A MACHINE LEARNING APPROACH TO PREDICT SOLAR RADIATION FOR SOLAR ENERGY BASED DEVICES

U.Divya
Computer Science and Engineering
Prathyusha Institute of Technology and Management
Chennai, India

Chitra Pasupathi
Computer Science and Engineering
Prathyusha Institute of Technology and Management
Chennai, India

Abstract: Solar energy is used in many applications, such as increasing water's temperature or moving electrons in a photovoltaic cell, agriculture planning, fuel production, electricity production, transport, architecture and urban planning, etc. Solar energy is secure, clean, and available on the Earth throughout the year. Its secure and clean applications are very important to the world, especially at a time of fossil fuel high costs and the critical situation of the atmosphere resulting from fossil fuel applications. Solar techniques include the use of photovoltaic systems, concentrated solar power and solar water heating to harness the energy. In this paper, prediction is focusing in the Southern part of India and the solar light will be available from 8 to 9 months in a year in this region. So to utilize the solar energy in an efficient way the prediction is done. To predict the availability of solar energy the machine learning Temporal Gaussian Process Regression(TGPR) method has been used. It provides better result and also more robust when compared with the methods like ELM, SVM, etc. The predicted values are used to measure and analyze the amount of energy that could be generated during a year in the southern region of India. This in turn can be utilized to identify the suitable solar based devices suitable for different locations.

Keywords: Solar energy, solar devices, Prediction, Temporal Gaussian Process Regression.

I. INTRODUCTION

Solar radiation data provides information on how much is the Sun potential at a location on the Earth during a specific time period. These data are very important for designing solar energy systems. Solar energy is an important source of renewable energy which is currently under expansion in many countries. Solar energy based systems estimate the solar energy production and this estimation requires the prediction of available solar energy[1]. Due to the high cost and installation difficulties in measuring solar energy, these data are not always available. Therefore, there is a demand to develop alternative ways of predicting these data.

The solar parameters collected from the meteorological centre are used in solar radiation prediction[3]. The parameters used in the prediction process are maximum air temperature, relative humidity, minimum air temperature, atmospheric air pressure, wind speed, wind direction, etc. Temporal Gaussian process regression method is used to predict the solar radiation data and it is proved to provide better results and also more accurate than the existing machine learning techniques like Extreme learning machine(ELM), Support Vector Regression(SVR), Artificial Neural Networks(ANN)[2].

Gaussian process regression method is used to interpolate the values and it is modeled by a Gaussian process governed by prior covariance. In this method, new data points can be constructed within the range of a discrete set of known data points. Under suitable assumptions on the priors, Gaussian process regression gives the best linear unbiased prediction of the values. Gaussian regression process can handle the limitation in the availability of training data[4]. In addition to the good numerical performance and stability, it can adopt very flexible kernel functions, and it provides confidence intervals for the predictions.

II. RELATED WORKS

S. Salcedo-Sanz, et al[5] discussed the performance of a novel Coral Reefs Optimization – Extreme Learning Machine (CRO–ELM) algorithm in a real problem of global solar radiation prediction. The work considered different meteorological data from the radiometric station at Murcia(southern Spain), both from measurements, radiosondes and meteorological models, and fully describes the hybrid CRO–ELM to solve the prediction of the daily global solar radiation from these data. The algorithm was designed in such a way that the

ELM solved the prediction problem, whereas the CRO evolved the weights of the neural network, in order to improve the solutions obtained.

J.L.Chen, et al[6] have presented the methods of monthly mean daily solar radiation estimation using support vector machines (SVMs) to examine the feasibility of SVMs in estimating monthly solar radiation using air temperatures.

Rasheed H. AL-Naimi, et al[7] developed an artificial neural network (ANN) model for estimating monthly mean daily global solar radiation of Baghdad city, Iraq. The results of the ANN models have been compared with measured data on the basis of root mean square error (RMSE), mean absolute error (MAE) and determination coefficient (R2). It is found that the solar radiation estimations by ANN are in good agreement with the measured values .Results obtained indicate that the ANN model can successfully be used for the estimation of monthly mean daily global solar radiation for Baghdad city.

