

Power Generation Forecast of Hybrid PV-Wind System

M.J. Sanjari, H.B. Gooi, *Senior Member, IEEE*, Nirmal-Kumar C. Nair, *Senior Member, IEEE*

Abstract— Due to their intermittency and unpredictability, increasing the penetration level of renewable energy (RE) resources to the power system leads to difficulties in operation. Reliable system operation requires a precise forecast of generated power by RE units. Photovoltaic (PV) and wind units are the significant portion of RE resources integrated into the power system. This paper proposes a forecast method for PV and wind generated power to achieve good prediction accuracy in different weather conditions. Not only is the relation between the wind and PV output power modeled, but the heat index (HI) is also taken into consideration as a useful meteorological variable to achieve the 15-min ahead precise expectation of PV/wind output power. Moreover, the input data is discretized in such a way that the best accuracy for the PV and wind power forecast is achieved. Comparing the results of the proposed method with the historical data recorded at actual PV and wind plants shows that the proposed forecast method results in high accuracy in PV and wind output power forecast. Moreover, the forecast model performance with HI consideration is compared with the model not using HI as an input variable.

Index Terms— Higher-order multivariate Markov Chain, photovoltaic power forecast, power generation uncertainty, power system management and control, wind power forecast.

ACRONYMS AND NOTATIONS

RE	Renewable energy
HI	Heat index
HMC	Higher-order Markov chain
HMMC	Higher-order multivariate Markov chain
DS	Data series
T	Temperature
RH	Relative humidity
AP	Atmospheric pressure
$\rho^{(m)}$	Monthly correlation index
$\bar{\rho}$	Mean of the linear correlation index
STD [ρ]	Standard deviation of the linear correlation index
$\sigma_i^{(m)}$	Standard deviation of i -th set of data in m -th month
$X^{(r)}$	r -th data series, $X^{(r)} = \{x_t^{(r)}\}$
$N^{(r)}$	Number of states related to r -th DS
$x_t^{(r)}$	Measured value of r -th DS at time t
$\tilde{x}_i^{(r)}$	i -th state margin related to the r -th DS

$\tilde{x}_i^{(r)}, \tilde{x}_{i+1}^{(r)}$	Lower and upper boundaries of $x_i^{(r)}$
$ x_i^{(r)} $	Length of i -th state related to r -th DS
$\delta_i^{(r)}$	Average intervals that $x_t^{(r)}$ stays in state $x_i^{(r)}$
s	Number of categorical DS
$NRMSE$	Normalized root mean square error
$c^{(r)}$	Current state in r -th DS
$\tilde{x}_i^{(r)}$	i -th state representative for r -th DS
$\pi_i^{(r)}$	State probability distribution of r -th DS at time t .
$\lambda_{i,j}^{(r)}$	Lag inter-sequence coefficient
$Q_{i,j}^{(r)}$	i -th step probability matrix: transition probabilities from states in j -th DS at time $t + 1 - i$ to states in r -th DS at time $t + 1$
$\pi^{(r)}$	Vector of stationery distribution related to r -th DS
$\tilde{\pi}^{(r)}$	Vector of estimated stationery distribution related to r -th DS
$\tilde{x}_{t+1}^{(r)}$	Forecasted value for the r -th DS at time $t + 1$

I. INTRODUCTION

INTEGRATION of REs to the power system has gained major interest in recent years due to the lack of carbon emission during their power generation cycles [1]. Being non-polluting, renewable and available in most sites and requiring no fuel, PV and wind have been the most significant parts of REs incorporated in the bulk power system [2]. To improve the economic sustainability and technical characteristics, the hybrid energy plants including different types of REs and energy storage systems are connected to the power system. The hybrid PV-wind power plant has been investigated as an economically and technically efficient scheme [3].

The wind and PV generated power has high variability, therefore the power system operation and control becomes more complex by increasing their penetration either in the form of hybrid renewable power plants or separate RE resource operations [4]. In line with this statement, forecasting the PV/wind energy production on regional scale would be crucial for the transmission system operators and managers [5] when making the medium-term control and long-term planning of the transmission system, especially regarding the cross-border power flows [6] or the power system with low marginal risk index [7]. This explains why much recent research has been devoted to the PV/wind generated power forecast [8]. Two approaches have been proposed in the PV/wind output power forecast. Some research studies have attempted to forecast the solar irradiance and wind speed and

M. J. Sanjari is with School of Engineering, Griffith University, Gold Coast, Australia, 4215, email: m.sanjari@griffith.edu.au

H. B. Gooi is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798 (e-mail: ehbgooi@ntu.edu.sg).

N. Nair is with the Department of Electrical and Computer Engineering, University of Auckland, New Zealand, 1142, e-mail: n.nair@auckland.ac.nz.

convert them to the power generated by PV/wind units by using the predefined mathematical models [9]. In this approach, some issues are not considered, e.g. tilt angle, control actions of the solar tracking system, shadow effect of trees, buildings and other generation units and nonlinear control actions of wind turbines. The other approach directly models the output power of REs [10]. Compared to the indirect PV/wind power prediction, this method leads to a higher level of accuracy because the nonlinear behavior of different elements in the RE site are modeled as the generated power is measured at the connecting point to the power system. Therefore, the direct RE power forecast is studied in this paper.

Various prediction methods have been presented to forecast the RE output power. The size of input variable vector is fixed in a number of proposed models meaning that a determined vector of measured variables are keyed in the model to forecast the PV/wind power generation in the next time interval. If a set of variable(s) is going to be added to the input vector to increase the forecast model accuracy, the model should be rebuilt, meaning the training procedure should be started from the beginning. This issue negatively affects the flexibility of the forecast models. In [11], the support vector machine (SVM)-based forecast is integrated into the finite-state Markov models and the proposed SVM-enhanced Markov model is applied to predict the wind power in a 10-min ahead horizon. The proposed method only uses the wind power measured samples to develop the model. In other words, additional variables cannot be exploited as input to the power forecast unit. Moreover, the incomplete data set cannot be handled by the proposed method in [11]. Another example is proposed in [12], which captures the regime-switching behavior with an approach relying on Markov-switching autoregressive (AR) models for the 1-step-ahead point forecasting of wind power time-series. Well-ordered data samples are assumed to be available during parameterization of the model coefficients. As a probabilistic approach, the first order Markov chain (MC) has been applied to PV/wind power modeling to predict data based on the last measured data sample in [11]. Since the first order MC is able to model only the normal fluctuations of the PV/wind power, it results in significant uncertainties and errors for PV/wind power modeling, especially in a short term forecast. As another example, in [13] a solar power prediction model based on various satellite images and SVM learning scheme is proposed. The historical satellite images are utilized to configure a large number of input and output data sets for the SVM learning to forecast motion vectors of clouds and finally to predict the power generated by the PV module in the next time interval.

