THE JOURNAL OF
COGNITIVE SYSTEMS
VOLUME 02 NUMBER 01 JUNE 2017

PREDICTION OF SOLAR RADIATION BASED ON MACHINE LEARNING METHODS

Seckin Karasu[1], Aytac Altan[1], Zehra Sarac[1], Rifat Hacioglu[1]
[1]Bülent Ecevit University, Zonguldak, Turkey

In this study, machine learning methods which linear regression and Gaussian process regression models are used to estimate the solar radiation on daily data set taken from the wind central in Zonguldak province in Turkey. The measured wind speed, temperature, pressure, humidity parameters together with solar radiation are used for the prediction process. In the prediction process, number of delay steps from 3 to 12 for these parameters are applied to the developed models. In order to determine the performance of the obtained model, the model is evaluated in terms of statistical error criteria such as MAE, MSE and RMSE. The least prediction error for the solar radiation prediction process is determined. It has been observed that Gaussian regression model approach provides a high performance to predict solar radiation with related to other measured parameters.

Index Terms —Solar radiation, prediction, linear regression, Gaussian process regression, machine learning

I. INTRODUCTION

MACHINE learning models are used to find a relationship in pattern recognition and classification problems where there is no representation between input and output, in data mining and prediction problems [1]. In machine learning models, there are supervised learning methods such as linear regression, nonlinear regression, artificial neural networks [2, 3], support vector machine, k-nearest neighbors. Beside them, there are unsupervised learning methods and ensemble learning methods [1].

In the literature, single and hybrid estimator models are constructed for the estimation of hourly solar radiation. In the first stage, Multi-Layer Perception (MLP), Autoregressive Moving Average (ARMA) and persistence models are established. In the second stage, these are combined with Bayesian rules, resulting in a 14% improvement in the estimate [4]. Polynomial Basis Function (PBF) and Radial Basis Function (RBF) based Support Vector Regression (SVR) method are used for daily solar radiation estimation. It is shown that the SVR method based on PBF with different statistical indicators has higher prediction performance for 1460 days solar radiation data [5]. Support Vector Machine (SVM) and Wavelet Transform (WT) methods have been combined to develop a hybrid estimation method. The proposed method is compared with other methods such as Artificial Neural Network (ANN), Genetic Programming (GP) and ARMA. It has been shown that the proposed prediction model gives a low error value in estimating the solar radiation for the different input parameters [6]. Genetic Algorithm (GA) and pruning method based on optimal NARX estimator based on optimal brain surgeon optimization methods are used for wind speed and solar radiation estimation. The wind speed and solar radiation are estimated for different time periods from 8 hours to 24 hours. It is mentioned that the proposed method can be applied in photovoltaic (PV) estimation and wind power generation [7]. In [8], monthly temperature, sunshine duration, meteorological data are used for solar radiation estimation. Adaptive Neuro-Fuzzy Inference System (ANFIS) is used in the estimation phase. It is stated that the proposed model is effective for practical applications.

In this study, solar radiation estimation is done by linear regression and Gaussian process regression methods using 1-year data consisting of daily time series. For wind speed, temperature, humidity and pressure variables, the coefficients obtained by the 1st order curve fitting method are calculated attached to the number of delay steps and are applied to the models. In the phase of model training, 10-fold cross-validation method is used. According to the model results, models created with statistical indicators such as MAE, MSE and RMSE are calculated for different delay step counts. It is seen that the minimum MSE value for estimating the solar radiation value is given by the Gaussian Process Regression model for delay step count 11.

This article is organized as follows: used methods and modelling are presented in Section II. Simulation and experimental results are given in Section III. Conclusions are finally discussed in Section IV.

II. USED METHODS AND MODELLING

In this study, one-year data consisting of daily data taken from meteorological station in Zonguldak province is used for solar radiation estimation. The maximum, minimum, mean and standard deviation values for wind speed, temperature, relative humidity, current pressure and solar radiation variables in the used data set are given in Table I. For solar radiation, hourly data values are summed to get daily data set.

