

CS293S: Internet of Things

# An In-Depth Analysis on Weather Data from CIMIS: Estimating Evapotranspiration (ET) Values

Nazmus Saquib   Udit Paul   Alex Ermakov   Santha Ramamoorthy

Graduate Students  
Department of Computer Science  
University of California Santa Barbara

March 13, 2019



# Outline

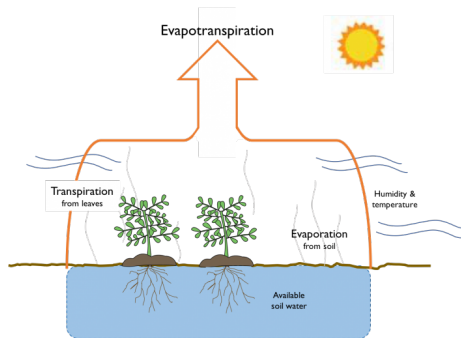
- 1 Introduction
- 2 Data Collection
- 3 Data Overview
- 4 Feature Selection
- 5 Regression Analysis
- 6 Nearest Neighbor Analysis
- 7 Demo
- 8 Questions

# Outline

- 1 Introduction
- 2 Data Collection
- 3 Data Overview
- 4 Feature Selection
- 5 Regression Analysis
- 6 Nearest Neighbor Analysis
- 7 Demo
- 8 Questions

# Introduction: Evapotranspiration (ET)

- Loss of water through:
  - 1 Evaporation and
  - 2 Transpiration
- Applications:
  - Irrigation scheduling
  - Water resource planning, etc.



# Introduction: CIMIS Weather Stations

- *California Irrigation Management Information System*
- 257 CIMIS stations all through California
  - 136 actively reports ET values
- Measures various weather parameters
- some directly influence ET
- Also measures (*calculates?*) ET

# Outline

- 1 Introduction
- 2 Data Collection**
- 3 Data Overview
- 4 Feature Selection
- 5 Regression Analysis
- 6 Nearest Neighbor Analysis
- 7 Demo
- 8 Questions

# Data Collection

- Publicly available API
- Reports both hourly and daily data
- A record contains 15 different features
- Current working dataset: data of last one year
- Certain analysis uses data from multiple years to capture seasonal variations

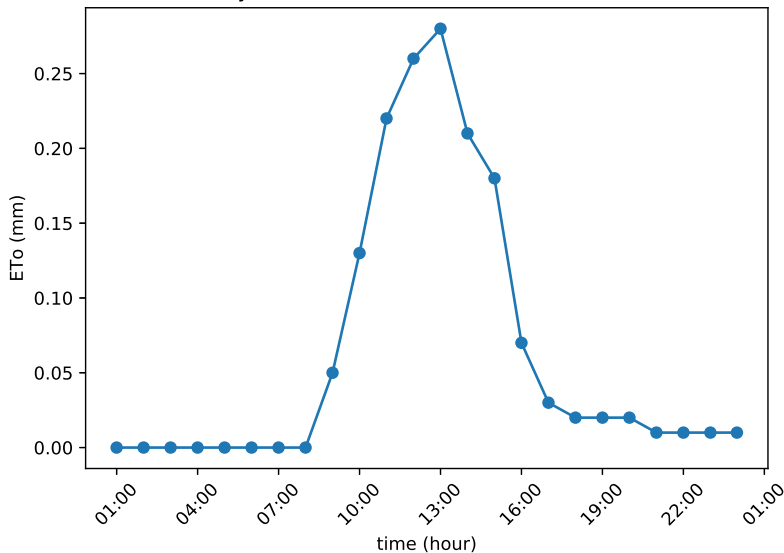
# Outline

- 1 Introduction
- 2 Data Collection
- 3 Data Overview**
- 4 Feature Selection
- 5 Regression Analysis
- 6 Nearest Neighbor Analysis
- 7 Demo
- 8 Questions