Adding new features/variables to the list of prediction model inputs can increase the accuracy of the forecast units. However, all the needed sensors are not installed at the time of site construction. In other words, with the technology growing over time, some variables/features would be provided later than the RE site accomplishment [10]. Moreover, the PV arrays and wind turbine characteristics are being continuously

changed over time and the RE peripheral conditions are also exposed to variations such as new building construction, tree condition changes, etc. Therefore, the forecast model should be time-adaptable to forecast RE power effectively over time. Failing to have this capability in forecasting methods [10], [11] leads to a higher level of forecast errors over time.

Another important issue is that in some cases, a coherent batch of historical data is not available for developing the forecast model, e.g. newly constructed PV/wind sites or cases in which there are not so many recorded data samples available or some samples are missed. Forecast models, which are developed based on a complete batch of historical data cannot be operative in this case [11]–[12].

A comprehensive model based on HMMC to forecast the 15-min ahead PV and wind output power is proposed in this paper, which is developed based on historical data of PV modules and wind turbine output power. The developed model predicts the PV/wind power variations based on the last data samples measured from the RE units while taking into account the generated power by the other RE unit(s) to consider the dependency between solar irradiation and wind speed. In other words, the proposed model considers both intra- and inter-relations between PV and wind power DS. This issue increases the power forecast accuracy because the solar irradiance and wind speed are correlated meteorological variables. Moreover, HI is also considered as a contributory meteorological variable to enhance the HMMC accuracy of the PV and wind power fluctuation forecast. HI is a nonlinear function of the ambient temperature and relative humidity reflecting a strong correlation with PV and wind output power. The model developing procedure proposed in this paper can be applied to an incomplete data set of the input variables. This feature is momentous in the case that some data samples are missed due to measurement problems or filtered out by the preprocessing data analysis to detect bad data. Parameters of the proposed method can be dynamically and adaptively updated with newly measured data points, i.e. PV and wind power and HI to capture the generation units' characteristic changes and peripheral variations. Moreover, the constructed model can exploit newly additive variable(s) to enhance the model performance. To capture the seasonality and daily non-stationarity of PV and wind power, time-adaptive sets of forecast model parameters are used, i.e. one special set of model parameters for each daily 4-hour epoch in each month. Based on output power measured from actual PV modules and wind turbines, the accuracy of the proposed forecast procedure is significantly improved.

The first contribution of this paper is to enhance the accuracy of PV/wind power output modeling by considering intra- and inter- relations between PV and wind power and also HI samples. The latter is an effective variable to increase the forecast algorithm accuracy, which is calculated in each time interval based on gauged instantaneous temperature and humidity.

The second contribution of this paper is proposing a flexible framework for forecasting PV/wind generated power in the case of disjointed batch of historical data. In other words, in

the cases where are few recorded data samples available or some samples are missed, the forecast model can be developed.

II. MATHEMATICAL FORMULATION OF HMMC

In this section, the notation and concept of HMMC are reviewed which are used in this paper to develop an effective tool for PV and wind power modeling.

A. Feature Selection for PV and Wind Power Modeling

In this section, the HMMC is applied to the problem of PV and wind output power modeling. To enhance the forecast model accuracy, the relationship between the PV/wind output power and also HI as an efficient meteorological data are considered in this paper.

To show the effectiveness of different meteorological variables on PV/wind power forecast, the monthly linear correlation index between the PV generated power, wind output and some meteorological data are calculated according to (1) in which $Y_1^{(m)}$ and $Y_2^{(m)}$ are sets of measured data samples in m -th month.

$$\rho^{(m)} = \frac{\text{cov}(Y_1^{(m)}, Y_2^{(m)})}{\sigma_1^{(m)} \sigma_2^{(m)}}, m = 1, 2, \dots, 12 \quad (1)$$

The PV/wind generated power, ambient temperature and relative humidity data measured in the year 2013 with the resolution of 15 min in an actual site in Belgium [14], [15] are used to calculate the linear correlation index. Moreover, HI, a nonlinear function of the T , and RH calculated as (2) [16], is considered as a descriptive meteorological index.

$$HI = \sum_{i,j=0}^3 C_{ij} (T)^i (RH)^j \quad (2)$$

where C_{ij} parameters, listed in [16], are the coefficients determined by different experimental results.

The mean and standard deviation of the monthly correlation indices are calculated as (3) and (4), respectively.

$$\bar{\rho} = \frac{1}{12} \sum_{m=1}^{12} \rho^{(m)} \quad (3)$$

$$\text{STD}[\rho] = \left[\sum_{m=1}^{12} \frac{(\rho^{(m)} - \bar{\rho})^2}{12} \right]^{\frac{1}{2}} \quad (4)$$

Statistical analysis of the linear correlation index between PV power and some meteorological data of the year 2013 is listed in Table I and the same statistical analysis of the linear correlation index between wind output power and some meteorological data of the year 2013 is listed in Table II. These two tables are used in section IV to prioritize several variables to be included in the forecast algorithm input.

As shown in these two tables, compared to the others, the correlations between HI and the PV and wind outputs are higher, which are reflected by the great mean values, i.e. -0.29 (bold font in gray background in Table I) and -0.58 (bold font in gray background in Table II), respectively. Moreover, the mentioned correlations remain quite unchanged during 12 months, which are reflected by low standard deviation values, i.e. 0.18 (bold font in white background in Table I) and 0.29

(bold font in white background in Table II), respectively. Therefore, HI is considered as an effective meteorological variable to enhance the PV/wind power forecast. Moreover, as shown in Table I, the mean value of the correlation between PV and wind output power is also high, i.e. -0.36, therefore the inter-relation between these two DS are taken into account in the RE power forecast and this explains the main motivation of using HMMC.

TABLE I
STATISTICAL ANALYSIS OF LINEAR CORRELATION INDEX BETWEEN PV POWER AND METEOROLOGICAL VARIABLES FOR YEAR 2013

Statistical index	wind output data	T	RH	HI	AP
$\bar{\rho}$	-0.36	0.19	-0.24	-0.29	0.02
STD $[\rho]$	0.17	0.24	0.31	0.18	0.21

TABLE II
STATISTICAL ANALYSIS OF LINEAR CORRELATION INDEX BETWEEN WIND POWER AND METEOROLOGICAL VARIABLES FOR YEAR 2013

Statistical index	PV output data	T	RH	HI	AP
$\bar{\rho}$	-0.36	0.47	-0.32	-0.58	0.06
STD $[\rho]$	0.17	0.41	0.38	0.29	0.17

As another case study to show the strong correlation between PV and wind power, one can refer to [10] which corroborates our findings in this section by analyzing the data of an actual PV system located in Salem, OR, USA [17].

