Yield Determinants and Prediction for California's Almond Orchards: Machine Learning Analytics

© Yufang Jin*, Bin Chen, Bruce Lampinen, and Patrick Brown

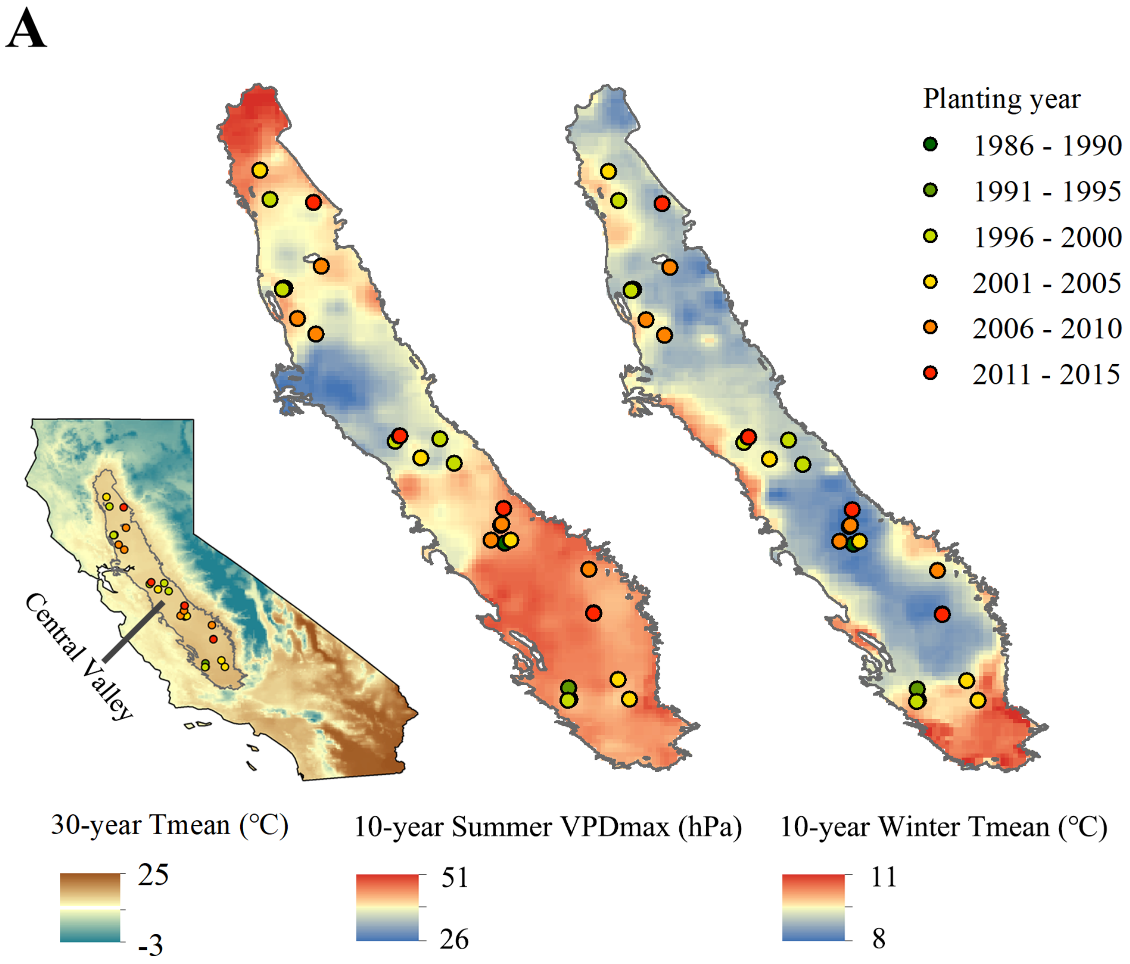
University of California, Davis

* yujin@ucdavis.edu

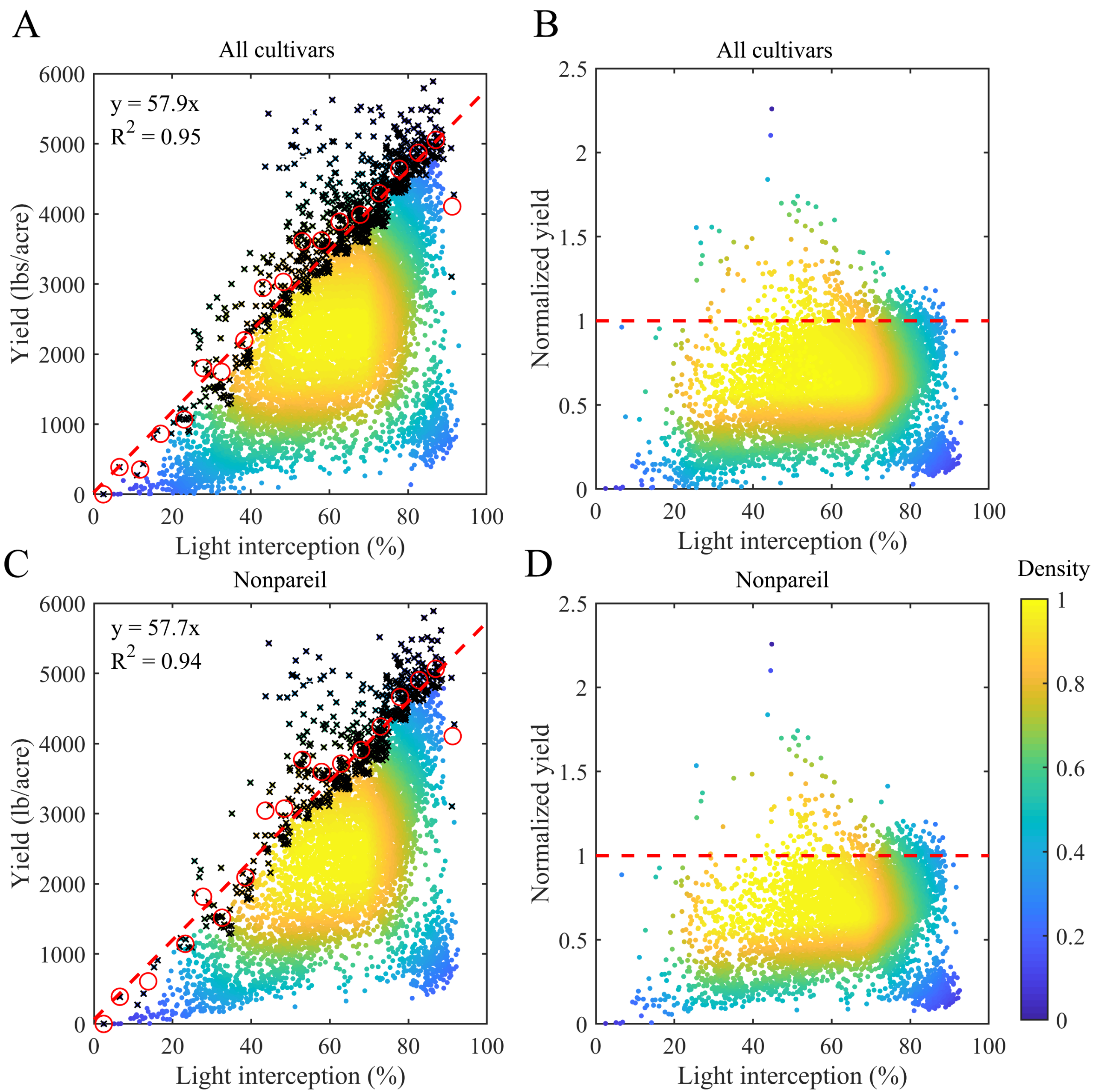
Motivations

- California produces about 80% of the world's almonds; Almond's acreage almost doubled since 2005, generating $5.6B in revenue.
- Large yield variation across the fields, and vulnerable to climate change and extreme.
- Degraded ground water quality due to nitrate leaching.

Understanding the determinants and improving prediction: key for optimizing production while reducing N input ……



Jin et al., Frontiers in Plant Science (2020)



- Overall, almond yield is highly dependent on the light interception.
- One percent of increase in light interception led to an increase of 57.9 lbs/acre in the potential yield.
- Many orchards did not reach the potential yield at a given light interception.

