Validation of Genetic Algorithm Optimized Hidden Markov Model for Short-term Photovoltaic Power Prediction

Victor Eniola*,**, Tawat Suriwong*‡, Chatchai Sirisamphanwong***,
Kasamsuk Ungchittrakool****,*****, Olatubosun Fasipe**,******

*School of Renewable Energy and Smart Grid Technology, Naresuan University, Phitsanulok 65000, Thailand

**Research and Consultancy Department, Energy Commission of Nigeria, Abuja 900211, Nigeria

***Department of Physics, Faculty of Science, Naresuan University, Phitsanulok 65000, Thailand

****Department of Mathematics, Faculty of Science, Naresuan University, Phitsanulok 65000, Thailand

*****Research Centre for Academic Excellence in Nonlinear Analysis and Optimization, Faculty of Science, Naresuan University, Phitsanulok 65000, Thailand

******Department of Civil Engineering, Faculty of Engineering, University of Benin, Benin City 300283, Nigeria

(enilav01@yahoo.com, tawats@nu.ac.th, chatchaisi@nu.ac.th, kasamsuku@nu.ac.th, fasipeo@gmail.com)

‡ Corresponding Author; Tawat Suriwong, School of Renewable Energy and Smart Grid Technology, Naresuan University, Phitsanulok 65000, Thailand, Tel: +66 5596 3180,

Fax: +066 5596 3182, tawats@nu.ac.th

Received: 18.04.2021 Accepted: 26.05.2021

Abstract- A substantial amount of renewable energy (RE)-based electrical power is generated over the last ten years due to global warming issues. Solar photovoltaic (PV) is being incredibly utilized because of its boundless quality. However, the inherent intermittency of PV power production at high penetration level to the grid leads to complications related grid reliability, stability and transportable unit of electric power. A viable approach to addressing this problem is to develop a reliable power forecast model for the short-term horizon related to scheduling and transmission. Based on an existing forecast model built on genetic algorithm (GA)-optimized hidden Markov model (HMM), this paper implements the model validation process using more recent input dataset. Model evaluation is based on the computation of normalized root mean square error (nRMSE). As the validation result, HMM+GA is sufficient to accurately forecast PV P_o under clear sky day (CSD) condition. Contrariwise, for cloudy days (CDs) presenting instantaneous changes in solar irradiance (G_s) between some hours of the day, HMM+GA adapted with a correction factor (ξ); articulated as HMM+GA+ξ; is adequate to forecast the P_o more precisely when the average change in the absolute value of G_s ($\left|\overline{\Delta G_s}\right|$) in the morning ($\left|\overline{\Delta G_s}\right|_m$) is greater than 128% and/or when $\left|\overline{\Delta G_s}\right|$ in the evening ($\left|\overline{\Delta G_s}\right|_e$) exceeds 90%. Particularly, the average nRMSE of 2.63% showed that HMM+GA with or without ξ are suitable techniques for forecasting PV P_o on an hourly basis. Therefore, the validation results are in harmony with those of the baseline models.

Keywords Prediction, photovoltaic, power production, short-term forecasting, validation.

Nomenclature

Acronym		*Symbol*	
ANNs	artificial neural networks	$\lvert\overline{\Delta G_s}\rvert$	average percentage change in absolute value of solar irradiance
CDs	cloudy days		
CSD	clear sky day	$\lvert\overline{\Delta G_s}\rvert_m$	average percentage change in absolute value of solar irradiance in the morning
DMoC	data monitoring and operation centre		
ELM	extreme learning machine	$\lvert\overline{\Delta G_s}\rvert_e$	average percentage change in absolute value of solar irradiance in the evening
EMA	expectation-maximization algorithm		
FFNN	feed-forward neural network	ξ	correction factor
GA	genetic algorithm	η	module efficiency
GPR	Gaussian process regression	α	temperature coefficient (power)
HMM	hidden Markov model	ξ_e	correction factor at evening
kW	kilowatts	ξ_m	correction factor at morning
MAPE	mean absolute percentage error	A_m	module area
MRE	mean relative error	G_s	solar irradiance
MW	megawatts	h_u	humidity
NBC	naïve Bayes classifier	i_i	illumination index
nRMSE	normalized root mean square error	n	number of observations
$nRMSE_{HMM}$	nRMSE of HMM	P_a	actual power
$nRMSE_{opt}$	nRMSE of optimized model	P_{act}	actual power output
PSO	particle swarm optimization	P_f	forecasted power
PV	photovoltaic	P_{HMM}	HMM power output
RBF	radial basis function	P_o	power output
RE	renewable energy	P_{opt}	optimized power output
SCADA	supervisory control and data acquisition	P_{rated}	rated power of PV system
STC	standard test condition	T_{amb}	ambient temperature
SVR	support vector machine	T_m	module temperature
VA	Viterbi algorithm	w	wind speed
		w_d	wind direction

1. Introduction

Because of fossil fuel depletion and climate issues, many incentives and energy regulations capable of advancing renewable energy (RE) deployment have been orchestrated in many countries. It is feasible to operate a 100% RE-based electric power grid [1]. Among the RE sources, solar photovoltaic (PV) can complement the conventional systems operating on fossil fuels. PV is incredibly utilized in locations with good solar resource because of its boundless quality and scalability. Additionally, PV systems are gaining popularity, considering their economic and environmental benefits [2]. With the falling prices of PV modules, it is projected that the PV power supply to modern electric power would increase further. However, the PV technology is confronted with some technical hitches predominantly at a high level of penetration where discontinuity is pronounced. Fluctuations in solar radiation received by PV panels is chiefly responsible for the unpredictability of PV power output [3]. This inherent unpredictability of PV power at higher level of penetration to the grid gives complications relating to a transportable unit of electric power and grid reliability in general [4]. It is one reason in developed countries why a high unit of electrical power is not allowed to be injected into the grid from RE sources. A viable approach to solving this problem is to develop a reliable power forecast model for the short-term horizon related to dispatching plan, scheduling and transmission [5-7]. Short-term PV power output forecasting benefits include improvement in grid security, enhancement of power system control, and determinable energy pricing in advance. With sound forecasting models, customers' dissatisfactions arising from power quality issues can be addressed.

PV power output prediction has been implemented using a number of techniques such as artificial neural networks (ANNs) [8], Gaussian process regression (GPR) [9], support vector regression (SVR) [10], extreme learning machine (ELM) [11], cloud modelling [12], Grey theory [13], random forests [6], naïve Bayes classifier (NBC) [14], hybrid approaches [15-17], and Markov processes [18]. Intending to achieve more reliable forecasts, Eniola et al., 2019 [18] built a genetic algorithm (GA)-optimized hidden Markov model (HMM)-based forecasting tool for hour-ahead prediction of the power output of a 1.2kW PV system installed at the School of Renewable Energy and Smart Grid Technology (SGtech), Naresuan University, Phitsanulok, Thailand. To further consider the dependability of the model, this study implements the validation of the forecast models using a more recent input dataset acquired from the PV supervisory control and data acquisition (SCADA) system.

