Low-Computational Voltage-Assessment Approach Considering Fine-Resolution Simulations for Distribution Systems With Photovoltaics

Karar Mahmoud, Mohamed Abdel-Nasser, and Matti Lehtonen

***Abstract*—The proliferation of photovoltaic (PV) has been increased in distribution systems worldwide. The intermittent PV generation can cause diverse operational problems in the grid, especially voltage deviations and violations. As a result, voltage assessment in distribution systems interconnected with PV, which has a heavy computational burden, is privileged to assist power utilities in decision-making. In this article, a low-computational and accurate voltage assessment approach with PV considering fine-resolution simulations (i.e., time-step of 1 s) is proposed. Specifically, the proposed approach can rapidly compute the voltage deviation in the whole distribution system and terminal voltages of PV units based on a data-driven model. This model is built using machine learning considering various scenarios of PV and load profiles. The proposed approach has the following features. 1) Its computational burden is very low compared to the widely used iterative-based methods. 2) It can handle the full data with the finest available resolution, yielding accurate voltage assessment. The proposed method has been applied for voltage assessment considering daily and annual simulations of different distribution systems interconnected with PV units. The simulation results manifest the high accuracy and computational speed of the proposed approach, especially for fine-resolution simulations.**

***Index Terms*—Data driven, distribution systems, low-computational, photovoltaic (PV), voltage assessment.**

I. Introduction

MODERN power systems incorporate decentralized small-scale distributed generations (DG) units attached to low/medium voltage distribution systems. To meet environmental and economic aspects, photovoltaic (PV) and wind DG units are widely installed to feed nearby commercial or residential loads with electricity [1]–[3]. Notably, power flow in distribution lines could be altered from unidirectional to bidirectional due to the injection of the intermittent power of renewable DG units [4], [5]. Power system planners and operators use computational tools to study the potential impacts of DG units that can occur at various times of the year, such as voltage violations and excessive power losses.

To analyze the impact of DG units on modern distribution systems, high time-resolution data should be used to accurately represent their intermittent output power as well as fluctuated loads. However, conventional snapshot tools are insufficient to study all possible impact of these DG units. Therefore, several quasi-static time series (QSTS) analysis methods have been proposed to study the time-dependent aspects of power flow, such as the interaction between load and PV output fluctuations and control devices [6], [7]. QSTS analysis aims to provide time-series analysis of distribution systems by conducting distribution power flow algorithms for all time instants within the studied period. In general, these QSTS analysis methods have a high computational burden, especially for annual or long-term simulations with fine-resolution data (i.e., time-step of 1 s). In [8], the challenges in reducing the computational burden of QSTS simulations were discussed, including the number of power flows to be computed, comprehensive simulation analysis with fine-resolution data, time dependence between time steps, and multiple valid power flow solutions. These challenges limit system operators to simulate the distribution system with low-resolutions data in order to reduce its computational burden on the expense of the accuracy, particularly when assessing the fluctuating voltage.

In the literature, several techniques have been proposed to reduce the computational burden of the voltage assessment in distribution systems with renewable generation. These techniques either reduce the number of power flows to be solved or shorten the computational time of power flow methods [9]. In [10], a generalized analytical technique to assess impacts of PV on radial distribution systems was presented. In [11], an index-based approach was proposed for evaluating the impact of PV with rich generation, taking into account the voltage rise phenomena as well as reverse power flow. In [12], a voltage assessment approach has been proposed for distribution systems with PV, where a variable-width sliding window is utilized. In [13], a voltage assessment method has been proposed based on regression and correction techniques with intermittent PV generation and fluctuated loads. A methodology for primary voltage assessment

Manuscript received May 10, 2020; revised July 24, 2020, September 11, 2020, December 12, 2020, and March 4, 2021; accepted May 7, 2021. Date of publication May 27, 2021; date of current version December 9, 2021. This work was supported by the Department of Electrical Engineering and Automation, Aalto University, Finland. *(Corresponding author: Karar Mahmoud.)*

Karar Mahmoud is with the Department of Electrical Engineering, and Automation, Aalto University, FI-00076 Espoo, Finland, and also with the Department of Electrical Engineering, Aswan University, Aswan 81542, Egypt (e-mail: karar.alnagar@aswu.edu.eg).

Mohamed Abdel-Nasser is with the Department of Computer Engineering and Mathematics, University Rovira i Virgili, 43007 Tarragona, Spain, and also with the Department of Electrical Engineering, Aswan University, Aswan 81542, Egypt (e-mail: mohamed.abdelnasser@aswu.edu.eg).

Matti Lehtonen is with the Department of Electrical Engineering and Automation, Aalto University, FI-00076 Espoo, Finland (e-mail: matti.lehtonen@aalto.fi).

Digital Object Identifier 10.1109/JSYST.2021.3079154

on distribution systems interconnected with high PV penetration has been proposed in [14] based on time-series simulations. The authors of [15] proposed a numerical approach for simulating PV impacts on distribution systems, utilizing complete generation and load models. The authors of [16] suggested to store and reassign power flow solutions during time-series simulations in distribution systems. In [17], a clustering technique was used to reduce the data resolution of load and DG scenarios, and calculated the power flow for the centroids of the clusters only. In [18], down-sampling and vector quantization methods were exploited to shorten the execution time of QSTS simulations. In [19], a variable step QSTS approach was proposed for fast analysis of PV impacts on distribution systems. Notably, most studies down sample, cluster, or quantize the fine-resolution data to reduce of the number of power flows required to complete long-term simulations (e.g., one year). However, the use of low resolution data leads to inaccurate simulations [8], [20], [21].

As demonstrated in the above-mentioned literature review, the assessment of PV impacts on the voltage profiles in distribution systems is essential for utilities. In this regard, annual fine-resolution simulations of the distribution system based on traditional iterative power flow methods is a highly time-consuming task. To cover the gap in the literature, we propose the employment of machine learning to the voltage assessment problem in distribution systems with PV. The proposed machine learning-based approach produces fast and accurate fine-resolution simulations of distribution systems with PV, which overcomes the limitations of the existing iterative approaches. Specifically, the proposed approach calculates the total voltage deviation (TVD) in the entire distribution system and the voltage at the point of common connection (PCC) of PV units based on machine learning techniques. Unlike the time-consuming iterative approaches, the proposed approach relies on machine learning, meaning it lightens the computational burden at the operation phase since the computational burden is shifted to the off-line training phase. Additionally, in contrast to existing approaches that reduce the resolution of data using down-sampling or clustering techniques to shorten the simulation time at the expense of accuracy, the proposed approach considers the whole data (intermittent load and PV profiles), which yields accurate and rapid simulations. Accordingly, the proposed data-driven-based approach is considered efficient for voltage assessment with intermittent PV generation considering fine-resolution simulations (i.e., time-step of 1 s). To demonstrate the effectiveness of the proposed approach, it has been tested on two distribution systems with PV considering daily and annual profiles at different time-resolutions, finding that the proposed approach reduces the computational burden of the voltage assessment to less than 1 s.

It is important to mention that the proposed approach facilitates performing accurate fine-resolution simulations for distribution systems interconnected with PV in a very short time. Accordingly, it can assist system planners to study the potential impacts of PV on voltage levels and system operators to monitor the operating voltage condition of the distribution system, which are cardinal for different control/automation tasks, such as feeder reconfiguration, restoration, and volt/VAr optimization.

