Predicting Evapotranspiration from Meteorological data & Vapor Pressure Gradient Error Predictions

by Dalyn McCauley

## Problem Statement, Motivation, Research Goals:

An essential component in the water balance is Evapotranspiration (ET) which is the processes of evaporation from open water and bare soil and transpiration from vegetation happening simultaneously. ET can be difficult to measure directly, unlike other components like runoff and precipitation, and so it is often estimated by meteorological data. Net radiation, air temperature, and relative humidity are driving factors in ET, but unfortunately the resolution required for precise ET estimates are often beyond the feasibility of affordable sensing equipment. Of particular concern is the error of relative humidity sensors, as the vapor pressure gradient is a vital aspect of the ET equation. I propose to analyze meteorological data to predict reference ET as well as attempt to predict when the data from the humidity sensors is unreliable, which is when the vapor pressure gradient is smaller than the bounds of uncertainty of the humidity sensors. Additionally, I am also interested in determining which of the two radiation sensors are more robust in estimating ET, quantum or thermopile. The goal of this analysis is to expediate the data pre-processing of meteorological data by flagging data that might be unreliable and ultimately develop an automated and low-cost approach to determining ET. If the algorithms designed in this project prove useful and robust, they will be incorporated into the programming of an Arduino based mobile weather stations. These low energy low cost stations can be placed in Vineyards or orchards to obtain a high-resolution map of ET across the landscape in hopes of fine-tuning crop water demand and limiting consumptive use of water in agriculture.

## Data Source and Description:

The data were collected at a Hazelnut Orchard in Amity, Oregon in summer of 2017 by Dr. Jason Kelley in the department of Soils and Water Systems. The data were collected over 85 days, from June 12, 2017 to September 9th, 2017. The data consists of ten variables, shown in Table 1. There are two radiation measurements, one from an Apogee quantum sensor and one from a Q7 thermopile sensors. The two humidity and temperature measurements at different heights provide the temperature and vapor gradients that drive ET. The wind measurement is from a sonic anemometer and is used to determine the aerodynamic effects on ET so the vapor being carried away from the measurement control volume can be determined.

## Literature Review and References:

Evapotranspiration can be measured in many different ways. I am using the energy balance approach in this project. The energy balance approach initially stemmed from H.L. Penman’s 1948 paper *Natural Evaporation from open water, bare soil and turf* whichdetails the assumptions and equations necessary to derive an equation for natural evaporation using only meteorological data. Penman defines two approaches, an aerodynamic approach and an energy balance approach. The only measured parameters necessary to calculate evaporation using Penman’s combination equation are mean air temperature, mean dewpoint (or relative humidity), mean wind velocity and mean duration of sunshine (radiation) (Penman, 1948) . Many assumptions are made throughout the derivation:

* Ideally restricted to a field after thorough wetting
* There is a zero-temperature gradient between Tsurface and Tair
* The changes in stored heat and heating of the test material surroundings is negligible over the period of several days

Monteith later revised the original Penman equation to build a fundamental equation for calculating reference evapotranspiration (“ASCE Manual 70 – Second Edition,” 2015). Reference ET represents the upper limit of ET given the available energy and meteorological measurements applied over a reference surface, often clipped grass or alfalfa. A crop coefficient factor is applied to the reference ET value to obtain the estimated ET from the vegetation of interest (R. G. Allen, Pereira, Howell, & Jensen, 2011). In this case, a crop coefficient for hazelnut orchards will be used.

## Preliminary EDA:

The initial dataset consists of nine variables, and 24,585 samples. Ten intermediate columns were created in order to calculate the reference ET, making the final dataset a 19 by 24585 matrix, equating to a size of 467,115. The variable descriptions and datatypes can be found in Table 1.

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Description** | **Measured/**  **Calculated** | **Data type** |
| TIMESTAMP | Date and time [YYYY:MM:DD hh:mm:ss] | Measured | datetime64[ns] |
| Q7corr | Corrected net radiation using thermopile technology [W/m2] | Measured | float64 |
| ApogeeWm2 | net radiation using quantum technology [W/m2] | Measured | float64 |
| HMP2m\_T\_avg | Averaged temperature at 2m [°C] | Measured | float64 |
| HMP5m\_T\_avg | Averaged temperature at 5m [°C] | Measured | float64 |
| HMP2m\_RH\_avg | Averaged relative humidity at 2m [%] | Measured | float64 |
| HMP5m\_RH\_avg | Averaged relative humidity at 5m [%] | Measured | float64 |
| Wspd | Horizontal windspeed from a sonic anemometer [m/s] | Measured | float64 |
| P | Air pressure [kPa] | Measured | float64 |
| air\_d | Air density [kg/m3] | Calculated | int64 |
| aero | Aerodynamic resistance [m/s] | Calculated | int64 |
| delta | Slope of the saturation vapor pressure -temperature curve [kPa/°C] | Calculated | int64 |
| e\_0 | Saturation vapor pressure [kPa] | Calculated | int64 |
| e\_a | Vapor pressure [kPa], | Calculated | int64 |
| gamma | Psychrometric constant [kPa/°C] | Calculated | int64 |
| Gflux | Ground heat flux [W/m2] | Calculated | int64 |
| VPG | Vapor pressure gradient [kPa], | Calculated | int64 |
| VPGError | Indicates when the vapor pressure gradient is too small to yield reliable ET estimate (0 when VPG > = 0.2, 1 when VPG < 0.2 \*error) | Calculated | int64 |
| ET\_ref | Calculated reference ET [mm/5min] | Calculated | int64 |

