This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2019.2904436, IEEE Transactions on Sustainable Energy

Development of a Markov-Chain-Based Solar Generation Model for Smart Microgrid Energy Management System

Reihaneh Haji Mahdizadeh Zargar, *and* Mohammad Hossein Yaghmaee, *Senior Member*, IEEE

***Abstract*—The growth of smart power grids has complicated the balancing of supply and demand, the control and management of power outages, and the reduction of grid costs. In order to strengthen power grid reliability and reduce the risk of failure in production resources, demand side management has taken on particular importance. Microgrids are dramatically expanding due to the potential benefits of an increasingly reliable, sustainable, efficient, and environmentally friendly energy supply from renewable energy sources. In recent years, the air pollution crisis has attracted much attention. Among the factors that aggravate this crisis, pollution from fossil fuels has been playing a significantly greater role. As a result, the trend towards renewable energies, such as solar power, has grown considerably. As a gradually introduced alternative to fossil fuels, solar energy presents many benefits as well as numerous challenges. One of the most critical issues is the unreliability of this energy source. The management solution is to define an appropriate model for predicting how solar energy works and how to accordingly manage its storage. The current paper proposes a solar generation model based on the Markov Chain (MC), by which the formation of a cluster of chains predicts the power generation of solar cells. In order to reduce the prediction error, Markov chains are continuously trained by historical data. Evaluation results confirm the accuracy of the proposed model.**

***Index Terms*—Cloud cover, energy management systems, Markov chain, microgrids, photovoltaic generation plants.**

NOMENCLATURE

Capital Letters

$AVG_{[t1,t2]}(error)$	mean value of error during interval [t1,t2]
BKN	cloud coverage- Broken = 5–7 oktas
CC_b	battery rechargeable capacity (kW)
CLR	cloud coverage- No cloud/Sky clear = 0 oktas
CMC	cluster of Markov chains Model
D_t	real cloud cover volume for time interval t (oktas)
DG	distributed Generation Sources- Solar Cells
$E_i(h)$	amount of power generated by the solar panel i in the time interval h (kW)
ED_i	electricity demand of request i (kW)
$EndHour_i$	end of time interval in sub-cluster i
$Er_t(i)$	prediction error of Markov chain i in time interval t
ESS	energy storage systems
FEW	cloud coverage- Few = 1–2 oktas
K	critical radiation point (W / m2)
L_d	length of dataset
LN	linear normalization function- $x_i / \sum_{j=1}^{z} x_j$
M	month of year
MAE	mean Absolute Error
MC	Markov chain
MC(k,l)	Markov chain l in sub-cluster k
MGCC	MicroGrid Central Controller
NWV_t	new weight vector generated according to the results of time interval t
OVC	cloud coverage- Overcast = 8 oktas, *i.e.*, full cloud coverage
$P_g^{max}(t)$	Maximum power allowed to receive from the distribution network in time interval t (kW)
$P_{ij}(k,l)$	transition probability from state i to state j for Markov chain l in sub-cluster k
P_t	array of MCs cloud cover predictions for time interval t
R_h	amount of solar radiation (W / m2)
RC	function of removing chain from cluster
SCT	cloud coverage- Scattered = 3–4 oktas
$SoC(t)$	state of charge in time interval t (%)
$StartHour_i$	start of time interval in sub-cluster i
TM	Markov chain transition matrix
V_{cp}	vectors array of states conditional probabilities
WV_t	weight vector for time interval t- indicates each MC prediction weight in computing weighted average. If used without indices, it means WV in the current time interval

Small Letters

c	number of solar panel cells
c_b	internal generated energy costs per unit of power
c_g	distribution network energy costs per unit of power
$c_{MG}(t)$	microgrid total costs in time interval t
$counter_i$	counter variable of loop i
e	efficiency of solar panels
h	hour of day
l	number of demand requests
m	number of sub-clusters in each cluster
n	number of Markov chains in each sub-cluster
s	vector of current state probabilities
$t_{ij}(k)$	transition from state i to state j in the hour k
weightedAvg	function of computing weighted average by getting V_{cp} and WV as inputs

Greek Letters

α	fraction of the electricity demands, which must be supplied from the domestic resources
β	fraction of the electricity demands, which must be

Reihaneh Haji Mahdizadeh Zargar and Mohammad Hossein Yaghmaee are with the Department of Computer Engineering, Ferdowsi University of Mashhad (FUM), Mashhad, Iran (e-mails: reihaneh_mahdizadeh@yahoo.com, hyaghmae@um.ac.ir).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2019.2904436, IEEE Transactions on Sustainable Energy

	supplied from the distribution network
$\pi(t)$	Markov chain vector of states probabilities in time interval t
ε_d	deletion threshold for chain weight
ε_m	chain merging threshold
ε_w	error threshold

I. Introduction

Microgrids are a collection of distributed generation, renewable energy sources, energy storage systems and local loads, which can operate in two modes, connected to the distribution network and islanded modes [1]. Stand-alone microgrids may never be connected to the main grid for economic or geographical reasons. Typically this type of microgrid is built in areas where the distribution infrastructure and transmission infrastructure have a high spatial distance. In a microgrid connected to the distribution network, the connection location is called the point of common coupling (PCC) [2], through which the energy required from the distribution network is received. The microgrid's most important feature is the ability to isolate itself from the distribution system when certain events occur, such as failures, voltage drops, and extinction. In connected to the distribution network mode, the main challenge is to minimize the cost of the microgrid; internal energy sources, such as solar cells, in addition to probabilistic behavior in the production, will incur costs. Along with internal resources, the distribution network can also supply a portion of the electricity demand by offering real time or time of use prices. Due to the costs each group can incur, the central controller will determine the source of the energy supply in order to reduce costs for a future period.

In recent years, the air pollution crisis has stirred up much controversy. Among the factors aggravating this crisis is the pollution caused by fossil fuels. Consequently, as a gradual alternative to fossil fuels, the trend towards renewable and new energies, such as solar energy, has grown significantly [3]. Although offering numerous advantages, solar energy also poses many challenges, the most important of which is its uncertainty as an energy source. This is basically due to this energy's random pattern of availability. The solution is to define a suitable model for predicting solar energy performance and for managing its battery storage systems. These issues are of particular importance in stand-alone microgrids, for which solar cells are one of the main sources of energy [4]. Furthermore, in microgrids employing solar energy as an auxiliary resource, the discussion of modeling this energy is of interest when attempting to reduce costs.

