International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161

Available at http://inpressco.com/category/ijcet

Research Article

Solar Energy Prediction using Least Square Linear Regression Method

Suruchi Dedgaonkar*, Vishal Patil, Niraj Rathod, Gajanan Hakare & Jyotiba Bhosale

Vishwakarma Institute of Information Technology, Savitribai Phule Pune University, Pune, Maharashtra, India

Accepted 01 Sept 2016, Available online 02 Sept 2016, **Vol.6, No.5 (Oct 2016)**

Abstract

A challenge with renewable energy prediction is that their power generation is intermittent and uncontrollable. But, prediction of renewable energy is important, because of variation in weather parameters and demand of energy at each location. The solar energy is an infinitely available source of energy. The amount of solar radiation varies at every location depending on the weather factors like temperature, rainfall, humidity, wind speed, etc. While manually developing sophisticated prediction models may be feasible for large-scale solar farms, developing them for distributed generation at millions of homes is a challenging problem. To address the problem, in this paper, we creating prediction models for solar power generation from National Data Centre (NDC) weather forecasts data using machine learning techniques.

Keywords: *Least square linear regression, solar energy predictions, machine learning.*

1. Introduction

These days energy consumption is increased as technology is increasing leading to new applications or appliances. But sources are limited, so we need another energy source to fulfill a need of energy.

Solar radiation varies with variation in weather factors like temperature, dew points, humidity, wind speed, etc. Thus it is important to find out the exact factors affecting radiation at a particular location. The correlation between different parameters is found by applying regression techniques. Various machine learning algorithms are applied to the weather dataset. The performance of the algorithm is compared. Finally, the efficient algorithm is selected to make the decision about location. The input parameters used in the datasets of the system are solar radiation, temperature, wind speed, cloud, dew, humidity and month.

2. Data Analysis

We analyze extensive traces of historical data from a weather station to correlate the weather metrics present in the forecast with the solar intensity, in megajoule per square meter, recorded by the weather station. Our analysis quantifies how each forecast parameter affects each other and the solar intensity.

3. Model Generation

We apply multiple machine learning techniques to derive prediction model for solar intensity using multiple forecast metrics, and then analyze the prediction accuracy of each model. We use machine learning techniques on a training data set of historical solar intensity observations and forecast to derive a function that computes future solar intensity for a given time horizon from a set of forecasted weather metrics.

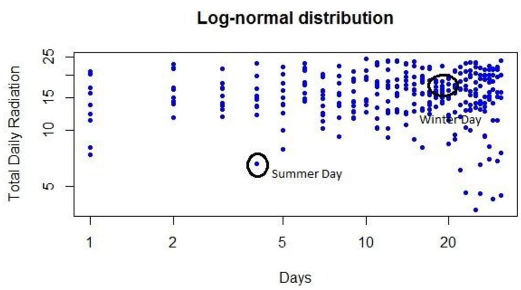
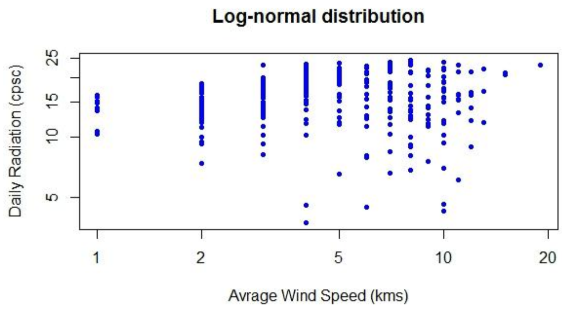


Fig. 1 Solar intensity shows seasonal variation with months of a year, although daily weather conditions also have a significant impact