The data were preprocessed by applying a moving 5 minute average to the 10Hz samples. This moving average smooths the data and reduces the sensor noise. Additional preprocessing work was done in deriving the value for evapotranspiration. The calculation method is outlined by Allen et al in the FAO 56 irrigation manual (R. Allen, 1998). The intermediate calculations and variables are shown and briefly described below.

The aerodynamic resistance is a function of the measurement height and the roughness length, and the surface resistance is a function of the stomatal resistance the leaf, rl, and the amount of vegetation area, known as the Leaf Area Index (LAI).

Air density is calculated from the ideal gas law:

Saturation vapor pressure is the pressure at which water vapor in the air condenses and is a function of temperature:

The actual vapor pressure is the pressure from water vapor in air and is a function of the saturation vapor pressure and the humidity, q:

The vapor pressure gradient is the difference between saturation vapor pressure and actual vapor pressure:

The slope of the saturation vapor pressure – temperature curve is defines as Δ and is a function of temperature and saturation vapor pressure:

The psychrometric constant relates the partial pressure water in the air to air pressure, where Mw is the molecular weight of water and L is the latent heat of evaporation:

Finally, the available energy (R – G), vapor pressure gradient (VPG) and resistance terms are combined to calculate the reference evapotranspiration rate for the given meteorological conditions:

The results of this preprocessing are shown in the Figure 1. The top image shows the daily cycle of ET over the course of the summer, and the bottom plot shows the cumulative ET expressed in mm. When this value is multiplied over an area, a volume of water lost to evapotranspiration can be estimated.

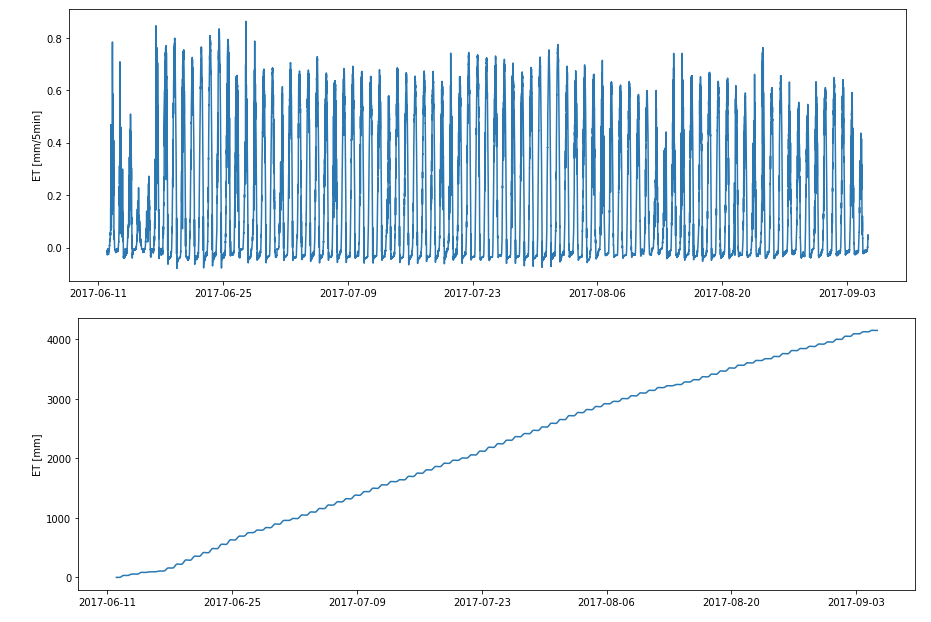


Figure 1

Modeling Process:

### Supervised Preliminary Results

Simple Linear regression was used as a baseline model for the supervised approach to estimating ET rates. From a first look, it performed curiously well. A training period of four days was used, from June 12, 2017 to June 18th, 2018. This period was selected because it had representative range in ET rates. For example, two days with very low ET were included to incorporate meteorological data from cloudy or cold days into the model. The model was fitted using only the measured meteorological data, which did not include the calculated components, Figure 2.