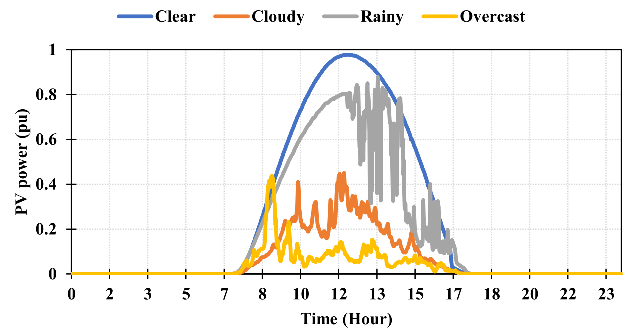

Fig. 1. PV generation profiles for different days.

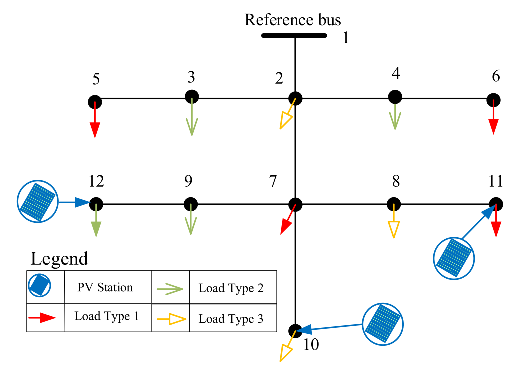

Fig. 2. Distribution system with different loads and PV.

Section II states the voltage assessment problem in the presence of PV. Section III describes the proposed machine learning-based voltage assessment approach. Sections IV and V present results and conclusion, respectively.

II. Problem Statement and Proposed Approach

The integration of PV units with high penetrations to distribution systems can be considered as a challenge for utilities. Such distributed PV generators have intermittent characteristics due to the fluctuations of the weather conditions, mainly solar irradiance and temperature. For example, Fig. 1 shows the output active power of PV during clear, cloudy, overcast, and rainy days. As shown, the PV generation profiles can have significant variations with respect to the different days. Furthermore, for all days (except the clear day), excessive fluctuations are noticed which requires fine-resolution simulations (in terms of seconds) to assess PV impacts on voltages. As a result, the voltage levels in the hosted distribution system [see Fig. (2)] are greatly alerted, and voltage assessment in the planning phase and even the operating phase of PV is essential. For this purpose, intensive time-series power flow solutions are required, which require heavy computational costs in terms of the CPU time. Driven by this limitation, we propose a fast approach for voltage assessment in distribution systems with PV. The proposed approach computes the PCC voltages, which represent terminal voltages of PV units, and the voltage deviation in the whole distribution system based on a data-driven model. This model is built using a machine learning technique considering various scenarios of PV and load profiles. Unlike the existing approaches that can takes hours to deliver annual simulations of distribution systems, the proposed data-driven approach can rabidly deliver such annual simulations within a fraction of second at a high accuracy rate.

III. Proposed Method

A. Data Structure

The objective of voltage assessment is to compute the voltages at PCCs and the total voltage deviation in distribution systems during all time instants of the studied period $\Omega_T = \{T_1, T_2, \ldots, T_{N_T}\}$. For a certain distribution system, $\Omega_{\text{Bus}} = \{B_1, B_2, \ldots, B_{N_B}\}$ and $\Omega_{PV} = \{PV_1, PV_2, \ldots, PV_{N_{PV}}\}$ symbolize the lists of all buses and PV buses, respectively. The generated power of each PV unit is denoted as S_{PV}. Besides PV, the distribution system involves different loads types, which represent the practical condition. To model these loads in the voltage assessment model, the time-series values of various load types, as well as the corresponding PV generation values can be structured in a matrix Θ, as follows:

$$\Theta = \begin{bmatrix} \alpha & \beta \end{bmatrix} \tag{1}$$

in which

$$\alpha = \begin{bmatrix} S_{L_1,T_1} & S_{L_2,T_1} & \cdots & S_{L_{N_L},T_1} \\ S_{L_1,T_2} & S_{L_2,T_2} & \cdots & S_{L_{N_L},T_2} \\ \vdots & \vdots & \ddots & \vdots \\ S_{L_1,T_{N_T}} & S_{L_2,T_{N_T}} & \cdots & S_{L_{N_L},T_{N_T}} \end{bmatrix} \tag{2}$$

$$\beta = \begin{bmatrix} S_{PV_1,T_1} & S_{PV_2,T_1} & \cdots & S_{PV_{N_{PV}},T_1} \\ S_{PV_1,T_2} & S_{PV_2,T_2} & \cdots & S_{PV_{N_{PV}},T_2} \\ \vdots & \vdots & \ddots & \vdots \\ S_{PV_1,T_{N_T}} & S_{PV_2,T_{N_T}} & \cdots & S_{PV_{N_{PV}},T_{N_T}} \end{bmatrix} \tag{3}$$

where α and β involve the profiles of various load types and PV units, respectively. In this article, we consider three load types: Residential, commercial, and industrial (details of load and PV generation profiles are described in Section V-A). Let i and j represent the rows and columns of Θ, respectively, the elements of Θ are normalized by dividing each column j with the corresponding nominal power, as follows:

$$\psi_{L_i,T_j} = S_{L_i,T_j} / S_{L_i,R} \forall i, j \in \alpha \tag{4}$$

$$\psi_{PV_i,T_j} = S_{PV_i,T_j} / S_{PV_i,R} \forall i, j \in \beta \tag{5}$$

where $S_{L_i,R}$ and $S_{PV_i,R}$ are, the rated power of load type i and nominal power of the ith PV unit, respectively. Then, a normalized matrix (Γ) can be expressed as follows:

$$\Gamma = \begin{bmatrix} \phi & \varphi \end{bmatrix} \tag{6}$$

in which

$$\phi = \begin{bmatrix} \psi_{L_1,T_1} & \psi_{L_2,T_1} & \dot{\psi} & \psi_{L_{N_L},T_1} \\ \psi_{L_1,T_2} & \psi_{L_2,T_2} & \dot{\psi} & \psi_{L_{N_L},T_2} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{L_1,T_{N_T}} & \psi_{L_2,T_{N_T}} & \cdots & \psi_{L_{N_L},T_{N_T}} \end{bmatrix} \tag{7}$$

$$\varphi = \begin{bmatrix} \psi_{PV_1,T_1} & \psi_{PV_2,T_1} & \cdots & \psi_{PV_{N_{PV}},T_1} \\ \psi_{PV_1,T_2} & \psi_{PV_2,T_2} & \cdots & \psi_{PV_{N_{PV}},T_2} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{PV_1,T_{N_T}} & \psi_{PV_2,T_{N_T}} & \cdots & \psi_{PV_{N_{PV}},T_{N_T}} \end{bmatrix} \tag{8}$$

where ϕ and φ represent the normalized matrix of loads and PVs, respectively.

B. Voltage Assessment Using Support Vector Regression

In general, all possible scenarios of PV output power and loads, and the corresponding TVD and PCC voltages can be considered as (input, output) training pairs. These pairs can be exploited to train a machine learning technique to learn a model (mapping function) that maps a possible scenario to TVD and PCC voltages. After the model is built, assessing the outputs for any new scenario of PV output power and load need simple operations with very short execution time. Indeed, there are several machine learning techniques that can be utilized to model voltage time-series data, such as support vector regression (SVR), regression trees, and neural networks. Indeed, the support vector machine (SVM) algorithm is a popular machine learning tool that offers solutions for both classification and regression problems. In this article, we employ SVR with the radial basis function (RBF) kernel to build the voltage assessment model, noting that the use of kernel trick (kernel functions) provides a way to create nonlinear regression models [22]. SVR with RBF kernel is efficient for handling nonlinear relationships between targets and input data by implicitly mapping their inputs into high-dimensional feature spaces. Another merit of the SVR technique is that its computational complexity does not depend on the dimensionality of data. Besides, it has a generalization capability with excellent prediction precision.