Manuscript received April 20, 2017; accepted May 25, 2017.
Corresponding author: Rifat Hacioglu (E-mail: hacirif@beun.edu.tr; ORCID® ISNI: 0000-0002-2480-0729).
Digital Object Identifier:

Table I. Statistical Values of Variables

Parameters	Max.	Min.	Mean	Std. Dev.
Wind Speed (m/s)	5	0	2.07	0.78
Temperature (°C)	26.2	0.2	15.03	7.17
Humidity (gr/m3)	99	31	75.66	14.40
Pressure (mbar)	1016	978	999	6.13
Solar radiation (W/m2)	22130	480	10864	6382

A. Multiple Linear Regression Model

A In the multiple linear regression model, the relationship between multiple independent variables ($x_1,\ x_2,\ x_3,\ x_4$) and a dependent variable ($\hat{y}$) is examined. The regression function used in this study is the first order, and it is assumed that each independent variable is a linear relationship with the dependent variable, as in (1)

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4 \tag{1}$$

where b_0 is the y-axis cut point of the modified regression curve, b_1 is the coefficient x_1 of the first guess variable, and b_2 is the coefficient x_2 of the first guess variable so on.

In this study, wind speed, temperature, humidity, pressure are used as independent variable ($x_1,\ x_2,\ x_3$ and x_4) respectively and solar radiation is used as a dependent variable ($\hat{y}$).

B. Gaussian Process Regression Model

In supervised learning, it is expected that similar predictor values x_i and response values y_i have close. In Gaussian processes this similarity is given by a related covariance function [9]. It is determined the covariance between two latent variables $f(x_i)$ and $f(x_j)$ for $i \neq j$. The signal length of the predictor is expressed as N.

The covariance function is expressed by different kernel functions such as Squared exponential, Exponential, Gibbs, Matérn and Rational quadratic kernel. The rational quadratic kernel is used in this study. The covariance function $k(x_i,x_j)$ can be expressed as $k(x_i, x_j|\hat{y})$ in (2) in terms of kernel parameters in the $\hat{y}$ vector.

$$k\left(x_i, x_j \mid \hat{y}\right) = \sigma_f^2 \left(1 + \frac{r^2}{2\alpha\sigma_l^2}\right)^{-\alpha} \tag{2}$$

$$r = \sqrt{\left(x_i - x_j\right)^T \left(x_i - x_j\right)} \tag{3}$$

σ_f and σ_l denote the signal standard deviation and characteristic length scale, respectively. α indicates rational quadratic exponent. The value of r in (3) is the Euclidean distance between x_i and x_j. It is possible to use a separate length scale (σ_m) for each predictor $m,\ m = 1, 2, \ldots, N$. For each predictor, covariance functions with separate length scale implement automatic relevance determination (ARD). In this case, the covariance function, $k(x_i, x_j|\hat{y})$ is expressed as in (4).

$$k\left(x_i, x_j \mid \hat{y}\right) = \sigma_f^2 \left(1 + \frac{1}{2\alpha}\sum_{m=1}^{N}\frac{\left(x_{im} - x_{jm}\right)}{\sigma_m^2}\right)^{-\alpha} \tag{4}$$

C. Feature Extraction and Normalization with Curve Fitting

Using (5), coefficients of curve fitting in the first order linear function are calculated taking into account the number of delay steps for wind speed, temperature, humidity and pressure values.

$$\hat{y} = \varphi_i x + \psi_i \tag{5}$$

Table II. Extracted Features and Their Labels

Parameters	Features	Feature label
Wind Speed	$[\varphi_1, \psi_1]$	F1 F2
Temperature	$[\varphi_2, \psi_2]$	F3 F4
Humidity	$[\varphi_3, \psi_3]$	F5 F6
Pressure	$[\varphi_4, \psi_4]$	F7 F8

The min-max normalization equation in (6) is used for the feature coefficients in Table II and solar radiation variable.

$$f_{i,new} = \frac{f_{i,old} - f_{min}}{f_{max} - f_{min}} \tag{6}$$

where f is the feature vector, f_i is i th element of feature vector, $f_{i,old}$ is the old value of i th element in feature vector, f_{max} is the maximum value of feature vector, f_{min} is the minimum value of feature vector, $f_{i,new}$ is the new value of i th element in feature vector. In Table I, the range values for all variables are converted to the range $[0, 1]$.

D. Evaluation of Model Accuracy

There is no single criterion for evaluating the performance of the model. Graphical representations can give an idea of model performance. Model performance can obtained by plotting the estimated time series on the real time series. Distribution of the error can be observed with the scatter plot [1].