M. Hassan, A. Bermak[8] have developed the model by integrating wireless sensing nodes with ambient energy harvesting capability to overcome limited battery power budget constraint and extending effective operational time of sensor network. An efficient algorithm for solar energy prediction based on additive decomposition (SEPAD) model was used. In this model, both seasonal and daily trends along with Sun's diurnal cycle are individually considered.

Jose Luis Guinon, et al[9] presented the use of Mathcad software for the implementation and analysis of the moving average and Savitzky-Golay filters. By means of the Mathcad software, moving average and Savitzky-Golay filters were successfully applied to the smoothing of photochemical and electrochemical reactor data.

J. Verrelst, et al[10] proposed a model for Retrieval of Chlorophyll Content From Imaging Spectroscopy Data. A new statistical method was proposed within the family of nonparametric Bayesian statistics, namely Gaussian Processes regression (GPR). GPR is simpler and more robust than their machine learning family members while maintaining very good numerical performance and stability. Other features include:

(i) GPR requires a relatively small training data set and can adopt very flexible kernels,

(ii) GPR identifies the relevant bands and observations in establishing relationships with a variable, and finally

(iii) along with pixelwise estimations GPR provides accompanying confidence intervals. They used GPR to retrieve *Chl* from hyperspectral reflectance data and evaluated the portability of the regression model to other images. Based on field *Chl* measurements from the SPARC dataset and corresponding spaceborne CHRIS spectra (acquired in 2003, Barrax, Spain), GPR developed a regression model that was excellently validated (r^2: 0.96, RMSE: 3.82 μg/cm^2).

III. DATA COLLECTION

This section includes a brief description of the variables involved in the prediction problem and the objective data available for the study. The data considered have been obtained from specific networks devoted to the study of solar radiation and from atmospheric soundings, which are usually not available in places out of radiometric networks. In this sense, the application of machine learning techniques in these data constitutes a novel contribution of this letter[14].

We considered the following predictive input meteorological variables.

- In order to take into account the two main processes involved in the solar radiation extinction, namely, scattering and absorption, the aerosol optical depth can be considered because aerosols are one of the most important particles responsible for scattering and some of them can absorb solar energy. Thus, the daily mean aerosol optical depth product obtained from a meteorological centre is used as an input parameter.
- Temperature is a measure of the air’s hotness or coldness and is the most measured quantity of the atmosphere. Temperature is probably is easiest weather measurement to understand. More specifically, temperature describes the kinetic energy, or energy of motion, of the gases that make up air. As gas molecules move more quickly, air temperature increases and it is used as one of the input parameter for solar radiation prediction.
- Humidity is the amount of water vapor in the air. Most atmospheric water vapor comes from evaporation of water from the ocean or other bodies of water. Water vapor in the air greatly affects the type of weather we have. It is the source of all clouds, fog, and precipitation. Water vapor also helps warm the air when it absorbs energy from the sun.
- Atmospheric pressure is a measure of the weight of air in atmosphere above us. Air is made up of molecules of elements in gaseous state and minute dust particles. The particles, having mass, naturally exert a force on everything below. Measuring changes in air pressure is a valuable tool to predicting the weather because weather patterns move around in regions of high and low pressures. For example, a high pressure area

usually indicates fair weather; a low pressure area usually indicates stormy weather. A sudden drop in air pressure often means an approaching storm.

- Wind is moving air caused by differences in air pressure. Air moves from an area of higher pressure to an area of lower pressure. If there were no wind, there wouldn't be much day-to-day difference in our weather. Winds bring in different air masses and therefore, different weather patterns. The greater the differences in air pressure, the greater the wind speed. The unit of measure for wind speed is the knot. Wind speeds are measured either going toward or away from the radar station that is recording the measurement. If the wind is going towards the radar, the wind speed is recorded as a negative number. If the wind is heading away from the radar, it is recorded as a positive number.

IV. SOLAR RADIATION PREDICTION

The Figure. 1 depicts the steps involved in daily global solar irradiation prediction system. The solar sensor collects the solar irradiation data. The collected parameters include air temperature, relative humidity, air pressure, wind speed, wind direction, global horizontal irradiance, diffuse normal irradiance, etc. Temporal Gaussian process regression method is used to predict the future data values using the given set of known data values based on time series. The predicted solar irradiation values are correlated with actual or observed solar irradiation values for accuracy based on root mean square error, mean absolute error and mean bias error. The amount of energy that could be generated using different sizes of photovoltaic cell can be measured and analyzed to identify the solar energy based devices with request to the power requirement.

