Electrical Power and Energy Systems 134 (2022) 107426

Contents lists available at ScienceDirect

International Journal of Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes

# A novel framework for hosting capacity analysis with spatio-temporal probabilistic voltage sensitivity analysis

Sai Munikoti [a,*], Mohammad Abujubbeh [a], Kumarsinh Jhala [b], Balasubramaniam Natarajan [a]

[a] Department of Electrical and Computer Engineering, Kansas State University, Manhattan, KS, USA
[b] CEEESA in the Energy Systems Division at Argonne National Laboratory, IL, USA

ARTICLE INFO

*Keywords:*
Photovoltaic (PV)
Probabilistic
Renewable
Analytical
Voltage Violations
Hosting capacity

ABSTRACT

Smart grids are envisioned to accommodate high penetration of distributed photovoltaic (PV) generation, which may cause adverse grid impacts in terms of voltage violations. Therefore, PV Hosting capacity is being used as a planning tool to determine the maximum PV installation capacity that causes the first voltage violation and above which would require infrastructure upgrades. Traditional methods of Hosting capacity analysis are scenario based and computationally complex as they rely on iterative load flow algorithms that require investigating a large number of scenarios for accurate assessment of PV impacts. Therefore, this paper presents a computationally efficient analytical approach to compute the probability distribution of voltage change due to random behavior of randomly located multiple distributed PVs. The proposed approach is based on Spatio-temporal probabilistic voltage sensitivity analysis that exploits both spatial and temporal uncertainties associated with PV injections. Thereafter, the derived distribution is used to quantify voltage violations for various PV penetration levels and subsequently determine the hosting capacity of the system without the need to examine large number of scenarios. Results of the proposed framework are validated via conventional load flow based simulation approach on the IEEE 37 and IEEE 123 node test systems.

## 1. Introduction

Power grid is undergoing significant changes to meet modern- energy demand in a more efficient manner. Integration of renewable energy sources, especially rooftop Photovoltaics (PVs) offers various solutions, including (1) low carbon footprint; (2) reduced operational cost; (3) ancillary services in terms of peak load shaving and voltage restoration at critical loads during contingencies. Therefore, many countries are aiming to meet a major portion of energy demand through renewable energy sources. For example, by 2050, USA, China, EU and India are projected to have 63%, 67%, 70% and 73% of their total energy use met through renewables, respectively [1]. Despite the aforementioned benefits, high penetration levels of PVs may impact the grid negatively in terms of voltage fluctuations and stability. This necessitates the need for thoroughly analyzing the grid in the presence of PVs to maximize their integration benefits. Therefore, utilities are interested in quantifying the maximum allowable PV penetration level in the system. In this regard, PV hosting capacity (HC) as a planning tool has attracted the attention of many researchers and practitioners. HC refers to the maximum amount of PV generation that can be accommodated in the distribution system, while keeping system operational constraints within their safe limits, without the need for infrastructure upgrades.

A comprehensive HC analysis monitors power quality, power loss, thermal overload, protection devices, and voltage deviation for different PV penetration levels. With increasing PV penetration levels in distribution systems, many operational issues have emerged, including voltage violations [2,3] which directly impact the HC. Therefore, the development of an accurate, yet computationally efficient HC analysis framework is essential to ensure efficient, economic, and reliable operation of the distribution system. Most of the existing methods for evaluating HC are scenario-based and require the execution of multiple load-flow runs for various PV allocation scenarios [4,5]. The drawbacks of scenario-based HC studies are (1) high computational complexity, which increases with the size of the network; (2) scenario-based and scenario-specific results that do not provide any performance guarantee for a more general case; and 3) very conservative results, typically based on worse case scenarios. As voltage is the primary concern for many utilities while determining the HC [6,4], voltage sensitivity analysis (VSA) can help identify voltage violations, which in turn can be used to compute the HC of the system. Traditional methods of VSA such as load

* Corresponding author.
*E-mail address:* saimunikoti@ksu.edu (S. Munikoti).

https://doi.org/10.1016/j.ijepes.2021.107426
Received 8 April 2021; Received in revised form 14 June 2021; Accepted 13 July 2021
Available online 6 August 2021
0142-0615/© 2021 Elsevier Ltd. All rights reserved.

flow based numerical approaches have been used by various researchers [7–9] to compute the HC. As numerical VSA methods are simulation-based, they possess the same demerits as Monte-Carlo based power flow methods. Additionally, at the planning stage, there could be scenarios where the grid operator may not be aware of the actual PV locations in the network. Under these scenarios of random power change at unknown locations of the grid, numerical VSA approaches would involve simulating a large number of scenarios to account for temporal and spatial uncertainties associated with power change. This would result in high computational complexity. To overcome the drawbacks of numerical methods, there are some limited analytical approaches for VSA that have been proposed in [10–14]. The analytical VSA paradigm enables quick, yet accurate, estimation of node vulnerability to voltage violations, however, these approaches do not apply to HC in their original formulations. For instance, authors in [13], provides an analytical formulation of voltage change for random power changes but the proposed approach is valid only for the balanced network. Similarly, in [14], the proposed analytical formulation only accounts for the temporal uncertainty whereas for Hosting capacity, one needs to incorporate both temporal and spatial uncertainty associated with PV generations in a distribution system. Motivated by these shortcomings of existing analytical approaches, this paper proposes a novel analytical spatio-temporal probabilistic voltage sensitivity analysis (ST-PVSA) framework for computing the probability distribution of voltage change. The proposed ST-PVSA is applicable to three phase unbalanced distribution systems and incorporates uncertainties associated with both power change (temporal randomness due to intermittent PV generation) and the locations of PV units (spatial randomness) in the system. The strength of this paradigm lies in its ability to account for spatio-temporal uncertainties associated with power changes in a computationally efficient manner. This fundamental theoretical result is then employed to simplify HC computation.

### 1.1. Related work

Authors in [15,16] have reviewed the literature related to HC and fundamentally classified the efforts into four major categories (deterministic, stochastic ([17,18]), optimization-based ([19]), and streamlined ([5])) based on the available data and the type of study to be performed. Most of these approaches for HC depend on numerical load flow-based methods [7,4,8,9], and involve the analysis of multiple PV deployment scenarios. For instance, a rigorous framework is developed in [17], where authors have incorporated both aleatoric (base voltages, solar PV production, local consumption) and epistemic uncertainty (installed capacity per customer, number of customers with solar PV, phase to which single-phase units are connected) of active distribution network individually. Similarly, in [4], a scenario is generated by randomly allocating PVs in the network, and then load flow is executed for each penetration level until a voltage violation is encountered. To cover all possible locations, the complete process is repeated for multiple scenarios, thereby presenting a huge computational burden. [8] assigns each feeder a minimum and maximum HC, corresponding to the most conservative and most optimistic HC value. However, the overvoltage risk within the range of two HC endpoint values is not quantified. Furthermore, authors in [5], propose a quasi-static-time-series (QSTS) based dynamic PV HC methodology. Here, power flow analysis is conducted on the load and PV data over one year, where the time duration of violation is also monitored along with the total violations count. For a real distribution model with thousands of nodes and one-second resolution data, an annual simulation could take a few days [5]. Furthermore, the PV and load uncertainties have significant influences upon HC values. As a result, probabilistic HC methods have gained attention [20–22]. The authors in [21] determine HC by incorporating uncertainties associated with PV, Wind turbines, and loads while. However, the approach does not consider the stochastic random distribution of DGs along the target feeder. Though the probabilistic approaches can effectively describe the uncertainty in fluctuations of PVs and loads, most of these approaches are simulation-based and thus are computationally inefficient. More importantly, the performance of these approaches relies heavily on the availability of data.

In addition to HC, numerical VSA methods have also been used to guide various grid applications such as voltage regulation, DER allocations, etc. [23,24]. For instance, authors in [23] propose a method for analyzing voltage variations due to PV generation fluctuations, considering a variety of factors. However, its dependency on the inefficient simulation method limits its applications to large scale distribution networks. Similarly, authors in [24] develop an optimization model for the electric vehicle charging schedule based on VSA approaches. Still, the requirements of iterative executions of power flow calculations and optimization models hinder its application in real-world scenarios. Thus, traditional methods suffer from high computational complexity and do not provide analytical insight into the underlying physics of the system. Therefore, to overcome the drawbacks of numerical methods, there are some limited analytical approaches for VSA that have been proposed. Authors in [10,25], develop an algorithm based on VSA which optimally manages active and reactive powers of DGs to keep the system voltages inside the limits. Here, instead of repeating load flow calculations to solve the optimization problem, a sensitivity matrix is used to conduct load flow computation in a non-iterative manner, reducing the computational burden significantly. However, the algorithms proposed are not properly validated with standard test systems. Authors in [11], have taken a probabilistic approach where smart meter data is used along with sensitivity analysis to define boundary values of various operation indices. This approach does not account for unbalanced load conditions. The authors in [14], develop a new probabilistic voltage sensitivity approach to quantify voltage change in a computationally efficient way and accounts for the temporal uncertainties associated with random power change at fixed locations of the grid. However, the assumption of a fixed location is not generic enough to account for unknown locations of PV installations. Therefore, in [13], authors provide a quick and efficient tool for estimating the distribution of voltage change in a distribution system with spatial randomness in PV installations. However, the proposed framework is valid only for single phase balanced networks and does not hold true for an unbalanced distribution system.

To summarize, probabilistic VSA represents a viable, low complexity, systematic approach to HC computation. Therefore, in this work, a generic computationally efficient framework for probabilistic voltage sensitivity is developed for an unbalanced distribution system that systematically accounts for both spatial and temporal uncertainty associated with PV generation. The ST-PVSA framework is then employed to efficiently determine the HC of an unbalanced distribution system. The proposed formulation helps grid operators prepare for the future modernized grid by providing new insights on the impact of high PV penetration on grid operations.

