PHILIPPINE ENGINEERING JOURNAL
PEJ 2021; Vol. 42, No. 2: 41-56

Received: 11 November 2020
Accepted: 23 October 2021

Short-Term Forecasting Model for Solar PV Power Output using LS-SVM

Russel John C. Gallano*
Electrical and Electronics Engineering Institute, College of Engineering,
University of the Philippines, Diliman, Quezon City, 1101, Philippines
**Corresponding Author: russel.gallano@eee.upd.edu.ph*

Abstract— *The use of renewable energy resources is becoming more prevalent nowadays, especially in distribution systems and microgrids. However, the variability of renewable energy output poses a challenge on the stability and resilience of the power system, particularly in balancing the supply with the load. An output forecast model is useful in this balancing, esp. in scheduling the supply power.*

Solar photovoltaic (PV) systems, commonly used as a distributed generator (DG), has a variable output that depends on external factors, such as temperature, irradiance, cloud cover, and so on. The lack of data about these external factors may hinder the accurate modelling and forecasting of solar PV output. This study attempts to develop a short-term forecast model of the output power of solar PV DGs using only historical solar PV output data. Least-Squares Support Vector Machine (LS-SVM) is used to establish the forecasting model and shows promising accuracy, even when used to forecast fluctuations in solar output.

Keywords— forecasting model, LS-SVM, short term forecast, solar power model, solar PV

I. INTRODUCTION

Electric power systems are continually evolving from a more traditional form to a more distributed, decentralized form. In its decentralized form nowadays, a power system utilizes small and distributed generators, apart from large and centralized power plants, to inject power into the grid to satisfy the power demand. These distributed generators (DGs) have smaller capacity than centralized generators and are directly located within the consumer area. DG technologies include wind, small-scale turbine generators, and solar photovoltaic (PV) [1]. DGs can be used to satisfy the growing demand for electricity but pose a challenge on the grid because of the daily output variability of some DG technologies.

In the Philippines, solar PV is commonly used in the electrification of remote areas and islands that are not connected to the grid, as well as in augmenting the supply power of grid-connected residential, commercial, and industrial loads. According to a study conducted by Solarplaza, a Dutch consultancy firm, the Philippines is the top user of solar photovoltaic systems for electricity generation among the developing Asian countries [2]. This growing contribution of solar PV onto the energy mix of the country, and of the world in general, would require more research into the effects of its daily variability on the power system and on the generation scheduling. Such studies fundamentally need more accurate models of solar PV that can be used to forecast its output.

There are several models of the power output of solar photovoltaic (PV) systems that are based on the various parameters of the PV technology and the environment wherein it is installed. Such models can be used to predict the output of solar PV systems and DGs in the short term (hourly to daily forecast) or medium term (weekly to monthly), providing system operators and planners enough information to forecast the future state of the power system that has solar PV. Short-term solar PV output forecast is particularly important in the hour-to-hour operations of the electricity market, given that the output of solar power generators is prioritized in the dispatch of power. Because of this prioritization over other generation technologies, especially coal- and gas-fired power plants, the solar power output helps dictate the electricity market prices. Furthermore, given that the solar power output is highly variable, it can affect the stability and reliability of the power system. Accurate forecasting of solar power output will help the system operator to allocate enough ancillary services, which would help ensure a stable power grid, and to optimally schedule generators to achieve adequate supply at optimally low prices on the intra- and inter-hour level.

Bulk of the forecasting models for solar power output is based on machine learning algorithms, which can learn through patterns without being explicitly programmed, or on linear regression, which develops forecasting model equations based on historical data. A study that developed a solar output forecast model based on Artificial Neural Network (ANN) used 12 variables for hourly and monthly forecasts. It concluded that solar irradiance, surface irradiance, and net top solar irradiance have the highest correlation with solar power output [3]. A different study used a model based on Nonlinear Autoregressive Neural Network to provide monthly and annual forecasts of solar power output from input parameters such as temperature, solar irradiance, and solar PV array area [4]. Another solar forecast model used a nonlinear autoregressive exogenous model, which predicts a time-series data, like solar power output, based on a set of historical data inputs [5]. The inputs themselves may also be time-series data that can also be predicted through different means.

Support Vector Machine (SVM) is a relatively new supervised machine-learning algorithm applicable to classification or regression problems. It is fast and accurate and can surpass the performance of other machine-learning algorithms. A specific SVM algorithm, least squares SVM (LS-SVM), has previously been used to forecast locational temperature and solar irradiation [6]. These two forecasted time-series data sets can then be used as independent variables in regression models of solar PV output.

