PREDICTING GROUND HEAT FLUX & HUMIDITY SENSOR ERROR FROM METEOROLOGICAL DATA

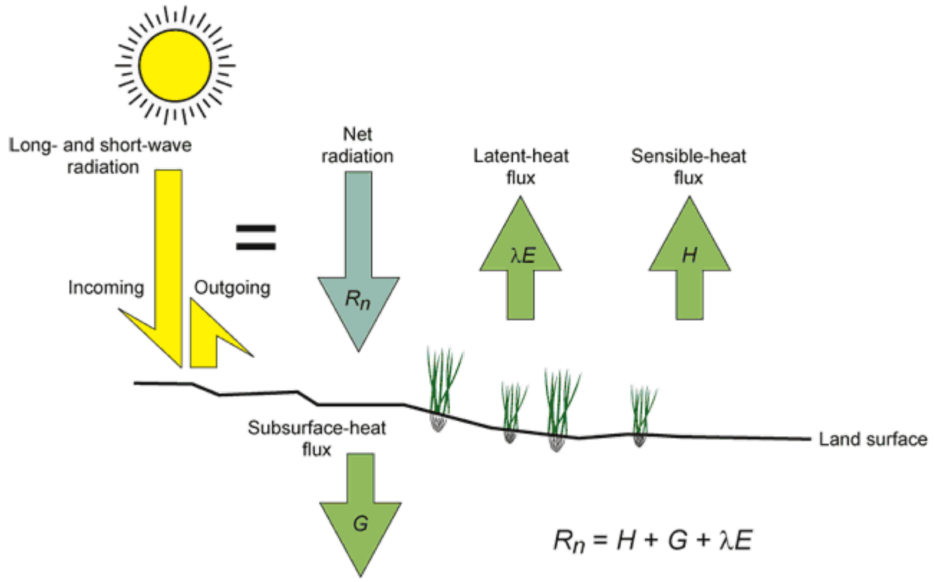
DALYN MCCAULEY

OUTLINE

- Data source and description
- Supervised
    - Ground Flux Regression
- Unsupervised Clustering
    - Error Classification
- Associations
    - Error Classification

RESEARCH QUESTION

- Can machine learning algorithms be trained to estimate ground heat flux from meteorological data to eliminate need for erroneous and expensive ground flux plates?
- Can machine learning algorithms be used to flag errors from Vapor Pressure Deficit (VPD) term in evapotranspiration estimate?



(Modified from DeMeo and others, 2003)

Heat flux plate

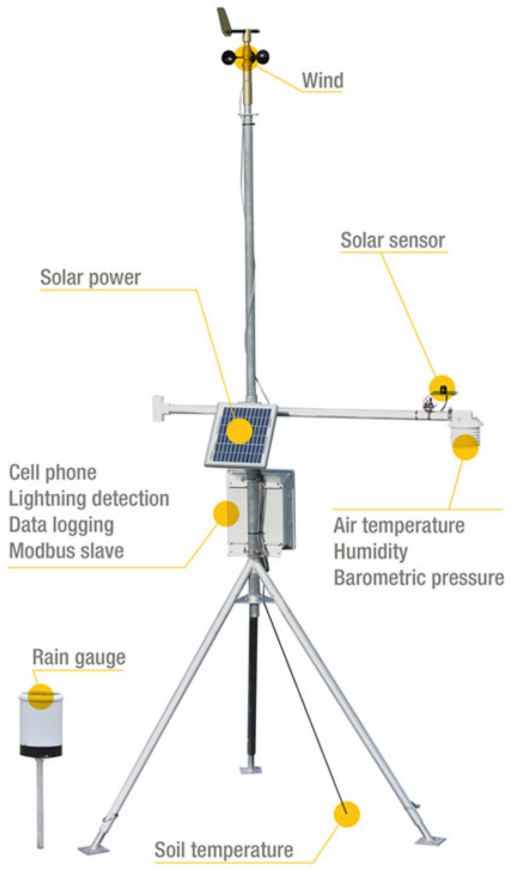
$$ET_{ref} = \frac{\Delta(R - G) + \rho C_p\left(\frac{VPD}{r_a}\right)}{\Delta + \gamma\left(1 + \frac{r_a}{r_s}\right)}$$

