Weather-driven predictions of solar energy

Adele Kuzmiakova, Gael Golas, Alex McKeehan | Stanford University
adele.kuzmiakova@stanford.edu, colasg@stanford.edu, mckeehan@stanford.edu

1 Introduction

Climate change and energy crisis have motivated the use and development of sustainable solar energy sources. Yet, **renewable energy sources are intermittent and weather-dependent**, and therefore, it is necessary to improve their predictions using hybrid ML algorithms that are weather-driven.

Questions:

1) Which algorithms are the most efficient for weather-derived predictions ?
2) What weather features are the most influential in the ML algorithms?

2 Methods

Raw data acquisition

- Weather inputs

The NOAA website [1-2] provides hourly weather measures from weather stations all over the US. To get the weather features at our location, we spatially averaged the measures over the 3 closest weather stations.

- Solar energy output

It was obtained from publicly-available repository of the University of Illinois campus.

Weather features	Unit	Weather features	Unit
Cloud coverage	% range	Relative humidity	%
Visibility	Miles	Wind speed	Mph
Temperature	°C	Station pressure	inchHg
Dew point	°C	Altimeter	inchHg

Data set

- Daily resolution: 6am – 5pm in 02/01/2016 - 09/31/2017
- Data set split : 80%/10%/10%
- **Training set : 6028** examples
- **Development and test set : 754** examples each

3 Results

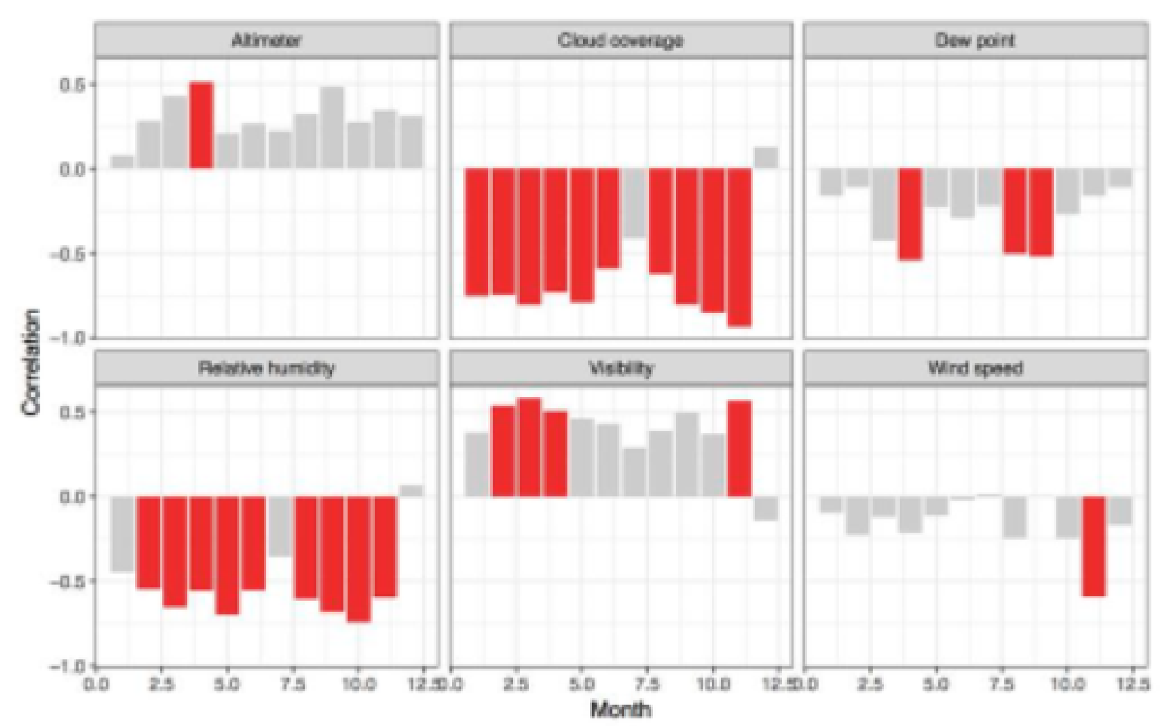

Figure 1: Time-based correlations between individual meteorological parameters and actual solar output. Red bars denote values where |r| > 0.5, corresponding to strong correlations.