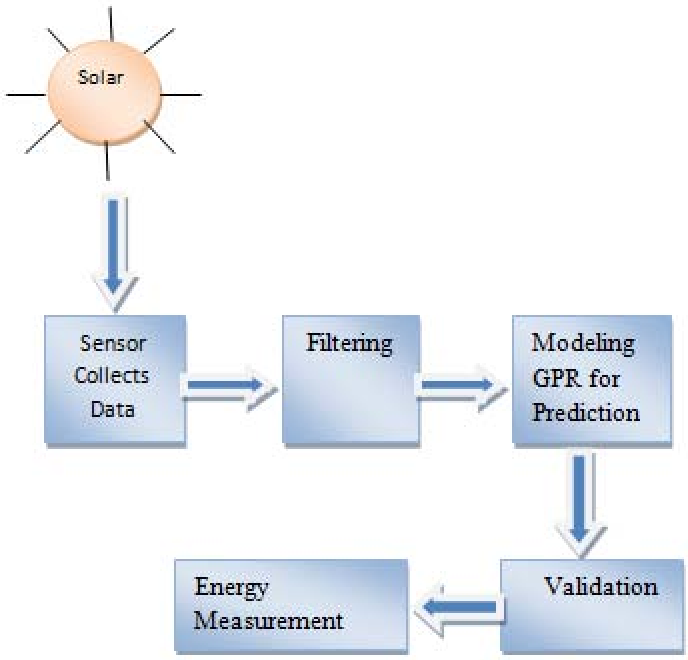

Figure 1. Solar Radiation Prediction.

A. Preprocessing

The global solar data around the Chennai region is collected for prediction process. The collected parameters like minimum air temperature, relative humidity, air pressure from the solar sensor obtained for five years is given as input to the preprocessing stage. The collected data set may contain too much of information out of which only a few data are required for daily global solar radiation prediction. Also, the use of noisy data will lead to failure in the prediction process. In preprocessing , moving average filtering technique is applied to remove the noise from the collected data values[11].

- *Moving average filter*

A calculation is made to analyze data points by creating a series of averages of different subsets of the full data set. A moving average is commonly used with time series data to smooth out the fluctuations in the data[9].

$$y = \text{filter}(b, a, x) \quad (1)$$

creates filtered data y by processing the data in vector x with the filter described by vectors a and b.The filter function is a general tapped delay-line filter, described by the difference equation

$$a(1)y(n)=b(1)x(n)+b(2)x(n-1)+\ldots+b(N_b)x(n-N_b+1) - a(2)y(n-1)-\ldots-a(N_a)y(n-N_a+1) \quad (2)$$

Here, n is the index of the current sample, N_a is the order of the polynomial described by vector a, and N_b is the order of the polynomial described by vector b. The output $y(n)$ is a linear combination of current and previous inputs, $x(n)x(n-1)...$, and previous outputs, $y(n-1)\ y(n-2)$.