2. Prediction Model Development

With six months of historical data comprising of ambient temperature (T_{amb}), wind speed (w), and solar irradiance (G_s) as inputs, the power output of a 1.2 kW PV system is forecasted. The PV system is installed at the Energy Park, SGtech, Naresuan University (Lat. 16°47′ N, Long. 100°16′ E), Thailand. It consists of twelve thin-film silicon modules manufactured by Kaneka Corporation in Japan. The system is equipped with SCADA and a monitoring device located at the data monitoring and operation centre (DMoC) from which historical and real-time data can be acquired. Figure 1 represents the system's equivalent circuit diagram and the electrical parameters and other information of the PV modules are presented in Table 1. It should be noted that electrical data is at standard test condition (STC): G_s 1000 W/m^2, spectrum air mass 1.5 and cell temperature 25 °C.

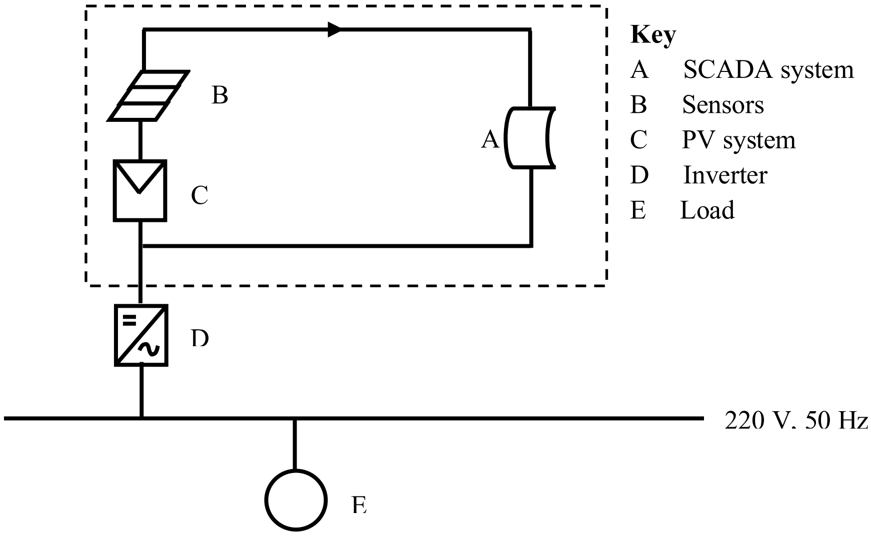

Fig. 1. Equivalent circuit diagram of the 1.2 kW PV system.

Table 1. Electrical specification of PV modules and other system information

S/N	Parameter	Value	Unit
1.	Array rating	1.2	kW
2.	Panel rating	100	W
3.	Number of panels	12	-
4.	Panel model	U-EA 100	-
5.	PV technology	Thin-film m-Si	-
6.	Maximum power (P_{max})	100	W
7.	Minimum value of P_{max}	95.0	W
8.	Open circuit voltage (V_{oc})	71.0	V
9.	Short circuit current (I_{sc})	2.25	A
10.	Voltage at P_{max} (V_{mpp})	53.5	V
11.	Current at P_{max} (I_{mpp})	1.87	A
12.	Temperature coefficient (power)	-0.35	%/K
13.	module efficiency (η)	8.2	%
14.	Dimension ($W \times L \times T$)	1210×1008×40	mm
15.	Array area (A_m)	14.64	m^2
16.	Module manufacturer	Kaneka corporation, Japan	-
17.	Inverter size/type	1×2.5, Leonics/Apollo G-303	kW
18.	Tracker	Nil	-
19.	Array tilt/Azimuth	Fixed, 17°/0	-

Before designing the forecast model, the forecast data is filtered to 1-hour timestamp followed by data refinement. Its purpose is to compensate for negative or missing data points by substituting them with their monthly average values. The error metrics used in this study for the forecast model performance evaluation necessitate preprocessing the dataset to exclude zero-value data occurring at night and early hours to avoid absurd error values at the validation phase. Afterwards, the entire dataset is partitioned into two quotas. The forecast model is trained using 95% of the data and the remainder is employed for model testing. In very short-term PV P_o forecasting, G_s and module temperature (T_m) are the best variables to accurately forecast swift PV energy variations due to T_m significant effect on voltage which consequently affects PV P_o [12, 19]. In this study, T_{amb} is used to compute T_m based on the mathematical transformation model expressed in Eq. (1) [12, 20].

$$T_m = 0.943T_{amb} + 0.028G_s - 1.528w + 4.3 \quad (1)$$

where T_m and T_{amb} are module and ambient temperature respectively, measured in °C, G_s is solar irradiance in W/m^2 and w is the wind speed in m/s.

Estimating parameters and training the Po forecast model using HMM requires determining the likelihood of observation sequence, predicting the next emission in the sequence of observations, and finding the most probable underlying explanation of the observation sequence. The problems aforementioned can be addressed using the forward-backward algorithm, Viterbi algorithm (VA) and the Baum-Welch algorithm, sometimes referred to as expectation-maximization algorithm (EMA) [21, 22]. As parameter categorization is required for the HMM-based forecast model design, G_s data is classified according to some rules. Table 2 exhibits the conventions adopted in categorizing G_s and the procedure resulted in five states representing the very clear sky, clear sky, partial cloud, cloudy and very cloudy. The emissions are then classed into three distinct levels: high, moderate, and low generations. HMM latent variables express to Markov chain and are discrete in nature. The next step is to equate inputs to observations and outputs to states so that the forecast model can learn from the output-input relationship. This process is known as supervised learning, and forecasts can be made based on models of observed data. The training dataset is sequenced to predict with the HMM, and the model parameters and the transition matrix are estimated. Notably, the computation of the transition matrix is carried out with the HMM simulation tool. After all input data have been preprocessed, the predefined state and emission are characterized according to the previous conventions. The learning process of the forecast model requires representing state and emission by G_s and historical P_o of PV, respectively. The next step is state and emission sequencing. With programming codes explicit to the HMM training process in our simulation tool and an optimal number of iterations of the EMA specified, each transition matrix element can be estimated. In this study, after specifying 500 iterations of the EMA in training the forecast model; the state probability distribution matrix A is as given below:

$$A = \begin{bmatrix} 0.514 & 0 & 0 & 0.470 & 0.016 \\ 0.487 & 0.500 & 0 & 0.013 & 0 \\ 0 & 0.569 & 0.066 & 0 & 0.365 \\ 0 & 0 & 0.504 & 0.402 & 0.094 \\ 0 & 0.038 & 0.024 & 0.376 & 0.562 \end{bmatrix}$$

Table 2. Classification of G_s

G_s	Class	State
> 800	very clear sky	5
≤ 800	clear sky	4
≤ 600	partial cloud	3
≤ 400	cloudy	2
≤ 200	very cloudy	1

