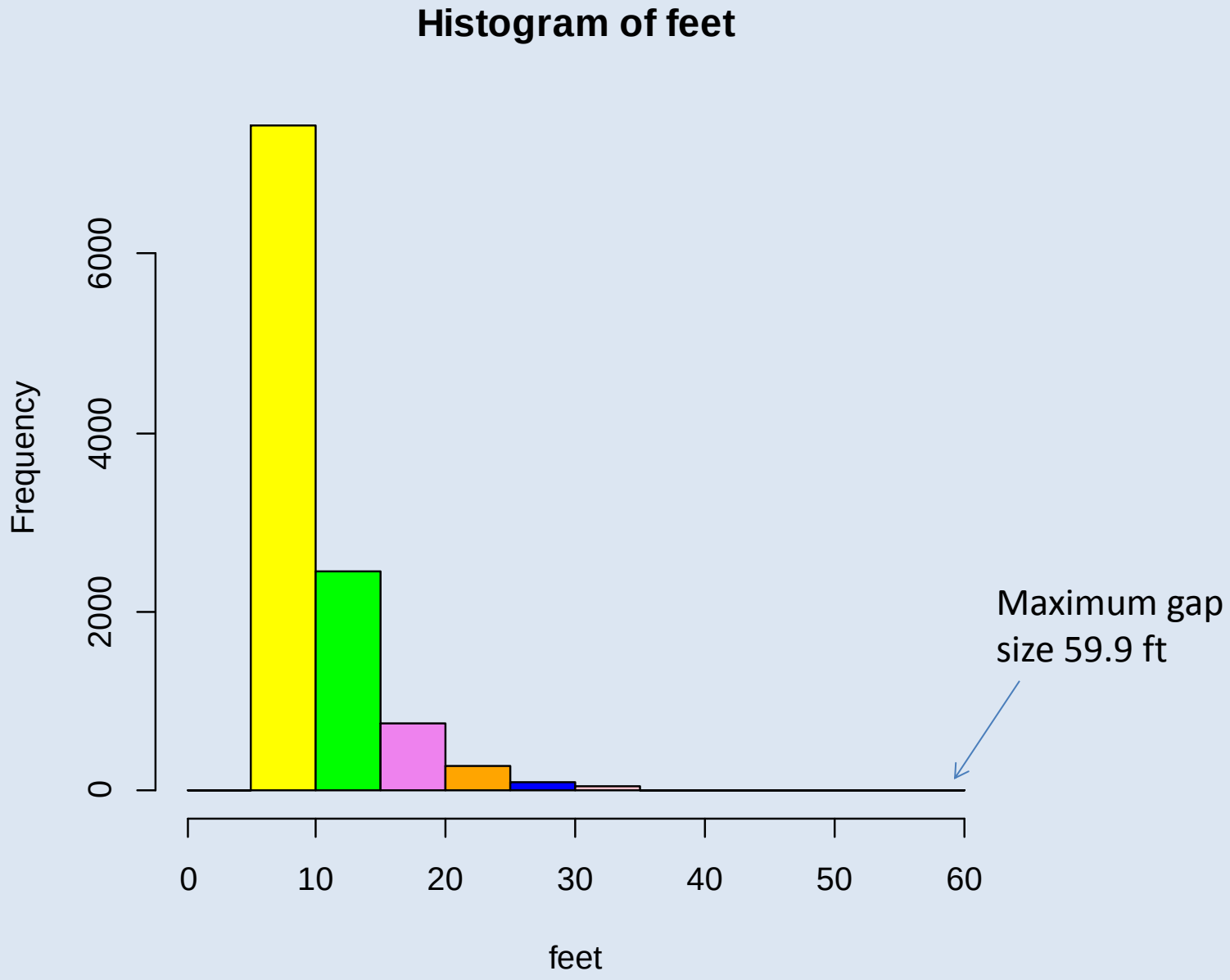
Gap size histogram for “Not Gappy” row



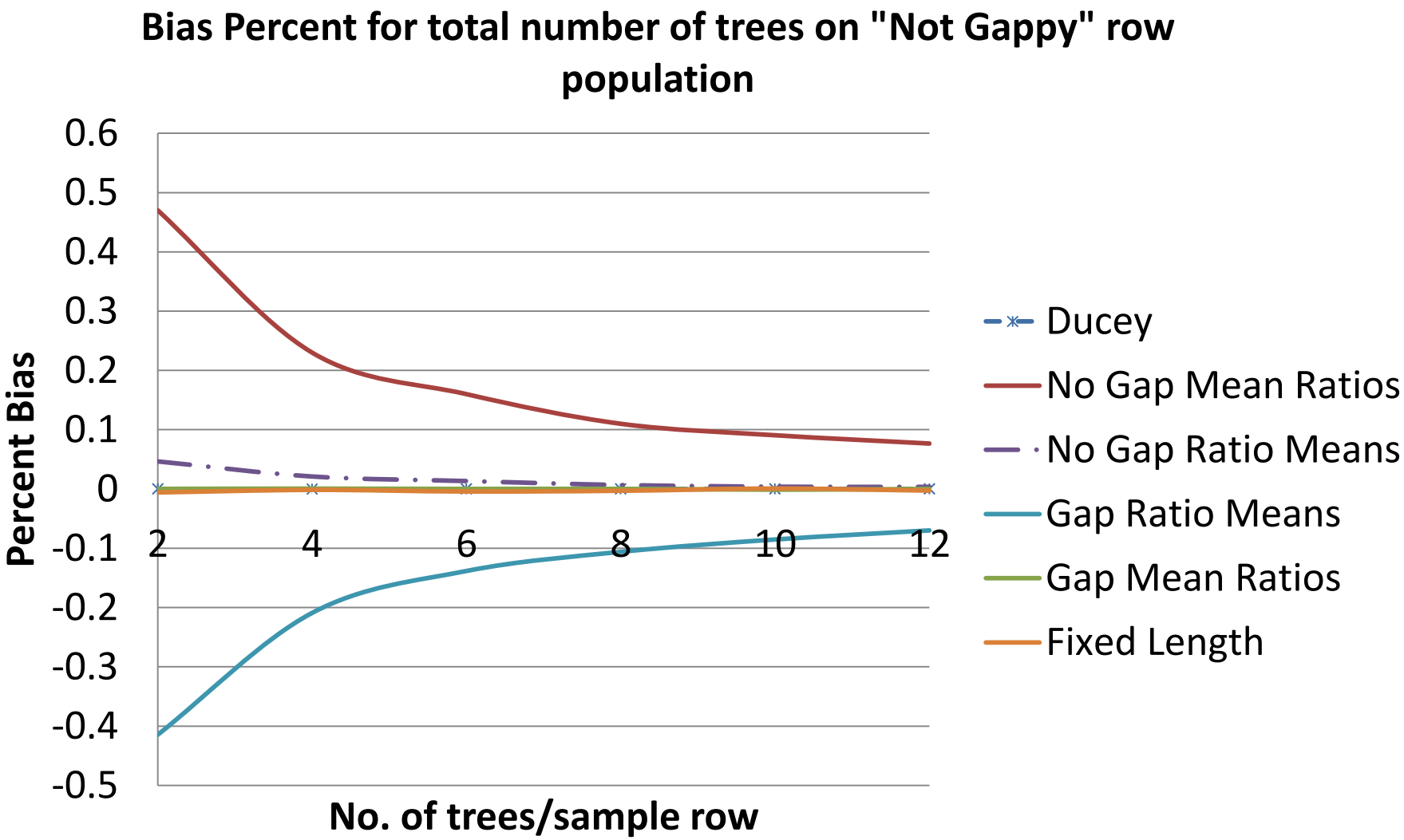
Gappy row population

- Target row length: 100,000 feet
- Mean inter-tree distance: 9 feet
- Standard deviation inter-tree distance: 7.4 feet