DATA DESCRIPTION

- Collected from a weather station in Corvallis, Oregon in Summer 2017
- 12 days of data averaged over 1 minute (17188 samples, 10 variables):
  - Solar radiation
  - Relative Humidity & Air Temperature
  - Wind Speed & Wind Direction
  - Soil temp at two depths
  - Soil water content, electrical conductivity, matric potential

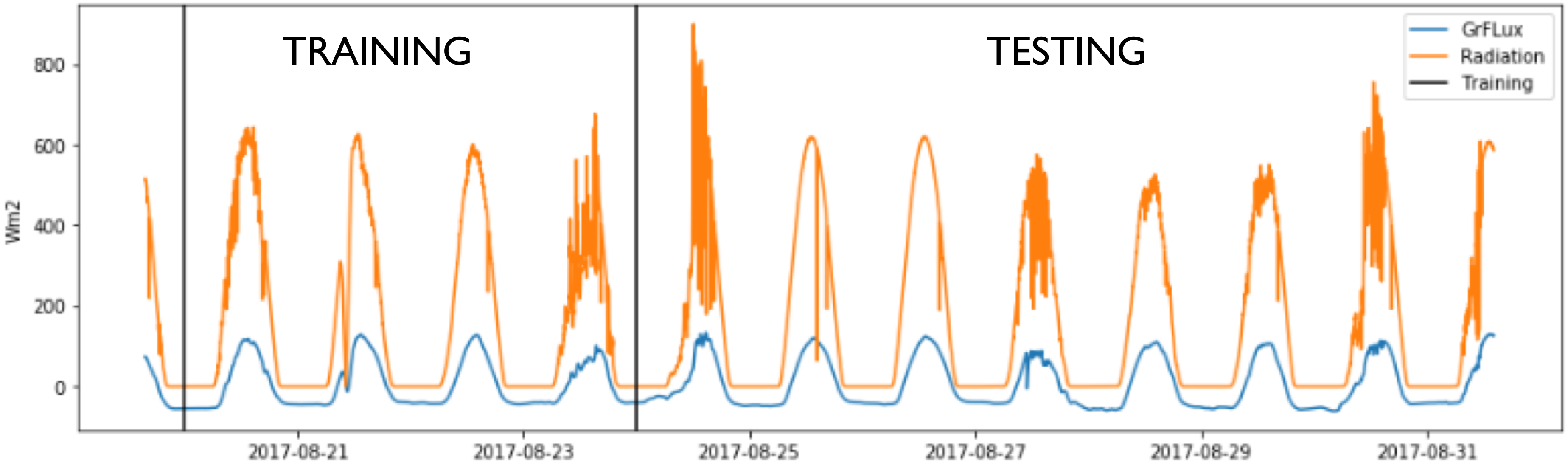
X (Variables) Y (Target)

TIMESTAMP	ApogeeWm2_Avg	HMP3m_RH_Avg	HMP3m_T_Avg	WDir	WSpd	Potential	Temp	VWC	Temp.1	EC	GrFlux_Avg
2017-08-19 15:38:00	514.2	23.59	27.10	124.8	5.751	-1510.5	29.7	0.128	29.7	0.009	73.82
2017-08-19 15:39:00	513.1	24.12	26.84	109.0	5.108	-1532.2	29.7	0.129	29.7	0.010	73.74
2017-08-19 15:40:00	514.1	24.43	26.83	109.9	6.627	-1518.9	29.7	0.128	29.7	0.009	73.47
2017-08-19 15:41:00	514.0	23.62	26.90	127.8	5.766	-1487.1	29.7	0.127	29.8	0.010	73.12
2017-08-19 15:42:00	512.4	23.58	26.95	121.6	4.851	-1495.3	29.7	0.127	29.8	0.008	72.78



