



ARTICLE

Dirichlet Process Mixture Model-Based Affine Voltage-Risk-Aware Planning of Distributed Photovoltaic Generation under Source-Load Uncertainty

Bin Lin¹, Jingmiao Huang¹, Quanmin Chen¹, Sihui Ke¹, Yusi Xu² and Yan Zhang^{2,*}

¹State Grid Longyan Power Supply Company of Fujian Electric Power Co., Ltd., Longyan, China

²College of Electrical Engineering and Automation, Fuzhou University, Fuzhou, China

*Corresponding Author: Yan Zhang. Email: zyww@fzu.edu.cn

Received: 08 July 2026; Accepted: 11 August 2026; Published: 18 October 2026

ABSTRACT: High penetration of distributed photovoltaic (PV) generation increases the impact of source-load uncertainty on voltage security and economic performance in distribution network planning. Planning methods that neglect uncertainty may underestimate voltage risks caused by PV and load fluctuations and may not fully reflect the performance differences among candidate PV schemes under different operating scenarios. To address this issue, this paper develops a distributed PV planning method that combines Dirichlet process mixture model (DPMM)-based scenario modeling with affine voltage risk evaluation. Historical PV and load data are organized into daily joint PV-load samples, and DPMM is used to identify latent source-load operating components. Based on these components, representative scenarios are constructed, from which occurrence probabilities, central profiles, and quantile-based fluctuation bounds are obtained. The scenario centers and fluctuation bounds are then converted into affine power injections, so that intra-scenario fluctuations can be propagated to nodal voltage intervals through affine power flow. In the proposed planning and evaluation framework, candidate PV siting and sizing schemes are assessed using scenario-weighted economic cost and affine voltage-risk indicators, and particle swarm optimization is used to update the planning decisions. Case studies on the IEEE 69-bus system and a practical 80-bus distribution network, together with annual Monte Carlo validation, indicate that the proposed method can incorporate uncertainty-induced voltage risk and multi-scenario economic performance into PV planning, thereby supporting a balanced trade-off between economic performance and voltage security.

KEYWORDS: Distributed photovoltaic planning; Dirichlet process mixture model; affine arithmetic; source-load uncertainty; voltage risk; scenario-based planning

1 Introduction

Driven by the low-carbon energy transition, photovoltaic (PV) generation has been increasingly integrated into distribution networks [1–3]. However, PV output is inherently intermittent and is affected by meteorological and geographical conditions, while load demand also varies over time. Under high PV penetration, the combined uncertainty of PV generation and load demand may change nodal power injections and aggravate operational issues such as voltage violations, reverse power flow, and network loss variation [4,5]. Since PV installation locations and capacities directly determine the spatial distribution of power injections and voltage profiles, distributed PV planning, especially PV siting and sizing decisions, should consider not only economic performance but also voltage-boundary risks under uncertain operating conditions [6,7].

Distributed PV siting and sizing jointly determine installation locations and capacity allocations and are generally formulated as a mixed-integer nonlinear optimization problem [8]. Existing studies have developed single and multi-objective planning models with objectives such as network loss reduction, voltage profile improvement, investment cost reduction, hosting capacity enhancement, and power quality improvement [9,10]. Recent studies have also considered coordinated PV-storage investment planning and DG allocation in unbalanced distribution systems [11–13]. Because these models usually involve nonlinear power flow constraints and discrete-continuous decision variables, metaheuristic and intelligent optimization algorithms have been widely applied to search for planning solutions [14,15]. Nevertheless, many existing planning models still evaluate candidate schemes using deterministic PV generation and load demand profiles, such as forecasted curves or typical daily profiles. Although such deterministic representations are convenient for planning calculation, they may underestimate voltage-boundary risk and operating-performance variations when actual source-load injections deviate from their nominal values [16,17].

To account for source-load uncertainty, various uncertainty-aware methods have been introduced into distribution network planning. Source-load uncertainty can be represented through random samples, representative scenarios, or probabilistic models [18–20]. Scenario-based stochastic and chance-constrained methods incorporate uncertainty through probability-weighted scenarios or probabilistic security constraints [21,22], whereas robust and distributionally robust methods address bounded and distributional uncertainty, respectively [23,24]. In scenario-based evaluation, using numerous operating samples increases the computational burden, whereas using only representative scenario centers may neglect intra-scenario fluctuations. Interval-based methods provide an alternative way to characterize intra-scenario fluctuations by describing uncertain inputs within prescribed bounds and estimating the corresponding range of system states. However, conventional interval analysis may produce conservative results because dependency information among uncertain variables is not fully retained. Affine arithmetic alleviates this limitation by representing uncertain variables using central values and shared noise elements, thereby retaining part of the dependence information during uncertainty propagation [25,26]. Affine power flow therefore provides a tractable means of estimating nodal voltage intervals under bounded source-load deviations.

For planning studies based on historical PV and load data, an important issue is how to identify representative source-load conditions and their associated fluctuation ranges. Conventional clustering methods, such as K-means, have been widely used for typical scenario extraction and scenario reduction in power system operation and planning [27,28]. However, these methods usually require the number of clusters to be specified in advance, which may introduce subjectivity when the actual number of PV-load operating patterns is unknown. The Dirichlet process mixture model (DPMM), as a nonparametric Bayesian mixture model, provides a data-driven way to identify latent operating components without fixing the component count beforehand [29–31]. Related DPMM-based models have also been applied to data-driven dispatch and distribution-network risk assessment [32,33]. For planning applications, the DPMM-derived latent components can be further organized into representative source-load scenarios, from which occurrence probabilities, central profiles, and quantile-based fluctuation bounds can be obtained.

For the planning problem considered in this study, scenario probabilities and scenario-specific bounds provide complementary uncertainty information. Scenario probabilities support the weighted evaluation of planning performance across representative operating conditions, whereas scenario-specific bounds describe source-load fluctuations within each condition and provide inputs for voltage-risk evaluation. Their coordinated use within a data-driven PV siting and sizing framework remains insufficiently explored.

Motivated by the above discussion, this paper proposes a DPMM-based affine voltage-risk-aware planning method for distributed PV siting and sizing under source-load uncertainty. Its principal methodological contribution lies in linking DPMM-based joint source-load scenario construction with scenario-dependent affine voltage-risk evaluation, thereby retaining both differences among representative operating conditions and fluctuations within each condition during candidate-scheme evaluation. Historical PV generation and load demand data are first organized into daily joint PV-load samples. DPMM is then used to identify latent source-load operating components and construct representative scenarios with occurrence probabilities, central profiles, and quantile-based fluctuation bounds. The scenario central profiles and fluctuation bounds are converted into affine source-load injections, whereas the occurrence probabilities are retained for multi-scenario planning evaluation. Affine power flow is then employed to propagate the intra-scenario fluctuations into nodal voltage intervals. During optimization, candidate PV siting and sizing schemes are evaluated by considering scenario-weighted economic cost and affine voltage-boundary risk. The resulting optimization problem is solved using particle swarm optimization (PSO).

The main contributions are summarized as follows.

- (1) A DPMM-based source-load scenario construction method is developed for distributed PV planning. Unlike predefined weather classification or clustering with a fixed number of clusters, the proposed scenario construction process identifies latent operating components from historical PV-load data and constructs representative scenarios with occurrence probabilities, central profiles, and quantile-based fluctuation bounds.
- (2) A scenario-dependent affine representation of source-load uncertainty is established for voltage-boundary risk evaluation. The DPMM-derived scenario centers are used as affine central values, while the intra-scenario fluctuation bounds are converted into affine radii, enabling scenario-level uncertainty propagation through affine power flow.
- (3) A multi-scenario economic and affine voltage-risk-aware PV siting and sizing model is formulated. Candidate planning schemes are evaluated by jointly considering scenario-weighted economic performance and voltage-boundary risk across representative source-load scenarios, and the model is solved using PSO and examined on the IEEE 69-bus system and a practical 80-bus distribution network through annual Monte Carlo validation.