### 1.2. Contributions

This work proposes a novel stochastic method of grid voltage sensitivity analysis which is used to calculate the likelihood of node voltage exceeding operational bounds. This framework is then employed to determine the PV Hosting capacity (HC) of an unbalanced distribution system. The key contributions of this paper are listed below:

- For the first time, the probability distribution of voltage change due to random change in complex power (temporal) across random locations (spatial) of a three phase unbalanced distribution grid is derived analytically.
- The proposed ST-PVSA method is used to (1) analyze the aggregate effect of spatial random distribution of PVs on the feeder voltage, and (2) determine the probability of node voltages exceeding allowable limits. Analytical results are validated using simulation on the IEEE 37-node and 123-node test systems.

- ST-PVSA involves a substantial change in the formulation compared to [14,13] since incorporating both spatial and temporal randomness in a three phase setting requires extensive mathematical and statistical analysis for obtaining the probability distribution of voltage change.
- The proposed ST-PVSA is employed to efficiently and accurately determine PV HC in a significantly faster way. For e.g., three orders faster in IEEE 123-node network relative to the existing load flow-based approach.

The rest of the paper is organized as follows: Section 2 discusses the typical simulation-based method for HC. Then, the probability distribution of voltage change with spatio-temporal uncertainties is derived in Section 3, followed by its validation in Section 4. In Section 5, the distribution is used to determine the HC and validated with a load flow-based approach, and finally, conclusions are provided in Section 6.

## 2. HC with simulation-based approach

This section describes a typical load flow based approach of determining the HC of the system. Here, the net power injection is increased in steps by allocating power to PVs located at random locations of the network. Then, load flow is executed for each penetration levels to track the number of node voltage violations throughout the network. This process is repeated for increasing penetration levels until the number of violations exceeds the threshold. The corresponding power (penetration level) is the HC for a particular PV deployment scenario. Thereafter, the complete process is repeated multiple times to cover all possible spatial distribution of the PV installations and the minimum capacity across all such scenarios is the final HC of the network. The scenario based analysis presents a huge computational burden due to the requirement of multiple load flow runs. Fig. 1 depicts the flow chart of the existing load flow based approach of computing HC [4].

Alternatively, this paper attempts to develop a probabilistic VSA approach that determines the HC in a computationally efficient manner. As mentioned in [7], a comprehensive analysis of PV distribution needs to monitor voltage, protection, power quality and control limits. However, voltage is the primary concern for many utilities [6,4]. So, similar to [4,5], this paper only considers voltage limits to determine the PV HC. The first step towards the probabilistic VSA approach for HC is to derive an analytical expression of voltage sensitivity due to random power change at random locations in the network, as presented in the next section.

## 3. ST-PVSA for random distribution of PVs

This section details the steps involved in the derivation of the probability distribution of voltage change at network nodes due to random power changes at random locations of the network. Throughout this paper, observation nodes are referred to those nodes where voltage change is observed and actor nodes are those where power changes. The change in complex voltage at any phase (say phase $a$) of observation node $O$ due to change in complex power at any phase of a single actor node $A$ is given by [14],

$$\Delta V_{OA}^{a} \approx -\left[\frac{\Delta S_A^{a^*} Z_{OA}^{aa}}{V_A^{a^*}} + \frac{\Delta S_A^{b^*} Z_{OA}^{ab}}{V_A^{b^*}} + \frac{\Delta S_A^{c^*} Z_{OA}^{ac}}{V_A^{c^*}}\right], \tag{1}$$

where, superscript $a, b$ and $c$ represent the three phases; this notation is used throughout the paper. $V_A^{a\star}$ and $\Delta S_A^{a}$ represent complex conjugate of voltage at phase $a$ and complex power change at actor node $A$, respectively; $Z_{OA}$ denotes the impedance matrix including self and mutual line impedance of the shared path between observation node and actor node from the source node. The subscript $A$ represents the actor node where power is varying. In [14], the authors use the Eqn. (1) to derive the distribution of voltage change. However, the distribution only incorporates the temporal uncertainty associated with power change, thereby prevents its uses in hosting capacity analysis. Therefore, here, we leverage Eqn. (1) to derive a generic probability distribution of voltage change, considering both temporal and spatial randomness in a three phase unbalanced network. In this regard, the complex voltage change in (1) is decomposed into real and imaginary parts as,

$$\Delta V_{OA}^{a} = \Delta V_{OA}^{a,r} + j\Delta V_{OA}^{a,i}. \tag{2}$$

For simplicity, the voltage change expression throughout the derivation is shown for a single phase (phase $a$). However, similar form and approach is applicable to other phases as well. On expanding power change ($\Delta S_A^{a\star} = \Delta P - j\Delta Q$) and impedance ($Z_{OA} = R + jX$) components in (1), the real ($\Delta V_{OA}^{a,r}$) and imaginary parts ($\Delta V_{OA}^{a,i}$) of voltage change at phase $a$ of the observation node $O$ can be written as,

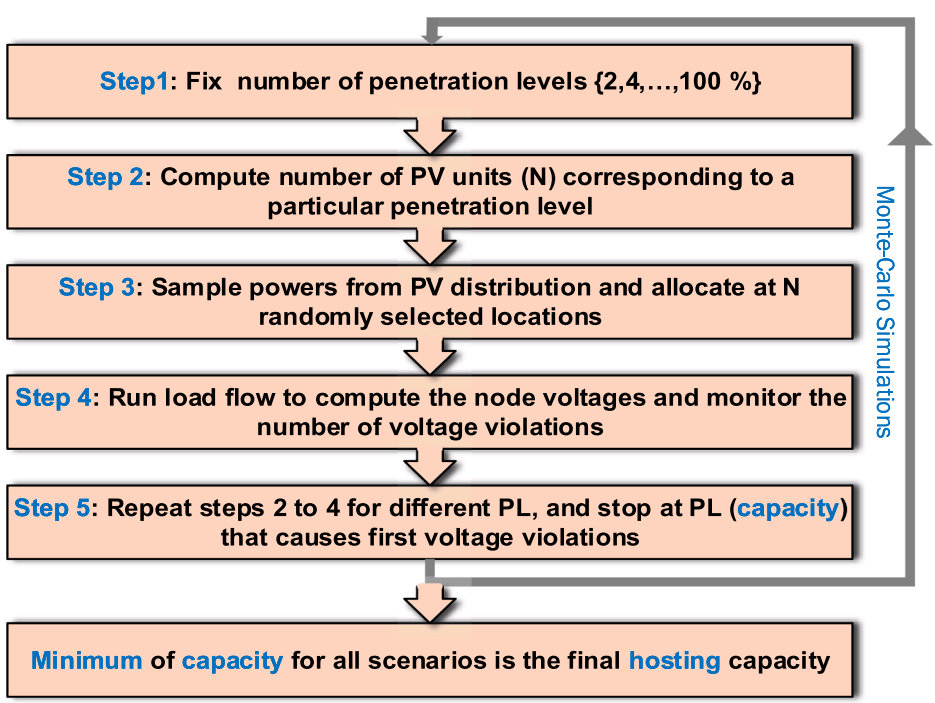


**Fig. 1.** Flowchart of Load flow based HC method.

$$\Delta V_{OA}^{a,r} = \sum_{h,u} \frac{-1}{|V_A^h|}\left[\Delta P_A^h\left(R_{OA}^u cos\left(\omega_A\right) - X_{OA}^u sin\left(\omega_A\right)\right) + \Delta Q_A^h\left(R_{OA}^u sin\left(\omega_A\right) + X_{OA}^u cos\left(\omega_A\right)\right)\right]$$
$$\Delta V_{OA}^{a,i} = \sum_{h,u} \frac{-1}{|V_A^h|}\left[\Delta P_A^h\left(R_{OA}^u sin\left(\omega_A\right) + X_{OA}^u cos\left(\omega_A\right)\right) + \Delta Q_A^h\left(X_{OA}^u sin\left(\omega_A\right) - R_{OA}^u cos\left(\omega_A\right)\right)\right] \quad (3)$$

where $h \in \widetilde{H}$ and $u \in \widetilde{U}$. The set $\widetilde{H} = \{a, b, c\}$ denotes different phases and the set $\widetilde{U} = \{aa, ab, ac\}$ represents phase sequence for the corresponding phase. $\Delta P_A^h$ and $\Delta Q_A^h$ are the active and reactive power changes, respectively. $R_{OA}^h$ and $X_{OA}^h$ are the resistance and reactance of shared path between the observation node $O$ and actor node $A$from the source node. $V_A^h$denotes the complex rated voltage of actor node $A$. The magnitude and angle of voltage at a particular phase, say phase $a$, of node $A$ are given by $|V_A^a|$ and $\theta_A^a$, respectively with reference to the slack bus. $\omega_A$ denotes the rated voltage angle of the actor node $A$. The detailed steps to obtain Eq. (3) from Eq. (2) are described in the Appendix. Line voltage of the network is always kept within permissible limits, and thus it is reasonable to assume the phase difference of 120° between the voltage angles of different phases with the same angle for all the node voltages of each phase. Based on this assumption, $\Delta V_{OA}^{a,r}, \Delta V_{OA}^{a,i}$ can be rewritten as:

$$\Delta V_{OA}^{a,r} = \left[\frac{-\Delta P_A^a R_{OA}^{aa}}{|V_A^a|} + \frac{\Delta P_A^b}{|V_A^b|}\left(\frac{R_{OA}^{ab}}{2} - \frac{\sqrt{3}X_{OA}^{ab}}{2}\right) + \frac{\Delta P_A^c}{|V_A^c|}\left(\frac{R_{OA}^{ac}}{2} + \frac{\sqrt{3}X_{OA}^{ac}}{2}\right) - \frac{\Delta Q_A^a X_{OA}^{aa}}{|V_A^a|} + \frac{\Delta Q_A^b}{|V_A^b|}\left(\frac{\sqrt{3}R_{OA}^{ab}}{2} + \frac{X_{OA}^{ab}}{2}\right) + \frac{\Delta Q_A^c}{|V_A^c|}\left(\frac{-\sqrt{3}R_{OA}^{ac}}{2} + \frac{X_{OA}^{ac}}{2}\right)\right] \quad (4)$$

$$\Delta V_{OA}^{a,i} = \left[\frac{-\Delta P_A^a X_{OA}^{aa}}{|V_A^a|} + \frac{\Delta P_A^b}{|V_A^b|}\left(\frac{\sqrt{3}R_{OA}^{ab}}{2} + \frac{X_{OA}^{ab}}{2}\right) + \frac{\Delta P_A^c}{|V_A^c|}\left(\frac{-\sqrt{3}R_{OA}^{ac}}{2} + \frac{X_{OA}^{ac}}{2}\right) + \frac{\Delta Q_A^a R_{OA}^{aa}}{|V_A^a|} + \frac{\Delta Q_A^b}{|V_A^b|}\left(-\frac{R_{OA}^{ab}}{2} + \frac{\sqrt{3}X_{OA}^{ab}}{2}\right) + \frac{\Delta Q_A^c}{|V_A^c|}\left(-\frac{R_{OA}^{ac}}{2} - \frac{\sqrt{3}X_{OA}^{ac}}{2}\right)\right] \quad (5)$$

