Forecasting the Annual Almond Crop Production in California

Project No.: **13-ABCBOD1-Wang**

Project Leader: Dr. Jane-Ling Wang
Department of Statistics
UC Davis
One Shields Avenue
Davis, CA 95616
530.752.2361
janelwang@ucdavis.edu

Project Researcher: Dr. Neil H. Willits
Senior Statistician, Statistical Laboratory
Department of Statistics
UC Davis
One Shields Avenue
Davis, CA 95616
530.554.9435
nhwillits@ucdavis.edu

Project Personnel and Cooperators:
Vic Tolomeo, USDA-NASS

Objectives:

The primary focus of this project is to answer the following questions:

1. What are the statistical operating characteristics of the existing methods for forecasting California almond production?

2. What are the changes that can be made to the existing methods for forecasting California almond production that could improve their accuracy and precision?

3. Can the Nonpareil production be forecast with better accuracy and precision?

Background and Discussion:

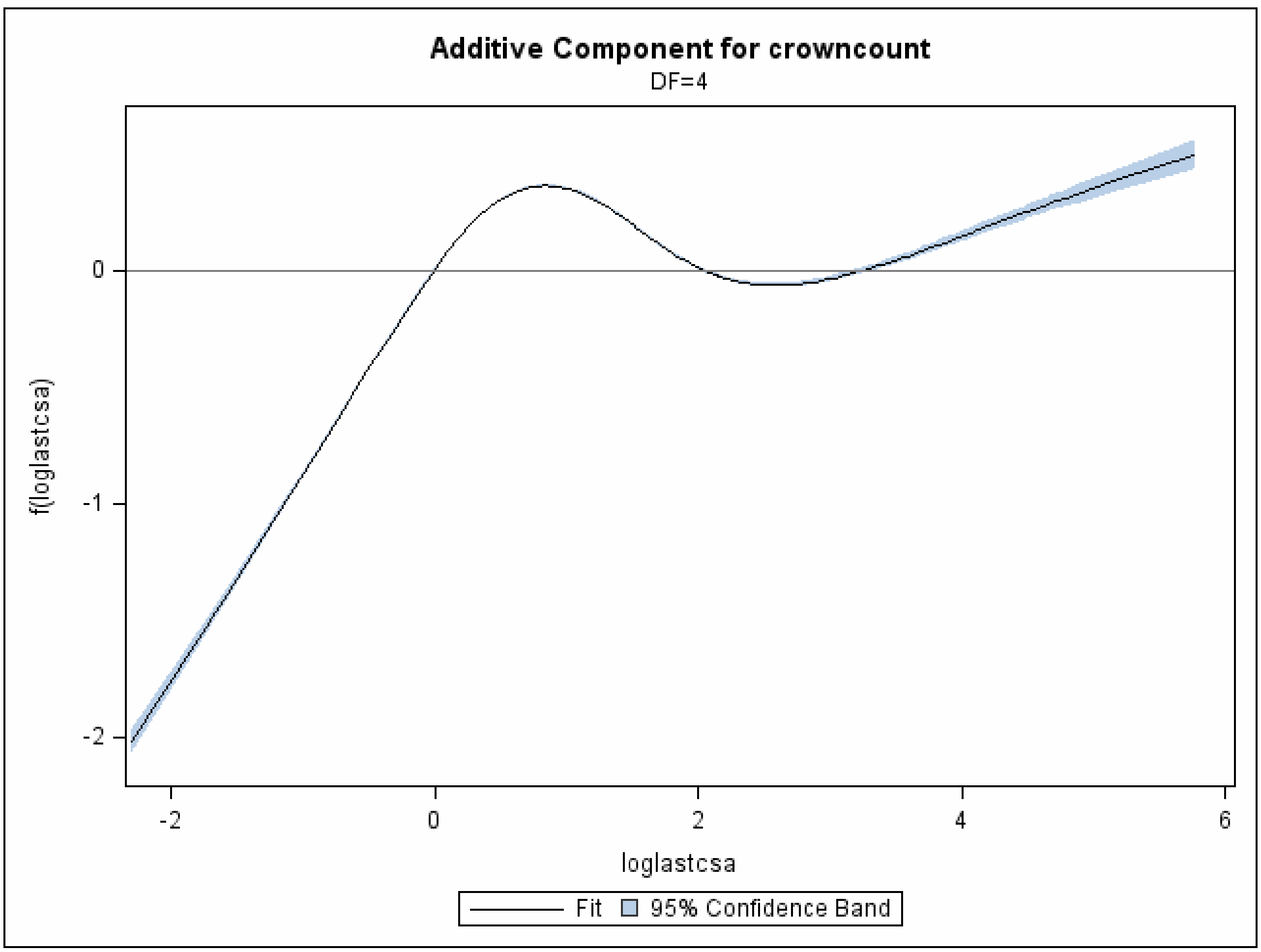
California is the largest almond producer in the world and the almond crop is a vital part of California's economy. It is thus crucial to have accurate predictions of the annual total crop for effective marketing and distribution of the crop.