2 DPMM-Based Source-Load Uncertainty Modeling and Affine Power Flow

The proposed method links data-driven source-load scenario construction, affine voltage-risk evaluation, and PV siting and sizing within a unified workflow. Historical PV generation and load demand are first organized into joint daily profiles, from which representative source-load scenarios are constructed with occurrence probabilities, central profiles, and fluctuation bounds. During optimization, each candidate PV configuration is evaluated under these scenarios using affine power flow, and the resulting economic cost and voltage-risk indicators are returned to update the planning decisions. The main stages and information exchange of the proposed method are shown in Fig. 1.

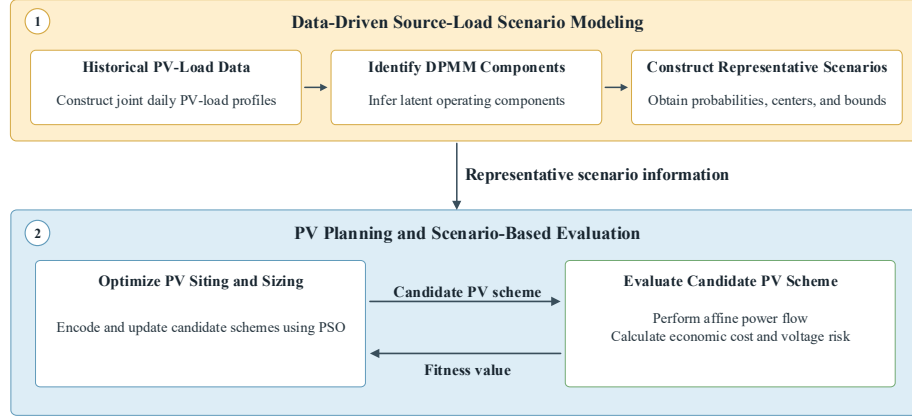


Figure 1: Overall framework of the proposed method.

2.1 DPMM-Based Source-Load Scenario Construction

Historical hourly PV generation and load demand records are organized into paired daily profiles to preserve their temporal correspondence and characterize joint source-load operating conditions. Let d denote the day index, t denote the time index within one day, and T denote the number of time periods in one day. To eliminate the influence of different PV capacities and load scales, PV output and load demand are normalized by the rated PV capacity and the base load, respectively. The normalized PV profile and load profile of the same day are combined into a daily joint PV-load sample:

$$\begin{cases} \rho_d = [\rho_{d,1}, \rho_{d,2}, \dots, \rho_{d,T}]^T \\ \lambda_d = [\lambda_{d,1}, \lambda_{d,2}, \dots, \lambda_{d,T}]^T \\ u_d = [\rho_d, \lambda_d] \end{cases} \quad (1)$$

where u_d is the daily joint source-load sample.

Daily PV and load profiles contain 24-h temporal information and have different numerical scales. To obtain a compact representation for joint scenario construction, each daily sample is mapped into a source-load feature space. After standardization, the PV-related and load-related feature vectors are weighted and concatenated as:

$$\begin{cases} f_d^{PV} = \mathcal{F}_{PV}(\rho_d) \\ f_d^L = \mathcal{F}_L(\lambda_d) \\ x_d = [w_{PV}\tilde{f}_d^{PV}, w_L\tilde{f}_d^L] \end{cases} \quad (2)$$

where $\mathcal{F}_{PV}(\cdot)$ and $\mathcal{F}_L(\cdot)$ denote the feature extraction mappings for PV and load daily profiles, respectively; f_d^{PV} and f_d^L are the PV and load feature vectors of day d ; $\tilde{f}_d^{PV}$ and $\tilde{f}_d^L$ are their standardized forms; w_{PV} and w_L are feature weights used to balance the influence of PV and load characteristics. The PV features describe the daily generation level and profile shape, while the load features characterize the main variation pattern of the daily load profile. Joint modeling preserves the daily correspondence between PV generation and load demand, so that the constructed scenarios can reflect combined source-load operating patterns rather than separate PV or load states.

Unlike PV-only profiles, which can often be associated with weather-related generation patterns, joint PV-load samples combine weather-dependent PV output with load demand influenced by both weather and consumption behavior. The latent patterns may therefore involve diverse PV-load combinations. DPMM is used to identify these latent components without fixing their number during the initial identification stage, and the identified components provide the basis for subsequent representative scenario construction.

The basic generative process of DPMM is expressed as follows [29]:

$$\begin{cases} G \sim DP(\alpha_D, G_0) \\ \theta_d | G \sim G \\ x_d | \theta_d \sim F(x_d | \theta_d) \end{cases} \quad (3)$$

where G is a random distribution following the Dirichlet process, G_0 is the base distribution, α_D is the concentration parameter, θ_d is the mixture component parameter of sample d , and $F(\cdot)$ is the conditional distribution of the sample.

The standardized source-load features are continuous and may exhibit multiple local patterns. Each latent component is therefore represented by a multivariate Gaussian distribution, whose mean and covariance describe the center and dependence structure of the corresponding feature group [32,33]:

$$x_d | (z_d = k, \mu_k, \Sigma_k) \sim \mathcal{N}(\mu_k, \Sigma_k) \quad (4)$$

where z_d is the latent component label of day d ; μ_k and Σ_k are the mean vector and covariance matrix of the k -th Gaussian mixture component.

Under the DPMM, the probability of assigning a sample to an existing component is proportional to the number of samples currently assigned to that component, while the probability of generating a new component is controlled by the concentration parameter [30]:

$$\begin{cases} p(z_d = k | z_{-d}) = \frac{n_{k,-d}}{N-1+\alpha_D} \\ p(z_d = \text{new} | z_{-d}) = \frac{\alpha_D}{N-1+\alpha_D} \end{cases} \quad (5)$$

where z_{-d} denotes the component labels of all samples except sample d , $n_{k,-d}$ is the number of samples assigned to component k excluding sample d , N is the total number of samples. Posterior inference is performed using a collapsed Gibbs sampler, in which the latent component assignments are updated iteratively and the concentration parameter is updated during sampling. The final component assignments provide the raw components for subsequent representative-scenario organization.

The latent components may represent detailed PV-load combinations, some of which contain only a few samples. For use in the proposed planning model, they are organized into representative scenarios according to sample size and similarity in the source-load feature space. This reduces the number of scenario evaluations while allowing empirical fluctuation bounds to be estimated from the samples assigned to each representative scenario. The similarity between two components is measured by the distance between their feature centers:

$$\begin{cases} \mathbf{m}_k = \frac{1}{|C_k|} \sum_{d \in C_k} x_d \\ d_{ki} = \|\mathbf{m}_k - \mathbf{m}_i\|_2 \end{cases} \quad (6)$$

where C_k is the set of daily samples assigned to the k -th latent component, $\mathbf{m}_k$ is the feature center of this component, and d_{kl} measures the feature-space distance between components k and l . A smaller d_{kl} indicates higher similarity between the two source-load operating components. Components with close feature centers can be represented by the same representative scenario when constructing the planning scenario set.

After component organization, each representative scenario corresponds to a set of original daily PV-load samples. Its occurrence probability is estimated from the sample frequency and used as the scenario weight for aggregating scenario-dependent economic cost and affine voltage risk in the subsequent planning evaluation:

$$\begin{cases} \mathcal{D}_k = \{d | z_d = k\} \\ \pi_k = \frac{|\mathcal{D}_k|}{N} \\ \sum_{k=1}^K \pi_k = 1 \end{cases} \quad (7)$$

where $\mathcal{D}_k$ is the set of days assigned to scenario k , and π_k is the occurrence probability of scenario k .

For each scenario, the representative PV and load profiles are recalculated from the original 24-h profiles of samples assigned to the same scenario:

$$\begin{cases} \bar{\rho}_{k,t} = \frac{1}{|\mathcal{D}_k|} \sum_{d \in \mathcal{D}_k} \rho_{d,t} \\ \bar{\lambda}_{k,t} = \frac{1}{|\mathcal{D}_k|} \sum_{d \in \mathcal{D}_k} \lambda_{d,t} \end{cases} \quad (8)$$

where $\bar{\rho}_{k,t}$ and $\bar{\lambda}_{k,t}$ are the representative PV output and load demand of scenario k at time t , respectively.

