Solar Irradiance Forecasting in Remote Microgrids Using Markov Switching Model

Ayush Shakya, *Student Member, IEEE*, Semhar Michael, Christopher Saunders, Douglas Armstrong, Prakash Pandey, Santosh Chalise, and Reinaldo Tonkoski, *Member, IEEE*

***Abstract*—Photovoltaic (PV) systems integration is increasingly being used to reduce fuel consumption in diesel-based remote microgrids. However, uncertainty and low correlation of PV power availability with load reduces the benefits of PV integration. These challenges can be handled by introducing reserve. However, this leads to increased operational cost. Solar irradiance forecasting helps to reduce reserve requirement, thereby improving the utilization of PV energy. This paper presents a new solar irradiance forecasting method for remote microgrids based on the Markov switching model. This method uses locally available data to predict one-day-ahead solar irradiance for scheduling energy resources in remote microgrids. The model considers past solar irradiance data, clear sky irradiance, and Fourier basis expansions to create linear models for three regimes or states: high, medium, and low energy regimes for days corresponding to sunny, mildly cloudy, and extremely cloudy days, respectively. The case study for Brookings, SD, USA, discussed in this paper, resulted in an average mean absolute percentage error of 31.8% for five years, from 2001 to 2005, with higher errors during summer months than during winter months.**

***Index Terms*—Clear sky irradiance (CSI), fourier basis expansion, global horizontal irradiance (GHI), markov switching model (MSM), solar irradiance forecasting.**

NOMENCLATURE

t	Time.
$CSI(t)$	Clear Sky Irradiance at time t.
AM	Air Mass.
T	Time period.
MSM	Markov Switching Model.
BIC	Bayesian Information Criterion.
BIC_K	BIC value of a model with K states.
$PRESS$	Predictive Residual Sum of Squares.
SE	Standard Error.
$RMSE$	Root Mean Square Error.
$MAPE$	Mean Absolute Percentage Error.
GHI	Global Horizontal Irradiance.
P_{PV}	Forecasted PV system power output.
P_{Rated}	Rated capacity of the PV array.
f_{PV}	PV system derating factor.
G_t	Forecasted irradiance at time t.
$G_{t,STC}$	Irradiance at Standard Test Conditions (STC).
K	Number of states or regimes.
k	State.
M_K	Model with K states.
$L(\boldsymbol{y}\|\hat{\theta}, M_k)$	Likelihood value of the model M_k at parameter estimate of $\hat{\theta}$.
d_K	Number of parameters for a model with K states.
p	Length of Fourier basis expansion for yearly variation in irradiance.
q	Length of Fourier basis expansion for daily variation in irradiance.
ϵ_k	Random noise component.
β	Coefficient of Fourier basis expansion.
$\beta_{k,1}$	Coefficient of $CSI(t)$ for state k.
$\beta_{k,1i}$	Coefficient of Fourier basis expansion for yearly variation in irradiance for state k.
$\beta_{k,2j}$	Coefficient of Fourier basis expansion for daily variation in irradiance for state k.
$\phi_m(t)$	Basis.
$\boldsymbol{y}$	Observations from univariate time series.
$\boldsymbol{A}$	Probability transition matrix.
$\boldsymbol{X}$	Design matrix.
$\boldsymbol{S}$	Vector of states.
θ	Vector of parameters.
S_t	State sequence.
$\hat{S}_t$	Estimated state sequence.
p_{ij}	Probability of transition from state i to state j.
e_t	Forecasting error.
y_t	Actual irradiance.
$\hat{y}_t$	Forecasted irradiance.
pbm	Parametric bootstrapping method.
fdh	Finite-difference approximation of the Hessian matrix.
$\tilde{\boldsymbol{\Psi}}_i$	Bootstrap estimated parameters.
$\hat{\boldsymbol{\Psi}}_i$	Original estimated parameter value.
B	Number of bootstrap samples.

Manuscript received March 19, 2016; revised August 6, 2016 and September 17, 2016; accepted November 3, 2016. Date of publication November 16, 2016; date of current version June 17, 2017. This work was supported by the SDBoR Competitive Research Grant Program 2015 and 2016 and in part by the Ph.D. Program, Department of Electrical Engineering, South Dakota State University. Paper no. TSTE-00214-2016.

A. Shakya, P. Pandey, S. Chalise, and R. Tonkoski are with the Department of Electrical Engineering and Computer Science, South Dakota State University, Brookings, SD 57007 USA (e-mail: ayush.shakya@sdstate.edu; prakash.pandey@sdstate.edu; santosh.chalise@sdstate.edu; reinaldo.tonkoski@sdstate.edu).

S. Michael, C. Saunders, and D. Armstrong are with the Department of Mathematics and Statistics, South Dakota State University, Brookings, SD 57007 USA (e-mail: semhar.michael@sdstate.edu; christopher.saunders@sdstate.edu; doug.armstrong@sdstate.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSTE.2016.2629974

I. Introduction

PHOTOVOLTAIC (PV) systems utilization is increasing rapidly in diesel-based remote microgrids to displace fuel, which is the main operational cost [1]. However, the load does not always correlate with PV power availability, and the full potential of PV cannot be utilized. A case study in Canada [2] showed that the installation of PV systems which supplied about 14% of the yearly energy needs of a PV-diesel hybrid microgrid resulted in 5% reduction in fuel consumption. A similar study showed 3.3% reduction in fuel consumption with the addition of PV systems which supplied about 14% of the yearly energy needs [3]. There are two basic challenges of PV power: uncertainty and variability. The challenges of integrating PV systems into the microgrids can be handled by using large reserves to ensure reliable operation. However, this large reserve requirement actually increases the operational cost because the diesel generators are now forced to operate at lower efficiencies to provide the required reserve margin. In such cases, PV forecasting is needed to schedule the energy resources and improve the PV energy utilization. With PV forecasting, the Energy Management Systems (EMS) can better manage the dispatchable resources to improve energy efficiency in the microgrid. The integration of renewable energy sources in a microgrid system with Distributed Intelligent EMS (DIEMS) was presented in [4]. Forecasted renewable energy was used in the energy management. With the use of forecasts in the EMS, the operating cost of the system decreased, while the battery lifetime improved. The total cost savings for the case studied was estimated to be 24.83% with DIEMS. Proper scheduling of microgrid components, such as battery, generator, and PV power output forecasting in the scheduling stage, is the key to reducing the cost of operation of remote microgrid [5].

PV power output is directly related to solar irradiance at ground level, so solar irradiance forecasting is the basis of solar power forecasting. The first and most important step in most of the solar power prediction systems is the forecasting of Global Horizontal Irradiance (GHI) [6], which is the total amount of irradiance falling on a surface horizontal to the surface of the earth.

Most of the solar irradiance forecasting techniques require various information, like weather and satellite data which may not be readily available in remote areas, or may require additional sensors. This paper proposes a novel solar irradiance forecasting technique for remote microgrids using the Markov Switching Model (MSM), which makes use of past solar irradiance data at the location of the microgrid for day-ahead forecasting. It considers the Clear Sky Irradiance (CSI) and the Fourier basis expansions to create linear models for three regimes or states: high, medium, and low energy regimes corresponding to sunny, mildly cloudy, and extremely cloudy days, respectively. The historical irradiance data is required only for fitting the model. Once it is fit, any other information like past irradiance data or weather and satellite data are not needed. So, this method is well-suited for remote microgrids, which do not have access to communication infrastructure.

This paper is organized as follows: Section II presents a brief description of various solar irradiance forecasting methods. Section III details the solar irradiance forecasting method using the MSM. Section IV discusses case studies, including error analysis. The conclusions are stated in Section V.

II. Solar Irradiance Forecasting Methods

The previous work on solar irradiance forecasting focused on the development of different models and forecasting approaches on the basis of time horizon, and applications [7]. The forecasting approaches may include Numerical Weather Prediction (NWP) models for obtaining forecasts from 6 hours to several days [8]; time series models for short time horizon of about 5 minutes to 6 hours based on on-site measurements [8], [9]; statistical models for short time window of about 6 hours based on satellite data, and cloud motion images [10].

