Predicting Agricultural Productivity Using Satellite Data and Machine Learning

Aakash Ahamed[1], Noah Dewar[1]

[1]Department of Geophysics, Stanford University

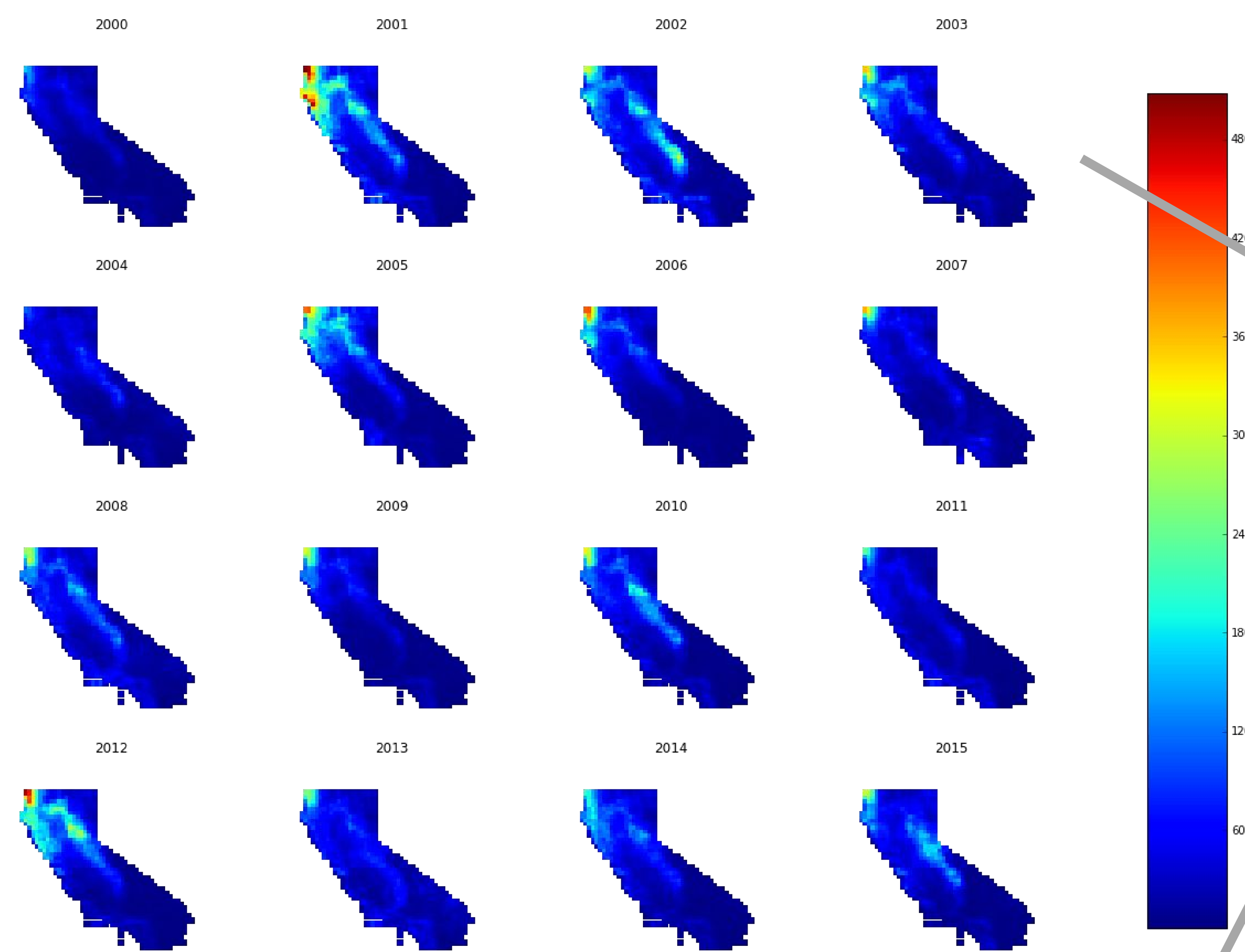
aahamed@stanford.edu ndewar@stanford.eduß