To describe intra-scenario PV and load fluctuations, empirical quantiles are used to construct the fluctuation bounds:

$$\begin{cases} \rho_{k,t}^{\min} = Q_{0.05}(\rho_{d,t} | d \in \mathcal{D}_k) \\ \rho_{k,t}^{\max} = Q_{0.95}(\rho_{d,t} | d \in \mathcal{D}_k) \\ \lambda_{k,t}^{\min} = Q_{0.05}(\lambda_{d,t} | d \in \mathcal{D}_k) \\ \lambda_{k,t}^{\max} = Q_{0.95}(\lambda_{d,t} | d \in \mathcal{D}_k) \end{cases} \quad (9)$$

where $\rho_{k,t}^{\min}$ and $\rho_{k,t}^{\max}$ are the lower and upper bounds of the normalized PV output in scenario k and $\lambda_{k,t}^{\min}$ and $\lambda_{k,t}^{\max}$ are the lower and upper bounds of the normalized load demand. $Q_{0.05}(\cdot)$ and $Q_{0.95}(\cdot)$ denote the 5% and 95% empirical quantiles, respectively.

Through the above process, the scenario construction procedure yields representative source-load scenarios with associated probabilities, central profiles, and intra-scenario fluctuation bounds, which provide the basis for scenario-weighted economic evaluation and scenario-dependent affine voltage-risk assessment.

2.2 Affine Representation of Scenario-Based Source-Load Uncertainty

Based on the representative scenarios obtained in Section 2.1, affine arithmetic is used to express the scenario central profiles and intra-scenario fluctuation bounds as affine uncertain source-load injections. Affine arithmetic is an interval-based uncertainty analysis method that represents an uncertain quantity as a central value plus a weighted combination of bounded noise elements. Unlike conventional interval arithmetic, which only records the lower and upper bounds of uncertain variables, affine arithmetic preserves part of the dependency information among variables through shared noise elements [34,35]. In affine arithmetic, an uncertain variable can be expressed as follows:

$$\hat{x} = x_0 + \sum_{m=1}^M x_m \varepsilon_m \quad (10)$$

where x_0 is the central value; each $\varepsilon_m \in [-1, 1]$ is a noise element representing the fluctuation of uncertainty source m ; and x_m is the partial deviation coefficient indicating the influence magnitude of each uncertainty.

The range of the affine variable is determined by the central value and the deviation coefficient. Its lower and upper bounds can be calculated using the following formula:

$$\begin{cases} \bar{x} = x_0 + \sum_{m=1}^M |x_m| \\ \underline{x} = x_0 - \sum_{m=1}^M |x_m| \end{cases} \quad (11)$$

The scenario center obtained in Section 2.1 is used as the affine central value, and the intra-scenario fluctuation range is converted into the coefficient of the affine noise term. According to the scenario profiles and fluctuation bounds, the fluctuation radii of PV output and load demand are defined as follows:

$$\begin{cases} \Delta \rho_{k,t} = \max\{\bar{\rho}_{k,t} - \rho_{k,t}^{\min}, \rho_{k,t}^{\max} - \bar{\rho}_{k,t}\} \\ \Delta \lambda_{k,t} = \max\{\bar{\lambda}_{k,t} - \lambda_{k,t}^{\min}, \lambda_{k,t}^{\max} - \bar{\lambda}_{k,t}\} \end{cases} \quad (12)$$

where $\Delta \rho_{k,t}$ and $\Delta \lambda_{k,t}$ are the fluctuation radii of PV output and load demand in scenario k at time t , respectively.

After the representative source-load profiles and fluctuation radii are obtained, they are converted into affine nodal power injections for power flow calculation. For node i , scenario k , and time t , the affine active power injections of PV generation and load demand are expressed as:

$$\begin{cases} \hat{P}_{i,k,t}^{PV} = S_i^{PV} (\bar{\rho}_{k,t} + \Delta \rho_{k,t} \varepsilon_{k,t}^{PV}) \\ \hat{P}_{i,k,t}^L = P_i^L (\bar{\lambda}_{k,t} + \Delta \lambda_{k,t} \varepsilon_{k,t}^L) \end{cases} \quad (13)$$

The corresponding reactive powers are determined according to the specified power factors of PV generation and load demand. Therefore, the affine net active and reactive power injections are written as:

$$\begin{cases} \hat{P}_{i,k,t} = \hat{P}_{i,k,t}^{PV} - \hat{P}_{i,k,t}^L \\ \hat{Q}_{i,k,t} = \eta^{PV} \hat{P}_{i,k,t}^{PV} - \eta^L \hat{P}_{i,k,t}^L \end{cases} \quad (14)$$

where η^{PV} and η^L are the reactive power coefficients determined by the PV and load power factors, respectively.

Finally, the affine complex power injection can be expressed as:

$$\begin{aligned}\hat{S}_{i,k,t} &= \hat{P}_{i,k,t} + j\hat{Q}_{i,k,t} \\ &= S_{i,0,t} + \sum_{k=1}^{N_\xi} S_{i,k,t} \varepsilon_k\end{aligned}\quad (15)$$

where $S_{i,0,t}$ is the central value of the nodal net complex power injection; $S_{i,k,t}$ is the deviation coefficient associated with different sources of uncertainty and ε_k is the corresponding noise element.

2.3 Affine Power Flow Propagation

To evaluate the voltage response under source-load uncertainty, the affine net power injections defined above are introduced into the power flow equations. Based on the nodal-admittance formulation in [25], the relationship between nodal currents and voltages is expressed as follows:

$$\begin{bmatrix} Y & Y_{in} \\ Y_{ni} & Y_{nn} \end{bmatrix} \begin{bmatrix} V \\ V_n \end{bmatrix} = \begin{bmatrix} \left(\frac{S}{V}\right)^* \\ \left(\frac{S_n}{V_n}\right)^* \end{bmatrix}\quad (16)$$

where Y denotes the nodal admittance matrix; V and S denote the nodal voltage and complex power, respectively; the subscript n represents non-slack buses.

Substituting the affine net power injections into the power-flow equations gives the iterative affine voltage formulation:

$$\hat{V} = Y^{-1} \left[\left(\frac{\hat{S}}{\hat{V}} \right)^* - Y_{ni} \hat{V}_n \right]\quad (17)$$

The iterative formulation in (17) is evaluated in affine form. At each iteration, the central values and affine coefficients of the bus voltages are updated using the scenario-specific affine power injections [35]. The network topology and model parameters are assumed to remain unchanged within each representative scenario, while the PV and load variations are confined to their corresponding scenario-specific bounds. The iteration terminates when the changes in the voltage bounds satisfy the prescribed convergence tolerance.

After convergence, the voltage magnitude at bus i can be expressed in affine form as follows:

$$\hat{V}_i = V_i^0 + \sum_{m=1}^M V_i^m \varepsilon_m\quad (18)$$

Accordingly, the lower and upper bounds of the nodal voltage magnitude are obtained as follows:

$$\begin{cases} \underline{V}_i = V_i^0 - \sum_{m=1}^M |V_i^m| \\ \overline{V}_i = V_i^0 + \sum_{m=1}^M |V_i^m| \end{cases}\quad (19)$$

This procedure yields affine representations of the bus voltages, including their central values and uncertainty ranges. Shared noise symbols retain the dependence among uncertain injections, and the resulting voltage intervals characterize the effects of intra-scenario source-load fluctuations. These intervals are subsequently used for voltage-risk evaluation in the planning model.

3 Distributed PV Planning Model Considering Scenario-Weighted Economic Cost and Affine Voltage Risk

Based on the representative source-load scenarios and affine power flow described in Section 2, the PV siting and sizing problem is formulated by considering affine voltage risk. PV installation locations and capacities affect the distribution of active and reactive power injections in the network and thus influence nodal voltages under source-load uncertainty. Accordingly, the planning objective includes the scenario-weighted economic cost under representative source-load scenarios and a voltage risk penalty evaluated using scenario-level affine voltage intervals.