A. Related Works

As the current paper's focus is the modeling of photovoltaic power generation, this section surveys research related to modeling and predicting the power produced by solar cells and electricity demands. The methods presented in the field of forecasting production and demands are grouped according to the predictive horizons, the purpose of the prediction, and the type of model [5]. Estimates of production and demand can be categorized according to the forecasting period, usually known as the forecast horizon, as follows:

- Very short-term forecast: In this model, the load/power generation forecast is for a few seconds, minutes, and hours. These models are usually used to control the flow [6], [7].
- Short-term forecast: This model predicts a few hours to several weeks. These models are usually for adapting production and demand [13], [14], [24]-[26].
- Medium-term and long-term forecasts: This kind of forecasting can vary from months to years. These models are usually for planning service provider organizations [10], [27].

The most common type of forecasting model is weekly, daily and hourly. The most significant difference between the three models mentioned above is the range of variables used. On the other hand, a model can be classified based on the number of variables predicted. Accordingly, there are two main groups. The first is a method that predicts only one value, such as the next day's load, the next peak day, the next day's total load, and so on. The second group is a method predicting several values, such as the peak load and total load. In order to solve the equations, the linear models are based on a combination of all the problem's features. The demand forecast includes a number of variables and nonlinear events that first need to be detected and then converted into the equation that forms the model. The linear models are divided into two groups: time-of-day models and dynamic models. Time-of-day models define load at any discrete point in time from a specified time series, with a prediction period. The model stores the load amount based on previous observations. Some models store the load curves from several weeks ago, while other models only store the previous week's load curves. Dynamic models assume that the load depends not only on the time of day, but also on weather variables and random inputs. Dynamic models include self-regression models and the ARMA model [8]-[10], [27]. Nonlinear models (based on the neural network) have fewer constraints than those of linear models. Neural networks are able to learn from experiments and generalize past observations into new ones. They can also capture the essential features of input variables. As a result, these networks offer some special advantages that have attracted much interest, such as adaptive learning, self-organization, error tolerance, real-time performance, and ease of integration into existing technologies. The types of non-linear models or the types of artificial neural networks include hybrid models that are monitoring or non-monitoring and neural networks with enhanced learning [11]-[14], [28], [29]. Some studies have exclusively employed neural networks, while a number of other studies have used neural networks in combination with other methods. The latter are known as hybrid models [15], [30]-[32]. In neural networks, there are two main modes for network training: learning with monitoring and learning without supervision. In the group of nonlinear models, some researchers have predicted the amount of solar power generation and electricity demands by applying classification methods on meteorology criteria such as temperature, humidity, dew point and global horizontal irradiance. In order to predict solar energy, [33] utilizes the naïve Bayes (NB) classification, in which they choose three

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2019.2904436, IEEE Transactions on Sustainable Energy

continuous variables such as temperature, relative humidity, and dew point. With the Markov chain model, the authors estimate the battery transition from one state of charge to another. [34] proposes a hierarchical clustering technique to classify historical load curves according to temperature and humidity. Via an unsupervised classification method, [35] identifies days by describing a common consumption behavior pattern. In [36], temperature and solar irradiance are vital features for classifying different operating conditions of the PV system. In addition, the study develops a higher-order Markov chain based on the categorized historical data of solar power in each operating condition. With the help of Markov chains, quantitative research has been conducted to model the production of solar cells [8], [17]-[23]. Aided by historical data, the research conducted by [17], [37] employs Markov chains for modeling the photovoltaic generation. The solar insolation conditions of the sun are considered as model states which take into account the effect of clouds. By emphasizing the importance of solar panel dimensions and battery size in the design of solar power stations, some literature [19] has proposed a multi-state Markov model for the hourly harvesting of solar energy. The model determines the cost of the optimal photovoltaic panel and battery size by considering a tolerable power outage. In [20], a prediction method for solar insolation and cloud cover is presented based on a homogeneous recursive Markov process with discrete states. The introduction of the present article briefly mentions [21] as a reference to the current state of solar insolation prediction methods published in 2013. Inspired by the method proposed in [37], [21] suggests a model for determining how sensor node behavior is represented by considering different levels of battery charging and different solar insolation conditions.

B. Contributions

1) Photovoltaic power generation modeling based on Markov chains clusters: to predict the amount of power produced by solar cells in a specific time interval, the current study employs Markov chains in the probabilistic modeling of solar cells. With the aid of a two-state chain or a separation of days, literature to date has initially indicated the solar radiation status as good and bad days or sunny and covered periods [18]-[20], [23]. As a sample, in reference [19]; the solar energy output for each day in a given month is computed and the days are sorted based on this energy. β% of the days with the lowest energy are termed "bad," and the rest, "good" days, and a two-state chain with good and bad states was proposed. In another research, the authors used a division (day and cloud/low Light) for modeling the battery charge based on Markov chains [23]. In [20], authors use a two-state model that was presented by the same author in previous work, a situation in which the sun is completely hidden behind the clouds and others where the sun is shining. The totality of the clear time span consists of n sunny periods separated by n covered periods. Compared to these articles that use a two-state chain, in the proposed Markov chain model, the present work defines the chain states based on the METAR criterion[1] and the state of cloud volume which has not yet been developed in any studies predicting the production rate of solar cells. In addition, for more accurate prediction, the current authors have clustered the chains based on seasons and the time of day.

2) Reduction of power generation prediction errors by clustering chains and assigning weights: The current study has improved the prediction of solar power generation through clusters of Markov chains. These clusters enable the prediction module to reduce errors by utilizing season and time of day factors, as well as weighing the historical data.

3) Model training for privatization based on the characteristics of the deployment environment: The present work continuously trains the proposed model based on feedback received from real cloud cover data. This training can help the model fit its deployment environment and model dynamics. In order to train the model, the prediction error of each Markov chain is used as a criterion for determining the weight of the chain.