Support Vector Machines employ what is called a kernel function in order to map and project the dataset to a higher dimension to obtain a better interpretation of the datapoints and to highlight underlying clusters and/or patterns. There are different types of kernel functions that can be used based on the intended application (e.g., clustering, categorization, pattern identification, etc.) and the type of dataset to be used. The commonly used SVM kernel functions include linear, polynomial, radial basis functions (RBF), and sigmoid. Linear kernel is the most basic type used for classification problems containing many features like text classification. Polynomial kernel is the least efficient and least accurate of all the aforementioned kernel functions when looking across different applications. Sigmoid kernel is more preferred for neural networks as it works similar to two-layer perception model of the neural network [7].

RBF is the most preferred and used kernel function for its easy design, good generalization from limited training data, and strong tolerance to input noise. Despite its complexity growing as more features are being explored, it is flexible as it can make a proper separation when there is little to no prior knowledge of the data (whether linear, nonlinear, radial, etc.), as seen in Figure 1.

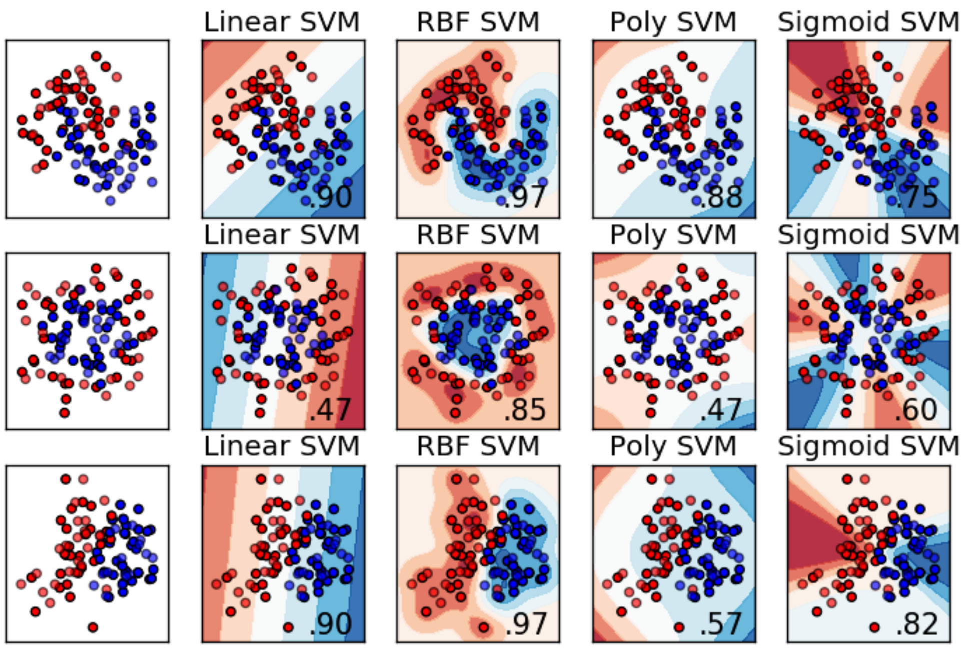

Figure 1. Comparison of the accuracies of commonly used SVM kernel functions in identifying relationships existing among datapoints [8]

This study aims to create a forecasting model using LS-SVM that can directly predict future solar PV power output, in kilowatt (kW), based solely on the historical solar PV power production. This method eliminates the need for several differing sets of data–temperature, irradiance, weather, etc.–that may not necessarily be compatible with one another in terms of temporal resolution, measurement accuracy, and dataset completeness. The use of the RBF kernel function, for its flexibility and high accuracy even with limited prior knowledge of the dataset, is suited for this study given the attempt of using only a single input: historical solar PV power output.

II. METHODOLOGY

The study has two parts: (i) modeling of the solar PV output through LS-SVM training, and (ii) solar PV power output forecast validation. For the first part, the LS-SVM will be trained using part of an obtained dataset to be able to forecast solar PV output forecasts with minimal error from the actual data. This will serve as the solar PV output forecasting model. The forecasting performance of resulting model will then be validated against a smaller part of the dataset. Metrics are used to quantify how accurate the forecasting model is.

2.1 LS-SVM Modeling

LS-SVM is a variation of SVM developed by Suykens et.al. that uses equality constraints instead of inequality constraints and a least square error term instead of the standard error term [7]. The primal cost function of LS-SVM is minimized and considers the following optimization problem in primal weight space:

$$\min \left\{ J(\omega, \xi) = \frac{1}{2}\omega^T\omega + \frac{1}{2}\gamma \sum_{i=1}^{N} \xi^2 \right\} \tag{1}$$

subject to:

$$y_i = \omega^T \varphi(x_i) + b + \xi_i \qquad i = 1,2, \dots, N \tag{2}$$

With $\varphi(x_i)$ as the kernel function which maps the input space to a higher dimension such that the data is linear in the higher dimensional feature space. The cost function J consists of a sum squared fitting error and a regularization term. γ represents the regularization parameter while b is the bias term. The primal weight space is of the form:

$$y = \omega^T \varphi(x) + b \tag{3}$$

The weight vector ω can have infinite dimensions, which makes the calculation of the weight vector from equation 3 nearly impossible. A solution to this problem is transforming and solving it in dual form instead of in primal space, defining the Lagrangian:

$$L(\omega, b, \xi, \alpha) = \frac{1}{2}\omega^T\omega + \frac{1}{2}\gamma \sum_{i=1}^{N} \xi^2 - \sum_{i=1}^{N} \alpha_i [\omega\varphi(x_i) + b + \xi_i - y_i] \tag{4}$$