3.1 Planning Variables and Scenario-Based Scheme Evaluation

The planning decision consists of two parts: the siting state and the installed PV capacity at candidate buses. The siting variable is a binary variable indicating whether a PV system is installed at candidate bus i . It is defined as:

$$y = [y_1, y_2, \dots, y_N], y_i \in \{0, 1\} \quad (20)$$

where N is the number of candidate buses. $y_i = 1$ if PV is installed at candidate bus i ; otherwise, $y_i = 0$.

The sizing variable is a continuous variable representing the installed PV capacity at candidate bus i . It is defined as:

$$s = [s_1, s_2, \dots, s_N], s_i \in [s_i^{\min}, s_i^{\max}] \quad (21)$$

where $s_i^{\min}$ and $s_i^{\max}$ are the minimum and maximum capacity at bus i .

For a candidate planning scheme $u = \{y_i, s_i\}$, the scenario-weighted economic cost and scenario-level affine voltage risk indicators are calculated based on the scenario-dependent affine representation established in Section 2. The obtained scenario-aggregated objective value is then used as fitness information for scheme comparison and subsequent particle updating.

3.2 Objective Function and Operating Constraints

The objective function of the proposed model consists of the scenario-weighted economic cost and the scenario-weighted affine voltage-boundary risk penalty. The economic cost evaluates the investment and operating performance of the PV planning scheme across representative source-load scenarios, while the affine voltage boundary risk penalty quantifies both central voltage violations and interval-boundary violations caused by intra-scenario source-load fluctuations.

3.2.1 Economic Cost Model

The economic cost mainly includes annualized PV investment cost, operation and maintenance (O&M) cost, scenario-dependent grid power transaction cost, and scenario-dependent network loss cost.

The annualized PV investment cost, which is independent of operating scenarios, is given by:

$$C_{inv} = (c_{num} \cdot \sum_{i=1}^N y_i + c_{cap} \cdot \sum_{i=1}^N s_i) \cdot \frac{r(1+r)^{T_{life}}}{(1+r)^{T_{life}} - 1} \quad (22)$$

where c_{num} and c_{cap} are the fixed installation cost coefficient and unit capacity investment cost coefficient, respectively; r is the discount rate; T_{life} is the project lifetime.

The annualized PV O&M cost of the installed PV capacity is expressed as:

$$C_{om} = c_{om} \cdot \sum_{i=1}^N s_i \quad (23)$$

where c_{om} is the unit O&M cost coefficient per capacity.

For each representative scenario, the grid power transaction cost is calculated according to the power exchanged between the distribution network and the upstream grid:

$$C_{grid,k} = \sum_{t=1}^T (c_b P_{buy,t}^k - c_s P_{sell,t}^k) \quad (24)$$

where c_b and c_s are the electricity purchase and selling prices, respectively; $P_{buy,t}^k$ and $P_{sell,t}^k$ are the purchased and sold power under representative scenario k at time t .

For each representative scenario, the network loss cost is expressed as:

$$\begin{cases} C_{loss,k} = c_e \cdot \sum_{t=1}^T L_{t,k} \\ L_{t,k} = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n G_g (V_i^2 + V_j^2 - 2V_i V_j \cos \theta_g) \end{cases} \quad (25)$$

where $L_{t,k}$ is the system loss under representative scenario k at time t ; c_e is the electricity price; n is the number of system buses.

Thus, the scenario-weighted total economic cost is expressed as:

$$C_{eco} = C_{inv} + C_{om} + \sum_{k=1}^K \pi_k (C_{grid,k} + C_{loss,k}) \quad (26)$$

Although the scenario-weighted economic cost reflects the comprehensive operating benefit of a planning scheme under representative source-load scenarios, it cannot directly characterize the voltage-boundary risk caused by intra-scenario PV and load fluctuations. Therefore, a scenario-dependent affine voltage-boundary risk index is further introduced.

3.2.2 Affine Voltage Risk Index

According to the scenario-based affine power flow propagation described in Section 2, the voltage magnitude at each bus, time period, and representative scenario can be expressed in affine form. To evaluate voltage security in the planning model, the voltage risk of each representative scenario is divided into central voltage risk and interval voltage risk. The former reflects voltage violations of the affine central

values under the scenario center, whereas the latter captures possible voltage-boundary violations caused by intra-scenario source-load fluctuations. For representative scenario k , the central voltage risk is defined as:

$$R_{V,k}^C = \sum_t \sum_i \left[\max(0, V_{\min} - V_{i,k}^0)^2 + \max(0, V_{i,k}^0 - V_{\max})^2 \right] \quad (27)$$

where $V_{\min}$ and $V_{\max}$ are the lower and upper voltage limits, respectively.

For representative scenario k , the affine interval voltage risk is defined as:

$$R_{V,k}^I = \sum_t \sum_i \left[\max(0, V_{\min} - \underline{V}_{i,k})^2 + \max(0, \bar{V}_{i,k} - V_{\max})^2 \right] \quad (28)$$

where the lower bound violation term reflects the potential undervoltage risk, and the upper bound violation term reflects the potential overvoltage risk caused by intra-scenario source-load fluctuations.

The comprehensive voltage risk of representative scenario k is then formulated as:

$$R_{V,k} = \alpha R_{V,k}^C + (1 - \alpha) R_{V,k}^I \quad (29)$$

where α is the weighting coefficient between the central voltage risk and the affine interval voltage risk within each representative scenario.

The voltage risk penalty is then expressed by:

$$\begin{cases} R_V = \sum_{k=1}^K \pi_k R_{V,k} \\ C_V = \lambda_V R_V \end{cases} \quad (30)$$

where λ_V is the voltage risk penalty coefficient. This risk term converts the scenario-weighted affine voltage risk into a planning penalty and is incorporated into the overall objective function.

3.2.3 Planning Constraints and Overall Objective Function

The number of PV installation buses is limited by:

$$\sum_{i=1}^N y_i \leq N_{\max} \quad (31)$$

where $N_{\max}$ is the maximum allowed number of PV installation buses.

The total installed PV capacity at each candidate bus and the total installed PV capacity are constrained by:

$$\begin{cases} y_i s_i^{\min} \leq s_i \leq y_i s_i^{\max} \\ \sum_{i=1}^N s_i \leq s_{MAX} \end{cases} \quad (32)$$

where s_{MAX} is the maximum total installed PV capacity.

A constraint penalty term is introduced to penalize infeasible solutions during the optimization process:

$$C_{pen} = \lambda_S \Delta S + \lambda_N \Delta N \quad (33)$$

where ΔS is the violation amount of the capacity constraints, ΔN is the violation amount of the maximum site number constraint; λ_S and λ_N are the corresponding penalty coefficients.

By combining the scenario-weighted economic cost, the scenario-weighted affine voltage risk penalty, and the constraint penalty, the overall objective function is formulated as:

$$\min F = C_{eco} + C_V + C_{pen} \quad (34)$$

3.3 PSO-Based Solution Method

The proposed planning model involves binary siting variables, continuous sizing variables, and nonlinear voltage risk evaluation based on affine power flow, which together form a mixed-integer nonlinear optimization problem. In the solution process, each particle represents a candidate PV siting and sizing scheme. The scenario-based affine evaluation process calculates the scenario-weighted economic cost, scenario-aggregated affine voltage risk, and constraint penalties of each candidate scheme, which are then combined into the fitness value. PSO is therefore adopted because its particle-based encoding can accommodate mixed discrete-continuous decision variables, and its fitness-driven update mechanism is suitable for the embedded affine evaluation process [13,15].

3.3.1 Particle Encoding and Update Strategy

Each particle represents a planning scheme. Its position vector consists of discrete siting variable y_i and continuous sizing variables s_i .

Continuous variables are updated using the standard PSO rule:

$$\begin{cases} v_i^{t+1} = w \cdot v_i^t + c_1 \cdot r_1 \cdot (p_{best,i} - x_i^t) + c_2 \cdot r_2 \cdot (g_{best} - x_i^t) \\ x_i^{t+1} = x_i^t + v_i^{t+1} \end{cases} \quad (35)$$

where v_i^t and v_i^{t+1} are the velocities of particle i at iterations t and $t + 1$; x_i^t is its current position; $p_{best,i}$ is its historical best position; g_{best} is the global best position.