II. System Model

A. Microgrid Components

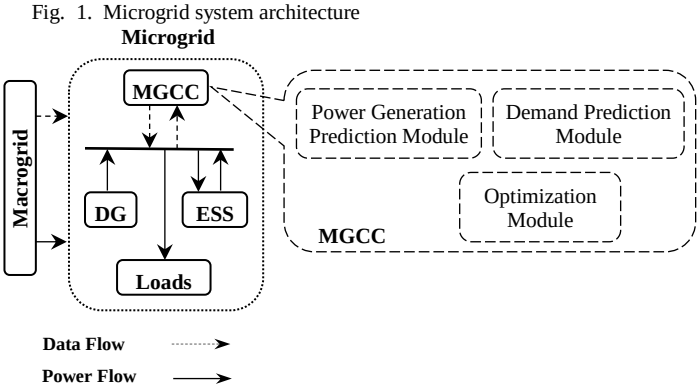
Each microgrid consists of a number of heterogeneous components such as power generation, electrical loads, and storage systems. Microgrid power generation may be from a range of variable distributed energy resources (DER's), including renewable and fossil fuelled generators. There are many forms of renewable energy; solar cells are an important source of renewable energy. Solar cells or photovoltaic cells convert the energy of light directly into electricity, which can be stored in batteries during peak production and be used in the production reduction times. A microgrid can supply its electricity demands with the help of internal resources such as solar cells or receive energy from a macrogrid in grid-connected mode and it can control the power flow at point of common coupling (PCC) by controlling the charging/ discharging of batteries. The present study assumes solar cells and macrogrids as power resources and local microgrid consumers as loads. Energy storage systems (ESS) are also used to increase the reliability of microgrid power resources. Reliability in power systems means the existence of energy at the time of need [38]. Fig. 1 shows the hypothesized microgrid structure. This structure models solar cells as Distributed Generation sources (DGs). The current study will expand microgrid resources in order to add other power sources, such as wind turbines, to future works. In each microgrid, the task of energy management is the responsibility of the microgrid central controller (MGCC). The microgrid central controller determines the amount of power recovered from each source according to production and demand conditions and the costs of each resource over the next period. In addition to considering the limitations of not creating a challenge for the main production

[1] METAR is a format for reporting weather information. A typical METAR contains data for the temperature, dew point, wind direction and speed, cloud cover and heights, visibility, and barometric pressure.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2019.2904436, IEEE Transactions on Sustainable Energy

network, the microgrid central controller attempts to reduce its costs via the energy management system. As illustrated in Fig. 1, the microgrid central controller consists of three modules: power generation prediction, demand prediction and optimization modules. The following sections provide the details of these modules.

Fig. 1. Microgrid system architecture



B. Solar Generation Model

The currently proposed strategy employs Markov chains to predict the amount of power produced by solar cells in future periods. The number of clouds in the sky directly affects power production and so a cloud-size pattern can predict the amount of available solar power.

To model solar generation, the present work suggests a Cluster of Markov Chains (CMC). Twelve clusters of Markov chains are formed as the first level clusters, of which each set represents one month of the year. Each set also consists of m sub-clusters of Markov chains as second level clusters, where m determines the number of divisions during a day. In each sub-cluster, the present work generates n Markov chains according to the historical data. The value of n increases over time as the historical dataset grows. Fig. 2 depicts each Markov chain structure, in which five states are possible: from clear sky to full cloud cover. As there are n Markov chains in each sub-cluster, $P_{ij}(k,l)$ is defined as a transition probability from state i to j for Markov chain l in sub-cluster k. Historical data helps form the transition matrix (TM) of each MC.

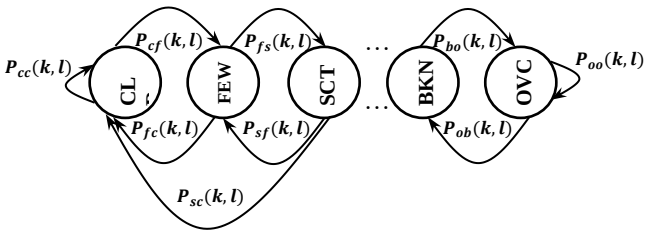


Fig. 2. An instance of Markov chains in clusters. For clarity, only some transitions are displayed.

A group of n Markov chains sharing the same properties (i.e. season and time of day) form a sub-cluster. Each sub-cluster may consist of a wide range of old or new historical data. Each cluster covers a full month of data and will model it. The overall scheme of the proposed model (CMC model) is shown in Fig. 3. At the beginning of the system, the model is formed with the help of a set of initial historical data. In order to produce predictions with acceptable error rates, the initial model needs to be updated and permanently maintained. Each Markov chain is represented by a series of historical data that determines the initial refinement of the chain transmission matrix. To cover the changes introduced by new data, these data will be continuously injected into the model at certain intervals. In other words, the model is trained on the basis of the data over time.

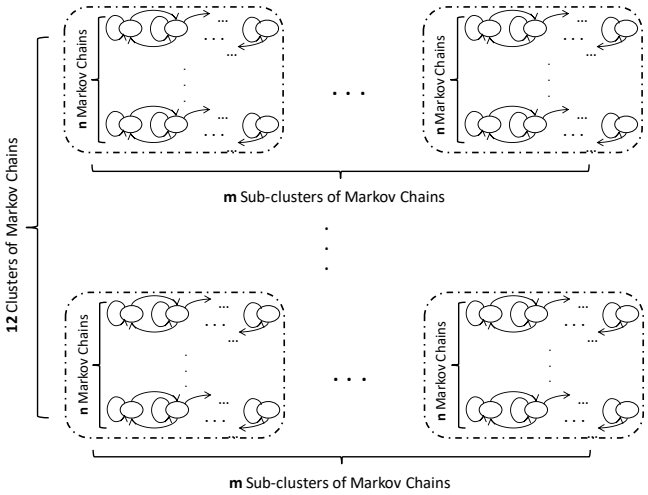


Fig. 3. Markov based solar generation model (CMC model). Each chain is shown in more detail in Fig. 2.

For determining the generation forecast of the future time interval, each chain provides a prediction value whose end result will be calculated by the weight assigned to each chain. The time for collecting the historical data of each chain initially determines the specific weight of each chain; we assigned more weight to the chains, which the corresponding cloud cover data are closer to the prediction period in terms of time. The new values of the weight vector elements (NWV) are determined through the maintenance module.

Algorithm 1 Training and Maintenance Module

function TrainAndMaintenance (P_t, WV_t, D_t)
Input: array of MCs cloud cover predictions for time interval t P_t, array of MCs weight vector WV_t, real cloud cover volume for time interval t D_t
Output: array of MCs new weight vector NWV_t

```
for counter1:=1 to n do
  Er_t[counter1] ← |P_t[counter1] - D_t|
end for
Er_t ← LN(Er_t)
for counter2:=1 to n do
  if AVG_[t-10,t](Er_t[counter2])> ε_w then
    power ← 1
  else power ← 2
  end
  NWV_t[counter2] ← WV_t[counter2] + (–1)^power
        *AVG_[t-10,t](Er_t[counter2])* WV_t[counter2]
  if NWV_t[counter2] < ε_d then
    RC(counter2)
  end
end for
```

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2019.2904436, IEEE Transactions on Sustainable Energy

After each run, the prediction module updates the weight of each chain based on the obtained error. In order to optimize the storage space and the complexity of the prediction module implementation during the growth of the model, chains having the same results as the difference of ε_m are merged while chains with a weight less than ε_d are removed from the sub-cluster. In the form of a pseudo code, Algorithm 1 presents the training and maintenance module.