The real (4) and imaginary (5) parts of the voltage change can further be represented in a simplified form as,

$$\Delta V_{OA}^{a,r} = (\mathbf{Z^{a,r}})^{\mathbf{T}}\Delta\mathbf{S}, \Delta V_{OA}^{a,i} = (\mathbf{Z^{a,i}})^{\mathbf{T}}\Delta\mathbf{S} \quad (6)$$

$$\mathbf{Z^{a,r}} = \begin{bmatrix} -R_{OA}^{aa} \\ \frac{R_{OA}^{ab}}{2} - \frac{\sqrt{3}X_{OA}^{ab}}{2} \\ \frac{R_{OA}^{ac}}{2} + \frac{\sqrt{3}X_{OA}^{ac}}{2} \\ -X_{OA}^{aa} \\ \left[1pt\right]\frac{\sqrt{3}R_{OA}^{ab}}{2} + \frac{X_{OA}^{ab}}{2} \\ \frac{-\sqrt{3}R_{OA}^{ac}}{2} + \frac{X_{OA}^{ac}}{2} \end{bmatrix}, \mathbf{Z^{a,i}} = \begin{bmatrix} -X_{OA}^{aa} \\ \frac{\sqrt{3}R_{OA}^{ab}}{2} + \frac{X_{OA}^{ab}}{2} \\ \frac{-\sqrt{3}R_{OA}^{ac}}{2} + \frac{X_{OA}^{ac}}{2} \\ R_{OA}^{aa} \\ \frac{-R_{OA}^{ab}}{2} + \frac{\sqrt{3}X_{OA}^{ab}}{2} \\ \frac{-R_{OA}^{ac}}{2} - \frac{\sqrt{3}X_{OA}^{ac}}{2} \end{bmatrix}, \Delta\mathbf{S} = \begin{bmatrix} \frac{\Delta P_A^a}{|V_A^a|} \\ \frac{\Delta P_A^b}{|V_A^b|} \\ \frac{\Delta P_A^c}{|V_A^c|} \\ \frac{\Delta Q_A^a}{|V_A^a|} \\ \frac{\Delta Q_A^b}{|V_A^b|}\frac{\Delta Q_A^c}{|V_A^c|} \end{bmatrix}$$

where $\mathbf{Z^{a,r}}$ and $\mathbf{Z^{a,i}}$ are the vectors incorporating shared path impedance terms corresponding to real and imaginary parts of voltage change, respectively. To represent the random variation of PV generation, the real and reactive power change is modeled as a random variable. Consistent with the prior efforts in modeling PV generation as a time series with a trend component and Gaussian noise [26,27], the power variation is assumed to be Gaussian. It is important to note that the framework is quite general to account for any arbitrary random variable with finite mean and variance. Therefore, the vector $\Delta\mathbf{S}$, which incorporate the terms corresponding to the ratio of power change and constant base voltages, can be expressed as Gaussian random vector $\Delta\mathbf{S} \sim \mathcal{N}(\mu_{\Delta S}, \sum_{\Delta S})$ with $\mu_{\Delta S}$ being mean vector, and covariance matrix $\sum_{\Delta S}$ as,

$$\begin{bmatrix} \sigma^2_{\frac{pa}{|V_A^a|}} cov\left(\Delta P_A^a/|V_A^a|, \Delta P_A^b/|V_A^b|\right) \ldots cov\left(\Delta P_A^a/|V_A^a|, \Delta Q_A^c/|V_A^c|\right) \\ \vdots \\ cov\left(\Delta P_A^a/|V_A^a|, \Delta Q_A^c/|V_A^c|\right) \quad \ldots \quad \sigma^2_{\frac{Qc}{|V_A^c|}} \end{bmatrix}$$

Here, the diagonal and off-diagonal elements indicate variance and covariance among the terms that are ratio of power changes and base voltages across different phases of actor nodes, respectively. The impedance of the shared line between a given observation node ($O$) and a random actor node can be modeled as a correlated random variable. The mean, variance and covariance of resistance $R_{OA}$ and reactance $X_{OA}$ corresponding to a given observation node $O$ can be estimated based on actual line impedance data. In addition, let $\mu_{Z^r}$ and $\mu_{Z^i}$ represent the mean of real ($Z^{a,r}$) and imaginary ($Z^{a,i}$) parts of impedance vector, respectively. The average is taken over all the nodes of the network with respect to the observation node. Similarly, $\sum_{Z^r}$ and $\sum_{Z^i}$ denote the covariance matrices of $Z^{a,r}$ and $Z^{a,i}$, respectively. The correlation coefficient between the shared path impedances for various actor nodes is computed based on network parameters. Particularly, the objective of this work is to derive the probability distribution of the magnitude of voltage change at an observation node due to random power variation of PVs located at random nodes, which will further be used to estimate the system HC. The probability distribution of real $\Delta V_{OA}^{a,r}$ and imaginary components $\Delta V_{OA}^{a,i}$ of the voltage change due to random spatial distribution of multiple PV units can be derived using the following steps:

**Step 1:** *Compute mean and variance of* $\Delta V_{OA}^{a,r}$ and $\Delta V_{OA}^{a,i}$ due to a single actor node:

Using (6), the mean of the voltage change can be expressed as the expectation of product of two terms, i.e., the shared path impedance vector ($Z^{a,r}$ for real and $Z^{a,i}$ for imaginary part) and power change vector $\Delta S$. As the terms in the product are mutually independent, the expectation of their product can be applied to individual terms separately yielding the mean of real ($\mu_r$) and imaginary ($\mu_i$) parts as,

$$\mu_r = E\left[Z_o^{(a,r)^T}\Delta S\right] = \mu_{Z_o^{a,r}}\mu_{\Delta S}$$
$$\mu_i = E\left[Z^{(a,i)^T}\Delta S\right] = \mu_{Z_o^{a,i}}\mu_{\Delta S} \tag{7}$$

Furthermore, the variance of real and imaginary parts of the voltage change can be computed as shown below,

$$\mathrm{Var}\left(\Delta V_{OA}^{a,r}\right) = E\left[\left(Z^{(a,r)^T}\Delta S\right)^2\right] - E\left[\left(Z^{(a,r)^T}\Delta S\right)\right]^2$$
$$\mathrm{Var}\left(\Delta V_{OA}^{a,i}\right) = E\left[\left(Z^{(a,i)^T}\Delta S\right)^2\right] - E\left[\left(Z^{(a,i)^T}\Delta S\right)\right]^2. \tag{8}$$

Since $Z_r^T$ and $\Delta S$ are independent, the expectation of their product can be written in terms of product of their individual expectation as,

$$E\left[Z^{(a,r)^T}\Delta S\Delta S^T Z^{a,r}\right] - \left(E\left[Z^{(a,r)^T}\right]E\left[\Delta S\right]\right)^2. \tag{9}$$

For simplicity, the equation for variance is shown for the real part of voltage change and a similar form exists for imaginary part. Now, using the properties of matrix trace, the variance of the real part can be rewritten as,

$$\begin{aligned}
&E\left[Tr\left(Z_r Z_r^T\Delta S\Delta S^T\right)\right] - \left(\mu_{Z^r}\mu_{\Delta S}\right)^2 = Tr\left(E\left[Z_r Z_r^T\right]E\left[\Delta S\Delta S^T\right]\right) - \left(\mu_{Z^r}\mu_{\Delta S}\right)^2\\
&= Tr\left[\left(\mu_{Z^r}\mu_{Z^r}^T + \sum_{Z^r}\right)\left(\mu_{\Delta S}\mu_{\Delta S}^T + \sum_{\Delta S}\right)\right] - \left(\mu_{Z^r}\mu_{\Delta S}\right)^2\\
&= Tr\left(\mu_{Z^r}\mu_{Z^r}^T\mu_{\Delta S}\mu_{\Delta S}^T\right) + Tr\left(\mu_{Z^r}\mu_{Z^r}^T\sum_{\Delta S}\right) + Tr\left(\sum_{Z^r}\mu_{\Delta S}\mu_{\Delta S}^T\right) + Tr\left(\sum_{Z^r}\sum_{\Delta S}\right) - \left(\mu_{Z^r}^T\mu_{\Delta S}\right)^2
\end{aligned} \tag{10}$$

Now, the term $Tr(\mu_{Z^r}\mu_{Z^r}^T\mu_{\Delta S}\mu_{\Delta S}^T)$ is rearranged to $(\mu_{Z^r}\mu_{\Delta S})^2$, that cancels the last term of (10). After applying trace operator, the variance of real part can be expressed as,

$$\mu_{Z^r}^T\sum_{\Delta S}\mu_{Z^r} + \mu_{\Delta S}^T\sum_{Z^r}\mu_{\Delta S} + Tr\left(\sum_{Z^r}\sum_{\Delta S}\right) \tag{11}$$