Matrix A is of order 5-by-5 as there are five discrete states. Element a_{ij} denotes the probability distribution of transitioning from state i to j. Thus, $a_{ij} \geq 0$ and $\sum_i^N a_{ij} = 1$ for $0 \leq i \leq 1$. The highest probable state sequence that is utilized to make the next hour P_o prediction is given by Viterbi deciphering. To make predictions, the model deploys the power formula expressed in Eq. (2) [23]. To obtain P_o at hour t+1, T_m and G_s at hour t are passed as inputs unto the forecast model. As determined in the HMM-based Po forecasting steps described in Fig. 2, the Po forecasting is carried out using the HMM toolbox in our simulation software.

$$P_o = \eta A_m G_s \left[1 - \alpha(T_m - 25)\right] \quad (2)$$

where P_o is the power output in kW, η is the module efficiency in %, A_m is module area in m^2, and α is the temperature coefficient (power) measured in %/K. The optimization of parameters and forecast model enhancement is built on GA. All input variables are initialized, and the fitness function is created. The fitness function is expressed as the sum of squares of the difference between fitted values and actual P_o (P_{act}). The GA-based optimization process entails passing a function handle to the fitness function alongside the number of variables in the problem. To ensure that the region of relevance is scrutinized by GA, preselected lower and upper bounds are passed as arguments after the number of variables is passed. The optimization process is terminated when the fitness value becomes lower than the function tolerance. Optimized parameters are used to modify the HMM to achieve a kind of GA-optimized HMM. At the model testing phase, abnormalities perceived to have ensued from abrupt changes in G_s are smoothened with a correction factor (ξ), which can either occur in the morning as (ξ_m) and/or evening time as (ξ_e). ξ is computed using an interior-point algorithm, a procedure that requires a fitness assignment and a constraint set by error definition, bounds whose upper value is fixed at the corresponding $P_{act,}$ and parameter initialization. Data analytics provided the basis upon which ξ must be applied. From the results, it can be deduced that when the average change in the absolute value of G_s ($\left|\overline{\Delta G_s}\right|$) is more than 128% in the morning time and/or when $\left|\overline{\Delta G_s}\right|$ in the evening exceeds 90%, then the application of ξ becomes inevitable. In this study, $\left|\overline{\Delta G_s}\right|$ in the morning and evening time are respectively articulated as $\left|\overline{\Delta G_s}\right|_m$ and $\left|\overline{\Delta G_s}\right|_e$. Nevertheless, the GA optimization process is considered non-iterative in prediction cases that necessitate the use of ξ. Finally, in line with this study's objective, which is to further investigate the dependability of our model by using a more recent input dataset obtained from the same SCADA system, the validation results are presented in the next section. The flowchart of the PV P_o forecast procedure described in this section is presented in Fig. 2.

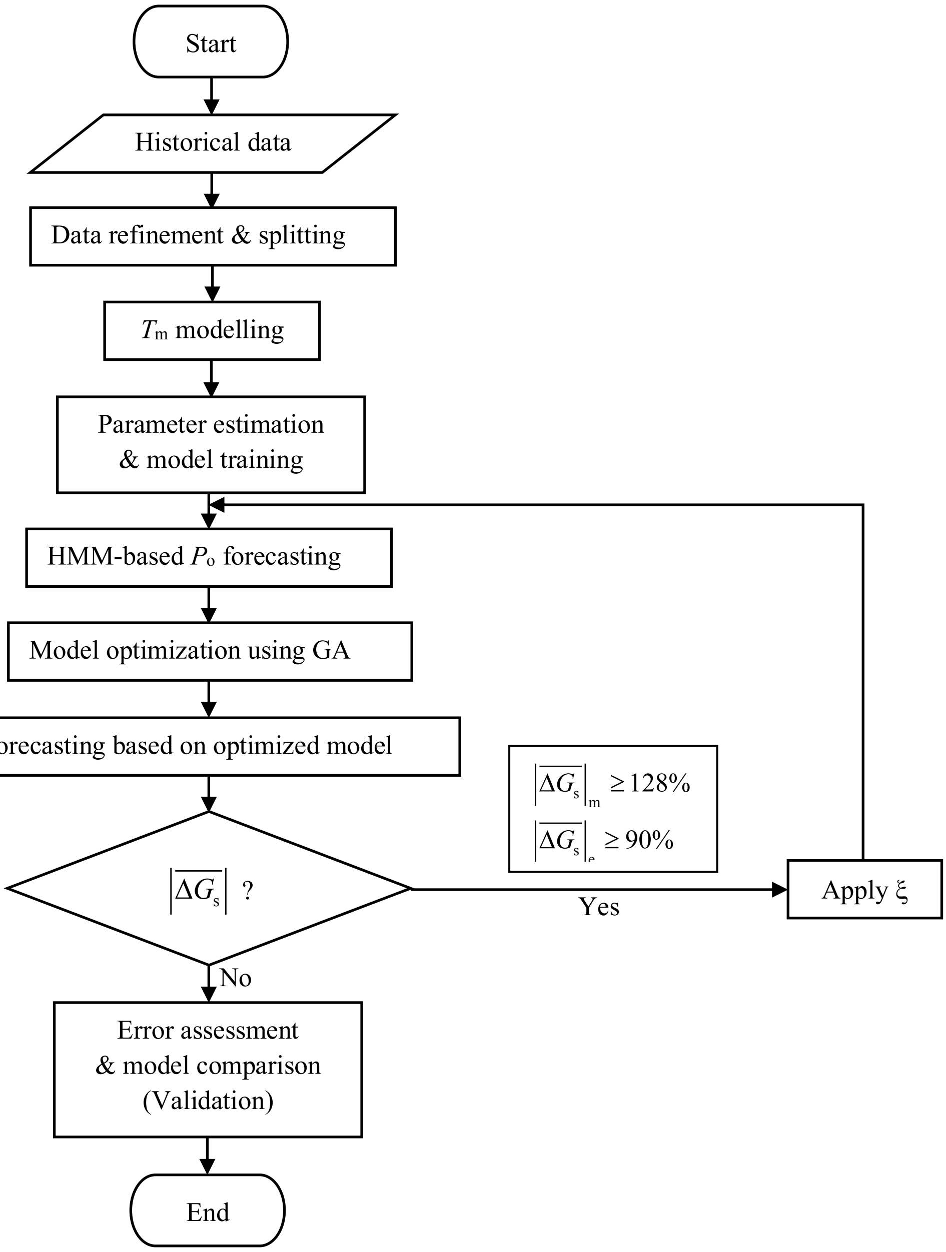

Fig. 2. The flowchart of the 1.2 kW PV P_o forecast procedure [18].