Solar irradiance forecasting methods can be broadly categorized as physical, statistical, Artificial Intelligence (AI), and hybrid methods [11]. Physical methods use PV models, NWP models and satellite images to predict solar irradiance, and PV generation [12], [13]. There are basically three physical methods: NWP, Total Sky Imagery (TSI), and satellite imagery methods for cloud observations [8], [14]. These methods utilize information such as weather data and satellite images for forecasting, which may not be readily available in remote areas.

Statistical methods are based on time series models. These methods primarily depend on historical data to train the statistical models for solar irradiance forecasting. Statistical methods include persistence forecasting, time series, and Model Output Statistics (MOS) methods. Persistence forecasting is a simple and effective forecasting method for short-term forecasting of about 1 hour [15], based on the current solar power output. As the name implies, it supposes that solar irradiance at time $t+1$ is best predicted by its value at time t i.e. $\hat{X}_{t+1} = X_t$. The time series method is used for long-term solar irradiance forecasting, which uses various models such as Autoregressive (AR) [16], Autoregressive Moving Average (ARMA) [15], and Autoregressive Integrated Moving Average (ARIMA) models [17]. A large amount of data is required to develop statistical correlations and ensure accuracy. This implies that statistical methods are not available immediately for locations without any prior local measurements. The MOS method interprets numerical model output and produces specific forecasts.

AI methods use advanced techniques like Artificial Neural Networks (ANNs) whereas hybrid methods are combination of physical, statistical, and AI methods [11]. Hybrid methods benefit from various forecasting models. Similar to statistical methods, these methods require large datasets.

Some of the solar irradiance forecasting methods available in the literature are shown in Table I. Most of these methods required meteorological data as input, which may not be readily available in remote areas. The errors in the meteorological data affect the quality of solar irradiance forecasts. In addition, weather information is usually descriptive but not quantified [18].

Among various models, MSMs have been popular in the wind power community, dealing with forecasting of output power on a large-scale offshore wind farms [25]. It was first introduced in Econometrics by [26]. The models have been commonly used in the field of speech recognition [27], and computational

TABLE I
DIFFERENT SOLAR IRRADIANCE FORECASTING TECHNIQUES

Forecasting methods	Data Required	Normalized Root Mean Squre Error (NRMSE) (%)	Mean Absolute Percentage Error (MAPE) (%)	Limitations
Physical	Weather data	11% to 17% [19]	–	Meteorological data are required which may not be available in remote locations.
		36% [20]	–	European Centre for Medium-Range Weather Forecasts (ECMWFs) may not be available in remote locations.
Artificial Neural Network (ANN)	Insolation, temperature, relative humidity	–	15.93% to 17.33% using Feed-forward Neural Network (FFNN), 15.1% to 17.21% using Radial Basis Function Neural Network (RBNN) and 15.20% to 16.86% using Recurrent Neural Network (RNN) [21]	Meteorological data are required which may not be available in remote locations.
	Mean daily solar irradiance, mean daily relative humidity, mean daily wind speed and daily air temperature	–	8.29% to 10.8% on sunny days, 6.36% to 8.89% on cloudy days and 24.16% to 54.44% on rainy days [22]	
	Weather data	10% to 16% [19]	–	
Hybrid	Maximum temperature, probability of precipitation and solar irradiance of previous similar days	–	3.295% [23] for a year from May 2012 to April 2013	Weather data may not be readily available in remote locations.
	Weather data, historical hourly PV power output data, historical aerosol index value, historical temperature, humidity, wind speed data	–	7.65% [24]	

biology [28]. First and second order Markov chains have been used to model wind speed or wind vector [29]–[31]. A Markov switching autoregressive model is introduced in [32] to describe a long-time series data of wind speed measurement on daily and interannual bases. Similarly, a Markov switching autoregressive model with time-varying coefficients has been developed for modeling and forecasting the short-term variations in wind generation [33]. A Markov switching vector autoregressive model has also been used for simulating wind speed and direction as in [34]. To the best knowledge of the authors of this article, the model has not been implemented for solar irradiance forecasting.

III. SOLAR IRRADIANCE FORECASTING USING MARKOV SWITCHING MODEL (MSM)

Three components are used for developing the MSM for forecasting solar irradiance: CSI, Fourier basis expansions, and past historical irradiance data. CSI and Fourier basis expansions are computed mathematically. The development of the model using these components is discussed in this section. Fig. 1 shows the block diagram for solar irradiance forecasting. It is explained in the following sections:

A. Clear Sky Irradiance (*CSI*)

CSI is the irradiance under a cloudless sky and can be expressed as the intensity of direct sunlight. It is defined as a function of the Air Mass (AM) using the following experimentally-based equation [35]:

$$CSI = 1367(0.7)^{AM^{0.678}}\ \mathrm{W/m^2} \tag{1}$$

where 0.678 and 0.7 are empirical numerical constants and AM is air mass, which is the length of the path that the solar radiation

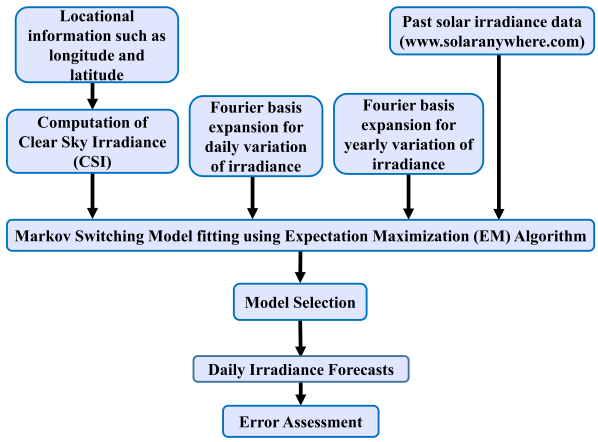

Fig. 1. Block diagram for solar irradiance forecasting using MSM.

covers to travel to the surface of the earth, as well as a function of the time of year and day and the latitude.

B. Fourier Basis Expansion

Fourier basis functions are used for the representation of periodic variation in the observed data. In this paper, the functions are used to explain the periodic characteristics of solar irradiance. The Fourier series [36] is given by

$$\hat{x}(t) = \sum_{m=0}^{n} \beta_m \phi_m(t) \tag{2}$$

That is,

$$\hat{x}(t) = \beta_0 + \beta_1 \sin \omega t + \beta_2 \cos \omega t + \beta_3 \sin 2\omega t + \beta_4 \cos 2\omega t + \beta_5 \sin 3\omega t + \beta_6 \cos 3\omega t \ldots \tag{3}$$

where $\phi_m(t)$ represents basis defined by $\phi_0(t) = 1$, $\phi_{2r-1}(t) = \sin(r\omega t)$ and $\phi_{2r}(t) = \cos(r\omega t)$.

The basis is periodic and the parameter ω defines the time period as $T = \frac{2\pi}{\omega}$. There are two variations in solar irradiance: daily and yearly variations. Solar irradiance is high during midday and low in the evening. This daily variation is represented using nine Fourier basis terms. The value of the coefficients are determined using model fitting. Similarly, the yearly variation of solar irradiance is represented by seven Fourier basis terms.

C. Markov Switching Model (MSM)

The MSM, also known as the regime switching model, is a mixture model in which the components that generate observations depend upon hidden states. These hidden states are considered to be a Markov process. The marginal distribution of the hidden Markov model can be a mixture of any known probability distribution. MSM involves multiple structures (or equations) representing different regimes to describe nonlinear time series behavior. This model allows switching between the regimes governed by a probabilistic process. Therefore, it is able to capture complex dynamic patterns. Generally, the unobservable state variable controls switching between the regimes, which commonly are assumed to follow Markov chain.