The conditions for optimal solutions are given as:

$$\frac{\partial L}{\partial \omega} = 0 \rightarrow \omega = \sum_{i=1}^{N} \alpha_i \varphi(x_i) \tag{5}$$

$$\frac{\partial L}{\partial \alpha_i} = 0 \rightarrow y_i = \omega\varphi(x_i) + b + \xi_i, i = 1, \dots \dots, N \tag{6}$$

$$\frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^{N} \alpha_i = 0 \tag{7}$$

$$\frac{\partial L}{\partial \xi_i} = 0 \rightarrow \alpha_i = \gamma\xi_i, i = 1, \dots \dots, N \tag{8}$$

After eliminating ω and ξ, the equations in matrix form are now obtained as

$$\begin{bmatrix} 0 & 1^T \\ 1 & \Omega + \frac{I}{\gamma} \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \tag{9}$$

where $\mathrm{y} = [\mathrm{y}_1, \ldots, \mathrm{y}_\mathrm{N}]$, $1 = [1, \ldots, 1]$, $\alpha = [\alpha_1, \ldots, \alpha_\mathrm{N}]$, and $\Omega =$ kernel function.

The kernel function to be used will be the Radial Basis Function which is expressed by:

$$K(x, y) = \exp\left(\frac{|x - x_i|^2}{2\sigma^2}\right) \tag{10}$$

2.2 Training of LS-SVM Model

Before training the data, the time-series data–in this case the historical solar PV power production–into a Hankel matrix, which is useful for training nonlinear function approximation. Assume that there is a matrix A as defined by equation 11.

$$A = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \\ d_1 & d_2 & d_3 \\ e_1 & e_2 & e_3 \\ f_1 & f_2 & f_3 \\ g_1 & g_2 & g_3 \end{bmatrix} \tag{11}$$

Depending on the number of rows chosen, matrix A will be transformed into matrix B. For example, if the chosen number is 3, row 1 to row 3 from matrix A will be the values for the first row of matrix B. Rows 2 to 4 of matrix A will be the values for the second row of matrix B. Continuing until reaching the final row of matrix A, matrix B will look like:

$$B = \begin{bmatrix} a_1 & a_2 & a_3 & b_1 & b_2 & b_3 & c_1 & c_2 & c_3 \\ b_1 & b_2 & b_3 & c_1 & c_2 & c_3 & d_1 & d_2 & d_3 \\ c_1 & c_2 & c_3 & d_1 & d_2 & d_3 & e_1 & e_2 & e_3 \\ d_1 & d_2 & d_3 & e_1 & e_2 & e_3 & f_1 & f_2 & f_3 \\ e_1 & e_2 & e_3 & f_1 & f_2 & f_3 & g_1 & g_2 & g_3 \end{bmatrix} \tag{12}$$

2.3 Data Set

The LS-SVM forecasting model is built using training data samples. One part of the input data shall be used as training set and another part as testing or validation set to test the model's accuracy in forecasting future values. Historical data used in this study came from the website Renewables.ninja. It collects weather data from the Modern-Era Retrospective analysis for Research Applications (MERRA) data set of the US National Aeronautics and Space Administration (NASA) [9]. Using that dataset, Renewables.ninja runs simulations to estimate hourly power output of solar PV panels located anywhere in the world. The simulation takes into account the following configuration: medium-size commercial panel, optimum panel placement, and 80kW peak capacity.

From this simulation, only a few weeks' worth of hourly solar PV power data was extracted to serve as the input data set, which will then be transformed into a Hankel matrix. Given the small dataset extracted, the matrix is divided into two equal partitions–one half for training, the other for testing. This 50/50 split for training and testing has been found to enough, esp. for cases that use small datasets, to achieve a desirable level of correct classification without leading the model to over-fitting [10].

2.4 Tuning

In the LS-SVM model, the regularization constant, gamma (γ), and the kernel parameter, σ^2, determine the trade-off between training error minimization and smoothness of the model. The kernel type chosen for this study is the Radial Basis Function kernel. The simplex optimization function is used. Simplex finds a local minimum function via the Nelder-Mead algorithm.

The tuning of the parameters is determined by using a state-of-the-art global optimization technique called Couple Simulated Annealing (CSA) where it determines suitable parameters according to some criterion. Afterwards, the parameters are provided to the simplex optimization technique for fine tuning.