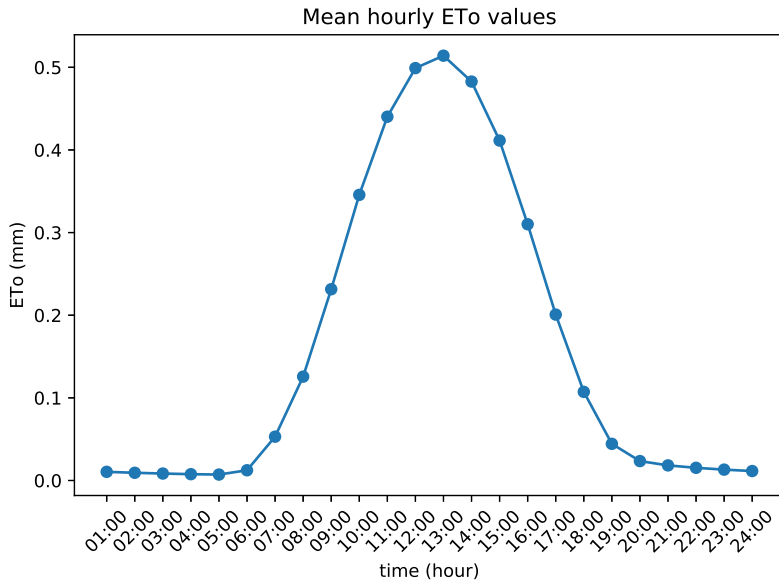


# Sample Hourly ET Values

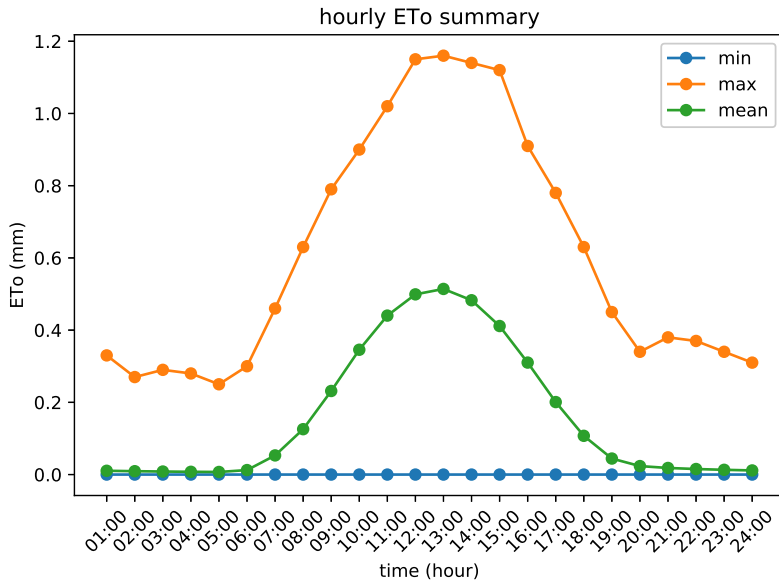
Hourly ETo values for station 2 on 2018-01-01

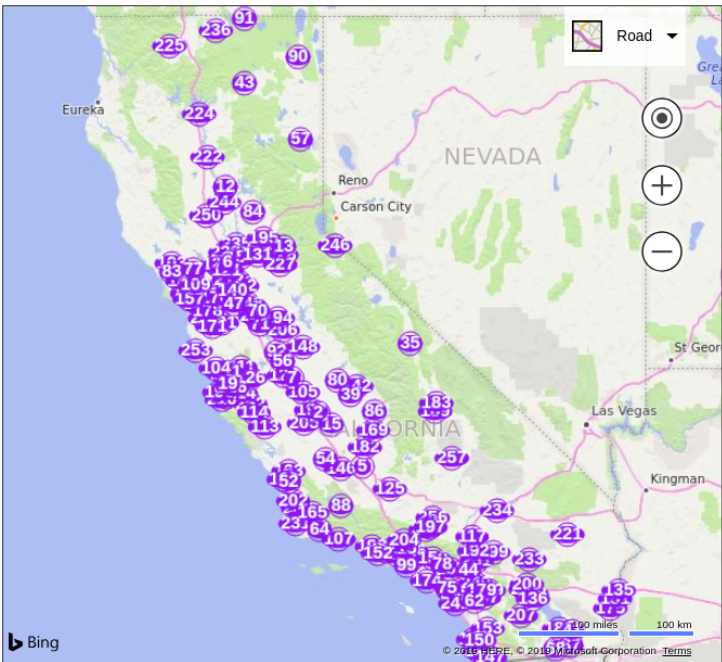


# Mean Hourly ET Values



# Min/Mean/Max Hourly ET Values

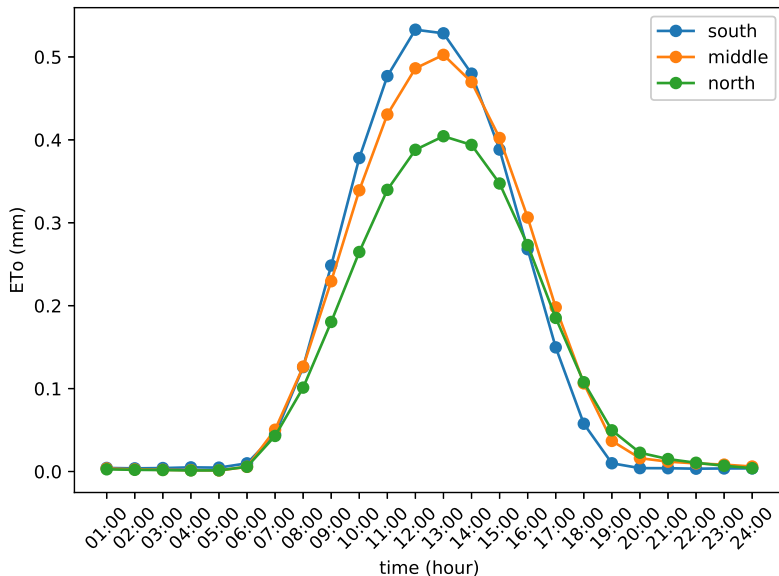




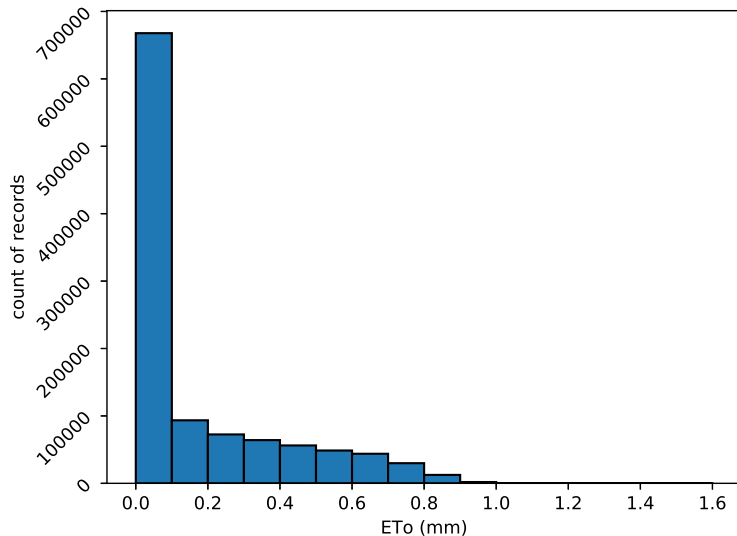
# Stations of Interest

- Station with lowest latitude  $LAT_{MIN}$  (south)
- Station with highest latitude  $LAT_{MAX}$  (north)
- Station with latitude closests to  $\frac{LAT_{MIN} + LAT_{MAX}}{2}$  (middle)