SUPERVISED REGRESSION – TRAIN/TEST SPLIT

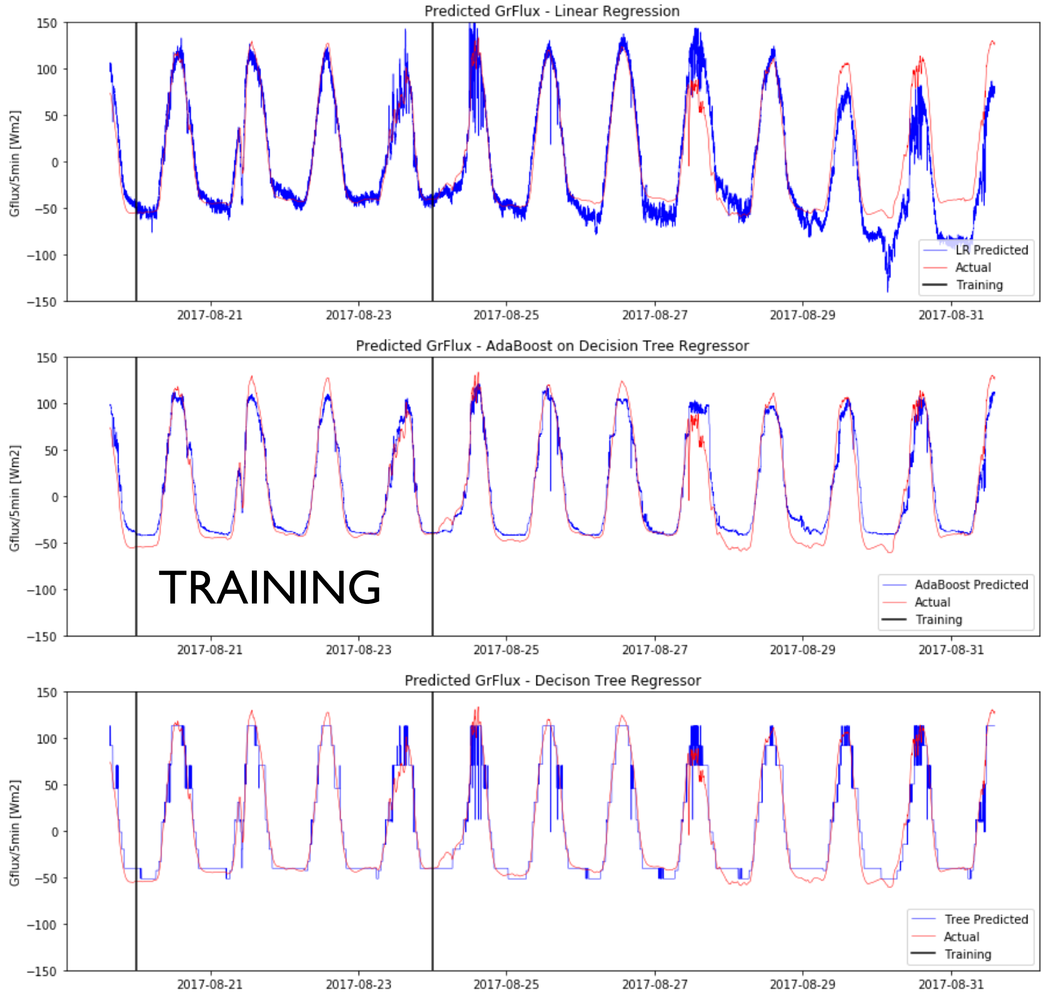
- Training Data: 4 days (8/20/18 – 8/24/18)
- Testing Data: 12 days (8/24/18 – 8/31/18)



