Effects of Solar and Wind Generation Integration on Feeder Hosting Capacity

Dimitra Apostolopoulou, Kyriakos Anastasopoulos, and Shay Bahramirad
Smart Grid & Technology
Commonwealth Edison Company
Chicago, IL, USA
{Dimitra.Apostolopoulou, Kyriakos.Anastasopoulos, Shay.Bahramirad}@ComEd.com

***Abstract*—With the increased penetration of distributed generation (DG) utilities are beginning to see impacts on their system, especially on the ability of a feeder to accommodate DG. In this paper we introduce a stochastic simulation framework to assess the effects on hosting capacity from solar and wind generation for various loading scenarios. The general approach includes the use of a k-means clustering algorithm for segmenting and grouping the raw wind, solar, and load data to define patterns and assign probabilities to each pattern. Monte Carlo simulations are adopted for calculating probabilistic outcomes for a variety of wind, solar, and load scenarios, with the use of a distribution planning software. The outcomes of the simulations, i.e., statistics of minimum and maximum feeder hosting capacity, are used to derive their probability distribution functions (pdfs). The pdfs of the minimum and maximum hosting capacity provide insights into the effects on loading from various wind and solar DG scenarios. The proposed framework is illustrated for a representative utility feeder.**

***Index Terms*—Feeder Hosting Capacity, Solar Generation, Wind Generation, Stochastic Simulation Framework.**

I. Introduction

With the rapid adoption of distributed generation (DG), it is becoming inherently challenging for utilities to assess how these applications will affect the distribution system. Until recently there have been relatively low amounts of DG installations and as a result most of the research has been focused towards developing techniques for determining the output of these systems and not their interconnection challenges. Market forces, government subsidies, and legislative mandates such as California's ambitious Renewable Energy Portfolio; as well as Fast-track integration initiatives such as the 15% screening process established by the Federal Electric Regulatory Council (FERC), are driving down costs for renewables and promoting their rapid adoption [1]. As a result of the short time frame well thought out interconnection requirements have not been established and utilities are limited to generally insufficient tools to assess the effects of these systems.

Particular concern is raised towards, solar and wind DG. Unlike other forms of DG, these generation types pose unique challenges due the variable and uncertain nature of their output. Conventional generation resources have relatively slow system fluctuations, and their energy output can be accurately predicted using day-ahead or hour-ahead forecasting techniques. Unlike these, wind and solar DG systems are characterized by more rapid and less predictable generation fluctuations over smaller time scales ranging from seconds to hours, and thus become difficult to predict [2]. As such, there exists a need for the development of more precise and robust modeling and analysis methods to understand wind and solar DG impacts on feeders from an interconnectivity standpoint.

One such analysis is determining hosting capacity of a feeder. Hosting capacity can be defined as the maximum amount of DG that can be accommodated without impacting system operation [3]. Hosting capacity addresses two of the major challenges that utilities are faced with when considering the integration of DG; overvoltage and overloading. Traditional feeder systems have been designed to flow radially and outward from the substation; the defining issue is the addition of local generation that will counteract the traditional unidirectional flow of a feeder contributing additional load to the system. As a result, local high voltage, load, and possible backfeed scenarios may emerge (e.g., [3], [4], [5], [6]). Detailed studies have shown that hosting capacity is rather complex and depends on a variety of different variables, such as DG location, size, and the individual feeder characteristics; factors which the 15% rule established by FERC does not take into account [1]. In light of the obvious need for developing better screening criteria to ensure grid reliability, many individual studies have been conducted utilizing stochastic methodologies in order to calculate and model hosting capacity. The key advantages of these methods include scalability and more realistic results.

In this paper we propose stochastic simulation framework that utilizes k-means clustering to determine patterns in the solar, wind, and load data. The k-means output provides a set of clusters and a probability of each cluster occurring. We sample the developed clusters for the solar, wind, and load data over a 24-hour window and run Monte Carlo simulations to determine the probability distribution function (pdf) of the hosting capacity of a specific feeder on a distribution planning software. The paper is outlined as follows. In Section II, we illustrate the development of the wind, solar, and load mathematical models as well as the clustering techniques used to group and analyze the data. In Section III, we describe the proposed stochastic simulation framework and sampling procedure. In Section IV, we illustrate the validity of the proposed framework through a representative utility feeder. In Section V, we make some concluding remarks.