Let $\boldsymbol{y} = y_1, \ldots, y_T$ be observations from univariate time series of length T and let $\boldsymbol{X} = \boldsymbol{X}_1, \ldots, \boldsymbol{X}_T$ denote the matrix of covariates. Then, based on the MSM model, at any time point, the observations are distributed according to a mixture model with K components (hidden states) $\boldsymbol{S} = (S_1, \ldots, S_T)$ where $S_t \in \{1, \ldots, K\}$ for $t = 1, \ldots, T$. The joint likelihood for $y_1, \ldots, y_T$ and the hidden states $S_1, \ldots, S_T$, given the covariates $\boldsymbol{X}$ and the vector of parameters $\boldsymbol{\theta}$ is given as

$$L(\boldsymbol{Y}, \boldsymbol{S}|\boldsymbol{X}, \boldsymbol{\theta}) = P(S_1 = i|\boldsymbol{X}_1) f_{S_1}(y_1|\boldsymbol{X}_1) \times \prod_{t=2}^{T} p_{\mathrm{ij}}(\boldsymbol{X}_t) f_{S_t}(y_t|\boldsymbol{X}_t) \quad (4)$$

where $f_{S_t}(y_t|\boldsymbol{X}_t)$ describes a vector of conditional densities $P(y_t|S_t = j, \boldsymbol{X}_t)$ for $j = 1, \ldots, K$. A matrix $\boldsymbol{A}$ is constructed from the transition probabilities $p_{\mathrm{ij}}(\boldsymbol{X}_t) = P(S_{t+1} = j|S_t = i; \boldsymbol{X}_t)$, representing the transition probability from state i to state j given covariate $\boldsymbol{X}_t$. The parameters are estimated through maximum likelihood approach. This is done either through the Expectation-Maximization (EM) algorithm [37] or numerical optimization of the corresponding log-likelihood using methods such as the Newton-Raphson. The EM algorithm is a maximum likelihood estimation technique that consists of two steps, namely, E-step and M-step. In the E-step, the sequence of states are estimated based on the model basis set and model parameters set. The M-step performs maximization of the complete-data likelihood resulting in adjusted model parameters. E-step and M-step are iterated until the given convergence criterion is met. Specifically, in the E-step, the best sequence of the states S_t for $t = 1, \ldots, T$ is estimated using forward-backward algorithm. This needs estimated parameters from the previous step of the iteration. At the M-step, the expected complete-data log likelihood function is then maximized with respect to the parameters given $\boldsymbol{y}$, $\boldsymbol{X}$ and estimated state sequence, $\hat{S}_t$. These two steps are iterated until a prespecified convergence criterion is met. One common criterion for convergence is specifying a threshold for the relative difference between consecutive log-likelihood values ($i.e.$, $\frac{\log L(\boldsymbol{Y}, \boldsymbol{S}^{(t)}|\boldsymbol{X}, \boldsymbol{\theta}^{(b)}) - L(\boldsymbol{Y}, \boldsymbol{S}^{(b-1)}|\boldsymbol{X}, \boldsymbol{\theta}^{(b-1)})}{|L(\boldsymbol{Y}, \boldsymbol{S}^{(b-1)}|\boldsymbol{X}, \boldsymbol{\theta}^{(b-1)})|} < \Delta$).

MSM is used to model the nonlinear pattern of solar irradiance. The irradiance at a time t is considered to have a latent state. The latent state variable affects the value of the solar irradiance at time t. The matrix of covariates ($\boldsymbol{X}$) consists of CSI, yearly Fourier component ϕ_{1i}, and daily Fourier component ϕ_{2j}. The linear model for each state k is written as

$$y_t(S_t = k) = \beta_0 + \beta_{k,1} CSI(t) + \sum_{i=1}^{p} \beta_{k,1i} \phi_{1i}(t) + \sum_{j=1}^{q} \beta_{k,2j} \phi_{2j}(t) + \epsilon_k(t) \quad (5)$$

where $CSI(t)$ is the clear sky irradiance at time t with the corresponding coefficient $\beta_{k,1}$. β_0 is the intercept of the model and $\beta_{k,1i}$ represents the coefficient corresponding to the i^{th} component of the Fourier basis function for yearly variation of solar irradiance at state k. Similarly, $\beta_{k,2j}$ represents the coefficient corresponding to the j^{th} component of the Fourier basis function for daily variation of solar irradiance at state k. The length of Fourier basis expansions for yearly and daily irradiance are p and q, respectively. $\epsilon_k(t)$ is a random noise component and is assumed to be a normally distributed random variable with zero mean and constant variance, σ_k^2. The probability transition matrix is another part of the MSM that needs to be estimated. It gives the probability of movement between the regimes. A system with K regimes will have a probability transition matrix of order $K \times K$, where a number of rows represents the current or "from" state and columns represent the final or "to" state. Probabilities close to one in the diagonal represent stable states. The equation given below shows the probability transition matrix of 3×3 order:

$$\boldsymbol{A} = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix}$$

The calculation of the elements of the probability transition matrix is done through maximum likelihood estimation using

$$p_{\mathrm{ij}} = \frac{Count(S_{t-1} = i \text{ and } S_t = j)}{Count(S_{t-1} = i)} \quad (6)$$

The estimated sequence of states obtained from the EM algorithm is used in the computation of (6) as the truth is unknown. The Fourier basis expansions and CSI values are used as covariates; and the coefficients and the probability transition matrix are estimated. The fitting of this model is done using the statistical R package, *MSwM* [38]. This results in $\hat{y}_t(S_t = k)$, the estimated solar irradiance at time $t = 1, \ldots, T$ for state $k = 1, \ldots, K$. For forecasting solar irradiance, we first use estimated parameters to find the predicted irradiance for each state. Then the best sequence is selected by looking at the first few hours of the day. The details of the forecasting procedure is given in Section IV.

The next point to consider is the assessment of the quality of the day ahead prediction using this model and the errors associated with the estimated parameters. This problem is approached in two ways. One approach is looking at the Standard Errors (SEs) of the estimated parameters. There are several methods proposed for finding them [39]. In this paper, the SEs are obtained using parametric bootstrapping method (*pbm*). The *MSwM* package also returns the SEs using the finite-difference approximation of the Hessian matrix; henceforth, referred to as *fdh*. However, as reported in [39], SEs based on *fdh* tend to be too small, which is confirmed for the case study presented in this paper. The estimated parameters and the corresponding SEs for the case study can be found in Appendix V. The second approach is by looking at the Predictive Residual Sum of Squares (PRESS). In Section IV, a modified PRESS statistic, the Root Mean Square Error (RMSE) is reported. This statistic is calculated by building the model using historical data up to a given time point and then forecasts for the next time point. Hence, the technique is evaluated on a set of data that was not used to build the model. The results show good performance of the forecasting technique.

D. Model Selection

Model selection is an important step to consider in model fitting. In this context, the optimal number of states, K, is one of the main parameters that must to be determined. In addition, note that different assumptions can be made to reduce the number of parameters to be estimated from data. For example, we can assume the yearly coefficients to be constant over different states and similarly to the other covariates. Bayesian Information Criterion (BIC) [40] is an index utilized to choose between the models being considered. The following equation is used for computing BIC for the a model with K states (M_K),

$$BIC_K = -2\log L(\boldsymbol{y}|\hat{\theta}, M_K) + d_K log N \tag{7}$$

where $log\ L(\boldsymbol{y}|\hat{\theta}, M_k)$ indicates the log likelihood value of the model at estimated parameter $\hat{\theta}$, K represents the number of states in the model, N is the sample size, and d_K is the number of parameters. For a model with varying parameters for all states, (8) estimates the number of parameters.

$$d_K = 2K + K(p+q) + K^2 \tag{8}$$

This information criterion takes into consideration the number of parameters and goodness of fit of the models. The criterion penalizes the number of parameters required for the goodness of fit of the model. Between the given models, the model with the smallest BIC value is chosen as the best fit with fewer parameters.

E. Model Validation and Computation of Errors

The model is validated with historical irradiance data in the location where forecasting is to be performed. Examples of validation are presented in Section IV. Two standard error metrics reported in the literature include RMSE and Mean Absolute Percentage Error (MAPE). RMSE gives the square root of the sum of the squares of the difference between hourly forecasted irradiance and hourly actual irradiance in a one day period. It can also be calculated for a whole year. RMSE gives more weight to the errors with larger values than to the errors with smaller values. MAPE gives the sum of the absolute values of the differences between hourly forecasted irradiance and hourly actual irradiance divided by the actual irradiance values. It measures size of the error in terms of percentage and gives the same weight to all the errors.