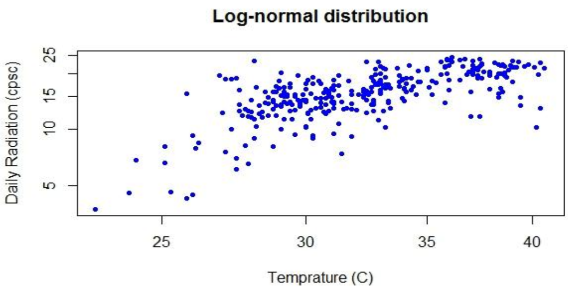
Fig. 1 shows how the day of the year affects solar intensity by charting the average solar intensity reading at noon per day over our 12 month monitoring period where day zero is January 1st, 2014. As expected, the graph shows that the solar intensity is lowest in January near the winter solstice and increases into the summer before decreasing after the vernal equinox. Additionally, the graph also implies that other conditions also have a significant

*Correspondinh author **Suruchi Dedgaonkar** is working as Professor and **Vishal Patil, Niraj Rathod, Gajanan Hakare & Jyotiba Bhosale** are students

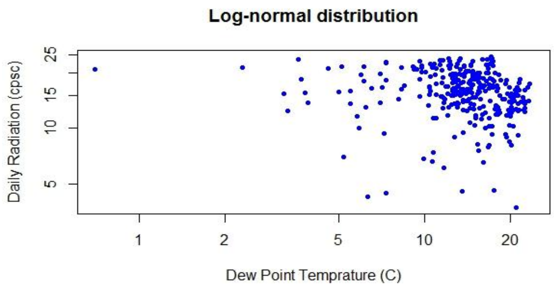
impact on solar intensity, since many days throughout the spring and summer have low solar intensity readings. The graph shows that solar intensity and the day of the year are roughly correlated most of the time, but not always, a summer day has higher solar intensity than a winter day. However, other factors must contribute to the solar intensity, since there are clearly some sunny winter days that record higher solar intensity readings than some cloudy summer days. To better understand correlations with other weather metrics, we model similar relationships for the other forecast metrics.



(a)



(b)

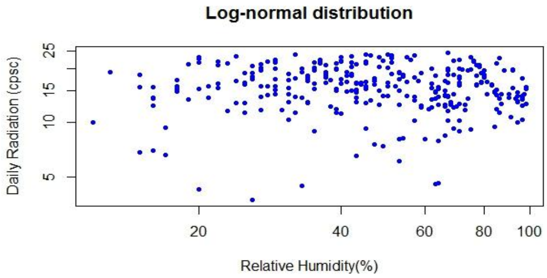


(c)

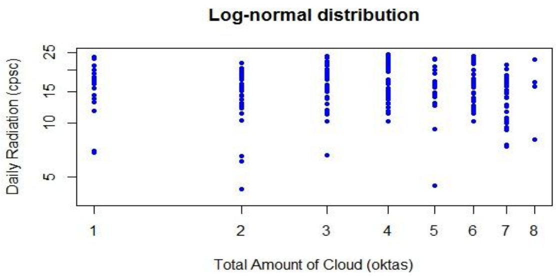
Fig 2 (a) Solar intensity and wind speed show little correlation (b) Solar intensity shows some correlation with temperature at high temperatures (c) with dew point at high dew points

For example, Fig. 2 shows that wind speed, dew point, and temperature are not highly correlated with solar intensity. Solar intensity varies almost uniformly from lower to higher values at any value of wind speed (a). Thus, wind speed has nearly zero correlation with solar intensity and its value is not indicative of the solar intensity or solar panel power generation.

Both temperature (b) and dew point (c) correlate with solar intensity at higher values. If the temperature or dew point is high, then the solar intensity is likely to be high. However, if the temperature or dew point is low, the solar intensity exhibits a more significant variation between high and low values. The results are intuitive. For example, in the summer a high temperature is often dependent on sunlight, while in the winter sunlight contributes less in raising the ambient temperature.



(a)



(b)

Fig. 3 Solar intensity generally decreases with increasing values of (a) relative humidity (b) total amount of cloud

In contrast, Fig. 3 shows that sky cover, relative humidity, and the amount of cloud have high negative correlations with solar intensity. In each case, as the value of the metric increases, the solar intensity reading generally decreases. However, as with the month of the year, there must be other factors that contribute to the solar intensity reading, since there are some days with a high sky cover, relative humidity, and precipitation probability, but a high solar intensity reading and vice versa. In addition to exhibiting complex relationships with solar intensity, each weather metric also exhibits a complex relationship with other weather metrics.

