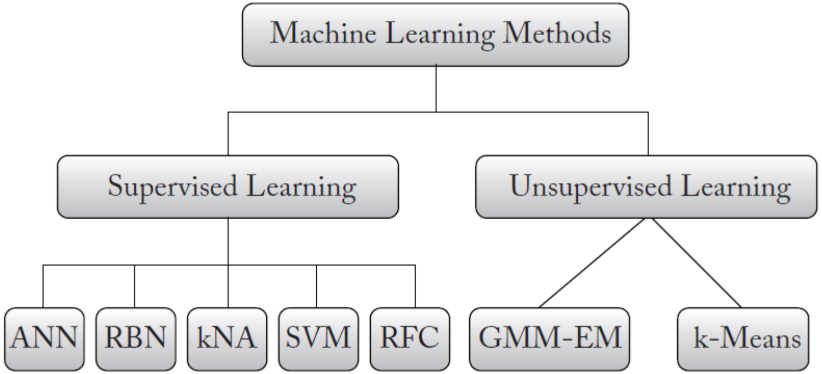
Surface Albedo Predictions Using Random Forests

Srinidhi Budhiraju, ASU CIDSE
Sameeksha Katoch, ECEE
Dr. Andreas Spanias, ECEE
Dr. Yiannis Tofis, Univ. of Cyprus KIOS

This work is partly sponsored by NSF REU Grant 1854273 and ASU SenSIP Center

Background Training in Machine Learning

- Learn about surface albedo and factors that impact it.
- Random forests vs neural networks.
- Python review.
- Google Colaboratory.



Motivation

- To elevate PV efficiency and robustness
- To reduce cost of maintenance
- To develop methods for predicting shading in PV Arrays

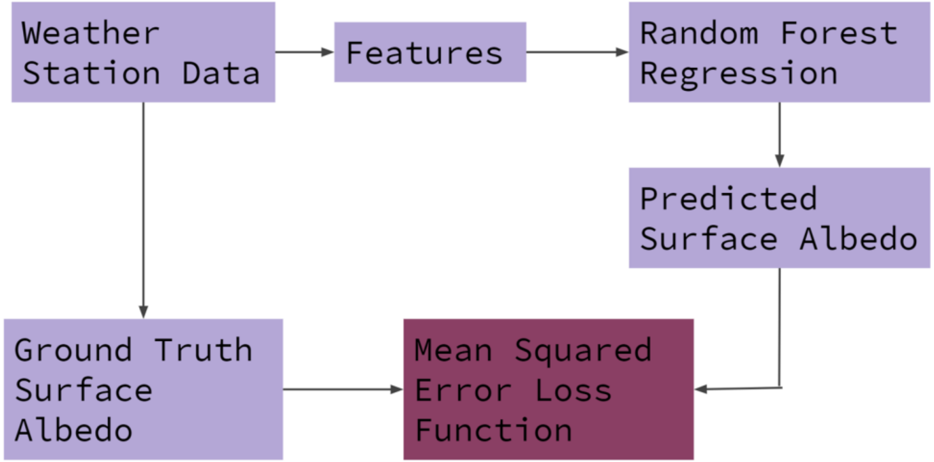
Introduction

- Climate change concerns increasing.

- PV systems are environmentally friendly.

- Problems
  - Fluctuations in power output
  - Power output -> irradiance -> surface albedo

- Challenges
  - Lot of data => Need to find way to make accurate predictions while also reducing training time.
  - Understanding conceptually why some techniques aren’t suitable for some ML algorithms



NSRDB Dataset

- Data collected at 30 minute intervals for one year.
- Data features include
  - **DHI**: The amount of solar radiance that is not received directly by the sun per unit area of surface.
  - **DNI**: The amount of radiation received directly by the sun per unit area of surface.
  - **GHI**: The total amount of radiation received per unit area of surface. It's the sum of DHI and DNI.
  - **Cloud Type**
  - **Dew Point**: The temperature that water must be cooled to in order to become saturated with water vapor.
  - **Solar Zenith Angle**: Sun rays angle from the vertical.
  - **Surface Albedo**: The fraction of sunlight reflected by a surface.
  - **Wind Speed**
  - **Precipitable Water**: Amount of water vapor in a column of the atmosphere.
  - **Wind Direction**
  - **Relative Humidity**: Percentage of water of vapor in the air.
  - **Temperature**
  - **Pressure**

