International Conference on Applied Energy 2019
Aug 12-15, 2019, Västerås, Sweden
Paper ID: 0894

DEVELOPMENT OF A FORECASTING MODEL FOR TROPICAL SOLAR ENERGY SYSTEMS USING SUPPORT VECTOR MACHINES

Alvin B. Culaba[1*], Aaron Jules R. Del Rosario[2], Aristotle T. Ubando[1]

[1]Mechanical Engineering Department and Center for Engineering and Sustainable Development Research, De La Salle University, 2401 Taft Avenue, Manila, Philippines 0922 (*Corresponding Author)
[2]Graduate Student, Mechanical Engineering Department, De La Salle University, 2401 Taft Avenue, Manila, Philippines 0922

ABSTRACT

Solar energy is a sustainable source that is favored in tropical areas where the resource intensity is high. However, developing accurate forecasting models, which is crucial in the optimal design and operation of tropical solar energy systems, has been challenging due to the high-dimensional nature of the system. This study presents a novel forecasting model for such systems using support vector machines. The proposed model was developed using a data-driven methodology. Model optimization and feature selection were applied to improve predicting accuracy. Modeling results show that the medium Gaussian function provided the most desirable balance between the accuracy and speed of the training functions. The previous hour observation was found to be the most significant input, while some variables considered in the initial model caused overfitting. The final model had superior accuracy over the initial models and those developed in the literature, thereby validating the effectiveness of the presented methodology.

Keywords: renewable energy, solar energy systems, energy system modeling, forecasting model, machine learning, support vector regression

1. INTRODUCTION

Renewable energy (RE) sources have been sought as a more sustainable solution to meet the increasing energy demand of cities. Solar energy is the most widely adopted urban RE source due to its low space requirement, maintenance, and noise. However, its downsides are the intermittency of its supply and unavailability at night [1].

In order to address this problem, proper planning must be conducted at the design and operation stage of solar energy systems. Forecasting models, which predict future outputs based on various variables (e.g., previous output, weather conditions), have been critical to solar energy system planning. The application of these models has led to an increase in efficiency and viability of solar energy systems [2]. This topic remains as an active area of research as researchers try to improve model performance and search for new areas of application.

Solar energy systems are favored in tropical areas where solar irradiance intensity is high. However, predicting solar energy system output in these areas is more challenging as such systems are sensitive to more weather variables than those in colder climates [3].

Forecasting models of solar energy systems based on machine learning techniques are gaining popularity in the literature as they require less computational effort than physical models but have higher accuracy than traditional time-series statistical models. Most machine learning models use artificial neural networks (ANN), but it was found that support vector machines (SVM) have a superior predicting ability in high-dimension datasets or those with more variables [4].

Despite the clear advantages of SVM as a machine learning algorithm in forecasting high-dimensional systems, there is a lack of studies applying this method to tropical solar energy systems. Thus, this study aims to contribute to this field of research by presenting the development of a forecasting model for tropical solar energy systems using the SVM method. Furthermore, the proposed model was improved using selection methods for the training function selection and input features. The performance of the final model was compared to the results of initial modeling and those obtained from previous studies [3].

2. THEORY

SVM is a supervised machine learning algorithm based on statistical learning theory and structural risk minimization. Support vector regression (SVR) is the implementation of SVM for function fitting and regression problems. The idea of SVR is to determine a linear regression function to approximate the output vector. For nonlinear problems, the data is first mapped into a high-dimensional feature space using a kernel trick, which uses polynomial, exponential, or radial functions. Then, the regression function is approximated through the minimization of a regularized function [5].

3. MATERIAL AND METHODS

The proposed forecasting model was developed and applied to a case study of a solar energy system in a tropical area. This section describes the case study and the methodology used to develop the model.

3.1 Case Study

The case study considers a solar energy system in a location with a tropical monsoon climate (i.e., Köppen climate classification category "*Am*"), covering sections in Central/South America, West/Central Africa, and South/Southeast Asia. The model predicts the Global Horizontal Irradiance (GHI), which is utilized by most solar energy systems. The input dataset contains the following variables: GHI, dry bulb temperature (DBT), dew point temperature (DPT), wind speed (WS), surface pressure (SP), cloud type, and clear sky GHI. The model was developed using the MATLAB R2019a software and a 16.0-GB Intel® Core™ i7-6500 CPU hardware.

3.2 Data Acquisition

The input dataset was acquired from the National Solar Radiation Data Base (NSRDB) of the National Renewable Energy Laboratory (NREL). The area of application has geographical coordinates of (25.77, -80.18) with a Köppen climate category *Am*. The dataset contains 17 years of 30-minute observations, covering the years from 1998 to 2014 [6].

3.3 Data Pre-Processing

The data pre-processing procedure used to make the data suitable for modeling is described as follows. First, the dataset was regularized into hourly intervals to reduce its size. Then, missing data points were filled using a jumping technique described in [7], which takes the mean of the corresponding data points for the same day of the previous and next years. Then, data outliers and noise were detected using the moving mean window method and cleaned through polynomial interpolation. Finally, the dataset was normalized to restrict the range of values of each variable between 0 and 1 and improve the quality and precision of the regression method [3].