Let $S = \{(x_i, y_i) | x_i \in \mathbb{R}^p, y_i \in \mathbb{R},\ i = 1, 2 \ldots, n\}$ is a dataset, in which x_i is an input and y_i is the corresponding target, and n is the number of instances. The SVR algorithm finds a function $f(x)$ that is flat as possible and maps x_i to y_i with ϵ deviations

$$f(x) = \langle w, x \rangle + b \text{ with } w \in \mathbb{R}^p \text{ and } b \in \mathbb{R}. \tag{9}$$

Here $\langle \rangle$ refers to the dot product in the p-dimensional space. Note that small values of w indicate the flatness of the solution. The regression problem can be expressed as optimization problem [23], [24]

$$\text{minimize } \frac{1}{2}|w|^2 + C \sum_{i=1}^{n} (\xi_i + \xi_i^*) \tag{10}$$

subject to

$$y_i - w.\Phi(x_i) - b \leq \varepsilon + \tilde{\zeta}_i \tag{11}$$

$$w.\Phi(x_i) + b - y_i \leq \varepsilon + \xi_i^* \tag{12}$$

$$\tilde{\zeta}_i, \xi_i^* \geq 0 \tag{13}$$

where ζ_i and ζ_i^* are slack variables. The parameter $C > 0$ is presented to make the linear solution feasible, noting that $C > 0$ controls the tradeoff between the flatness of $f(x)$ and the amount to which deviations larger than ϵ. The optimization problem is usually solved in its dual formulation using Lagrange multipliers

$$\text{minimize } \frac{1}{2}\sum_{l=1}^{n}\sum_{j=1}^{N}(\alpha_i - \alpha_i^*) \cdot \Phi(x_i, x_j) \cdot (\alpha_j - \alpha_j^*)$$

$$-\sum_{i=1}^{N}(\alpha_i + \alpha_i^*) \cdot \varepsilon + \sum_{i=1}^{n}(\alpha_i - \alpha_i^*) \cdot y_i \tag{14}$$

subject to

$$\sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0; \alpha_i, \alpha_i^* \geq 0. \tag{15}$$

The following simplified expression for SVR output can be obtained after solving for the postive Lagrange multiplier pairs $(\alpha_i - \alpha_i^*)$:

$$\hat{y}(x) = \sum_{l=1}^{N}(\alpha_l - \alpha_i^*) \cdot K(x, x_i) + b. \tag{16}$$

Note that $K(x_i, x_j) = \Phi(x_i).\Phi(x_j)$ refers to the SVR kernel. Here, we use RBF kernel, which can be defined as follows: $K(x_i, x_j) = \exp(-\frac{|x_i - x_j|^2}{\gamma^2})$.

To implement the voltage assessment approach by the SVR technique, training dateset which involves input and output vectors are required. For this purpose, synthetic data are constructed using different possible scenarios of PV and load power values.To do so, the normalized load and PV power values are sampled within the higher and lower boundaries $(\eta^{\min}, \eta^{\max})$. Let us define the two vectors $\boldsymbol{X}_{Lu}$ and $\boldsymbol{X}_{PVv}$ that contain the sampled values of load type u and PV unit v, respectively, as follows:

$$\mathbf{X_{L_u}} = \begin{bmatrix} \eta_{L_u}^{\text{Min}} \\ \eta_{L_u}^{\text{Min}} + 1 \times R_{L_u} \\ \eta_{L_u}^{\text{Min}} + 2 \times R_{L_u} \\ \vdots \\ \eta_{L_u}^{\text{Max}} \end{bmatrix} \tag{17}$$

$$\mathbf{X_{PV_v}} = \begin{bmatrix} \eta_{PV_v}^{\text{Min}} \\ \eta_{PV_v}^{\text{Min}} + 1 \times R_{PV_v} \\ \eta_{PV_v}^{\text{Min}} + 2 \times R_{PV_v} \\ \vdots \\ \eta_{PV_v}^{\text{Max}} \end{bmatrix} \tag{18}$$

where R_{PV_v} and R_{L_v} are uniform sampling rates of PV and load power values, respectively. Then, a matrix $\mathbf{X}$ can be built by considering all possible combinations of $\mathbf{X_{L_u}}$ and $\mathbf{X_{PV_v}}$ elements, in which each combination represents a possible operation scenario. Once the input matrix $\mathbf{X}$ is built, the corresponding output matrix $\mathbf{Y}$ is generated. For this purpose, a power flow tool can be utilized to compute the voltages and total voltage deviations (i.e., $\mathbf{Y}$) for each possible scenario (i.e., each raw of $\mathbf{X}$). In this article, a backward forward sweep (BFS) power flow method proposed in [25] is utilized. The solution steps of this power flow method involves: 1) computing current injections at all nodes considering PV and load power, 2) computing the line power flows, and 3) updating voltages by developed quadratic-based line models. These three steps are repeated till the solutions are converged.

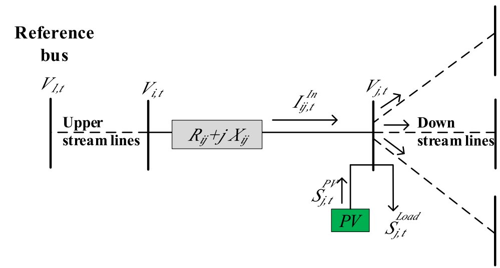

Fig. 3. Model of distribution branches interconnected with PV.

Fig. 3 shows a general branch model of distribution systems interconnected with PV. For a distribution system with various branches, the $\mathbf{Y}$ matrix can be expressed as follows:

$$\mathbf{Y} = \begin{bmatrix} \mathbf{V^{PV}} & \mathbf{TVD} \end{bmatrix} \tag{19}$$

in which

$$\mathbf{V^{PV}} = \begin{bmatrix} V_{1,1}^{PV} & \cdots & V_{1,N_{PV}}^{PV} \\ \vdots & \vdots & \vdots \\ V_{N_s,1}^{PV} & \cdots & V_{N_s,N_{PV}}^{PV} \end{bmatrix} \tag{20}$$

$$\mathbf{TVD} = \begin{bmatrix} \text{TVD}_1 \\ \vdots \\ \text{TVD}_{N_s} \end{bmatrix} \tag{21}$$

$$\text{TVD}_i = \sum_{b \in \Omega_{\text{Bus}}}^{NB} (1 - V_{b_i})^2. \tag{22}$$

The input data ($\mathbf{X}$) and output data ($\mathbf{Y}$) are employed for constructing the SVR model. Then, daily or yearly data (PVs and loads) can be fed into the trained SVR model to rapidly calculate the corresponding PCC voltages and TVD ($\mathbf{Y^{Est}}$). It is worth noting that the data used to train the voltage assessment model are generated by considering the upper and lower operating conditions of the load and PV units $(\eta^{\min}, \eta^{\max})$ in the distribution system under study. Therefore, any actual PV and load power will be within these operational limits. Note that in the assessment phase, realistic datasets for PV and load profiles, described in the result section, are utilized for evaluating the proposed approach. These datasets are considered as inputs for the trained machine learning model which in turn estimates the corresponding output voltage and TVD.

Notably, the performance of the proposed approach can be maintained under the reconfiguration of the distribution network with tie and sectionalizing switches. In general, the topology of the distribution system is decided depending on the load and PV generation conditions. Accordingly, for each load and PV scenario, the topology of the distribution network is assigned and incorporated in the matrix $\mathbf{X}$. By doing so, the voltage assessment model will be updated to consider the reconfiguration of the network.