III. Energy Scheduling

The inputs of the energy scheduling module are the load demands, power generation amount, energy purchase price from the distribution network, and the cost of energy generation. In the energy scheduling module, the maximum allowable power purchase amount and the state of charge of the energy storage system are applied as constraints on the optimization problem.

A. Supply and Demand Prediction

The energy generated by solar cells depends on a variety of factors, most notably the amount of cloud cover in the installation area of the panels. The amount of power generated by solar panel i in time interval h can be calculated by the following equation [17]:

$$E_i(h) = c.\begin{cases}\frac{e}{K}.R_h^2 & 0 < R_h < K \\ e.R_h & R_h > K\end{cases} \quad (1)$$

where c, e, K and R_h respectively represent the number of cells in the panel, the respective efficiency, the critical radiation point (W / m^2), and the amount of solar radiation. Cloud historical data is generally accessible via weather stations. These data can be used to predict future cloud cover. The currently proposed model consists of $n \times m \times 12$ MCs. Each MC state space consists of five states, namely S:{CLR, FEW, SCT, BKN, OVC}. A series-related historical dataset based on Equation (2) calculates the MC transition matrix. In Equation (3), t_{i*} signifies the transfer from state i to other states.

$$TM = \begin{bmatrix} P_{cc} & P_{cf} & P_{cs} & P_{cb} & P_{co} \\ P_{fc} & P_{ff} & P_{fs} & P_{fb} & P_{fo} \\ P_{sc} & P_{sf} & P_{ss} & P_{sb} & P_{so} \\ P_{bc} & P_{bf} & P_{bs} & P_{bb} & P_{bo} \\ P_{oc} & P_{of} & P_{os} & P_{ob} & P_{oo} \end{bmatrix} \quad (2)$$

$$P_{ij} = \sum_{k=1}^{L_d} t_{ij}(k) \,/ \sum_{k=1}^{L_d} t_{i*}(k) \;, for\ i,j \in S \quad (3)$$

$$t_{ij}(k) = \begin{cases}1 & \text{exists transition from i to j } in\ hour\ k \\ 0 & otherwise\end{cases} \quad (4)$$

$$t_{i*}(k) = \begin{cases}1 & \text{exists transition that has started from i} \\ 0 & otherwise\end{cases} \quad (5)$$

The probabilities of the transition matrix are calculated according to the series-related historical dataset, and will be changed based on changes in the dataset. In Equation (3), k signifies the hour of day in the dataset and its values are the integers that are determined by the size of the dataset. As shown in Equation (4), t_{ij} indicates the existence or absence of a transition from state i to state j. Given the probabilities vector of the current state and the transition matrix of each MC in time interval t-1, the states probability of time interval t is obtained according to Equation (6). In other words, through the chain transition matrix, the state of the system can be predicted in the next iteration.

$$\pi(t) = \pi(t-1) * TM \quad (6)$$

In order to predict the volume of the cloud over for the next time interval, the sub-cluster is first determined according to the time of day and season. The selected sub-cluster consists of n chains which provide n predictions. Prediction values are aggregated based on the weight vector of the chains and so the final value of the cloud cover prediction is determined. Algorithm 2 presents the supply prediction module in pseudo code form.

Algorithm 2 Supply Prediction Module

```
function prediction(h, M, s)
Input: hour of day h, month of year M, vector of current state probabilities s
Output: vectors array of states conditional probabilities V_cp
for counter1:=1 to m do
  if StartHour_counter1 ≤ h ≤ EndHour_counter1 then
    for counter2:=1 to n do
      V_cp[counter2] ← s*TM[M][counter1][counter2]
    end for
    V_cp ← weightedAvg (V_cp, WV)
    return V_cp
  end
end for
```

The present study has applied the Linear Prediction filter Coefficients (LPC) method to estimate the load demand. The LPC method determines the coefficients of a forward linear predictor by minimizing the least squares prediction error. By receiving the x and p entries, the LPC function determines the pth-order linear predictor coefficients that predict the current value based on actual values of time series x in the past.

$$\hat{x}(n) = -a(2)x(n-1) - a(3)x(n-2) - \ldots - a(p+1)x(n-p) \quad (7)$$

In Equation (7), p is the order of the prediction filter polynomial.

B. Optimization

The present work's objective function is to minimize the cost of the microgrid (c_{MG}) by taking into account the limits of solar production and the ceiling for receiving power from the macrogrid for a given time period. The optimization problem is solved for $(T - t_k)$ time intervals starting from t_k to T at each execution time. In the optimization problem, t_k determines the starting point of time intervals that can take on different values depending on the scheduled start hour of the day.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2019.2904436, IEEE Transactions on Sustainable Energy

$$Min\sum_{t=t_k}^{T} c_{MG}(t)$$

$$= Min\sum_{t=t_k}^{T}[c_g(t)\sum_{i=1}^{l}\beta_i(t)ED_i(t) + c_b(t)\sum_{i=1}^{l}\alpha_i(t)ED_i(t)] \quad (8)$$

$$S.t.$$

$$\sum_{i=1}^{l}\alpha_i(t)ED_i(t) \le SoC(t)CC_b \quad (9)$$

$$\sum_{i=1}^{l}\beta_i(t)ED_i(t) \le P_g^{\max}(t) \quad (10)$$

$$\forall i,t \mid i \in \{1,2,...,l\}, t \in \{t_k, t_{k+1},...,T\} : \alpha_i(t) + \beta_i(t) = 1 \quad (11)$$

$$0 \le \alpha_i(t) \le 1 \quad (12)$$

$$0 \le \beta_i(t) \le 1 \quad (13)$$

The costs of the microgrid (c_{MG}), are divided into two parts: the internal energy generation costs (c_b) and the cost of purchasing energy from the distribution network (c_g) in each time interval t. The decision variables in the proposed optimization problem are α_i and β_i which determine the fractions of the electricity demands (ED_i) that must be supplied from domestic resources or the distribution network, respectively. Inequalities (9) and (10) will also impose a storage capacity limitation (i.e. available storage power in time interval t, that is determined by the state of charge in time interval t (SOC(t)) and the rechargeable battery capacity (CC_b) in kW) and a limitation on the power received from the distribution network (P_g^{max}) in time interval t, respectively. In order to determine the supply source of the total amount of electricity demands, the sum of α_i and β_i shall be equal to 1 for each electricity demand i in time interval t according to Equation (11). In addition, the values of α_i and β_i must be between 0 and 1 (inequalities (12) and (13)). Limiting the purchase of energy from a macrogrid is critical for preventing electricity demand peaks when macrogrid electricity prices fall. Also, due to the goals of creating microgrids, which are distributed energy generation and the ability to separate and isolate themselves from the distribution system at the time of an event, such as failure, voltage drop, and outage, the energy purchase plays a significant role in regulating microgrid behavior so that it stays in line with such goals. The optimization problem will be solved for a day ahead. Each day is divided into $(T - t_k)$ time slots. Based on the prediction made for the amount of power generation and demand in each time interval, the problem is initialized and solved.