Introduction

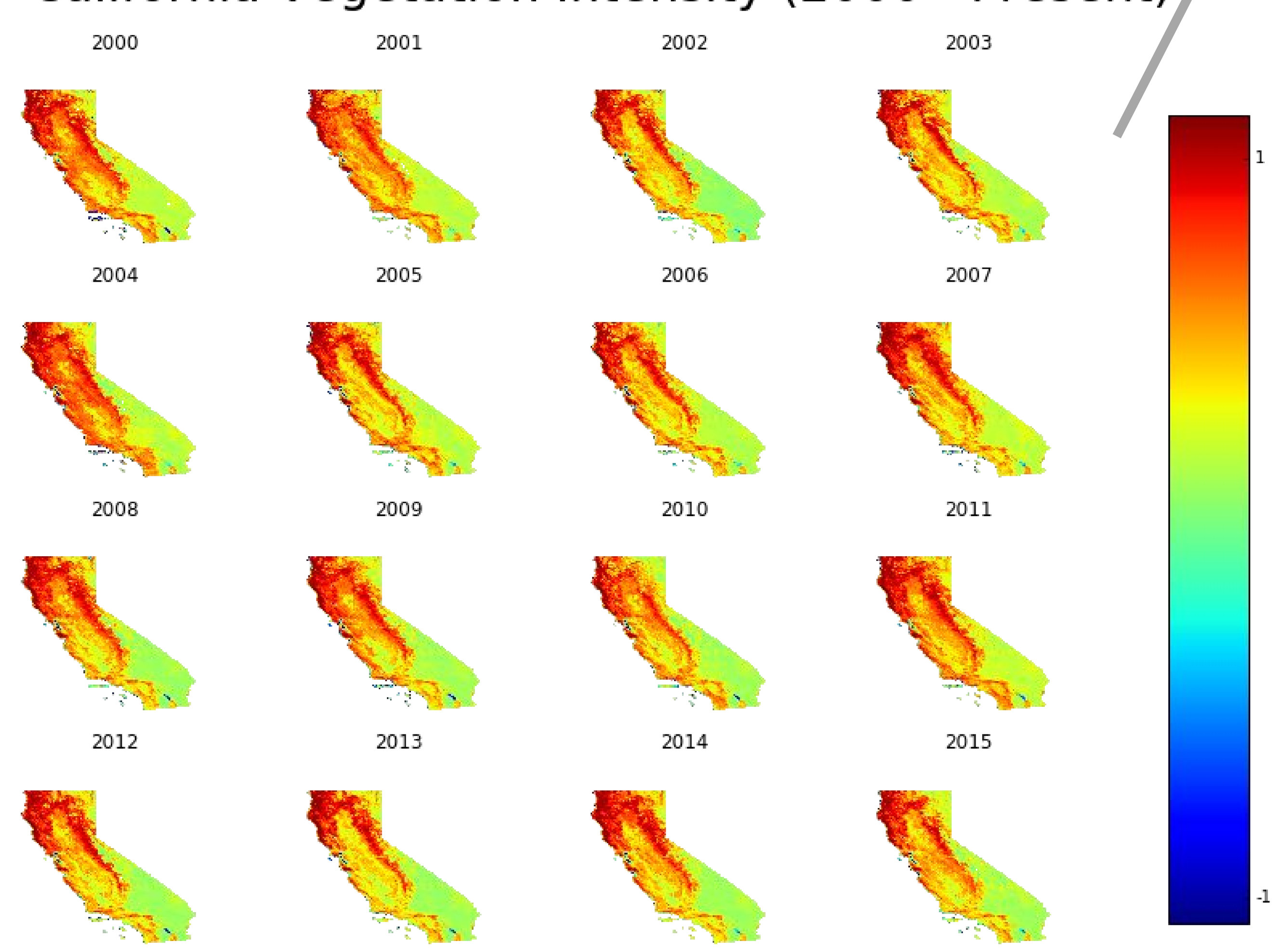
California is the most productive agricultural region in the United States, producing over 46 billion USD of agricultural commodities in 2013 (NASS, 2013). In the wake of the 2010 – 2016 drought, improved statistical models to predict agricultural yields can greatly benefit water managers and growers as well as economic agencies seeking to subsidize farm production and regulate commodity prices. Accurately modelling the complex relationships between agricultural yield and parameters derived from satellite imagery (e.g. vegetation intensity, precipitation, surface water) facilitates real time data driven management decisions at the county scale. In this study we explore the possibility of agricultural yield prediction from satellite imagery using machine learning, and report test accuracies for various algorithms applied to satellite data