II. Power System Model

In this section, we develop the mathematical modeling used to describe the behavior of the distribution system including DG. More specifically, we model the wind and solar output and present the model representing the load and network.

A. Wind Generation

The analysis of the impacts of wind generation into a specific feeder requires a wind generation model with the capability to represent the wind speed and power output with the appropriate level of detail. We develop a model that explicitly represents the variability and intermittency characteristics of the wind generation. To this end, we focus on the wind speed modeling, capture patterns in the wind data by using clustering techniques, and determine the wind generation output [7].

1) Wind Speed Model: The shape of the daily wind power output depends directly on the daily wind speed pattern. Similar daily wind patterns may occur multiple times during the period under consideration. To identify the various wind speed patterns, we group the days whose wind speed patterns have similar "shapes" into a class. We partition the day into $H = 24$ non-overlapping time intervals to analyze the wind speed data. We collect hourly data for a D number of days, usually for a one-year period; thus $D = 365$. We denote by $\mathscr{D}$ the set that contains the days of the period under consideration $\mathscr{D} = \{1, \ldots, D\}$. We denote by $v_{d,h}$ the wind speed at day d and hour h. Thus, we formulate the wind speed vector for day d as $v_d = [v_{d,1}, v_{d,2}, \ldots, v_{d,24}]^T$. Then, we construct the matrix V that contains the wind speed data for every hour for all the days D during the period under consideration. Thus, we have

$$V = \begin{bmatrix} v_{1,1} & v_{1,2} & \cdots & v_{1,24} \\ v_{2,1} & v_{2,2} & \cdots & v_{2,24} \\ \vdots & \vdots & \ddots & \vdots \\ v_{D,1} & v_{D_2} & \cdots & v_{D,24} \end{bmatrix} = \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_D^T \end{bmatrix}. \tag{1}$$

2) Clustering of Wind Speed Data: We wish to group the collected data into groups that have similar patterns and define the probability of these patterns occurring. To this end, we apply clustering techniques in the wind speed data given in V to determine wind speed patterns. We use the k-means method, whose detailed description may be found in [8]. In this paper, we provide a brief description of the algorithm. We begin by k typical daily wind speed vectors and denote the set of class centers. We then compute the relative Euclidean distance between the cluster points and class centers and assign new cluster centers depending on the calculated distance between cluster points and original centers so as to minimize the distance between all members in a cluster.

Once we have identified the k classes, $\mathscr{C}_{W_1}, \ldots, \mathscr{C}_{W_k}$, composed of the similar days of wind speed data, we have defined k daily wind speed patterns. We interpret each class $\mathscr{C}_{W_c}$, $c = 1, \ldots, k$, as a realization of the set of hourly wind speed random variable $\nu_{c,h}$ for $h = 1, \ldots, 24$. The mean value of each random variable $\mu_{W_{c,h}}$ is defined as

$$\mu_{W_{c,h}} = \frac{1}{|\mathscr{C}_{W_c}|} \sum_{v_d \in \mathscr{C}_{W_c}} v_{d,h}, \tag{2}$$

where $|\mathscr{C}_{W_c}|$ is the number of wind speed vectors v_d that are part of class $\mathscr{C}_{W_c}$. For each class $\mathscr{C}_{W_c}$ we construct the vector that represents the daily wind speed pattern by $\mu_{W_c} = [\mu_{W_{c,1}}, \ldots, \mu_{W_{c,24}}]^T$. We wish to assign a probability of each realization, i.e., class, occurring. To this end we associate with each class $\mathscr{C}_{W_c}$ the probability p_{W_c}, which is equal to

$$p_{W_c} = \frac{|\mathscr{C}_{W_c}|}{D}. \tag{3}$$

3) Wind Generation Output Model: We have established wind speed patterns and their associated probability of occurrence. Next, we need to define the wind generation output based on a certain daily wind speed pattern. To this end, we denote by v_c the cut-in wind speed, v_r the rated wind speed, v_o the cut-out wind speed, and P_{W_r} the maximum or rated output power of a wind unit. We use the following function for the wind output curve

$$P_{W_h}(v_{d,h}) = \begin{cases} 0 & 0 \leq v_{d,h} < v_c \\ a + b v_{d,h}^3 & v_c \leq v_{d,h} < v_r \\ P_{W_r} & v_r \leq v_{d,h} < v_o \\ 0 & v_{d,h} \geq v_o \end{cases}. \tag{4}$$