Determinants for the normalized yield by light interception (based on Random Forest)

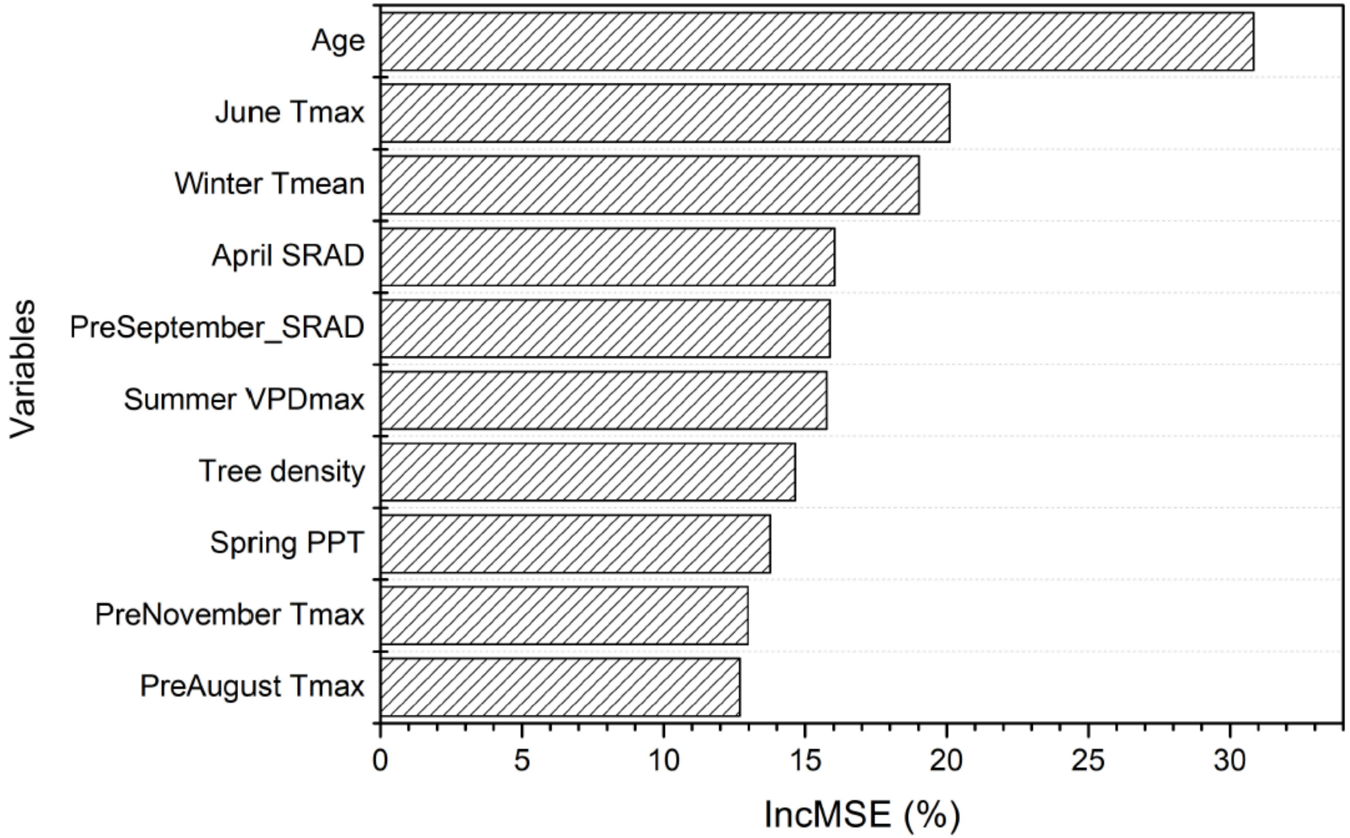
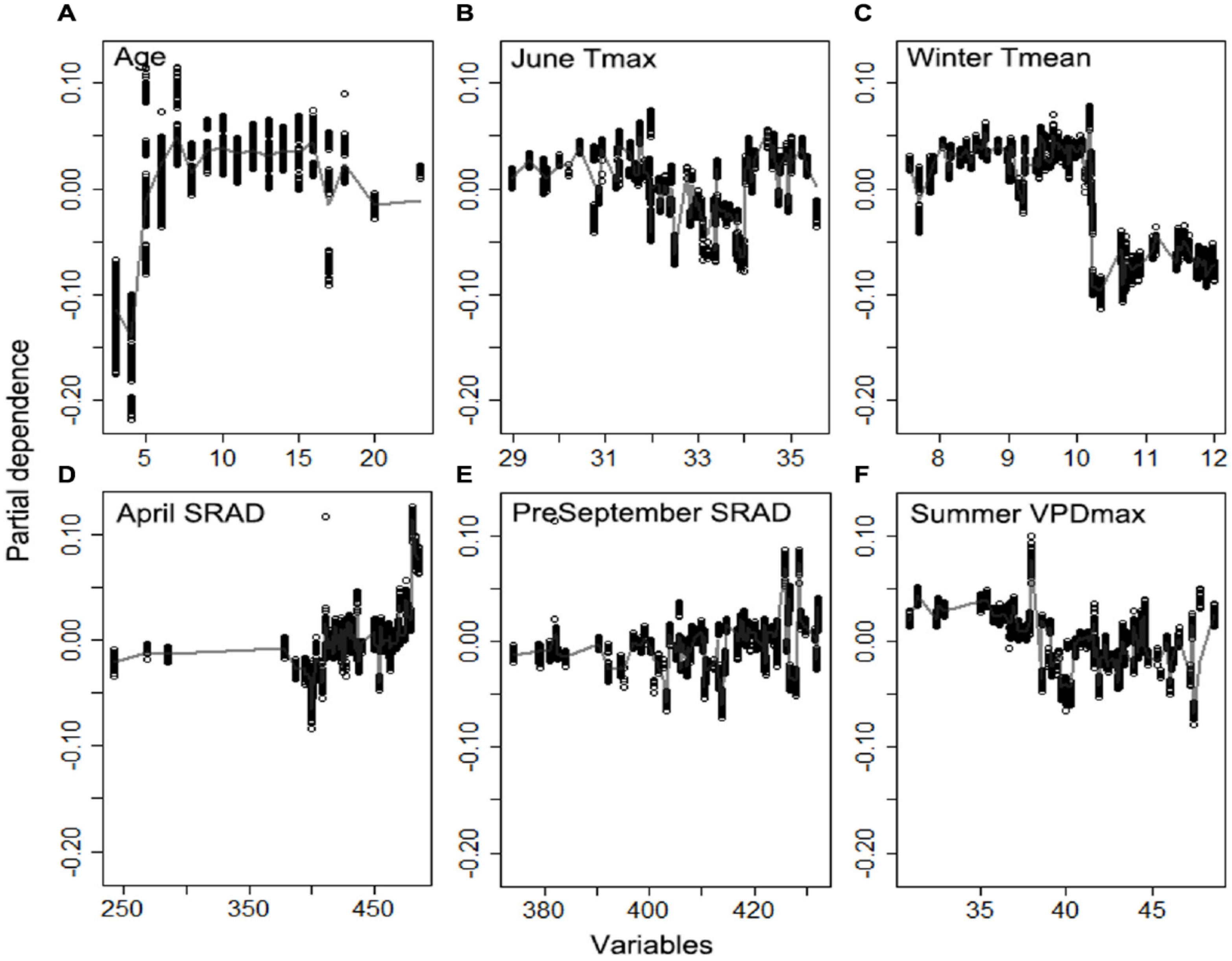