2.5 Validation

The values produced by the model are then validated by using the Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE). The RMSE is computed using the following:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2} \tag{13}$$

where Y_i is the actual or observed value and $\hat{Y}_i$ is the forecasted value for the time *i* = 1 to N.

RMSE is a widely used method in measuring differences between forecast values, in which lower RMSE value indicates a better fit. It is considered to be the most important criterion for fit if the main purpose of the model is forecast. Given it squares the deviations, it prevents positive and negative deviations from cancelling out one another and exaggerates large errors.

The MAE metric, sometimes referred to as Mean Average Deviation (MAD), primarily provides the average distance between the forecast and actual data. This provides an insight on how well the model fits the forecast into the pattern of the actuals. MAE is expressed as

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|Y_i - \hat{Y}_i| \tag{14}$$

III. RESULTS AND ANALYSIS

In order to have a levelled comparison of the observed data and the forecasted data, as well as simplify the magnitude of the data presentation, they have been normalized to 1. Two forecasting models were developed using LS-SVM: day-ahead forecast model and week-ahead forecast model.

3.1 Data Handling

Renewables.ninja was used to simulate solar generation in the Metro Manila area to get hourly solar PV output power that will be treated as observed data. The simulation accounts for the following configuration: medium-size commercial solar PV panel, optimum panel placement, and 80kW peak capacity. It also considers an absence of sunlight from 6PM to 5AM every day. Only solar irradiation from NASA's MERRA were considered by the simulator. The set of data points from March was considered for this study as the solar power output reached peak value during this month.

For the day-ahead forecasting model, the input data used yielded a 144 x 25 Hankel matrix with a lag value of 24. The lag value is essentially the number of inputs that are considered to forecast a future value. Note that there is no definitive way to determine the appropriate lag value for an application. It is usually set through trial-and-error.

The chosen lag value for the day-ahead model conveniently yields two 72 x 25 matrices–one for training and the other for testing. The partitions also represent three days' worth of data individually, i.e., 72 hours. This set of 72 datapoints will be the input to the LS-SVM forecasting model for the day-ahead forecast tests.

For the week-ahead forecast model, the data set covered 360 points. Similar to the day-ahead forecast model, a lag value of 24 was used. This value corresponds to 24 hours prior to the day being predicted. For a short-term forecast model, this lag value may be enough.

When the input data was transformed, the resulting Hankel matrix had a size of 336 x 25, which can be divided into two 168 x 25 matrices for training and testing sets. The resulting sets each have 7 days' worth of hourly data points. This resulting dataset containing a total of 168 hours of data (7 days) will be the input to the forecasting model for the week-ahead testing phase.

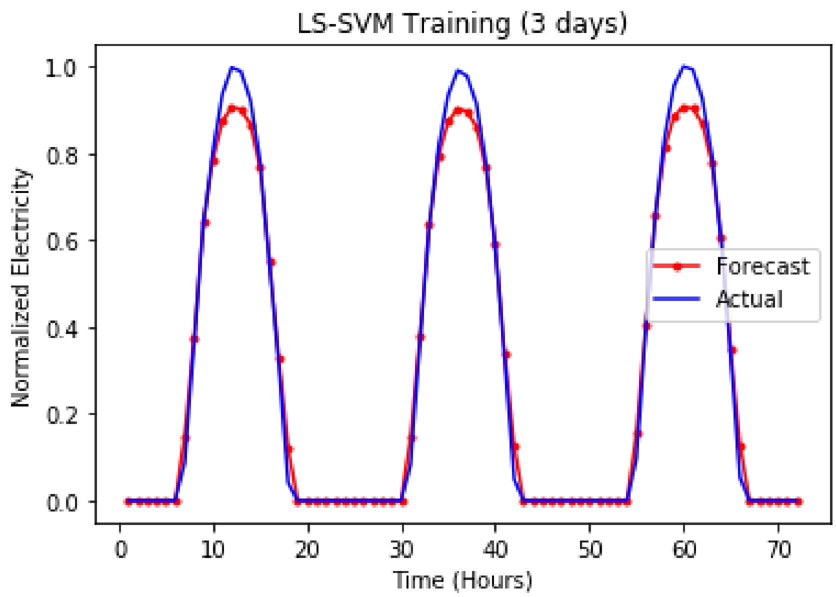

Figure 2. Plot of actual and predicted data in the training phase for day-ahead forecast model

3.2 Day-Ahead Forecast Model

The results of the training phase for the day-ahead forecast model are shown in Figure 2. Note that the values in the plots are already normalized.