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T} e_t^2} \tag{9}$$

$$MAPE = \frac{1}{T}\sum_{t=1}^{T} \left|\frac{e_t}{y_t}\right| \times 100\% \tag{10}$$

where $e_t = y_t - \hat{y}_t$ with y_t and $\hat{y}_t$ representing the actual and forecasted irradiance at time t, respectively. Here, T is the number of hours in a day or in a year.

After obtaining forecasted solar irradiance, the forecasted PV system power output [41], P_{PV} in kW can be obtained by using

$$P_{\mathrm{PV}} = P_{\mathrm{Rated}} f_{\mathrm{PV}} \frac{G_t}{G_{t,STC}} \tag{11}$$

where P_{Rated} is the rated capacity of the PV array in the microgrid in kW, f_{PV} is the PV system derating factor, G_t is the forecasted irradiance at time t in W/m^2, and $G_{t,STC}$ is the irradiance at Standard Test Conditions (STC) in W/m^2. Here the effect of temperature on the PV array is not considered. The forecasted PV power output is used in the EMS of the remote microgrid.

IV. Case Study

A. Dataset

Solar radiation data for the past 15 years were obtained from SolarAnywhere [42], a web-based service that offers hourly GHI, Direct Normal Irradiance (DNI), and Diffuse Horizontal Irradiance (DHI) for various locations within the USA and which dates from 1998 to 2013. The website considers ground based solar resource data as well as satellite based solar resource data and reduces uncertainty in solar resource assessment [43]. The ground-measured data is collected using ground stations and then the solar resource is evaluated via long term satellite-based irradiance measurement. The area resolution of the solar irradiance data was 100 km^2 and the time resolution was 1 hour. Only GHI was used for data analysis.

A case study was considered for Brookings, SD which has a latitude of 44.3° N and longitude of 96.8° W. Fig. 2 shows variation in solar irradiance from 2001 to 2011. The solar irradiance for these ten years illustrate the periodic variation in solar irradiance in the yearly data. Thus, the Fourier basis expansion is used to capture the yearly periodic variation in solar irradiance.

The variation in solar irradiance for the year 2001 is shown in Fig. 3. The daily peaks are high during the summer season and low during the winter season. Fig. 3 also shows the CSI for the same year in the background. As expected, the CSI is always greater than the observed solar irradiance. However, the

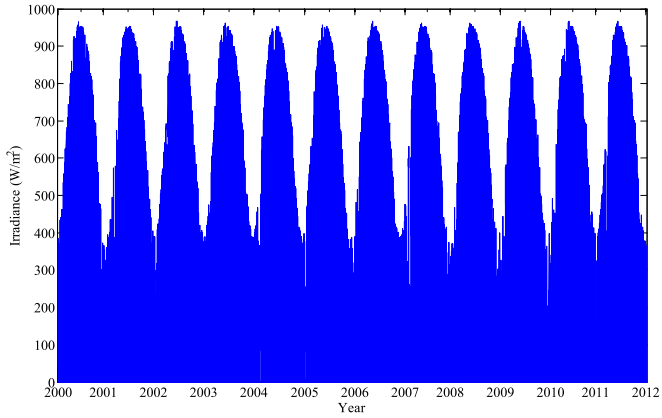

Fig. 2. Solar irradiance for years from 2001 to 2010.

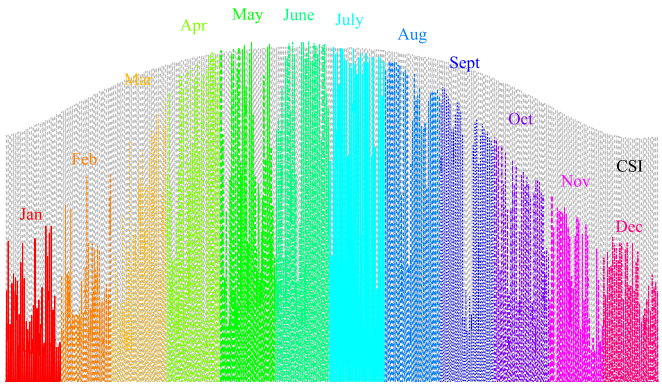

Fig. 3. CSI for a year along with actual irradiance for the year 2001.

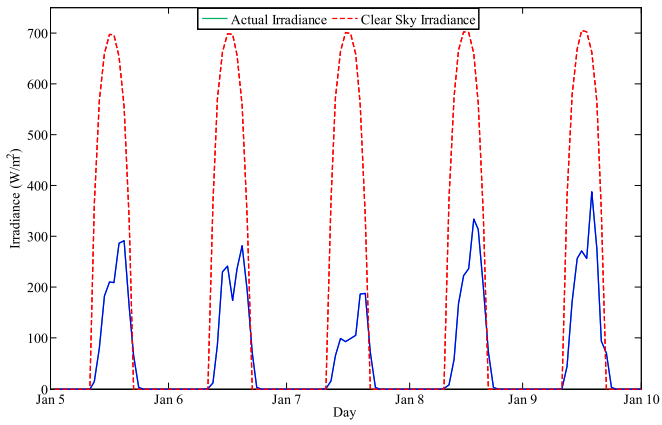

Fig. 4. CSI and actual solar irradiance from January 6 through January 10, 2001.

same seasonal variation is observed in CSI as the actual solar irradiance.

Taking a closer look, Fig. 4 depicts the variation in solar irradiance from January 6 to January 10 in the year 2001. It can be observed that the variation in solar irradiance is periodic in nature which implies that solar irradiance is at a maximum during the day, a minimum during the morning and evening, and zero at night. Thus, Fourier basis expansion can be used to capture this periodic nature of irradiance.

TABLE II
BIC SCORES FOR FIT MODELS

Yearly Fourier Coefficients	Daily Fourier Coefficients	K = 2	K = 3
Constant	Constant	157,869	153,992
Constant	Varying	155,663	149,715
Varying	Constant	157,464	153,294
Varying	Varying	151,878	145,037

Only hourly solar irradiance was considered for forecasting purpose although there may exist intra-hour variations in the irradiance data. In a microgrid system, the intra-hour variations can be handled by using reserves. The hourly irradiance forecasts are used in the scheduling stage of the EMS in the microgrid system. The deviations in the forecasted and actual irradiance in real-time is compensated in the dispatching stage using batteries and gensets [44].

B. Calculation of CSI

For solar irradiance forecasting, the CSI must be calculated using (1). Fig. 4 illustrates the variation in the calculated CSI for January 6 through January 10, 2001. It can be observed that the CSI is zero at the beginning of the day and increases to the maximum value during the day and then decreases during the evening, illustrating daily variation of the irradiance. The cycle repeats for each day. CSI is similar for all years except leap years, which have one extra day. The figure also shows the variation in both actual irradiance and the CSI from January 6 through January 10. The difference exists because the CSI does not consider various components such as clouds, humidity, pollution, etc., which have major roles in the absorption of solar irradiance.

C. MSM Fit and Forecasting

For fitting and testing the models, a free software environment for statistical computation and graphics, R version 3.1.3 was used on a computer with 64-bit Windows 7 operating system with 8 GB RAM and Intel(R) Core(TM)2 Quad CPU at 2.83 GHz. GHI data from years 2000 to 2002 were considered. Eight models with $K = 2$ and $K = 3$ were fitted and tested. Table II compares the models using BIC scores. The model that minimizes BIC score was selected. The model with $K = 3$ states, and varying daily and yearly Fourier coefficients was found to have the lowest BIC score of 145,037. Thus, this model was selected. The three states were designated as high, medium, and low energy state, each corresponding to sunny, mildly cloudy, and extremely cloudy days, respectively.