IV. Results

A. Photovoltaic Generation Prediction

For the purpose of forming Markov chain clusters, the current research employed data collected from the meteorological stations provided by Iowa Environmental Mesonet (IEM) [16]. The meteorological data (temperature, wind speed, cloud volume, etc.) of various stations around the world are collected by IEM and these provide users with a scale of about one hour (most stations offer a scale of one hour and a few number of stations provide a scale of one minute). First, the present work performed an initial refinement of data from IEM which included digitizing METAR values for cloud cover, considering the maximum cloud cover (worst case) for slots whose data scale is less than an hour, and adding data for slots without cloud cover information due to the cloud cover rate in the previous slot. In the end, the current study utilized refined data to form Markov chains. Fig. 4 shows the unconditional probability mass functions for one meteorological station for the hours from 7 a.m. to 9 a.m. on the first of every month for a period of six years. As shown in Fig. 4, these probability functions follow a similar pattern. The horizontal axis states are equivalent to Markov chain state space S :{ CLR, FEW, SCT, BKN, OVC}, respectively.

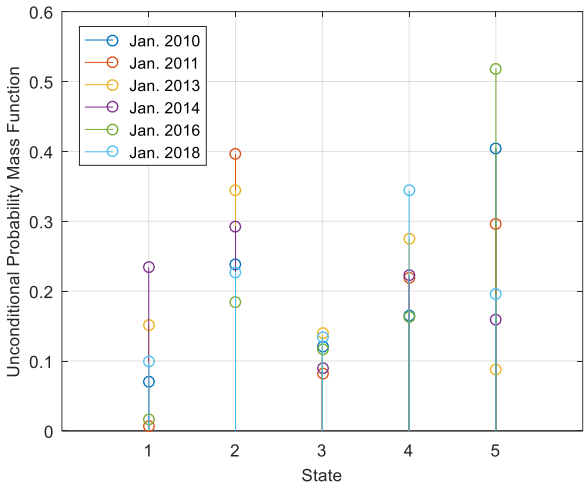


Fig. 4. Unconditional probability mass functions (pmf) for first month of 6 years cloud cover datasets.

The present research created a probability mass function by averaging the probability functions obtained from datasets of the years between 2010 and 2015 for one meteorological station. It then compared this probability function with the probability mass function obtained for the year 2018 during the same month and hours. Fig. 5(a) presents the obtained functions. If the difference between these two probability function values is considered as an error, then the mean absolute error will be 0.064878 in this case. In a similar comparison, when calculating the mean probability function, the current study assigned weights to each of the probability functions that are proportional to the time of collecting the corresponding cloud cover data. As shown in Fig. 5(b), in this case, the value obtained for MAE is lower (MAE = 0.057949576). Therefore, the weight vector assigned to different Markov chains can play a critical role in reducing the prediction error and, accordingly, the weight vector is continuously updated in the maintenance module. Fig. 5(c) provides the calculated pmf graph according to the trained chain weight vectors after 10 iterations. The authors assume $\varepsilon_w = 0.09$ as the error threshold for determining new weight vector values of the Markov chains.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2019.2904436, IEEE Transactions on Sustainable Energy

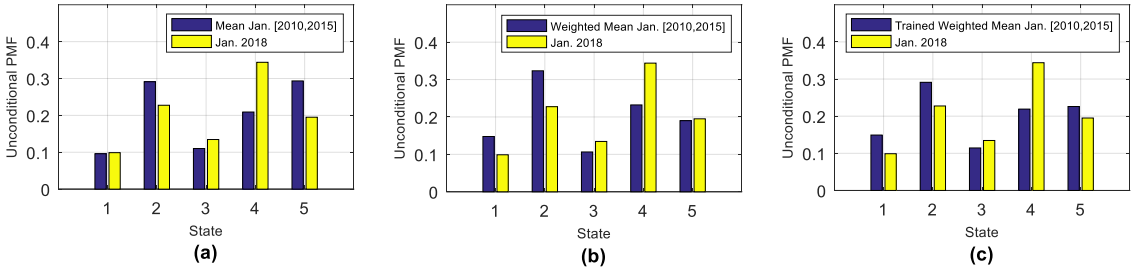


Fig. 5. Probability mass functions comparison (a) between mean of years 2010 to 2015 pmfs and year 2018 pmf (MAE=0.057949576) (b) between weighted mean of years 2010 to 2015 pmfs and year 2018 pmf (MAE=0.064878) (c) between trained weighted mean of years 2010 to 2015 pmfs and year 2018 pmf (MAE=0.057995585).

B. Load Demand Prediction

In order to choose a method for predicting the amount of energy required, the present work analyzed the home power consumption data of the UK. These data were extracted from the data collection for household use and recorded from domestic houses in the southeast of the UK. This dataset provided power consumption data separately for each house. Given the size of the microgrid, electricity demands in Equation (8) are considered either as one home's demands or as the cumulative effect of the electricity demands of homes and other subscribers of a region. Fig. 6(a) presents the sample of household power consumption for 30,000 hours. The 175-hour subset of these data is examined individually. Fig. 6(c) draws the graph of the equivalent polynomial curve of these data with a degree of 5. To extract the relationship between these data, the current work examined the autocorrelation of the cumulative power consumption of 1,000 homes between November 2012 and April 2015. Fig. 7(a) to Fig. 7(d) shows the autocorrelation of power consumption data for one day, 48 hours, one week, and one month, respectively. The highest autocorrelation observed was in 1 and 24 lags. The LPC method predicted the power consumption. Fig. 8 provides the power consumption prediction results for the data in Fig. 6(b). The mean absolute error was 24.5499 in this experiment. It was repeated for 100 sub-sets of dataset and the prediction MAE is an average of 31.5.

C. Optimization Module

The present study divides each day into 24 time slots and so this solves the proposed optimization problem for the next 24 slots. For each time slot, the estimation of the state of charge, $\mathrm{SoC(t)}$, is obtained based on the CMC's model's prediction of the solar generation amounts.