3. Model Validation Results

In this study, the forecast model validation is implemented; using input dataset for March to June 2019; based on the computation of normalized root mean square error (nRMSE) and mean absolute percentage error (MAPE) given as follows:

$$\mathrm{nRMSE} = \frac{1}{P_{\mathrm{rated}}}\sqrt{\frac{1}{n}\sum\nolimits_{i=1}^{n}(P_{\mathrm{a,1}} - P_{\mathrm{f,i}})^2} \tag{3}$$

$$\mathrm{MAPE} = 100 \times \frac{1}{n}\sum\nolimits_{i=1}^{n}\frac{|P_{\mathrm{a,i}} - P_{\mathrm{f,i}}|}{P_{\mathrm{a,i}}} \tag{4}$$

where P_{rated} is the PV rated power measured in kW, P_a and P_f are actual and forecasted power respectively, measured in kW, and n is the number of observations. The two error metrics compare the forecasted P_o and P_{act} values. Forecast models with good performance should present low nRMSE and MAPE. Figure 3a presents the results of P_o model validation of the day 14.03.2019 using HMM and HMM+GA. The HMM P_o, expressed as P_{HMM}, overshoot is noticeable almost over the entire hour of the day. The peak overshoot is about 20-24% of P_{act} occurring at the hour of 13.00. Improving the P_o forecasting close to the P_{act}, requires predicting P_o with HMM+GA; expressed as P_{opt} and the results are seen to match almost with the P_{act}, except for hours between 11.00 and 13.00 indicating slight under-forecast. To consider the error of HMM and HMM+GA (Fig. 3b), the values of $nRMSE_{opt}$ are well below that of $nRMSE_{HMM}$. The HMM is observed to over-forecast the data points with an ensemble nRMSE of 7.44%, whereas the ensemble nRMSE of GA-integrated HMM is reduced to 2.44%.

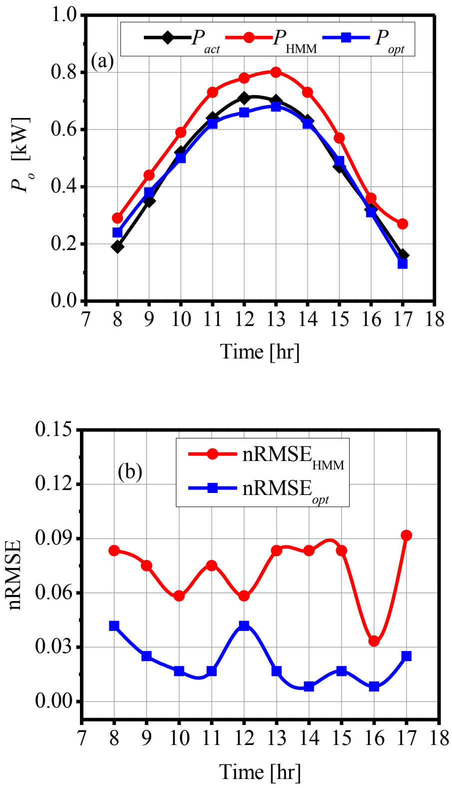

Fig. 3 (a) P_o forecast and (b) nRMSE of models on 14.03.2019 using HMM and HMM+GA.

Similarly, Fig. 4a presents the results of P_o model validation of the day 19.03.2019 based on HMM and HMM+GA. The overshoots of P_{HMM} manifest between 11.00 and 15.00 hours. The peak overshoot is about 12-15% of P_{act} occurring at the hour of 13.00. Improving the P_o forecasting close to the P_{act}, requires predicting P_o with P_{opt}, and the results can be observed to reasonably approximate the P_{act}. Considering the error of HMM and HMM+GA (Fig. 4b), the values of $nRMSE_{opt}$ are noticeably lower than that of $nRMSE_{HMM}$. The HMM is observed to over-forecast the data points with an ensemble nRMSE of 5.69%, whereas the ensemble nRMSE of GA-integrated HMM is reduced to 1.62%.

The HMM and HMM+GA P_o forecast validation on 21.03.2019, is as shown in Fig. 5a. The P_o forecasting using HMM is slightly higher above P_{act}, particularly between 10:00 and 13:00. The over-forecast of the HMM is reduced with the HMM+GA model which forecasts the P_{opt} to match almost with the P_{act}. Error consideration based on nRMSE (Fig. 5b) shows that $nRMSE_{opt}$ values are well below those of $nRMSE_{HMM}$. The HMM gives a maximum nRMSE of about 7% between 9:00 and 10:00, whereas the optimized model presents a maximum nRMSE value of nearly 5% at about 9:00 hour. The ensemble $nRMSE_{HMM}$ of 4.43% as against 2.71% for $nRMSE_{opt}$ further justifies the overshooting nature of the HMM.

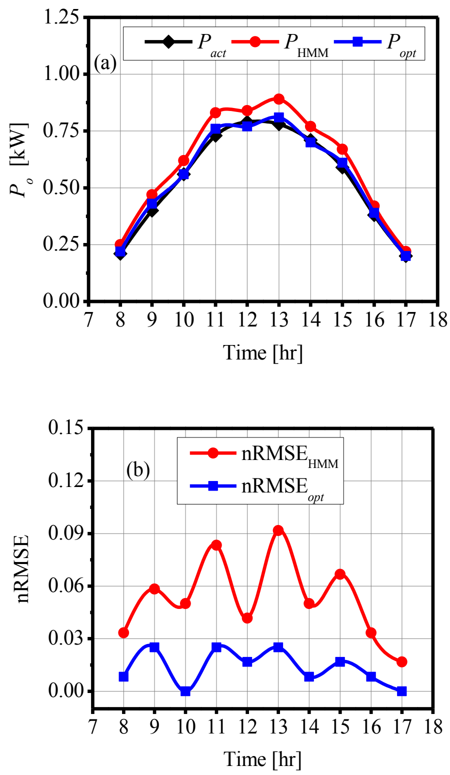

Fig. 4 (a) P_o forecast and (b) nRMSE of models on 19.03.2019 using HMM and HMM+GA.

The HMM and HMM+GA P_o forecast validation on 17.05.2019, is as shown in Fig. 6a. The P_o output forecasting using HMM is higher above P_{act}, over the entire time considered. The over-forecast of the HMM is reduced with the HMM+GA model which forecasts the P_{opt} more precisely. Error consideration based on nRMSE (Fig. 6b) shows that $nRMSE_{opt}$ values are well below those of $nRMSE_{HMM}$. The HMM gives a maximum nRMSE of about 14% about 11:00 hour, whereas the optimized model presents a maximum nRMSE value of nearly 4% around the hours of 9:00 and 16:00. The ensemble $nRMSE_{HMM}$ of 8.68% as against 2.85% for $nRMSE_{opt}$ further explains the overshooting nature of the HMM.