The probability transition matrix for this model was calculated to be:

$$\begin{bmatrix} 0.854 & 0.115 & 0.000 \\ 0.146 & 0.735 & 0.105 \\ 0.000 & 0.120 & 0.895 \end{bmatrix}$$

In the calculated matrix, the diagonal elements are closer to 1 indicating that the states are stable. The matrix shows that

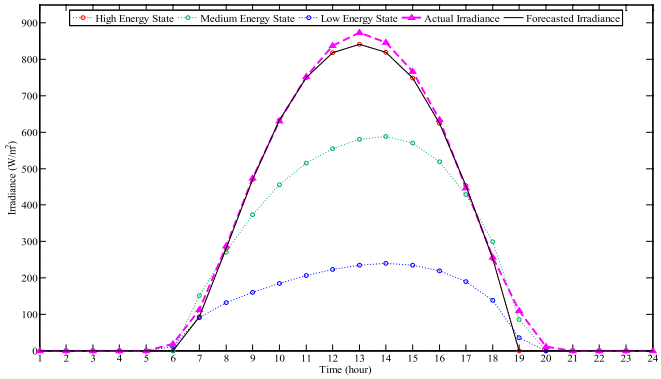

Fig. 5. Irradiance variation for August 15, 2011.

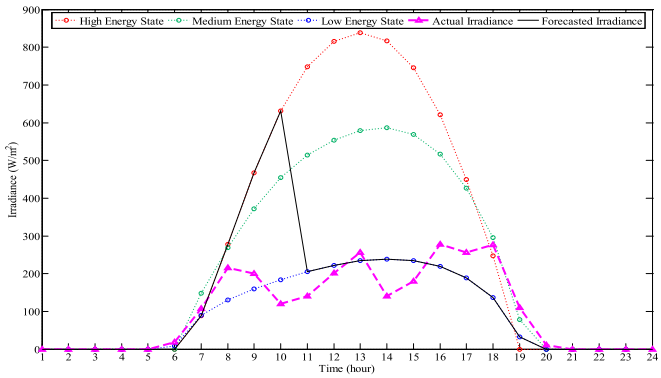

Fig. 6. Irradiance variation for August 16, 2011.

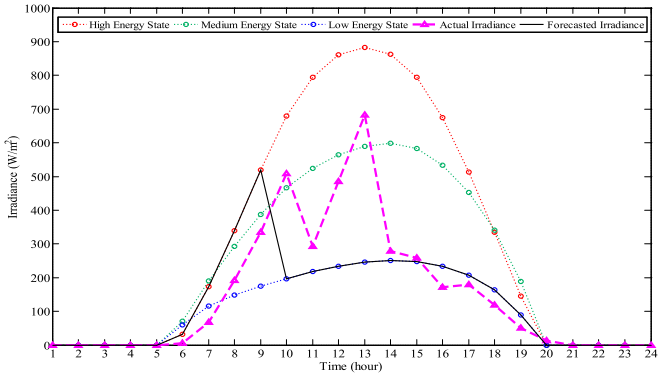

Fig. 7. Irradiance variation for July 23, 2012.

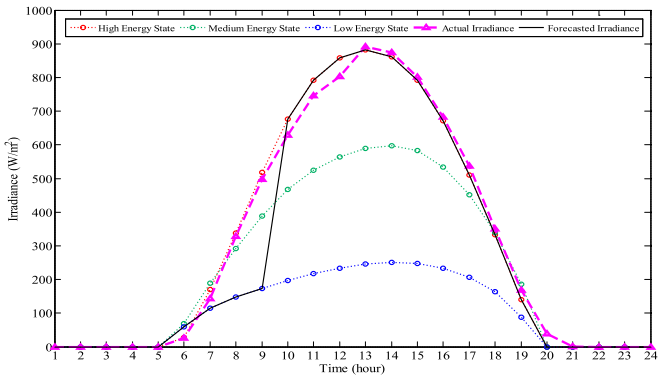

Fig. 8. Irradiance variation for July 24, 2012.

the probability of solar irradiance remaining in the high energy regime is 0.854, which is higher than the probability of transitioning from the high energy regime to both the medium and low energy regimes. In fact, the diagonal elements are higher than the off-diagonal elements, indicating that irradiance is more likely to maintain regime for the next hour. For example, if the morning irradiance is high, then it is predicted to remain high for the rest of the day.

Forecasting of solar irradiance is done by comparing the first four hours of actual irradiance data in the morning with each of the three energy states. The least square error method is used for this purpose. The regime with the least sum of squares of errors will be the forecasted irradiance for the rest of the day. However, for the first four hours of forecasting, the energy state on the previous day is used. For example, if the previous day's solar irradiance forecasting state was in the high energy regime, then the first four hours irradiance on the next day is predicted to remain in the high energy regime.

As an example for the process of the forecasting algorithm, solar irradiance for August 16, 2011 was considered. Fig. 5 shows the irradiance prediction for August 15, 2011. The red, green, and blue dotted curves indicate irradiance for high, medium, and low energy states, respectively. The magenta (dashed) and black (solid) curves indicate actual and forecasted irradiance, respectively. During this day, high energy state was observed. As a result, the forecasted irradiance for the first four hours of August 16 remained in the high energy state as shown in Fig. 6. The actual irradiance plot for the first four hours of the day was closer to the low energy state, which caused the forecasted irradiance to remain in the low energy state. However, the irradiance for the first four hours in the morning was forecasted on the basis of the state of August 15. The RMSE, and MAPE for this day were calculated to be 127.2 W/m^2 and 47.5%, respectively. The RMSE and MAPE over a one year period (8760 hours) for the year 2011 were computed to be 106.8 W/m^2 and 31.9%, respectively.

Similarly, as a second example, consider the forescasts for July 24, 2012, which is a leap year. The first four hours of irradiance were predicted to be in the low energy regime as the irradiance on the previous day, i.e. July 23, 2012, was predicted to lie in low energy regime as shown in Fig. 7. Fig. 8 shows the plot of solar irradiance for July 24. The actual irradiance plot for the first four hours of the day was closer to the high energy state, which caused the forecasted irradiance to remain in the high energy state. The RMSE, and MAPE for this case were calculated to be 79.4 W/m^2 and 21.7%, respectively. The RMSE and MAPE over a one year period (8784 hours) was computed to be 108.4 W/m^2 and 31.9%, respectively.

D. Error Assessment

The boxplots of monthly RMSE for the year 2001 is shown in Fig. 9. The RMSE is comparatively higher during the summer

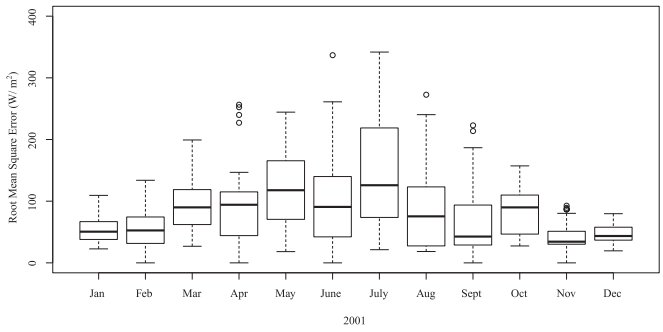

Fig. 9. Boxplots of monthly RMSE for the year 2001.

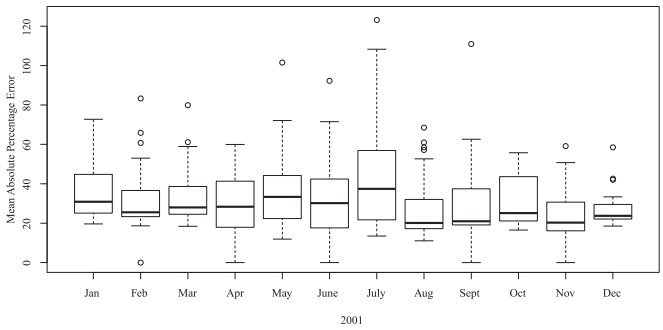

Fig. 10. Boxplots of monthly MAPE for the year 2001.