Based on the aforementioned explanations, three DS are considered in this paper to construct a comprehensive model for wind and solar power forecast.

DS 1: Measured wind output power (kW) in each time interval, i.e. $X^{(1)}$; DS 2: Measured PV generated power (kW) in each time interval, i.e. $X^{(2)}$; DS 3: Calculated HI, i.e. $X^{(3)}$ in each time interval based on the measured RH (%) and T (°C) in that time interval according to (2).

The layout of the forecast system inputs and outputs is shown in Fig. 1.

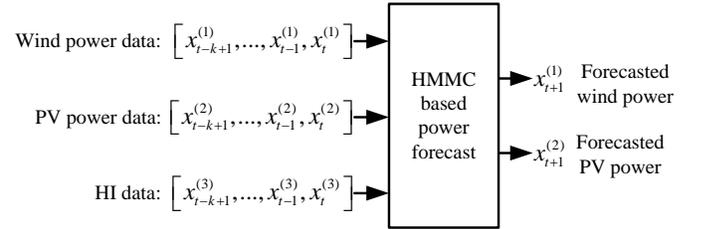


Fig. 1. Schematic of the forecast system and its input and output data

As shown in Fig. 1, the discretized PV/wind power and HI are considered as forecast unit inputs to predict the wind and PV power in the next time interval.

B. Data Discretization

Since 15-min ahead forecasting is considered in this paper, the DS continuous data should be quantized to form a discrete HMMC. Assume that there are s categorical continuous DS, i.e. wind output, PV power and HI. These data sequences can be modeled by a discrete HMMC including $N^{(r)}$ possible states for each DS, i.e. $\{1, 2, \dots, N^{(r)}\}$, $r = 1, \dots, s = 3$. Each DS, i.e. $X^{(r)} = \{x_t^{(r)}\}$, $1 \leq t \leq T$ is discretized based on the following states definition.

$$x_t^{(r)} = \left[\tilde{x}_i^{(r)}, \tilde{x}_{i+1}^{(r)} \right], i \in \{1, \dots, N^{(r)}\}, r = 1, \dots, s \quad (5)$$

It should be noted that $\tilde{x}_1^{(r)}$ and $\tilde{x}_{N+1}^{(r)}$ are minimum and maximum values observed in the r -th DS, respectively. It is

stated that the r -th DS is at i -th state when the measured value of the variable related to the r -th DS at time t , $x_t^{(r)}$ lies in the i -th interval, i.e. $\tilde{x}_i^{(r)} \leq x_t^{(r)} \leq \tilde{x}_{i+1}^{(r)}$.

The first step to design a HMMC-based forecast model is data discretization, which is carried out by determining appropriate intervals as indicated in (5). The selection of interval boundaries, i.e. $\tilde{x}_i^{(r)}$, is a challenging procedure because it affects the ability of HMMC to capture the PV/wind power dynamics. One approach attempts to divide the entire range of data to intervals with equal length, i.e. $|\tilde{x}_i^{(r)}| = \Delta x^{(r)}; i = 1, \dots, N^{(r)}; r = 1, \dots, s$ where $\Delta x^{(r)}$ is the equal length for all the states related to the r -th DS. This approach can be simply implemented. However, due to its drawbacks which has been mentioned in [4], in this paper, the continuous data of three DS, i.e. wind power, PV output and HI is discretized in such a way that an equal time of staying is achieved for all the states. Therefore, the following criteria should be satisfied [18].

$$\delta_i^{(r)} = \delta^{(r)}, i = 1, \dots, N^{(r)}, r = 1, \dots, s \quad (6)$$

$\delta^{(r)}$ is the equal time of staying for the r -th DS states, which will be determined optimally in section IV. In this paper, $x_i^{(r)}$ intervals are determined by solving (6) numerically after defining an appropriate value for $\delta^{(r)}$.

After the state design step, the representative value for each state is determined so that the following mean square error is minimized.

$$\hat{x}_i^{(r)} = \arg \min_x (x - x_i^{(r)})^2, x_i^{(r)} \in [\tilde{x}_i^{(r)}, \tilde{x}_{i+1}^{(r)}], 1 \leq t \leq T \quad (7)$$

After designing the states, the discrete HMMC-based model can be developed which is explained in the next subsection.

C. HMMC Model Development

It is assumed that the state probability distribution of a sequence in the next time interval depends on the state probability distribution of all the sequences (including the mentioned one) at the present time and $k - 1$ previous temporal intervals. HMMC is applied to model the transition among states in subsequent time intervals, as follows [19]:

$$\pi_{i+1}^{(r)} = \sum_{j=1}^{s=3} \sum_{i=1}^k \lambda_{i,j}^{(r)} Q_{i,j}^{(r)} \pi_{i+1-i}^{(r)}, r = 1, 2, 3 \quad (8)$$

The constraints of $\sum_{j=1}^{s=3} \sum_{i=1}^k \lambda_{i,j}^{(r)} = 1$ and $\lambda_{i,j}^{(r)} \geq 0$ ($1 \leq i \leq k, 1 \leq r, j \leq 3$) should be satisfied about λ elements.

The elements of $Q_{i,j}^{(r)}$ and $\lambda_{i,j}^{(r)}$ should be estimated to model the DS and their relationship by using HMMC considering the observation window with the length of k . $Q_{i,j}^{(r)}$ is estimated by normalizing the matrix of $F_{i,j}^{(r)}$ over each column, where $F_{i,j}^{(r)}$ is the transition frequency from the states in the j -th DS occurring at i previous temporal steps to the ones in the r -th DS. Clearly $Q^{(r)}$ is a $N^{(r)} \times N^{(r)}$ matrix.

In addition, the parameter λ should be estimated to develop the model. The HMMC converges to the stationary distribution, π , regardless of the initial distribution function as follows [18].

$$\lim_{t \rightarrow \infty} \pi_t^{(r)} = \pi^{(r)} \quad (9)$$

It should be noted that $\tilde{\pi}^{(r)}$ can be estimated by computing the proportion of the occurrence of each state in the r -th sequence.