- Minimum inter-tree distance: 3.6 feet
- Maximum inter-tree distance: 59.9 feet
- Total number of trees: 11,109

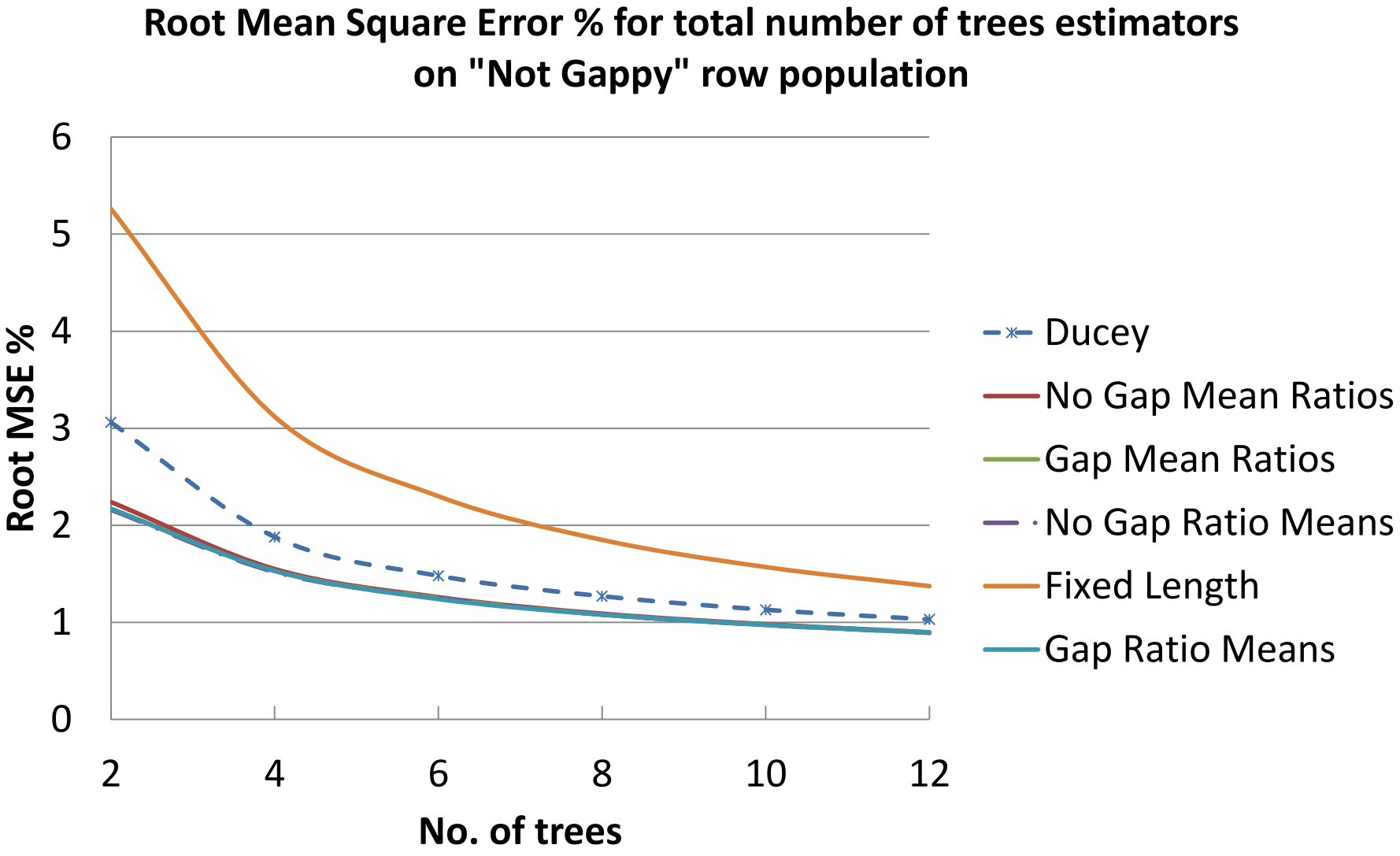
Gap size histogram for “Gappy” row



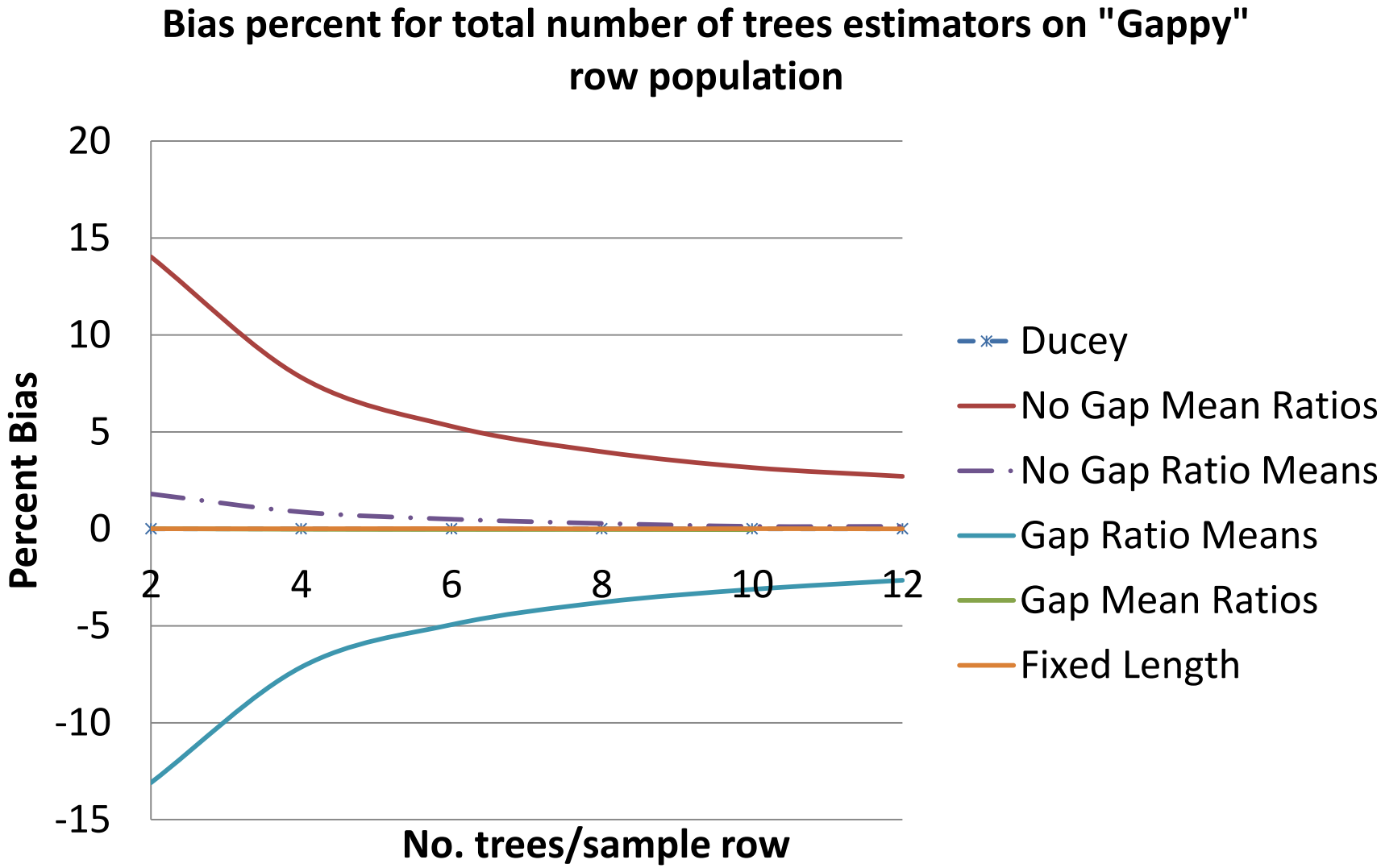
