**Predicting Corn, Wheat and Soybean Yield using Meteorological Data and Feature Importance Comparison among Cultures**

Maura M. Tokay

Problem Definition

Precipitation and temperature are the two main characteristics of weather and climate, they also play an important role on the outcome of crop yield. In this study the goal is to try to predict crop yield for corn, wheat and soybeans using meteorological data using machine learning predictive models like Lasso, Decision Tree Regressor and Random Forest Regressor. A feature analyses will also be conducted to understand the importance that each meteorological parameter have on the outcome of the model. It will be interesting to verify if the feature importance will be the same for each crop or not. Once we can quantify the importance of weather parameters listed below for crop yield we can analyze the impact of climate change on food security.

Dataset

The dataset used is the result of a long-term agroecological research project in Beltsville, Maryland, the Farming Systems Project (FSP). The data available for analyses span 20 years, 1996-2016.

The grain yield dataset has the following information:

* Crop
* Growing Season
* System
* System Name
* Plot
* Grain Yield (kg/ha)
* Grain Yield (lb/A)
* Grain Yield (bu/A)
* Harvest Date
* Notes

The meteorological data was measured daily on the same site where crops were cultivated and covers the period of January 1st, 1996 to December 31st, 2016. The dataset has the following information:

* Year
* Julian Day
* Month
* Day of Month
* Average Temperature (⁰C)
* Maximum Temperature (⁰C)
* Minimum Temperature (⁰C)
* Average Humidity (%)
* Maximum Humidity (%)
* Minimum Humidity (%)
* Average Radiation (w/m²)
* Daily light integral (mol/m² day)
* Mean Wind (m/s)
* Maximum Wind (m/s)
* Total Rain (mm)