FIGURE 3 | Variable importance from the random forest model of yield gap, as measured by the increase in mean-square-error (IncMSE) of predictions when excluding each variable.

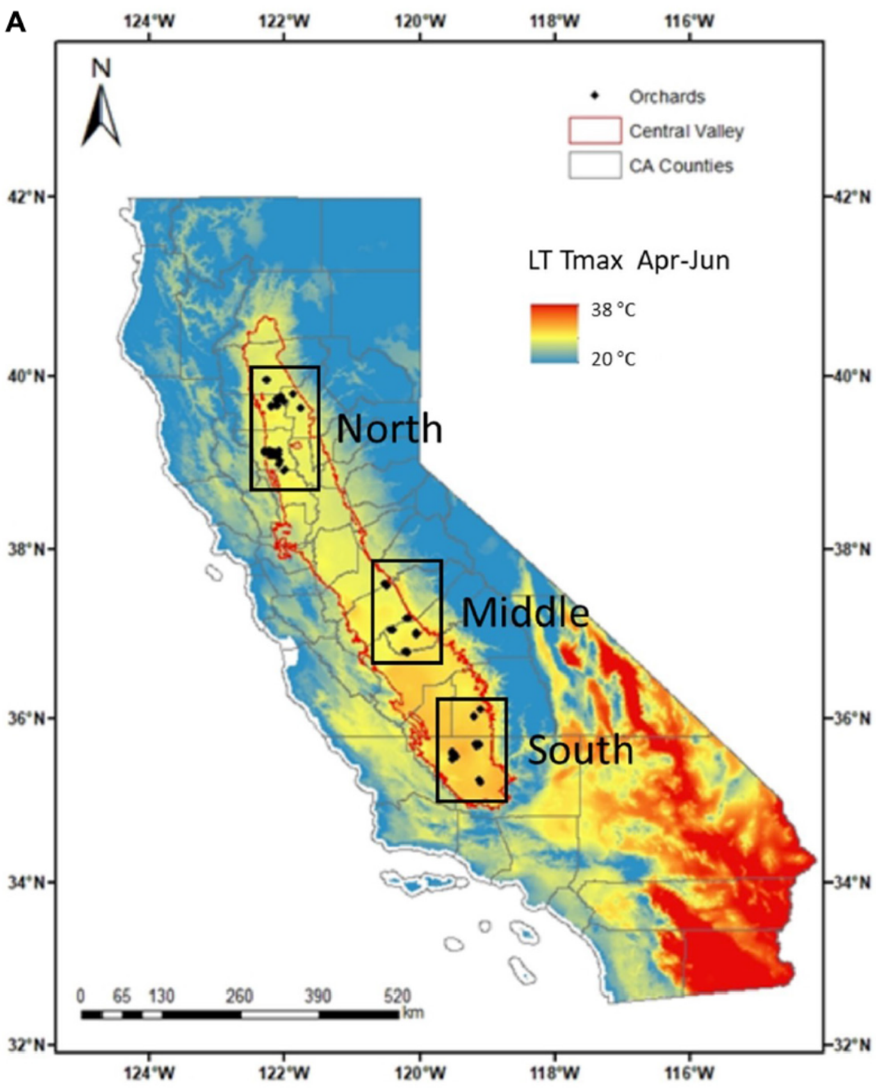
Partial dependence of the normalized yield



frontiers in Plant Science

ORIGINAL RESEARCH
published: 18 July 2019
doi: 10.3389/fpls.2019.00809

Check for updates



OPEN ACCESS

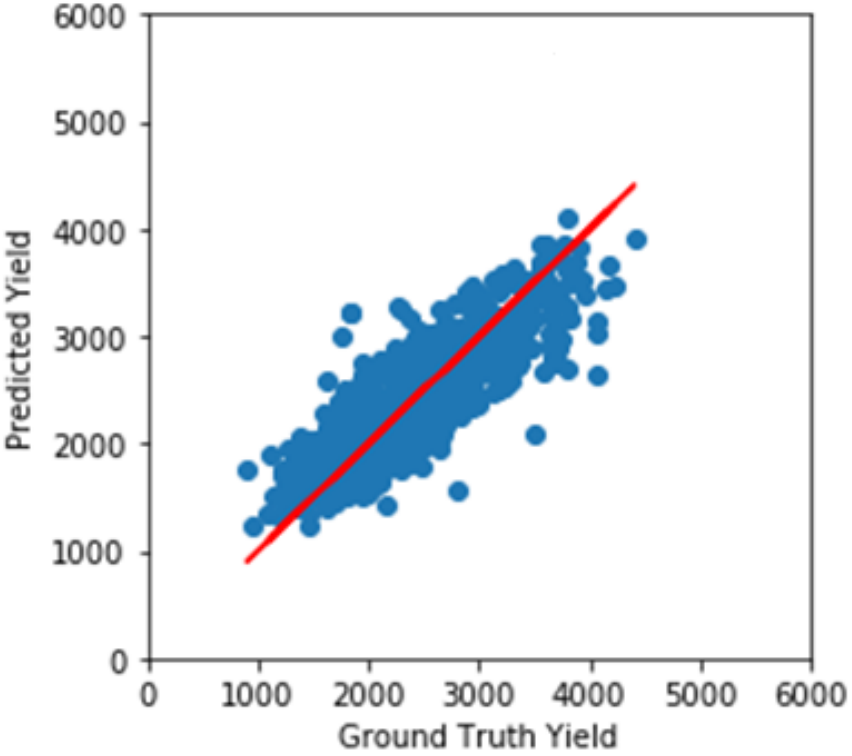
California Almond Yield Prediction at the Orchard Level With a Machine Learning Approach