Considering (8) and (9), the following equation can be written:

$$\tilde{\pi}^{(r)} \approx \sum_{j=1}^{s=3} \sum_{i=1}^k \lambda_{i,j}^{(r)} Q_{i,j}^{(r)} \tilde{\pi}^{(r)}, r = 1, \dots, s \quad (10)$$

Parameter λ should be determined in such a way that the developed model shows the long-term pattern of the data samples. Therefore, the following linear optimization problem is formed to estimate the parameter λ .

$$\begin{aligned} \text{Min}_{\lambda_{i,j}^{(r)}} & \left| \tilde{\pi}^{(r)} - \sum_{j=1}^{s=3} \sum_{i=1}^k \lambda_{i,j}^{(r)} Q_{i,j}^{(r)} \tilde{\pi}^{(r)} \right| \\ \text{subject to:} & \sum_{j=1}^{s=3} \sum_{i=1}^k \lambda_{i,j}^{(r)} = 1 \\ & \lambda_{i,j}^{(r)} \geq 0 \quad (r, j = 1, 2, 3, i = 1, \dots, k) \end{aligned} \quad (11)$$

The objective function in (11) addresses the difference between the long-term pattern of DS achieved by data investigation, i.e. $\tilde{\pi}^{(r)}$ and the developed model, i.e.

$$\sum_{j=1}^{s=3} \sum_{i=1}^k \lambda_{i,j}^{(r)} Q_{i,j}^{(r)} \tilde{\pi}^{(r)}.$$

Equation (11) is solved in this paper by using the interior-point method implemented in the optimization toolbox of MATLAB software. The globally optimal solution is achieved by solving (11) because it is a linear convex optimization problem.

Considering the aforementioned procedure for HMMC development, λ and Q are two model parameters. Parameter Q reflects the short-term variation of a DS as it is calculated based on investigating transitions of states in subsequent time intervals. Referring to (10), parameter λ is calculated such that the developed model matches the long-term pattern of the DS. Therefore, λ reflects the long-term dynamic of inter- and intra-DS relations.

Remark: As discussed in this section, when $s > 1$ and $k > 1$, a HMMC is developed to consider the inter- and intra-relations among s series of data, where k samples of each of them are considered in each step. If $s = 1$ and $k > 1$, the developed model is called HMC in which only interdependency among data samples of a sequence is taken into consideration.

III. HMMC-BASED PV AND WIND POWER FORECAST

Once a HMMC-based forecast model is developed as explained in section II, it can be used to forecast the PV and wind outputs in the next time interval, i.e. 15 min ahead by using k last data samples observed in each DS according to the following equation:

$$\tilde{x}_{t+1}^{(r)} = \sum_{i=1}^k \sum_{j=1}^s \lambda_{i,j}^{(r)} [\text{column } c^{(r)} \text{ of } Q_{i,j}^{(r)}]^T [\hat{x}_1^{(r)} \dots \hat{x}_N^{(r)}] \quad (12)$$

Equation (12) calculates $\tilde{x}_{r+1}^{(r)}$, considering that the current state in the r -th DS is $c^{(r)}$, ($1 \leq c^{(r)} \leq N^{(r)}$).

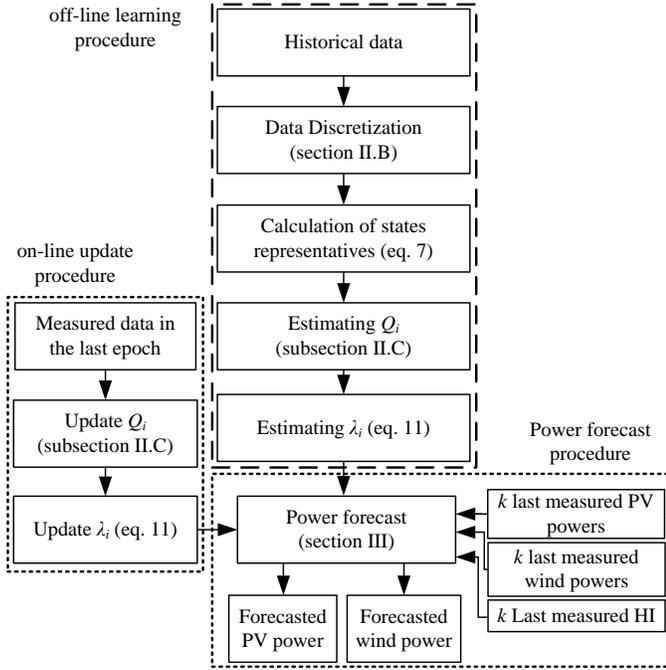


Fig. 2. Flowchart of proposed forecast method.

Moreover, the parameters of the forecast system, i.e. λ and Q can be updated dynamically by using the last measured data samples after each epoch, i.e. 4 h, so that the updated values of the parameters are used to forecast the power of RE units according to (12). It should be noted that the process of parameter updating can be done in parallel with the calculations related to power forecast. Therefore, it does not interfere with the calculations related to the power forecast procedure. The flowchart of the proposed forecast method and procedure of model parameter update is shown in Fig. 2.

IV. CASE STUDIES OF REAL WIND/PV POWER DATA

In this section, the PV/wind generated power, ambient temperature and relative humidity data measured and collected in Belgium are used to test the proposed method effectiveness for the 15-min ahead power forecast [14], [15]. The instantaneous values of the mentioned variables are recorded every 15 min for the years 2013 and 2014. The monitored capacity of PV and wind generated power ratings are 2915.88 MW and 592.20 MW, respectively. In this section, the data of the year 2013 is used to train the HMMC and the data of the year 2014 is used to test the performance of the proposed method. Separate HMMC based forecast models are developed for each daily 4-hour epoch in each month to capture the seasonality and daily non-stationarity of PV and wind power. In other words, time-adaptive sets of coefficients are used in the proposed forecast method such that different PV/wind power patterns can be accurately modeled. The PV output power of 8:00-16:00 is considered in this paper because the generated power by the PV plant is low/very fluctuating on the other time intervals in some seasons. The proposed

methods for HMMC-based model development and power forecast are implemented in MATLAB software.

A. Scenario 1: HMC-based PV/Wind Power Forecast

In this section, the PV output and wind generated power are modeled by constructing two separate forecast models based on the higher-order Markov Chain.

1) States Design

Wind generated power, PV output power and HI are used in this paper as three DS. These sequences include continuous data, which should be quantized by determining appropriate intervals as stated in (5). Since δ plays a major role in the states design step and it affects the forecasting accuracy, the state intervals are determined by solving (6) numerically for a predefined value of $\delta^{(r)}$. In order to find the best value for δ , the states are designed by using different values for $\delta^{(r)}$, i.e., $\delta^{(r)} = 1, 1.5, \dots, 5$ min, $r = 1, 2, 3$ and the best forecast model is picked. After the state design step, the probability transition matrix, Q , is determined according to the method described in section II.C by investigating the batch of data samples, i.e. wind/PV power and HI. To construct the model completely, the parameters λ should be calculated by solving the linear programming problem (11) via the MATLAB optimization toolbox.