Following the same steps from Eqs. (9)–(11), the variance of imaginary part of voltage change can be written as,

$$\mu_{Z^i}^T\sum_{\Delta S}\mu_{Z^i} + \mu_{\Delta S}^T\sum_{Z^i}\mu_{\Delta S} + Tr\left(\sum_{Z^i}\sum_{\Delta S}\right) \tag{12}$$

**Step 2:** *Compute covariance between real $\Delta V_{OA}^{a,r}$ and imaginary $\Delta V_{OA}^{a,i}$ parts of voltage change*:

The covariance between the real and imaginary parts of voltage change can be expressed as:

$$\begin{aligned}
Cov\left(\Delta V_{OA}^{a,r}, \Delta V_{OA}^{a,i}\right) &= E\left(\Delta V_{OA}^{a,r}\Delta V_{OA}^{a,i}\right) - E\left(\Delta V_{OA}^{a,r}\right)E\left(\Delta V_{OA}^{a,i}\right)\\
&= E\left[Z_A^{(a,r)^T}\Delta S_A Z_A^{(a,i)^T}\Delta S_A\right]
\end{aligned}$$

$\boldsymbol{Z}^{(a,r)^T}\Delta\boldsymbol{S}$ and $\boldsymbol{Z}^{(a,i)^T}\Delta\boldsymbol{S}$ are expanded using Eqn. (4) and (5) to express covariance as the expectation of following term,

$$\begin{aligned}
&E\Bigg[\Bigg[\frac{\Delta P_A^a}{|V_A^a|}\left(-R_{OA}^{aa}\right) + \frac{\Delta P_A^b}{|V_A^b|}\left(\frac{R_{OA}^{ab}}{2} - \frac{\sqrt{3}X_{OA}^{ab}}{2}\right) + \frac{\Delta P_A^c}{|V_A^c|}\left(\frac{R_{OA}^{ac}}{2} + \frac{\sqrt{3}X_{OA}^{ac}}{2}\right)\\
&- \frac{\Delta Q_A^a}{|V_A^a|}\left(X_{OA}^{aa}\right) + \frac{\Delta Q_A^b}{|V_A^b|}\left(\frac{\sqrt{3}R_{OA}^{ab}}{2} + \frac{X_{OA}^{ab}}{2}\right) + \frac{\Delta Q_A^c}{|V_A^c|}\left(-\frac{\sqrt{3}R_{O1}^{ac}}{2} + \frac{X_{O1}^{ac}}{2}\right)\Bigg]\times\\
&\Bigg[\frac{\Delta P_A^a}{|V_A^a|}\left(-X_{OA}^{aa}\right) + \frac{\Delta P_A^b}{|V_A^b|}\left(\frac{\sqrt{3}R_{OA}^{ab}}{2} + \frac{X_{OA}^{ab}}{2}\right) + \frac{\Delta P_A^c}{|V_A^c|}\left(\frac{-\sqrt{3}R_{OA}^{ac}}{2} + \frac{X_{OA}^{ac}}{2}\right)\\
&+ \frac{\Delta Q_A^a}{|V_A^a|}\left(R_{OA}^{aa}\right) + \frac{\Delta Q_A^b}{|V_A^b|}\left(-\frac{R_{OA}^{ab}}{2} + \frac{\sqrt{3}X_{OA}^{ab}}{2}\right) + \frac{\Delta Q_A^c}{|V_A^c|}\left(-\frac{R_{OA}^{ac}}{2} - \frac{\sqrt{3}X_{OA}^{ac}}{2}\right)\Bigg]\Bigg]
\end{aligned} \tag{13}$$

Terms inside the expectation operator are cross multiplied as,

$$\begin{aligned}
&\frac{\rho_{p^a}\sigma_{p^a}^2}{|V_A^a|^2}\mu_{R^{aa}}\mu_{X^{aa}} - \frac{\rho_{q^a}}{|V_A^a|^2}\sigma_{q^a}^2\mu_{R^{aa}}\mu_{X^{aa}} +\\
&\frac{\rho_{p^b}}{|V_A^b|^2}\sigma_{p^b}^2\left(0.43\mu_{R^{ab}}^2 - 0.5\mu_{R^{ab}}\mu_{X^{ab}} - 0.43\mu_{X^{ab}}^2\right) +\\
&\frac{\rho_{p^c}}{|V_A^c|^2}\sigma_{p^c}^2\left(-0.43\mu_{R^{ac}}^2 - 0.5\mu_{R^{ac}}\mu_{X^{ac}} + 0.43\mu_{X^{ac}}^2\right) +\\
&\frac{\rho_{q^b}}{|V_A^b|^2}\sigma_{q^b}^2\left(-0.43\mu_{R^{ab}}^2 + 0.5\mu_{R^{ab}}\mu_{X^{ab}} + 0.43\mu_{X^{ab}}^2\right) +\\
&\frac{\rho_{q^c}}{|V_A^c|^2}\sigma_{q^c}^2\left(0.43\mu_{R^{ac}}^2 + 0.5\mu_{R^{ac}}\mu_{X^{ac}} - 0.43\mu_{X^{ac}}^2\right) +\\
&\frac{\rho_{p^a q^a}}{|V_A|^2}\sigma_{p^a}\sigma_{q^a}\left(-\mu_{R^{aa}}^2 + \mu_{X^{aa}}^2\right) +\\
&\frac{\rho_{p^b q^b}}{|V_A^b|^2}\sigma_{p^b}\sigma_{q^b}\left(0.5\mu_{R^{ab}}^2 + \sqrt{3}\,\mu_{R^{ab}}\mu_{X^{ab}} - 0.5\mu_{X^{ab}}^2\right) +\\
&\frac{\rho_{p^c q^c}}{|V_c^a|^2}\sigma_{p^c}\sigma_{q^c}\left(0.5\mu_{R^{ac}}^2 - \sqrt{3}\,\mu_{R^{ac}}\mu_{X^{ac}} - 0.5\mu_{X^{ac}}^2\right)
\end{aligned} \tag{14}$$

where, $\rho_{p^h}$ and $\rho_{q^h}$ denote the correlation coefficients of active power and reactive power change among the same phase of different actor nodes with $h$ representing the corresponding phase term ($h = \{a,b,c\}$), respectively. $\rho_{p^h q^h}$ denotes the correlation coefficient between the active and reactive power within the same phase. Similarly, $\sigma_{p^h}^2$ and $\sigma_{q^h}^2$ depict the variance of active power and reactive power change, respectively. For random impedance part, $\mu_{R^k}$ and $\mu_{X^k}$ denote the mean of shared path resistance and reactance between all the nodes and a certain observation node, respectively, with $k$ representing the corresponding self/mutual

impedance terms ($k = aa,ab,ac,ba,bb,bc,ca,cb,cc$). It is important to note that all the defined parameters with respect to power change are user defined and usually set based on historical data, whereas, the parameters corresponding to shared path impedance are computed based on the network specifications.

**Step 3:** *Compute covariance between* $\Delta V_{OA1}^{a,(r,i)}$ *and* $\Delta V_{OA2}^{a,(r,i)}$:

The covariance between the real component of complex voltage change caused by two different PVs located at actor nodes $A1$ and $A2$ can be calculated as:

$$\begin{aligned} Cov(\Delta V_{OA1}^{a,r}, \Delta V_{OA2}^{a,r}) &= E(\Delta V_{OA1}^{a,r}\Delta V_{OA2}^{a,r}) - E(\Delta V_{OA1}^{a,r})E(\Delta V_{OA2}^{a,r}) \\ &= E\left[\boldsymbol{Z}_{A1}^{(a,r)^T}\Delta \boldsymbol{S}_{A1}\boldsymbol{Z}_{A2}^{(a,r)^T}\Delta \boldsymbol{S}_{A2}\right] \end{aligned} \tag{15}$$

corresponding covariance for the imaginary part of voltage change.

**Step 4:** *Compute mean and variance of* $\Delta V_{OA}^{a,r}$ *and* $\Delta V_{OA}^{a,i}$ *due to randomly distributed multiple actor nodes*:

The mean value of real and imaginary parts of voltage change due to randomly distributed multiple actor nodes are:

$$\begin{aligned} E\left[\Delta V_O^{a,r}\right] &= \mu_r = E\sum_{A=1}^{N}\Delta V_{OA}^{a,r} = N\mu_{Z_o^r}\mu_{\Delta S} \\ E\left[\Delta V_O^{a,i}\right] &= \mu_i = E\sum_{A=1}^{N}\Delta V_{OA}^{a,i} = N\mu_{Z_o^i}\mu_{\Delta S} \end{aligned} \tag{18}$$

Further, the variance of real and imaginary parts of the net voltage change can be expressed as,

$$\begin{aligned} \mathrm{Var}\left[\Delta V_O^{a,r}\right] &= \sigma_r^2 = Var\sum_{A=1}^{N}\left(Z_A^{(a,r)^T}\Delta S\right) = NVar\left(\boldsymbol{Z}^{(a,r)^T}\Delta \boldsymbol{S}\right) + 2\sum_{I<J}Cov\left(\Delta V_{OI}^{a,r}, \Delta V_{OJ}^{a,r}\right) \\ \mathrm{Var}\left[\Delta V_O^{a,i}\right] &= \sigma_i^2 = Var\sum_{A=1}^{N}\left(Z_A^{(a,i)^T}\Delta S\right) = NVar\left(\boldsymbol{Z}^{(a,i)^T}\Delta \boldsymbol{S}\right) + 2\sum_{I<J}Cov\left(\Delta V_{OI}^{a,i}, \Delta V_{OJ}^{a,i}\right) \end{aligned} \tag{19}$$