Discrete variables are updated by mapping the velocity to a probability via a sigmoid function:

$$\begin{aligned} x_i^{t+1} &= \begin{cases} 1 & \text{if } rand() < \sigma(v_i^{t+1}) \\ 0 & \text{otherwise} \end{cases} \\ \sigma(v_i^{t+1}) &= \frac{1}{1 + e^{-v_i^{t+1}}} \end{aligned} \quad (36)$$

Inertia weight w is updated with the strategy below:

$$w(m) = w_{\max} - (w_{\max} - w_{\min}) \cdot \left(\frac{m}{M} \right) \quad (37)$$

where m is the current iteration count, M is the maximum iteration count, $w_{\max}$ and $w_{\min}$ are the maximum and minimum inertia weights.

The fitness function used in the PSO search is expressed as:

$$Fitness = F = C_{eco} + C_V + C_{pen} \quad (38)$$

3.3.2 Fitness Evaluation and Optimization Procedure

For each particle representing a candidate planning scheme, its fitness is evaluated through the following steps:

Step 1: Based on the representative source-load scenarios obtained in Section 2 and the decoded siting and sizing variables, the affine PV generation, load demand, and nodal net power injections are constructed for each scenario.

Step 2: Affine power flow is performed using these uncertain injections to obtain the central values and fluctuation ranges of the nodal voltages.

Step 3: The fitness value is calculated from the scenario-weighted economic cost, central voltage risk, affine interval voltage risk, and constraint penalties.

After all particles have been evaluated, their states and the best solutions are updated according to the calculated fitness values. The fitness evaluation and particle update are repeated until the convergence criterion is satisfied, after which the optimal PV configuration is obtained. This embedded evaluation enables PSO to compare candidate schemes by considering both multi-scenario economic performance and affine voltage risk under source-load uncertainty. The detailed solution procedure is illustrated in Fig. 2.

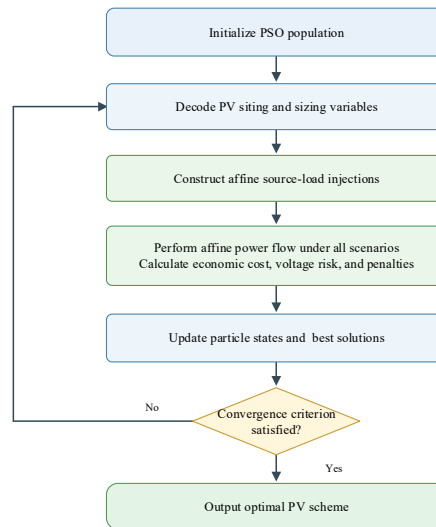


Figure 2: Detailed solution procedure of the proposed PV planning method.

4 Case Studies and Results

Case studies are conducted on the IEEE 69-bus distribution system and a practical 80-bus distribution network in Fujian Province, China. The two systems are selected to represent different voltage risk characteristics. The IEEE 69-bus system is used to examine the planning results in an undervoltage prone feeder, while the practical 80-bus system is used to evaluate the trade-off between economy and voltage risk under limited upper voltage margin.

4.1 Case Setup

4.1.1 Economic and Algorithmic Parameters

The investment cost of PV systems is converted into an annualized cost and incorporated into the objective function. The main economic and algorithmic parameters are listed in Table 1.

Table 1: Economic and algorithmic parameters.

Parameter	Value
Fixed installation cost	20×10^4 CNY/site
Unit capacity investment cost	350×10^4 CNY/MW
Annual O&M cost	5×10^4 CNY/(MW·year)
Purchasing electricity price	0.05×10^4 CNY/MWh
Selling electricity price	0.01×10^4 CNY/MWh
PV lifetime	20 years
Discount rate	6%
Allowable voltage range	0.95–1.05 p.u.
Population size	70
Maximum stagnation iteration	15
Central-interval voltage risk weight, α	0.4
Affine voltage risk penalty coefficient, λ_V	15

4.1.2 DPMM-Based Representative Source-Load Scenarios

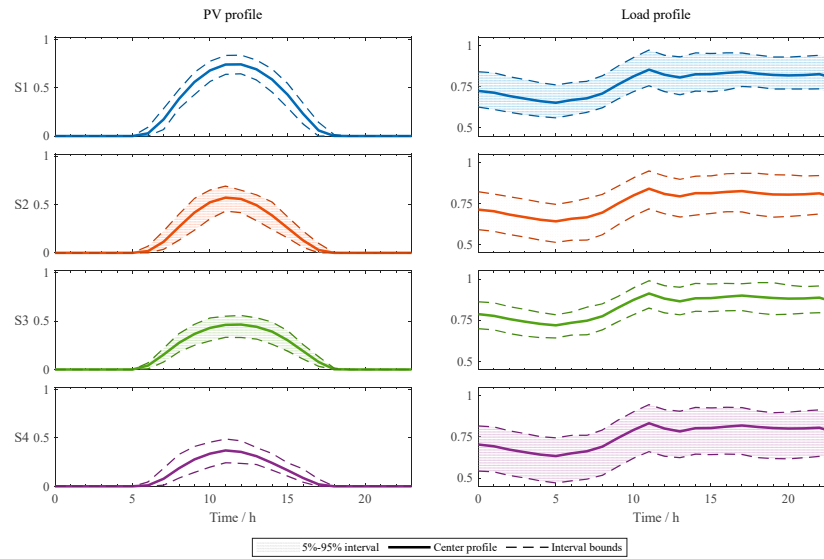
The source-load uncertainty data used in the case studies are constructed from hourly PV generation and load demand records collected from a region in Fujian Province, China [36,37]. The PV data are normalized according to the rated PV capacity, while the load data are converted into normalized load coefficients. The annual hourly data are organized into daily PV-load samples, where each sample consists of a 24-h PV output profile and the corresponding 24-h load demand profile.

DPMM is first applied to the joint PV-load samples to identify latent operating components without specifying the number of components in advance. For the reported case studies, the concentration parameter was initialized according to the number of daily samples and subsequently updated during sampling. The DPMM was run for 500 iterations, and the component assignments from the final iteration were retained as the raw DPMM result. The use of joint samples preserves the temporal correspondence between PV generation and load demand on the same day and the identified components therefore provide a data-driven basis for constructing representative planning scenarios. Since the raw DPMM components may include low-probability groups or highly similar PV-load patterns, they are further organized before being used in affine calculation. The organization process considers both sample size and feature-space similarity, allowing components with close source-load characteristics to be represented by the same planning scenario. Unlike K-means clustering, where the final number of clusters is specified in advance, this procedure first identifies latent source-load components and then constructs the final scenario set according to their statistical support and similarity. After this process, four representative source-load scenarios are obtained. For each scenario, the central profile and the 5%–95% empirical quantile interval are recalculated from the original 24-h PV and load profiles assigned to that scenario.

The representative source-load scenarios used for planning are summarized in Table 2. The corresponding scenario profiles and 5%–95% fluctuation intervals are shown in Fig. 3. In the figure, the solid lines denote the scenario centers, while the shaded regions and dashed lines represent the intra-scenario fluctuation intervals. The obtained scenario centers, occurrence probabilities, and fluctuation bounds are used as the uncertainty inputs for the subsequent affine evaluation.

Table 2: Representative source-load scenarios obtained by DPMM-based modeling.

Scenario	Days	Probability	Operating Pattern
S1	223	0.611	High PV generation and moderate load demand
S2	75	0.205	Moderate PV generation and moderate load demand
S3	27	0.074	Moderate PV generation and high load demand
S4	40	0.110	Low PV generation and low load demand

**Figure 3:** Representative source-load profiles and fluctuation intervals.

4.1.3 Comparative Schemes and Monte Carlo Validation

Four planning schemes are considered for comparison:

Case 1: No-PV baseline. No new PV is installed, and this case is used to represent the original operating condition of the system.