IV. Solution Process and Evaluation Metrics

A. Solution Process

The solution process of the proposed approach is presented in Algorithm 1. It starts with reading the data of the distribution system under study including line parameters. Further, the algorithm reads all data of various load types and their corresponding profiles. Another important variable to be entered to the algorithm is the data of the PV units including the locations, sizes, and their generation profiles. Once these data are entered into the algorithm, the vectors $\mathbf{X_L}$ and $\mathbf{X_{PV}}$ are constructed according to (17) and (18) based on the adopted uniform sampling rates R_{PV} and R_L, respectively. Then the $\mathbf{X}$ matrix is built as explained in the previous subsection, and so the corresponding $\mathbf{Y}$ matrix represented in (19) is computed via the power flow method.

To build the voltage assessment model, both $\mathbf{X}$ and $\mathbf{Y}$ are fed into the SVR technique. Given daily or annual profile with a certain resolution of PV and load power data, for each time instants of the studied period $\Omega_T = \{T_1, T_2, \ldots, T_{N_T}\}$, the corresponding TVD and PCC voltages are computed using the voltage assessment model. Notably, the proposed approach considers the whole dataset avoiding down-sampling or clustering techniques that may loose information about the actual status of voltages in the distribution system. This voltage assessment model lightens the computational burden because most of the computational burden has been shifted to the off-line training process. Therefore, the use of the trained voltage assessment can perform fine-resolution simulations rapidly. Note that the BFS power flow method [25] which can deliver accurate results is utilized for computing the actual values of the voltage and TVD in the distribution system.

B. Performance Evaluation Metrics

To prove the validity of the proposed approach, we use five metrics to measure how much the computed TVD and voltages are close to the exact values. The metrics used are normalized mean squared error (NMSE), root mean squared error (RMSE), mean absolute error (MAE), mean absolute relative error (MARE), and coefficient of correlation (COC), formulated as follows:

$$\text{NMSE}_i = \frac{1}{N_T} \sum_{t=t_1}^{t_{N_T}} \frac{(|D_{i,t}^P| - |D_{i,t}^E|)^2}{a_1 * a_2} \tag{23}$$

$$\text{RMSE}_i = \sqrt{\frac{1}{N_T} \sum_{t=t_1}^{t_{N_T}} \left(|D_{i,t}^P| - |D_{i,t}^E|\right)^2} \tag{24}$$

Algorithm 1: Implementation of Proposed Method.

Read distribution system parameters and various load profiles considering their distribution among the buses.
Read data of PV including: Locations, sizes, and generation profiles.
Build the $\mathbf{X_L}$ and $\mathbf{X_{PV}}$ matrices according to (17) and (18).
Construct the matrix $\mathbf{X}$ by considering all possible combinations of $\mathbf{X_L}$ and $\mathbf{X_{PV}}$ elements.
Generate $\mathbf{Y}$ matrix with respect to $\mathbf{X}$ as described in (19).
Build the SVR model using the input matrix $\mathbf{X}$ and the target matrix $\mathbf{Y}$.
Read daily or annual PV and load profiles during the studied period.
while $t \leq N_T$
    Compute $\mathbf{Y_t^{Est}}$ using the SVR model.
    Save PCC voltages and TVD.
end
    Print results.

$$\text{MAE}_i = \frac{1}{N_T} \sum_{t=t_1}^{t_{N_T}} |D_{i,t}^P - D_{i,t}^E| \tag{25}$$

$$\text{MARE}_i = \frac{1}{N_T} \sum_{t=t_1}^{t_{N_T}} \frac{|D_{i,t}^P - D_{i,t}^E|}{D_{i,t}^E} \tag{26}$$

$$\text{COC}_i = \frac{N_T \sum_{t=t_1}^{t_{N_T}} (D_{i,t}^E D_{i,t}^P) - (\sum_{t=t_1}^{t_{N_T}} D_{i,t}^E)(\sum_{t=t_1}^{t_{N_T}} D_{i,t}^P)}{\sqrt{Z1_{i,t} Z2_{i,t}}} \tag{27}$$

where D^E and D^P refer to the actual values computed by the BFS power flow method and predicted values obtained by the trained model, respectively, $a_1 = \frac{1}{N_T} \sum_{t=t_1}^{t_{N_T}} D_{i,t}^P$, $a_2 = \frac{1}{N_T} \sum_{t=t_1}^{t_{N_T}} D_{i,t}^E$, $Z1_{i,t} = N_T \sum_{t=t_1}^{t_{N_T}} (D_{i,t}^E)^2 - (\sum_{t=t_1}^{t_{N_T}} D_{i,t}^E)^2$, and $Z2_{i,t} = N_T \sum_{t=t_1}^{t_{N_T}} (D_{i,t}^P)^2 - (\sum_{t=t_1}^{t_{N_T}} D_{i,t}^P)^2$.

V. Results and Discussions

A. Test System and Studied Cases

The proposed voltage assessment approach with PV has been implemented on MATLAB2019 with 2.4-GHz PC with 16.00 GB of RAM. The IEEE 33-bus distribution test feeder has been utilized for evaluating the performance of the proposed approach (Fig. 4). The nominal load data and parameters of the lines are given in [26]. In this article, three PV units are assumed to be distribution at buses 6, 17, and 32. To simulate the real situation where various load types exit, we consider three load types (i.e., residential, commercial, and industrial) to be allocated among the buses of the test feeder. We study two cases on the IEEE 33-bus test feeder interconnected to the three PV units to prove the efficacy of the proposed voltage assessment approach.

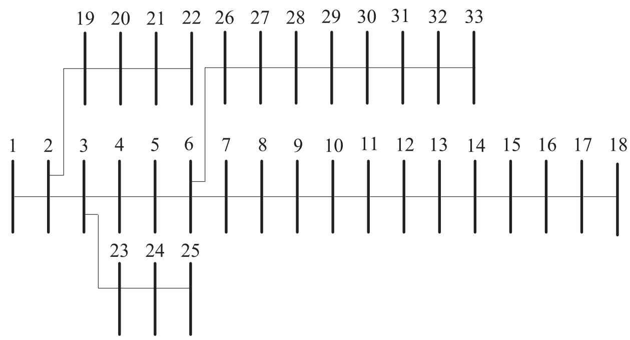

Fig. 4. IEEE 33-bus MV distribution test system.

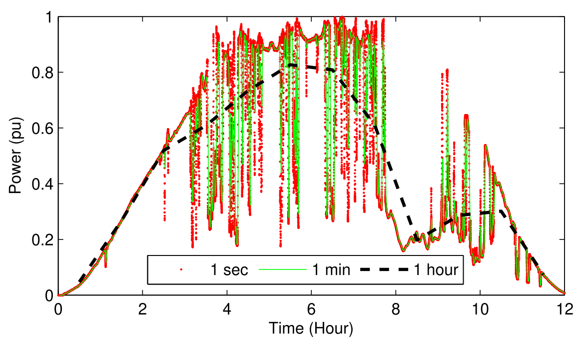

Fig. 5. PV generation profile with different time resolutions.

1) *Case 1–Daily case*: The voltage assessment is performed for a fluctuated profile of PVs with 1-s, 30-s, 1-min, 15-min, 30-min, and 1-h time resolutions.
2) *Case 2–Yearly case*: The voltage assessment is performed for a whole year of PVs with 30-s, 1-min, 15-min, 30-min, and 1-h time resolutions.

The BFS method is used to validate the results and the performance of the proposed method with the two cases. With case 1, the daily generation profiles of PVs and profiles of the three load types are given in [27] and [28], respectively. Fig. 5 shows generation profiles of PV with 1-s, 1-min and 1-h resolutions. As noticed, the variation with the lower resolutions (1 s and 1 min) are much higher than the 1-h resolutions. Therefore, the voltage assessment should be performed at the finest resolution of data (huge number of samples), which requires a fast computational approach to get the result in a reasonable time. With case 2, the profiles of the various load types and PV units for the year are taken from [29] and [30], respectively. In the assessment phase, these daily and yearly datasets are entered to the trained machine learning model for estimating the corresponding output voltage and TVD.