SUPERVISED REGRESSION – GROUND FLUX PREDICTIONS

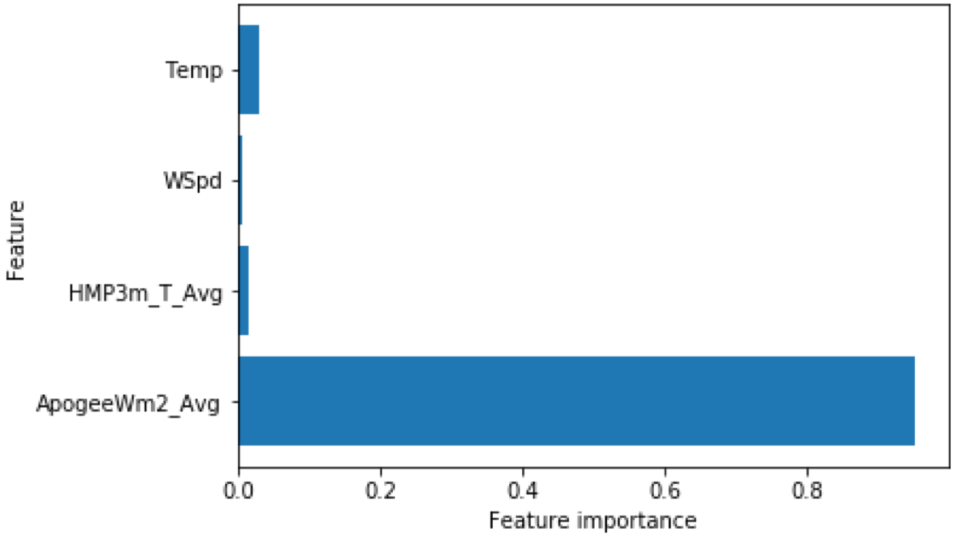
- Decision Tree & RF Regressor had the highest testing accuracy, $R^2 = 0.952$
- Most models perform poorly on cloudy days and at night

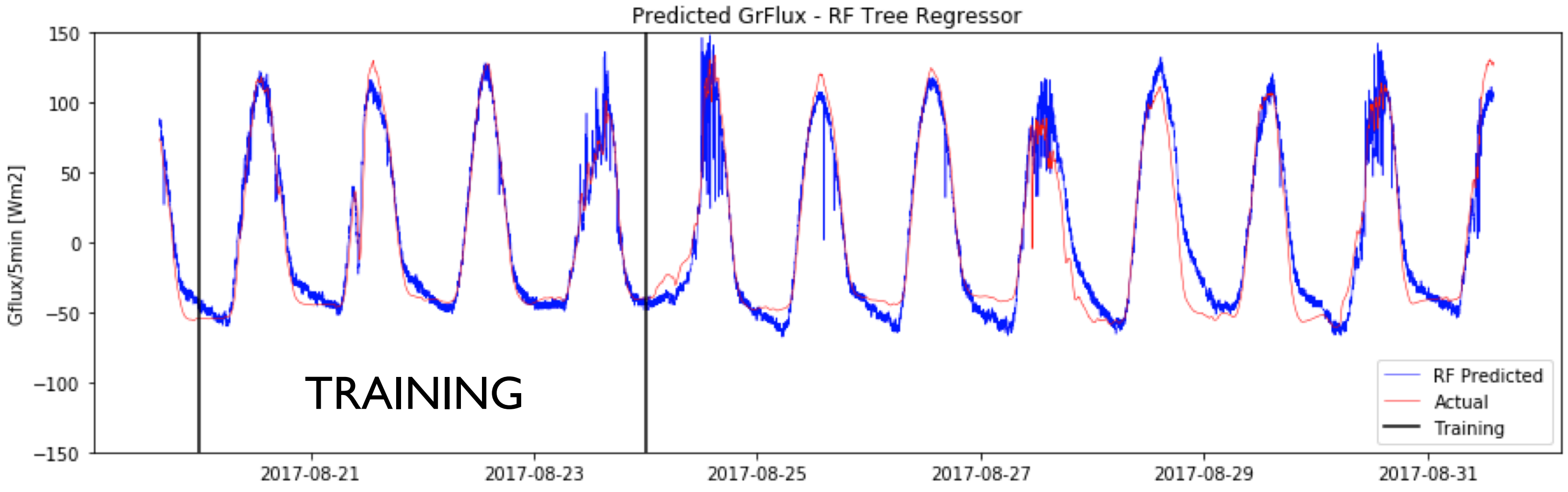
Model	Model Parameters	Train	Test
Linear Regression	-	0.957	0.845
Ridge Regression	Alpha = 10	0.952	0.945
Lasso Regression	Alpha = 1.0	0.949	0.950
Decision Tree Regressor	Depth = 4	0.962	0.952
AdaBoost	Decision Tree	0.973	0.942
Random Forest Regressor	Depth = 4	0.962	0.952