Case 2: Deterministic planning. PV siting and sizing are optimized using the scenario central profiles and their occurrence probabilities, without considering intra-scenario PV and load fluctuations.

Case 3: Robust planning. PV siting and sizing are optimized using conservative combinations of the upper and lower bounds of PV generation and load demand within each representative scenario.

Case 4: Affine-based planning. PV siting and sizing are optimized using the proposed scenario-weighted economic and affine voltage-risk-aware formulation.

The four cases are designed to distinguish the original no-PV condition, deterministic planning based on scenario central profiles, conservative bound-based uncertainty treatment, and affine voltage-risk evaluation. Cases 2–4 use the same representative source-load data, decision variables, economic model, technical constraints, candidate buses, and PSO settings, and differ in how intra-scenario uncertainty is treated during candidate-scheme evaluation.

Since the compared planning schemes adopt different uncertainty treatments, their optimization objective values are not directly comparable. Therefore, a unified annual Monte Carlo validation is performed after planning. For each scheme, the obtained PV siting and sizing result is fixed, and annual source-load samples are generated according to the representative scenario probabilities and intra-scenario fluctuation ranges. The same validation samples are then used to calculate the average annual cost, network loss cost,

voltage quantiles, and voltage violation depth and violation rate. In this way, the long-term operating performance of different schemes can be compared under a consistent uncertainty setting.

In the Monte Carlo validation, voltage violation depth and violation rate are used to quantify voltage security. The violation rate is calculated as the proportion of validation samples in which the voltage exceeds the allowable limit. The violation depth is defined as follows:

$$\begin{aligned} D_{\text{under}} &= \sum_t \sum_i \max(0, V_{\min}^{\text{lim}} - V_{i,t}) \\ D_{\text{over}} &= \sum_t \sum_i \max(0, V_{i,t} - V_{\max}^{\text{lim}}) \end{aligned} \quad (39)$$

where D_{under} and D_{over} denote the cumulative undervoltage and overvoltage depths over all buses and time periods for one validation sample, respectively. The reported Monte Carlo results are averaged over all validation samples.

4.2 Results on the IEEE 69-Bus System

4.2.1 System Description and Parameter Settings

The IEEE 69-bus distribution system is a typical radial distribution network, shown in Fig. 4. The base power is 10 MVA, the total active load is 3.8021 MW, and bus 1 is the slack bus. The original system has relatively low voltage at downstream buses. Therefore, this case is used to examine the behavior of different planning schemes in an undervoltage prone feeder, where PV integration tends to improve both economic performance and voltage level. In this case, buses 11, 21, 27, 50, 61, 65, and 68 are selected as PV candidate buses. The total PV capacity limit is set to 3.5 MW.

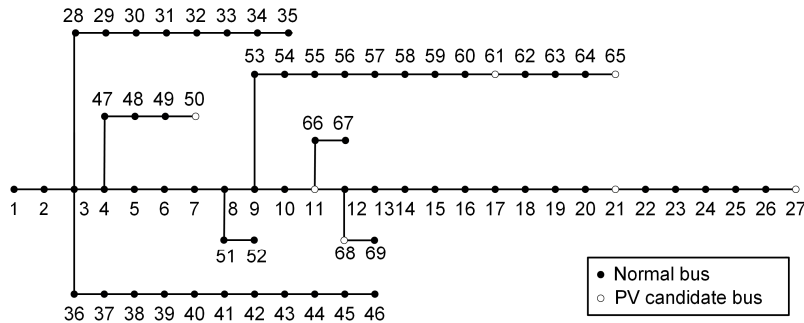


Figure 4: Topology of the IEEE 69-bus distribution system with PV candidate buses.

4.2.2 Planning Results

The planning results obtained by different methods are listed in Table 3.

Table 3: Planning results of the IEEE 69-bus system.

Scheme	Total Capacity/MW	Number of Sites	Annualized PV Cost/ 10^4 CNY	PV Location and Capacity/MW
Case 1: No-PV	0.00	0	0	-
Case 2: Deterministic	3.5	3	129.532	21:0.687; 61:1.875; 65:0.938
Case 3: Robust	3.5	2	127.789	61:1.770; 65:1.730
Case 4: Affine-based	3.5	2	127.789	61:1.924; 65:1.576

All three PV planning schemes reach the total capacity limit of 3.50 MW, indicating that additional PV capacity is favored in this undervoltage prone system. However, the siting patterns are not identical. The deterministic scheme selects buses 21, 61, and 65, whereas the robust and affine-based schemes concentrate the installations at buses 61 and 65. This indicates that the treatment of source-load uncertainty affects the spatial allocation of PV units to some extent.

The similarity between the robust and affine-based schemes is related to the operating characteristics of the IEEE 69-bus system. Since the dominant issue is insufficient downstream voltage support, PV generation at downstream buses can reduce upstream power purchase, decrease network losses, and raise voltage levels simultaneously. As a result, the economic benefit and undervoltage mitigation effect have consistent directions, and the operating performance differences among the PV schemes are expected to be limited in the subsequent validation.

4.2.3 Monte Carlo Validation and Analysis

To compare the long-term operating performance of different planning schemes under the same uncertainty setting, all schemes are tested using a unified annual Monte Carlo validation process. The validation results are listed in Table 4.

Table 4: Monte Carlo validation results of the IEEE 69-bus system.

Scheme	Average Annual Cost/ 10^4 CNY	P5 Minimum Voltage/p.u.	Undervoltage Depth	Undervoltage Violation Rate/%
Case 1: No-PV	1338.612	0.916	529.643	97.922
Case 2: Deterministic	1158.223	0.932	57.149	21.124
Case 3: Robust	1158.981	0.934	49.649	18.371
Case 4: Affine-based	1158.743	0.934	49.784	18.454

Table 4 shows that the no-PV case exhibits a clear undervoltage problem, with a P5 minimum voltage of 0.916 p.u., an undervoltage depth of 529.643, and an undervoltage violation rate of 97.922%. After PV integration, all three planning schemes reduce the average annual cost and substantially decrease the undervoltage depth. This indicates that the IEEE 69-bus system is an undervoltage prone case in which PV integration provides both economic benefit and voltage support.

The performance differences among the three PV planning schemes are relatively limited. The deterministic scheme retains a slight economic advantage, whereas the robust and affine-based schemes provide modest improvements in lower-tail voltage performance. The cost-undervoltage trade-off is weak because PV integration improves economic performance and alleviates undervoltage in the same direction.

Fig. 5 compares the minimum voltage quantiles of different planning schemes. Compared with the no-PV case, the voltage quantiles of all PV planning schemes shift upward, indicating improved low-voltage performance after PV integration. However, the quantile differences among the deterministic, robust, and affine-based schemes are small. The robust and affine-based schemes show slightly better lower-tail voltage performance, which is consistent with their lower undervoltage depths in Table 4.

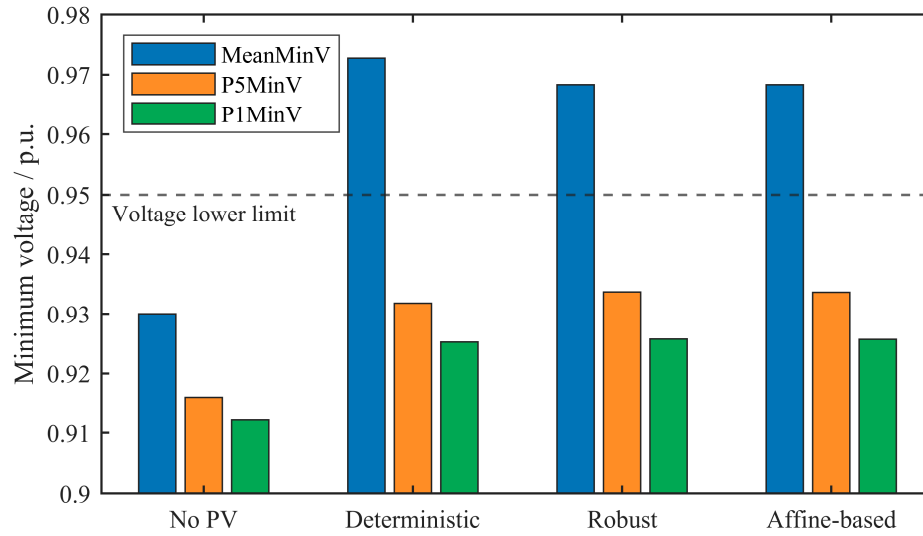


Figure 5: Minimum-voltage quantile comparison on the IEEE 69-bus system.