Not Gappy Bias, estimators for total number of trees based on a sample size of *n*=10 row locations, 1 million simulations or exact



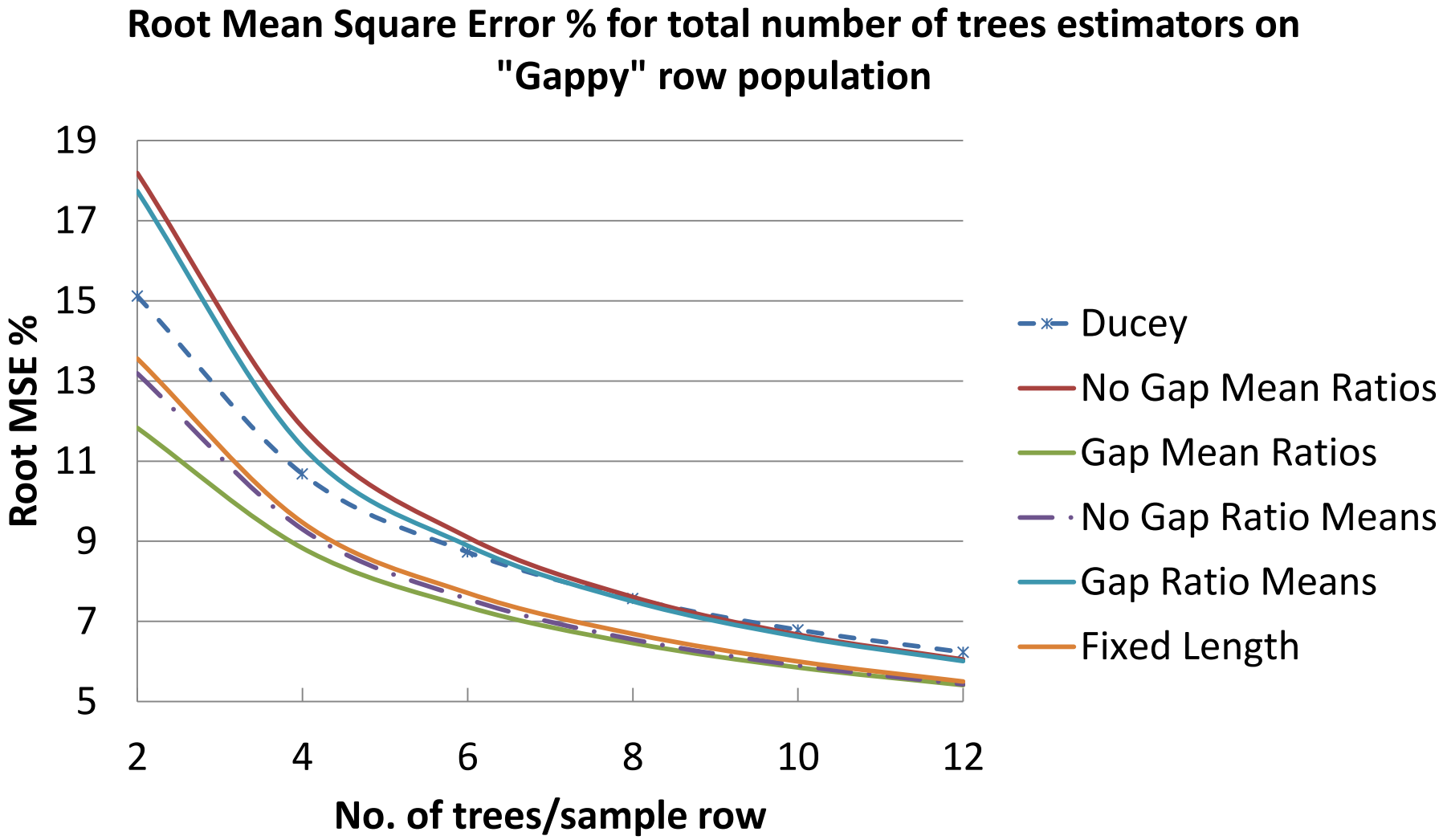
Not Gappy Root Mean Square Error%, estimators for