The LPC method also predicts the amount of electricity demands, $ED(t)$. In Equation 5, one of the most critical terms in power planning is the price of electricity purchased from the macrogrid. The pricing models are divided into five categories: Time of Use Pricing (TOUP), Critical Peak Pricing (CPP), Extreme Day Pricing (EDP), Extreme Day CPP (ED-CPP), and Real Time Pricing (RTP). The Time of Use (TOUP) and Real Time (RTP) pricing differ in that the former fixes the price and time periods in advance while the latter does neither of these two in advance. Thus, TOU rates can be considered static while RTP rates are dynamic, even though before the feature time-varying prices. Other rate designs bridge the gap between these two. TOU, CPP, EDP, and ED-CPP can be grouped into the same category although these models act differently in determining pricing details and announcing peak hours of consumption. The current research has assumed here that macrogrid provides energy according to the TOUP model. In this model, for peak times, the price is 2.5 times the average electricity price; at off-peak times, the price will be one-third of the average electricity price. Here, the peak hours of use are set from 12 to 18 and 20 to 23, while the off-peak hours are from 23 to 7 in the morning.

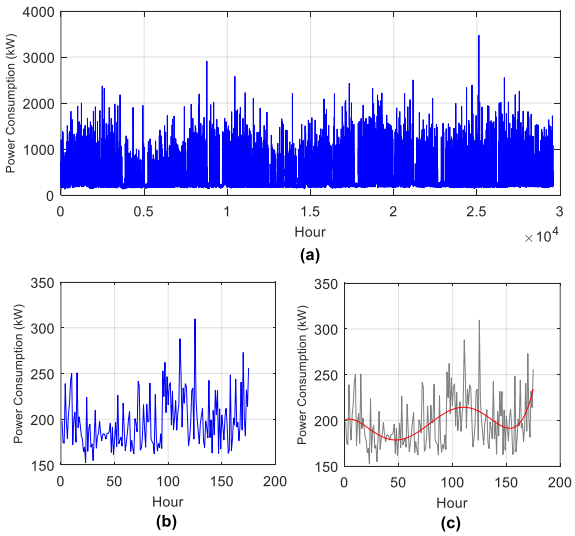


Fig. 6. (a) Sample residential power consumption for 3×10^4 hours. (b) Power consumption on Mondays (175 hours). (c) Polynomial curve estimation for chart of part (b).

The present work has divided the basic electricity demands of microgrid subscribers, such as lighting, into five electricity demand groups, namely ED_1 to ED_5. For these five groups, the LPC method estimates the electricity demand for the next day. Fig. 9(a) presents the ratio of α to β in aggregate for these requests. Fig. 9(b) and Fig. 9(c) show two samples of the power source allocation for electricity demands. As seen in Fig. 9(a), at consumption peak times, the maximum power allocation is made from the batteries. Also, due to the reduction in the battery charge state at the end of the day, electricity demands have been powered by macrogrid during the hours that solar cells produced more power and energy cost is less, to reduce the total cost of supplying power. The current research repeated this experiment for 30 days. The results indicate a 12% average reduction in energy costs.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2019.2904436, IEEE Transactions on Sustainable Energy

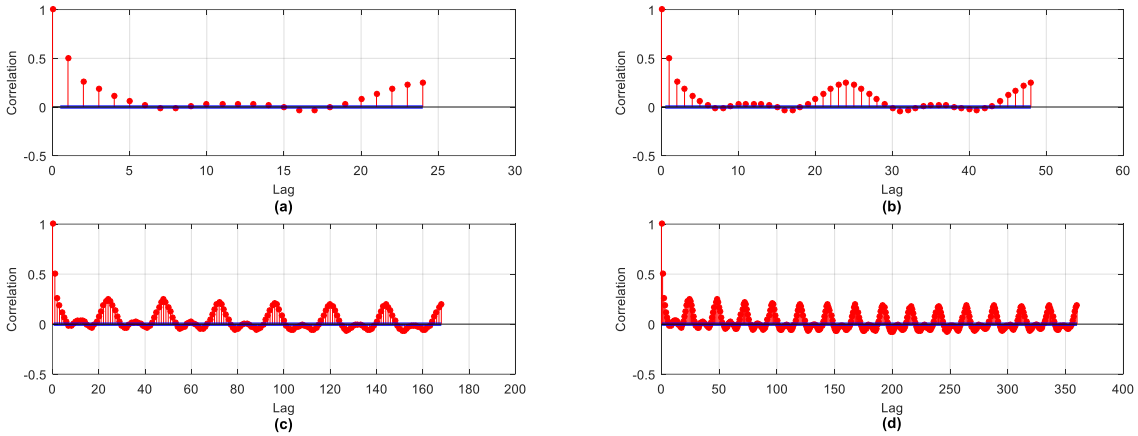


Fig. 7. Autocorrelation graph of the residential power consumption data. (a) During a day. (b) Between two days in a row. (c) During a week. (d) During a month.

This cost reduction is due to planning for the whole day and considering the status of resources and costs in all time slots. TABLE I compares the total energy costs over a 6-day period for the proposed power source allocation method and that of a simple method.

TABLE I
COST REDUCTION DUE TO OPTIMIZING POWER RESOURCE ALLOCATION

Iteration		1	2	3	4	5	6
Total cost	Optimized Resource Allocation	1306	1324	1041.6	1061	1140	978.5
	Simple Resource Allocation	1433.3	1551.5	1162.5	1195.1	1310.5	1158.6

The simple method determines the source of power solely on the basis of each time slot's energy price. In other words, the decision is only made on the basis of current conditions and the amount of solar power generation; the battery state of charge in the next time slots do not have any effect on the decision's results. TABLE I divides the costs by the average electricity price per kilowatt. As shown in TABLE I, in comparison with the simple power source allocation method, the proposed method lowered the total energy cost from 8.8 to 15.5 percent.

V. Conclusion

There is considerable interest in the design of microgrid systems that are able to schedule power resources and load demands. In recent years, the air pollution crisis has received special attention. Among the factors that exacerbate this crisis is the large role played by fossil fuels. As an alternative to fossil fuels, there has emerged a significantly growing trend towards gradually introduced new and renewable energies, such as solar energy. Along with its many advantages, the use of solar energy poses many challenges; one of the most important of which is its uncertainty. This uncertainty means that the availability of this energy follows a probabilistic pattern. The present study employs Markov chains to model solar cell power generation and also clusters these to achieve a lower prediction error rate in comparison to the same model without clustering.