Preprocessing NSRDB Dataset

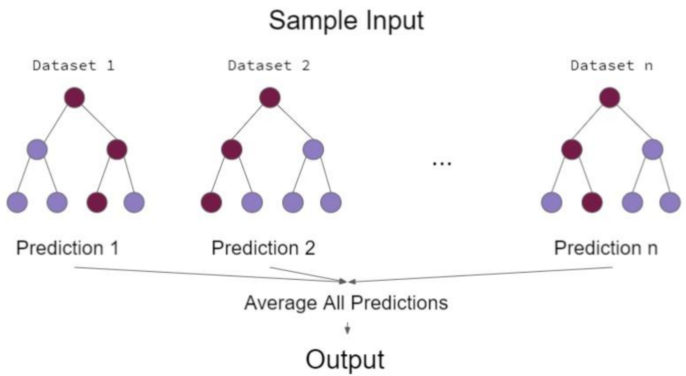
	Source	Location ID	City	State	Country	Latitude	Longitude	Time Zone	Elevation	Local Time Zone	Clearsky DHI Units	Clearsky DNI Units	Clearsky GHI Units	Dew Point Units	DHI Units	DNI Units	GHI Units
0	NSRDB	77853	-	-	-	33.41	-111.9	-7	366	-7	w/m2	w/m2	w/m2	c	w/m2	w/m2	w/m2
1	Year	Month	Day	Hour	Minute	DHI	DNI	GHI	Clearsky DHI	Clearsky DNI	Clearsky GHI	Cloud Type	Dew Point	Solar Zenith Angle	Fill Flag	Surface Albedo	Wind Speed
2	2015	1	1	0	0	0	0	0	0	0	0	4	0	167.58	0	0.158	1.0
3	2015	1	1	0	30	0	0	0	0	0	0	4	0	169.6	0	0.158	0.9
4	2015	1	1	1	0	0	0	0	0	0	0	4	0	167.8	0	0.158	0.8

	DHI	DNI	GHI	Clearsky DHI	Clearsky DNI	Clearsky GHI	Cloud Type	Dew Point	Solar Zenith Angle	Fill Flag	Surface Albedo	Wind Speed	Precipitable Water	Wind Direction	Relative Humidity	Temperature	Pressure
0	-1.0	0	-1.0	0	0	0	4	0	167.58	0	-0.888889	1.0	0.868	254.9	99.22	0	960.0
1	-1.0	0	-1.0	0	0	0	4	0	169.60	0	-0.888889	0.9	0.857	254.9	99.27	0	960.0
2	-1.0	0	-1.0	0	0	0	4	0	167.80	0	-0.888889	0.8	0.847	230.1	95.09	0	960.0
3	-1.0	0	-1.0	0	0	0	4	0	163.38	0	-0.888889	0.7	0.854	230.1	95.12	0	960.0
4	-1.0	0	-1.0	0	0	0	4	0	157.87	0	-0.888889	0.6	0.861	209.3	93.39	0	960.0

- Remove unnecessary columns from dataset.
- One-hot encoding on cloud type feature.
- Feature scaling (normalizing the data).
- Split into train/test dataset.

Random Forest Regressor

- Procedure for training dataset
  - For N trees
    - Pick k data points from training set
    - Build a decision tree for those k data points, where certain feature will be evaluated at each node
  - For new data point, have each tree make a prediction, average all the predictions, and assign average to new data point