Here, we provide the parameters of the proposed voltage assessment approach. $\eta_{PV_v}^{\text{Min}}$ and $\eta_{L_u}^{\text{Min}}$ values are 0.0 and 0.3, respectively, while $\eta_{PV_v}^{\text{Max}}$ and $\eta_{L_u}^{\text{Max}}$ values are 1.0. R_{L_u} and R_{PV_v} values are adopted to be 0.05. Accordingly, the sample numbers in $\mathbf{X_{L_u}}$ and $\mathbf{X_{PV_v}}$ equal 15 and 21, respectively. Since the number of load types (u) is 3, the number of samples in $\mathbf{Y}$ is 70 875 according to the description of (19).

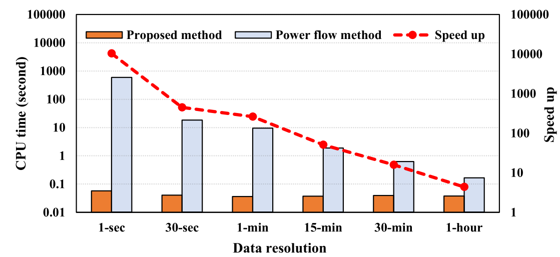

Fig. 6. Computational performance of the proposed and existing methods for daily simulations with different resolutions.

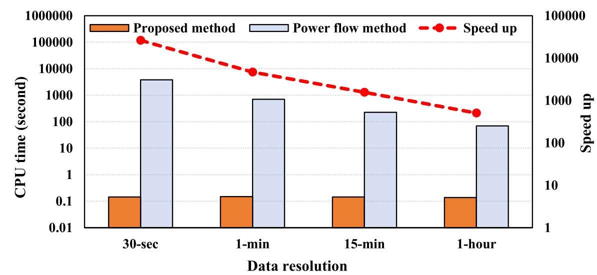

Fig. 7. Computational performance of the proposed and existing methods for year simulations with different resolutions.

B. Evaluation of the Computational Burden

Here, the computational burden (in terms of the CPU time) of the proposed voltage assessment approach with the two cases at various resolutions is studied. Figs. 6 and 7 show the computational burden of the proposed voltage assessment approach and the exact BFS method with case 1 and 2, respectively. Besides, the corresponding speedup of the proposed approach is shown at the different time resolutions. Note that the *y*-axis in Figs. 6 and 7 has the log scale. These figures demonstrate that computational burden of the iterative power flow method is heavy, especially with the high resolutions of data (i.e., massive number of samples). At the finest resolution of case 1 (1 s), the CPU time of the iterative power flow method is close to 1000 s while it is about 10 000 s at the finest resolution of case 2 (30 s). This analysis emphasizes that the iterative power flow methods are not computationally efficient for assessing voltages with high-resolution data because they require extensive calculation processes at each time instant of the simulation. On the one hand, the long execution time of the iterative power flow methods with high data resolutions limits the number of simulations to be done by system operators. On the other hand, the use of rough resolutions of data reduces the CPU time values of the iterative power flow methods on the expense of the accuracy of the voltage assessment since they do not utilize the full representation of fluctuated voltages.

Unlike the existing voltage assessment approaches based on the iterative power flow methods, the proposed approach has a very low computational burden without applying any downsampling or clustering technique to reduce the resolution of the data, and thus, benefiting the full representation of PV and load profiles. In both cases, Figs. 6 and 7 show that the CPU time

TABLE I
ACCURACY OF THE PROPOSED METHOD FOR DAILY SIMULATIONS WITH DIFFERENT RESOLUTIONS

Item	Voltage Errors					TVD Errors				
Error	NMSE	RMSE	MAE	MARE	COC	NMSE	RMSE	MAE	MARE	COC
1-sec	0.000113	0.000062	0.000053	0.000055	0.999944	0.000086	0.000242	0.000193	0.002581	0.999967
30-sec	0.000114	0.000062	0.000053	0.000055	0.999943	0.000087	0.000244	0.000195	0.002605	0.999967
1-min	0.000112	0.000061	0.000052	0.000054	0.999944	0.000088	0.000245	0.000196	0.002613	0.999966
15-min	0.00011	0.000062	0.000052	0.000054	0.999945	0.000091	0.000252	0.000202	0.002662	0.999965
30-min	0.000115	0.000063	0.000053	0.000055	0.999943	0.000084	0.000242	0.000190	0.002492	0.999971
1-hour	0.000101	0.00006	0.000051	0.000053	0.999949	0.000066	0.000217	0.000168	0.002220	0.999976

TABLE II
ACCURACY OF THE PROPOSED METHOD FOR YEAR SIMULATIONS WITH DIFFERENT RESOLUTIONS

Item	TVD errors					Voltage errors				
Error	NMSE	RMSE	MAE	MARE	COC	NMSE	RMSE	MAE	MARE	COC
30-sec	0.000371	0.000245	0.000191	0.004115	0.999815	0.00049	0.000081	0.000065	0.000068	0.999755
1-min	0.000370	0.000245	0.000190	0.004113	0.999816	0.00049	0.000081	0.000066	0.000068	0.999755
15-min	0.000369	0.000245	0.000190	0.004110	0.999816	0.000489	0.000081	0.000065	0.000068	0.999755
1-hour	0.000371	0.000246	0.000192	0.004158	0.999815	0.000484	0.000080	0.000065	0.000067	0.999758

of the proposed voltage assessment approach is almost 0.1 s at the different time resolutions, which is much lower than those of the iterative method. We also compute the speedup of the proposed approach as follows: speedup $= T_i/T_p$, where T_p and T_i are the CPU time of the proposed and iterative power flow approaches, respectively. In the case of the rough resolution of data (1 h), the speedup of the proposed method is about 10 with the daily simulations, and about 900 with annual simulations. The red dashed lines shown in Figs. 6 and 7 demonstrate that the speedup values of the proposed approach significantly increase with the use of high resolution of data. Interestingly, the speedup of the proposed approach with cases 1 and 2 at the finest data resolutions is about 10 000. This analysis emphasises the low computational burden of the proposed voltage assessment approach with both fine and rough resolutions of data.

C. Accuracy Evaluation

In this subsection, the accuracy of the proposed approach to compute the TVD and PCC voltages of PV is demonstrated for the daily and annual case studies with the different resolutions. In Tables I and II, the performance of the proposed approach in terms of NMSE, RMSE, MAE, MARE, and COC evaluation metrics, is shown for case 1 and case 2, respectively, at the different resolutions of the data. Note that the best value of NMSE, RMSE, MAE, and MARE evaluation metrics is zero, while the best value of COC is one. As noticed in Tables I and II, the values of NMSE, RMSE, MAE, and MARE achieved by the proposed approach are close to zero at all resolutions with the two cases. With case 1, the proposed approach achieves a RMSE value of 0.000062 at 1-s data resolution while it achieves a RMSE value of 0.000245 at 30-s data resolution with case 2. Further, the proposed approach obtains COC values almost equal 0.999 (COC$\approx$1) for all resolution of the daily and annual cases.

It is worth noting that fine-resolution simulations can provide a more detailed analysis for different aspects, most importantly voltage fluctuations. For the sake of simplicity and space limitations, in Fig. 8 we show the voltage at bus 6 and TVD computed by the proposed and iterative approaches at only 1-s, 1-min, and 1-h data resolutions for case 1. Notably, the fluctuation of the TVD and the PCC voltage in the case of 1 s are much higher than those in the case of 1 h. In other words, the voltage fluctuations cannot be noticed as shown in Fig. 8(a) that has a resolution of 1 h; However, the use of a finer resolution (1 s) in Fig. 8(c) reveals the existing fluctuations of the voltage during the studied period. As noticed, the voltage and the TVD values computed by the proposed approach are almost equal to the the exact values computed by the iterative power flow approach, even at the finest resolution.