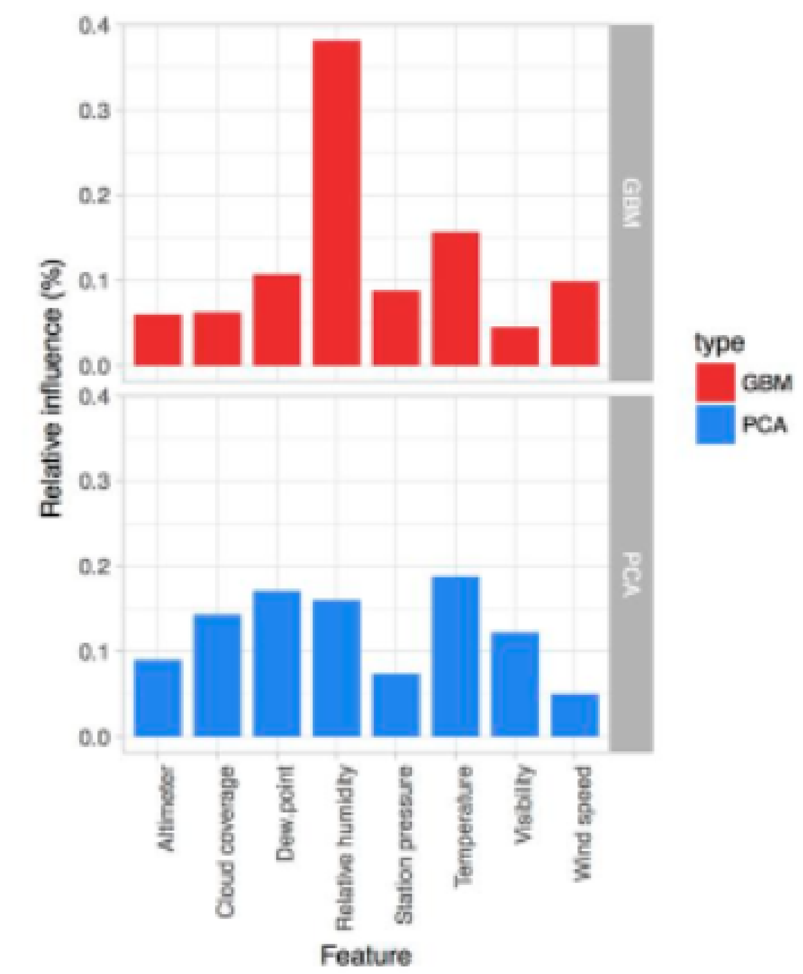

Figure 2: Relative influence of meteorological features in the Generalized Boosting Model (GBM; red) and weighted linear regression with PCA (PCA; blue).