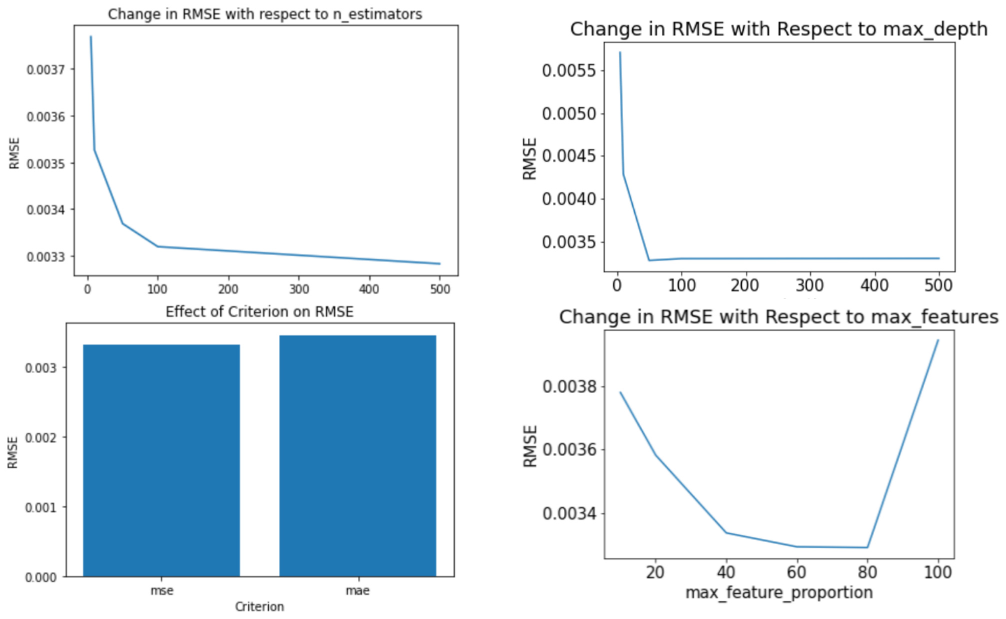


- Parameters
  - n_estimators, criterion, max_depth, max_features

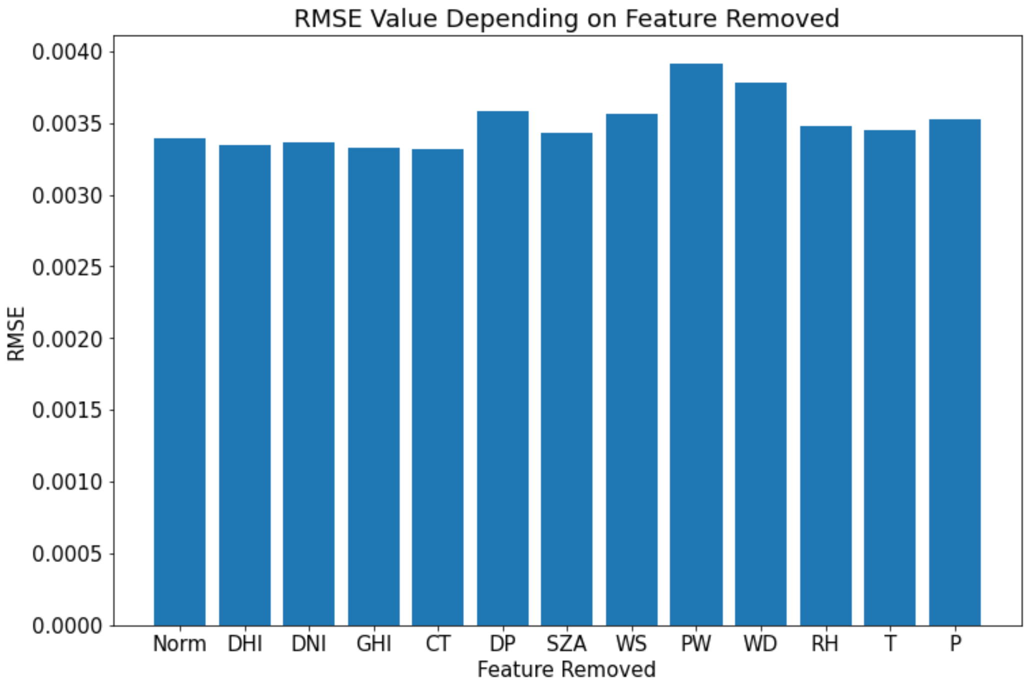
Hyperparameter Tuning

- Hyperparameter tuning for the number of trees in the random forest algorithm, the function used for the quality of the split, the maximum depth of the tree, and the maximum number of features.