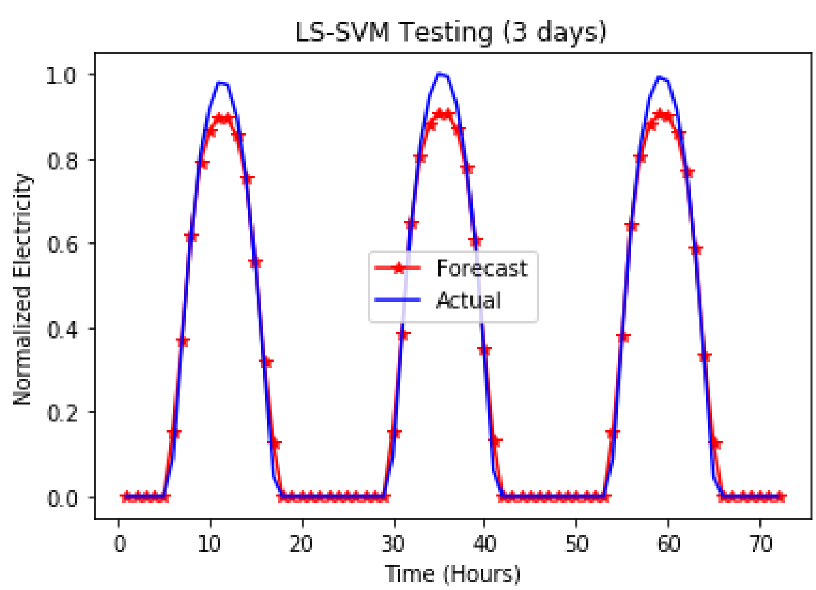

Figure 3. Plot of actual and predicted data for the testing phase for day-ahead forecast model

The blue curve represents the actual data set used for training, whereas the orange curve represents the predicted data through the supervised learning method. It should be noted that on the third day, the LS-SVM had its largest discrepancy from the observed data. This may be due to the seemingly declining trend of the peak during the first two days. Still, it is evident that the training results show that the LS-SVM forecast model was trained enough to grasp the trend of the input data. This observation is further supported quantitatively by the accuracy metrics, as stated in Table 1. This is quite a significant result because the metrics are leaning close to zero, which implies that high model accuracy was developed during the training phase.

As for the validation of the accuracy of the model, it was used to forecast three consecutive days in the testing phase. The result of this forecast was then compared to the testing data set for this test case. It can be seen from Figure 3 that the model follows the regularly oscillating trend and is accurate in forecasting the level of power output. The quantitative results in Table 1 supports this observation. It is important to remember that the values being compared are normalized and so the performance metrics are scaled to that.

Table 1. Daily Solar LS-SVM forecasting model results for a three-day duration

	Training Phase	Testing Phase
MAE	0.02094	0.020497
RMSE	0.03742	0.03656

For the testing phase results, the model accurately forecasted the solar power output for the three consecutive days. In fact, the model performed better in the testing phase than in the training phase. This can be attributed to the regularity in the periodic trends of the input data that the LS-SVM model was able to follow.

3.3 Week-Ahead Forecast Model

For the week-ahead forecast model, the same methodology was applied to a different, larger input dataset. The results of the training phase are shown in Fig. 4. As seen from the figure, the model was able to follow the trend of the actual data and has produced magnitudes that are close to the actual.

The week-ahead forecast model was then run to predict the solar power generation for the succeeding seven days, the results of which were compared to the actual production data. Figure 5 shows the results of this testing phase. When focusing on the plot, it can be noticed that the forecasts had bigger errors around the daily peaks, which would have increased the error of the forecasts. However, at times other than the peak, the forecast curve tightly follows the actual data. Because of this characteristic, the model is highly accurate. This is supported by the quantitative results indicated in Table 2.

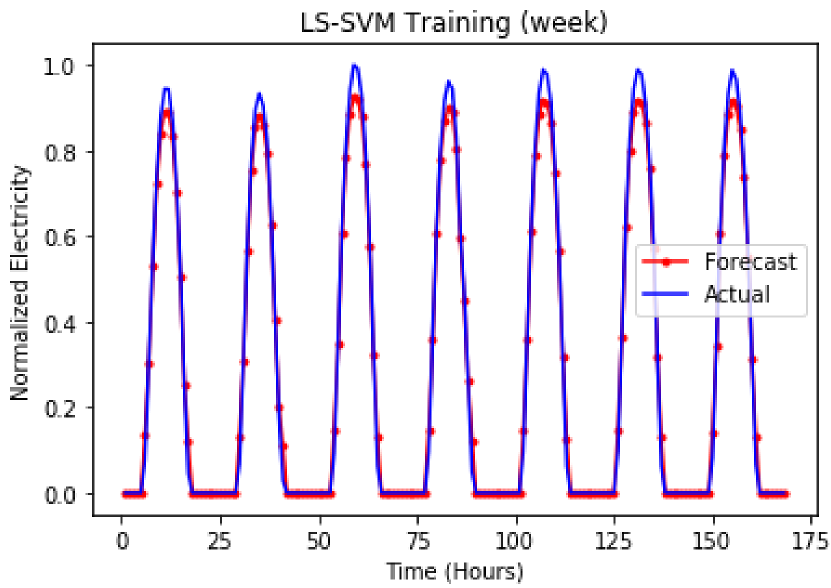

Figure 4. Plot of actual and predicted data for the training phase for week-ahead forecast model

There are two things that should be highlighted regarding these results for the day- and week-ahead models. First, the results show that only a small dataset is needed to accurately model solar power output with regularity. Second, this finding, however, is only limited to regular, periodic trends in historical and future data.