3.4 Feature Engineering

Features from the dataset were extracted based on those that are significant in modeling the system as determined by previous studies. The following features were extracted. First is the clear sky index (CSI), which is the ratio of the actual GHI and clear sky GHI. Second is the solar zenith angle, which is a function of both temporal (i.e., the hour of the day, and the day of the year) and spatial (i.e., latitude) variables. Third are temporal features, which include the hour and the month. Fourth are lagged variables, which includes the previous hour, previous day, and previous year observations obtained from the historical dataset [8].

3.5 Model Training and Validation

The model dataset was first partitioned into training and testing sets, which consists of 70% and 30% of the data, respectively. The training set was used by the model for learning and validation, while the test set was used to evaluate the model after training. The former set was used to train the model through the Regression Learner application of MATLAB. A holdout validation procedure, which is recommended for large datasets, was done in this study. In this method, 25% of the training data is held out during the learning stage and is used to validate the trained model and compute for the error metrics [9]. The performance of the model was measured using the following commonly used metrics in regression: root mean square error (RMSE) and coefficient of determination (R^2). The training time and prediction speed of the training model were also computed.

3.6 Model Optimization

Multiple SVR models were developed using different kernel functions, namely: linear, quadratic, cubic, fine Gaussian, medium Gaussian, and coarse Gaussian functions. The hyperparameters of the training models were tuned by the software using the Bayesian optimization algorithm [9]. The final model used in the evaluation stage was chosen based on the performance and speed of the trained models.

3.7 Feature Selection

A backward stepwise regression procedure was conducted to determine the relative significance of individual variables in the model. This method was done by eliminating input variables from the model one at a time and computing the difference in the performance of the original model to the new model [10].

3.8 Model Testing and Evaluation

The final model was developed using the chosen kernel function and significant input variables. The test set was used for the evaluation of the developed model. Finally, the performance of the final model was compared to the training model and models from previous studies using the normalized RMSE metric [3].

4. RESULTS AND DISCUSSION

This section presents and discusses the results of developing the model for the case study.

4.1 Data Description

The dataset consists of 298,032 data points, with six numerical variables and one categorical variable. The numerical variables are statistically summarized in Table 1. The categorical variable, cloud type, has a following range of values: {clear, probably clear, fog, water, mixed, opaque ice, cirrus, overlapping, overshooting, unknown}.

Table 1. The statistical description of the numerical variables in the acquired dataset.

Variable	Unit	Min	Max	Mean	Standard Deviation
GHI	W/m^2	0	1065	222.38	304.27
DBT	°C	5.0	32.0	24.77	3.53
DPT	°C	1.0	27.0	20.50	4.06
WS	m/s	0	25.7	4.22	1.87
SP	mbar	970	1030	1012.02	4.59
Clear Sky GHI	W/m^2	0	1065	266.13	340.02

4.2 Model Selection Results

The performance metrics of the different SVR training models based on the selected kernel functions are shown in Table 2.

Table 2. Performance and speed comparison of different SVR models trained in the study.

Kernel Function	RMSE	R^2	Training Time [s]	Prediction Speed [/s]
Linear	0.0549	0.97	975.53	~2700
Quadratic	0.0203	1.00	1640.2	~13000
Cubic	0.0161	1.00	4152.1	~35000
Fine Gaussian	0.0940	0.91	2351.9	~1200
Medium Gaussian	0.0169	1.00	2578.9	~14000
Coarse Gaussian	0.0217	1.00	2666.1	~5300

The results show that the cubic kernel function yields the lowest error and highest prediction speed; however, it takes the longest time to train. The second-best in accuracy and speed is the medium Gaussian function but has significantly less training time. This performance can be advantageous when the model is physically deployed to solar energy systems, where model learning must be done in real-time for efficient operation. Due to this consideration, the medium Gaussian kernel function was chosen for the final model.

4.3 Feature Selection Results

The change in model performance from the sequential removal of each variable is shown in Table 3.

Table 3. The change in model performance caused by removing each variable one at a time.

Variable	ΔRMSE	Variable	ΔRMSE
DBT	-0.7%	Zenith Angle	+2.1%
DPT	-0.8%	Hour	-136%
WS	+0.7%	Month	-40.8%
SP	+1.5%	Previous Hour	-137%
Cloud Type	-0.2%	Previous Day	-3.4%
CSI	-58.0%	Previous Year	-3.0%

The results show that the accuracy of the model increased when the wind speed, pressure, and zenith angle variables were removed from inputs. These results mean that model overfitting is present when these inputs are included. Thus, these variables were removed from the final model.

4.4 Development of the Final Model

After developing the final model, it was evaluated using the test set. A comparison of the predicted and actual values over a 24-hour sample period from the evaluation of the final model is shown in Fig. 1.