total number of trees based on a sample size of *n*=10 row locations, 1 million simulations or exact



Gappy Bias, estimators for total number of trees based on a sample size of *n*=10 row locations, 1 million simulations or exact



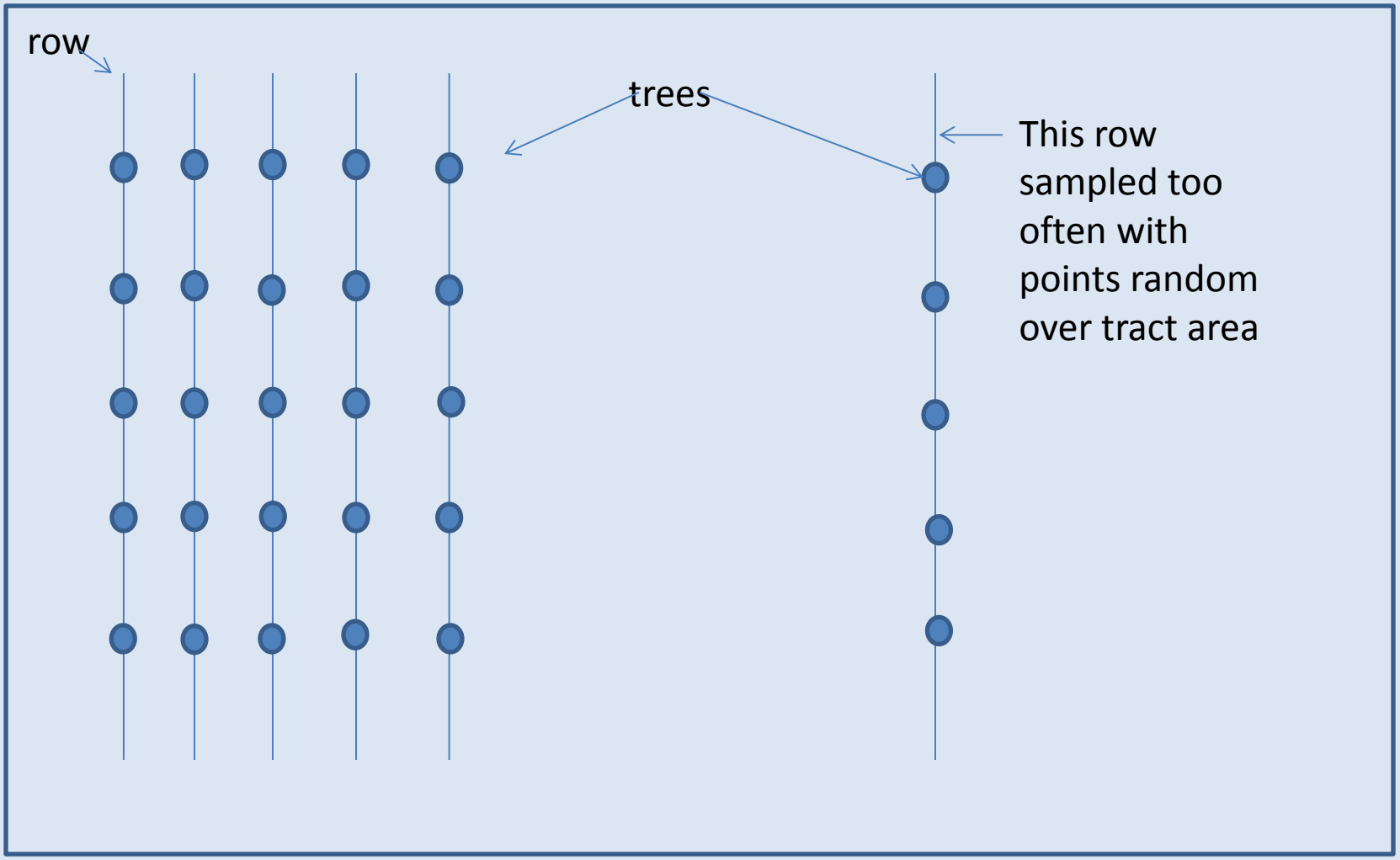
Gappy Root Mean Square Error%, estimators for total number of trees based on a sample size of *n*=10 row locations, 1 million simulations or exact



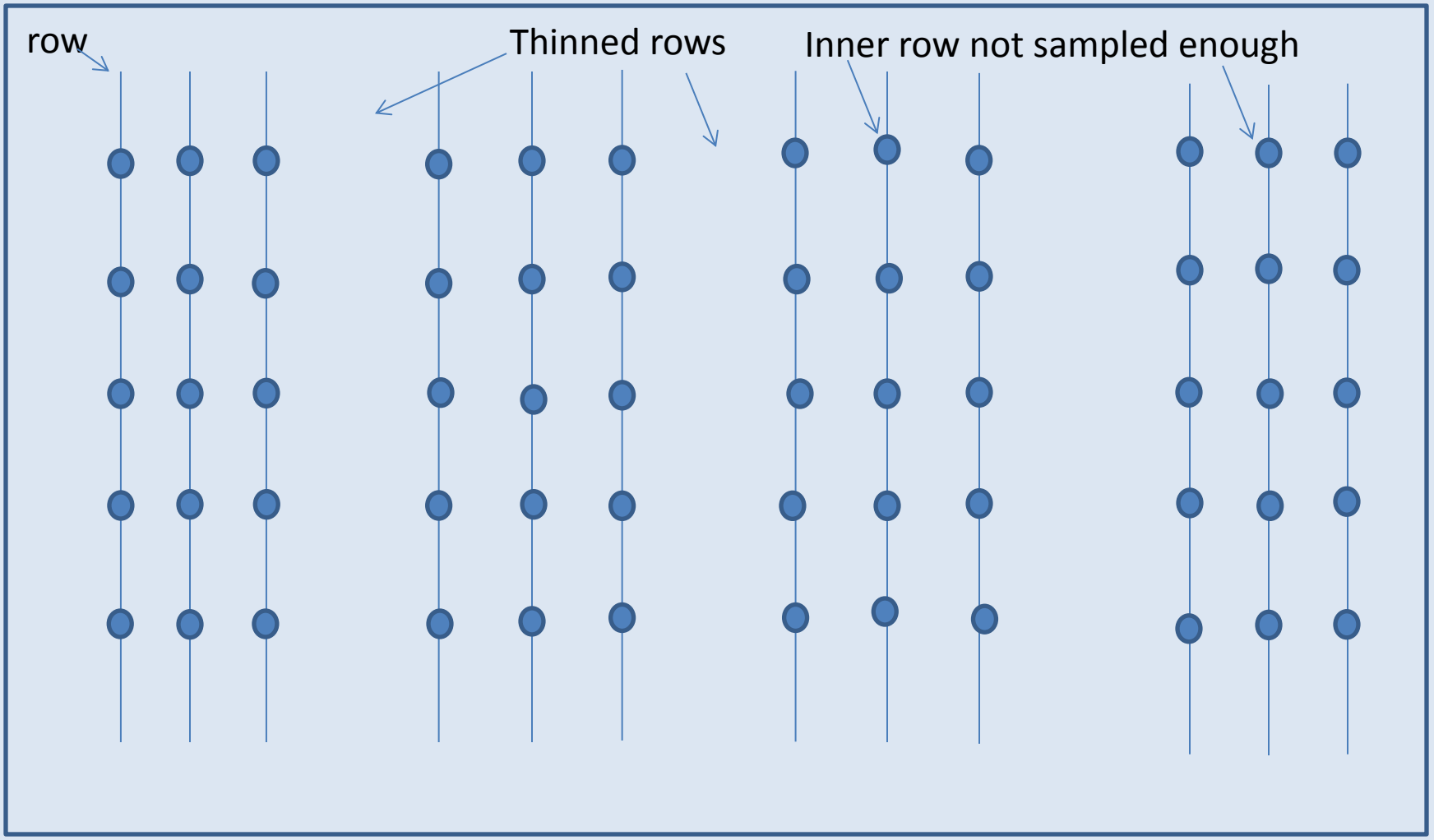
Simulation Summary

- Not Gappy row – all methods had bias < 0.5%
- Gappy row – The “No Gap Mean of Ratios” and the “Gap Ratio of Means” had substantial bias >5% for 6 sample trees or less, down to 2.5% for 12 trees
- Fixed row length – highest RMSE% for “Not Gappy” but 3rd lowest for “Gappy” row
- “Gap Mean of Ratios” and “No Gap Ratio of Means” - low bias and RMSE% both Gappy and Not Gappy row.
- Ducey not most precise method but unbiased any spatial arrangement. Precision not substantially greater.