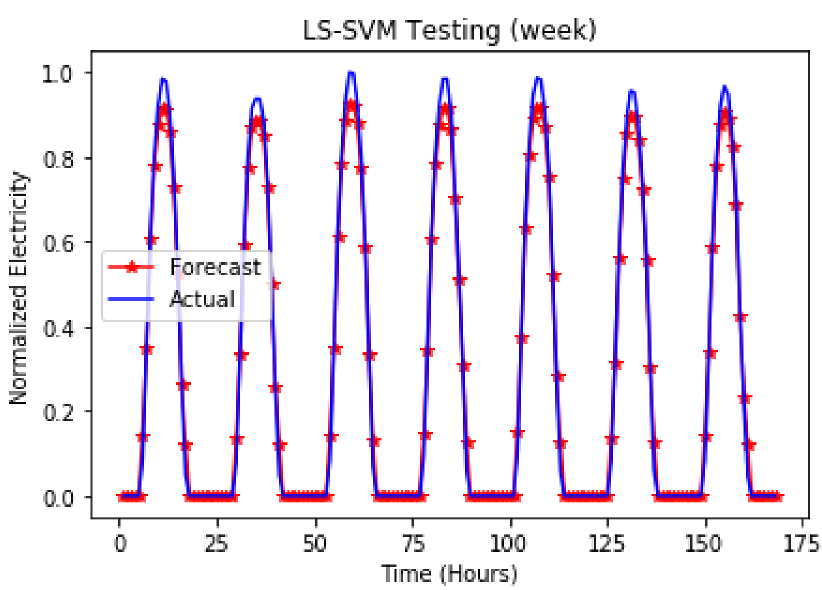

Figure 5. Plot of actual and predicted data for the testing phase for week-ahead forecast model

In a more realistic sense, the solar power output has more irregularity and intermittency throughout the day because of a lot of factors, such as cloud cover, weather changes, presence of shade, and so on. If some of those factors were included in modelling with a small dataset, the system would most likely provide forecasts with high variance and possibly lower accuracies.

Table 2. Summary of Week-ahead Solar LS-SVM forecasting model results

	Training Phase	Testing Phase
MAE	0.01798	0.01814
RMSE	0.03020	0.03043

3.4 Robustness Analysis

To test the robustness of the model in relation to the intermittency of the solar power generation, a robustness test case was implemented. In this test, the day- and week-ahead forecast models were validated against a test dataset that was incorporated with random fluctuations. These random fluctuations were also normalized to 1 and were subtracted from the dataset. Given that it is possible to have negative differences, the absolute value of the resulting difference was used as the testing data. Figure 6 shows sample plot of the daily solar output power with fluctuations.

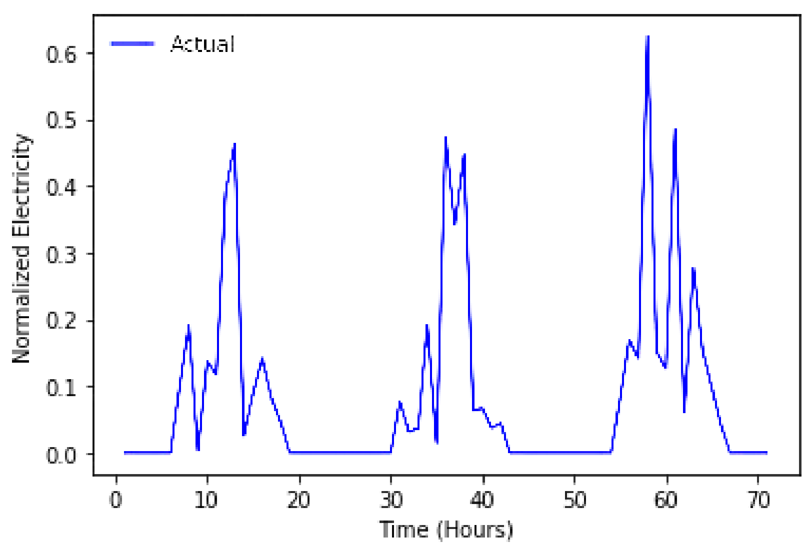

Figure 6. Sample plot of three-day solar output power data incorporated with fluctuations

The objective of this process is to mimic the intermittency of the solar power output caused by factors like cloud cover and temperature that were not reflected in the input data from Renewables.ninja. If the model is robust enough, it should be able to forecast the fluctuations with an accuracy that is comparable to that of when the data is smooth and periodic, as in the previous tests.