B. Solar Generation

The approach developed to represent the solar output generation is similar to that of wind generation described in Section II-A. We collect hourly solar generation output data for a period of D days. The output of solar generation for day d at hour h is denoted by $P_{S_{d,h}}$. We construct the solar output vector for day d as $P_{S_d} = [P_{S_{d,1}}, P_{S_{d,2}}, \ldots, P_{S_{d,24}}]^T$. Then, we construct the matrix P_S that contains the solar generation output for every hour for all the days D during the period under consideration. Thus, we have

$$P_S = \begin{bmatrix} P_{S_{1,1}} & P_{S_{1,2}} & \cdots & P_{S_{1,24}} \\ P_{S_{2,1}} & P_{S_{2,2}} & \cdots & P_{S_{2,24}} \\ \vdots & \vdots & \ddots & \vdots \\ P_{S_{D,1}} & P_{S_{D,2}} & \cdots & P_{S_{D,24}} \end{bmatrix} = \begin{bmatrix} P_{S_1}^T \\ P_{S_2}^T \\ \vdots \\ P_{S_D}^T \end{bmatrix}. \tag{5}$$

We use the clustering algorithm described in Section II-A to derive the classes and associated probabilities for the solar generation output. Thus, we have the classes $\mathscr{C}_{S_1}, \ldots, \mathscr{C}_{S_k}$, composed of the similar days of solar generation output. We interpret each class $\mathscr{C}_{S_c}$, $c = 1, \ldots, k$, as a realization of the set of hourly solar generation output random variable $\rho_{c,h}$ for $h = 1, \ldots, 24$. The mean value of each random variable $\mu_{S_{c,h}}$ is defined as

$$\mu_{S_{c,h}} = \frac{1}{|\mathscr{C}_{S_c}|} \sum_{P_{S_d} \in \mathscr{C}_{S_c}} P_{S_{d,h}}, \tag{6}$$

where $|\mathscr{C}_{S_c}|$ is the number of solar generation vectors P_{S_d} that are part of class $\mathscr{C}_{S_c}$. For each class $\mathscr{C}_{S_c}$ we construct the

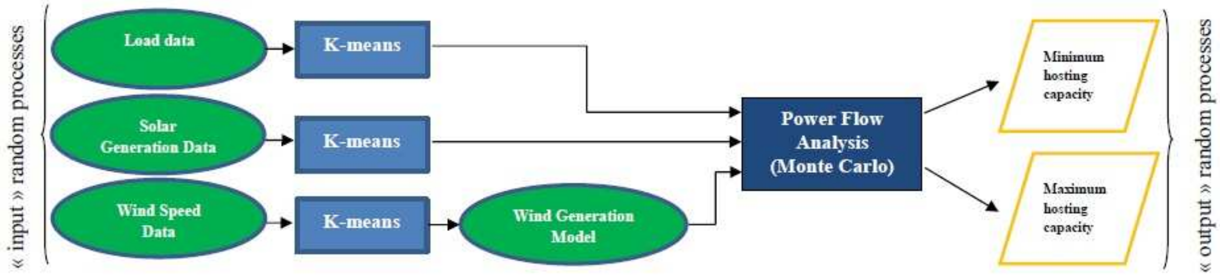

Fig. 1: Conceptual structure of the proposed stochastic framework.