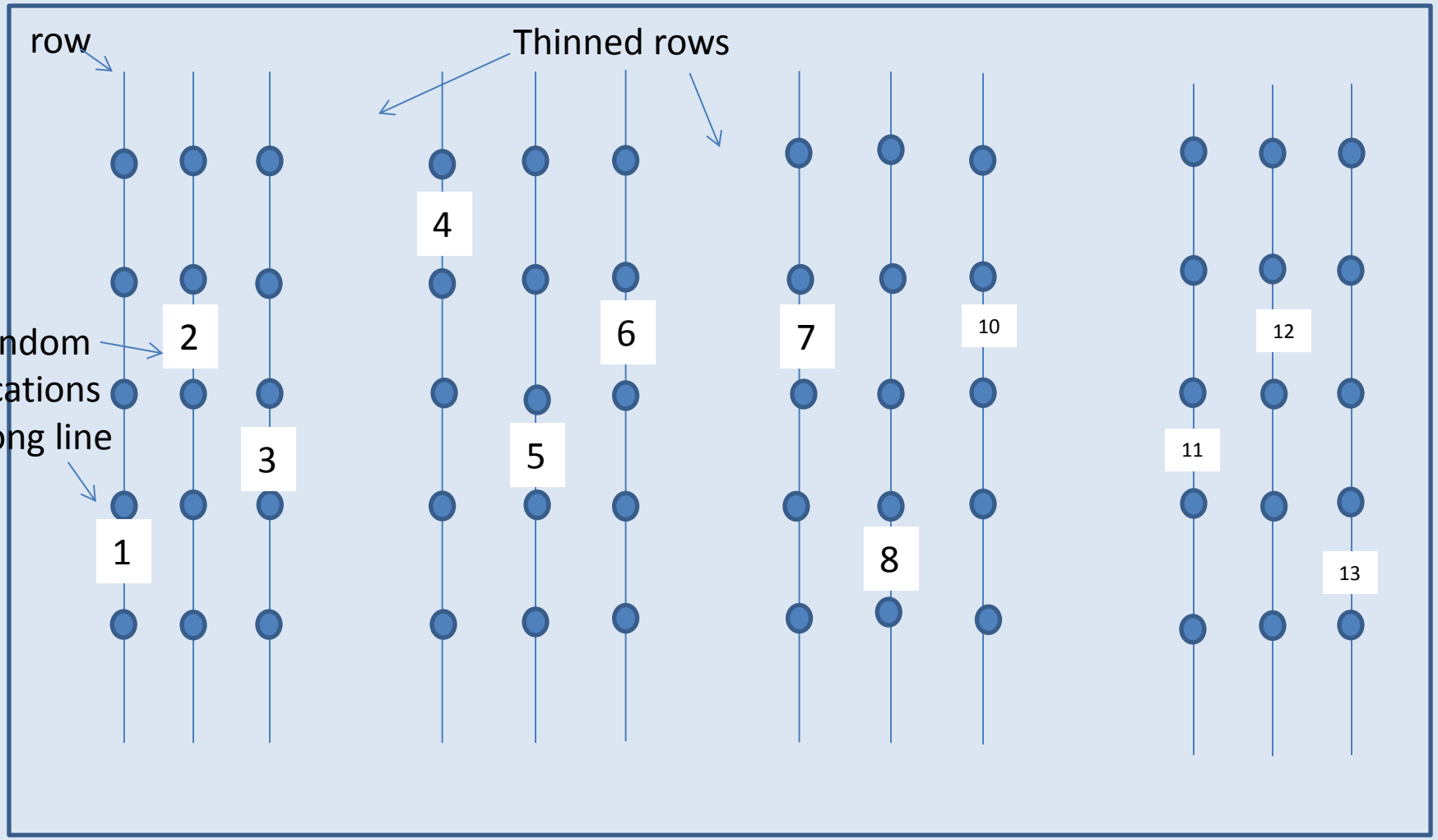
Random number location along line vs. closest random point in tract area



Random number location along line vs. closest random point in tract area



Random number location along line (uniform on (0, L)) – generate GPS location? And photo overlay?



Conclusions

- RMSE – Gap Mean of Ratios and No Gap Ratio of Means best
- Substantial BIAS Gappy row – Gap Ratio of Means and No Gap Mean of Ratios
- Ducey's method - guaranteed unbiased any spatial distribution - slightly higher RMSE
- Place random (or systematic) sample points along ROW
- Random point in 2D – possible bias row thinned plantations.