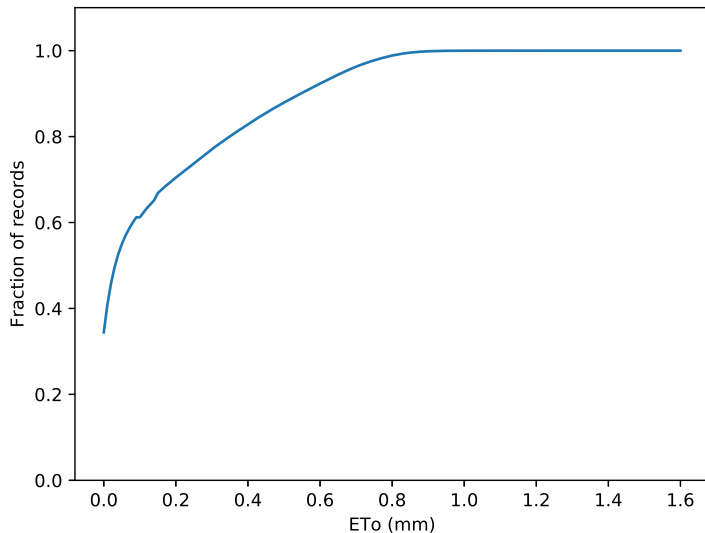
# Mean Hourly ET Values of Stations of Interest



# Histogram of ET Values



# Empirical CDF of ET Values





# Outline

- 1 Introduction
- 2 Data Collection
- 3 Data Overview
- 4 Feature Selection**
- 5 Regression Analysis
- 6 Nearest Neighbor Analysis
- 7 Demo
- 8 Questions

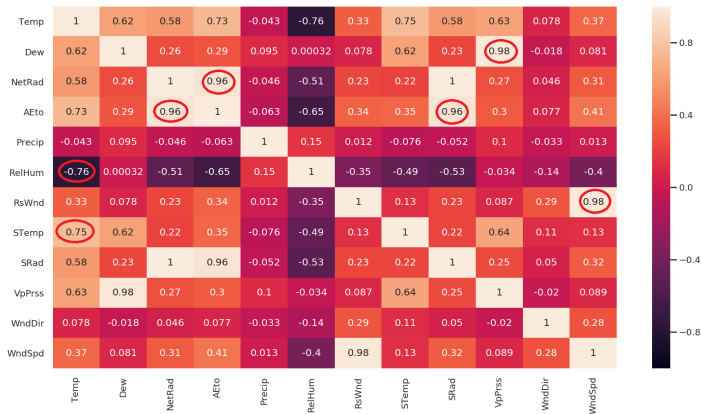
# Correlation Analysis

To determine the most important features in our data, we first of all perform a correlation analysis. The result of the analysis is demonstrated in the figure overleaf.

From the figure we can see that a lot of parameters are highly correlated with some other features. A very good example of it is the parameter Dew which has 98 percent correlation with the parameter Vapor Pressure.

# Correlation Analysis - Result

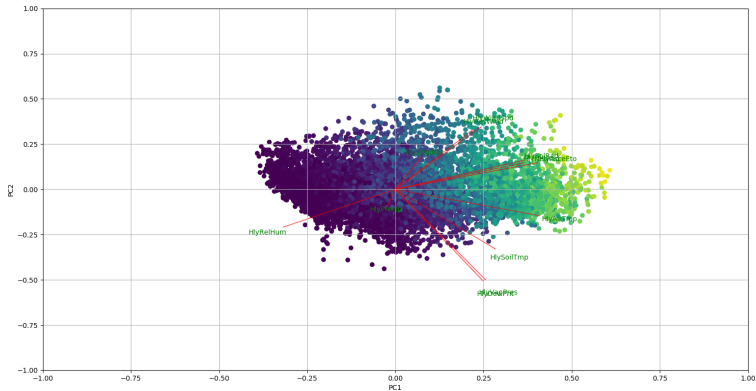
Correlation between the parameters



We then visualize the data using first two components of the Principal Component Analysis (PCA). PCA is a commonly used tool that reduces the dimensionality of a dataset by determining the subset of features that captures the highest amount of variance within the data. The biplot diagram overleaf presents the relative "importance" of each of our feature on the over all data(length of each feature line emphasize its importance). Based on these analysis, we proceed to select 6 most important features namely: Temperature, Radiation, Precipitation, Vapor Pressure, Wind direction and Wind speed.