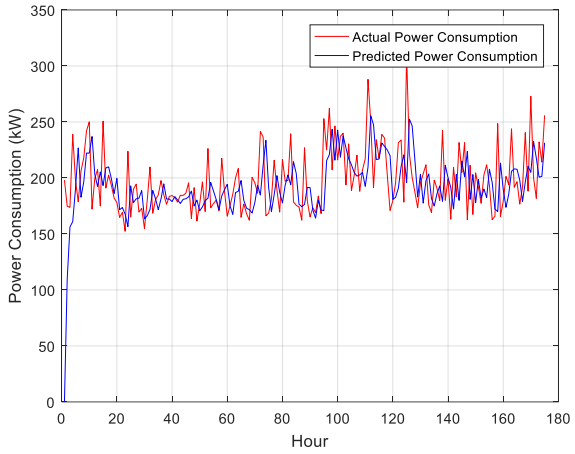


Fig. 8. Prediction results of the power consumption using LPC method.

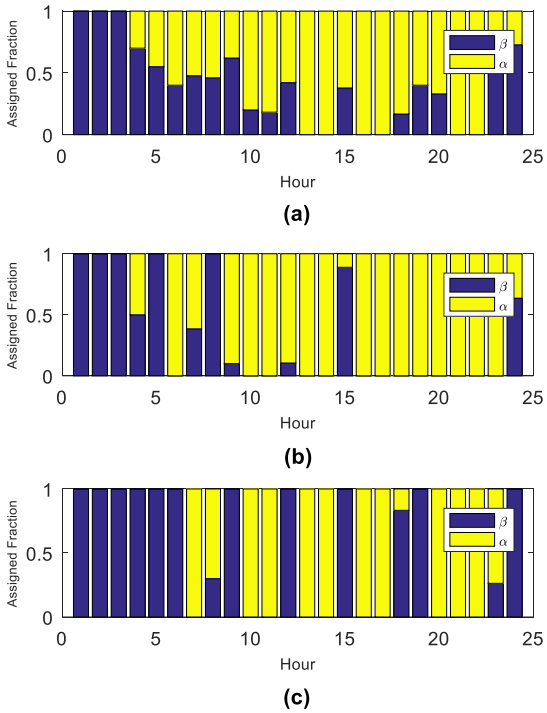


Fig. 9. Results of optimizing power source allocation for a day. (a) Five groups of electricity demands in aggregate. (b) Electricity demand group 2 (ED_2). (c) Electricity demand group 5 (ED_5).

The CMC model is continuously trained with newly collected data (both a weight vector and chains). The results indicate that the proposed training and maintenance module will reduce the forecast error rate over time. In future works, the authors intend to reduce the error rate by conducting further analysis on the final amount of solar power generation calculated from the predicted values obtained from each of the Markov chains, model training, and the maintenance module. In addition, they plan to expand microgrid power generation resources for the purposes of adding other power sources, such as wind turbines. In order to achieve results closer to those of real conditions in future works, the authors will utilize real time pricing for modeling the macrogrid energy cost.

REFERENCES

[1] N. Hatziargyriou, Microgrids, Architectures and Control, Chichester, West Sussex, United Kingdom: Wiley, 2014.

[2] S. X. Chen, H.B.Gooi, and M. Q. Wang. (March 2012). Sizing of Energy Storage for Microgrids. IEEE Transactions on Smart Grid, vol. 3, no. 1, pp. 142 – 151.

[3] C. Wan, J. Zhao, Y. Song, Z. Xu, J. Lin, and Z. Hu. (December 2015). Photovoltaic and Solar Power Forecasting for Smart Grid Energy Management. CSEE Journal of Power and Energy Systems, vol. 1, no. 4, pp. 38 - 46.

[4] M. Datta, T. Senjyu, A. Yona, T. Funabashi, and C. Kim. (March 2009). A Coordinated Control Method for Leveling PV Output Power Fluctuations of PV–Diesel Hybrid Systems Connected to Isolated Power Utility. IEEE Transactions on Energy Conversion, vol. 24, no. 1, pp. 153 – 162.

[5] L. Hernandez, C. Baladr´on, J. M. Aguiar, B. Carro, A. J. Sanchez-Esguevillas, J. Loret, and J. Massana. (Third Quarter 2014). A Survey on Electric Power Demand Forecasting: Future Trends in Smart Grids, Microgrids and Smart Buildings. IEEE Communications Surveys & Tutorials, vol. 16, no. 3, pp. 1460 – 1495.

[6] Z. Dong, D. Yang, T. Reindl, and W. M. Walsh. (March 2014). Satellite image analysis and a hybrid ESSS/ANN model to forecast solar irradiance in the tropics. Energy Conversion and Management, vol. 79, pp. 66–73.

[7] X. G. Agoua, R. Girard, and G. Kariniotakis. (April 2018). Short-Term Spatio-Temporal Forecasting of Photovoltaic Power Production. IEEE Transactions on Sustainable Energy, vol. 9, no. 2, pp. 538 – 546.

[8] J. Leithon, T. J. Lim, and S. Sun. (April 2016). Renewable energy management in cellular networks: An online strategy based on ARIMA forecasting and a Markov chain model. IEEE Wireless Communications and Networking Conference.

[9] I. Colak, M. Yesilbudak, N. Genc, and R. Bayindir. (March 2016). Multi-period Prediction of Solar Radiation Using ARMA and ARIMA Models. IEEE 14th International Conference on Machine Learning and Applications (ICMLA).

[10] H. Cheng, W. Cao, and P. Ge. (Aug. 2012). Forecasting Research of Long-Term Solar Irradiance and Output Power for Photovoltaic Generation System. Fourth International Conference on Computational and Information Sciences.

[11] Q. Liu and Q. Zhang. (June 2016). Accuracy Improvement of Energy Prediction for Solar-Energy-Powered Embedded Systems, IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 24, no. 6, pp. 2062 – 2074.

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

[13] A. Asrari, T. X. Wu, and B. Ramos. (April 2017). A Hybrid Algorithm for Short-Term Solar Power Prediction—Sunshine State Case Study. IEEE Transactions on Sustainable Energy, vol. 8, no. 2, pp. 582 – 591.

[14] K. Y. Bae, H. S. Jang, and D. K. Sung. (March 2017). Hourly Solar Irradiance Prediction Based on Support Vector Machine and Its Error Analysis. IEEE Transactions on Power Systems, vol. 32, no. 2. pp. 935 – 945.

[15] H. Nazaripouya, B. Wang, Y. Wang, P. Chu, H. R. Pota, and R. Gadh. (May 2016). Univariate time series prediction of solar power using a hybrid wavelet-ARMA-NARX prediction method. IEEE/PES Transmission and Distribution Conference and Exposition (T&D).

[16] Iowa State University, "Iowa Environmental Mesonet Search," [Online]. Available: https://mesonet.agron.iastate.edu/request/download. phtml. [Accessed: Jan. 15, 2018].