vector that represents the daily solar generation output pattern by $\mu_{S_c} = [\mu_{S_{c,1}}, \ldots, \mu_{S_{c,24}}]^T$. The probability of each class $\mathscr{C}_{W_c}$ occurring is p_{S_c}, and is equal to

$$p_{S_c} = \frac{|\mathscr{C}_{S_c}|}{D}. \tag{7}$$

C. Load Modeling

For each of the three phases $\phi = 1, 2, 3$ we collect historical data over D days. The load for day d at hour h for phase ϕ is denoted by $P_{L^\phi_{d,h}}$. We combine the loads at the three phases and construct the load vector for day d as

$$\begin{aligned} P_{L_d} \quad = \quad & [P_{L^1_{d,1}}, \ldots, P_{L^1_{d,24}}, P_{L^2_{d,1}}, \ldots, P_{L^2_{d,24}}, \\ & P_{L^3_{d,1}}, \ldots, P_{L^3_{d,24}}]^T. \end{aligned} \tag{8}$$

We stack all the phases together to maintain the load correlation between phases. Then, we construct the matrix P_L that contains the load for every hour for all the days D during the period under consideration. Thus, we have $P_L = [P^T_{L_1}, P^T_{L_2}, \ldots, P^T_{L_D}]^T$.

Same as in the previous sections we calculate the classes $\mathscr{C}_{L_1}, \ldots, \mathscr{C}_{L_k}$, composed of the similar days of load. We interpret each class $\mathscr{C}_{L_c}$, $c = 1, \ldots, k$, as a realization of the set of the load random variable $\ell_{c,h}$ for $h = 1, \ldots, 24$. The mean value of each random variable $\mu_{L_{c,h}}$ is defined as

$$\mu_{L_{c,h}} = \frac{1}{|\mathscr{C}_{L_c}|} \sum_{P_{L_d} \in \mathscr{C}_{L_c}} P_{L_{d,h}}, \tag{9}$$

where $|\mathscr{C}_{L_c}|$ is the number of load vectors P_{L_d} that are part of class $\mathscr{C}_{L_c}$. For each class $\mathscr{C}_{L_c}$ we construct the vector that represents the daily three-phase load pattern by $\mu_{L_c} = [\mu_{L_{c,1}}, \ldots, \mu_{L_{c,24}}]^T$. The probability of each class $\mathscr{C}_{L_c}$ occurring is p_{L_c}, and is equal to

$$p_{L_c} = \frac{|\mathscr{C}_{L_c}|}{D}. \tag{10}$$

D. Distribution Network

We consider a power distribution system with N buses indexed by $\mathscr{N} = \{0, 1, \ldots, N-1\}$. Let node 0 represent the point of common coupling (PCC). Such networks are mostly radial and the network topology can be described by a connected tree, the edge set of which is denoted by $\mathscr{E}$, where $(i, k) \in \mathscr{E}$ if i is connected to k by a line. We denote by $\mathscr{P}_n = \{a_n, b_n, c_n\}$ the phases at bus $n \in \mathscr{N}$. Let $V^\phi_n = |V^\phi_n| \angle \theta^\phi_n$ (I^ϕ_n) denote the bus n voltage (current) at phase ϕ. Similarly, let $P^\phi_{L_n}$ and $Q^\phi_{L_n}$ denote the active and reactive power demanded by a wye-connected load at bus n. We denote by $P^\phi_{S_n}$ ($P^\phi_{W_n}$) the solar (wind) output at bus n. The active power flow equation for the distribution system is given by

$$P^\phi_{S_n} + P^\phi_{W_n} - P^\phi_{L_n} = V^\phi_n I^{\phi^\star}_n, \forall \phi \in \mathscr{P}_n, n \in \mathscr{N}. \tag{11}$$

In this paper we use a distribution planning tool to solve the power flow. We include (11) to show how the uncertainty and variability in the output of solar and wind generation affects the power flow in distribution systems.

III. Proposed Stochastic Simulation Framework

The proposed framework is based on a steady-state analysis that emulates the integration of wind and solar DG on a selected feeder. We choose the 24-hour window to capture the time-dependent, variable, and intermittent nature of the wind and solar resources and the correlation between the chronological load and the uncontrollable DG output. The effects of DG on a feeder's hosting capacity are affected by its location as well as its output. In order to accurately assess the DG impacts on the distribution systems and capture the uncertainty in the size and the location of potential DG installations, in the wind and solar generation clustering we include a class that has zero elements for all hours, and assign a probability of that class occurring. We do so to capture the event that several buses might not have DG.