Table 3. Day- and week-ahead mean and standard deviation of the RMSE and MAE in robustness test

Metric		Day-Ahead Model	Week-Ahead Model
RMSE	average	0.04395	0.04186
	st. dev.	0.00305	0.00240
MAE	average	0.02926	0.02635
	st. dev.	0.00261	0.00232

Given that this test involves random values, the generation of datasets and the testing phase of the models were iterated for ten times. The RMSE and MAE metrics from all iterations were then averaged, as indicated in Table 3.

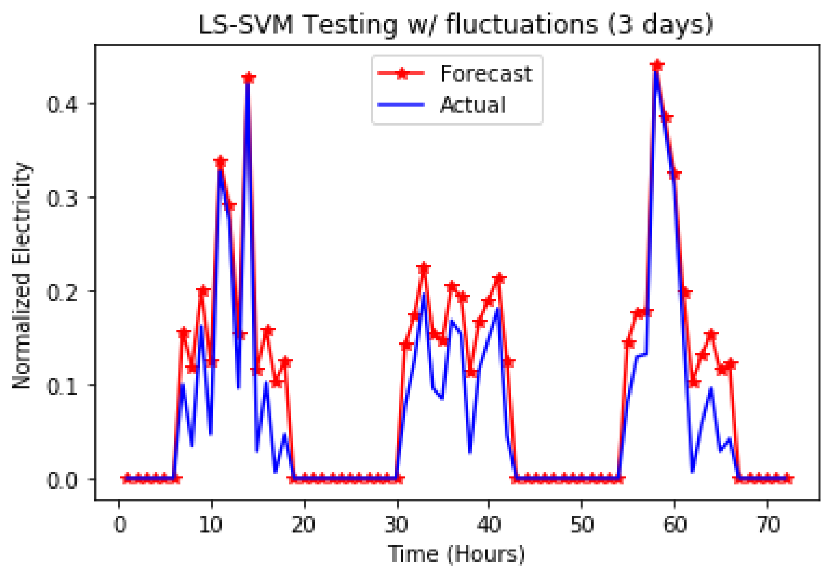

Figure 7. Plot of actual vs. forecasted values for day-ahead validation with fluctuations

As hypothesized, the performance metrics of both models increased when applied to the robustness test because the models were developed with smooth, periodic training data. But based on the average values of the performance metrics of the day- and week-ahead models, both models still perform quite well despite the fluctuations in the testing dataset. In addition, the standard deviation of the performance metrics across several iterations shows that there is small variance in the accuracy of the model. This shows that the models are robust enough to handle sudden changes in the historical data and still continue giving accurate forecasts. Fig. 6-7 show the comparison plots for a sample iteration run for each model.

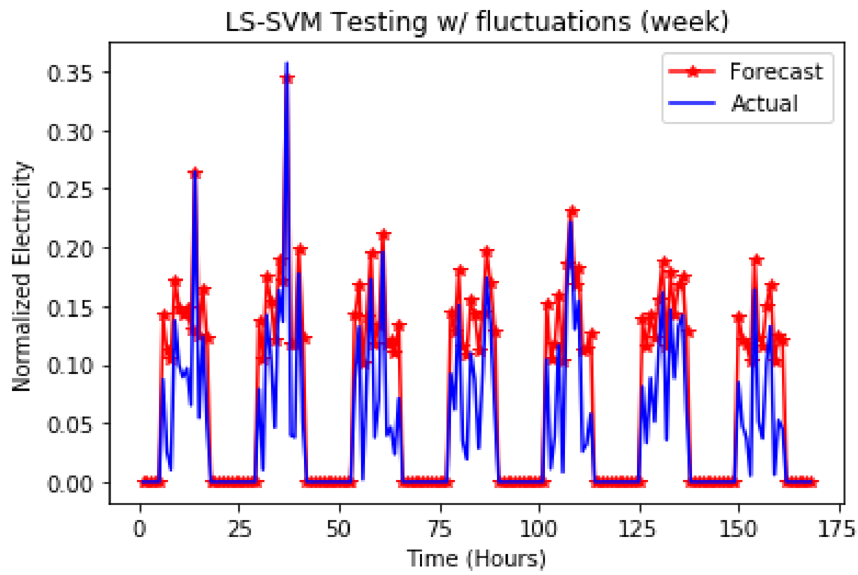

Figure 8. Plot of actual vs. forecasted values for week-ahead model validation with fluctuations

3.5 Peak Value Forecast

It can be noticed that the results of the forecasting model in the base case day- and week-ahead tests have higher deviations as the peak hours of solar power production is reached. This shows that the forecast model can benefit from: (i) a longer set of input, i.e., wider window of historical solar power output to be transformed into a Hankel matrix; and (ii) its application to a timeframe with more discernible variations. Note that the inaccuracy during peak hours is higher for the day-ahead forecast test than the week-ahead test. This is because using the LS-SVM model to forecast inter-day solar output would extract less features from the limited input of 72 datapoints, unlike the use of the model for week-ahead forecasting wherein more datapoints in the input are available to potentially yield more features.