Satellite Data

Annual California Precipitation, mm (2000 - Present)



California Vegetation Intensity (2000 - Present)



Training Data and Preprocessing

- County level crop production data (production aggregated by crop type) for California from the United States Department of Agriculture from 2000 - 2017 were the response variable, Input features were satellite data.

- Optical satellite data were filtered to eliminate non-agricultural areas, and averaged during the course of the growing season (February - September).

Methods

1) Naive Bayes - multinomial event model

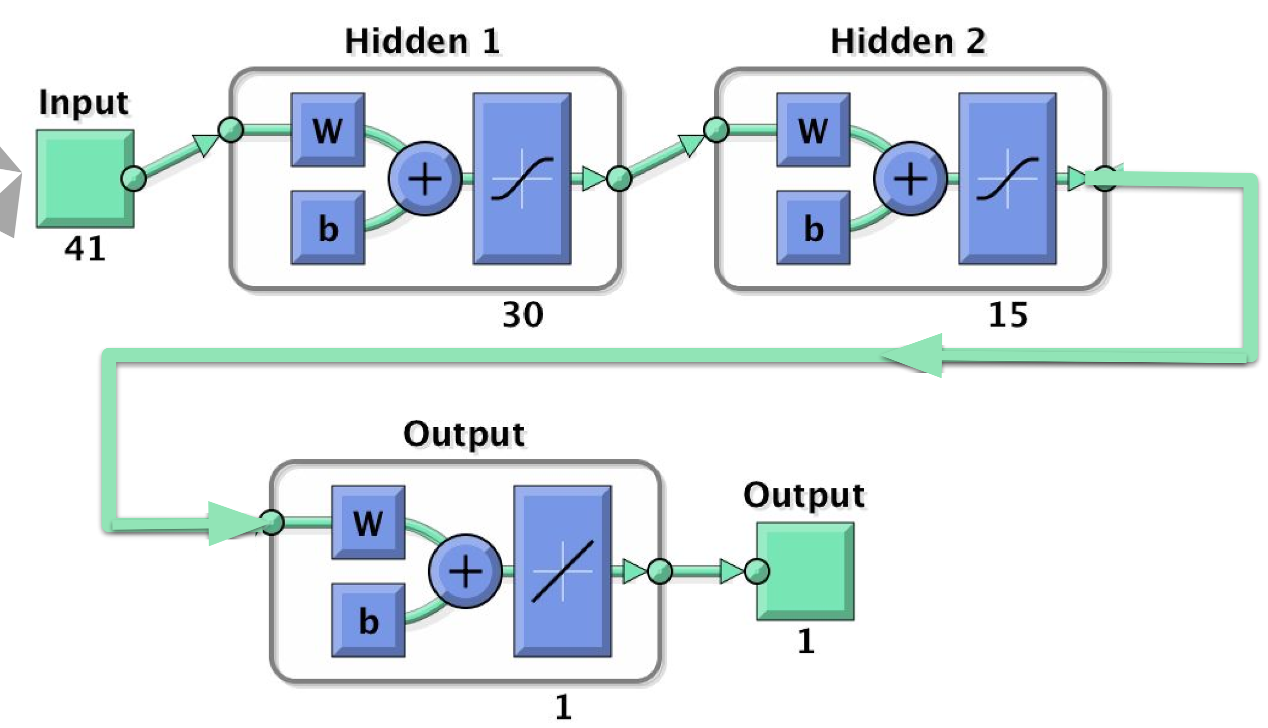
$$\phi_{k|y=1} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n_i} 1\{x_j^{(i)} = k \wedge y^{(i)} = 1\} + 1}{\sum_{i=1}^{m} 1\{y^{(i)} = 1\} n_i + |V|}$$

