Capstone Project Proposal
For GEOG 596A

Penn State | College of Earth and Mineral Sciences Master of Geographic Information Systems
Author: Jason Biagio | jrb430@psu.edu
Advisor: Dr. Richard Marini | rpm12@psu.edu
December 2019
A bit about me

- I work as a geospatial developer/analyst for Setton Farms. My primary job function is to track, evaluate, and model orchard health of over 200,000 acres of citrus, hazelnuts, and pistachios in the western United States, primarily in California. My focus usually begins with remote sensing of high-resolution, multi-spectral satellite imagery, performing a series of standardized and proprietary analytic indices and resulting in a determination of health, or prescription of improvement. Due to the sensitivity of my work (regarding pistachios), I am performing a focused study on the specific crop of hazelnuts.
A branch of science dealing with the relations between climate and periodic biological phenomena (such as bird migration or plant flowering) or periodic biological phenomena that are correlated with climatic conditions (Phenology | Definition of Phenology by Merriam-Webster n.d.). In horticulture, phenology generally refers to a growth stage, such as bloom, the timing of which is influenced by temperature.
Model

A system of postulates, data, and inferences presented as a mathematical description of an entity or state of affairs -also- a computer simulation (see ‘simulation’: the imitative representation of the functioning of one system or process by means of the functioning of another) based on such a system [“climate model”] (Model | Definition of Model by Merriam-Webster n.d.).
Principal Component Regression (PCR)

In principal components regression, we first perform principal components analysis (PCA) on the original data, then perform dimension reduction by selecting the number of principal components (m) using cross-validation or test set error, and finally conduct regression using the first m dimension reduced principal components (7.1 - Principal Components Regression (PCR) | STAT 508 n.d.).
Jupyter Notebooks

Is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations, and narrative text. Uses include data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more (Project Jupyter | Home n.d.).
Problem Statement

Forecasting harvest date and yield (phenological model) of selected hazelnut orchards within Oregon’s Willamette Valley using Principal Component Regression (PCR) on highly multivariate data within a novel Python-based Jupyter Notebook.
Project Location
Research Questions to be answered:

1. Can an accurate phenological model for hazelnuts (and a template for other deciduous species) be developed using highly multivariate (yet sparse) historical data?

2. Can a free and open-source Python-powered Jupyter Notebook be used in lieu of a more traditional statistical software package (such as SAS) for statistical analysis and visualization?
So, why hazelnuts?
Background

- Man has always attempted to model the physical world in mathematical simulations/models (O'Connor and Robertson n.d.)

- Phenology is just another model – ‘Agricultural Meteorology’ (Helmut Lieth (auth.) 1974)


- Never all at once
Methods/Approach

- Massive Data Collection
- Perform PCR on inputs
- Develop Phenological Model for forecasting
- Contain all logic in transferable Jupyter Notebook
Methods/Approach - cont.

- Historical yields (from the Oregon Hazelnut Industry)
- Historical weather/climate data will be obtained from *Agrimet, *Hydromet (* = closest station) and NOAA's Regional Climate Center's (RCCs) ACIS (Applied Climate Information System). (Mehlenbacher 1991; Heide 1993; Pope et al. 2015; Rahemi and Pakkish 2009; Luedeling, Zhang, Luedeling, et al. 2009)
- Pollination Period
- Budburst/bloom dates for previous years
- Harvest dates for previous years
- Vegetative Indices (NDVI, GNDVI, OSAVI, etc.)
Methods/Approach - cont.

- MN = Minimum Daily Air Temperature (F)
- MX = Maximum Daily Air Temperature (F)
- MM = Mean Daily Air Temperature (F)
- Budburst/bloom dates for previous years
- Chilling hours
- Growing Degree Days (GDD) (calculated)
- ET = Evapotranspiration Kimberly-Penman (in)
- PC = Accumulated Precipitation Since Recharge/Reset (in)

- PP = Daily (24 hour) Precipitation (in)
- PU = Accumulated Water Year Precipitation (in)
- SR = Daily Global Solar Radiation (langley's)
- TA = Mean Daily Humidity (%)
- YM = Mean Daily Dewpoint Temperature (F)
- UA = Daily Average Wind Speed (mph)
- UD = Daily Average Wind Direction (deg az)
- WG = Daily Peak Wind Gust (mph)
Anticipated Results

1. Successfully develop a rough phenological model used to forecast potential harvest dates and yields for hazelnuts in the Willamette Valley, OR.

2. Successfully memorialize project process and methods within a Jupyter Notebook; to be used in future studies.
Proposed Timeline

- Now
- Collect Data
- Perform Study
- Present Findings
- April/May 2020
Proposed Applications/Future Work

1. Basis for other studies of similar deciduous phenology

2. Modular notebook approach could allow data to be evaluated from several different techniques if proposed does not produce a usable model

3. Modular Notebook could serve as the basis for more exhaustive evaluations such as local or cloud-based machine learning efforts
Possible Venues

1. Potential presentation for the Hazelnut Industry of Oregon
2. Potential presentation for the International Nut & Dried Fruit Council
3. Potential presentation for the American Pistachio Growers (APG) Annual convention, February 2020
4. Potential presentation as published white paper
5. No definitive venue has been decided upon as of the writing of the draft and the date of this presentation


References


- n.d. 7.1 - Principal Components Regression (PCR) | STAT 508. https://newonlinecourses.science.psu.edu/stat508/lesson/7/7.1.
References

Thank you for listening