4.3 Results on the Practical 80-Bus Distribution System

4.3.1 System Description and Parameter Settings

To further examine the performance of different planning schemes under limited upper-voltage margin and overvoltage risk, a practical 10 kV 80-bus distribution network from a city in Fujian Province, China, is used for the second case study. The network topology and PV candidate buses are shown in Fig. 6. The system has a base power of 10 MVA, a total active load of 1.4197 MW, and bus 1 is the slack bus. Without new PV integration, the maximum voltage of the system is 1.0250 p.u., which does not exceed the voltage limit but indicates a limited upper voltage margin. Buses 12, 35, 36, 40, 60, 73, 77, and 79 are selected as PV candidate buses. The total PV capacity limit is set to 3.00 MW, and the maximum number of PV installation sites is set to 4.

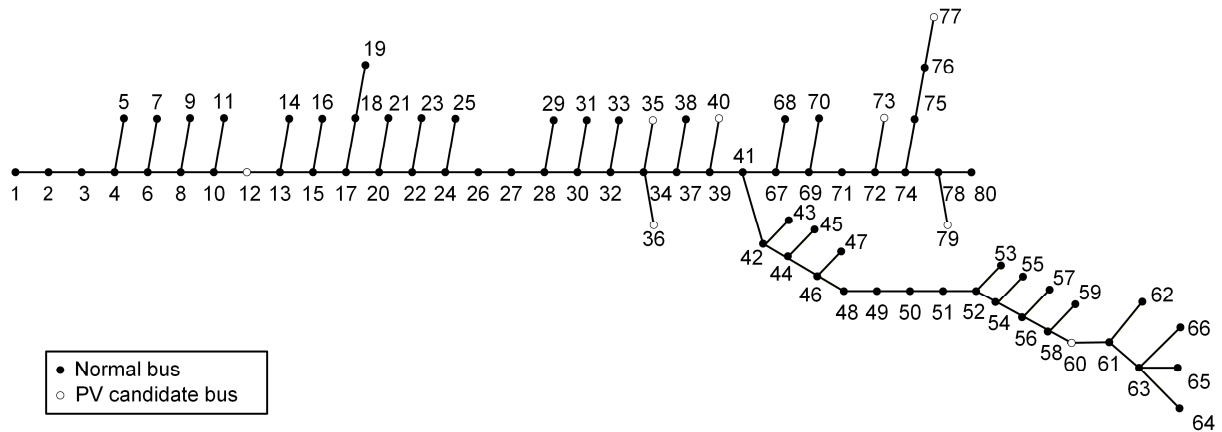


Figure 6: Topology of the practical 80-bus distribution network with PV candidate buses.

4.3.2 Planning Results

The planning results of the practical 80-bus distribution system are shown in Table 5.

Table 5: Planning results of the practical 80-bus system.

Scheme	Total Capacity/MW	Number of Sites	Annualized PV Cost/ 10^4 CNY	PV Location and Capacity/MW
Case 1: No-PV	0.00	0	0.00	—
Case 2: Deterministic	2.55	3	95.793	35:0.750; 60:0.850; 77:0.950
Case 3: Robust	2.30	4	88.658	12:0.250; 35:0.750; 60:0.850; 79:0.450
Case 4: Affine-based	2.338	3	88.250	35:0.750; 60:0.638; 77:0.950

In contrast to the common capacity limit reached in the IEEE 69-bus system, the three planning methods produce different total capacities and siting patterns in the practical 80-bus system. The deterministic scheme installs 2.55 MW of PV at buses 35, 60, and 77, which is the largest total capacity among the three PV schemes.

The robust scheme reduces the total installed PV capacity to 2.30 MW and distributes PV units over four buses, namely buses 12, 35, 60, and 79. This indicates a more conservative allocation strategy under source-load uncertainty. The affine-based scheme installs 2.338 MW at buses 35, 60, and 77. Compared with the deterministic scheme, it keeps the main installation buses but reduces the capacity at bus 60. These differences show that the treatment of intra-scenario uncertainty affects both the total installed capacity and the spatial allocation of PV units. Their long-term economic and voltage-security performance is examined below using the same Monte Carlo samples.

4.3.3 Monte Carlo Validation and Analysis

The practical 80-bus system is further evaluated using the unified annual Monte Carlo validation procedure described in Section 4.1.3. The validation results are listed in Table 6.

Table 6: Monte Carlo validation results of the practical 80-bus system.

Scheme	Average Annual Cost/ 10^4 CNY	P95 Maximum Voltage/p.u.	Overvoltage Depth	Overvoltage Violation Rate/%
Case 1: No-PV	484.269	1.025	0	0
Case 2: Deterministic	400.353	1.050	4.300	4.784
Case 3: Robust	402.774	1.044	0.013	0.107
Case 4: Affine-based	400.836	1.044	0.003	0.086

Without PV integration, the system has an average annual cost of 484.269×10^4 CNY, and no overvoltage occurs. After PV integration, all three planning schemes reduce the average annual cost because local PV generation decreases upstream power purchase and network losses. However, unlike the IEEE 69-bus system, the economic benefit of PV integration is accompanied by different levels of overvoltage risk.

Table 6 reveals a clear cost-overvoltage trade-off among the three PV planning schemes. The deterministic scheme achieves the lowest annual cost but exhibits the highest upper-tail voltage risk, indicating that evaluation based only on scenario central profiles may underestimate the effect of intra-scenario source-load fluctuations. Both the robust and affine-based schemes substantially reduce this risk. Compared with the robust scheme, the affine-based scheme maintains similar voltage-security performance while retaining an annual cost closer to that of deterministic planning.

Fig. 7 compares the maximum voltage quantiles of different planning schemes. The deterministic scheme has the highest upper tail voltage level, and its P99 maximum voltage exceeds the upper voltage

limit. By contrast, the robust and affine-based schemes reduce the upper tail voltage quantiles, indicating better voltage security performance under uncertain operating conditions.

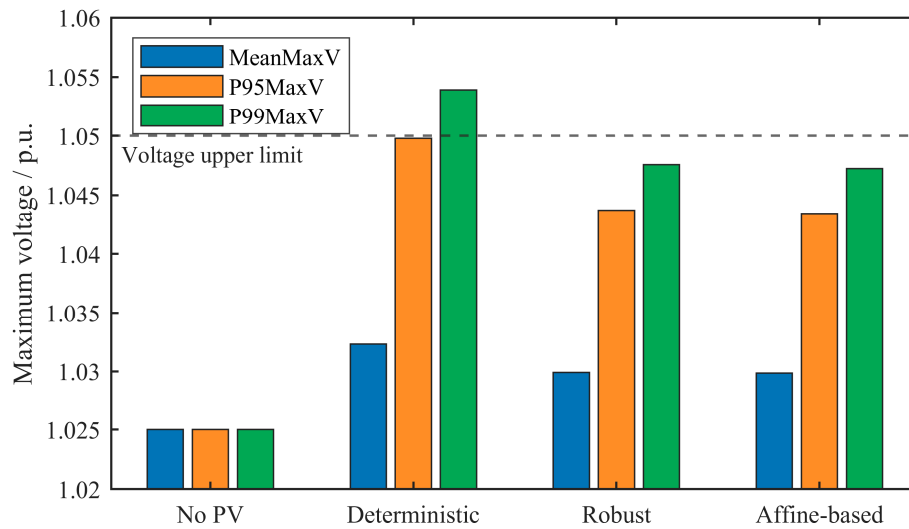


Figure 7: Maximum-voltage quantile comparison on the practical 80-bus system.