$$\phi_{k|y=0} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n_i} 1\{x_j^{(i)} = k \wedge y^{(i)} = 0\} + 1}{\sum_{i=1}^{m} 1\{y^{(i)} = 0\} n_i + |V|}$$

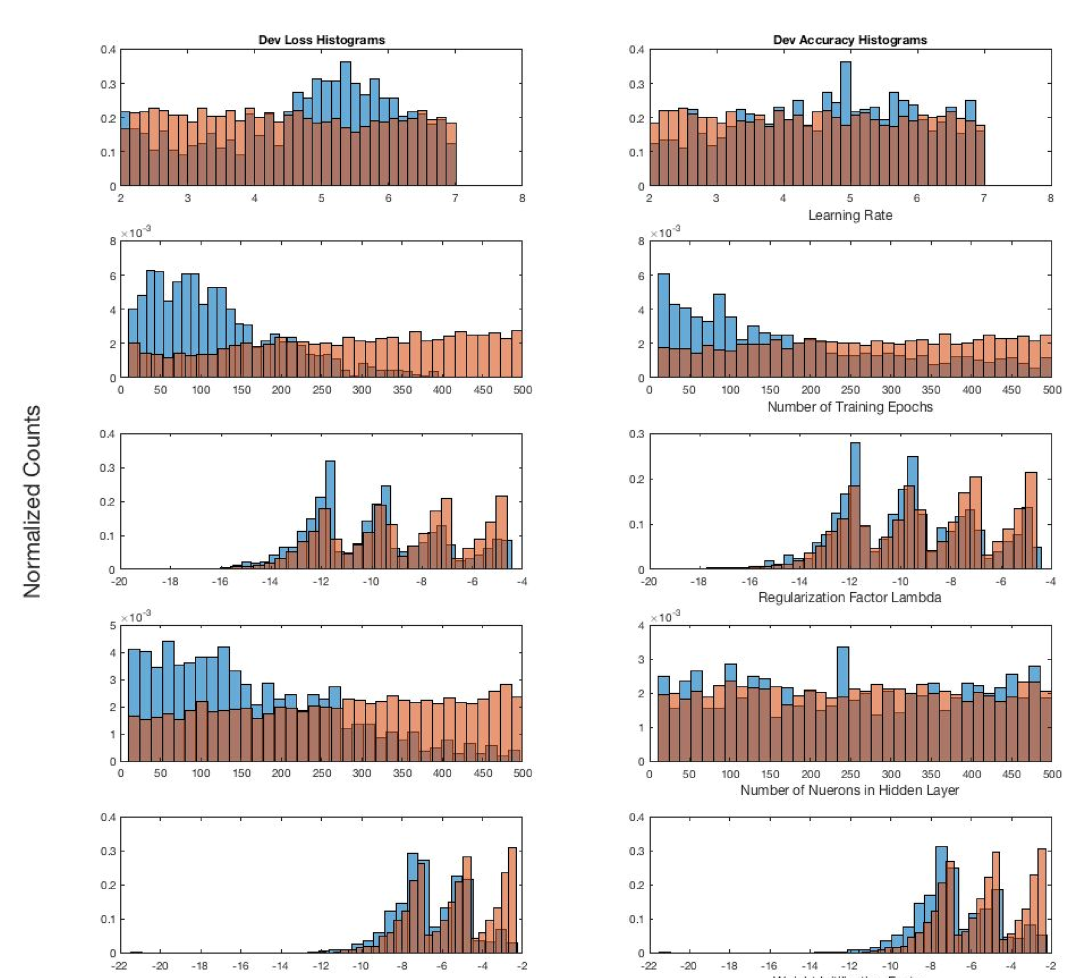
2) Neural Networks

- **Levenberg–Marquardt**
- **Variable Learning Rate Gradient Descent**
- **Bayesian Regularization [BEST]**