Looking at the robustness analysis, it can be seen that although the model is similarly less accurate during peak hours, the model can easily follow the trend of the highly variable benchmark data. This is an important feature of an intermittent energy resource, like solar PV, that forecasting models should be able to discern and predict to provide reliable projections.

IV. CONCLUSION

In this study, short-term solar PV output forecasting models were created for day- and week-ahead forecasts using LS-SVM. The models were developed by using historical data of the generated solar PV power, which reduces the required data for modelling. Both models show promising accuracy in forecasting the following day's or following week's solar PV power output, as evidenced by the computed RMSE and MAE values. This can be attributed to the input data for training and testing being smooth and periodic.

The results of this robustness test show that the models are robust enough against variability in the input despite being trained with smooth, periodic data. This presents an advantage over traditional regression models as those types of models typically experience increasing deviations from the actual trend once exposed to data variability. Furthermore, there was no need to retrain the LS-SVM models to account for fluctuations in the data. Regression models might require recalibration to handle such fluctuations, esp. when such models were developed using smooth, periodic data.

Given they were trained using data from March, the models should be tested against data that correspond to months or weeks that receive reduced solar irradiation. An improvement of the model is to increase its accuracy, especially for the day-ahead forecast, by training it with more training datasets with varying levels of peak solar output values. Despite performing quite well in the robustness tests, the models can further be developed to use additional training data that account for the intermittency of received solar radiation due to cloud cover and other factors.

NOMENCLATURE

Symbol	Description	Units
A	dataset matrix	[-]
α	Lagrange multiplier	[-]
B	Hankel-transformed dataset matrix	[-]
b	bias term	[-]
γ	regularization parameter	[-]
$J(\omega,\xi)$	primal cost function	[-]
K	RBF	[-]
$L(\omega,b,\xi,\alpha)$	Lagrangian function	[-]
ξ	slack variable	[-]
σ	variance	[-]
$\varphi(x)$	kernel function	[-]
ω	weight vector	[-]
Ω	kernel function (Lagrangian)	[-]
x	input data	[-]
y	classification function	[-]
Y_i	actual or observed value	[-]
$\hat{Y}_i$	forecasted value	[-]

Subscripts, Superscripts, and Abbreviations

exp abbreviation refers to exponential function
i subscripts refer to a data point
MAE refers to mean average error
RMSE refers to root-mean-square error
T superscripts refer to matrix transposition

REFERENCES

[1] Barker P, Mello R. 2000. Determining the impact of distributed generation on power systems: Part 1 - Radial distribution systems. Proceedings of the IEEE Power Engineering Society Transmission and Distribution Conference. pp. 1645 – 1656. IEEE. doi: 10.1109/PESS.2000.868775.

[2] Domingo R. 2018. PH 5th worldwide, top in Asia in solar power, Inquirer.net Retrieved from https://business.inquirer.net/243407/ph-5th-worldwide-top-asia-solar-power.

[3] Abuella R, Chowdhury B. 2015. Solar power forecasting using artificial neural networks. 2015 North American Power Symposium (NAPS); Charlotte, USA. IEEE. pp. 1-5. doi: 10.1109/NAPS.2015.7335176.

[4] Bimenyimana S, Norense G, Asemota O, Lingling L, Li L. 2017. Output power prediction of photovoltaic module using nonlinear autoregressive neural network. Journal of Energy, Environmental & Chemical Engineering. 2: 32-40. doi: 10.11648/j.jeece.20170204.11.

[5] Boussaada Z, Curea O, Ahmed R, Camblong H, Najiba M. 2018. A nonlinear autoregressive exogenous (NARX) neural network model for the prediction of the daily direct solar radiation. Energies. 11: 620. doi: 10.3390/en11030620.

[6] Potot J. 2019. Nonlinear modeling of solar power parameters using support vector machine for short- and medium-term solar power forecasting. Quezon City: University of the Philippines Diliman.

[7] Suykens J, van Gestel T, de Brabanter J, de Moora B, Vandewalle J. 2002. Least- squares support vector machines. Singapore: World Scientifics. p. 29-147.

[8] Wittman F (GitHub Gist). 2016. Retrieved from https://gist.github.com/WittmannF/60680723ed8dd0cb993051a7448f7805 on Jan 2021.

[9] Renewables.ninja. 2019. Retrieved from https://www.renewables.ninja/about on 31 July 2020.

[10] Xu Y, Goodacre R. 2018. On splitting training and validation set: a comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. Journal of Analysis and Test. 2: 249–262. doi: https://doi.org/10.1007/s41664-018-0068-2