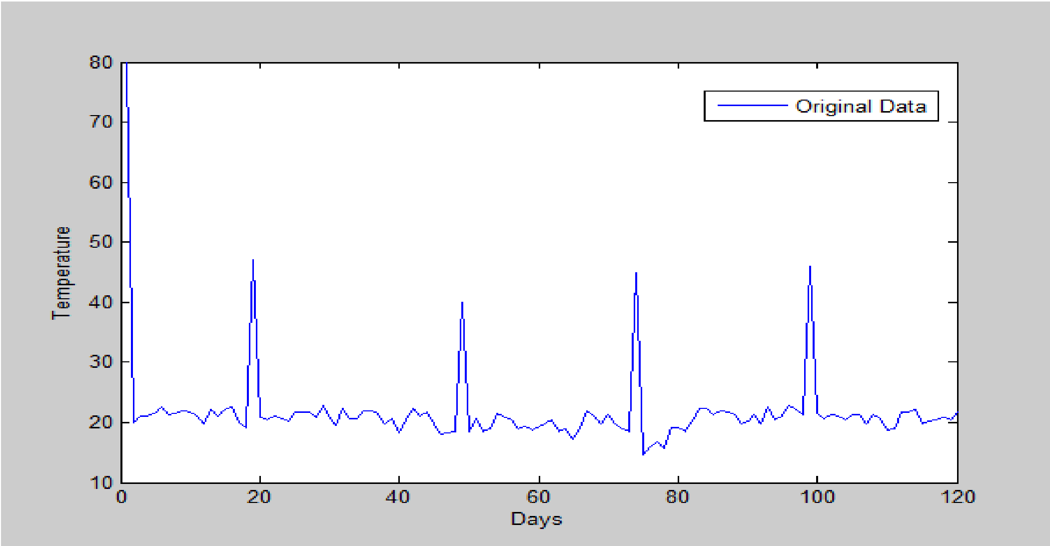

Figure 2. Maximum Air temperature before preprocessing

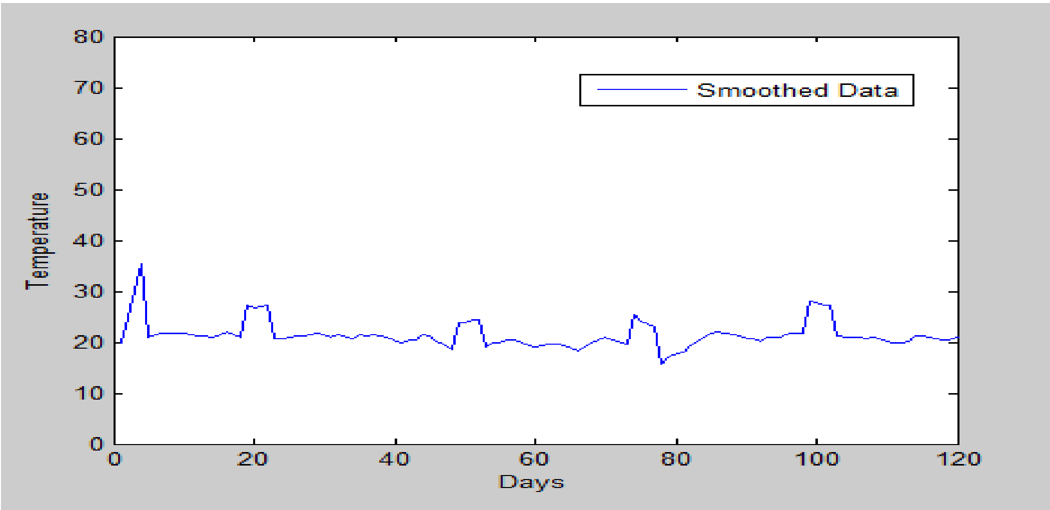

Figure 3. Maximum Air Temperature After Preprocessing

The Figure. 2 shows the erroneous data of maximum air temperature which is analyzed and filtered as shown in Figure. 3.

B. Prediction

Solar energy prediction is done using Gaussian process regression model. It is a method of interpolation for which the interpolated values are modeled by a Gaussian process governed by prior covariance. In this method, new data points can be constructed within the range of a discrete set of known data points. Under suitable assumptions on the priors, Gaussian process regression gives the best linear unbiased prediction of the values[12]. The main idea is to predict the value of the function at a given point by computing a weighted average of the known values of the function in the neighborhood of the point.

It is assumed that for a Gaussian process f observed at coordinates x, the vector of values f(x) is just one sample from a multivariate Gaussian distribution of dimension equal to number of observed coordinates |x|. Therefore under the assumption of a zero-meaned distribution,

$$f(x) \sim N(0, K(\theta,x,x')), \tag{3}$$

where K(θ,x,x') is the covariance matrix between all possible pairs (x,x') for a given set of hyperparameters θ. Maximizing this marginal likelihood towards θ provides the complete specification of the Gaussian process f.

Having specified θ making predictions about unobserved values f(x*) at coordinates x* is then only a matter of drawing samples from the predictive distribution

$$p(y^*|x^*,f(x),x) = N(y^*|A,B) \quad (4)$$

where K(θ,x*,x) is the covariance of between the new coordinate of estimation x* and all other observed coordinates *x* for a given hyperparameter vector θ, K(θ,x,x') and f(x) are defined as before and K(θ,x*,x*) is the variance at point *x** as dictated by *θ*.