Table 1: Correlation Matrix table

	R	T
Radiation (R)	1	0.896
Temp. (T)	0.896	1
Humidity(H)	-0.865	-0.880
Month(M)	-0.436	-0.310
Wind Speed(Ws)	0.196	0.209
Amt. of Cloud(C)	-0.468	-0.283
Dew points(Dp)	-0.265	-0.183

H	M	Ws	C	Dp
-0.865	-0.436	0.1962	-0.468	-0.265
-0.88	-0.31	0.2091	-0.283	-0.183
1	0.277	-0.216	0.355	0.338
0.277	1	-0.008	0.471	0.317
-0.216	-0.008	1	0.651	0.707
0.355	0.471	0.651	1	0.91
0.3387	0.317	0.707	0.917	1

Table 1 shows correlation coefficients for each weather metric using the Pearson product-moment correlation coefficient, which divides the covariance of the two variables by the product of their standard deviations. The higher the absolute value of the correlation coefficient, the stronger the correlation between the two weather metrics a positive correlation indicates an increasing linear relationship, while a negative correlation indicates a decreasing linear relationship. The complex relationships between weather metrics and solar intensity shown in this table motivate our study of automated prediction models.

4. Prediction Model

We represent both observational and forecast weather metrics as a time-series that changes due to changing weather patterns and seasons. As the previous section shows, solar intensity depends on multiple weather metrics, which complicates the task of developing an accurate prediction model. The high dimensionality of the time-series data motivates our study of regression methods to develop solar intensity prediction models. The machine learning techniques automatically output a function that computes solar intensity from the 6 weather metrics, as well as the day of the year. We use the remaining 2 months of our data set to test the model's accuracy. One benefit of using machine learning to automatically generate prediction model that, in general, the more training data that is available, the more accurate the model.

The generated model is a simple function that computes solar intensity from multiple weather metrics including the day of the year.

SolarRadiation = F (Months, Temperature, DewPoint, WindSpeed, TotalAmountofCloud, Humidity)

F is the function that we determine using different regression methods. We preserve the units of each metric. We represent each month as a value between 1 and 12, temperature in degrees Celsius, wind speed in Miles per hour, Total amount of cloud in oktas, and humidity in percentage between 0% and 100%. However, before applying any regression techniques below we normalize all feature data to have zero mean and unit variance. To quantify the accuracy of each model, we use the Root Mean Squared Error (RMS-Error) between predicted solar intensity at any time and the actual solar intensity observed. RMS-Error is a well-known statistical measure of the accuracy of values predicted by a time-series model with respect to the observed values. An RMS-Error of zero indicates that the model exactly predicts solar intensity three hours in the future. The closer the RMS-Error is to zero the more accurate the model's predictions.

5. Least Square Regression Method

We first apply a linear least squares regression method to predict solar intensity. Linear least squares regression is a simple and commonly-used technique to estimate the relationship between a dependent or response variable, e.g., solar intensity, and a set of independent variables or predictors. The regression minimizes the sum of the squared differences between the observed solar intensity and the solar intensity predicted by a linear approximation of the forecast weather metrics. Applying the linear least squares method to the eight months of training data yields the prediction model below, with coefficients for each metric.

Solar Radiation = 3.77 * Temperature – 56.04 * Humidity + 8.77 * Wind Speed + 30.67 Cloud Index + 10.04 * Months + 16.38 * Dew Points.

5.1 Implementation

Least square regression is a method for finding a line that summarizes the relationship between the two variables, at least within the domain of the explanatory variable x.

$$Y = a + bx$$

Where

$$b = r\frac{SDy}{SDx}$$

$$a = \bar{Y} - b\bar{X}$$

Also,

Slope = $(N\Sigma XY - (\Sigma X)(\Sigma Y)) / (N\Sigma X^2 - (\Sigma X)^2)$
Where, b = the slope of the regression line
a = the intercept point of the regression line and the y-axis
$\bar{X}$ = Mean of x values.
$\bar{Y}$ = Mean of y values
SDx = Standard Deviation of x
SDy = Standard Deviation of y

To Find,
Least Square Regression Equation

Step 1:
Count the number of given x values.
N = 12

Step 2:
Find XY, X^2 for the given values.