SUPERVISED REGRESSION – GROUND FLUX PREDICTIONS

- Most important feature in predicting ground flux is radiation (ApogeeWm2)
- Failure at night most likely due to no informative radiation data

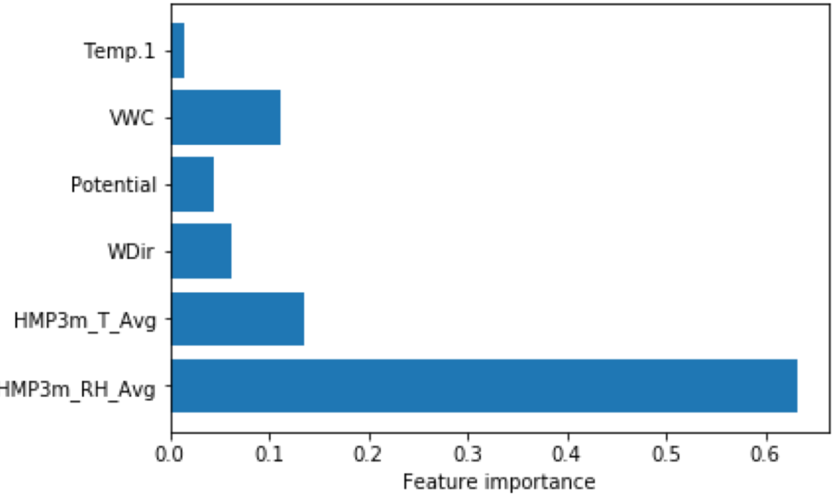


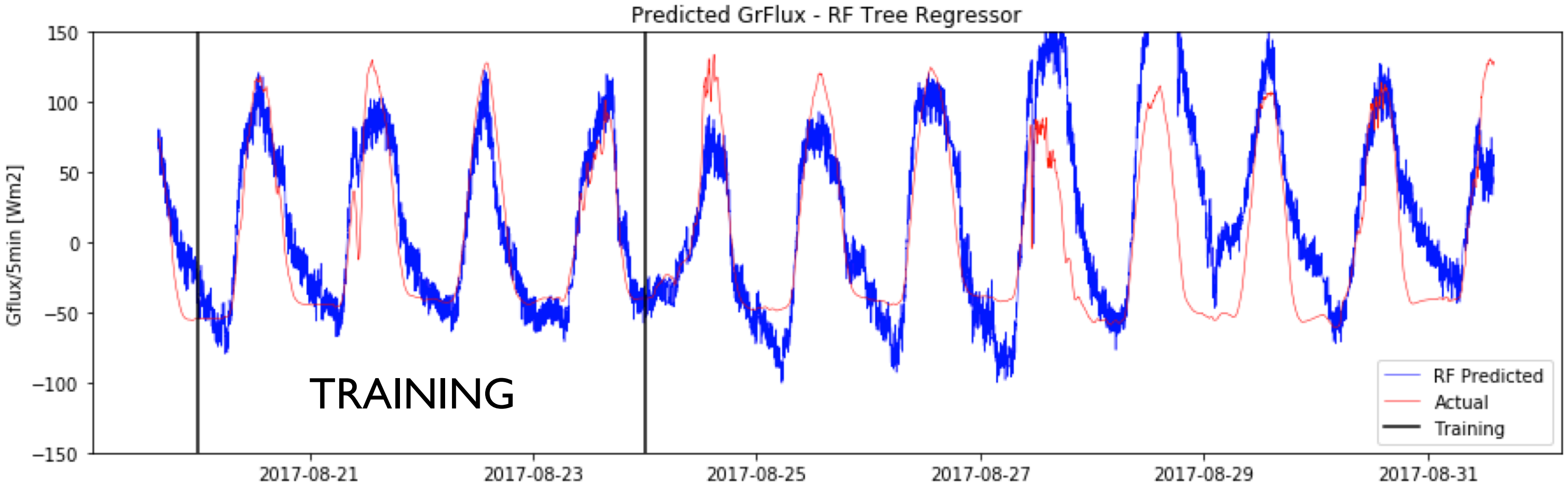


SUPERVISED REGRESSION – GROUND FLUX PREDICTIONS W/O RADIATION DATA

- Model accuracies significantly decline when radiation data is removed from predictors
- Relative humidity and soil data drive ground flux predictions when radiation data not available