Using Eq. (4), $\boldsymbol{Z}^{(a,r)^T}\Delta \boldsymbol{S}$ can be expanded for both the actor nodes in the following way,

$$\begin{aligned} &\left[\left[\frac{\Delta P_1^a}{|V_1^a|}\left(-R_{O1}^{aa}\right) + \frac{\Delta P_1^b}{|V_1^b|}\left(\frac{R_{O1}^{ab}}{2} - \frac{\sqrt{3}X_{O1}^{ab}}{2}\right) + \frac{\Delta P_1^c}{|V_1^c|}\left(\frac{R_{O1}^{ac}}{2} + \frac{\sqrt{3}X_{O1}^{ac}}{2}\right)\right.\right. \\ &\left. - \frac{\Delta Q_1^a}{|V_1^a|}X_{O1}^{aa} + \frac{\Delta Q_1^b}{|V_1^b|}\left(\frac{\sqrt{3}R_{O1}^{ab}}{2} + \frac{X_{O1}^{ab}}{2}\right) + \frac{\Delta Q_1^c}{|V_1^c|}\left(-\frac{\sqrt{3}R_{O1}^{ac}}{2} + \frac{X_{O1}^{ac}}{2}\right)\right] \times \\ &\left[\frac{\Delta P_2^a}{|V_2^a|}\left(-R_{O2}^{aa}\right) + \frac{\Delta P_2^b}{|V_2^b|}\left(\frac{R_{O2}^{ab}}{2} - \frac{\sqrt{3}X_{O2}^{ab}}{2}\right) + \frac{\Delta P_2^c}{|V_2^c|}\left(\frac{R_{O2}^{ac}}{2} + \frac{\sqrt{3}X_{O2}^{ac}}{2}\right)\right. \\ &\left.\left. - \frac{\Delta Q_2^a}{|V_2^a|}X_{O2}^{aa} + \frac{\Delta Q_2^b}{|V_2^b|}\left(\frac{\sqrt{3}R_{O2}^{ab}}{2} + \frac{X_{O2}^{ab}}{2}\right) + \frac{\Delta Q_2^c}{|V_2^c|}\left(-\frac{\sqrt{3}R_{O2}^{ac}}{2} + \frac{X_{O2}^{ac}}{2}\right)\right]\right] \end{aligned} \tag{16}$$

For simplicity, actor nodes $A1$ and $A2$ are denoted by subscript 1 and 2, respectively. Like (14), the terms inside the expectation operator is cross multiplied to express covariance,

$$\begin{aligned} &\frac{\rho_{p^a}}{|V_A^a|^2}\sigma_{p^a}^2\mu_{R^{aa}}^2 + \frac{\rho_{q^a}}{|V_A^a|^2}\sigma_{q^a}^2\mu_{X^{aa}}^2 + \\ &\frac{\rho_{p^b}}{|V_A^b|^2}\sigma_{p^b}^2\left(0.25\mu_{R^{ab}}^2 - 0.86\mu_{R^{ab}}\mu_{X^{ab}} + 0.75\mu_{X^{ab}}^2\right) + \\ &\frac{\rho_{p^c}}{|V_A^c|^2}\sigma_{p^c}^2\left(0.25\mu_{R^{ac}}^2 + 0.86\mu_{R^{ac}}\mu_{X^{ac}} + 0.75\mu_{X^{ac}}^2\right) + \\ &\frac{\rho_{q^b}}{|V_A^b|^2}\sigma_{q^b}^2\left(0.75\mu_{R^{ab}}^2 + 0.86\mu_{R^{ab}}\mu_{X^{ab}} + 0.25\mu_{X^{ab}}^2\right) + \\ &\frac{\rho_{q^c}}{|V_A^c|^2}\sigma_{q^c}^2\left(0.75\mu_{R^{ac}}^2 - 0.86\mu_{R^{ac}}\mu_{X^{ac}} + 0.25\mu_{X^{ac}}^2\right) - \\ &\frac{\rho_{p^aq^a}}{|V_A^a|}\sigma_{p^a}\sigma_{q^a}\left(2\mu_{R^{aa}}\mu_{X^{aa}}\right) - \\ &\frac{\rho_{p^bq^b}}{|V_A^b|^2}\sigma_{p^b}\sigma_{q^b}\left(0.86\mu_{R^{ab}}^2 - \mu_{R^{ab}}\mu_{X^{ab}} - 0.86\mu_{X^{ab}}^2\right) + \\ &\frac{\rho_{p^cq^c}}{|V_c^c|^2}\sigma_{p^c}\sigma_{q^c}\left(-0.86\mu_{R^{ac}}^2 - \mu_{R^{ac}}\mu_{X^{ac}} + 0.86\mu_{X^{ac}}^2\right) \end{aligned} \tag{17}$$

The correlation coefficients and variances are same as defined in Eq. (14). Now, following the same steps from (15)–(17), yields

Now, by invoking Lindeberg-Feller central limit theorem, it can be shown that the real and imaginary parts of voltage change follow non zero mean Gaussian distribution with mean and variance as stated in Eqs. (18) and (19), respectively. As the square of non zero mean Gaussian variable follows non-central chi-square distribution [28], the distribution of the squared magnitude of $\Delta V_O^a$ is the sum of dependent non-central chi-square variables.

$$|\Delta V_O^a|^2 \sim \sigma_r^2\chi_1^2\left(\mu_r^2\right) + \sigma_i^2\chi_1^2\left(\mu_i^2\right) \tag{20}$$

where $\sigma^2$ and $\mu^2$ are the weight and non centrality parameters of non central chi square distribution with one degree of freedom corresponding to both real and imaginary parts of the voltage change. The sum of weighted non-central chi-square distributions can then be approximated with a scaled non-central chi-square with weight $\lambda$, non-centrality parameter $w$, and $v$ degrees of freedom as shown below [28]:

$$|\Delta V_O^a|^2 \sim \lambda\chi_v^2\left(w\right), \tag{21}$$

$$\begin{aligned} \lambda &= \frac{\sigma_r^4\left(1+2\mu_r^2\right) + \sigma_i^4\left(1+2\mu_i^2\right)}{\sigma_r^2\left(1+2\mu_r^2\right) + \sigma_i^2\left(1+2\mu_i^2\right)} \\ w &= \frac{\left(\sigma_r^2\mu_r^2 + \sigma_i^2\mu_i^2\right)\left(\sigma_r^2 + \sigma_i^2 + 2\sigma_r^2\mu_r^2 + 2\sigma_i^2\mu_i^2\right)}{\sigma_r^4 + \sigma_i^4 + 2\sigma_r^4\mu_r^2 + 2\sigma_r^4\mu_i^2} \\ v &= \frac{\left(\sigma_r^2 + \sigma_i^2\right)\left(\sigma_r^2 + \sigma_i^2 + 2\sigma_r^2\mu_r^2 + 2\sigma_i^2\mu_i^2\right)}{\sigma_r^2 + \sigma_i^2 + 2\left(\sigma_r^4\mu_r^2\right) + 2\left(\sigma_i^4\mu_i^2\right)} \end{aligned} \tag{22}$$

Since the square root of a non-central chi-square random variables follows a Rician distribution [28], the magnitude of voltage change will follow a Rician distribution:

$$|\Delta V_O^a| \sim Rician(k, \sigma) \tag{23}$$

where $k = \sqrt{w}$ and $\sigma = \sqrt{\lambda}$. The magnitudes of voltage changes for other phases follow a similar expression with the respective phase values. If the power variation is assumed to follow a zero-mean Gaussian distribution, which is a typical assumption used in many prior works [26,27], $\mu_{\Delta S}$ vanishes from the mean (Eqs. (7)) and variance (Eqs. (11) and (12)) equations of voltage change. This eventually leads to zero value for $\mu_r$

and $\mu_i$. Again, by invoking Lindeberg-Feller central limit theorem, one can show that the real and imaginary parts of the voltage change follow zero-mean normal distributions as,

$$\Delta V_O^{a,r} \overset{D}{\sim} \mathcal{N}\left(0, \sigma_r^2\right), \quad \Delta V_O^{a,i} \overset{D}{\sim} \mathcal{N}\left(0, \sigma_i^2\right) \tag{24}$$

The square of the magnitude of voltage change follows a gamma distribution [29], and subsequently, the magnitude of voltage change follows a Nakagami distribution [30],

$$|\Delta V_O^a| \sim Nakagami(m, \omega), \tag{25}$$

where parameter $\theta = 2(\sigma_r^4 + \sigma_i^4 + 2c^2)/(\sigma_r^2 + \sigma_i^2)$, shape parameter $m = (\sigma_r^2 + \sigma_i^2)/\theta$, scale parameter $\omega = \sqrt{m\theta}$, and $c$ being the covariance between the real and imaginary parts of voltage change. In the next sections, the proposed ST-PVSA method is first validated using simulations, and then it is employed to estimate PV HC in a efficient manner.