$NRMSE$ of the forecasted PV and wind power is used in this paper as a useful index to evaluate the forecast method accuracy which can be calculated according to (13) based on the actual and estimated values of a DS, i.e. $x(t)$ and $x_{est}(t)$, respectively.

$$NRMSE = \frac{1}{x_{\max}} \sqrt{\frac{1}{T} \left(\sum_{t=1}^T [x(t) - x_{est}(t)]^2 \right)} \quad (13)$$

Since the aim of this paper is to forecast the PV and wind power, the optimal value of $\delta^{(r)}$ is chosen to reach the minimum of the $NRMSE$ values related to the wind and PV power. The $NRMSE$ related to HI forecast is not considered because HI prediction is not the objective of this paper.

Using the data of the year 2013, the power forecast model is developed. Different values of $NRMSE$ are shown in Fig. 3 versus the variation of δ . As shown in this figure, the $NRMSE$ related to wind (PV) power forecast decreases with increasing the parameter δ until it reaches $\delta^{(1)} = 2.5$ min ($\delta^{(2)} = 1.5$ min) while increasing the δ more than 2.5 min (1.5 min) leads to an increase in the $NRMSE$ index of wind (PV) power forecast. If a large value is selected for δ , the long-term trend of the output power variations is modeled and the short-term variations cannot be captured accurately which lead to a larger error. Conversely, if the value of δ is small, the forecast model concentrates on modeling the short-term variations and the long-term characteristic of output power dynamics cannot be modeled well, which results in a lower forecast accuracy. Therefore, the best values of δ are selected as $\delta^{(1)} = 2.5$ min and $\delta^{(2)} = 1.5$ min for wind and PV power data discretization, respectively based on Fig. 3 and they are applied to the data discretization as explained in section II.B. The results of Fig. 3 are achieved as the optimal value of $\delta^{(3)} = 3.5$ min is used for the HI data discretization.

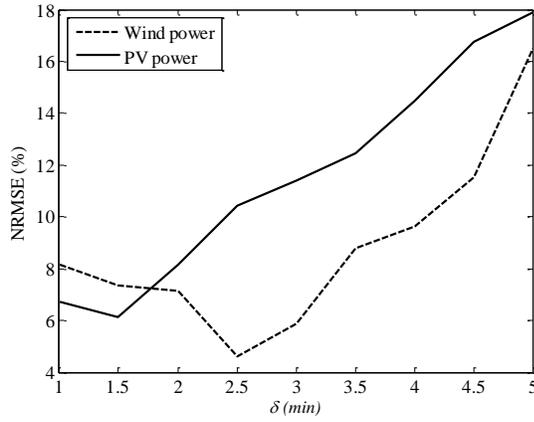


Fig. 3. $NRMSE$ as the function of the parameter δ

2) Power Forecast Model Development

In this section, the simplest form of the HMC, i.e. set of two separate higher-order Markov chains is used to forecast PV output and wind power individually. Only the temporal dynamics of PV/wind power are captured and the other interdependencies are neglected. The wind and PV power DS is discretized considering $\delta^{(1)} = 2.5$ min and $\delta^{(2)} = 1.5$ min, respectively which are related to minimum forecast error as shown in Fig. 3. With $k = 10$, the 10 last observed samples to forecast the next one are considered. State transition matrices are estimated according to the procedure stated in subsection II.C and λ is calculated by solving (11) (assuming $s = 1$ in this scenario) via the optimization toolbox of MATLAB software. It should be noted that there are two different λ parameters, each of them related to one DS as shown in Fig. 4. As shown in Fig. 4, the dependency between the samples in the first and second DS decreases with increasing the time lag parameter showing that the forecasted power of PV and wind depends more on the recently measured data sample than the previous ones.

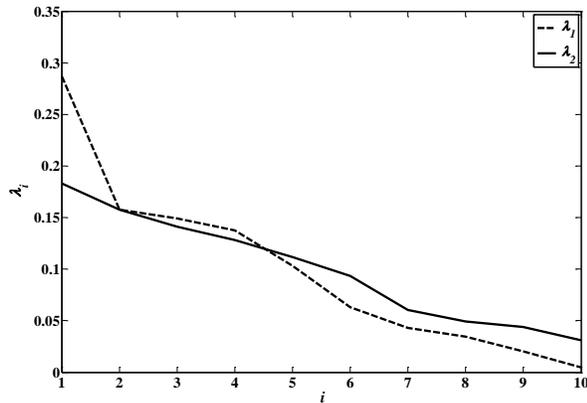


Fig. 4. Intra series lag coefficients in scenario 1

Remark:

The effect of k values on the HMMC accuracy is listed in Table III. In the second and fourth columns of Table III, the calculated $NRMSE$ related to PV and wind power are listed, respectively. In the third and fifth column of Table III, the accuracy improvement of forecast method, i.e. decrease in $NRMSE$ is calculated based on (14) and listed for

that value of k .

$$\Delta NRMSE(k) = NRMSE(k-1) - NRMSE(k), k > 1 \quad (14)$$

TABLE III
 $NRMSE$ FOR DIFFERENT VALUES OF k FOR PV AND WIND POWER FORECAST OVER THE DATA OF THE YEAR 2014

k	$NRMSE_{PV}$ (%)	$\Delta NRMSE_{PV}$ (%)	$NRMSE_{wind}$ (%)	$\Delta NRMSE_{wind}$ (%)
1	14.101	--	35.738	--
2	11.004	3.098	28.655	7.083
3	9.564	1.440	22.281	6.374
4	8.384	1.179	16.765	5.516
5	7.728	0.656	13.069	3.696
6	7.135	0.593	10.041	3.028
7	6.673	0.462	7.516	2.525
8	6.403	0.270	5.934	1.582
9	6.287	0.116	4.98	0.954
10	6.270	0.017	4.73	0.250
11	6.265	0.005	4.691	0.039
12	6.262	0.004	4.604	0.087
13	6.258	0.003	4.543	0.061
14	6.256	0.003	4.492	0.051
15	6.254	0.002	4.460	0.032

B. Scenario 2: HMMC based PV/Wind Power Forecast Model

In this section, the inter-dependencies of the PV output, wind generated power and HI variable are captured by constructing the HMMC-based forecast model. In this scenario, it is assumed that the HI data samples are provided from the year 2013 onward and they can be used in the training procedure. Considering $k = 10$, the state transition probability matrices, i.e. $Q_{i,j}^{(r)}$ and lag inter-sequence coefficients, i.e. $\lambda_{i,j}^{(r)}$ are estimated according to the approach proposed in subsection II.C to construct the power forecast model.