The National Agricultural Statistics Service of the USDA (NASS-USDA) has developed and uses a plan to forecast the annual crop production together with its associated prediction interval. Previous predictions have been satisfactory but the Almond Board of California would like to improve the precision of the forecast intervals and develop an improved plan to predict Nonpareil production.

The research in the past year (second year of a three year project) focused on the second and to some extent the third objective.

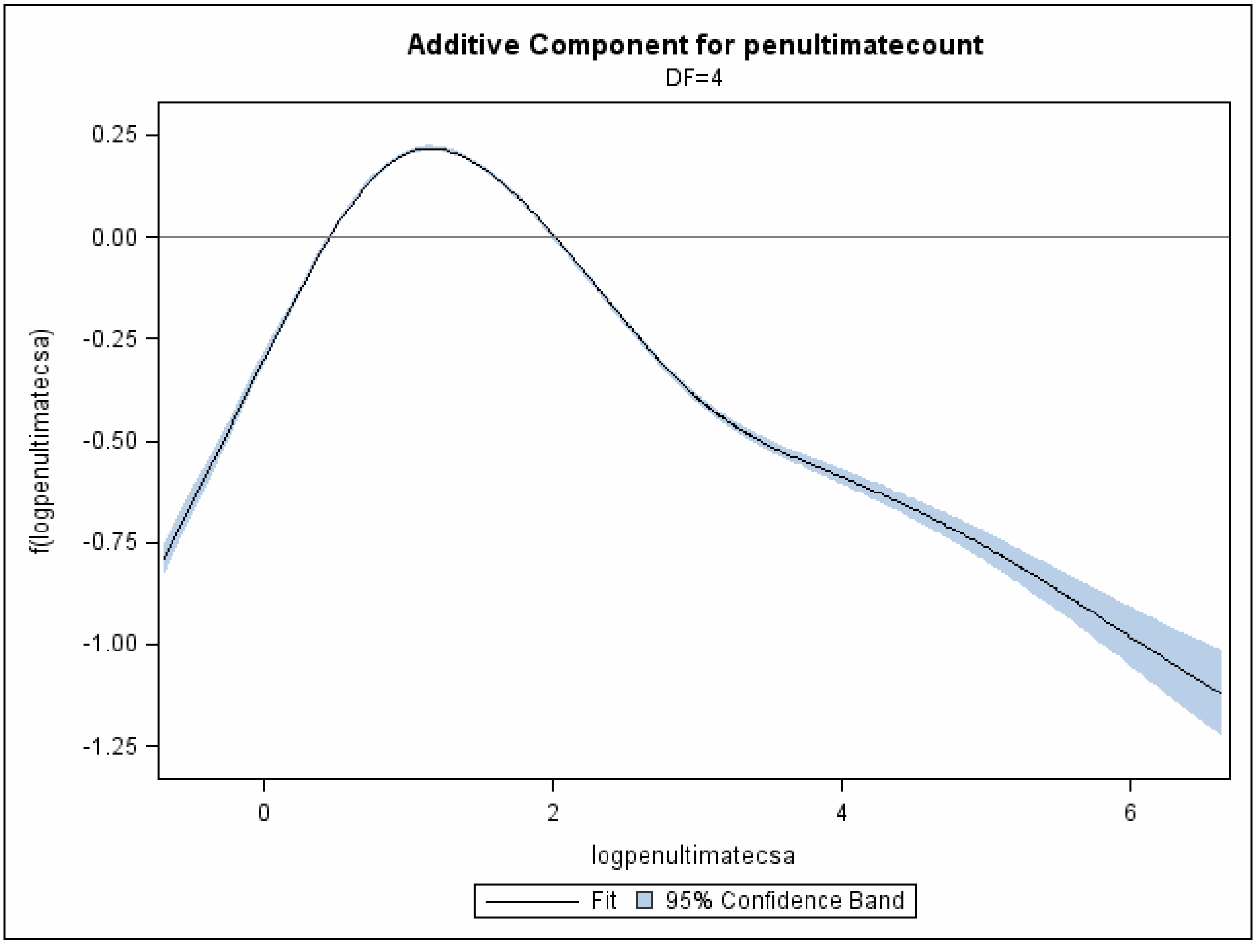
Crop estimates are based on three components: an estimate of the number of nuts per tree, an estimate of the number of trees per acre, and an estimate of the number of acres planted statewide in a given variety. It can be shown mathematically that the relative impact of each of these components on the percent error in the overall crop estimate will be proportional to the percent (or relative) error in each component. Thus the potential for improvement in the overall estimate would most likely come through improving the estimation of the components that are more difficult to estimate accurately.