Network Architecture



Parameter Optimization with Monte Carlo Method

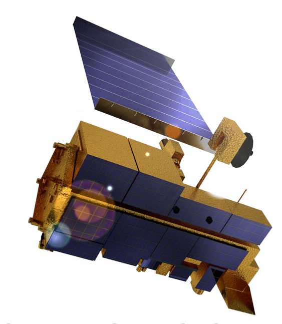


Tools and Technologies Used

python

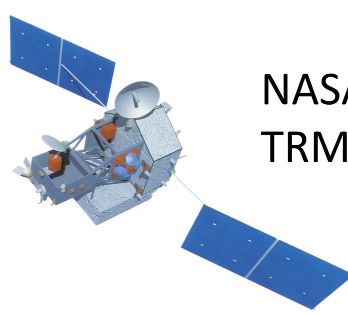
MATLAB

Google Earth Engine

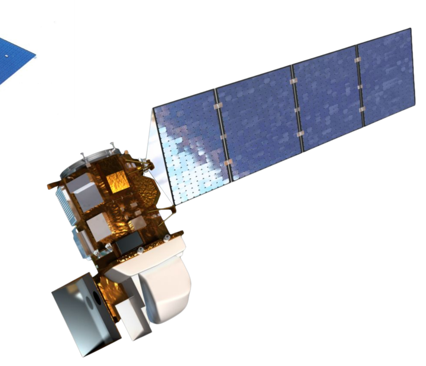


NASA MODIS Satellites

TensorFlow (Planned)