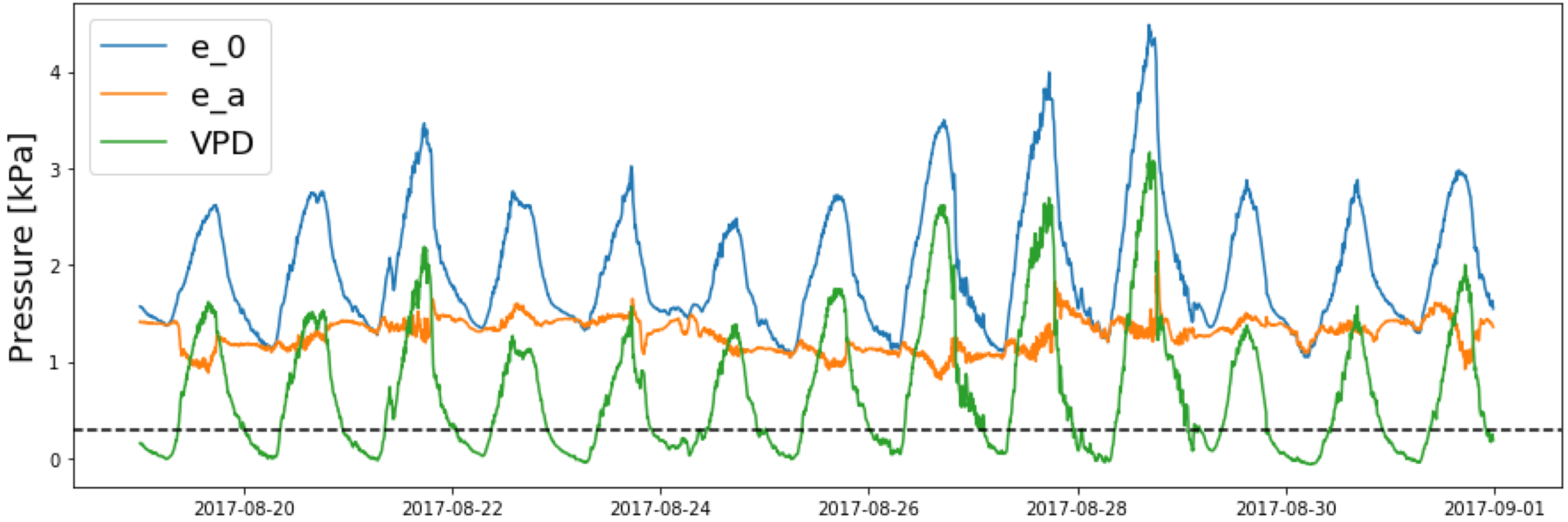
	Train	Test
Lin Reg	0.899928	0.219161
Ridge	0.853141	0.533537
Lasso	0.774098	0.354022
Tree	0.891886	0.613077
AdaBoost	0.954872	0.697086
RF	0.891886	0.613077





UNSUPERVISED CLUSTERING – PREDICTING UNCERTAINTY

- Predicting when Vapor Pressure Deficit (VPD) is within the uncertainty bounds of humidity sensor from meteorological data
- When (VPD) is less than the resolution of the humidity sensor, the data is flagged
  - VPD < 0.3%, VPDError = 1
  - VPD > 0.3%, VPDError = 0