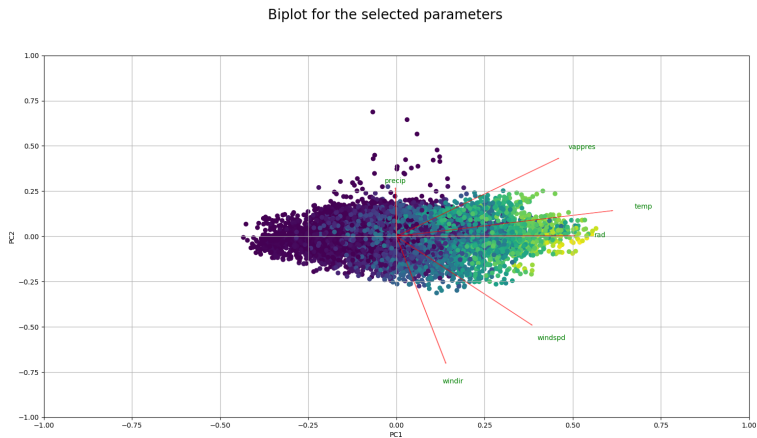
# Biplot-all

Biplot for all parameters



# Biplot-selected

Figure below shows the biplot view of the selected features.

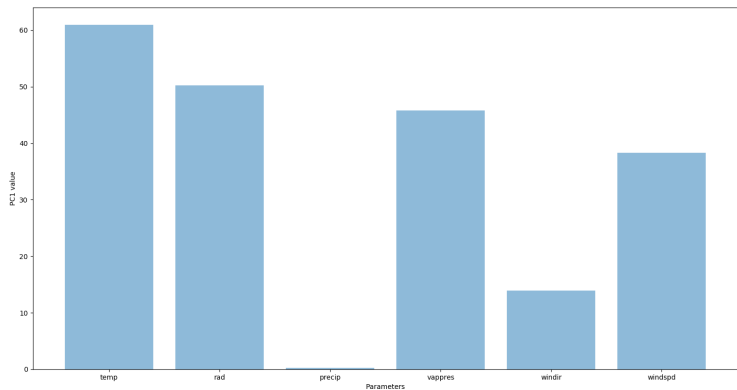


We finally check the eigenvalues of our selected features on the first principal component eigenvectors. The figure overleaf shows the value each of our parameter possess in this analysis (the values have been multiplied by 10.)

Based on these values, for the rest of our analysis for this project, we use the features, Temperature, Radiation, Vapor Pressure and Wind Speed (the four most important parameters according to PCA.)

# PC1 values

First order principal component values





# Outline

- 1 Introduction
- 2 Data Collection
- 3 Data Overview
- 4 Feature Selection
- 5 Regression Analysis**
- 6 Nearest Neighbor Analysis
- 7 Demo
- 8 Questions

# Estimation of ET Values

Given a set of features, can we estimate ET?

- Which features to choose?
- *How well* is our estimate?

Given a set of features, can we estimate ET?

- Which features to choose?
- *How well* is our estimate?

# (CIMIS) Penman Monteith Equation for Calculating ET

$$ET_o = \frac{\Delta(R_n - G)}{\lambda[\Delta + \gamma(1 + C_d u_2)]} + \frac{\gamma \frac{37}{T_a + 273.16} u_2 (e_s - e_a)}{\Delta + \gamma(1 + C_d u_2)}$$

Ultimately depends on four weather features

- Solar net radiation
- Vapor pressure
- Air temperature
- Wind speed

# (CIMIS) Penman Monteith Equation for Calculating ET

$$ET_o = \frac{\Delta(R_n - G)}{\lambda[\Delta + \gamma(1 + C_d u_2)]} + \frac{\gamma \frac{37}{T_a + 273.16} u_2 (e_s - e_a)}{\Delta + \gamma(1 + C_d u_2)}$$

Ultimately depends on four weather features

- Solar net radiation
- Vapor pressure
- Air temperature
- Wind speed

# Regression Results

Features	Mean Squared Error	$R^2$ Value
HlyAirTmp,HlyNetRad,HlyVapPres,HlyWindSpd	0.000970123960314	0.9812940161
HlyAirTmp,HlyNetRad,HlyVapPres	0.00130358866256	0.9747612206
HlyAirTmp,HlyNetRad,HlyWindSpd	0.00131186536214	0.9745279825
HlyAirTmp,HlyNetRad	0.00173654973306	0.9665370047
HlyNetRad,HlyVapPres,HlyWindSpd	0.00248645097725	0.9520098573
HlyNetRad,HlyWindSpd	0.0024909080494	0.9516599092
HlyNetRad,HlyVapPres	0.00302176798112	0.9410658003
HlyNetRad	0.00304665078019	0.9409558541
HlyAirTmp,HlyVapPres,HlyWindSpd	0.0236668111725	0.540318481
HlyAirTmp,HlyWindSpd	0.0242823252297	0.5285606181
HlyAirTmp,HlyVapPres	0.026563048828	0.4850281600
HlyAirTmp	0.0278295291341	0.4597101537
HlyVapPres,HlyWindSpd	0.0407552684279	0.2088275258
HlyWindSpd	0.0412914020576	0.1961185540
HlyVapPres	0.0510006461517	0.0128578989