The simulation emulates the three-phase unbalanced power flow of a distribution system feeder for various outputs of DG. Specifically, in each simulation period, we emulate the power flow output over a 24-hour period. The modeling of the highly variable DG, which is uncertain, is in terms of a collection of discrete random variables (d.r.v.). The probability distribution function of the d.r.v. is established from the k-means clustering algorithm. We make use of independent Monte Carlo simulations [9, p. 10], and construct multiple independent and identically distributed (i.i.d.) sample paths for each of the wind, solar, and load random variables. For each simulation run, we use one sample path for each one of

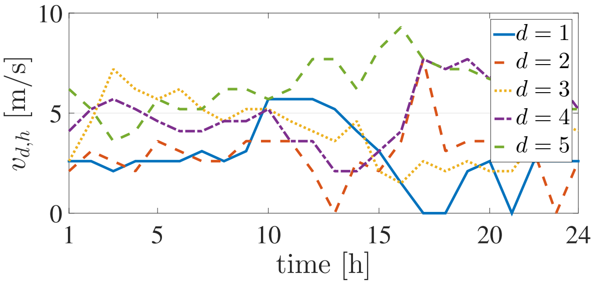

Fig. 2: Hourly wind speed for 5 days.

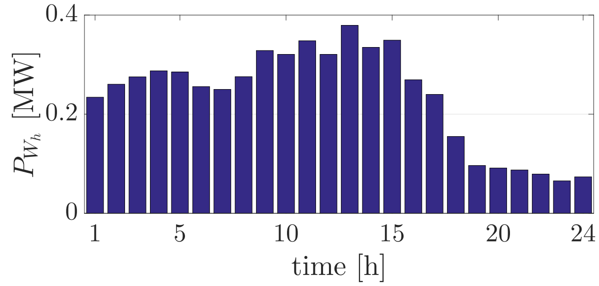

Fig. 4: Hourly wind output for cluster $\mathscr{C}_1$.

the input random variables to calculate system performance metrics. A conceptual structure of the proposed framework is depicted in Fig. 1.

The output of the power flow generates the voltage at each bus. We gather the maximum voltages of each 24-hour run, and compare them with the ANSI voltage limit to determine which penetration levels in each realization might cause problems in the feeder [10, p .2]. The minimum hosting capacity refers to the penetration level where the first violation is observed. The maximum hosting capacity refers to the penetration level where all the maximum voltages exceed the ANSI voltage limit. At each 24-hour simulation run we generate realizations of the minimum or the maximum hosting capacity. We use these realizations to approximate the statistics of the minimum and the maximum hosting capacity of a feeder. We utilize the statistics and build an empirical pdf. However, the result of the pdf has the form of a step function due to the finite amount of data. In order to smooth the pdf we use the kernel density estimation (KDE) method [11]. In this regard, the outputs of the proposed stochastic simulation framework are smoothed pdfs of the minimum and maximum hosting capacity.

IV. Numerical Results

We performed extensive testing of the proposed framework to validate the results. In this paper, we illustrate the application of the framework with a representative utility feeder, which contains 118 nodes. We collected historical for the wind speed, solar generation, and load data over a one-year period. For the solar generation output we utilize an open source NREL software [12]. We demonstrate how the wind generation output was determined. In Fig. 2 we depict the wind speed at a particular site for 5 random days. We use the k-means algorithm to group the wind speed data into 10 clusters that have similar patterns and define the probability of such a pattern occurring. In Fig. 3, we depict the 10 classes, $\mathscr{C}_{W_1}, \ldots, \mathscr{C}_{W_{10}}$, composed of the similar days of wind speed data. In order to define the wind generation output P_W based on a certain daily wind speed pattern, we use the following function for the wind output curve

$$P_W = \begin{cases} 0 & , 0 \leq v < 3 \\ -2.6 \cdot 10^{-3} + 8.8 \cdot 10^{-4} v^3 & , 3 \leq v < 12 \\ 1.5 & , 12 \leq v < 20 \\ 0 & , v \geq 20 \end{cases}, \quad (12)$$

where v is the wind speed. The wind generation for cluster $\mathscr{C}_{W_1}$ is depicted in Fig. 4. A similar approach is conducted for the solar and load data, as described in Section II.