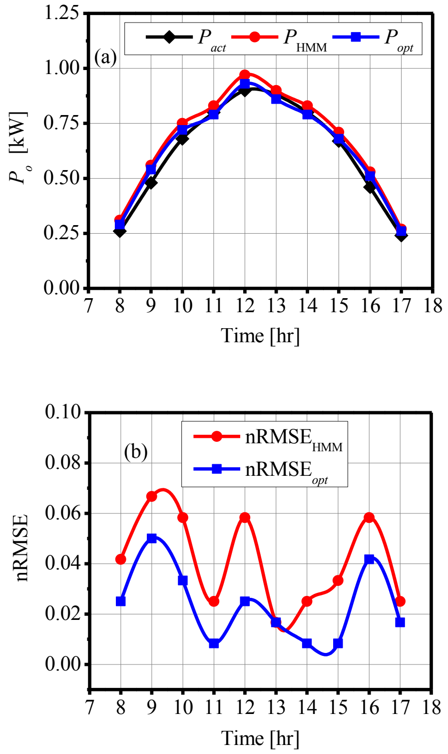

Fig. 5 (a) P_o forecast and (b) nRMSE of models on 21.03.2019 using HMM and HMM+GA.

Figure 7a presents the results of P_o model validation of day 27.03.2019 based on HMM and HMM+GA. As a CD, the P_{act} fluctuates over the entire hours of the day. The P_{opt} is closer to P_{act} than P_{HMM}, particularly between the hours of 10.00 and 15.00 with forecast peaks at about 10.00 and 13.00 hours. Nonetheless, P_{HMM} and P_{opt} do not approach P_{act} at 17:00 due to the influence of instantaneous change in G_s. According to their nRMSE curve (Fig. 7b), both nRMSEs present the highest values. In addition, $nRMSE_{opt}$ and $nRMSE_{HMM}$ have the highest values of about 34-36% and 44-46% respectively occurring at 17.00 hour. This confirms that HMM and HMM+GA models have a limitation for instantaneous changes in G_s. Based on the criterion established in the forecast model development section for the use of ξ in the evening time, the abnormality was corrected with a value of 0.25.

Following the use of ξ, both P_{HMM} and P_{opt} present more reasonable P_o curves in Fig. 8a. To determine the influence ξ-adapted HMM and ξ-adapted HMM+GA have on the nRMSE (Fig. 8b), it can be observed that $nRMSE_{HMM}$ and $nRMSE_{opt}$ at 17.00 reduced close to about 3% and 1% respectively. The reduced peaks of $nRMSE_{HMM}$ and $nRMSE_{opt}$ and their respective ensemble nRMSE values decreased to 4.97% and 2.61%, further reinforce the importance of ξ. To compare HMM and HMM+GA adapted with and without ξ (Fig. 7 and Fig. 8), the abnormalities and nRMSE are significantly reduced with the use of ξ, but also the values of nRMSE of ξ-adapted HMM and HMM+GA are also less fluctuating than without the ξ. The nRMSE of HMM+GA+ξ relatively maintains a range between 0-5%.

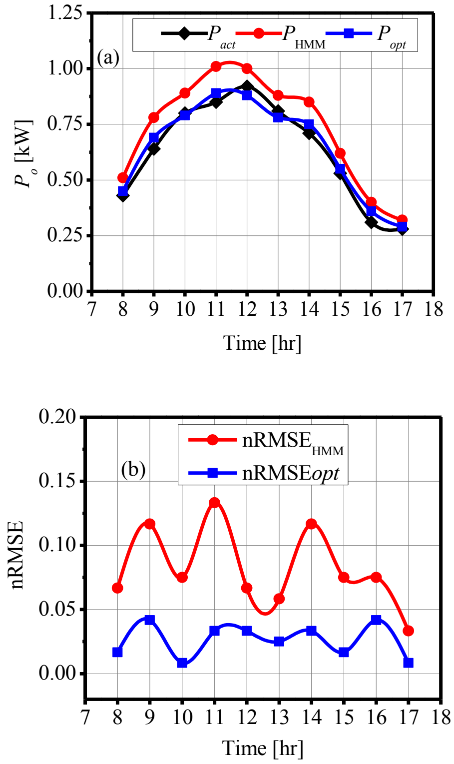

Fig. 6 (a) P_o forecast and (b) nRMSE of models on 17.05.2019 using HMM and HMM+GA.

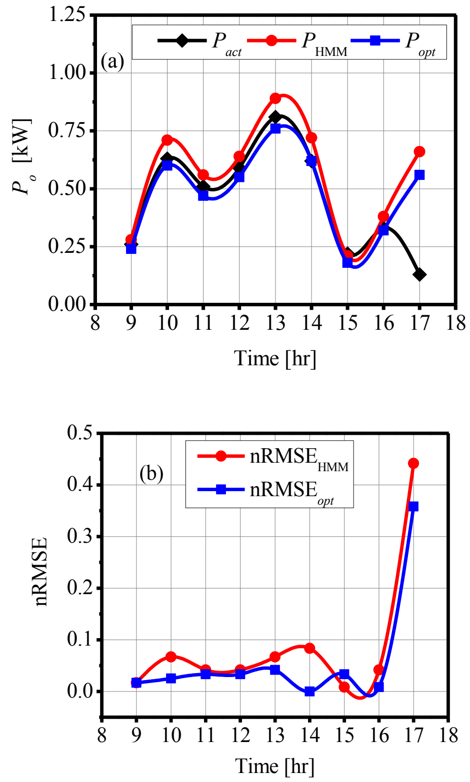

Fig. 7 (a) P_o forecast and (b) nRMSE of models on 27.03.2019 using HMM and HMM+GA.

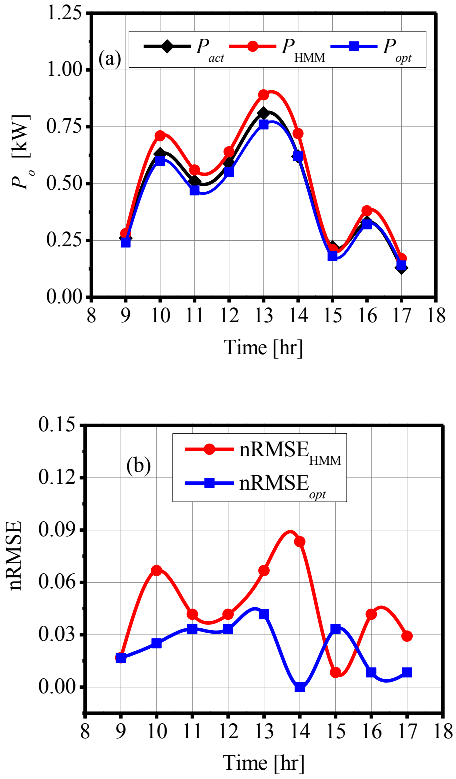

Fig. 8 (a) P_o forecast and (b) nRMSE of models on 27.03.2019 using HMM+ξ and HMM+GA+ξ.