# Regression Results

Features	Mean Squared Error	$R^2$ Value
HlyAirT <sub>mp</sub> ,HlyNetRad,HlyVapPres,HlyWindSpd	0.000970123960314	0.9812940161
HlyAirT <sub>mp</sub> ,HlyNetRad,HlyVapPres	0.00130358866256	0.9747612206
HlyAirT <sub>mp</sub> ,HlyNetRad,HlyWindSpd	0.00131186536214	0.9745279825
HlyAirT <sub>mp</sub> ,HlyNetRad	0.00173654973306	0.9665370047
HlyNetRad,HlyVapPres,HlyWindSpd	0.00248645097725	0.9520098573
HlyNetRad,HlyWindSpd	0.0024909080494	0.9516599092
HlyNetRad,HlyVapPres	0.00302176798112	0.9410658003
HlyNetRad	0.00304665078019	0.9409558541
HlyAirT <sub>mp</sub> ,HlyVapPres,HlyWindSpd	0.0236668111725	0.540318481
HlyAirT <sub>mp</sub> ,HlyWindSpd	0.0242823252297	0.5285606181
HlyAirT <sub>mp</sub> ,HlyVapPres	0.026563048828	0.4850281600
HlyAirT <sub>mp</sub>	0.0278295291341	0.4597101537
HlyVapPres,HlyWindSpd	0.0407552684279	0.2088275258
HlyWindSpd	0.0412914020576	0.1961185540
HlyVapPres	0.0510006461517	0.0128578989

# Outline

- 1 Introduction
- 2 Data Collection
- 3 Data Overview
- 4 Feature Selection
- 5 Regression Analysis
- 6 Nearest Neighbor Analysis**
- 7 Demo
- 8 Questions



# Nearest Neighbor Analysis

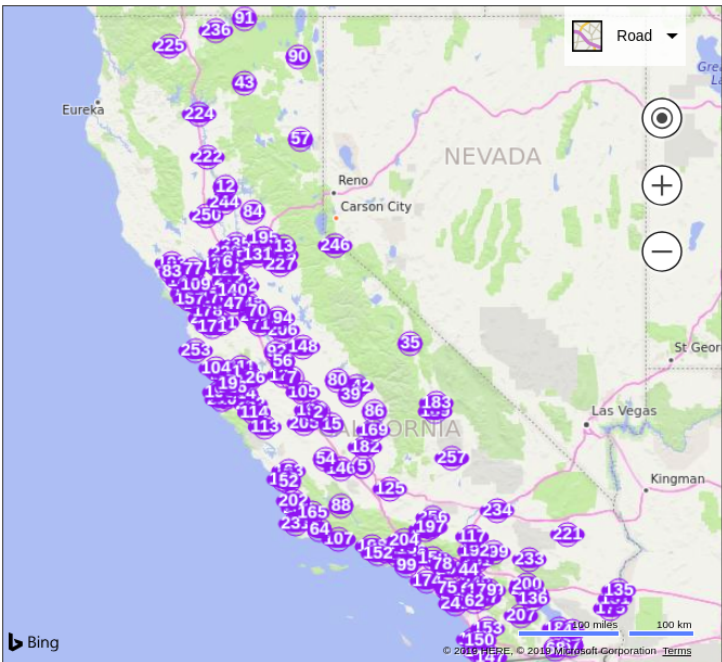
Given the ET value of  $k$  nearest stations of a place, can we estimate ET?

- Arithmetic mean of  $k$  values
- Inverse Distance Weighted (IDW) average of  $k$  values

# Nearest Neighbor Analysis

Given the ET value of  $k$  nearest stations of a place, can we estimate ET?

- Arithmetic mean of  $k$  values
- Inverse Distance Weighted (IDW) average of  $k$  values

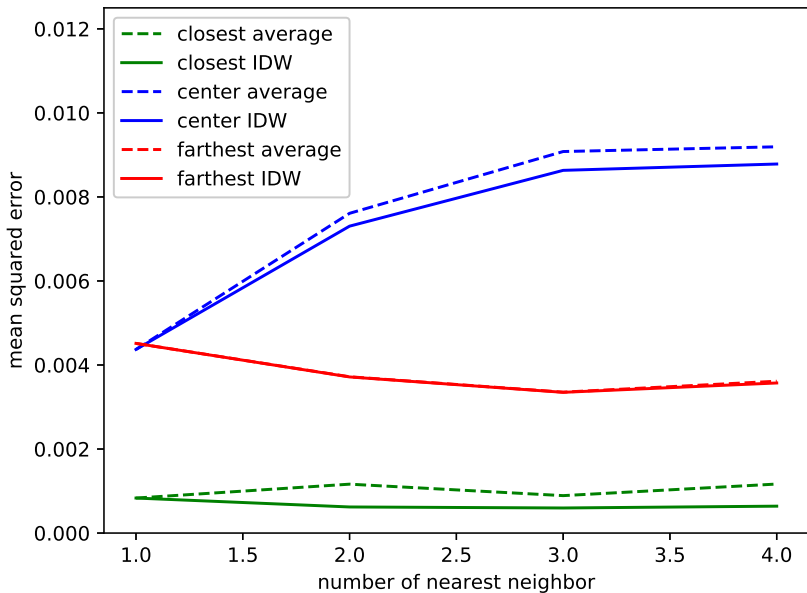


# Stations of Interest

- Station with lowest distance  $D_{MIN}$  to nearest neighbor
- Station with highest distance  $D_{MAX}$  to nearest neighbor
- Station with nearest neighbor at a distance closest to  $\frac{D_{MIN} + D_{MAX}}{2}$

# Nearest Neighbor Results

Station Number	Num of Neighbors	MSE for Average	MSE for IDW
129	1	0.000832971114168	0.000832971114168
234	1	0.00437018526497	0.00437018526497
57	1	0.00451400872516	0.00451400872516
129	2	0.00116361600992	0.000620877927137
234	2	0.00761026004119	0.00730456269316
57	2	0.00371994564336	0.0037154634375
129	3	0.000890784115612	0.000596760525931
234	3	0.00908058999082	0.00863260116925
57	3	0.00335367604618	0.00334925237208
129	4	0.00116647617403	0.00063999172153
234	4	0.00919325287807	0.00878339044833
57	4	0.00361403432169	0.00357201358681



# Nearest Neighbor with Sensor Values

What if we have sensor values from nearby stations instead of only ET values?