D. Voltage Assessment With Different PV Penetrations

Here, the proposed voltage assessment approach is applied to the 33-bus test feeder considering different PV penetrations. This analysis is vital to show the variation of the voltage level with PV size in order to find the optimal PV size while avoiding the voltage violations. To do so, we apply the proposed approach to rapidly compute the voltage at PCC and the voltage deviation. The three PV units are located at buses mentioned in Section V-A. In this analysis, the finest resolution of the annual data is considered, where the PV size is changed from 0 to 10 000 kW.

Fig. 9 shows the variation of PCC voltage with the PV size and the average voltage deviation during the year against the PV size. In Fig. 9(a), the horizontal dashed line represents the maximum voltage limit (1.1 pu) while the vertical dashed line represents the critical PV size at which the PCC voltage exceeds the voltage limit during the year. It should be noted that each blue curve represents the voltages computed using the proposed approach for a particular sample of the annual data with the studied PV penetrations. As shown, the critical penetration equals 4200 kW, noting that any PV size larger than this limit yields excessive voltage violations. Besides, the average voltage deviation given in Fig. 9(b) demonstrates that the PV size can contribute in minimizing the voltage deviations until 1200-kW PV size at which the corresponding lowest voltage deviation is 0.05836. Beyond this limit, the the average voltage deviation is increased with the PV penetration. This analysis

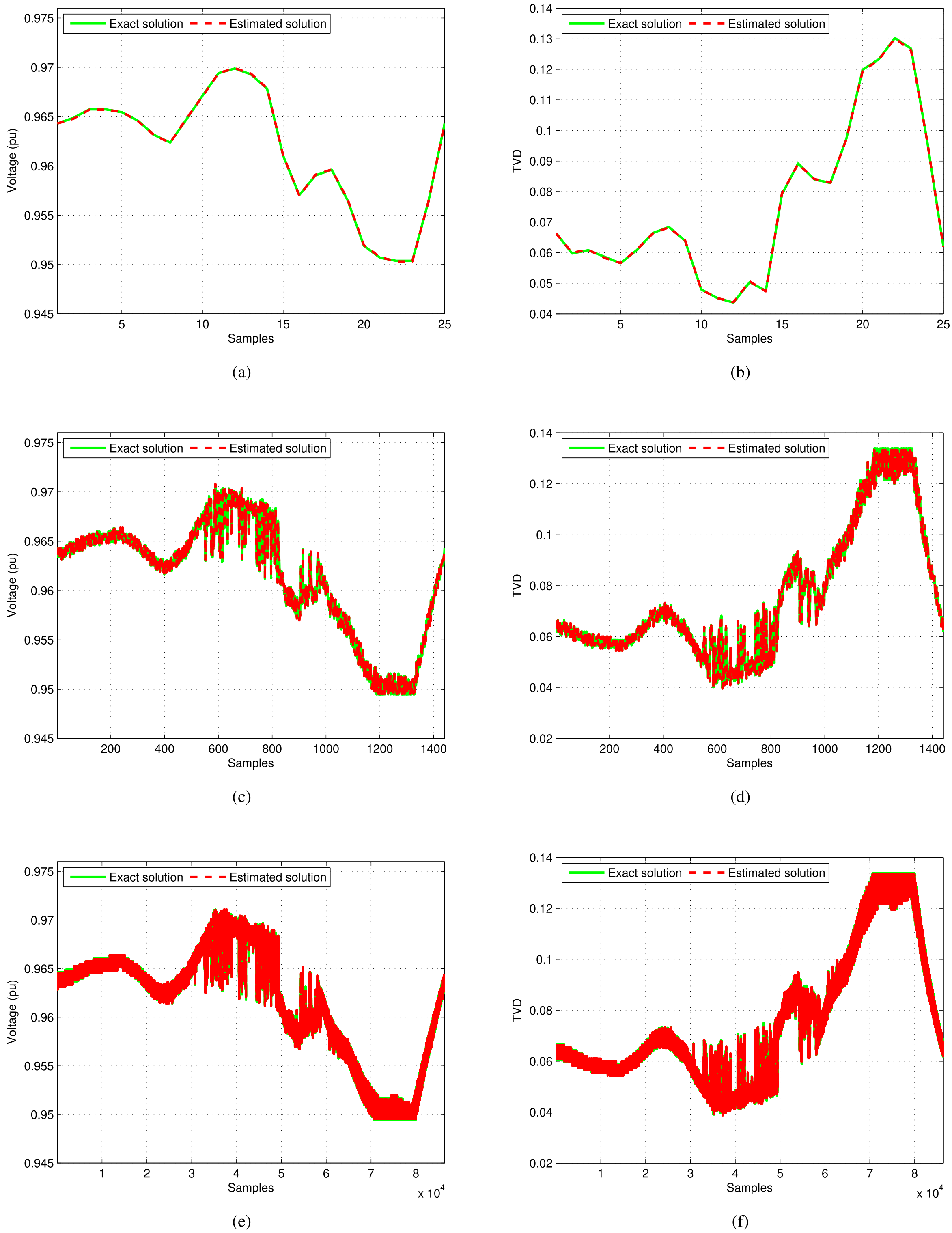

Fig. 8. Voltage at bus 6 and TVD computed by the proposed and iterative approaches at different resolutions for case 1. (a) Voltage at 1-h resolution. (b) TVD at 1-h resolution. (c) Voltage at 1-min resolution. (d) TVD at 1-min resolution. (e) Voltage at 1-s resolution. (f) TVD at 1-s resolution.

affirms the applicability of the proposed approach not only for time-series simulations but also for determining the most proper PV size with respect to the PCC voltage and the total voltage deviation.

E. Validation on Larger Distribution System

To further verify the effectiveness of the proposed approach, we test it on the 119-bus distribution system [Fig. (10)] inter-

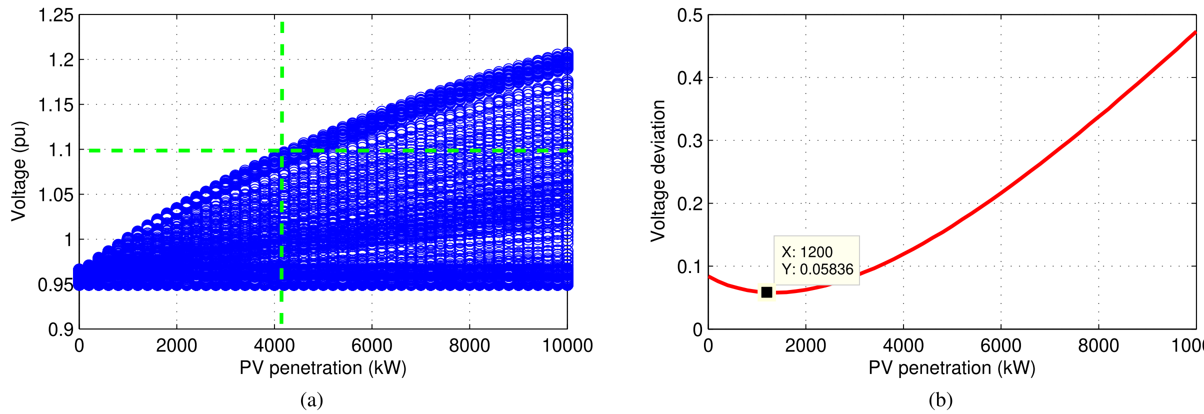

Fig. 9. Variation of the PCC voltage and the average voltage deviation during the year with PV penetration. (a) PCC voltage. (b) Average voltage deviation.