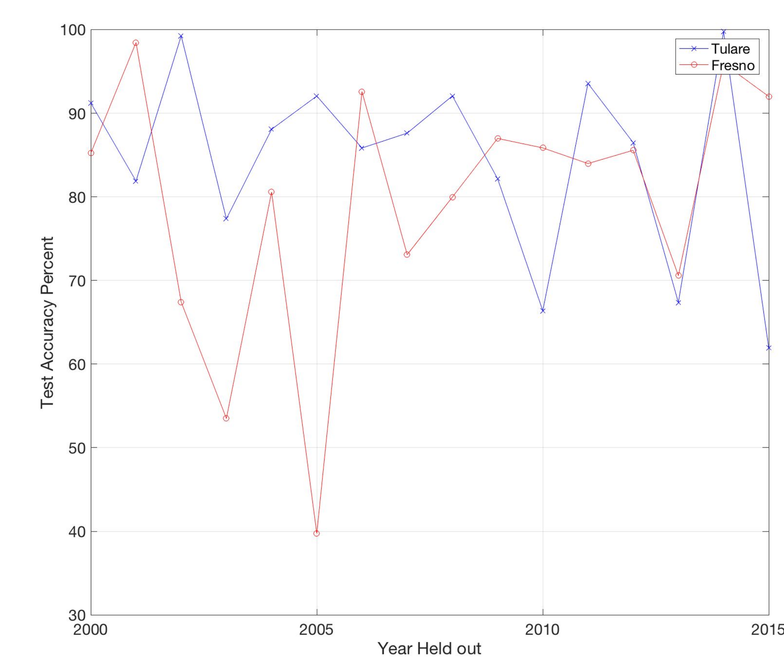
NASA / JAXA TRMM Satellites

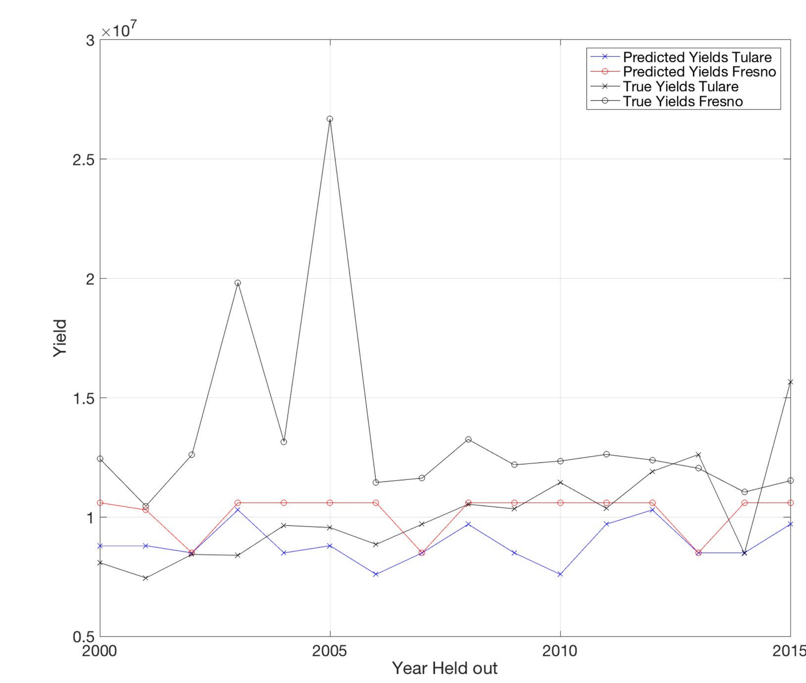


NASA Landsat (Planned)

