Evaluation of sampling methods and statistical models for predicting fruit load on individual *Coffee arabica* trees.

Sebastian Castro
Nathan Warner
Manuel Blanco
Mark Chandler
Thomas Dietsch
Natalia Ureña
Ernesto Méndez

Models which predict total fruit load in a coffee tree exist. Although conceptually sane, they are based on the assumption that the number of fruit per branch (fruit lateral^{-1}) can be well described using a normal distribution. This assumption may not always be correct. The questions guiding our research were: 1) which is the most suited distribution to describe the fruit lateral^{-1} in a coffee tree, 2) are fruit lateral^{-1} affected by the location of the lateral?,3) which is the error associated to the mean fruit lateral^{-1} estimated with varying sampling intensities and schemes? and, 4) how accurate are least square regression models in predicting the tree fruit load?

Data was collected during 2008 in 15 coffee plantations found in the Los Santos Region, Costa Rica. In 60 coffee plants, the count of fruits per lateral was recorded and grouped by lateral.

The *Kolmogorov –Smirnov* test indicated that the observed distribution of fruit lateral^{-1} did not significantly differ from a log normal distribution with the observed parameters in 58 out of 60 plants. The *D* statistic obtained from this test ranged from 0.0779 to 0.333. Linear mix models indicated that the location of the lateral significantly affected fruits lateral^{-1}. This relationship was best described by a quadratic polynomial function. A computer program designed for resampling without replacement allowed us to describe the error of the mean fruits lateral^{-1} estimated using different sampling intensities. Sampling schemes which considered the location of the lateral showed lower error rates.

Linear square regressions meant to predict fruit tree^{-1} were fitted. These were performed using different sampling intensities and the number of productive laterals per tree. R^2 of these models ranged from 0.81 to 0.92. Our research contributes with improved sampling schemes and statistical models which build up in predictive schemes developed in the past.