Data of PV/wind power and HI in the year 2013 is used to develop the forecast model while the data measured in the year 2014 is used to evaluate the performance of the developed model. A useful parameter to reflect the performance of the proposed method for PV/wind power forecast is the correlation index between the forecasted values and their respective measured ones, which can be calculated according to (1). The correlation indices related to PV and wind power in each epoch are listed in Table IV. Using scenario 1, the correlation indices vary in the interval of [83.54%, 84.27%] and [71.73%, 78.61%] for PV and wind power forecast, respectively while they lie in the margins of [96.12%, 98.06%] and [88.47%, 92.41%], respectively in the case of applying scenario 2. The stronger relation is seen between the measured and forecasted values of the generation when the relation between PV and wind output power is taken into consideration in scenario 2.

The median of the correlation indices are achieved as 97.58% and 91.62% by applying scenario 2 for the PV and wind power forecast, respectively, showing that frequently accurate forecasts are achieved for the set of PV and wind systems.

As shown in Table IV, the PV output power is taken into consideration in the time interval of 8:00-16:00. Therefore, the

correlation between the actual and forecasted values of PV power is not calculated in the other time intervals.

TABLE IV
CORRELATION INDICES BETWEEN FORECASTED AND ACTUAL VALUES OF PV AND WIND POWER

Time	PV output power		Wind output power	
	Scenario 1	Scenario 2	Scenario 1	Scenario 2
00:00- 3:59	--	--	72.48 %	88.86 %
4:00- 7:59	--	--	74.11 %	90.64 %
8:00- 11:59	83.54 %	96.12 %	78.61 %	91.92 %
12:00- 15:59	84.27 %	98.06 %	77.83 %	92.41 %
16:00- 19:59	--	--	75.42 %	89.75 %
20:00- 23:59	--	--	71.73 %	88.47 %

To show the scattering of forecast errors, the cumulative distribution function of the *NRMSE* related to the PV and wind forecast procedure is used as a good index, calculated according to the following equation:

$$CDF(x) = \int_{E=-\infty}^{E=x} \text{Prob.}(NRMSE = E) dE \quad (15)$$

where the *NRMSE* is the forecast error calculated by (13) and *Prob. (.)* is the probability function.

Regarding the CDF of PV forecast error shown in Fig. 5, almost 76% of the annual PV forecast errors have an *NRMSE* lower than 4% and 95% of them have an *NRMSE* lower than 4.5%. Moreover, as shown in Fig. 5, 90% of the wind forecast errors have an *NRMSE* less than 3.5%. Fig. 5 depicts the low dispersion of the PV and wind forecast errors in different daily times, months and weather conditions.

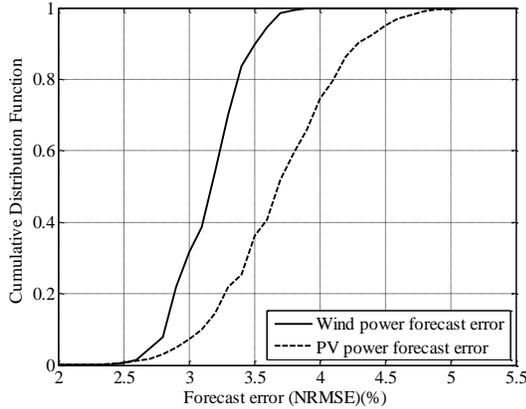


Fig. 5. Cumulative distribution of the PV and wind power forecast error

The hourly values of PV and wind forecasted *NRMSE*s are statistically analyzed and their mean and standard deviation are shown in Fig. 6. It should be noted that 24 points show the hourly mean and standard deviation of wind forecast *NRMSE* and 8 points indicate hourly mean and standard deviation of the PV forecast *NRMSE* related to 8:00-16:00. As shown in Fig. 6, the wind power can be forecasted by using the proposed method with a lower error and lower standard deviation compared to the PV power prediction. However, both of them have reasonable values of *NRMSE* showing that the proposed method can be effectively used in the PV and wind power forecast.

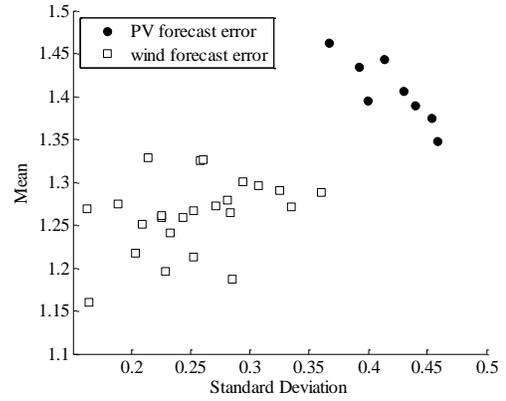


Fig. 6. Hourly mean and standard deviation of PV/wind power forecast error

C. Scenario 3: Input Vector Modification

In this section, the effect of increasing the dimension of the input vector of the forecast algorithm is investigated. At the first stage, the variations of PV output and wind generated power are modeled by constructing the simplest form of the HMMC-based framework, which uses just the PV and wind power samples as its input. In each of the next stages, one of the other variables, i.e. ambient temperature, relative humidity, atmospheric pressure and heat index, is added to the list of input variables to investigate their effect on enhancing the accuracy of the forecast model. It should be noted that the variables are added in each step based on their priority calculated in Table I and Table II.

Data of the year 2013 is used to develop the forecast model while the data measured in the year 2014 is used to evaluate the performance of the developed model. As listed in Table V, yearly values of $NRMSE_{PV}$ and $NRMSE_{wind}$ will reflect the performance of the proposed method for the PV/wind power forecast. As listed in Table V, adding the additional variables to the forecast method will enhance the PV/wind forecasting accuracy because the *NRMSE* values regarding steps 1-4 are less than that of the base case, i.e. considering PV/wind power as input variables. However, the HI is the most effective additional variable, which can be considered in the input vector in addition to PV and wind power.

TABLE V
NRMSE FOR DIFFERENT VARIABLES ADDED TO LIST OF INPUT OVER THE DATA OF THE YEAR 2014

Step	List of input variables	$NRMSE_{PV}$	$NRMSE_{wind}$
Base-case	PV power and wind power	6.10	4.17
1	PV power, wind power and ambient temperature	5.31	3.49
2	PV power, wind power and relative humidity	5.14	3.71
3	PV power, wind power and atmospheric pressure	6.02	3.97
4	PV power, wind power and heat index	3.90	2.59

To show the proposed framework ability to work with higher order input the largest size of input vector, i.e. a vector of PV and wind power, heat index, temperature, relative humidity and atmospheric pressure, is used as input to the forecast algorithm which leads to the $NRMSE_{PV}=3.17$ and $NRMSE_{wind}=2.21$.