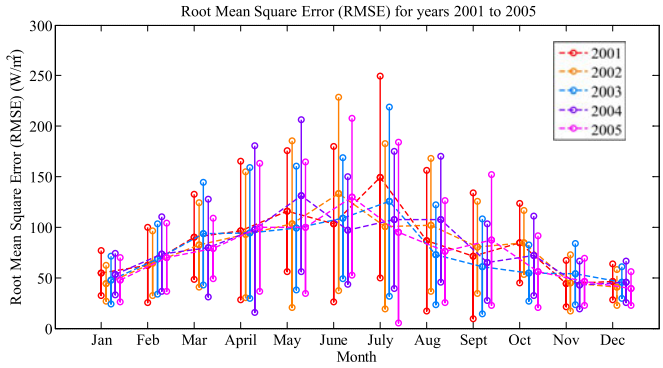

Fig. 11. RMSE for years 2001 to 2005.

season, since total daylight hours are higher during this season. Higher daylight hours imply a higher number of forecasts and hence higher error. The highest and lowest RMSE were calculated to be 341.8 W/m^2 and 17.5 W/m^2, respectively. The RMSE throughout the whole year was computed to be 105.8 W/m^2. However, monthly MAPE, being relative error metric, is similar during all the seasons as illustrated in Fig. 10. The MAPE throughout the whole year was calculated to be 32.9%.

RMSE from 2001 to 2005 is shown in Fig. 11. Overall, the trend for all these years was similar with higher RMSE during summer seasons and lower RMSE during winter seasons. The upper point in each month indicates the monthly average RMSE plus one standard deviation, while the lower point indicates monthly average RMSE minus one standard deviation. Table III

TABLE III
RMSE FOR YEARS 2001 TO 2005

Year	RMSE throughout the year (W/m^2)	Maximum RMSE (W/m^2)	Minimum RMSE (W/m^2)	MAPE throughout the year (%)
2001	105.8	341.8 (July)	17.5 (June)	32.9
2002	102.9	360.5 (May)	12.4 (Nov)	31.6
2003	96.2	348.5 (July)	14.5 (July)	30.7
2004	100.0	365.8 (May)	14.2 (Dec)	32.3
2005	97.5	306.7 (July)	13.3 $Oct)	31.7

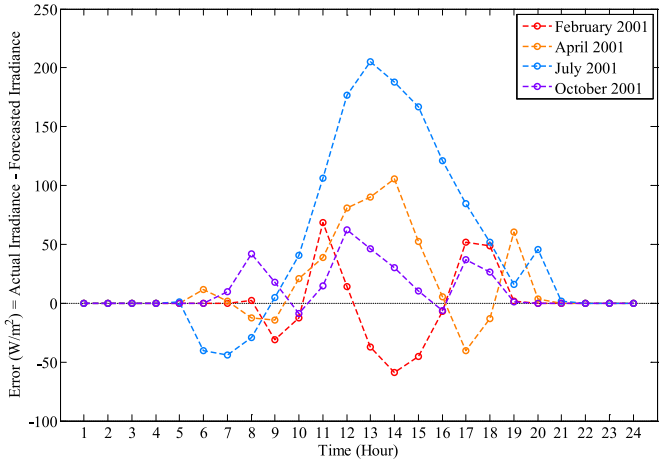

Fig. 12. Plot of difference between hourly actual and forecasted irradiance for four different months of year 2001.

shows RMSE throughout the year, maximum RMSE, minimum RMSE, and MAPE for the years 2001 to 2005.

The hourly error plots for four different months of a year are shown in Fig. 12 for four different seasons: February for winter, April for spring, July for summer and October for fall for the year 2001. The error is fairly high for the summer month. The model tends to overestimate irradiance in February and underestimate it in July. In future work, this trend among different months in a year can be taken into consideration to minimize the forecasting error. Likewise, in this paper, only Fourier basis expansions have been considered to represent periodicity in irradiance. However, in future studies, other basis expansions can be considered to find one that produces a lower error rate.

V. CONCLUSION

A novel solar irradiance forecasting technique using the Markov Switching Model has been developed. The model uses historical irradiance data and produces a day ahead forecast. MSM was used to capture dynamic behavior of solar irradiance by permitting switching between the states. Based on the rotation of the earth around the sun and its own axis, periodicity is observed in the data. This was captured by Fourier basis expansions. The deterministic irradiance value (CSI) and the Fourier basis expansions were used as independent variables, resulting in a fitted irradiance value for each state. A case study was conducted using readily available historical irradiance data for Brookings, SD. The best model, which was

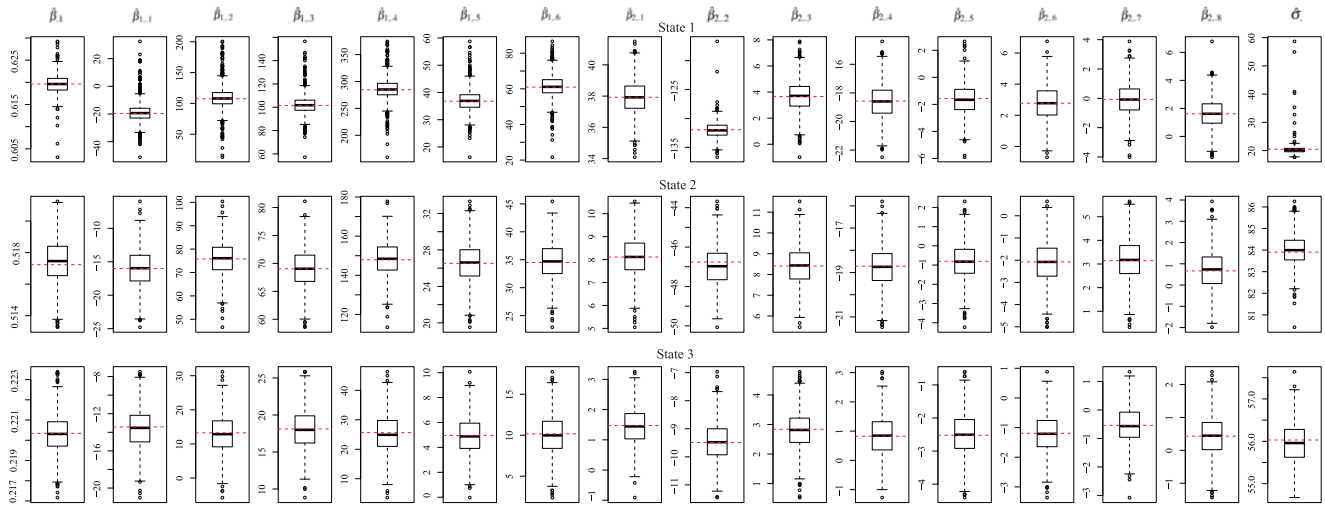

Fig. 13. The boxplots show the distribution of the bootstrap estimated parameters ($\tilde{\mathbf{\Psi}}_i$). The red dashed line indicates the original estimated parameter value ($\hat{\mathbf{\Psi}}_i$).

TABLE IV
ESTIMATED PARAMETERS OBTAINED BY FITTING MSM TO THE SOLAR IRRADIANCE DATA

	$\hat{\beta}_{.1}$	$\hat{\beta}_{1,.1}$	$\hat{\beta}_{1,.2}$	$\hat{\beta}_{1,.3}$	$\hat{\beta}_{1,.4}$	$\hat{\beta}_{1,.5}$	$\hat{\beta}_{1,.6}$	$\hat{\beta}_{2,.1}$
State 1	0.62 *(0.002)*	−19.66 *(7.44)*	107.53 *(19.70)*	101.55 *(8.48)*	284.59 *(21.20)*	36.79 *(4.06)*	60.95 *(6.51)*	37.92 *(1.07)*
State 2	0.52 *(0.01)*	−16.00 *(2.84)*	75.80 *(7.35)*	69.04 *(3.47)*	147.92 *(8.72)*	26.53 *(2.15)*	34.51 *(3.32)*	8.11 *(0.89)*
State 3	0.22 *(0.001)*	−13.42 *(2.16)*	13.25 *(5.58)*	18.11 *(2.66)*	25.65 *(6.52)*	4.96 *(1.54)*	10.16 *(2.41)*	1.47 *(0.61)*
	$\hat{\beta}_{2,.2}$	$\hat{\beta}_{2,.3}$	$\hat{\beta}_{2,.4}$	$\hat{\beta}_{2,.5}$	$\hat{\beta}_{2,.6}$	$\hat{\beta}_{2,.7}$	$\hat{\beta}_{2,.8}$	$\hat{\sigma}_.$
State 1	−131.95 *(1.46)*	3.65 *(1.16)*	−18.57 *(1.20)*	−1.57 *(1.14)*	2.77 *(1.15)*	−0.07 *(1.09)*	1.61 *(1.04)*	20.47*(2.14)*
State 2	−46.76 *(0.99)*	8.41 *(0.92)*	−18.71 *(0.90)*	−0.80 *(0.94)*	−2.08 *(0.93)*	3.14 *(0.88)*	0.67 *(0.95)*	83.91 *(0.70)*
State 3	−9.49 *(0.67)*	2.84 *(0.64)*	0.83 *(0.68)*	−2.52 *(0.63)*	−1.19 *(0.65)*	−0.54 *(0.62)*	0.43 *(0.61)*	56.03 *(0.47)*