Apart from these, there are indicators that represent mathematical model performance. In Equation (7-9), Mean Absolute Error (MAE), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) are given.

$$MAE = \frac{1}{N}\sum_{n=1}^{N}\left|e_n\right| \tag{7}$$

$$MSE = \frac{1}{N}\sum_{n=1}^{N}e_n^2 \tag{8}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N} e_n^2} \qquad (9)$$

The error is expressed

$$e_n = y_n - \hat{y}_n \qquad (10)$$

where y_n is observed for a given time n and $\hat{y}_n$ is the predicted time series. The Pearson linear correlation coefficient, which is an indicator of the relationship between input and output for $m = 4$, is calculated as

$$R = \frac{\sum_{i=1}^{m}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{m}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{m}(y_i - \bar{y})^2}} \qquad (11)$$

when the R value is 1, it is said that there is a linear relationship. When the R value is 0, there is a nonlinear relationship between input and output.

III. SIMULATION AND EXPERIMENTAL RESULTS

The one-year wind speed, temperature, relative humidity, current pressure and solar radiation data from the daily time series from the meteorological station are filtered by 10-day average filter in the smoothing process. The curve fitting coefficients are obtained from each feature by the first order curve fitting method. In this study, solar radiation estimation is performed by using different numbers of delay steps for linear regression and Gaussian process regression models such as wind speed, temperature, relative humidity and current pressure values as input, solar radiation value as output.

The generated models are compared with graphical representations, error indicators and the best prediction model is chosen for number of delay steps from 3 to 12. The simulation is performed in MATLAB. The errors that occur for two different estimation models are given in Figure 1-3.

When the MAE, MSE and RMSE values are examined, it is observed that the error values decrease when the number of delay steps is increased. In the Gaussian process regression method, it is seen that the error amount is less than the linear regression method. The minimum MSE value is obtained for the Gaussian process regression model when the number of delay step is 11.

The solar radiation value estimated by the model obtained by the Gaussian Process Regression Method and the actual value are shown in Figure 4. When the estimated values of the obtained model are examined, it is seen that the error amount is very low. The relationship between the actual values and the predicted values is shown in scatter plot in Figure 5. This relation is expressed by first order linear regression with $y = 0.99x + 0.0033$ equation. R is calculated as 0.99792, and it is concluded that the model is successful in predicting the actual solar radiation values.

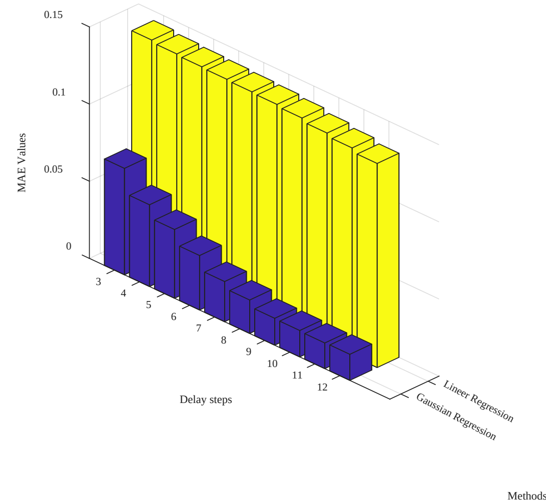

Fig. 1. The MAE change due to the number of delay steps for the two estimation methods.

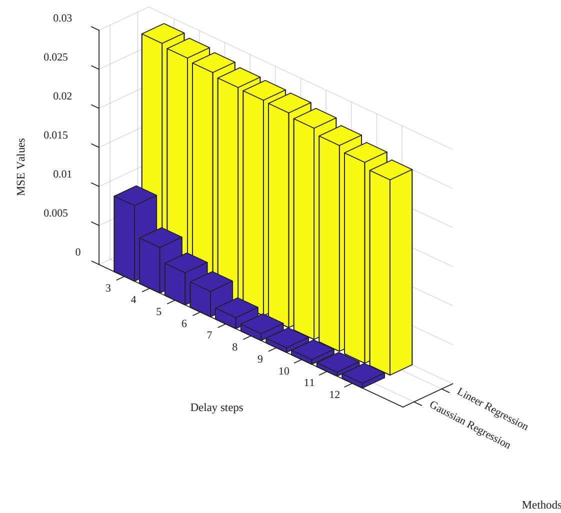

Fig. 2. The MSE change due to the number of delay steps for the two estimation methods.