The “nuts per tree” estimate is based on sampling data. The first stage in this process is selecting a representative (random) sample of orchards and then random trees within the selected orchards, then the number of nuts on those trees are estimated by counting the nuts on a “random path” through the tree and extrapolating that count to an estimate of the nuts on the entire tree. This extrapolation relies on the assumption that the number of nuts on a branch is proportional to the cross sectional area of the branch, which would be true if the fitted curve below (fit to Nonpareil data) were linear with a slope of one.



This assumption appears to be valid for the smaller branches, but the nut counts for larger branches will tend to be overestimated if this assumption is used. Thus the formula that extrapolates the count on a random path into an estimate of the nuts on the entire tree may benefit from an adjustment when branch cross sectional areas are large. The above curve was fit to the outermost branches of a tree.

The need for an adjustment is even greater for interior branches, as shown by this graph:



Both of the above graphs were fit to the data for Nonpareil trees. Qualitatively, the results are similar for most other varieties, but there are still apt to be statistically significant but subtle differences among varieties. Thus the exact form of the adjusted nut count estimates for large branches may vary by almond variety. This could improve the crop estimate particularly for the Nonpareil crop.

The estimation of the number of acres planted in a given variety is more problematical, since acreage enters or leaves production every year and records of those changes become available only after the harvest has occurred. There have been changes over time in these numbers, which appear to be well-approximated by linear trends, at least over a period of a decade or so. However, the annual percent change in acres differs substantially from one variety to another, ranging between a 13% annual increase in Monterey acreage, to a 1.7%

increase for Nonpareil, to a 20% decrease per year for Merced. Thus any adjustments for these yearly changes need to be “customized” to the particular variety of almond.

By comparison, the estimation of the number of trees per acre is more straightforward, since those numbers in large follow standardized patterns within a given variety and location.

Based on the above, the research has and will address the following areas and questions:

1. The assumed proportional relationship between branch cross-sectional area (CSA) and nut counts. Empirically, large CSA branches don’t have as many nuts as this relationship would imply, causing overestimation of nut counts. (work largely completed)

2. The distribution of nuts within a tree between canopy (outermost or terminal branches) and interior. In using random path sampling, a larger proportion of the tree interior is sampled than for the tree canopy, since as a path moves outward from the trunk, only one branch is sampled at each junction. Thus while roughly 60% of counted nuts were from terminal branches, it’s estimated that over 85% of total nuts are from those branches. (work largely completed)

3. The role of tree age in crop load. For a branch of a given CSA, nut counts are largest when a tree is young, since the subsequent increases (growth) in branch CSA aren’t matched by proportionate increases in nut counts. The purpose of this investigation is to determine whether the magnitude of this phenomenon is similar between almond varieties. (largely completed)

4. Uniformity of nut weights between canopy and interior. Since nuts are disproportionately sampled from the tree interior, estimation of crop characteristics (average nut weight, nut sizes) could be impacted if those characteristics vary between the canopy and the interior. To gain a better understanding of this, nuts are being collected according to their location within a tree, rather than in an aggregate fashion. (additional data being collected)

5. More appropriate relative weighting of nut counts to minimize variance, bias while taking nonuniformity of random path sampling into account. Random paths are chosen in a way that large branches are more likely to be sampled than smaller ones. Since they’re infrequently sampled, the nuts from the small branches that do get sampled need to be given greater weight in the overall estimate since they “represent” a large number of small branches that went unsampled. Thus in addition to the weighting determined in Question 1, this aspect will require an additional adjustment to the relative weighting given to different branches within a random path. (work started, to be a major focus in year 3)

6. Can an adjustment for alternate bearing in longitudinally-monitored trees improve estimation? The annual survey is conducted so that a given tree is included every year over a period of years. This allows better estimation of changes in nut load, but the interpretation of those changes will differ, depending on whether the tree is going from an “on” year to an “off” year, or vice versa. (Topic for year 3)

Research Effort Recent Publications:

No publications have been submitted based on this research because the statistical analyses do not warrant publication, and the data on which the analyses are based are confidential and cannot be disseminated.

References Cited:

None.