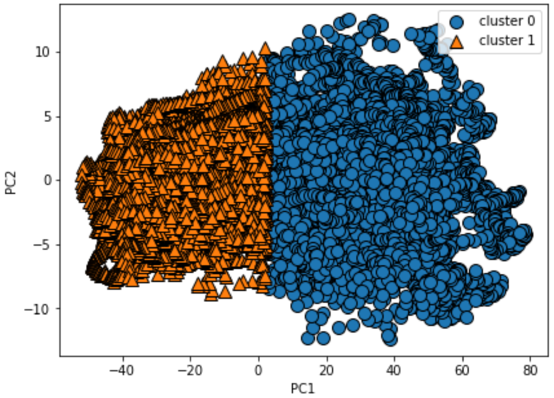
$$ET_{ref} = \frac{\Delta(R-G) + \rho C_p \left(\frac{VPD}{r_a}\right)}{\Delta + \gamma\left(1 + \frac{r_a}{r_s}\right)}$$

UNSUPERVISED CLUSTERING – PREDICTING UNCERTAINTY

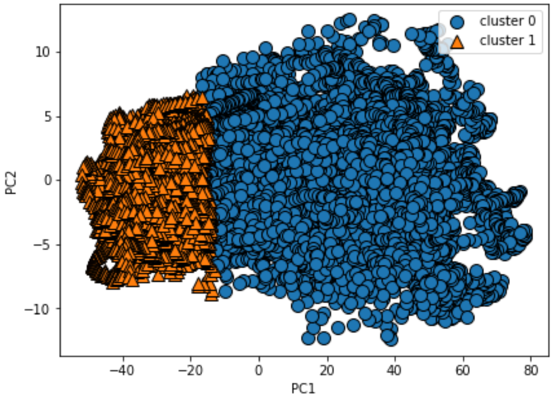
X (Variables)

TIMESTAMP	HMP2m_T_Avg	HMP5m_T_Avg	HMP2m_RH_Avg	HMP5m_RH_Avg	WSpd_Avg	P	VPGError	ETref
2017-06-12 00:05:00	12.20	12.36	74.68	71.78	1.450	100.81	0	-0.015278
2017-06-12 00:10:00	12.19	12.35	75.29	72.56	1.325	100.81	0	-0.012733
2017-06-12 00:15:00	12.11	12.31	75.88	72.71	0.825	100.81	0	-0.009972
2017-06-12 00:20:00	12.27	12.45	74.86	71.57	3.450	100.81	0	-0.017495
2017-06-12 00:25:00	12.17	12.36	75.98	72.57	3.200	100.81	0	-0.023142

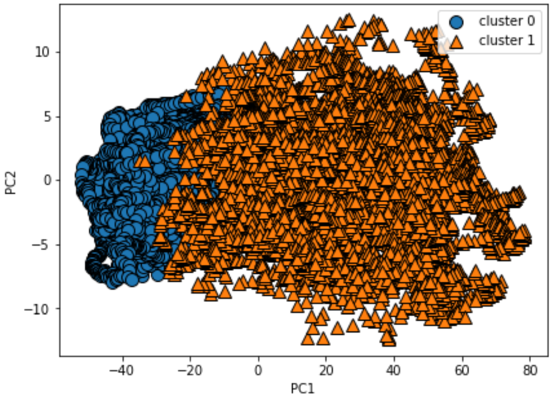
Kmeans, n = 2



HC, n = 2



Model Based, n = 2, Covariance = Full



UNSUPERVISED CLUSTERING – RESULTS

- All clustering models tend to over predict error, high false positive rate
- K-means clustering has highest mutual info score

```
print('Kmeans {:}'.format(normalized_mutual_info_score(data4.VPDError, data4.KM)))
print('HC {:}'.format(normalized_mutual_info_score(data4.VPDError, data4.HC)))
print('GMM {:}'.format(normalized_mutual_info_score(data4.VPDError, data4.GM)))
```

```
Kmeans 0.4714450912005556
HC 0.3662264925635367
GMM 0.3160661025089997
```

```
pd.crosstab(data4.VPDError,data4.HC,margins=True)
```

HC	0	1	All
VPDError			
0	6102	9572	15674
1	8911	0	8911
All	15013	9572	24585

```
pd.crosstab(data4.VPDError,data4.KM,margins=True)
```

