Yield Determinants and Prediction for California’s Almond Orchards Based on Machine Learning Analytics

Yufang Jin, Bin Chen, Bruce Lampinen, and Patrick Brown
University of California, Davis, CA, USA (yujin@ucdavis.edu)

Agricultural productivity is subject to various stressors, including abiotic and biotic threats, many of which are exacerbated by a changing climate. The productivity of tree crops, such as almond orchards, is particularly complex. Moreover, the State of California has implemented legislatively mandated nitrogen (N) management strategies of all growers statewide to minimize nitrogen losses to the environment, and almond growers must now apply N in accordance with the estimated yield in early spring. To understand and mitigate these threats requires a collection of multi-layer large data sets, and advanced analytics is also critical to integrate these highly heterogeneous datasets to generate insights about the key constraints on the yields at tree and field scales. Here we used machine learning approaches to predict orchard-level yield and examine the determinants of almond yield variation in California’s almond orchards, based on a unique 10-year dataset of field measurements of light interception, remote sensing metrics, and almond yield, along with meteorological data. We found that overall the maximum almond yield was highly dependent on light interception, e.g., with each one percent increase in light interception resulting in an increase of 57.9 lbs/acre in the potential yield. Light interception was highest for mature sites with higher long term mean spring incoming solar radiation, and lowest for younger orchards and when March maximum temperature was lower than 19 °C. However, at any given level of light interception, actual yield often falls significantly below full yield potential, driven mostly by tree age, temperature profiles in June and winter, and summer maximum vapor pressure deficit (VPDmax). The full random forest model was found to explain 82% (±1%) of yield variation, with a RMSE of 480±9 lbs/acre. When excluding light interception from the predictors, overall orchard characteristics (such as age, location and tree density) and key meteorological variables could still explain 78% of yield variation. The model analysis also showed that warmer winter conditions often limited mature orchards from reaching maximum yield potential and higher summer VPDmax significantly limited the yield. Our findings through the machine learning approach improved our understanding of the complex interaction between climate, canopy light interception, and almond nut production. The demonstrated relatively robust predictability of almond yield, driven by “big data”, also provides quantitative information and guidance to make informed orchard nutrient management decisions, allocate resources, determine almond price targets, and improve market planning.