## 4. Validation of ST-PVSA

The proposed probability distribution of the voltage change is validated on the modified IEEE 37-node test system. The nominal voltage of the test system is 4.8 kV. The actual distribution of the magnitude of the voltage change is obtained using Newton–Raphson based sensitivity analysis method, and the theoretical distribution is obtained using the proposed method of ST-PVSA. A scenario is considered for simulation where 9 PV units are located at random locations in the distribution system. The power at the actor nodes, i.e., the nodes injected with PVs, varies randomly due to fluctuations in PV generation. For illustration, 9 actor nodes are chosen where change in PV generation at a particular time instant is modeled as a zero mean Gaussian random variable. However, ST-PVSA is valid for any number of actor nodes with any arbitrary distribution of power variation. Typically, unbalance in the distribution system is caused by single phase loads. Therefore, unbalance in our experiments is achieved by employing single-phase and two-phase loads in the standard three-phase test networks. The base loads are the same as provided in the distribution system analysis subcommittee report [31]. Unbalance can also be induced by unequal power change across different phases of the system. However, the magnitude of power change needs to be strong enough which also depend on the base loads. The covariance matrix $\sum_{\Delta S}$ captures the spatial correlation of PV generation, which exists because of geographical proximity as PVs in the same region typically exhibit same generation profile. The diagonal elements of the covariance matrix contain variances that depend on the size of PV units and the off-diagonal elements capture the effect of geographical proximity of these PV units. In our simulation, the variance of change in real power ($\Delta P$) is set to 5 kW and the variance of change in reactive power ($\Delta Q$) is set to 0.5 kVar. The values of the correlation coefficients $\rho_{p^h}, \rho_{q^h}$ and $\rho_{p^h q^h}$ are set to 0.2, 0.2, and -0.5, respectively for all the phases. Variance can be set to zero for nodes with no PVs. Now, for random impedance part, the mean and variance of resistance and reactance between random actor node and observation node 9 is calculated from data of the IEEE 37-node test system. The value of correlation coefficient between resistance and reactance is 0.99. Fig. 2 compares the actual distribution of the magnitude of voltage change with the proposed ST-PVSA case. The actual distribution of $\Delta V_9$ is obtained by randomly varying powers of all actor nodes at phase-a and subsequently, voltage change at node 9 is computed by using Newton–Raphson based method. Further, Monte-Carlo simulations (MCS) are incorporated to capture the uncertainties associated with the power changes. Here, voltage changes are computed for one million MCS. The scaled histogram of $|\Delta V_9|$ is depicted through the orange curve in the Fig. 2. The theoretical distribution computed with Eqs. (18) and (19) is shown by blue curve in Fig. 2. It can be observed that the probability distribution computed using the proposed method is very close to the actual simulated distribution with 0.18 as Jensen-Shannon distance. Further, the execution time of our method to calculate the voltage change distribution in both the 37-node and 123-node networks are within 1 min, whereas the time exceeds 120 min in the classical load flow based method. Thus, ST-PVSA is order 2 faster compared to the conventional approach. This experiment demonstrates the effectiveness of the proposed ST-PVSA approach.

## 5. ST-PVSA for PV hosting capacity

This section presents the methodology for computing HC with the proposed ST-PVSA approach. As ST-PVSA provides the probability distribution of voltage change at a node due to random power changes at random locations of the network, it suffices to identify voltage violations for different PV penetration levels. The procedure to determine HC begins with fixing the number of penetration levels say from 1% to 100% level at an increment of 1%. Then, the number of PV units ($N_k$) that need to be integrated for each penetration level is computed using Eqn. (26). $N_k$ is determined statistically based on the distribution of real PV sizes. The size of actual PV installations in the state of California, USA is collected from the California dataset [32]. Fig. 3 depicts the scaled histogram of PV sizes which approximately follows a gamma distribution. The penetration level is divided into various bands based on the percentage of total demand. For instance, $k$ varies from 1 to 5 for 5 bands, i.e., $(0-20\%), (21-40\%) \ldots (81-100\%)$. A unique $N_k$ is defined for each band such that the same number of PV units is used for all penetration levels in that particular band. In each band $k$, the power injection increases with the increasing penetration levels at $N_k$ random locations. This is logical in a sense that it is not necessary to increase the number of PV units for simulating increasing penetration level rather it can be achieved by increasing the power injection in the existing PVs.

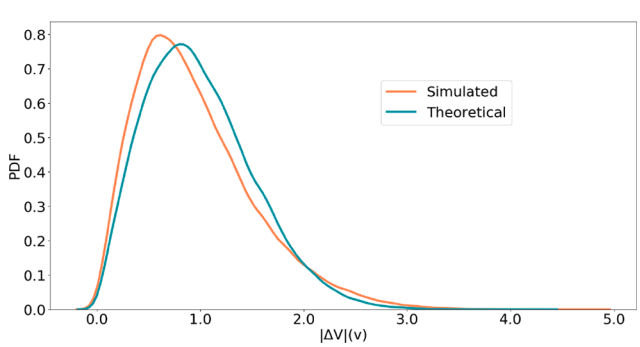


**Fig. 2.** Distribution of voltage change at node 9.

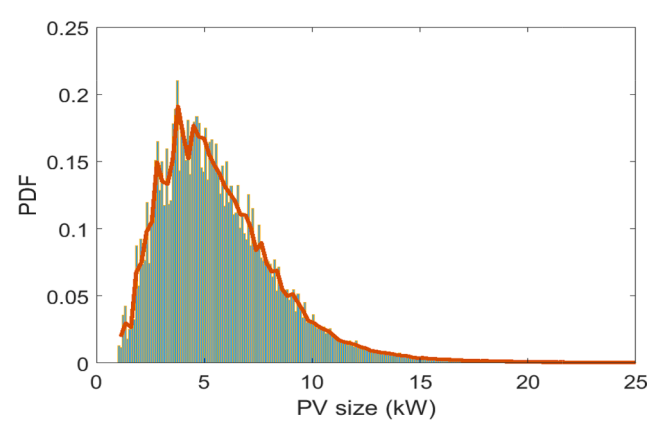


**Fig. 3.** Distribution of PV sizes from California dataset.

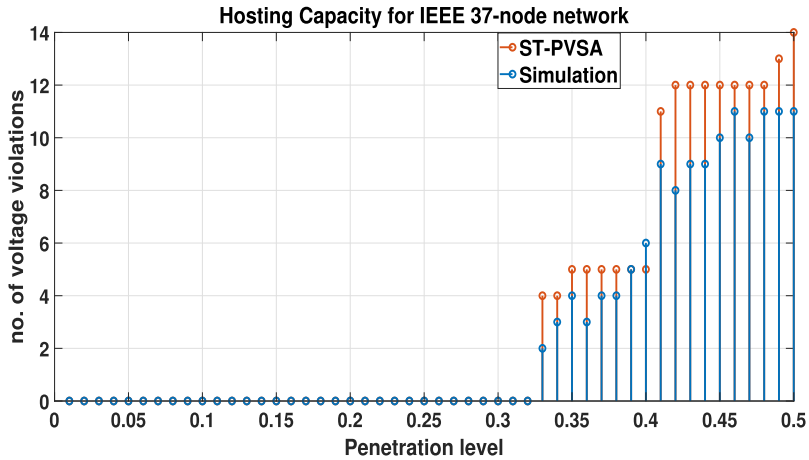


**Fig. 4.** Variation of violations count with Penetration levels.

However, the power injections cannot be increased beyond a certain limit due to the restriction of PV size. Therefore, $N_k$ increases as we move to the higher penetration band. $N_k$ for a particular penetration band $k$ is computed as following:

$$N_k = \frac{\text{Mean penetration level for band k}}{\text{Max PV size}} \tag{26}$$

where "Max PV size" comes from the PV size distribution and the mean penetration level is the average power injection for band $k$. In the third step, $N_k$ is used to obtain the mean ($\mu^l_{\Delta S}$) of power change vector $\Delta S$ (Eqn. 4) for each penetration level $l$, such that $\mu^l_{\Delta S} N^l \approx P_l$. Here, $P_l$ and $N^l$ are the net power injection and PV units for penetration level $l$, respectively. The complex voltage change due to power injection is added to the base voltage to get the future voltage. Following the same arguments as mentioned in Theorem 2 of [33], the distribution of future voltage is shown to follow Rician with the parameters as defined in Eq. (22). The mean of real ($\mu_r$) and imaginary ($\mu_i$) parts of voltage change (18) are added to the corresponding parts of base voltage to get the mean value of the future voltage. The mean future voltage is then plugged in the derived Eqn. (23) to find the distribution of future voltage at all nodes of the network. Nodes that have a probability of voltage violation greater than the threshold are classified as highly vulnerable nodes, and violations are reported. For illustration, a violation is recorded when the probability of voltage violations is more than 0.5. 0.5 is unbiased and gives equal preference to both detection and non-detection of violations. The complete process is repeated for increasing penetration levels until the algorithm encounters the first violation. The corresponding penetration level is the HC of the system. "First voltage violation" refers to the situation when we observe voltage violations in the system for the very first time while increasing the PV penetration level. In this paper, the minimum penetration level for which the violation is observed for the first time is considered as the hosting capacity. Algorithm 1 provides the pseudo-code of ST-PVSA approach to compute HC.

**Algorithm 1.** Proposed ST-PVSA method to compute Hosting capacity

1: Fix number of penetration levels $(1, 2, \ldots 100\%)$
2: Calculate number of PVs for a particular penetration level using Eq. (26).
3: Compute mean and variance of power change vector corresponding to a particular penetration level.
4: Use ST-PVSA to compute node voltages and track total number of voltage violations.
5: Repeat steps 2 to 4 for different penetration levels.
6: The penetration level that causes first voltage violations is the hosting capacity.

To evaluate the performance of ST-PVSA in determining the HC, load flow based HC is used as a benchmark. Similar to the ST-PVSA approach, the PV penetration level is fixed from 1% to 100% level at 1% increment. For each penetration level, Monte Carlo simulations are repeated $10k$ times thereby creating one million different PV deployment scenarios. For illustration purposes, the loads on the test network are chosen as reported in the IEEE PES distribution system analysis subcommittee report [31]. However, the proposed method is generic enough to accommodate other loading scenarios such as daytime (10 am-2 pm) maximum load and daytime minimum. Finally, for each penetration level, $N_k$ locations are selected randomly to allocate PV units and load flow is executed to track the voltage violations. For IEEE 37-node network, the number of PV units for each of the five penetration level bands are $5, 10, 20, 25$ and 30. The power is increased from 10 kW (1% penetration level) to 1100 kW (100% level) in steps of 11 kW. Fig. 4 depicts the variation of violations count with increasing penetration levels. It can be observed that the proposed STPVSA approach is 100% accurate in estimating HC of IEEE 37-node test network. In other words, the penetration level predicted by ST-PVSA aligns well with those computed from the load flow based approach (i.e., lies within the range of load flow based HC values). Further, to demonstrate the scalability of the proposed method, the HC analysis is also validated on the IEEE 123-node network.