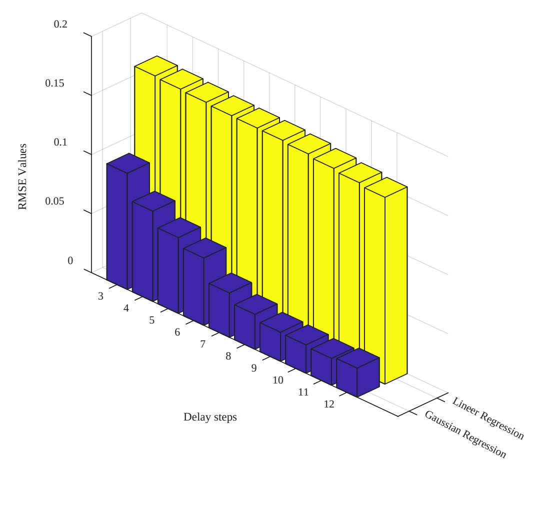

Fig. 3. The RMSE change due to the number of delay steps for the two estimation methods.

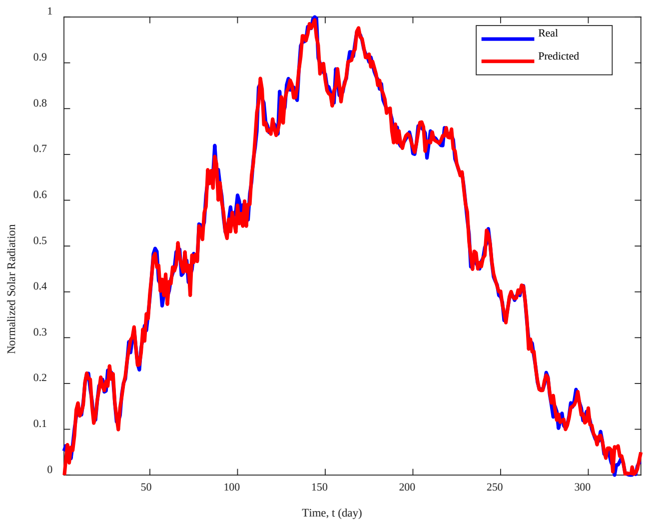

Fig. 4. Solar radiation prediction by Gaussian process regression method.

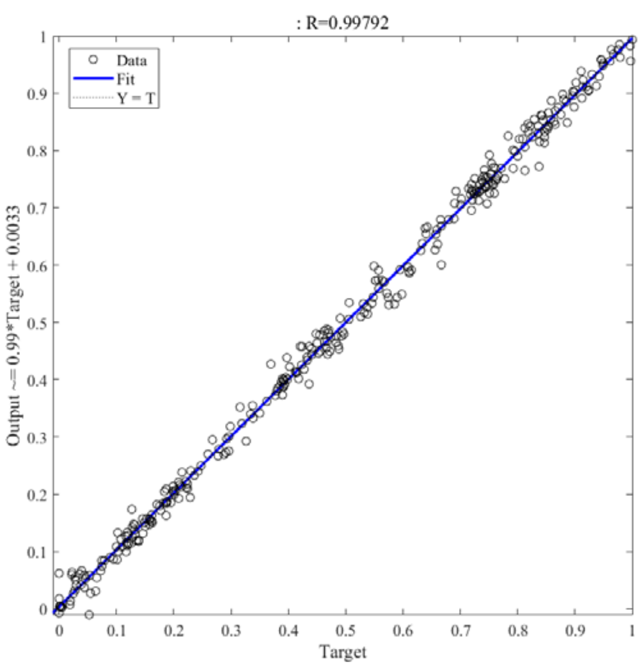

Fig. 5. Scatter plot for real and predicted points.

IV. CONCLUSIONS

In this article, linear regression and Gaussian process regression methods are used to develop a solar radiation prediction model that gives the least error. One-year wind speed, temperature, pressure, humidity and solar radiation values data consisting of one daily time series taken from meteorological station in Zonguldak province are used for solar radiation estimation. Coefficients are generated by first order curve fitting for variables outside the solar radiation. In the solar radiation prediction process, number of delay steps from 3 to 12 for these parameters are applied to the created models. In training phase, in order to measure the model performance, a 10-fold cross-validation method is applied independently of the data. It is determined that the Gaussian process regression method has a lower MSE value than the linear regression method. For the model and model parameters for which the best result is obtained, the MAE value is calculated as 0.016620, the MSE value is 0.000514 and the RMSE value is 0.022674. The predicted solar radiation versus real solar radiation for the given minimum error model is plotted. Also, it is observed that the performance is high by scatter plot.