Once the patterns are determined, we sample them and carry out 100 simulation runs to estimate the output random processes of maximum and minimum hosting capacity through a power flow simulation on the distribution planning software. In Fig. 5, we depict the histogram of the maximum hosting capacity. We smooth the discontinuities with the KDE, and depict in Fig. 6 the pdf of the maximum hosting capacity. Similarly, we calculate the pdf of the minimum hosting capacity (see Fig. 7). Based on these values, utilities may decide what confidence level is acceptable and how much risk they are willing to undertake in DG installation. In this specific example, a moderate-risk approach would be to allow a penetration of 10,000 kW. If the 15% rule was used in this specific example then the hosting capacity would be 647 kW. Thus, we may see that the heuristic approach provides very conservative results in this particular feeder, whereas our proposed framework provides more realistic results as depicted in Figs. 6, 7. A stochastic method that evaluates the

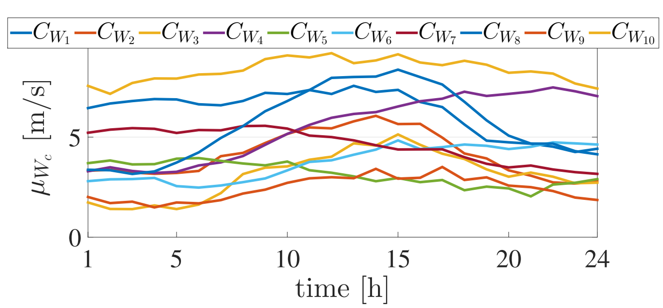

Fig. 3: Wind speed clusters.

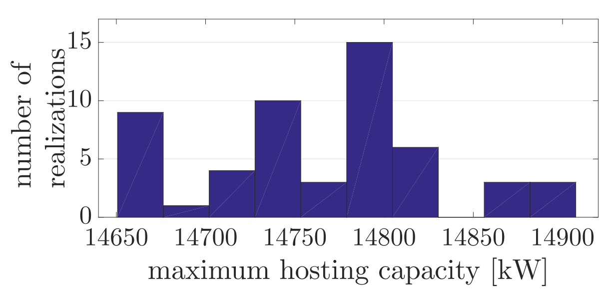

Fig. 5: Histogram of maximum hosting capacity.

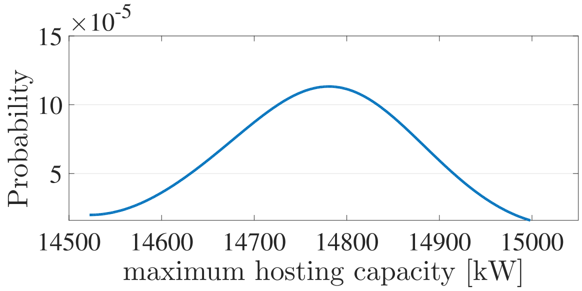

Fig. 6: Probability distribution function of maximum hosting capacity.

hosting capacity of a feeder during minimum load conditions determines that the minimum hosting to be $4,000$ kW and maximum hosting capacity $10,200$ kW. The minimum load conditions may be considered as a "worst" case scenario. Therefore, a more detailed approach as the one described in this paper is beneficial for utilities planning.

V. Distribution System Practical Applications

While determining the hosting capacity allows utilities to determine the amount of DG that specific feeders can accommodate, it also provides additional wide reaching benefits in areas of risk management, investment deferment, and possible optimal DG deployment to minimize outages and improve reliability to the customer.

One such example is for interconnection purposes since it is yet unclear who pays for the additional upgrades needed to interconnect to the system. The proposed framework may be used to determine the interconnection charges based on a calculated risk factor for the addition of each DG unit. Furthermore, the proposed framework may also allow determine where and when issues might arise as a result of DG units deployment. Knowledge such as this could be utilized for planning purposes in order to identify the ideal location for DG installations so as to minimize necessary upgrades; or to take a more proactive approach into the management of the distribution system. Lastly, as an increasing amount of DG and storage technologies are connected to the grid, there is an opportunity to utilize them to minimize the number and the duration of outages. The proposed framework can present an analytic approach to determine which DG should be utilized without causing adverse effects elsewhere in the system.