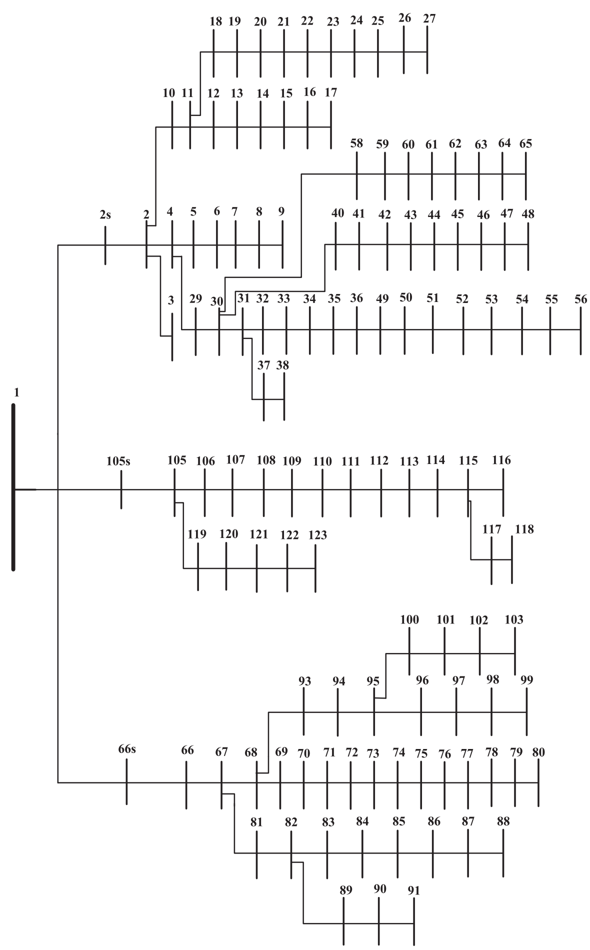

Fig. 10. 119-bus MV distribution test system.

TABLE III
ACCURACY OF THE PROPOSED METHOD FOR YEAR SIMULATIONS WITH DIFFERENT RESOLUTIONS FOR THE 119-BUS DISTRIBUTION SYSTEM

Error	NMSE	RMSE	MAE	MARE	COC
1-hour	0.000317	0.000317	0.000251	0.000959	0.999722
15-min	0.00032	0.000316	0.000250	0.000954	0.999719
1-min	0.000321	0.000317	0.000251	0.000955	0.999719
30-sec	0.00032	0.000317	0.000251	0.000956	0.999719

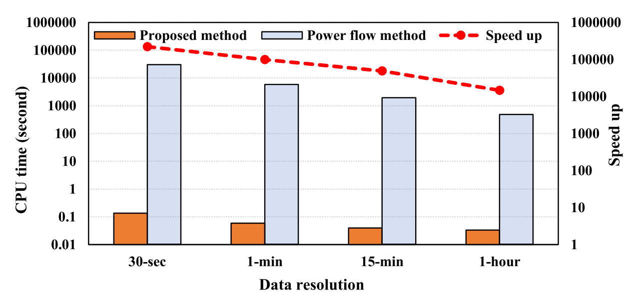

Fig. 11. Computational performance of the proposed and existing methods for year simulations with different resolutions for the 119-bus distribution system.

connected to three PV units [31]. Table III shows the maximum NMSE, RMSE, MAE, MARE values, and the minimum COC values of the TVD and voltages by the proposed approach with yearly simulations in the 119-bus distribution system. As noticed the values of NMSE, RMSE, MAE, and MARE achieved by the proposed approach are close to zero at all resolutions while the proposed approach obtains COC values almost equal 0.999 for all resolution. Fig. 11 shows the computational burden of the proposed voltage assessment approach and the exact BFS method with yearly simulations in the 119-bus distribution system. With time-resolutions of 30 s, 1 min, 15 min, and 1 h, the CPU time of the proposed approach is almost 0.1 s, which is much lower than the ones of the exact method. Interestingly, the speed-up values for this larger distribution system is higher than those of the 33-bus test system. This trend reveals that the proposed approach is efficient for voltage assessment in large-scale distribution systems.

VI. CONCLUSION

In this article, we have proposed a fast and accurate voltage assessment approach in distribution interconnected with PV using fine-resolution simulations. The proposed approach is based on data-driven models using machine learning techniques considering various scenarios of PV and load profiles. The main merit of the proposed approach is that it can calculate the total voltage deviation and terminal voltages of PV units accurately in a very short time. To demonstrate the effectiveness of the proposed approach, it has been tested on distribution systems interconnected with PV units considering daily and annual simulations. The simulation results emphasize the high precision and insignificant computational burden of the proposed approach compared to existing iterative-based methods with both fine- and rough- resolution simulations. In both daily and yearly cases, the CPU time of the proposed voltage assessment approach is almost 0.1 s at the different time resolutions, which is much lower than those of the iterative method (e.g., 10 000 s with the finest resolution of the yearly case). In terms of 5 evaluation metrics, the proposed approaches achieves accurate TVD and PCC voltages, specifically it obtains very small NMSE, RMSE, MAE, MARE values, and COC close to one. The proposed approach can play a vital role to assist power utilities in decision-making relating to assessing the influence of PV on voltages levels in distribution systems with high resolution of data. In the future, the proposed approach will be extended to include other DG types and energy storage systems.

References

[1] A. Blakers, M. Stocks, B. Lu, C. Cheng, and R. Stocks, "Pathway to 100% renewable electricity," *IEEE J. Photovolt.*, vol. 9, no. 6, pp. 1828–1833, Nov. 2019.

[2] I. Mauleón, "Assessing PV and wind roadmaps: Learning rates, risk, and social discounting," *Renewable Sustain. Energy Rev.*, vol. 100, pp. 71–89, 2019.

[3] S. Xia, Z. Ding, T. Du, D. Zhang, M. Shahidehpour, and T. Ding, "Multi-time scale coordinated scheduling for the combined system of wind power, photovoltaic, thermal generator, hydro pumped storage and batteries," *IEEE Trans. Ind. Appl.*, vol. 56, no. 3, pp. 2227–2237, May/Jun. 2020.

[4] K. Mahmoud and M. Abdel-Nasser, "Fast yet accurate energy-loss-assessment approach for analyzing/sizing PV in distribution systems using machine learning," *IEEE Trans. Sustain. Energy*, vol. 10, no. 3, pp. 1025–1033, Jul. 2019.

[5] M. Aly, E. M. Ahmed, and M. Shoyama, "Thermal and reliability assessment for wind energy systems with DSTATCOM functionality in resilient microgrids," *IEEE Trans. Sustain. Energy*, vol. 8, no. 3, pp. 953–965, Jul. 2017.

[6] *IEEE Approved Draft Guide to Conducting Distribution Impact Studies for Distributed Resource Interconnection*. IEEE P1547.7/D11, Feb. 2014.

[7] R. F. Arritt and R. C. Dugan, "Value of sequential-time simulations in distribution planning," in *Proc. IEEE Rural Electric Power Conf.*, 2013, pp. C 2-1–C2-5.

[8] J. Deboever, X. Zhang, M. J. Reno, R. J. Broderick, S. Grijalva, and F. Therrien, "Challenges in reducing the computational time of QSTS simulations for distribution system analysis," Sandia Nat. Lab., Albuquerque, NM, USA, SAND2017-5743, 2017.

[9] X. Zhang, S. Grijalva, M. J. Reno, J. Deboever, and R. J. Broderick, "A fast quasi-static time series (QSTS) simulation method for PV impact studies using voltage sensitivities of controllable elements," in *Proc. IEEE 44th Photovoltaic Specialist Conf.*, 2017, pp. 1555–1560.