Figure 9a presents the result comparison of P_o forecast models for 24.05.2019 using HMM+ξ and HMM+GA+ξ on cloudy sky condition. The computed value of ξ used to adapt the abnormality occurring at 17.00 is 0.25. P_o forecasted with HMM+ξ presents overshoot noticeably between 11:00 and 16:00. To improve the P_o close to the P_{act}, P_{opt} was predicted based on HMM+GA+ξ model, which is observed to forecast P_o more accurately. To consider the forecast error (Fig. 9b), the ensemble $nRMSE_{opt}$ values of 2.36% for the HMM+GA+ξ is lower than the 8.67% $nRMSE_{HMM}$ of the HMM+ξ; especially the value of HMM+GA+ξ relatively maintains a range between 0-4%.

Figure 10a presents the results of P_o forecast for the day 19.06.2019 based on cloudy sky condition using HMM+ξ and HMM+GA+ξ models. The abnormalities occurring at 8:00 and 17.00 hours are adjusted with ξ values of 0.40 and 0.25, respectively. P_{HMM} and P_{opt} present a good agreement with P_{act} exclusive of the hours between 10.00 – 12.00 and 14.00 hour. The improvement in P_o prediction with HMM+GA+ξ can be perceived by considering the nRMSE curves shown in Fig. 10b. Although the under-forecast of HMM+GA+ξ between 10.00 and 12.00 presents a maximum $nRMSE_{opt}$ of approximately 7.5%. Notwithstanding, $nRMSE_{HMM}$ globally peaks at around 8.5% and HMM+GA+ξ also presents a lower ensemble nRMSE value of 3.82% against HMM+ξ whose error value is 4.94%.

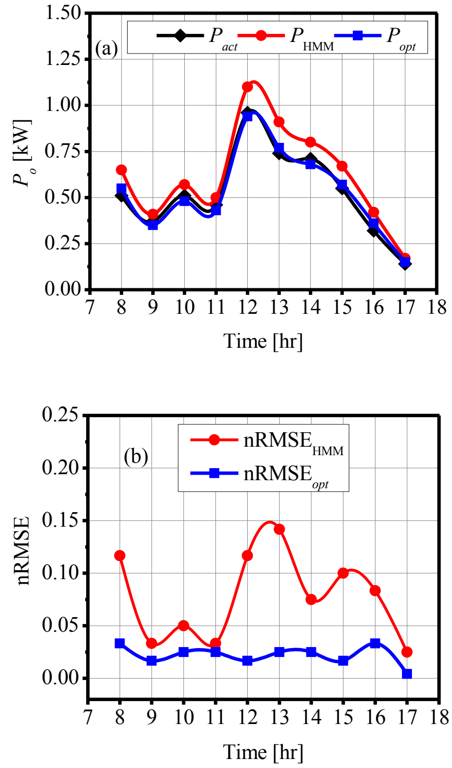

Fig. 9 (a) P_o forecast and (b) nRMSE of models on 24.05.2019 using HMM+ξ and HMM+GA+ξ.

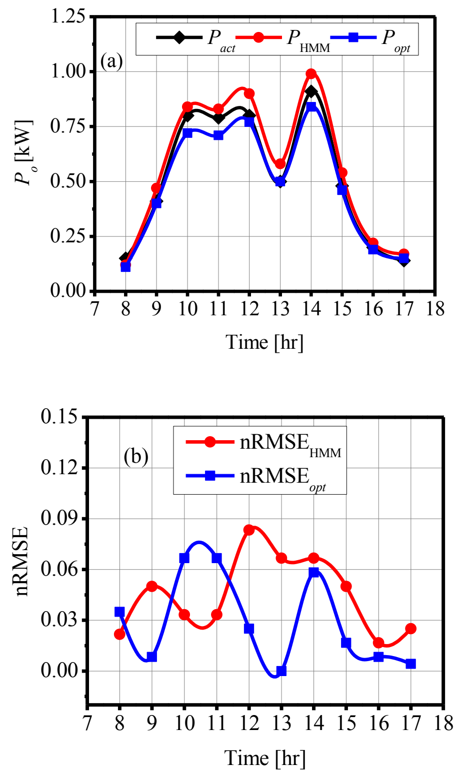

Fig. 10 (a) Po forecast and (b) nRMSE of models on 19.06.2019 using HMM+ξ and HMM+GA+ξ.

At the validation step, the HMM and GA-optimized HMM performance with or without ξ on hour-ahead P_o forecasting of the PV system under different conditions of G_s are synopsized in Table 3. The computation of nRMSE and MAPE error metrics consolidate the prediction strength of the standalone model (HMM) and GA-enhanced model (HMM+GA). Both nRMSE and MAPE decreased when GA is integrated with HMM, corresponding to the class of the day under CSD consideration. This reflects PV power forecasting with GA-integrated HMM has a higher P_o prediction capability than ordinary HMM, as the results of the optimized forecast parameters. In CD consideration, the use of ξ; based on the established forecast criteria in section 2 further improves the accuracy of the forecast as expressed in the percentage of nRMSE and MAPE. These criteria were developed based on data analytics. As the instrumental decision support, if $\left|\overline{\Delta G_s}\right|_m$ is more than 128%, ξ in the range of 0.33 – 0.41 is suitable. Then again, if $\left|\overline{\Delta G_s}\right|_e$ exceeds 90%; applicable ξrange between 0.24 - 0.35. The HMM with or without ξ presents the average nRMSE and MAPE larger than HMM+GA with or without ξ. Besides, the average nRMSE and MAPE of HMM+GA with or without ξ is 2.63% and 6.05%. As a result, the incorporation of GA and ξ are able to improve the forecasting accuracy of the hour-ahead P_o of the PV system based on the optimized forecast parameters. For all the days used in the validation process, the proposed method outpaced the ordinary model (HMM) with or without ξ. It can be observed, from Table 3, that ξ is inconsequential for a typical CSD; and for CDs in which ξ is determined, its adoption may be in the morning and/or evening time.