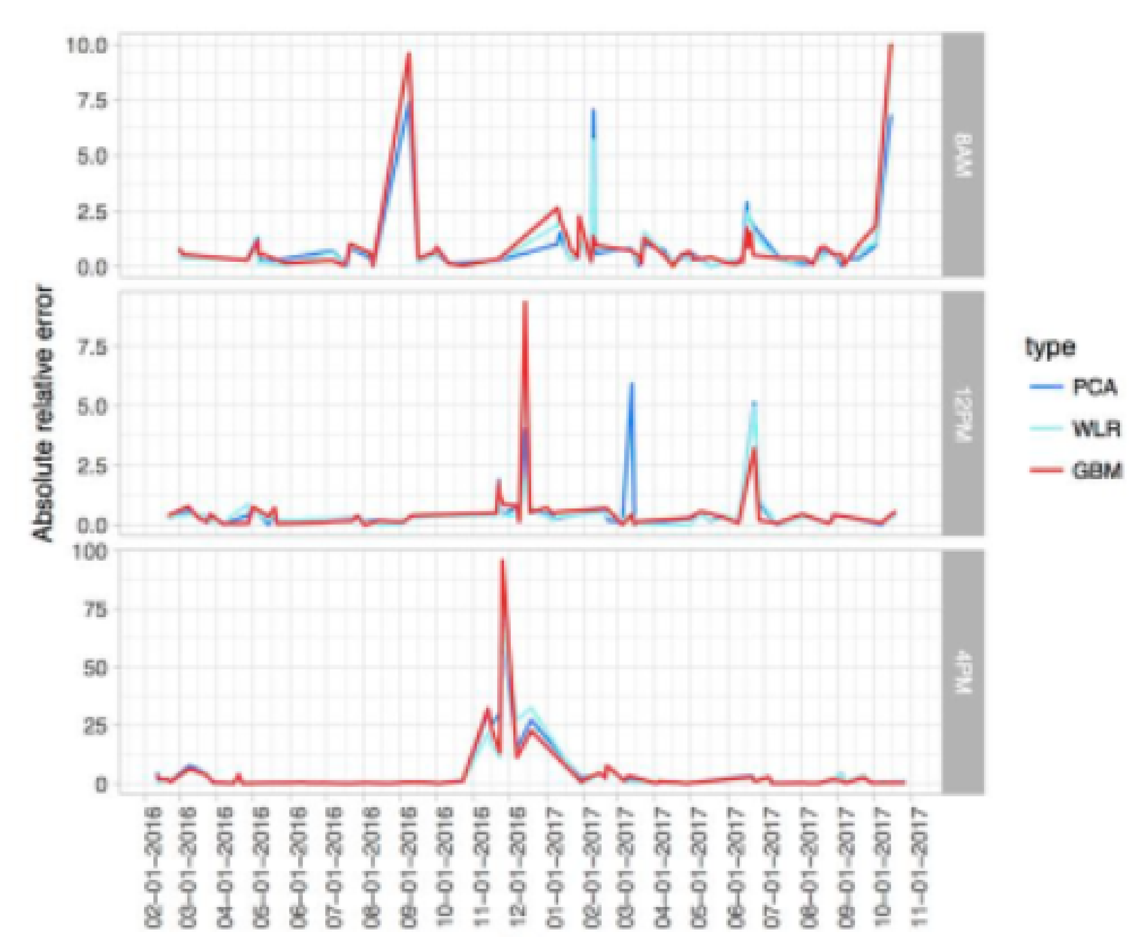

Figure 3: Absolute relative error from PCA-based regression (PCA; blue), ordinary weighted linear regression (WLR; cyan), and generalized boosting random forest model (GBM; red). Vertical panels correspond to errors at 8AM (morning), 12PM (noon), and 4PM (afternoon).

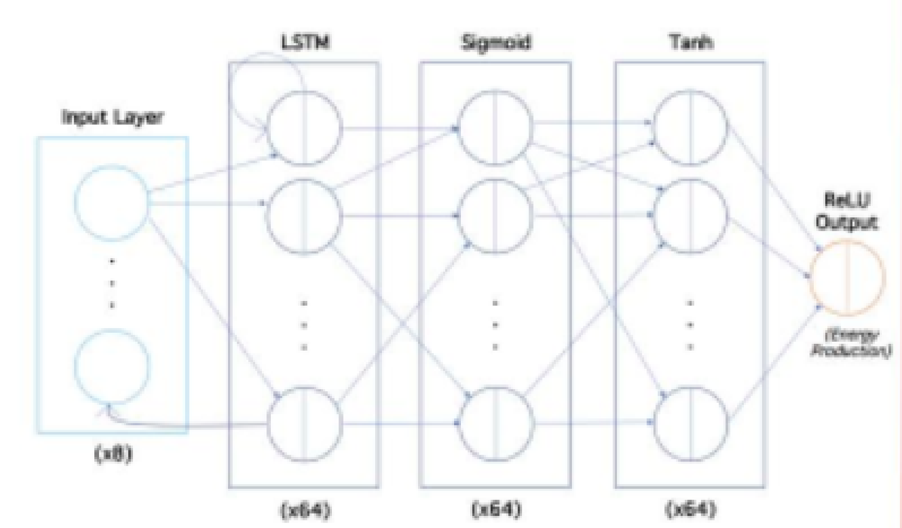

Figure 4: Graphic depicting flow of hidden layers in optimized neural network. Including an LSTM layer vastly improved performance, while the nonlinear tanh and sigmoid layers greatly reduced error levels in comparison to standard linear activation functions [3].

Model	No. of param	Training error	Test error
PCA-based weighted regression	5	0.866*10^6	1.077*10^6
Ordinary weighted linear regression	8	0.802*10^6	1.011*10^6
Generalized boosting method (random forest)	3	0.856*10^6	1.240*10^6
Neural networks	*	1.160*10^6	1.183*10^6

* Parameters optimized: batch size, no.of epochs, no. of hidden layers, type of hidden layers, no. of neurons, dropout rates, optimization type, and learning rate

4 Conclusions

- Most influential features for predicting solar energy: **relative humidity and temperature**
- Weighted linear regression (with or without PCA) and Random Forest are too simple of a model to predict the complexity of weather-driven phenomena.
- Most error contribution comes from winter months (Nov - Jan) where the actual solar output is very low.
- When applying the WLR model on the "non-winter" datase: our median percentage test error decreased from 41% to 35%.

5 References

[1] http://www.ncdc.noaa.gov/
[2] https://github.com/sborgenson/local-weather
[3] Oeda S, Kurimoto I and Ichimura T 2006 Time series data classification using recurrent neural network with ensemble learning *Lecture Notes Comput. Sci.* 4253 742-8