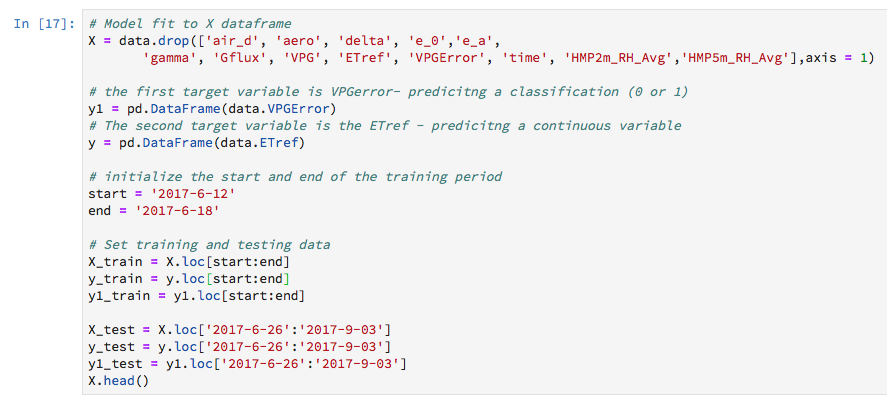


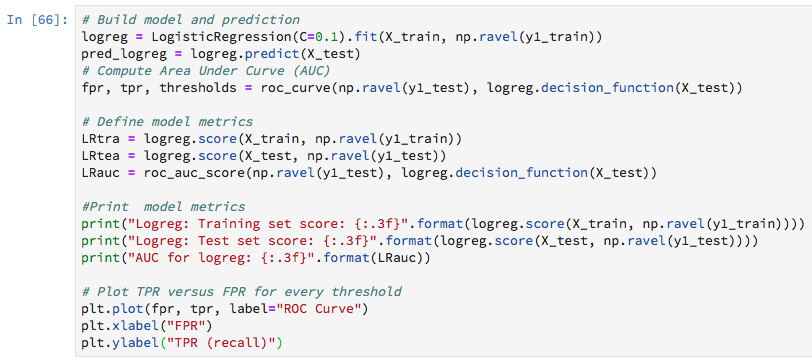
Figure 2

The figure below shows the predicated versus the calculated ET rate, and also the predicted versus calculated cumulative ET over the course of the summer. The fit was very high with an R2 value of 0.95. The main discrepancies between the two series are seen in the inability of the model to accurately predict the daily peaks of ET, and also the fact that linear regression allows for negative values. Physically, negative values of ET are not possible. This will be one aspect of the model to improve for the final report. Even though the model has a high accuracy, the error is still great enough to create significant differences in the total water lost to evaporation over time.

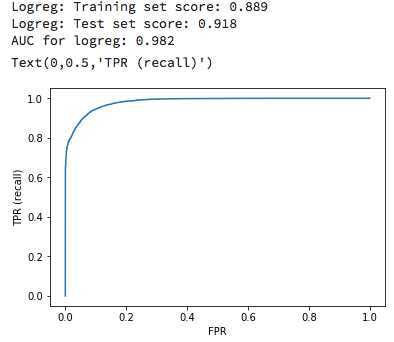
### 

Figure 3

Predicting reference ET was not the only goal of this project, there was also some success in using supervised learning to classify when the data was unreliable due to the vapor pressure gradient. The same training period was used for the VPG error classification. Four classification models were used for the preliminary analysis, K-nearest neighbor classifier, Logistic Regression, Gradient Boosting Classifier and Decision Tree Classifier. The best performer was logistic regression with a training accuracy of 92% and an AUC value of 0.98. The code and results for this model are shown in Figure 4 and Figure 5, respectively.