Table 1 presents the HC values computed with the proposed ST-PVSA based approach and existing load flow-based method for a various number of scenarios. It is worth noting that for each PV penetration level, ST-PVSA needs to be run once (independent of scenarios), whereas multiple simulations are required for convergence in the load flow-based

**Table 1**
Hosting capacity with Load flow and ST-PVSA.

| Test Network | Scenarios* | 1 k (%) | 10 k (%) | 30 k (%) | 40 k % | 50 k % | 60 k% |
|---|---|---|---|---|---|---|---|
| IEEE 37 | Load Flow | 33 | 32 | 31 | 31 | 31 | 31 |
| | ST-PVSA | | | 33 | | | |
| IEEE 123 | Load Flow | 44 | 43 | 42 | 41 | 40 | 40 |
| | ST-PVSA | | | 39 | | | |

**Table 2**
Execution time with Load flow and ST-PVSA.

| Test Network | Scenarios* | 1 k (min) | 10 k (min) | 30 k (min) |
|---|---|---|---|---|
| IEEE 37 | LF | 1.93 | 18.8 | 55.7 |
| | ST-PVSA | | 1.13 | |
| IEEE 123 | LF | 43.93 | 400.2 | 1195.6 |
| | ST-PVSA | | 3.91 | |

approach. For the IEEE-37 node network, ST-PVSA yields a HC of 33% which lies in the range of values computed with load flow based approach. Similarly, for the IEEE-123 test network, the ST-PVSA based HC value is 39%, which again intersects with that of load flow's approach. Furthermore, the proposed approach is also evaluated for a balanced load case in the IEEE 37-node test network. The estimated value of HC turns out to be 41%, whereas 42% is obtained with 30 k simulations in the conventional approach. This demonstrates the generalizability of our method. Additionally, HC seems to decrease for an unbalanced case compared to a balanced one although all the factors (power change and network parameters) remain unchanged. This is because of non-uniformity in voltages across the buses which increases the probability of extreme voltages leading to violations in a relatively earlier stage compared to a balanced load scenario.

Along with the high estimation accuracy, ST-PVSA offers a significant advantage in terms of computational complexity. Table 2 represents the execution time of scenarios simulated in 1 for the two test networks. All experiments are conducted in a machine with an Intel-i7 processor and 16 Gb RAM. It can be inferred from Table 2 that in IEEE 123-node test network, the ST-PVSA is three orders faster than the load flow-based approach, and the gap will further increase as the network size grows.

The above-discussed experiments demonstrate the efficacy of ST-PVSA for a typical snapshot type HC. Further, it will be more effective for a dynamic HC, which is relatively a new way of analyzing HC of distribution systems. Dynamic HC is not based on worst-case snapshot power flows. It requires probabilistic screens that consider the uncertainty around the time-series input variables, like hourly PV productions and building loads. Power flow analysis is conducted on large time-series data of load and PV on an hourly basis. For a real distribution model with thousands of nodes and one-second resolution data, simulations could take a few days [5]. Furthermore, the PV and load uncertainties have significant influences on hosting capacity values. Under this type of dynamic analysis, the proposed approach could work very efficiently by accurately capturing voltage violations in an acceptable amount of time. The performance of ST-PVSA in dynamic HC will be investigated as part of our future work.

## 6. Conclusion

This work presents an analytical approach to compute the probability distribution of voltage change at a particular node as a function of random change in power at random locations of the network due to distributed PV units. The proposed approach is validated with a conventional load flow based approach in two different test networks namely IEEE 37 and IEEE 123. The estimated probability distribution matches with the baseline to a high degree of accuracy (as demonstrated with a low Jensen-Shannon distance of 0.18). The computational complexity is also reduced by an order of 3 compared to the conventional approach. Our framework can be applied to analyze stochastic operations of a power distribution system. One of the use cases is shown by employing the proposed method to determine the hosting capacity of the system without investigating multiple scenarios. Our method is fairly accurate in identifying the hosting capacity and offers huge advantage in terms of computational efficiency. In the IEEE 123-node test network, proposed method is an order two faster compared to conventional load flow based approach and this gap will further increase as the network size grows. As part of future work, we plan to extend our framework for dynamic HC involving continuous time series data of load and PV with the system allowing violations of small duration in accordance with the ANSI standards.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgement

This material is based upon work partly supported by the Department of Energy, Office of Energy Efficiency and Renewable Energy (EERE), Solar Energy Technologies Office, under Award # DE-EE0008767 and National science foundation under award # 1855216.

## Appendix A

The change in complex voltage at any phase (say phase $a$) of observation node $O$ due to change in complex power at any phase of a single actor node $A$ is given as [14],

$$\Delta V_{OA}^{a} \approx -\left[\frac{\Delta S_A^{a*} Z_{OA}^{aa}}{V_A^{a*}} + \frac{\Delta S_A^{b*} Z_{OA}^{ab}}{V_A^{b*}} + \frac{\Delta S_A^{c*} Z_{OA}^{ac}}{V_A^{c*}}\right], \tag{27}$$

On expanding the complex power and shared path impedance terms, we get the following equation,

$$\begin{aligned} \Delta V_{OA}^{a} &\approx -\left[\frac{\left(\Delta P_A^a - j\Delta Q_A^a\right)\left(R_{OA}^{aa} + jX_{OA}^{aa}\right)}{V_A^{a\star}} + \ldots\right], \\ &\approx -\left[\frac{\left(\Delta P_A^a R_{OA}^{aa} + Q_A^a X_{OA}^{aa}\right) + j\left(\Delta P_A^a X_{OA}^{aa} - R_{OA}^{aa}\Delta Q_A^a\right)}{\left(V_A^{a,r} - jV_A^{a,i}\right) + \left(\Delta V_A^{a,r} - \Delta V_A^{a,i}\right)} + \ldots\right], \end{aligned} \tag{28}$$

On normalizing the numerator and denominator, (28) reduces to

$$\Delta V_{OA}^{a} \approx -\left[\frac{\left(\Delta P_A^a R_{OA}^{aa} + \Delta Q_A^a X_{OA}^{aa}\right) + j\left(\Delta P_A^a X_{OA}^{aa} - R_{OA}^{aa}\Delta Q_A^a\right)}{\left(V_A^{a,r} + \Delta V_A^{a,r}\right) - j\left(V_A^{a,i} + \Delta V_A^{a,i}\right)} * \frac{\left(V_A^{a,r} + \Delta V_A^{a,r}\right) + j\left(V_A^{a,i} + \Delta V_A^{a,i}\right)}{\left(V_A^{a,r} + \Delta V_A^{a,r}\right) + j\left(V_A^{a,i} + \Delta V_A^{a,i}\right)} + \ldots\right], \tag{29}$$

Normalization segregates (29) into real and imaginary parts as shown below,

$$\begin{aligned}\Delta V_{OA}^{a} \approx -\Bigg[&\frac{\left(\Delta P_A^a R_{OA}^{aa} + \Delta Q_A^a X_{OA}^{aa}\right)\left(V_A^{a,r} + \Delta V_A^{a,r}\right) - \left(\Delta P_A^a X_{OA}^{aa} - R_{OA}^{aa}\Delta Q_A^a\right)\left(V_A^{a,i} + \Delta V_A^{a,i}\right)}{\left(V_A^{a,r} + \Delta V_A^{a,r}\right)^2 + \left(V_A^{a,i} + \Delta V_A^{a,i}\right)^2} - \ldots \\ &+ j\frac{\left(\Delta P_A^a R_{OA}^{aa} + \Delta Q_A^a X_{OA}^{aa}\right)\left(V_A^{a,i} + \Delta V_A^{a,i}\right) + \left(\Delta P_A^a X_{OA}^{aa} - R_{OA}^{aa}\Delta Q_A^a\right)\left(V_A^{a,r} + \Delta V_A^{a,r}\right)}{\left(V_A^{a,r} + \Delta V_A^{a,r}\right)^2 + \left(V_A^{a,i} + \Delta V_A^{a,i}\right)^2} - \ldots\Bigg].\end{aligned} \tag{30}$$

The real part of the voltage change can be expressed as,

$$\Delta V_{OA}^{a,r} \approx -\left[\frac{\left(\Delta P_A^a R_{OA}^{aa} + \Delta Q_A^a X_{OA}^{aa}\right)\left(V_A^{a,r}\right)\left(1 + \frac{\Delta V_A^{a,r}}{V_A^{a,r}}\right)}{(V_A^{a,r})^2\left(1 + \frac{\Delta V_A^{a,r}}{V_A^{a,r}}\right)^2 + (V_A^{a,i})^2\left(1 + \frac{\Delta V_A^{a,i}}{V_A^{a,i}}\right)^2} - \frac{\left(\Delta P_A^a X_{OA}^{aa} - \Delta Q_A^a R_{OA}^{aa}\right)\left(V_A^{a,i}\right)\left(1 + \frac{\Delta V_A^{a,i}}{V_A^{a,i}}\right)}{(V_A^{a,r})^2\left(1 + \frac{\Delta V_A^{a,r}}{V_A^{a,r}}\right)^2 + (V_A^{a,i})^2\left(1 + \frac{\Delta V_A^{a,i}}{V_A^{a,i}}\right)^2} - \ldots\right], \tag{31}$$

Using the same assumptions as in [14], (31) can further be simplified as,

$$\Delta V_{OA}^{a,r} \approx -\left[\frac{\left(\Delta P_A^a R_{OA}^{aa} + \Delta Q_A^a X_{OA}^{aa}\right)\left(V_A^{a,r}\right)}{\left(V_A^{a,r}\right)^2 + \left(V_A^{a,i}\right)^2} - \frac{\left(\Delta P_A^a X_{OA}^{aa} - \Delta Q_A^a R_{OA}^{aa}\right)\left(V_A^{a,i}\right)}{\left(V_A^{a,r}\right)^2 + \left(V_A^{a,i}\right)^2} - \ldots\right]. \tag{32}$$