MSE decreases according to CIMIS Penman Equation

What if we have sensor values from nearby stations **along with local air temperature?**

MSE decreases even further

# Nearest Neighbor with Sensor Values

What if we have sensor values from nearby stations instead of only ET values?

MSE decreases according to CIMIS Penman Equation

What if we have sensor values from nearby stations **along with local air temperature?**

MSE decreases even further



# Nearest Neighbor with Sensor Values

What if we have sensor values from nearby stations instead of only ET values?

MSE decreases according to CIMIS Penman Equation

What if we have sensor values from nearby stations **along with local air temperature?**

MSE decreases even further

# Nearest Neighbor with Sensor Values

What if we have sensor values from nearby stations instead of only ET values?

MSE decreases according to CIMIS Penman Equation

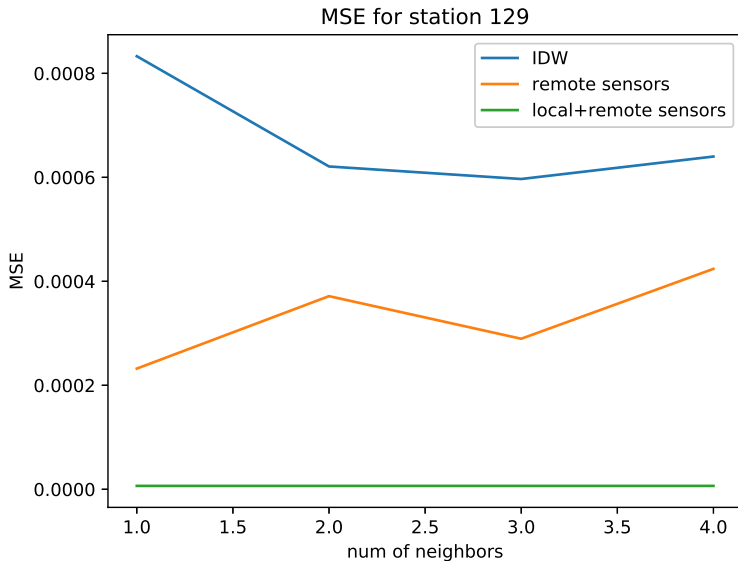
What if we have sensor values from nearby stations **along with local air temperature?**

MSE decreases even further

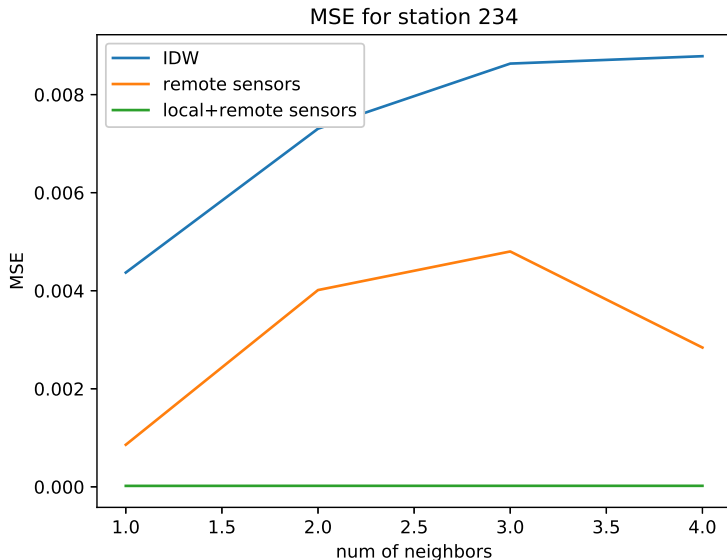
## Nearest Neighbor with Sensor Values (*Cntd.*)

Stn No	Num of Nbrs	MSE IDW	MSE	MSE Local
234	1	0.00437018	0.00085818	0.00001899
234	2	0.00730456	0.00401320	0.00002029
234	3	0.00863260	0.00480052	0.00002029
234	4	0.00878339	0.00284048	0.00002034
129	1	0.00083297	0.00023209	0.00000650
129	2	0.00062087	0.00037144	0.00000649
129	3	0.00059676	0.00028950	0.00000649
129	4	0.00063999	0.00042391	0.00000650
57	1	0.00451400	0.00044228	0.00000982
57	2	0.00371546	0.00051686	0.00000982
57	3	0.00334925	0.00031521	0.00000982
57	4	0.00357201	0.00042077	0.00000983