Italicized Values in Parentheses Represent Corresponding SEs Estimated by *pbm*

selected based on BIC had three regimes: high, medium and low energy regimes for days, corresponding to sunny, mildly cloudy and extremely cloudy days, respectively. The MAPE for the case study in Brookings for years 2001 through 2005 ranged between 30.7%-32.9% indicating that the model is not biased towards the data used for fitting. The average MAPE throughout the five years, was calculated to be 31.8%. The RMSE was found to be higher during summer months as there are more hours of sunlight in the summer. This forecasting technique mostly applies to remote locations where sophisticated forecasting techniques such as those based on satellite images and numerical weather prediction are not options. This method uses publicly and freely available past irradiance data for model fitting and once fitting is done, any further information is not required. The fitted model can be used for forecasting a number of years.

APPENDIX

In Section III-C, the MSM model was presented and the fitting procedure was described. In this section, the estimated parameters of the model for the case study presented in Section IV is reported along with the corresponding standard errors. The standard errors are calculated using *pbm*. The following steps were used:

Step 1: Fit the MSM to the given data and find the estimated set of parameters, $\hat{\mathbf{\Psi}}$. This includes the coefficients of the linear model ($\hat{\boldsymbol{\beta}}_k$) and variance $\hat{\sigma}_k^2, \forall k = 1, \ldots, K$ and the probability transition matrix $\hat{\boldsymbol{A}}$.

Step 2: Using $\hat{\mathbf{\Psi}}$, generate new dataset of size N (same size as the original data). For this, using each linear equation, $\hat{y}_k$, is generated. Then a state sequence $\hat{S}$ is found using the probability transition matrix $\hat{\boldsymbol{A}}$ to obtain $\hat{y}$.

Step 3: Fit the MSM model to the data generated in Step 2 ($\hat{y}$) and record the estimated parameters, say $\tilde{\mathbf{\Psi}}_i$.

Step 4. Repeat Steps 3 and 4 B times, where B is the number of bootstrap samples. Then obtain $\tilde{\mathbf{\Psi}}_1, \ldots, \tilde{\mathbf{\Psi}}_B$.

Step 5. Calculate the standard deviation of the obtained estimated parameters.

A size $B = 1000$ bootstrap samples is used here. The side-by-side boxplots in Fig. 13 show how the bootstrap sample's estimated parameters are distributed around the original estimated parameter values ($\hat{\Psi}$). The results of the *pbm* are presented in Table IV. The italicized values within parentheses represent corresponding standard errors estimated by *pbm*. The coefficient of CSI seem to exhibit the least variability. The approximate 95% confidence interval constructed from these SEs will be $estimate \pm 1.96 \times SE$ [39].

REFERENCES

[1] A. Dehamna and P. Asmus, "Energy storage for microgrids," Navigant Research, *Tech. Rep.*, 2014. [Online]. Available: http://atargroup.com/blog/wp-content/uploads/2014/05/ESMG-14-Navigant-Research.pdf

[2] S. Pelland, D. Turcotte, G. Colgate, and A. Swingler, "Nemiah valley photovoltaic-diesel mini-grid: System performance and fuel saving based on one year of monitored data," *IEEE Trans. Sustain. Energy*, vol. 3, no. 1, pp. 167–175, Jan. 2012.

[3] R. Tonkoski, *Impact of High Penetration of Photovoltaics on Low Voltage Systems and Remedial Actions*. Saarbrücken, Germany: LAP LAMBERT Academic, 2014.

[4] S. Chakraborty, M. D. Weiss, and M. G. Simoes, "Distributed intelligent energy management system for a single-phase high-frequency AC microgrid," *IEEE Trans. Ind. Electron.*, vol. 54, no. 1, pp. 97–109, Feb. 2007.

[5] S. Chalise and R. Tonkoski, "Day ahead schedule of remote microgrids with renewable energy sources considering battery lifetime," in *Proc. 11th IEEE/IAS Int. Conf. Ind. Appl.*, 2014, pp. 1–5.

[6] E. Lorenz, T. Scheidsteger, J. Hurka, D. Heinemann, and C. Kurz, "Regional PV power prediction for improved grid integration," *Prog. Photovolt., Res. Appl.*, vol. 19, no. 7, pp. 757–771, 2011.

[7] L. Ciabattoni, G. Ippoliti, S. Longhi, M. Cavalletti, and M. Rocchetti, "Solar irradiation forecasting using RBF networks for PV systems with storage," in *Proc. IEEE Int. Conf. Ind. Technol.*, Mar. 2012, pp. 699–704.

[8] H. M. Diagne, D. Mathieu, P. Lauret, and M. David, "Solar irradiation forecasting: State-of-the-art and proposition for future developments for small-scale insular grids," in *Proc. World Renewable Energy Forum*, 2012, pp. 1–8.

[9] G. Reikard, "Predicting solar radiation at high resolutions: A comparison of time series forecasts," *Sol. Energy*, vol. 83, no. 3, pp. 342–349, 2009.

[10] E. Lorenz, A. Hammer, and D. Heinemann, "Short term forecasting of solar radiation based on satellite data," in *Proc. ISES Europe Sol. Congr.*, 2004, pp. 841–848.

[11] C. Wan, J. Zhao, Y. Song, Z. Xu, J. Lin, and Z. Hu, "Photovoltaic and solar power forecasting for smart grid energy management," *CSEE J. Power Energy Syst.*, vol. 1, no. 4, pp. 38–46, 2015.

[12] R. Perez *et al.*, "Comparison of numerical weather prediction solar irradiance forecasts in the US, Canada and Europe," *Sol. Energy*, vol. 94, pp. 305–326, 2013.

[13] E. Geraldi, F. Romano, and E. Ricciardelli, "An advanced model for the estimation of the surface solar irradiance under all atmospheric conditions using MSG/SEVIRI data," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 8, pp. 2934–2953, Aug. 2012.

[14] S. Pelland, J. Remund, J. Kleissl, T. Oozeki, and K. De Brabandere, "Photovoltaic and solar forecasting: State of the art," *Int. Energy Agency, Rep. IEA PVPS T14–01:2013*, 2013.

[15] R. Huang, T. Huang, R. Gadh, and N. Li, "Solar generation prediction using the ARMA model in a laboratory-level micro-grid," in *Proc. 3rd IEEE Int. Conf. Smart Grid Commun.*, 2012, pp. 528–533.

[16] P. Bacher, H. Madsen, and H. A. Nielsen, "Online short-term solar power forecasting," *Sol. Energy*, vol. 83, no. 10, pp. 1772–1783, 2009.

[17] R. Perdomo, E. Banguero, and G. Gordillo, "Statistical modeling for global solar radiation forecasting in Bogotá," in *Proc. 35th IEEE Photovolt. Spec. Conf.*, 2010, pp. 2374–2379.

[18] S. Jafarzadeh, M. S. Fadali, and C. Y. Evrenosoglu, "Solar power prediction using interval type-2 TSK modeling," *IEEE Trans. Sustain. Energy*, vol. 4, no. 2, pp. 333–339, Apr. 2013.

[19] Y. Huang, J. Lu, C. Liu, X. Xu, W. Wang, and X. Zhou, "Comparative study of power forecasting methods for PV stations," in *Proc. Int. Conf. Power Syst. Technol.*, 2010, pp. 1–6.

[20] E. Lorenz, J. Hurka, D. Heinemann, and H. G. Beyer, "Irradiance forecasting for the power prediction of grid-connected photovoltaic systems," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 2, no. 1, pp. 2–10, Mar. 2009.