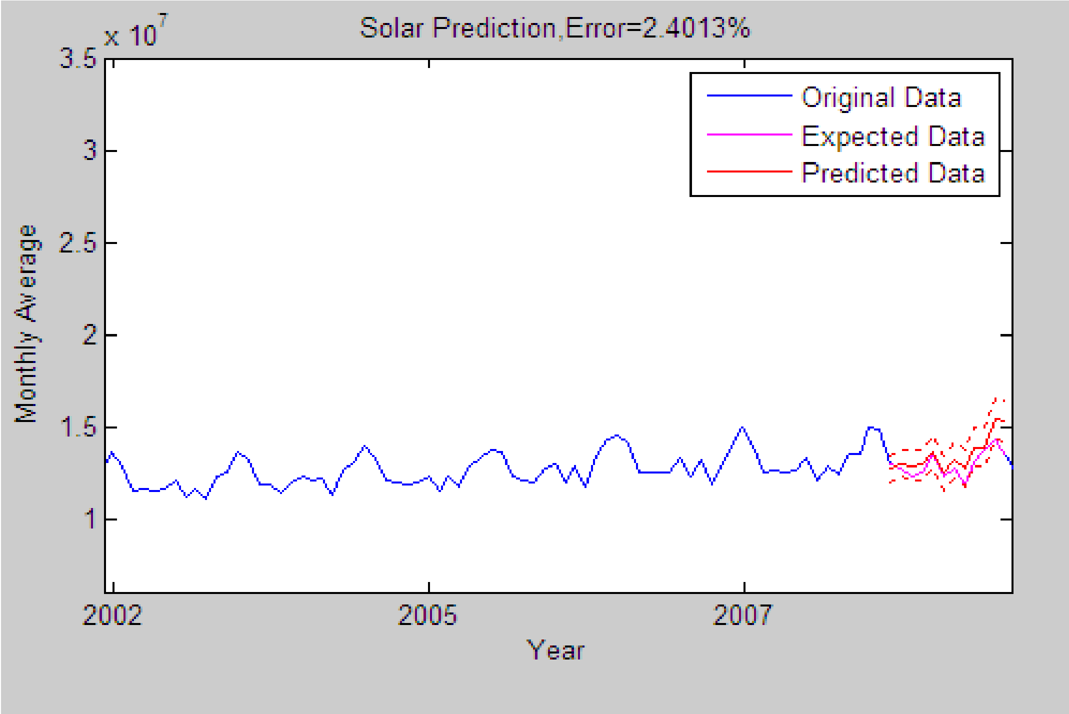

Figure 4. Monthly Average Solar Radiation Prediction

The Figure. 4 shows the monthly average solar radiation prediction for one year with an error of 4.5998%.

It is important to note that practically the posterior mean estimate f(x*) is just a linear combination of the observations f(x); in a similar manner the variance of f(x*) is actually independent of the observations f(x). A known bottleneck in Gaussian process prediction is that the computational complexity of prediction is cubic in the number of points |x| and as such can become unfeasible for larger data sets. Works on sparse Gaussian processes, that usually are based on the idea of building a representative set for the given process f.

C. Validation

The predicted solar irradiation values are compared with actual or observed solar irradiation values for accuracy based on root mean square error or root mean square deviation (RMSD), mean absolute error(MAE) and mean bias error(MBE)[13].

- *RMSD:*

The root-mean-square deviation (RMSD) or root-mean-square error (RMSE) is a frequently used measure of the differences between predicted values and actually observed values.

$$\text{RMSD} = \sqrt{\frac{\sum_{t=1}^{n}(\hat{y}_t - y_t)^2}{n}} \quad (5)$$

The RMSD of predicted values $\hat{y}_t$ for times *t* of a regression's dependent variable y is computed for *n* different predictions as the square root of the mean of the squares of the deviations.

- *MAE:*

The mean absolute error (MAE) is a quantity used to measure how close predictions are to the eventual outcomes. The mean absolute error is given by

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|f_i - y_i| = \frac{1}{n}\sum_{i=1}^{n}|e_i| \quad (6)$$

The mean absolute error is an average of the absolute errors $e_i=|f_i-y_i|$, where f_i is the prediction and y_i the true value.