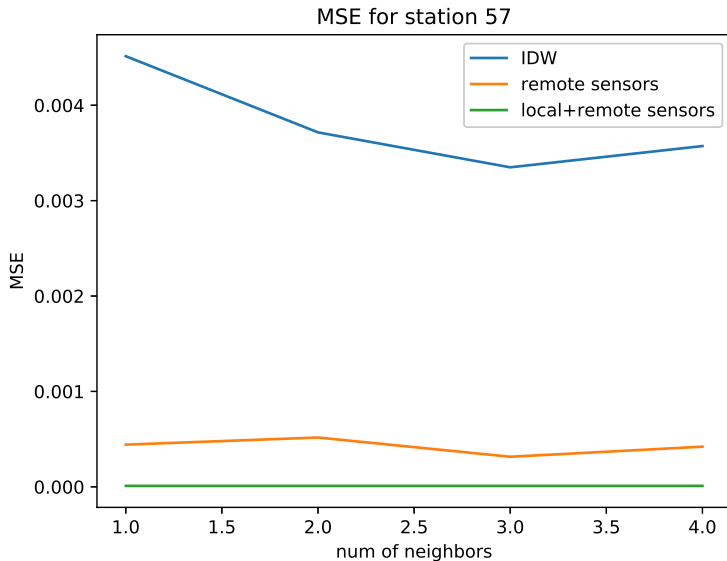
# Nearest Neighbor with Sensor Values (*Cntd.*)



# Nearest Neighbor with Sensor Values (Cntd.)



# Nearest Neighbor with Sensor Values (*Cntd.*)



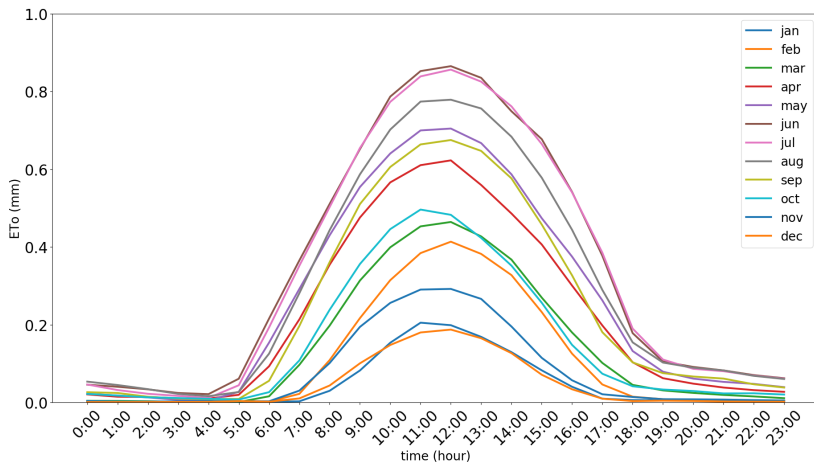
# Outline

- 1 Introduction
- 2 Data Collection
- 3 Data Overview
- 4 Feature Selection
- 5 Regression Analysis
- 6 Nearest Neighbor Analysis
- 7 Demo**
- 8 Questions

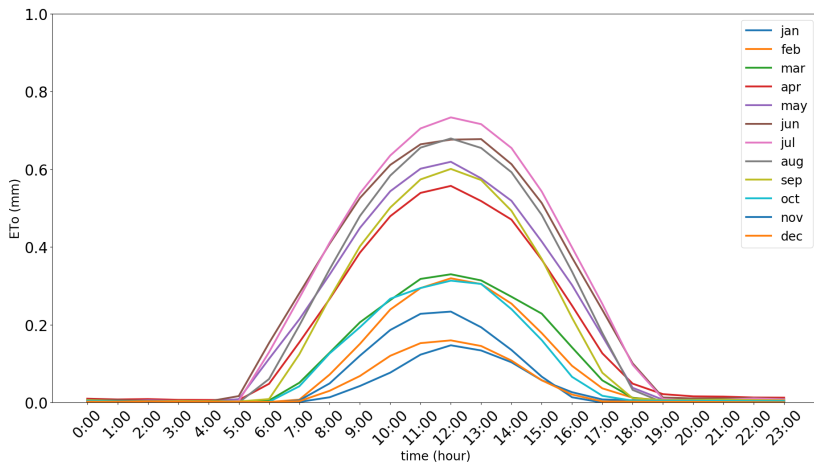
# DEMO



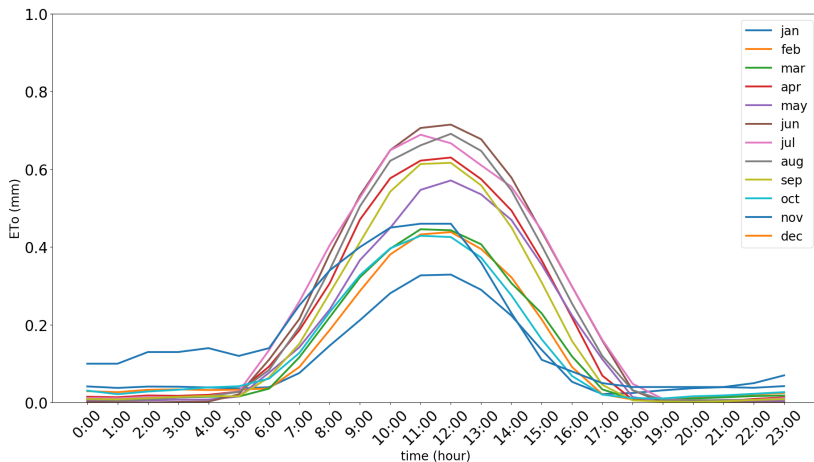
# Normal 12-month Graph for Station 2 in 2016