[21] A. Yona, T. Senjyu, A. Y. Saber, T. Funabashi, H. Sekine, and C.-H. Kim, "Application of neural network to one-day-ahead 24 hours generating power forecasting for photovoltaic system," in *Proc. Int. Conf. Intell. Syst. Appl. Power Syst.*, 2007, pp. 1–6.

[22] C. Chen, S. Duan, T. Cai, and B. Liu, "Online 24-h solar power forecasting based on weather type classification using artificial neural network," *Sol. Energy*, vol. 85, no. 11, pp. 2856–2870, 2011.

[23] H. T. Yang, C. M. Huang, Y. C. Huang, and Y. S. Pai, "A weather-based hybrid method for 1-day ahead hourly forecasting of PV power output," *IEEE Trans. Sustain. Energy*, vol. 5, no. 3, pp. 917–926, Jul. 2014.

[24] J. Liu, W. Fang, X. Zhang, and 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.

[25] P. Trombe, Pierre-Julien and H. Madsen, "A general probabilistic forecasting framework for offshore wind power fluctuations," *Energies*, vol. 5, no. 3, pp. 621–657, 2012.

[26] J. D. Hamilton, "A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica, J. Econometric Soc.*, vol. 57, pp. 357–384, 1989.

[27] L. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

[28] R. Durbin, S. R. Eddy, A. Krogh, and G. Mitchison, *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[29] A. D. Sahin and Z. Sen, "First-order Markov chain approach to wind speed modelling," *J. Wind Eng. Ind. Aerodyn.*, vol. 89, no. 3, pp. 263–269, 2001.

[30] A. Shamshad, M. Bawadi, W. W. Hussin, T. Majid, and S. Sanusi, "First and second order Markov chain models for synthetic generation of wind speed time series," *Energy*, vol. 30, no. 5, pp. 693–708, 2005.

[31] F. Y. Ettoumi, H. Sauvageot, and A.-E.-H. Adane, "Statistical bivariate modelling of wind using first-order Markov chain and Weibull distribution," *Renewable Energy*, vol. 28, no. 11, pp. 1787–1802, 2003.

[32] P. Ailliot and V. Monbet, "Markov-switching autoregressive models for wind time series," *Environ. Model. Softw.*, vol. 30, pp. 92–101, 2012.

[33] P. Pinson, H. Madsen, P. E. Sørensen, and N. A. Cutululis, "Adaptive modelling of offshore wind power fluctuations," in *Proc. Nordic Wind Power Conf.*, 2007, pp. 1–9.

[34] K. Kazor and A. S. Hering, "Assessing the performance of model-based clustering methods in multivariate time series with application to identifying regional wind regimes," *J. Agricultural, Biol., Environ. Statist.*, vol. 20, no. 2, pp. 192–217, 2015.

[35] A. B. Meinel and M. P. Meinel, *Applied Solar Energy: An Introduction*. Reading, MA, USA: Addison-Wesley, 1976.

[36] J. O. Ramsay, *Functional Data Analysis*. Hoboken, NJ, USA: Wiley, 2006.

[37] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood for incomplete data via the EM algorithm (with discussion)," *J. Roy. Statist. Soc. B*, vol. 39, pp. 1–38, 1977.

[38] J. A. Sanchez-Espigares and A. Lopez-Moreno, *MSwM: Fitting Markov Switching Models, R package version 1.2*, 2014. [Online]. Available: https://CRAN.R-project.org/package=MSwM

[39] I. Visser, M. E. Raijmakers, and P. Molenaar, "Confidence intervals for hidden Markov model parameters," *Brit. J. Math. Statist. Psychol.*, vol. 53, no. 2, pp. 317–327, 2000.

[40] G. Schwarz, "Estimating the dimensions of a model," *Ann. Statist.*, vol. 6, pp. 461–464, 1978.

[41] HOMER Energy, *HOMER Pro Version 3.7 User Manual*, 2016. [Online]. Available: http://www.homerenergy.com/pdf/HOMERHelpManual.pdf, Accessed on: Nov. 14, 2016.

[42] Clean power research, "Solaranywhere," 2015. [Online]. Available: http://www.solaranywhere.com, Accessed on: Apr. 18, 2015.

[43] A. Kankiewicz, W. Kirkland, J. Dise, E. Wu, and R. Perez, "Solar 2014: Reducing solar project uncertainty with an optimized resource assessment tuning methodology, " in *Proc. ASES Annu. Conf.*, 2014, pp. 1–6.

[44] S. Chalise, J. Sternhagen, T. M. Hansen, and R. Tonkoski, "Energy management of remote microgrids considering battery lifetime," *Elect. J.*, vol. 29, no. 6, pp. 1–10, 2016.

Ayush Shakya (S'14) received the Bachelor's degree in electrical engineering from the Institute of Engineering, Tribhuvan University, Kirtipur, Nepal, in 2012 and the M.S. degree in electrical engineering from South Dakota State University, Brookings, SD, USA, in 2016. His research interests include microgrid energy management system, power system generation, transmission and distribution, renewable energy forecasting, and optimization. He received the Nepal Vidhya Bhusan (Gold Medal) as an honor from the Government of Nepal for showing outstanding academic performance in his undergraduate studies.

Semhar Michael received the Graduate degree in applied mathematics from the University of Asmara, Asmara, Eritrea, in 2006, the M.Sc. degree in mathematics from The University of North Dakota, Grand Forks, ND, USA, in 2009, and the Ph.D. degree in applied statistics from The University of Alabama, Tuscaloosa, AL, USA, in 2015. She is currently an Assistant Professor of statistics in the Mathematics and Statistics Department at South Dakota State University. Her research interests include finite mixture and Markov switching models and applications, natural language processing, specifically text data clustering and its applications.

Christopher Saunders received the Ph.D. degree in statistics from the University of Kentucky, Lexington, KY, USA, before accepting an Intelligence Community Postdoctoral Fellowship. He is a Mathematical Statistician with the MITRE Corporation; he also holds an associate professorship of statistics at South Dakota State University. His current research focus is related to high-dimensional pattern recognition and approximate Bayesian inference with applications to time series analysis and forensic identification.

Douglas Armstrong received the B.A. degree in statistics from the University of Minnesota-Morris, Morris, MN, USA, in 2011 and the Master's degree in statistics from South Dakota State University, Brookings, SD, USA, where upon defending his Masters research "Modeling Common Carp Under-Ice Movement using Hierarchical Markov Simulation," he entered into the Ph.D. program. He immediately began assisting in several projects involving in Markov switching and mixture modeling, pattern analysis, and simulation.

Prakash Pandey received the B.E. degree from the Institute of Engineering, Tribhuvan University, Kirtipur, Nepal, and the M.S. degree in electrical engineering from South Dakota State University, Brookings, SD, USA, in 2010 and 2015, respectively. He is currently working at Open Access Technology International, Inc., Minneapolis, MN, USA. His research interests include renewable energy, forecasting methods, and transmission system planning.

Santosh Chalise received the B.E. degree from the Institute of Engineering, Tribhuvan University, Kirtipur, Nepal, the M.S. degree in electrical engineering from Northern Arizona University, Flagstaff, AZ, USA, and the Ph.D. degree from South Dakota State University, Brookings, SD, USA, in 2004, 2012, and 2016, respectively. He is currently working at Open Access Technology International, Minneapolis, MN, USA. His research interests include microgrid operation, distributed generation, renewable energy management, battery lifetime management, and power system optimization.

Reinaldo Tonkoski (S'04–M'11) received the B.A.Sc. degree in control and automation engineering in 2004 and the M.Sc. degree in electrical engineering in 2006 from Pontifcia Universidade Catlica do RS, Porto Alegre, Brazil, and the Ph.D. degree in 2011 from Concordia University, Montreal, QC, Canada. He was with CanmetENERGY, Natural Resources Canada, from January 2009 to January 2010, where he worked on projects related to the grid integration of renewable energy sources. He is currently an Assistant Professor in the Electrical Engineering and Computer Science Department, South Dakota State University, Brookings, SD, USA. He has authored more than sixty technical publications in peer-reviewed journals and conferences. His research interests include grid integration of renewable energy systems, distributed generation, power quality, and power electronics.