Table 4 presents a synopsis of the previous studies on short-term PV P_o forecasting considered. To compare with other models developed in former studies, Lahouar et al., 2017 reported a 24 hour-ahead P_o forecasting of a 500 kW_p PV system, using Random forests based on bagging algorithm with and without G_s; given a MAPE of 28.97% in April [6] as one of the results. In the study conducted by Raza et al., 2017 [24], Autoregressive predictor; Radial Basis Function network enhanced with Particle Swarm Optimization (RBF+PSO), and PSO-augmented Feed-forward Neural Network (FFNN+PSO) were deployed as a multivariate ensemble framework to make seasonal 24 hours and 7 days-ahead P_o prediction of a 2.14 MW_p PV plant in Australia. The findings present an nRMSE of 9.55% and 9.51%, in the spring season, for CSD and CD, respectively. Zhong et al., 2017 predicted the power production volume of a PV system based on PSO boosted-multivariable Grey theory method; and the model validation with PSO gives a Mean Relative Error (MRE) reducing from 7.14% to 3.53%, equivalent to approximately 51% decrease [13]. In a previous study carried out in Thailand by Eniola et al., 2019, the power output of a 1.2 kW_p PV system was forecasted an hour-ahead based on the following models: HMM, HMM+GA, HMM+ξ, and HMM+GA+ξ; depending on the class of day under consideration. The authors reported an average nRMSE of 2.33% for HMM+GA with or without ξ, maximum MAPE of 12.33% occurring in April, maximum nRMSE of 2.55% and 4.29% for CSD and CD respectively, and HMM+GA together with or without ξ presenting around 54% decline in nRMSE at the testing phase. The testing results reflected that HMM+GA predicts P_o for CSDs more precisely, whereas HMM+GA+ξ gives the best P_o forecast for CDs, supporting the consideration of their proposed prediction model as a suitable method for hour-ahead P_o forecasting of the PV plant [18]. The results of this current study built on the technique developed by Eniola et al., 2019 [18], and proven with more recent input parameters; indicated an average nRMSE of 2.63% for HMM+GA with or without ξ, maximum MAPE of 7.95% occurring in March, maximum nRMSE of 2.85% for CSD and 3.82% for CD, and HMM validation with GA together with or without ξ gives about 58% decline in nRMSE. Considering the trend of the present results discussed above, it is worth mentioning that our validation results are in good agreement with those obtained at the model testing phase reported by Eniola et al. [18].

Accordingly, as HMM+GA is capable of predicting the P_o for CSD more accurately, whereas HMM+GA+ξ is efficient in practically forecasting PV P_o under CD condition; GA-reinforced HMM with or without ξ can be considered a reasonable method particularly for the hour-ahead prediction of the P_o of a PV system. With this model, energy planning and management can be ameliorated. Emphatically, the present model can be relevantly deployed in practical cases at locations with comparable meteorological data with Thailand's. Nonetheless, model implementation in sites with different weather patterns may necessitate the use of no less than six months of the historical dataset in retraining the proposed forecast model. Furthermore, the approach of this study is deemed practically effective and beneficial considering that out-of-sample data obtained from a PV power plant in Thailand have been deployed as inputs. For all the conditions reflected, the model behavior is in good agreement with the current validation results. As solar resource is highly stochastic, the models articulated in the present study can be employed to precisely guesstimate the P_o of PV in advance. It can be utilized for decision-making on electric power transmission, electric load drop or gain, vis-à-vis having reasonable control overpower quality problems, including frequency deviation from the rated value when demand is higher than supply. Besides, in decentralized electricity markets with power bidding processes, plant owners and grid operators may incur additional cost in the form of penalty if they fail to supply electric power within stipulated tolerance bands. This extra cost can be minimized with the models proposed in this study.

Although, this research did not consider the effect of seasonal variation on forecast model performance. Nevertheless, previous tests and this experimental validation using out-of-sample datasets recorded at different times in year 2018 and 2019 showed that the model performances are consistent. It is not incredible that the model results will be stable if the training data size is six months or more and the meteorological variables follow a similar distribution. Also, the impact of the order of the state probability distribution matrix A on the forecast model behavior is not investigated. However, if the number of state increases from 5 to a positive integer N, matrix A will remain a square matrix, all the element across each row of matrix A sum up to 1; each element of matrix A ranges between 0 and 1 inclusive; but the order of matrix A will increase from 5-by-5 to N-by-N. In such a case, since more states will translate to smaller state interval, it may not be incorrect to anticipate an entirely distinct forecast model capability. Therefore, future work would examine the effect of the order of the state probability distribution matrix on forecast model performance.

Table 3. Forecast model validation performance for March to June 2019

Date	Class	Models	ξ_m	ξ_e	nRMSE [%]		MAPE [%]	
					P_{HMM}	P_{opt}	P_{HMM}	P_{opt}
14.03.2019	CSD	HMM/HMM+GA	n/a	n/a	7.44	2.44	24.84	7.95
19.03.2019	CSD	HMM/HMM+GA	n/a	n/a	5.69	1.62	12.39	3.02
21.03.2019	CSD	HMM/HMM+GA	n/a	n/a	4.43	2.71	9.74	5.87
17.05.2019	CSD	HMM/HMM+GA	n/a	n/a	8.68	2.85	16.79	5.56
27.03.2019	CD	HMM+ξ/HMM+GA+ξ	n/a	0.25	4.97	2.61	12.37	6.91
24.05.2019	CD	HMM+ξ/HMM+GA+ξ	n/a	0.25	8.67	2.36	18.35	5.57
19.06.2019	CD	HMM+ξ/HMM+GA+ξ	0.40	0.25	4.94	3.82	12.33	7.47
Average					6.40	2.63	15.26	6.05

Table 4. Forecast model result comparisons

Researchers	Lahouar et al., [6].	Raza et al., [24].	Zhong et al., [13].	Eniola et al., [18].	Present study
Location	Tunisia	Australia	China	Thailand	Thailand
Study year	2017	2017	2017	2019	2019
PV technology	--	--	--	Thin-film Si	Thin-film Si
PV Capacity	500kW$_p$	2.14MW$_p$	--	1.2kW$_p$	1.2 kW$_p$
Training data	G_s	G_s	G_s	G_s	G_s
	--	--	T_{amb}	T_{amb}	T_{amb}
	w	w	--	w	w
	w_d	--	--	--	--
	T_m	T_m	--	T_m	T_m
	--	P_o	--	P_o	P_o
	h_u	h_u	--	--	--
	--	--	i_i	--	--
Method(s)	Random forests using bagging algorithm with and without G_s.	Autoregressive predictor + (RBF+PSO predictor) + (FFNN+PSO predictor).	Multivariable Grey theory + PSO.	HMM HMM+GA HMM+ξ HMM+GA+ξ	HMM HMM+GA HMM+ξ HMM+GA+ξ
Data size	--	12 months	--	6 months	6 months
Forecast horizon	24 hours	24 hours	24 hours	1 hour	1 hour
nRMSE	--	9.55% - CSDs 9.51% - CDs	--	2.33% avg. 2.55% max. -CSDs 4.29% max. - CDs	2.63% avg. 2.85% max. - CSDs 3.82% max. - CDs
MAPE	28.97% - April	--	--	12.33% max. -April	7.95% max. - March
% Reduction	--	--	MRE (~51%)	nRMSE (~54%)	nRMSE (~58%)

4. Conclusion

Herein, the present study further validates the method for hour-ahead P_o forecasting of a PV system based on HMM and HMM+GA together with or without correction factor (ξ), as projected by Eniola et al., 2019 [18]. For CSDs, HMM+GA can predict the P_o with high accuracy. Whereas days corresponding to CDs require ξ to adapt HMM+GA when $\left|\overline{\Delta G_s}\right|_m \geq 128\%$ and/or $\left|\overline{\Delta G_s}\right|_e \geq 90\%$. In all days used in the validation process, HMM+GA and HMM+GA+ξ give more potential forecast than HMM and HMM+ξ. The trend of the results from this study showed that our validation results are in harmony with those obtained at the model testing phase reported by Eniola et al. [18]. Considering the forecast models' average nRMSE and MAPE of 2.63% and 6.05%, respectively, GA-optimized HMM with or without ξ reflects a suitable technique to forecasting PV P_o on an hourly basis. Grid operators and PV power plant owners could implement this model to address customer's dissatisfaction over power quality issue, reduce the cost of energy reserve, and determine the cost of energy beforehand.