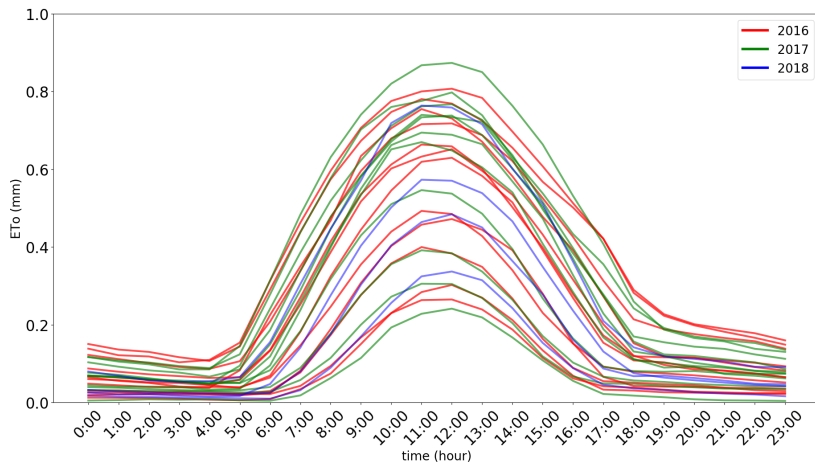
# 12-month Graph for Station 12 in 2016



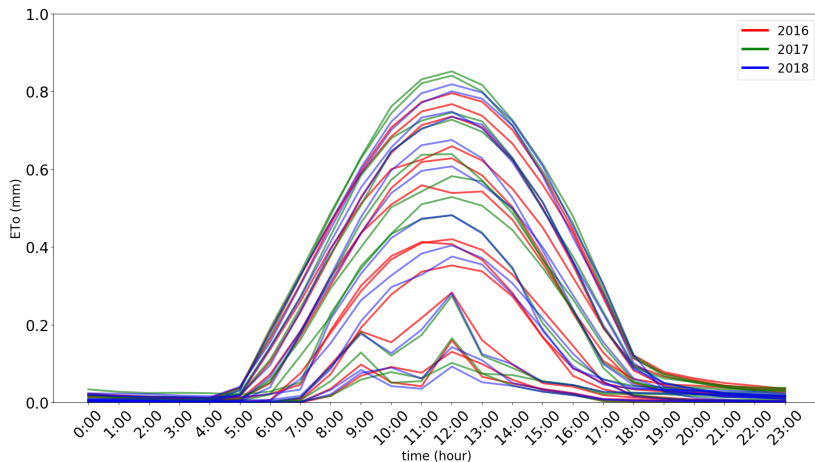
# 12-month Graph for Station 62 in 2018



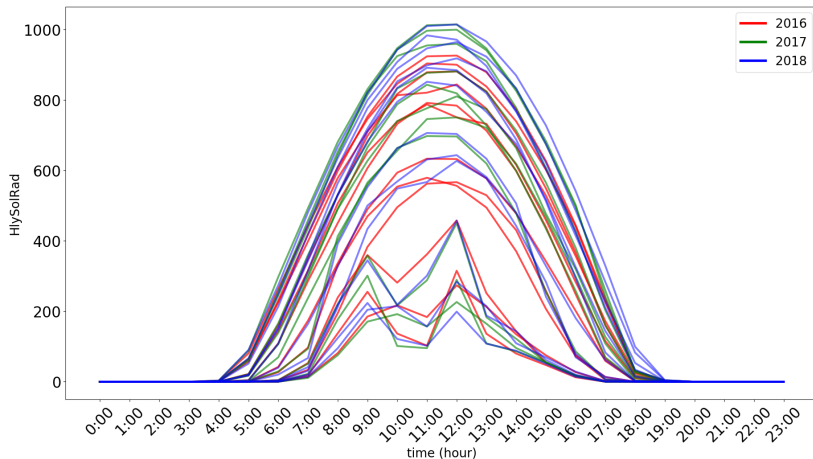
# 36-month Graph for Station 2



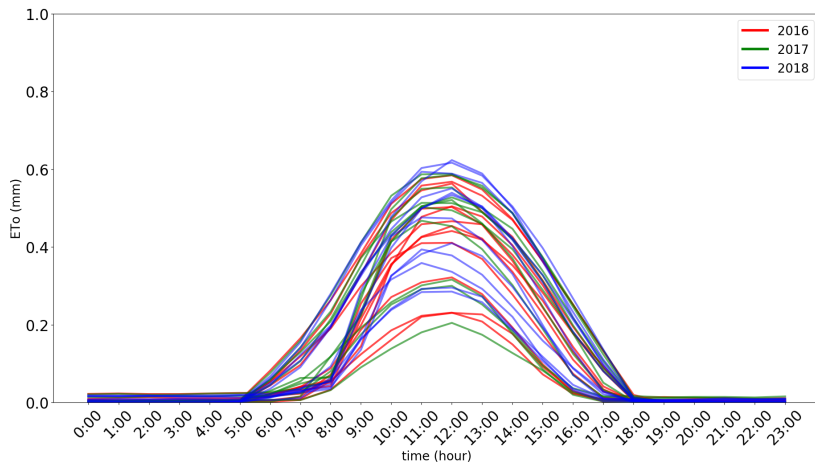
# 36-month Graph for Station 7



# 36-month Solar Graph for Station 7



# 36-month Graph for Station 202



# Outline

- 1 Introduction
- 2 Data Collection
- 3 Data Overview
- 4 Feature Selection
- 5 Regression Analysis
- 6 Nearest Neighbor Analysis
- 7 Demo
- 8 Questions**



# Questions?