- Create RMSE plot for each of these



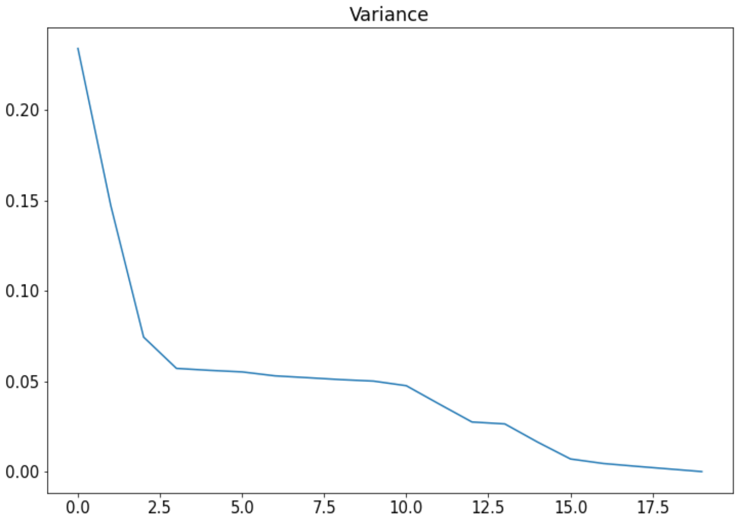
Feature Ranking



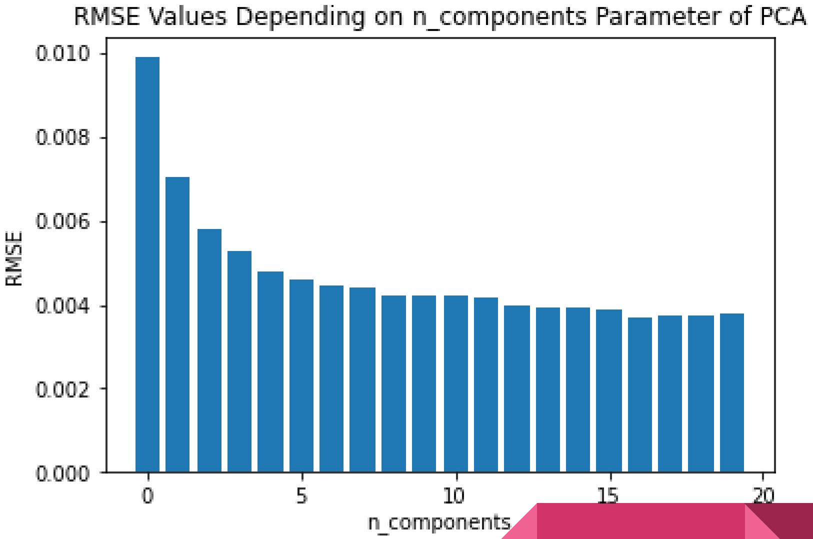
- Manual feature importance by removing individual features
- Use the one provided by sklearn
  - Did not match well with the manual feature importance graph

Principal Component Analysis

Variance Due to Each Component



RMSE vs Number of Components



Conclusion/Future Work

Conclusion

- 4 most important features for surface albedo prediction
  - Precipitable Water
  - Wind Direction
  - Dew Point
  - Wind Speed

- Lowest RMSE achieved: 0.00370587

Future Work

- How the features relate to each other

- PCA vs Autoencoder

References

1. S. Rao, S. Katoch, V. Narayanaswamy, G. Muniraju, C. Tepedelenlioglu, A. Spanias, P. Turaga, R. Ayyanar, and D. Srinivasan, “Machine Learning for Solar Array Monitoring, Optimization, and Control,” Synthesis Lectures on Power Electronics, vol. 7, no. 1, pp. 1–91, 2020.

1. B. Marion, “Albedo Data Sets for Bifacial PV Systems,” 2020 47th IEEE Photovoltaic Specialists Conference (PVSC), 2020.

1. C. T. Clack, “Modeling Solar Irradiance and Solar PV Power Output to Create a Resource Assessment Using Linear Multiple Multivariate Regression,” Journal of Applied Meteorology and Climatology, vol. 56, no. 1, pp. 109–125, 2017.

1. Y. Kotak, M.S. Gul, T. Muneer, and S.M. Ivanova, ``Investigating the Impact of Ground Albedo on the Performance of PV Systems," CIBSE, Technical Symposium. London, UK 16-17, April 2015.

1. M. Blondel, M. Brucher, L. Buitinck, D. Cournapeau, N. Dawe, S. Du, V. Dubourg, E. Duchesnay, A. Fabisch, V. Fritsch, S. Ghosh, A. S. Gollonet, C. Gorgolewski, J. Grobler, B. Holt, A. Joly, T. R. Jones, K. Kastner, M. Kumar, R. Layton, W. Li, P. Losi, G. Louppe, V. Michel, J. Millman, A. Passos, F. Pedregosa, P. Prettenhofer, V. Rajagopalan, J. Schreiber, J. Vanderplas, D. Warde-Farley, and R. Weiss, “Scikit-learn: Machine Learning in {P}ython,” Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.