Zhou Zhang[1,2], Yufang Jin[1], Bin Chen[1] and Patrick Brown[3]*

[1] Department of Land, Air and Water Resources, University of California, Davis, Davis, CA, United States, [2] Biological Systems Engineering, University of Wisconsin–Madison, Madison, WI, United States, [3] Department of Plant Sciences, University of California, Davis, Davis, CA, United States

California's almond growers face challenges with nitrogen management as new legislatively mandated nitrogen management strategies for almond have been implemented. These regulations require that growers apply nitrogen to meet, but not exceed, the annual N demand for crop and tree growth and nut production. To accurately predict seasonal nitrogen demand, therefore, growers need to estimate block-level almond yield early in the growing season so that timely N management decisions can be made. However, methods to predict almond yield are not currently available. To fill this gap, we have developed statistical models using the Stochastic Gradient Boosting, a machine learning approach, for early season yield projection and

Almond Yield Prediction (for individual block)



Early season prediction by the gradient boosting tree model agreed well with the grower reports.

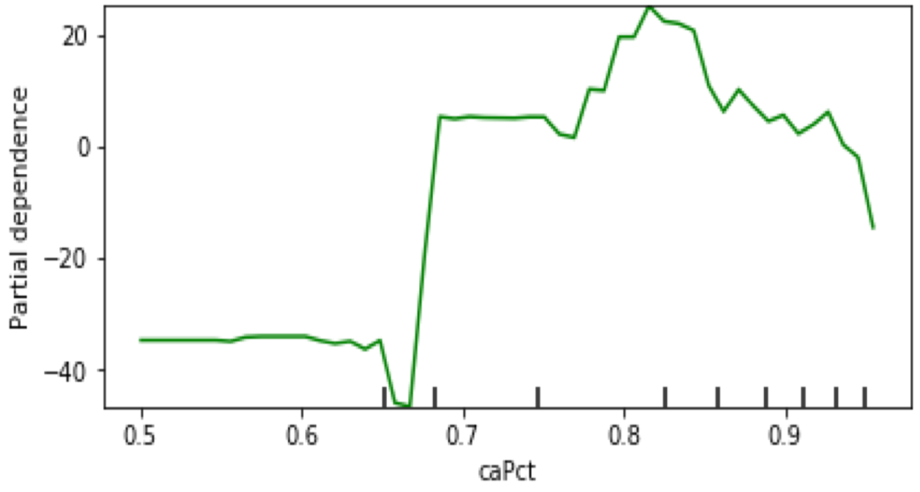
TABLE 2 | Comparison of the performance of five machine learning approaches for orchard-level almond yield prediction, when using the full set of input variables.

Prediction time	Machine learning approach	R^2	RMSE (kg/ha)	RPIQ
Early season	Linear regression	0.58 (0.04)	422 (4.1)	2.19 (0.19)
	Support vector regression	0.51 (0.04)	460 (16.5)	2.01 (0.16)
	Artificial neural network	0.50 (0.05)	474 (26.1)	1.96 (0.16)
	Random Forest	0.69 (0.04)	364 (14.8)	2.55 (0.28)
	Stochastic gradient boosting	**0.71 (0.04)**	**352 (15.2)**	**2.64 (0.33)**
Mid-season	Linear regression	0.59 (0.05)	416 (6.1)	2.23 (0.22)
	Support vector regression	0.52 (0.04)	453 (15.5)	2.05 (0.17)
	Artificial neural network	0.48 (0.04)	473 (7.6)	1.96 (0.15)
	Random Forest	0.69 (0.04)	365 (13.9)	2.54 (0.27)
	Stochastic gradient boosting	**0.71 (0.04)**	**355 (12.3)**	**2.62 (0.29)**