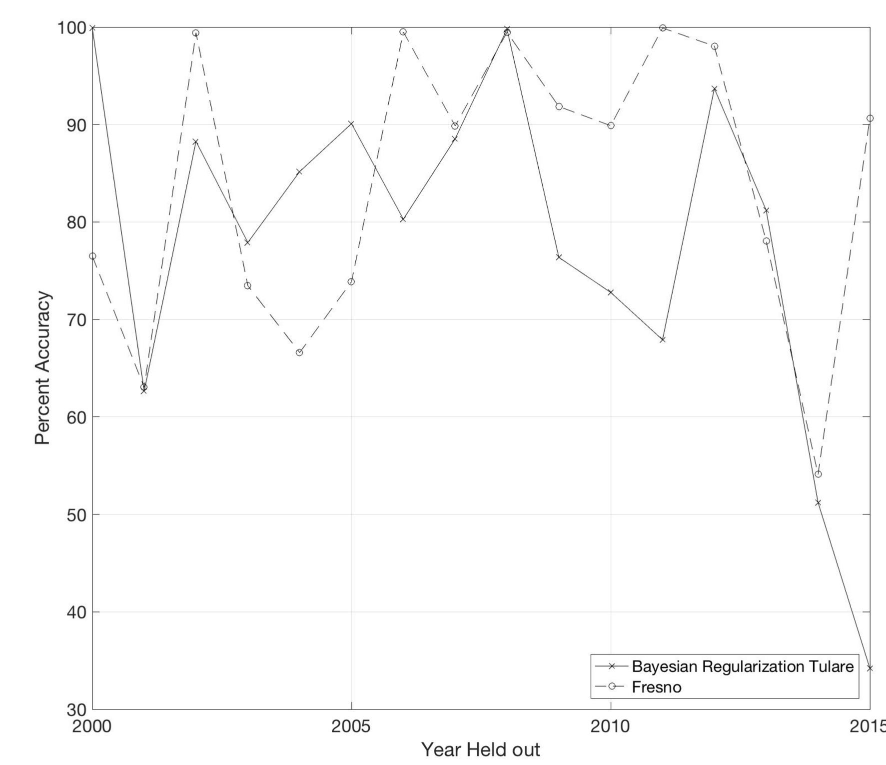
Results

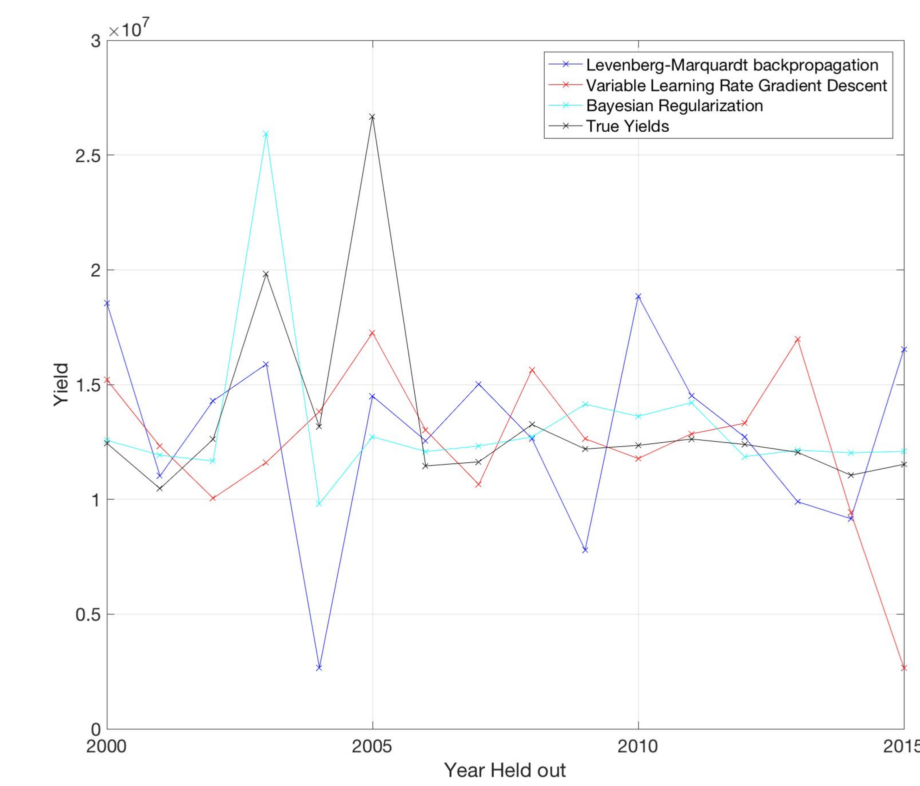
1) Naive bayes - Leave one out cross validation - 82% accuracy





2) Neural Network - Leave one out cross validation - <u>85 - 90% accuracy</u>





Discussion and Future Work

Preliminary results of hold one out cross validation are promising, with overall accuracies of 82% for Naive Bayes and 85-90% for the Neural Network (with Bayesian regularization as the training function). However, there is significant potential for improvement due to a number of factors. Models have been trained and tested on 2 of the most prolific agricultural counties in order to save time and judge efficacy. There are 58 counties in California, out of which ~30 are significant agricultural regions. Interestingly, test accuracies are higher when models are trained on a per-county basis, rather than an intra-county basis. We expect that as additional counties are added, test accuracy will increase. Further, only 2 satellite derived variables, precipitation and vegetation fluorescence, were considered in the input feature space. Better results could be achieved through the addition of additional satellites (e.g. Landsat), additional variables (e.g. land surface temperature), and different aggregation schemes (e.g. averaging data monthly or seasonally). To summarize, despite favorable initial performance for simple models, enhanced performance could be achieved by (1) optimizing model parameters, (2) adjusting the preprocessing of input features, and (3) adding additional satellites and satellite derived variables as input features.

References

California Agricultural Statistics, Crop Year 2013 USDA, NASS, Pacific Region. https://www.nass.usda.gov/Statistics_by_State/California/Publications/California_Ag_Statistics/CA_Ag_Overview.pdf

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, *353*(6301), 790-794

You, J., Li, X., Low, M., Lobell, D., & Ermon, S. (2017). Deep Gaussian Process for Crop Yield Prediction Based on Remote Sensing Data. In *AAAI* (pp. 4559-4566).