Similarly, the imaginary part of the voltage change can be written as,

$$\Delta V_{OA}^{a,i} \approx -\left[\frac{\left(\Delta P_A^a X_{OA}^{aa} - \Delta Q_A^a R_{OA}^{aa}\right)\left(V_A^{a,r}\right)}{\left(V_A^{a,r}\right)^2 + \left(V_A^{a,i}\right)^2} + \frac{\left(\Delta P_A^a R_{OA}^{aa} + \Delta Q_A^a X_{OA}^{aa}\right)\left(V_A^{a,i}\right)}{\left(V_A^{a,r}\right)^2 + \left(V_A^{a,i}\right)^2} - \ldots\right], \tag{33}$$

Eqs. (32) and (33) are rearranged by taking the common factor with power terms as shown below,

$$\begin{aligned}\Delta V_{OA}^{a,r} &\approx -\left[\frac{\Delta P_A^a\left(R_{OA}^{aa} V_A^{a,r} - X_{OA}^{aa} V_A^{a,i}\right)}{\left(V_A^{a,r}\right)^2 + \left(V_A^{a,i}\right)^2} + \frac{\Delta Q_A^a\left(X_{OA}^{aa} V_A^{a,r} + R_{OA}^{aa} V_A^{a,i}\right)}{\left(V_A^{a,r}\right)^2 + \left(V_A^{a,i}\right)^2} - \ldots\right] \\ \Delta V_{OA}^{a,i} &\approx -\left[\frac{\Delta P_A^a\left(R_{OA}^{aa} V_A^{a,i} + X_{OA}^{aa} V_A^{a,r}\right)}{\left(V_A^{a,r}\right)^2 + \left(V_A^{a,i}\right)^2} + \frac{\Delta Q_A^a\left(X_{OA}^{aa} V_A^{a,i} - R_{OA}^{aa} V_A^{a,r}\right)}{\left(V_A^{a,r}\right)^2 + \left(V_A^{a,i}\right)^2} - \ldots\right]\end{aligned} \tag{34}$$

Finally, the real and imaginary part of base voltages are expressed in polar magnitude form (i.e., $V_A^{a,r} = \left|V_A^a\right| cos(\omega_A)$ $V_A^{a,i} = \left|V_A^a\right| sin(\omega_A)$) which reduces (34) to,

$$\begin{aligned}\Delta V_{OA}^{a,r} &\approx -\left[\frac{\Delta P_A^a\left(R_{OA}^{aa} cos(\omega_A) - X_{OA}^{aa} sin(\omega_A)\right)}{\left|V_A^a\right|} + \frac{\Delta Q_A^a\left(X_{OA}^{aa} cos(\omega_A) + R_{OA}^{aa} sin(\omega_A)\right)}{\left|V_A^a\right|} - \ldots\right] \\ \Delta V_{OA}^{a,i} &\approx -\left[\frac{\Delta P_A^a\left(R_{OA}^{aa} sin(\omega_A) + X_{OA}^{aa} cos(\omega_A)\right)}{\left|V_A^a\right|} + \frac{\Delta Q_A^a\left(X_{OA}^{aa} sin(\omega_A) - R_{OA}^{aa} cos(\omega_A)\right)}{\left|V_A^a\right|} - \ldots\right].\end{aligned} \tag{35}$$

## References


[1] IRENA, "Global energy transformation: A roadmap to 2050," 2018.

[2] Alyami S, Wang Y, Wang C, Zhao J, Zhao B. Adaptive real power capping method for fair overvoltage regulation of distribution networks with high penetration of pv systems. IEEE Transactions on Smart Grid 2014;5(6):2729–38.

[3] Olivier F, Aristidou P, Ernst D, Van Cutsem T. Active management of low-voltage networks for mitigating overvoltages due to photovoltaic units. IEEE Transactions on Smart Grid 2015;7(2):926–36.

[4] Ding F, Mather B. On distributed pv hosting capacity estimation, sensitivity study, and improvement. IEEE Transactions on Sustainable Energy 2016;8(3):1010–20.

[5] A.K. Jain, K. Horowitz, F. Ding, N. Gensollen, B. Mather, and B. Palmintier, "Quasi-static time-series pv hosting capacity methodology and metrics," in 2019 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT). IEEE, 2019, pp. 1–5.

[6] Tonkoski R, Turcotte D, El-Fouly TH. Impact of high pv penetration on voltage profiles in residential neighborhoods. IEEE Transactions on Sustainable Energy 2012;3(3):518–27.

[7] J. Smith and M. Rylander, "Stochastic analysis to determine feeder hosting capacity for distributed solar pv," Electric Power Research Inst., Palo Alto, CA, Tech. Rep, vol. 1026640, pp. 0885–8950, 2012.

[8] Dubey A, Santoso S. On estimation and sensitivity analysis of distribution circuit's photovoltaic hosting capacity. IEEE Trans. Power Syst. 2016;32(4):2779–89.

[9] M. Rylander, J. Smith, and W. Sunderman, "Streamlined method for determining distribution system hosting capacity," in 2015 IEEE Rural Electric Power Conference. IEEE, 2015, pp. 3–9.

[10] Brenna M, De Berardinis E, Foiadelli F, Sapienza G, Zaninelli D. Voltage control in smart grids: An approach based on sensitivity theory. Journal of Electromagnetic Analysis and Applications 2010;2010.

[11] V. Klonari, B.B. Zad, J. Lobry, and F. Vallée, "Application of voltage sensitivity analysis in a probabilistic context for characterizing low voltage network operation," in 2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS). IEEE, 2016, pp. 1–7.
[12] S. Munikoti, K. Jhala, K. Lai, and B. Natarajan, "Analytical voltage sensitivity analysis for unbalanced power distribution system," in 2020 IEEE Power & Energy Society General Meeting (PESGM). IEEE, 2020, pp. 1–5.
[13] K. Jhala, B. Natarajan, and A. Pahwa, "Probabilistic voltage sensitivity analysis (pvsa) for random spatial distribution of active consumers," in 2018 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT). IEEE, 2018, pp. 1–5.
[14] Munikoti S, Natarajan B, Jhala K, Lai K. Probabilistic voltage sensitivity analysis to quantify impact of high pv penetration on unbalanced distribution system. IEEE Trans. Power Syst. 2021:1.
[15] Mulenga E, Bollen MH, Etherden N. A review of hosting capacity quantification methods for photovoltaics in low-voltage distribution grids. International Journal of Electrical Power & Energy Systems 2020;115:105445.
[16] M. Zain ul Abideen, O. Ellabban, and L. Al-Fagih, "A review of the tools and methods for distribution networks' hosting capacity calculation," Energies, vol. 13, no. 11, p. 2758, 2020.
[17] Mulenga E, Bollen MH, Etherden N. Solar pv stochastic hosting capacity in distribution networks considering aleatory and epistemic uncertainties. International Journal of Electrical Power & Energy Systems 2021;130:106928.
[18] Ali A, Mahmoud K, Lehtonen M. Maximizing hosting capacity of uncertain photovoltaics by coordinated management of oltc, var sources and stochastic evs. International Journal of Electrical Power & Energy Systems 2021;127:106627.
[19] Takenobu Y, Yasuda N, Minato S-I, Hayashi Y. Scalable enumeration approach for maximizing hosting capacity of distributed generation. International Journal of Electrical Power & Energy Systems 2019;105:867–76.
[20] Xu Q, Kang C, Zhang N, Ding Y, Xia Q, Sun R, Xu J. A probabilistic method for determining grid-accommodable wind power capacity based on multiscenario system operation simulation. IEEE Transactions on Smart Grid 2014;7(1):400–9.
[21] Al-Saadi H, Zivanovic R, Al-Sarawi SF. Probabilistic hosting capacity for active distribution networks. IEEE Trans. Industr. Inf. 2017;13(5):2519–32.
[22] Mulenga E, Bollen MH, Etherden N. Distribution networks measured background voltage variations, probability distributions characterization and solar pv hosting capacity estimations. Electric Power Systems Research 2021;192:106979.
[23] Yan R, Saha TK. Voltage variation sensitivity analysis for unbalanced distribution networks due to photovoltaic power fluctuations. IEEE Trans. Power Syst. 2012;27 (2):1078–89.
[24] Kang S, Kim J, Park J-W, Baek S-M. Reactive power management based on voltage sensitivity analysis of distribution system with high penetration of renewable energies. Energies 2019;12(8):1493.
[25] B.B. Zad, J. Lobry, and F. Vallée, "A centralized approach for voltage control of mv distribution systems using dgs power control and a direct sensitivity analysis method," in 2016 IEEE International Energy Conference (ENERGYCON). IEEE, 2016, pp. 1–6.
[26] J. Vasilj, P. Sarajcev, and D. Jakus, "Pv power forecast error simulation model," in 2015 12th International Conference on the European Energy Market (EEM). IEEE, 2015, pp. 1–5.
[27] Jhala K, Krishnan V, Natarajan B, Zhang Y. Data-driven preemptive voltage monitoring and control using probabilistic voltage sensitivities. In: 2019 IEEE Power Energy Society General Meeting (PESGM); 2019. p. 1–5.
[28] Mathai AM, Provost SB. Quadratic forms in random variables: theory and applications. Dekker; 1992.
[29] Chuang L-L, Shih Y-S. Approximated distributions of the weighted sum of correlated chi-squared random variables. Journal of Statistical Planning and Inference 2012;142(2):457–72.
[30] M. Nakagami, "The m-distribution–a general formula of intensity distribution of rapid fading," in Statistical methods in radio wave propagation. Elsevier, 1960, pp. 3–36.
[31] Kersting WH. Radial distribution test feeders. IEEE Trans. Power Syst. 1991;6(3): 975–85.
[32] California solar statistics. [Online]. Available: https://californiasolarstatistics.org.
[33] M. Abujubbeh, S. Munikoti, and B. Natarajan, "Probabilistic voltage sensitivity based preemptive voltage monitoring in unbalanced distribution networks," in 2020 North American Power Symposium (NAPS). IEEE, 2020, pp. 1–6.