D. Scenario 4: Forecast Model Update

In scenario 2, the forecast model is built based on the complete training set of the data, i.e. data of PV and wind output power and HI in the year 2013. To show the capability of the model parameters update, in this scenario, it is assumed that the HI data samples are not available in the year 2013 and the process of recording HI values was initiated from the commencement of the year 2014. This means that the constructed model is a HMMC with two DS, i.e. PV and wind output power based on data of the year 2013. The third DS, i.e. the HI values set, is added to the list of prediction model inputs after model development. This means that the constructed model should be expanded such that the third DS can be used as an additional predictor. Moreover, the model parameters should be updated as HI data arrives. In this scenario, the proposed model coefficients, i.e. Q and λ are updated at the end of each epoch of the year 2014 when a new set of data samples, i.e. PV and wind output power and HI become available. In other words, the proposed model is sequentially updated and the PV/wind forecast is carried out based on the updated version of the forecast unit.

In Fig. 7, $NRMSE_{PV}$ and $NRMSE_{wind}$ versus time are shown in the year 2014. In this scenario, it is assumed that there is no batch of HI historical data in the year 2013. Therefore, the forecast error is high at the beginning of 2014 ($NRMSE_{wind} = 4.37\%$ and $NRMSE_{PV} = 5.34\%$) as the power forecast is carried out by using the forecast model which is developed based on only two DS, i.e. wind power and PV generation in the year 2013. The forecast model accuracy will be enhanced over time as more HI samples are provided. Another reason for decreasing forecast error is that some of the characteristics of PV and wind units vary over time and some peripheral changes occurred which are taken into consideration by updating the forecast model parameters as newly measured PV and wind power is available.

As shown in Fig. 7, $NRMSE_{PV}$ decreases from 5.34% to 3.90% (26.96% reduction in error) by considering the update capability of the proposed method to add the HI values, which are provided from the beginning of 2014 onwards to the predictor input list. The other error index, $NRMSE_{wind}$ decreases from 4.37% to 2.59% (40.73% reduction in error).

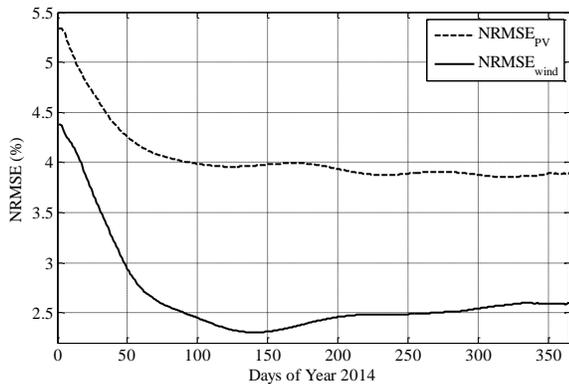


Fig. 7. Total $NRMSE$ related to scenario 4

To compare the performance of the proposed method to forecast the PV/wind power, some newly proposed methods are applied to the PV and generated power data samples [11],

[12]. AR models are adopted from [12] by considering one regime, and then the support of the Gaussian distribution is truncated into $[0, P_{max}]$. Specifically, the order of the AR models is determined by using the partial autocorrelation functions of the PV/wind power time series [20]. The SVM-enhanced Markov model is also provided by integrating the forecast done by the SVM into the first order Markov models as explained in [11].

The $NRMSE$ values related to PV and wind power forecast are listed in Table VI. The naïve forecast model is applied as a benchmark. It is the estimating technique in which the last period measurements are used as the next period forecast, without adjusting them or attempting to establish causal factors. It is used only for comparison with the forecasts generated by the more sophisticated techniques [21]. As shown in this table, the adaptable HMMC-based method outperforms the other ones, which reflects its capability in modeling the dynamic variations, and stochastic nature of the output power of PV and wind plants.

TABLE VI
 $NRMSE$ OF DIFFERENT METHODS FOR PV AND WIND POWER FORECAST OVER THE DATA OF THE YEAR 2014 IN SCENARIO 4

Forecast Model	Error (%)	
	$NRMSE_{PV}$	$NRMSE_{wind}$
Naïve forecast model	15.19	17.73
HMMC	3.90	2.59
SVM-enhanced MC	4.16	3.81
HMC (scenario 1)	6.08	4.64
Auto regression based power forecast	5.11	4.71

E. Scenario 5: Incomplete Batch of Historical Data

In scenarios 2 and 3, the forecast model is developed based on a complete batch of historical data of PV/wind power and HI. In some real cases, a number of samples are not measured accurately or they are not transmitted/recorded successfully. Moreover, the data preprocessing may omit some data samples when they are detected as bad data samples. In conventional model construction methods, the procedure of data recovery should be carried out to make the set of historical data ready for model construction. However, the proposed method can be applied to the incomplete set of historical data and no missed data detection/recovery procedure is needed to prepare the data samples.

In scenario 5, the performance of the proposed method to capture the PV/wind power variations from the incomplete set of historical data is evaluated.

In scenario 5, 5% of the randomly selected data samples of PV/wind power and HI are excluded from the historical data set. Based on the incomplete batch data, the forecast model is developed as stated in section II and its performance index, i.e. $NRMSE$ is calculated as stated in (13). It should be noted that these steps are repeated 1000 times and the mean of the $NRMSE$ values are calculated as
$$\overline{NRMSE} = \frac{1}{1000} \sum_{i=1}^{1000} NRMSE$$

which are listed in Table VII.

The mentioned procedure is also repeated for 10% and 15% of the data points drop of the original data set and the mean of the error results are listed in Table VII. As shown in this table, the proposed method outperforms the other ones when applied

on the incomplete set of historical data. This issue reflects its capability in modeling the dynamic variations, and the stochastic nature of the output power of PV and wind plants.