Step 3:
Now, Find ΣX, ΣY, ΣXY, ΣX2 for the values
$\Sigma X = 78$
$\Sigma Y = 188.68$
$\Sigma XY = 650$
$\Sigma X^2 = 1168.14$

Step 4:
Substitute the values in the above slope formula given.
Slope (b) = $(N\Sigma XY - (\Sigma X)(\Sigma Y)) / (N\Sigma X^2 - (\Sigma X)^2)$
$= ((12)*(650)-(78)*(188.8))/((12)*(1168.14)-(78)^2)$
$= (7800 - 14726.4) / (14017.68 - 6084)$
$= -0.8730$

Step 5:
Now, again substitute in the above intercept formula given.
Intercept (a) = $(\Sigma Y - b(\Sigma X)) / N$
$= (188.68 - (-0.8730)(78))/12$
$= 21.3978$

Step 6:
Then substitute these values in regression equation formula

Regression Equation(y) = a + bx = 21.3978 - 0.8730x

Suppose if we want to calculate the approximate y value for the variable x = 13 then, we can substitute the value in the above equation
Regression Equation(y) = a + bx = 21.3978 - 0.8730(13) = 10.0488

Similarly, we calculate others and predicted this equation of Solar Radiation.

Equation:
Solar Radiation = 3.77 * Temperature – 56.04 * Humidity + 8.77 * Wind Speed + 30.67 Cloud Index + 10.04 * Months + 16.38 * Dew Points.

We verify the prediction accuracy using our test dataset for the remaining months of the year. We cross-validate the regression model with the training dataset (from Mar and Apr 2014) and verify its prediction accuracy using the testing dataset (May 2014). We observed that RMS error is 1.3085. The linear regression models work with accuracy 71% for 30 tested values with 61 trained values.

Conclusion

Prior prediction models for solar energy harvesting have been based primarily on the immediate past (A. Kansal, 2007) (D. Noh, 2009) (C. Moser, 2009) unfortunately, the methods are unable to predict changes in weather patterns in advance. Since weather forecasts from the NDC are based on aggregations of multiple data sources from across the country, they are able to provide advance warning. We show that the relationship between these forecast weather metrics and solar intensity is complex. Thus, we automatically derive linear regression prediction model which is 71% accurate from historical solar intensity and forecast data for radiation. The accuracy of the system will increase with more training.

Our results indicate that automatically generating accurate models that predict solar intensity, and hence energy harvesting of solar arrays, from weather forecasts is a promising area. We find that models derived using linear least squares outperform a past predicts future models and a simple model based on sky condition forecasts from prior work (N. Sharma, 2010) and is a promising area for increasing the accuracy of solar power generation prediction with more training, which is essential to increase the fraction of renewables in the grid. Moving forward, we plan on using our prediction models to better match renewable generation to consumption in both smart homes and data center's that utilize on-site solar arrays to generate power.

References

Sharma N., Gummeson J., Irwin D., and Shenoy P. (June 2010), Cloudy Computing: Leveraging Weather Forecasts in Energy Harvesting Sensor Systems, in SECON.

Kansal A., Hsu J., Zahedi S., and Srivastava M. (Sept. 2007), Power Management in Energy Harvesting Sensor Networks, in Transactions on Embedded Computing Systems.

Noh D., Wang L., Yang Y, Le H., and Abdelzaher T.(June 2009), Minimum Variance Energy Allocation for a Solar-powered Sensor System, in DCSS.

Moser C. (March 2009), Power Management in Energy Harvesting Embedded Systems, Ph.D. Thesis, ETH Zurich.

Sharma N., Gummeson J., Irwin, D. and Shenoy P., Predicting Solar Generation From Whether Forecasts Using Machine Learning, University of Massachusetts Amherst 01003