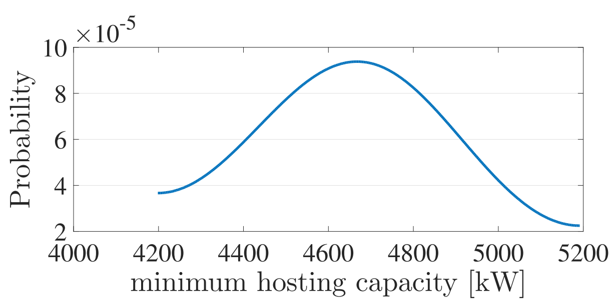

Fig. 7: Probability distribution function of minimum hosting capacity.

VI. Conclusions

In this paper, we presented a stochastic simulation framework for the determination of a feeder's hosting capacity. To this end, we developed probabilistic models for the wind and solar generation, and load profiles by using historical data and the k-means clustering algorithm. Next, we used the probabilistic models for the solar, wind, and load data, and fed them into a distribution planning software, which runs the power flow, to determine the voltages at each bus. We collected the maximum voltages of each wind, solar, and load realization to determine statistics for the feeder's minimum and maximum hosting capacity. We used the KDE technique to obtain a smoothed pdf of the minimum and maximum hosting capacity.

In the numerical result section, we applied the proposed framework into a representative 118-node feeder, showed the derivation of the probabilistic models with the k-means clustering technique, and used the KDE technique to calculate the pdfs of the minimum and maximum hosting capacity. We compared our findings with that of the 15% process, and showed the process to be quite conservative and that it did not exemplify a good perspective for planning. We also proposed the utilization of the proposed framework towards other related applications for the distribution systems.

References

[1] M. Coddington *et al.*, "Updating interconnection screens for pv system integration," National Reneable Energy Laboratory, Tech. Rep. NREL/TP-5500-54063, Feb. 2012.

[2] E. K. Hart, E. D. Stoutendburg, and M. Z. Jacobson, "The potential of intermittent renewables to meet electric power demand: Current methods and emerging analytical techniques," in *Proceedings of IEEE*, vol. 100, Feb. 2012, pp. 322–334.

[3] J. Smith *et al.*, "A new method for characterizing distribution system hosting capacity for distributed energy resources: A streamlined approach for solar photovoltaics," Electric Power Research Institute, Tech. Rep. 3002003278, Dec. 2014.

[4] K. Coogan, M. J. Reno, S. Grijalva, and R. J. Broderick, "Locational dependence of pv hosting capacity correlated with feeder load," pp. 1–5, Apr. 2014.

[5] (2015, Accessed Aug.) California renewable portfolio standards. [Online]. Available: http://www.cpuc.ca.gov/PUC/energy/Renewables/index.htm

[6] F. J. Ruiz-Rodriguez, J. C. Hernandez, and F. Jurado, "Probabilistic load flow for radial distribution networks with photovoltaic generators," *IET Renewable Power Generation*, vol. 6, no. 2, pp. 110–121, Mar. 2012.

[7] N. Maisonneuve and G. Gross, "A production simulation tool for systems, with integrated wind energy resources," *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 2285 –2292, Nov. 2011.

[8] C. D. Mannin, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge University Press, 2008.

[9] J. Kleijnen, *Statistical techniques in simulation - Part 1 and 2*. Marcel Dekker, Inc. New York,, 1974.

[10] PacifiCorp, *Voltage Level and Range*, ser. Discussion Paper Series. Engineering Handbook, 2007.

[11] J.-N. Hwang, S.-R. Lay, and A. Lippman, "Nonparametric multivariate density estimation: A comparative study," *IEEE Transactions on Signal Processing*, vol. 42, no. 10, pp. 2795–2810, Oct. 1994.

[12] (2015, Accessed Aug.) NREL's pvwatts calculators. [Online]. Available: http://pvwatts.nrel.gov/index.php