TABLE VII

NRMSE RELATED TO USE OF INCOMPLETE DATA SET FOR PV AND WIND POWER FORECAST OVER THE DATA OF THE YEAR 2014 IN SCENARIO 5

Percent of missed data (%)	Mean of Error (%), Over 1000 repetitions	HMMC	SVM-enhanced MC	Auto regression based power forecast
5	$NRMSE_{PV}$	4.18	6.25	8.64
	$NRMSE_{wind}$	2.74	4.28	7.69
10	$NRMSE_{PV}$	4.56	9.95	11.64
	$NRMSE_{wind}$	3.69	5.98	9.59
15	$NRMSE_{PV}$	5.50	17.54	20.59
	$NRMSE_{wind}$	3.89	9.91	17.49

V. CONCLUSION

In this paper, the higher-order multivariate Markov Chain was developed to forecast the PV and wind generated power. The time-adaptive stochastic correlation between the wind and PV output power was taken into consideration in the proposed method, which works based on the observation window of the last measured data samples to follow the pattern of PV/wind power variations. Moreover, to enhance the forecast accuracy, two issues are considered: the HI index was utilized as an additional meteorological variable and the procedure of dynamic update of the model parameters was applied to the PV/wind power forecast. Compared to scenario 1, in which the PV and wind power interdependency is not considered, the HMMC-based method models the correlation between PV and wind output power and HI and leads to higher forecast accuracy (scenario 2). The other advantage of applying the proposed framework to forecast the PV/wind power is that the dimension of the input vector of the forecast algorithm can be increased in a flexible manner to reach better performance of the forecast algorithm. This issue is simulated in scenario 3 with adding one of the variables, i.e. ambient temperature, relative humidity, atmospheric pressure and heat index, in different steps to the list of forecast algorithm input. Another advantage of using the proposed method is its adaptability, which was simulated in scenario 4 such that lower forecast error is achieved, compared to the condition in which the forecast model parameters cannot be updated dynamically (scenario 2). Finally, scenario 4 is the result of the proposed method, which matches well with the measurement data recorded from the actual PV and wind plants. To make the proposed method more applicable to real conditions in scenario 5, the proposed method was applied to the incomplete set of historical data to show the proposed method capability to model the PV/wind power variation in the case some samples are missed.

References

[1] A. Bonfiglio, M. Brignone, F. Delfino, R. Procopio, "Optimal control and operation of grid-connected photovoltaic production units for voltage support in medium-voltage networks," *IEEE Trans. Sustain. Energy*, vol.5, no.1, pp.254-263, Jan. 2014.

[2] P. García, C.A. García, L.M. Fernández, F. Llorens, F. Jurado, "ANFIS-Based Control of a Grid-Connected Hybrid System Integrating Renewable Energies, Hydrogen and Batteries," *IEEE Trans. Ind. Inf.*, vol. 10, no. 2, pp. 1107-1117, May 2014.

[3] N. Etherden, V. Vyatkin, M. H. J. Bollen, "Virtual Power Plant for Grid Services Using IEC 61850," *IEEE Trans. Ind. Inf.*, vol. 12, no. 1, pp. 437-447, Feb. 2016.

[4] M. J. Sanjari; H. B. Gooi, "Probabilistic Forecast of PV Power Generation Based on Higher-Order Markov Chain," in *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2942 – 2952, 2017.

[5] M.J. Sanjari, G.B. Gharehpetian, "Game Theoretic Approach to Cooperative Control of Distributed Energy Resources in Isolated Microgrid Considering Voltage and Frequency Stability", *Neural Computing and Applications*, vol. 25, issue 2, pp 343-351, 2014.

[6] M. Marinelli, P. Maule, A. N. Hahmann, O. Gehrke, P. B. Nørgrd and N. A. Cutululis, "Wind and Photovoltaic Large-Scale Regional Models for Hourly Production Evaluation," *IEEE Trans. Sustain. Energy*, vol. 6, no. 3, pp. 916-923, July 2015.

[7] M. Asensio and J. Contreras, "Stochastic Unit Commitment in Isolated Systems With Renewable Penetration Under CVaR Assessment," *IEEE Transactions on Smart Grid*, vol. 7, no. 3, pp. 1356-1367, May 2016.

[8] K.D. Orwig, M.L. Ahlstrom, V. Banunarayanan, J. Sharp, J.M. Wilczak, J. Freedman, S.E. Haupt, J. Cline, O. Bartholomy, H.F. Hamann, B.M. Hodge, C. Finley, D. Nakafuji, J. Peterson, D. Maggio, M. Marquis, "Recent Trends in Variable Generation Forecasting and Its Value to the Power System," *IEEE Trans. Sustain. Energy*, vol.6, no.3, pp.924-933, Jul. 2015.

[9] Y. M. Atwa, E. F. El-Saadany, "Annual Wind Speed Estimation Utilizing Constrained Grey Predictor," *IEEE Trans. Energy Con.*, vol. 24, no. 2, pp. 548-550, June 2009.

[10] J. Liu, W. Fang, X. Zhang, C. Yang, "An improved photovoltaic power forecasting model with the assistance of aerosol index data," *IEEE Trans. Sustain. Energy*, vol.6, no.2, pp.434-442, Apr. 2015.

[11] L. Yang, M. He, J. Zhang and V. Vittal, "Support-Vector-Machine-Enhanced Markov Model for Short-Term Wind Power Forecast," *IEEE Trans. Sustain. Energy*, vol. 6, no. 3, pp. 791-799, Jul. 2015.

[12] P. Pinson and H. Madsen, "Adaptive modelling and forecasting of offshore wind power fluctuations with Markov-switching autoregressive models," *J. Forecasting*, vol. 31, no. 4, pp. 281-313, 2012.

[13] H. S. Jang, K. Y. Bae, H. S. Park and D. K. Sung, "Solar Power Prediction Based on Satellite Images and Support Vector Machine," *IEEE Transactions on Sustainable Energy*, vol. 7, no. 3, pp. 1255-1263, July 2016.

[14] Wind and Solar-PV Power Generation Data [Online]. <http://www.elia.be/en/about-elia>

[15] Historical Weather Type Data [Online]. <http://mesonet.agron.iastate.edu/>

[16] R.B. Stull, "Meteorology for scientists and engineers" second edition, Brooks Cole, 1999.

[17] Solar Radiation Monitoring Laboratory, University of Oregon. (2013). Salem Photovoltaic Data [Online]. Available: <http://solardat.uoregon.edu>

[18] Q. Zhang and S. A. Kassam, "Finite-state Markov model for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 11, pp. 1688-1692, Nov. 1999.

[19] Wai-Ki Ching, Michael K. Ng, Eric S. Fung, "Higher-order multivariate Markov chains and their applications", *Linear Algebra and its Applications*, vol. 428, no. 2-3, pp 492-507, 2008.

[20] M. H. Hayes, *Statistical Digital Signal Processing, and Modeling*. Hoboken, NJ, USA: Wiley, 1996.

[21] L. M. Carvalho, J. Teixeira and M. Matos, "Modeling wind power uncertainty in the long-term operational reserve adequacy assessment: A comparative analysis between the Naïve and the ARIMA forecasting models," *2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, Beijing, 2016, pp. 1-6.