Figure

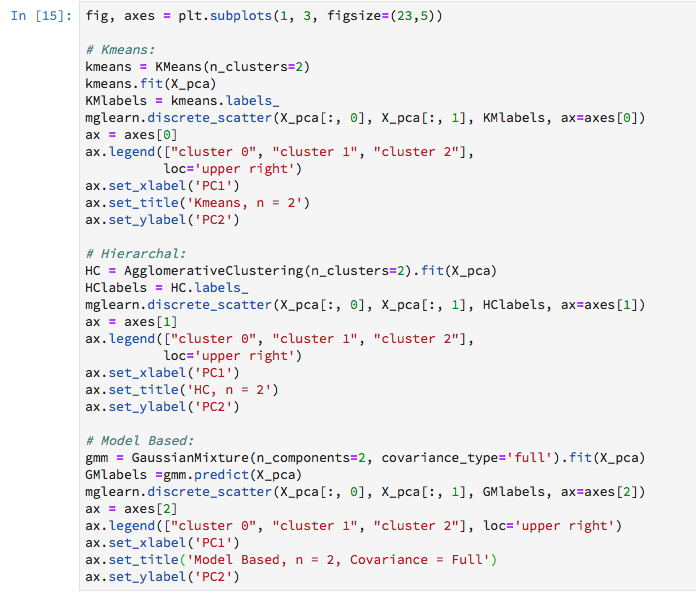


Figure

The successful results of the logistic regression indicate that with a training period of four days the model would be able to take in meteorological data and flag data that is most likely to have too small of a VPG to be reliable. From here a simple filtering could be applied to get rid of bad data before computing and accumulating reference ET results.

### Unsupervised Preliminary Results

Three unsupervised models were used to classify whether or not the vapor pressure gradient is too small to give reliable data, K-means clustering, Hierarchal clustering and gaussian mixture model based clustering. Principal component analysis was used first, and the first three principle components were kept. All three models were set to create two clusters, in hopes of classifying the data into two groups, good data and bad data. All three models performed very poorly. The best result was from k-means clustering where there was a 32% success rate in identifying the unreliable data, this of course means that it also falsely identified around 70% of the data. The code and results are shown in Figure 6 and 7, respectively.



Figure

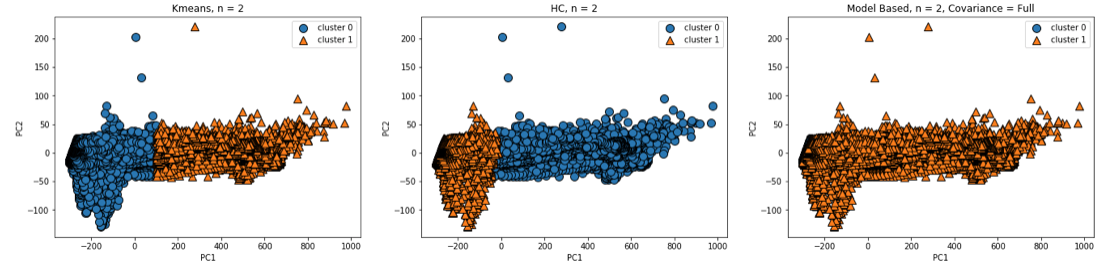


Figure 7

## Project Progress, Timeline, and Achievement:

The progress is progressing as to be expected. The initial data preprocessing and reference ET calculations took longer than expected, around two weeks. The supervised learning models were developed quite easily, but the unsupervised clustering has been challenging. The rest of the semester will be used to optimize the models outlined in this paper.

Major achievements made this far are as follows:

1. Calculating reference ET from meteorological data.
2. Learning how to use pandas time series data as an index for easy plotting and data partitioning.
3. Supervised linear regression predicts reference ET very well.
4. Supervised classification predicts VPG error very well.
5. Clustering analysis for VPG error prediction improved from a true positive rate of 20% to 32%.

The significance of my work thus far is that I can successfully predict evapotranspiration from a training period of only three days and I can predict with sufficient accuracy when the meteorological data collected will lead to erroneous ET measurements.

## Conclusions and Possible Future Work:

In conclusion, the preliminary results of this project are promising. There is still a lot of room for improvements especially in the unsupervised modeling. There is one objective of my project that might change due to data availability. This report outlines the result in predicating reference ET, which is a calculated variable. I am trying to obtain *actual* ET values from an eddy covariance system. Eddy covariance is a direct measurement of the water vapor in the air above vegetation. I think this will be a more challenging variable to predict because it is not directly related via the Penman-Monteith equation to the meteorological data.

## References:

Allen, R. (1998). FAO Irrigation and Drainage Paper No. 56.

Allen, R. G., Pereira, L. S., Howell, T. A., & Jensen, M. E. (2011). Evapotranspiration information reporting: I. Factors governing measurement accuracy. *Agricultural Water Management*, *98*(6), 899–920. https://doi.org/10.1016/j.agwat.2010.12.015

ASCE Manual 70 – Second Edition: Evaporation, Evapotranspiration and Irrigation Requirements. (2015). In *2015 ASABE / IA Irrigation Symposium: Emerging Technologies for Sustainable Irrigation - A Tribute to the Career of Terry Howell, Sr. Conference Proceedings* (pp. 1–16). American Society of Agricultural and Biological Engineers. https://doi.org/10.13031/irrig.20152143358

Penman, H. L. (1948). Natural Evaporation from Open Water, Bare Soil, and Grass. *JSTOR*, 120–145.