[10] J. Hernandez, F. Ruiz-Rodriguez, and F. Jurado, "Technical impact of photovoltaic-distributed generation on radial distribution systems: Stochastic simulations for a feeder in spain," *Int. J. Elect. Power Energy Syst.*, vol. 50, pp. 25–32, 2013.

[11] M. Hasheminamin, V. G. Agelidis, V. Salehi, R. Teodorescu, and B. Hredzak, "Index-based assessment of voltage rise and reverse power flow phenomena in a distribution feeder under high PV penetration," *IEEE J. Photovolt.*, vol. 5, no. 4, pp. 1158–1168, Jul. 2015.

[12] M. Alam, K. Muttaqi, and D. Sutanto, "An approach for online assessment of rooftop solar PV impacts on low-voltage distribution networks," *IEEE Trans. Sustain. Energy*, vol. 5, no. 2, pp. 663–672, Apr. 2014.

[13] K. Mahmoud and M. Abdel-Nasser, "Efficient SPF approach based on regression and correction models for active distribution systems," *IET Renewable Power Gener.*, vol. 11, no. 14, pp. 1778–1784, 2017.

[14] H. Mirzaee, V. Zamani, and F. Katiraei, "A methodology for primary voltage assessment on highly PV penetrated distribution feeders by using SCADA and AMI," in *Proc. IEEE Power Energy Soc. Innov. Smart Grid Technol. Conf.*, 2019.

[15] J. Widén, E. Wäckelgård, J. Paatero, and P. Lund, "Impacts of distributed photovoltaics on network voltages: Stochastic simulations of three swedish low-voltage distribution grids," *Electric Power Syst. Res.*, vol. 80, no. 12, pp. 1562–1571, 2010.

[16] J. Deboever, S. Grijalva, M. J. Reno, and R. J. Broderick, "Fast quasi-static time-series (QSTS) for yearlong PV impact studies using vector quantization," *Sol. Energy*, vol. 159, pp. 538–547, 2018.

[17] A. Pagnetti and G. Delille, "A simple and efficient method for fast analysis of renewable generation connection to active distribution networks," *Electric Power Syst. Res.*, vol. 125, pp. 133–140, 2015.

[18] C. D. López, B. Idlbi, T. Stetz, and M. Braun, *Shortening quasi-static time-series simulations for cost-benefit analysis of low voltage network operation with photovoltaic feed-in*. Universitätsbibliothek Dortmund, 2015.

[19] B. Mather, "Fast determination of distribution-connected PV impacts using a variable time-step quasi-static time-series approach: Preprint," *Nat. Renewable Energy Lab.*, Golden, CO, USA, Tech. Rep. NREL/CP-5D00-67769, 2017.

[20] A. Nguyen *et al.*, "High PV penetration impacts on five local distribution networks using high resolution solar resource assessment with sky imager and quasi-steady state distribution system simulations," *Sol. Energy*, vol. 132, pp. 221–235, 2016.

[21] B. L. Norris and J. H. Dise, "High resolution PV power modeling for distribution circuit analysis," *Nat. Renewable Energy Lab.*, Golden, CO, USA, Tech. Rep. NREL/SR-5D00-60476, 2013.

[22] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proc. 5th Annu. Workshop Comput. Learn. Theory*, 1992, pp. 144–152.

[23] H. Drucker, C. J. Burges, L. Kaufman, A. J. Smola, and V. Vapnik, "Support vector regression machines," in *Proc. Adv. Neural Inf. Process. Syst.*, 1997, pp. 155–161.

[24] U. Thissen, R. Van Brakel, A. De Weijer, W. Melssen, and L. Buydens, "Using support vector machines for time series prediction," *Chemometrics Intell. Lab. Syst.*, vol. 69, no. 1/2, pp. 35–49, 2003.

[25] K. Mahmoud and N. Yorino, "Robust quadratic-based BFS power flow method for multi-phase distribution systems," *IET Gener., Transmiss. Distrib.*, vol. 10, no. 9, pp. 2240–2250, 2016.

[26] M. Kashem, V. Ganapathy, G. Jasmon, and M. Buhari, "A novel method for loss minimization in distribution networks," in *Proc. IEEE Int. Conf. Elect. Utility Deregulation Restructuring Power Technol.*, 2000, pp. 251–256.

[27] Open Distribution System Simulator. OpenDSS, 2015. [Online]. Available: https://sourceforge.net/

[28] R. Taleski and D. Rajicic, "Energy summation method for energy loss computation in radial distribution networks," *IEEE Trans. Power Syst.*, vol. 11, no. 2, pp. 1104–1111, May 1996.

[29] "Southern California Edison 2010 static load profiles" Accessed: Jan. 9, 2017. [Online]. Available: https://www.sce.com/wps/portal/home/regulatory/load-profiles

[30] T. Stoffel and A. Andreas, "NREL solar radiation research laboratory (SRRL): Baseline measurement system (BMS); Golden, Colorado (Data)," *Nat. Renewable Energy Lab.*, Golden, CO, USA, *NREL/DA-5500-56488*, 1981.

[31] K. Mahmoud, M. M. Hussein, M. Abdel-Nasser, and M. Lehtonen, "Optimal voltage control in distribution systems with intermittent PV using multiobjective Grey-Wolf-Lévy optimizer," *IEEE Syst. J.*, vol. 14, no. 1, pp. 760–770, Mar. 2020.

Karar Mahmoud received the B.Sc. and M.Sc. degrees in electrical engineering from Aswan University, Aswan, Egypt, in 2008 and 2012, respectively, and the Ph.D. degree from Graduate School of Engineering, Hiroshima University, Hiroshima, Japan, in 2016.

Since 2019, he has been a Postdoctoral Researcher with the Prof. M. Lehtonen's Power Systems and High Voltage Engineering Group, School of Electrical Engineering, Aalto University, Espoo, Finland. Since 2010, he has been with Aswan University, where he is currently an Associate Professor with the Department of Electrical Engineering. He has authored or coauthored more than 90 publications in top-ranked journals, including IEEE journals, international conferences, and book chapters. His research interests include power systems, renewable energy sources, smart grids, distributed generation, optimization, applied machine learning, IoT, Industry 4.0, and high voltage.

Dr. Mahmoud is a Topic Editor for *Sensors* journal (MDPI), since 2021.

Mohamed Abdel-Nasser received the Ph.D. degree in computer engineering from the University Rovira i Virgili, Tarragona, Spain, in 2016.

He has participated in several projects funded by the European Union and the Government of Spain. He has authored and coauthored more than 70 papers in international journals and conferences. His research interests include the application of machine learning and deep learning to several real-world problems, including smart grid analysis, time-series forecasting, and medical image analysis.

Dr. Abdel Nasser was the recipient of the Marc Esteva Vivanco Prize for the Best Ph.D. Dissertation on Artificial Intelligence, in 2017. He is currently serving as a Guest Editor for *Forecasting* with a special issue entitled "Advanced Forecasting Methods with Applications to Smart Grids."

Matti Lehtonen received the master's and Licentiate degrees in electrical engineering from the Helsinki University of Technology, Espoo, Finland, in 1984 and 1989, respectively, and the D.Tech. degree from the Tampere University of Technology, Tampere, Finland, in 1992.

He was with VTT Energy, Espoo, Finland, from 1987 to 2003, and since 1999, he has been a Full Professor and the Head of Power Systems and the High Voltage Engineering Group, Aalto University, Espoo, Finland. His research interests include power systems planning and assets management, power systems protection including earth fault problems, harmonic related issues, high voltage systems, power cable insulation, and polymer nanocomposites.

Dr. Lehtonen is currently an Associate Editor for *Electric Power Systems Research* and *IET Generation, Transmission, and Distribution*.