[17] M. H. K. Tushar, C. Assi, M. Maier and M. F. Uddin. (January 2014). Smart Microgrids: Optimal Joint Scheduling for Electric Vehicles and Home Appliances. IEEE Transactions on Smart Grid, vol. 5, no. 1, pp. 239-250.

[18] N. Kakimoto, S. Matsumura, K. Kobayashi, and Masaki Shoji. (January 2014). Two-State Markov Model of Solar Radiation and Consideration on Storage Size. IEEE Transactions on Sustainable Energy, vol. 5, no. 1, pp. 171 - 181.

[19] V. Chamola and B. Sikdar. (October 2015). A Multistate Markov Model for Dimensioning Solar Powered Cellular Base Stations. IEEE Transactions on Sustainable Energy, vol. 6, no. 4, pp. 1650-1652.

[20] H. Morf. (December 2014). Sunshine and cloud cover prediction based on Markov processes. Solar Energy, vol. 110, pp. 615–626.

[21] Photovoltaic and Solar Forecasting: State of the Art. IEAPVPS Task 14, Subtask 3.1.Report IEA-PVPS T14–01: 2013. ISBN 978-3-906042-13-8.

[22] J. Song, V. Krishnamurthy, A. Kwasinski and R. Sharma. (April 2013) Development of a Markov-Chain-Based Energy Storage Model for Power Supply Availability Assessment of Photovoltaic Generation Plants. IEEE Transactions on Sustainable Energy, vol. 4, no. 2, pp. 491-500.

[23] M. A. Murtaza, and M. Tahir. (2013). Optimal Data Transmission and Battery Charging Policies for Solar Powered Sensor Networks using Markov Decision Process. IEEE Wireless communication and Networking Conference.

[24] Y. Zhang and G. Luo. (October–November 2015). Short term power load prediction with knowledge transfer. Information Systems, vol. 53, pp. 161-169.

[25] C. E. Borges, Y. K. Penya, and I. Fern´andez. (August 2013). Evaluating Combined Load Forecasting in Large Power Systems and Smart Grids, IEEE Transactions on Industrial Informatics, vol. 9, no. 3, pp. 1570 - 1577.

[26] N Amjady, Fa. Keynia, and H. Zareipour. (December 2010). Short-Term Load Forecast of Microgrids by a New Bilevel Prediction Strategy, IEEE Transactions on Smart Grid, vol. 1, vo. 3, pp. 286 – 294.

[27] J. Xie, T. Hong, and J. Stroud. (September 2015). Long-Term Retail Energy Forecasting With Consideration of Residential Customer Attrition, IEEE Transactions on Smart Grid, vol. 6, no. 5, pp. 2245 – 2252.

[28] A.S. Khwaja, M. Naeem, A. Anpalagan, A. Venetsanopoulos, and B. Venkatesh. (August 2015). Improved short-term load forecasting using bagged neural networks, Electric Power Systems Research, vol. 125, pp. 109–115.

[29] A. RazaKhan, A. Mahmood, A. Safdar, Z. A.Khan, and N. AhmedKhan.(February 2016). Load forecasting, dynamic pricing and DSM in smart grid: A review. Renewable and Sustainable Energy Reviews, vol. 54, pp. 1311–1322.

[30] N. Liu, Q. Tang, J. Zhang, W. Fan, and J. Liu. (September 2014). A hybrid forecasting model with parameter optimization for short-term load forecasting of micro-grids. Applied Energy, vol.129, pp. 336–345.

[31] A. Bracale, P. Caramia, G.Carpinelli, A. R.D. Fazio, and P. Varilone. (December 2013). A Bayesian-Based Approach for a Short-Term Steady-State Forecast of a Smart Grid. IEEE Transactions on Smart Grid, vol. 4, no. 4, pp. 1760 – 1771.

[32] L. Suganthi, S. Iniyan, and A. Samuel. (August 2015). Applications of fuzzy logic in renewable energy systems – A review. Renewable and Sustainable Energy Reviews, vol. 48, pp. 585–607.

[33] Y. Kwon, A. Kwasinski, and A. Kwasinski, "Microgrids for base stations: Renewable energy prediction and battery bank management for effective state of charge control", presented at the IEEE International Telecommunications Energy Conference (INTELEC), Osaka, Japan, Oct. 18-22 2015.

[34] P. Zhang, X. Wu, X. Wang, and S. Bi. (September 2015). Short-term load forecasting based on big data technologies. CSEE Journal of Power and Energy Systems, vol. 1, no. 3, pp. 59 - 67.

[35] M. Chaouch. (January 2014). Clustering-Based Improvement of Nonparametric Functional Time Series Forecasting: Application to Intra-Day Household-Level Load Curves. IEEE Transactions on Smart Grid, vol. 5, no. 1, pp. 411 - 419.

[36] M. J. Sanjari, H. B. Gooi. (July 2017). Probabilistic Forecast of PV Power Generation Based on Higher Order Markov Chain. IEEE Transactions on Power Systems, vol. 32, no. 4, pp. 2942 - 2952.

[37] D. Niyato, E. Hossain and A. Fallahi. (February 2007). Sleep and Wakeup Strategies in Solar-Powered Wireless Sensor/Mesh Networks:

Performance Analysis and Optimization. IEEE Transactions on Mobile Computing, vol. 6, no. 2, pp. 221-236.

[38] K. Moslehi and R. Kumar. (June 2010). A Reliability Perspective of the Smart Grid. IEEE Transactions on Smart Grid, vol. 1, no. 1, pp. 57 - 64.

Reihaneh Haji Mahdizadeh Zargar is a PhD student at the Computer Engineering Department, Ferdowsi University of Mashhad (FUM). She received M.Sc. degree in software computer engineering from Ferdowsi University of Mashhad in 2013. Her research interests are in the area of Microgrid energy management systems, solar generation modeling, and Markov chain models.

Mohammad Hossein Yaghmaee Moghaddam (SM'12) received his B.S. degree from Sharif University of Technology, Tehran, Iran in 1993, and M.S. and Ph.D degrees in communication engineering from Tehran Polytechnic (Amirkabir) University of Technology in 1995 and 2000, respectively. November 1998 to July1999, he was with Network Technology Group (NTG), C&C Media research labs., NEC corporation, Tokyo, Japan, as visiting research scholar. September 2007 to August 2008, he was with the Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, USA as the visiting associate professor. July 2015 to September 2016, he was with the electrical and computer engineering department of the University of Toronto (UoT) as the visiting professor. Currently, he is a full professor at the Computer Engineering Department, Ferdowsi University of Mashhad (FUM). His research interests are in Smart Grid Communication, Software Defined Networking (SDN) and Network Function Virtualization (NFV).