KM	0	1	All
VPDError			
0	4273	11401	15674
1	8911	0	8911
All	13184	11401	24585

ASSOCIATION

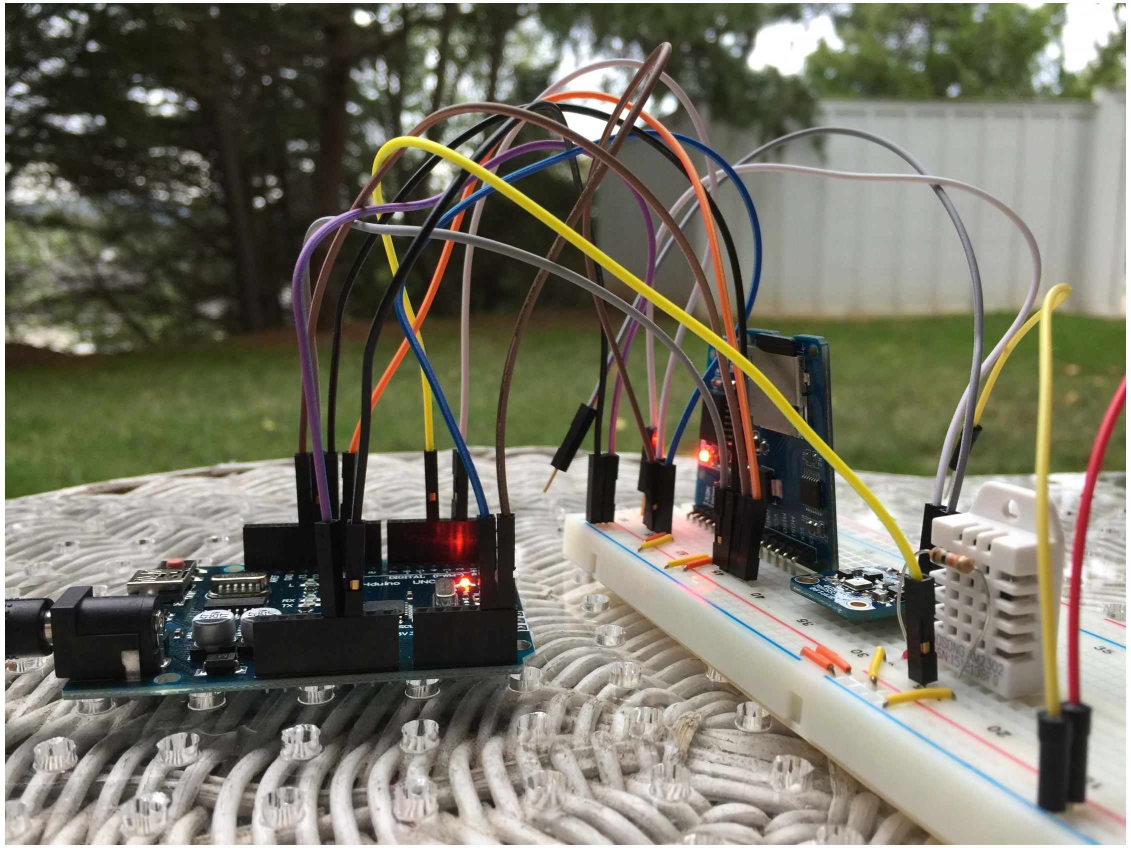
- Created categorical variables based on time of day (night, sunrise, day, sunset)
- Given it's night time, there is 24% chance that the VPD error will occur

TIMESTAMP	VPDError	night	sunrise	day	sunset
2017-06-12 00:05:00	0	1	0	0	0
2017-06-12 00:10:00	0	1	0	0	0
2017-06-12 00:15:00	0	1	0	0	0
2017-06-12 00:20:00	0	1	0	0	0
2017-06-12 00:25:00	0	1	0	0	0

	support	itemsets
0	0.362457	(VPDError)
1	0.334798	(night)
2	0.456376	(day)
3	0.240472	(night, VPDError)

FUTURE WORK

- Use machine learning algorithms to train multiple micro Arduino based weather stations
- Used to understand micro-climates within orchards/vineyards as a tool for assessing plant productivity and irrigation management



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.