Cotton Incorporated funded project - Quarterly Report

Leonardo Bastos

Assistant Professor, University of Georgia

This quarterly report describes progress towards the final goal of utilizing the USDA long-term cotton variety trial and publicly available weather data to understand the historical effects of weather on cotton fiber quality.

# Achieved tasks in the last quarter

During this quarter, the following was achieved:

* Downloaded, processed, and merged USDA cotton variety trial data sets into one data frame.
* Appended each study site with the geographical coordinates of the centroid of nearest town.
* Used site centroid to retrieve historical weather data from [Dayment](https://daymet.ornl.gov):
  + Temporal resolution: daily
  + Temporal extent: from 1980 through 2020
  + Spatial resolution: 1 x 1 km
  + Primary weather variables:
    - Day length (h)
    - Precipitation (mm)
    - Solar radiation (W/m2)
    - Minimum air temperature (C)
    - Maximum air temperature (C)
    - Vapor pressure deficit (Pa)

|  |
| --- |
| Figure 1. Distribution of daily primary weather variables (summarized by the growing season). dayl is day length (h), prcp is precipitation (mm), srad is solar radiation (W/m2), tmax is maximum air temperature (C), tmin is minimum air temperature (C), vp is vapor pressure deficit. |

* Used primary weather variables to derive secondary weather variables:
  + Temperature amplitude (C)
  + Growing degree days

|  |
| --- |
| Figure 2. Distribution of daily secondary weather variables (summarized by the growing season). tamp is temperature amplitude, gdd is growing degree days. |

* Summarized primary and secondary weather variables by summing (precipitation, solar radiation, growing degree days) or averaging (all others) across different temporal scales including:
  + Every month
  + Every 2 months
  + Every 4 months
  + Growing season (April through November)

|  |
| --- |
| Figure 3. Distribution of minimum air temperature summarized at different temporal scales. |

* Extracted the best linear unbiased estimators (BLUEs) of one fiber quality parameter (strength) across all site-years in the data set.

|  |
| --- |
| Figure 3. Distribution of cotton fiber strength best linear unbiased estimator. |

|  |
| --- |
| Figure 4. Cotton fiber strength time-series. |

* Merged summarized weather variables with fiber strength BLUEs in the same data set for machine learning training:
  + Number of unique sites: 69
  + Number of years: 41
  + Number of site-years: 1,103
  + Number of summarized weather predictor variables: 151

# Future tasks

* Train random forest models in a machine learning framework (data split, hyperparameter optimization, predictive power assessment) to predict cotton fiber strength as a function of weather. A different model will be trained for each of the summarized temporal scales (from monthly to entire growing season), and one including all temporal scales in same model.
* Once the workflow has been developed for cotton fiber strength, similar analysis will be conducted for the remaining fiber quality variables (length, uniformity, micronaire).