Crop yield prediction - A machine learning based approach

Prajna Kandarpa

March 24, 2016

University of Waterloo

why should we predict crop yields?

CYP Uses

- Urgent need for crop yield increase to sustain increasing global population [1]
- **Crop yield predictions aim to reduce yield gap**
- Help inform governmental policies and achieve food security
- Crop yield varies geospatially and temporally in a non-linear fashion (duh)
- *curious trend observed globally - most crop yields tend to plateau at 70-80% of Yp at regional and national levels*

Literature review

Status Quo

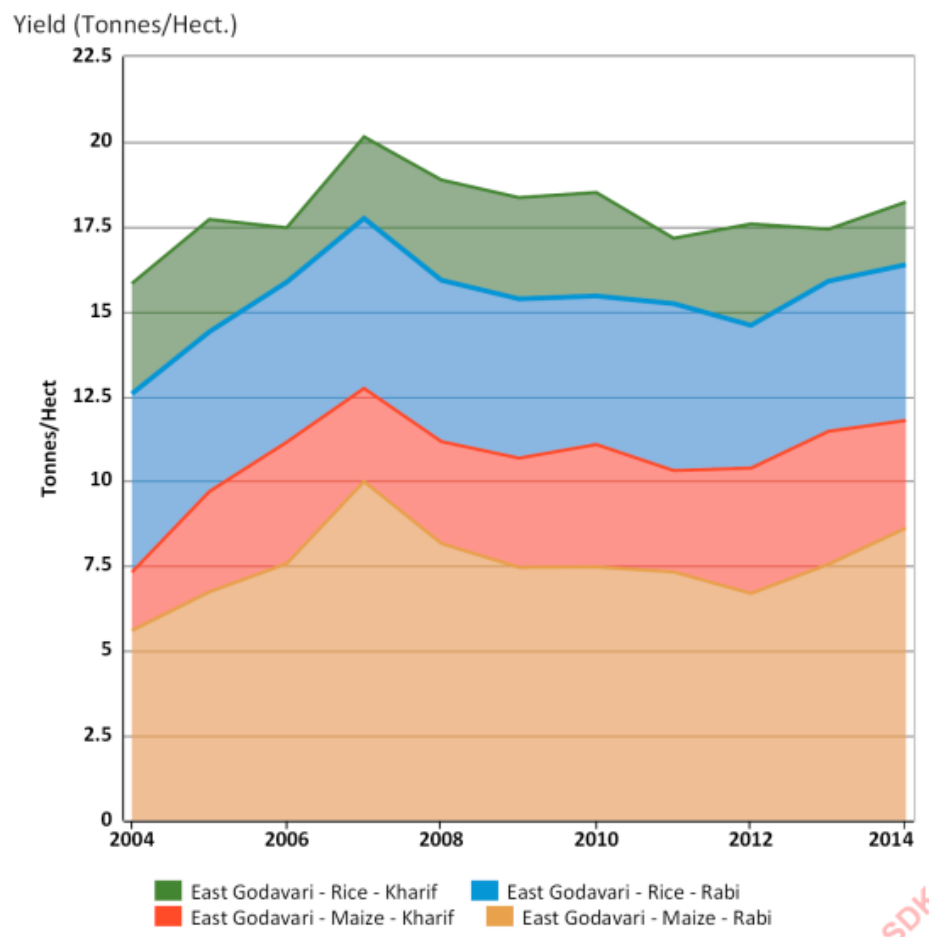
How do current methods work?

- advanced dynamic models of crop production and socio-economic factors.
  - CERES-Wheat, CERES-Maize, CERES-Barley, SOYGRO for legumes, AFRCWHEAT2 [2]
- use decades of research on crop physiology and reproduction, agronomy, and soil science, among other disciplines
- Need to collect data through multi-disciplinary experiments
- Research that seeks to emulate model results using statistical and ML models - goal of this project

Data Sources and Preprocessing

Crop selection

Rice, Maize - Subset of data for a single district(region)



Data Sources and Preprocessing

Data Sources

- weather data from 1872-2014 (regional precipitation levels for the country of India)
- Crop yield data from 1959-2014

Feature Selection

Weather - **regional resolution** 2 agricultural seasons in India - Kharif and Rabi

- precipitation levels
- temperature
- humidity

Physiographic data

- area under cultivation
- irrigation levels

Prediction Accuracy

Output = seasonal crop yield at regional level

- Standard regression error measures
- Root Mean Squared Error (RMSE)
- Mean Absolute Percentage Error (MAPE)

Analysis Methods

Machine learning for CYP

machine learning is statistics minus any checking of models and assumptions - Brian D. Ripley

- can we achieve results similar to advanced crop models using fewer variables?
- ML approaches are data driven - require lesser understanding of the underlying physical systems
- analogous to financial time series forecasting

Modeling Techniques

Time series prediction involves measuring *co-movement*

co-movement is a measure of correlation between covariation, detrended cross correlations with frequency, trends, seasonality and uncertainty of the data

- Convolutional Neural Networks with 3-7 hidden layers - experiment with number of layers and number of nodes in each layer
- MARS - Multiple Adaptive Regressive Splines
- Support Vector Regression

Steps

$$\hat{y}[t+1] = f(y[t-k], ..., y[t])$$

All analysis is done in R using *caret*

1. Cleaned weather, yield, land datasets into a single dataframe
2. Pick important predictors using covariance measures
3. Data binning, filtering, k-fold cross validation
4. train all 3 models
5. Prediction cross validation across all evaluated models

Extras

Depending on performance of NN, SVR and MARS models, build a crop recommendation tool for farmers that factors in historical performance and regional crop yields

Assess inferential viability of neural network by examining hidden layer outputs

Crop Recommendation tool requires yield prediction models for each crop

References

- D. B. Lobell and M. B. Burke, "On the use of statistical models to predict crop yield responses to climate change," *Agricultural and Forest Meteorology*, vol. 150, no. 11, pp. 1443–1452, 2010. [Online]. Available: http://dx.doi.org/10.1016/j.agrformet.2010.07.008
- A. Gonzalez-Sanchez, J. Frausto-Solis, and W. Ojeda-Bustamante, "Predictive ability of machine learning methods for massive crop yield prediction," *Spanish Journal of Agricultural Research*, vol. 12, no. 2, pp. 313–328, 2014. [Online]. Available: http://revistas.inia.es/index.php/sjar/article/view/4439