Acknowledgements

This research is fully supported by Thailand International Development Cooperation Agency (TICA) scholarship. The authors are grateful to Royal Thai Government. Also, we appreciate Naresuan University and the School of Renewable Energy and Smart Grid Technology (SGtech), Phitsanulok, Thailand for their kind cooperation.

References

[1] A. Alzahrani, P. Shamsi, M. Ferdowsi, and C. Dagli. "Solar irradiance forecasting using deep recurrent neural networks", 2017 IEEE 6th International Conference on Renewable Energy Research and Applications (ICRERA), pp. 988-994, 5-8 Nov. 2017.

[2] E. Ndiaye, A. Ndiaye, M. Tankari, and G. Lefebvre. "Adaptive Neuro-Fuzzy Inference System Application for The Identification of a Photovoltaic System and The Forecasting of Its Maximum Power Point", 2018 7th International Conference on Renewable Energy Research and Applications (ICRERA), pp. 1061-1067, 14-17 Oct. 2018.

[3] S. Ech-Charqaouy, "Integration of Decentralized Generations into the Distribution Network-Smart Grid Downstream of the Meter", Int. J. Smart Grid, Vol. 4, No. 1, pp. 17-27, 2020.

[4] H. Pedro, R. Inman, and C. Coimbra, Mathematical methods for optimized solar forecasting, Cambridge: Woodhead Publishing, 2017, ch. 4.

[5] Z. Chen, S. Che, Y. Xu, and D. Yin, "Prediction Method of PV Output Power Based on Cloud Model", J. Eng., Vol. 2017, No. 13, pp. 1519-1523, 2017.

[6] A. Lahouar, A. Mejri, and J. Slama. "Importance based selection method for day-ahead photovoltaic power forecast using random forests", 22017 International Conference on Green Energy Conversion Systems (GECS), pp. 1-7, 23-25 March 2017.

[7] A. Almadhor. "Performance prediction of distributed PV generation systems using Artificial Neural Networks (ANN) and Mesh Networks", 2018 International Conference on Smart Grid (icSmartGrid), pp. 88-91, 4-6 Dec. 2018.

[8] V. Sharma, D. Yang, W. Walsh, and T. Reindl, "Short term solar irradiance forecasting using a mixed wavelet neural network", Renew. Energy, Vol. 90, No. 2016, pp. 481-492, 2016.

[9] L. Fen, L. Chunyang, Y. Yong, Y. Quanquan, Z. Jinbin, and W. Lijuan, "Short-term photovoltaic power probability forecasting based on OLPP-GPR and modified clearness index", J. Eng., Vol. 2017, No. 13, pp. 1625-1628, 2017.

[10] A. Sa'ad, H. Zied, and A. Nyoungue. "A day-ahead Multi-Approach Machine Learning Technique for Photovoltaic Power Forecasting", 2020 9th International Conference on Renewable Energy Research and Application (ICRERA), pp. 257-262, 27-30 Sept. 2020.

[11] P. Tang, D. Chen, and Y. Hou, "Entropy method combined with extreme learning machine method for the short-term photovoltaic power generation forecasting", Chaos, Solitons & Fractals, Vol. 89, pp. 243-248, 2016.

[12] F. Barbieri, S. Rajakaruna, and A. Ghosh, "Very short-term photovoltaic power forecasting with cloud modeling: A review", Renew. Sustain. Energy Rev., Vol. 75, pp. 242-263, 2017.

[13] Z. Zhong, C. Yang, W. Cao, and C. Yan, "Short-Term Photovoltaic Power Generation Forecasting Based on Multivariable Grey Theory Model with Parameter Optimization", Math. Probl. Eng., pp. 1-9, 2017.

[14] R. Bayindir, M. Yesilbudak, M. Colak, and N. Genc. "A Novel Application of Naive Bayes Classifier in Photovoltaic Energy Prediction", 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 523-527, 18-21 Dec. 2017.

[15] M. Giorgi, P. Congedo, M. Malvoni, and D. Laforgia, "Error analysis of hybrid photovoltaic power forecasting models: A case study of mediterranean climate", Energy Convers. Manag., Vol. 100, pp. 117-130, 2015.

[16] A. Eseye, J. Zhang, and D. Zheng, "Short-term photovoltaic solar power forecasting using a hybrid Wavelet-PSO-SVM model based on SCADA and Meteorological information", Renew. Energy, Vol. 118, pp. 357-367, 2018.

[17] M. Colak, M. Yesilbudak, and R. Bayindir, "Daily Photovoltaic Power Prediction Enhanced by Hybrid GWO-MLP, ALO-MLP and WOA-MLP Models Using Meteorological Information", Energies, Vol. 13, No. 4, 2020.

[18] V. Eniola, T. Suriwong, C. Sirisamphanwong, and K. Ungchittrakool, "Hour-ahead Forecasting of Photovoltaic Power Output based on Hidden Markov Model and Genetic Algorithm", Int. J. Renew. Energy Res., Vol. 9, No. 2, pp. 933-943, 2019.

[19] S. Dubey, J. Sarvaiya, and B. Seshadri, "Temperature Dependent Photovoltaic (PV) Efficiency and Its Effect on PV Production in the World – A Review", Energy Procedia, Vol. 33, pp. 311-321, 2013.

[20] N. Savvakis and T. Tsoutsos, "Performance assessment of a thin film photovoltaic system under actual Mediterranean climate conditions in the island of Crete", Energy, Vol. 90, pp. 1435-1455, 2015.

[21] J. Joshi, K. Tankeshwar, and S. Srivastava, "Hidden Markov Model for Quantitative Prediction of Snowfall and Analysis of Hazardous Snowfall Events over Indian Himalaya", J. Earth Syst. Sci., pp. 126: 033, 2017.

[22] L.R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition", Proc. IEEE, Vol. 77, No. 2, pp. 257-286, 1989.

[23] J. Wang, R. Ran, and Y. Zhou, "A Short-Term Photovoltaic Power Prediction Model Based on an FOS-ELM Algorithm", Appl. Sci., Vol. 7, pp. 423, 2017.

[24] M. Raza, M. Nadarajah, and C. Ekanayake. "A multivariate ensemble framework for short term solar photovoltaic output power forecast", 2017 IEEE Power & Energy Soc. Gen. Meet., DOI: 10.1109/PESGM.2017.8274676, pp. 1-5, 2017.