ACKNOWLEDGMENT

This study was supported by Bülent Ecevit University (BAP Project No: 2012-17-15-01 and 2014-75737790-01). The authors would like to thank Bülent Ecevit University for their support. The study is selected from International Engineering Research Symposium-UMAS 2017 (Duzce University).

REFERENCES

[1] C. Voyant, G. Notton, S. Kalogirou, M.L. Nivet, C. Paoli, F. Motte, and A. Fouilloy, “Machine learning methods for solar radiation forecasting: A review”, Renewable Energy, 105, 2017, pp. 569-582.

[2] S. Karasu, A. Altan, Z. Saraç, and R. Hacıoğlu, “Prediction of wind speed with non-linear autoregressive (NAR) neural networks”, IEEE 25th Signal Processing and Communications Applications Conference, 2017, pp. 1-4.

[3] S. Karasu, A. Altan, Z. Saraç, and R. Hacıoğlu, “Estimation of fast varied wind speed based on NARX neural network by using curve fitting”, International Conference on Advanced Technology & Sciences (ICAT’17), 2017, pp. 551-557.

[4] C. Voyant, C. Darras, M. Muselli, C. Paoli, M.L. Nivet, and P. Poggi, “Bayesian rules and stochastic models for high accuracy prediction of solar radiation”, Applied Energy, 114, 2014, pp. 218-226.

[5] K. Mohammadi, S. Shamshirband, M.H. Anisi, K.A. Alam, and D. Petkovic, “Support vector regression based prediction of global solar radiation on a horizontal surface”, Energy Conversion and Management, 91, 2015, pp .433-441.

[6] K. Mohammadi, S. Shamshirband, C.W. Tong, M. Arif, D. Petkovic, and S. Ch, “A new hybrid support vector machine–wavelet transform approach for estimation of horizontal global solar radiation”, Energy Conversion and Management, 92, 2015, pp. 162-171.

[7 A. Di Piazza, M.C. Di Piazza, and G. Vitale, “Solar and wind forecasting by NARX neural networks”, Renewable Energy and Environmental Sustainability, 39, 2016, pp. 1-5.

[8] L. Olatomiwa, S. Mekhilef, S. Shamshirband, and D. Petkovic, “Adaptive neuro-fuzzy approach for solar radiation prediction in Nigeria”, Renewable and Sustainable Energy Reviews, 51, 2015, pp. 1784-1791.

[9] C.E. Rasmussen and C.K.I. Williams, “Gaussian processes for machine learning”, MIT Press, 2006.

[10] A.G. Wilson, “Covariance kernels for fast automatic pattern discovery and extrapolation with Gaussian processes”, University of Cambridge, 2014.

[11] R.M. Neal, “Bayesian Learning for Neural Networks”, Toronto University Doctorate Thesis, 1995, p. 118.

BIOGRAPHIES

Seçkin Karasu is a research assistant at the Department of Electrical Electronics Engineering, Bülent Ecevit University in Turkey. He recieved the BSc and M.Sc. degrees in the department of Electrical Electronics Engineering from Bülent Ecevit University in 2011 and 2013, respectively. He is a Ph.D. candidate. His research interests are signal processing, image processing, machine learning and power systems.

Aytaç Altan is a lecturer at the Department of Electrical Electronics Engineering, Bülent Ecevit University in Turkey. He received the BSc and M.Sc. degrees in the department of Electrical Electronics Engineering from Anadolu University in 2011 and 2016, respectively. He is a Ph.D. candidate. His research interests are signal processing, machine learning, robotic systems, adaptive control and system identification.

Zehra Saraç is a Professor at the Department of Electrical Electronics Engineering at the Bülent Ecevit University in Turkey. She received the M.Sc. and Ph.D. degrees in the department of Electronics Engineering from Gebze Technical University 1996 and 2002, respectively. Her research interests are electromagnetic waves, signal processing, image processing, opto-electronics, holography, interferometry and microwave photonics.

Rıfat Hacıoğlu is an Assistant Professor at the Department of Electrical Electronics Engineering at the Bülent Ecevit University in Turkey. He received the M.Sc. and Ph.D. degrees in the department of Electrical Electronics Engineering from Illinois Institute of Technology in 1996 and 2002, respectively. His research interests are signal processing, control applications, linear and nonlinear system identification, parametric model estimation approaches, microprocessor based control systems, image identification problems.