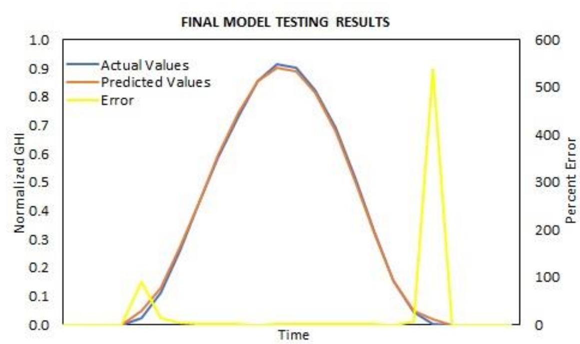

Fig. 1. Sample final model testing results.

The results show that the developed final model provides accurate predictions for the unbiased test set. The normalized RMSE metric decreased from 1.7238% in training to 1.6929% in testing, which further validates the accuracy of the model. The performance of the proposed model was also better than those developed in the literature, with normalized RMSE values ranging from 7.19%-19.67% for ANN models and 2.74%-10.24% for SVM models. The downside of the proposed model, however, is its significantly inaccurate prediction at the extremes of the sunshine hours. Nevertheless, these values are low in magnitude and have a relatively less economic impact during implementation.

5. CONCLUSION

This study presented a novel forecasting model for tropical solar energy systems using SVM. A dataset acquired from NREL NSRDB was used as model input, and the model was developed in MATLAB. The methodology involved data pre-processing, feature engineering, model optimization, and feature selection procedures to improve the performance of the proposed model. The results show that the medium Gaussian kernel function balances the tradeoff between predictive accuracy and training speed of the forecasting model, which is advantageous when applied to physical systems. Also, some initial variables, such as wind speed and surface pressure, caused model overfitting. The final model was trained using the chosen kernel function and predictors to maximize accuracy and avoid overfitting, then evaluated using a test set unused during training. The evaluation of the final model showed an increase in predicting accuracy compared to the initial models. Also, the proposed model in the study had superior performance compared to those developed in previous studies utilizing ANN and SVM algorithms. The only weakness of the proposed model was in predicting outputs at the extremes of the sunshine hours. Future studies can improve pre-processing and feature engineering methods to increase model accuracy, such as removing non-sunshine hours and extracting other input variables. Also, potential applications of the model to optimize the design and operation of tropical solar energy systems must be sought and demonstrated.

ACKNOWLEDGMENT

The second author thanks the Department of Science and Technology-Science Education Institute (DOST-SEI) and the Engineering Research and Development for Technology (ERDT) consortium for the financial assistance his pursuit of a degree in Master of Science in Mechanical Engineering.

REFERENCES

[1] V. Khare, S. Nema and P. Baredar, "Solar–wind hybrid renewable energy system: A review," *Renewable and Sustainable Energy Reviews,* vol. 58, pp. 23-33, May 2016.

[2] H. Thieblemont, F. Haghighat, R. Ooka and A. Moreau, "Predictive control strategies based on weather forecast in buildings with energy storage system: A review of the state-of-the-art," *Energy and Buildings,* vol. 153, pp. 485-500, 15 October 2017.

[3] U. K. Das, K. S. Tey, M. Seyedmahmoudian, S. Mekhilef, M. Y. I. Idris, W. Van Deventer, B. Horan and A. Stojcevski, "Forecasting of photovoltaic power generation and model optimization: A review," *Renewable and Sustainable Energy Reviews,* vol. 81, no. 1, pp. 912-928, January 2018.

[4] Z. Ramedani, M. Omid, A. Keyhani, S. Shamshirband and B. Khoshnevisan, "Potential of radial basis function based support vector regression for global solar radiation prediction," *Renewable and Sustainable Energy Reviews,* vol. 39, pp. 1005-1011, November 2014.

[5] P. Vrablecová, A. B. Ezzeddine, V. Rozinajová, S. Šárik and A. K. Sangaiah, "Smart grid load forecasting using online support vector regression," *Computers & Electrical Engineering,* vol. 65, pp. 102-117, January 2018.

[6] M. Sengupta, Y. Xie, A. Lopez, A. Habte, G. Maclaurin and J. Shelby, "The National Solar Radiation Data Base (NSRDB)," *Renewable and Sustainable Energy Reviews,* vol. 89, pp. 51-60, June 2018.

[7] A. T. Eseye, J. Zhuang and D. Zheng, "Short-term photovoltaic solar power forecasting using a hybrid Wavelet-PSO-SVM model based on SCADA and Meteorological information," *Renewable Energy,* vol. 118, pp. 357-367, April 2018.

[8] A. Alzahrani, P. Shamsi, C. Dagli and M. Ferdowsi, "Solar Irradiance Forecasting Using Deep Neural Networks," *Procedia Computer Science,* vol. 114, pp. 304-313, 2017.

[9] MathWorks, "Regression learner App," MathWorks, 9 March 2017. [Online]. Available: https://www.mathworks.com/help/stats/regression-learner-app.html. [Accessed 31 May 2019].

[10] B. Yildiz, J. I. Bilbao and A. B. Sproul, "A review and analysis of regression and machine learning models on commercial building electricity load forecasting," *Renewable and Sustainable Energy Reviews,* vol. 73, pp. 1104-1122, June 2017.