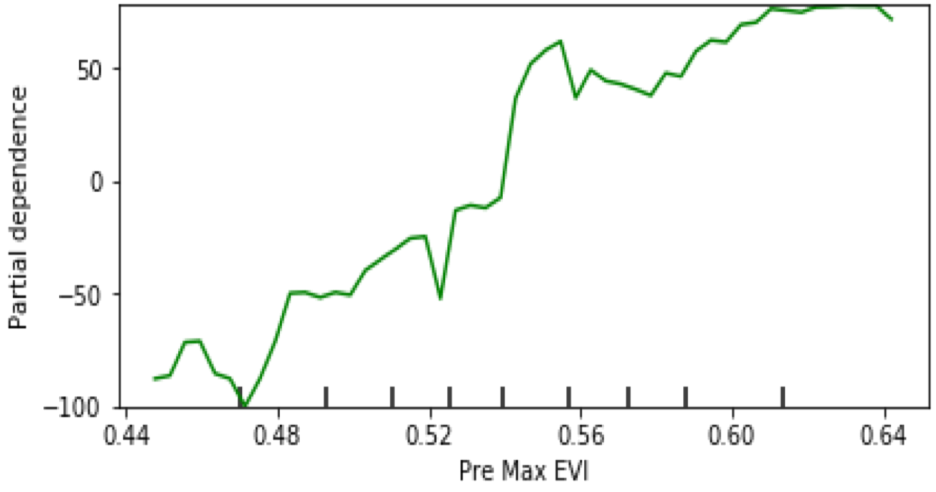
A four-fold cross-validation strategy was used for model building and testing; both mean values and standard deviations (in parenthesis) of R^2, RMSE, and RPIQ are presented here based on the comparison of the prediction and the independent testing data.

Impact of Remote Sensing Metrics

Canopy Cover



Previous year's maximum EVI

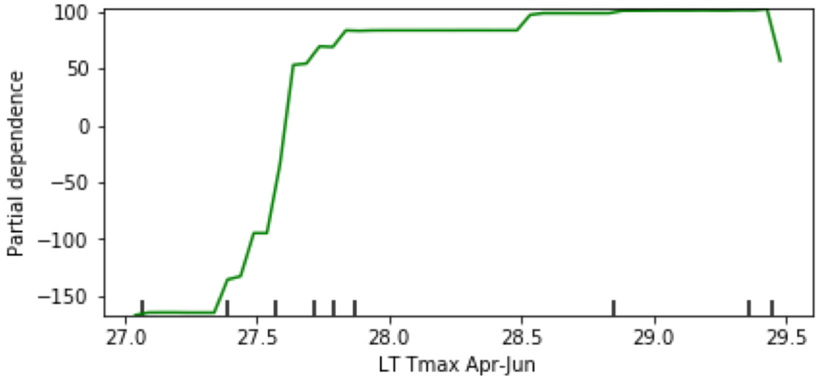


- For mature orchards, canopy cover and Enhanced Vegetation Index (Landsat) are both positively correlated to the yield.

Zhou et al., Frontiers in Plant Science (2019)

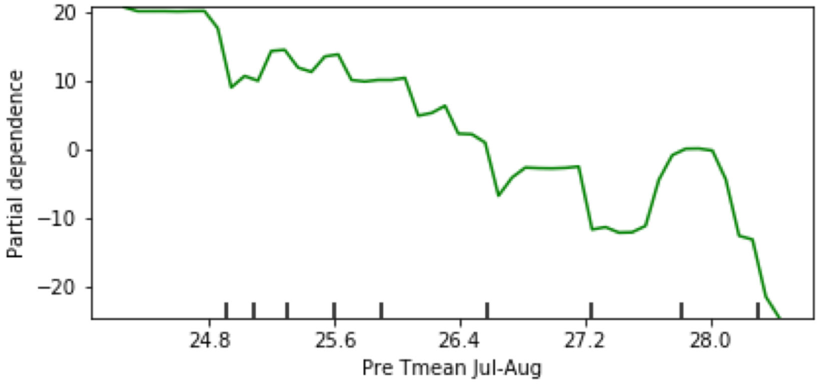
Impact of climate variables

Long Term Tmax Apr-Jun



PPT Feb-Mar

Last year's summer Tmean



- **Long term mean Tmax during April ~June enhanced the yield.**
- **Precipitation during blooming time reduced the almond yield.**
- **Previous year's summer Tmean had a negative impact on the yield.**

Summary

- Canopy characteristics, long term climate, and short term weather affected the yield at individual almond orchard.
- Machine learning approaches, such as random forest and gradient boosting, provided insights about the importance of key drivers on almond yield.
- The machine models also predicted the yield reasonably well in both early and mid-season, and thus can be used to guide the data-driven N management.