TABLE 1. RMSE AND MAE BETWEEN THE ACTUAL AND THE ESTIMATED DAILY SOLAR IRRADIATION OF REGRESSION MODELS

METHOD	RMSE	MAE
ELM	4.22	2.11
SVR	3.23	2.12
GPR	2.15	1.27
TGPR	0.865	0.724

V. CONCLUSION

This paper has presented GPR for global solar radiation prediction based on the solar parameter values collected for the southern part of India. Gaussian process regression method based on time series provides more accurate, bias and robust values when compared with other machine learning techniques like Support Vector Machine(SVM) and Artificial Neural Networks(ANN). And the accuracy is measured based on root mean square deviation(RMSD) and mean absolute error(MAE). Solar energy prediction provides support for measuring and analyzing solar energy. This information is used to identify the suitable solar energy based devices that can be implemented with particular panel size at a specific location.

REFERENCES

[1] Rogiros D. Tapakis and Alexandros G. Charalambides "Monitoring Cloud Motion in Cyprus for Solar Irradiance Prediction", Conference Papers in Energy, 2013.
[2] Sancho Salcedo-Sanz, Casanova-Mateo, Jordi Munoz-Mari" Prediction of Daily Global Solar Irradiation Using Temporal Gaussian Processes"IEEE Geosci. Remote Sens.,vol.11,No.11,2014.
[3] Bj orn Wol, Elke Lorenz, Oliver Kramer" Statistical Learning for Short-Term Photovoltaic Power Predictions", 2012.
[4] J. Verrelst, L. Alonso, J. P. Rivera, J.Moreno, and G. Camps-Valls, “Gaussian process retrieval of chlorophyll content from imaging spectroscopy data,” IEEE J. Sel. Topics Appl.Earth Observ. Remote Sens., vol. 6, no. 2,pp. 867–874, Apr. 2013.
[5] S. Salcedo-Sanz, " Daily global solar radiation prediction based on a hybrid Coral Reefs Optimization – Extreme Learning Machine approach", Solar Energy 105 (2014).
[6] J. L.Chen, H. B. Liu,W.Wu, and D. T. Xie, “Estimation of monthly solar radiation from measured temperatures using support vector machines—A case study,”Renew. Energy, vol. 36, no. 1, pp. 413–420, Jan. 2011.
[7] Rasheed H. AL-Naimi, Ali M. AL-Salihi, Dher I. Bakr, “Neural network based global solar radiation estimation using limited meteorological data for Baghdad, Iraq”, International Journal of Energy and Environment (IJEE), Volume 5, Issue 1, 2014.
[8] M. Hassan, "Solar Harvested energy prediction algorithm for wireless sensors," Quality Electronic Design(ASQED), 2012 4th Asia Symposium on, vol., no., pp.178,181,10-11 July 2012.
[9] José Luis Guiñón, Emma Ortega, José García-Antón, Valentín Pérez-Herranz" Moving Average and Savitzki-Golay Smoothing Filters Using Mathcad ", International Conference on Engineering Education, 2007.
[10] J. Verrelst, L. Alonso, J. P. Rivera, J.Moreno, and G. Camps-Valls, “Gaussian process retrieval of chlorophyll content from imaging spectroscopy data,” IEEE J. Sel. Topics Appl.Earth Observ. Remote Sens., vol. 6, no. 2,pp. 867–874, Apr. 2013.
[11] Omar W. Ahmed , Rami Qahwaji , Tufan Colak ,Paul A. Higgins , Peter Prediction Using Advanced Feature Extraction, Machine Learning, and Feature Selection"Solar Phys,2011.
[12] T.Gallagher, D. Shaun Bloomfield " Solar Flare Joao Paulo Coelho and José Boaventura-Cunha" Long Term Solar Radiation Forecast UsingComputational Intelligence Methods", Applied Computational Intelligence and So Computing, 2014.
[13] M. Ant´on , R. Rom´an, A. Valenzuela , F. J.Olmo3,4, and L. Alados-Arboledas, "Directsuntotal ozone data from a spectroradiometer: methodology and comparison with satellite observations" , Atmos. Meas. Tech., 6, 637– 647, 2013.
[14] http://e_missions.net/wvstorm
[15] http://photovoltaic-software.com/PV-solar-energy-calculation.php