Whereas Table 6 and Fig. 7 present system-level voltage statistics, Fig. 8 shows the spatial and temporal distributions of the node-hour P95 voltage for the four cases. For each case, the P95 voltage is calculated at every bus and PV-active hour from annual Monte Carlo realizations generated under the same uncertainty setting. A common color scale and a 1.05 p.u. reference plane are used to facilitate comparison.

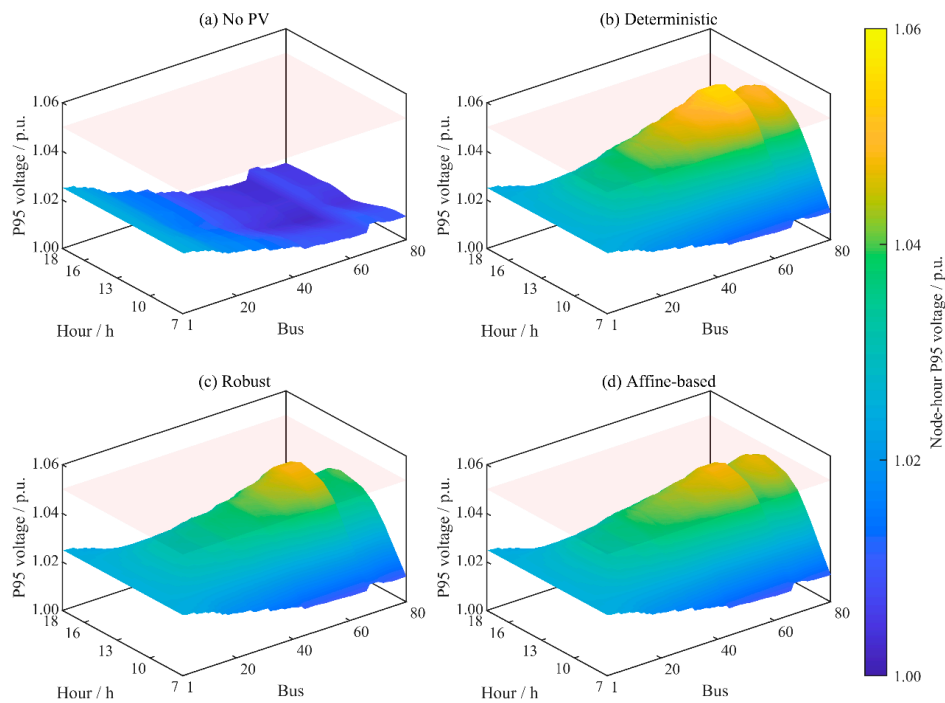


Figure 8: Node-hour P95 voltage surfaces for different planning cases on the practical 80-bus distribution network.

The no-PV case remains below the upper voltage limit. Deterministic planning produces the most pronounced voltage rise and the widest high-voltage region during PV-active hours, whereas the robust and affine-based schemes reduce both the magnitude and the spatiotemporal extent of the voltage rise. The node-hour voltage distributions are consistent with the system-level indicators in Table 6 and the voltage-quantile comparison in Fig. 7.

4.4 Discussion

Overall, the two case studies indicate that the benefit of uncertainty-aware PV planning depends on the voltage characteristics of the distribution network. In the IEEE 69-bus system, PV integration improves both economic performance and undervoltage conditions, resulting in limited differences among the planning schemes. In the practical 80-bus network, PV integration introduces a clearer trade-off between annual cost and overvoltage risk.

The proposed framework is particularly relevant to feeders where PV integration creates a non-negligible overvoltage risk. By combining probability-weighted operating scenarios with affine propagation of intra-scenario fluctuations, it allows the trade-off between economic performance and voltage security to be considered without repeated Monte Carlo sampling during each optimization iteration.

The scenario construction depends on the representativeness of the available PV and load records and may require updating as source-load patterns evolve. The current formulation focuses on annual economic performance and steady-state voltage security. Future work will extend the framework to multi-year source-load data, coordinated planning with flexible resources, and additional engineering constraints.

5 Conclusions

This paper develops an affine voltage risk-aware distributed PV planning method under source-load uncertainty. Historical PV and load data are organized into daily joint samples and DPMM is used to identify latent source-load operating components for representative scenario construction. The scenario probabilities are used to aggregate scenario-dependent economic cost and voltage risk, while the scenario central profiles and intra-scenario fluctuation bounds are converted into affine source-load injections for voltage interval evaluation. The main conclusions are as follows:

- (1) The DPMM-based scenario construction process provides a data-driven description of source-load operating patterns. By using daily joint PV-load samples, the method preserves the correspondence between PV generation and load demand. The representative scenarios obtained after scenario organization provide occurrence probabilities for scenario-weighted evaluation and central profiles with fluctuation bounds for affine uncertainty propagation.
- (2) The affine voltage risk evaluation links intra-scenario source-load fluctuations with voltage-boundary violation assessment. By converting scenario centers and fluctuation bounds into affine injections, the planning model can consider both central voltage violations and affine interval-boundary violations.
- (3) The case studies indicate that the effect of uncertainty-aware planning is related to the voltage risk characteristics of the distribution system. In the IEEE 69-bus system, the trade-off between economy and voltage risk is relatively limited because PV integration reduces annual cost and alleviates undervoltage simultaneously. In the practical 80-bus system, the affine-based scheme reduces overvoltage risk compared with deterministic planning while maintaining a comparable annual cost, providing a more balanced trade-off between economic performance and voltage-risk mitigation. From a practical perspective, the proposed framework provides utility planners with a consistent basis for comparing

candidate PV configurations and selecting siting and sizing schemes that balance annual cost and voltage-security performance as PV penetration increases.

Acknowledgement: Not applicable.

Funding Statement: This work was funded by the Science and Technology Project of State Grid Fujian Electric Power Co., Ltd., grant number 52136024000P.

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization, Yan Zhang; software, Yusi Xu; investigation, Jingmiao Huang; data curation, Quanmin Chen; writing—original draft preparation, Bin Lin; writing—review and editing, Sihui Ke; project administration, Bin Lin; funding acquisition, Jingmiao Huang. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: Due to institutional and engineering confidentiality restrictions, the practical distribution network data are not publicly available. The PV data and provincial load data are openly available from the cited public data sources. The main simulation settings and parameters are provided within the article. Additional data that support the findings of this study are available from the corresponding author upon reasonable request and with permission from the data provider.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare that this study received funding from State Grid Fujian Electric Power Co., Ltd., grant number 52136024000P. Authors Bin Lin, Jingmiao Huang, Quanmin Chen, and Sihui Ke are employed by State Grid Longyan Power Supply Company of Fujian Electric Power Co., Ltd. The funder had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results. All authors declare that the research was conducted in the absence of any other commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

PV	Photovoltaic
DPMM	Dirichlet process mixture model
PSO	Particle swarm optimization
O&M	Operation and maintenance

References

1. Alwez MA, Jasni J, MohdRadzi MA, Azis N. Adaptive reactive power control for voltage rise mitigation on distribution network with high photovoltaic penetration. *Renew Sustain Energy Rev.* 2025;207:114948. [[CrossRef](#)].
2. Zhou L, Shao Z, Chen F, Cheng G, Lin C, Chen Z, et al. A physics-informed topology-adaptive graph convolutional network for harmonic source location in three-phase distribution networks with renewable energy integration. *IEEE Trans Power Deliv.* 2026;1–17. [[CrossRef](#)].
3. Zhou L, Shao Z, Cheng G, Lin J, Chen F. Interval harmonic state estimation of three-phase asymmetric distribution network by integrating multi-source data. *Prot Control Mod Power Syst.* 2026;11(2):62–76. [[CrossRef](#)].
4. Diahovchenko I, Morva G, Chuprun A, Keane A. Comparison of voltage rise mitigation strategies for distribution networks with high photovoltaic penetration. *Renew Sustain Energy Rev.* 2025;212:115399. [[CrossRef](#)].
5. Majeed IB, Nwulu NI. Reverse power flow due to solar photovoltaic in the low voltage network. *IEEE Access.* 2023;11:44741–58. [[CrossRef](#)].
6. Li H, Li L. Bilevel planning of distribution networks with distributed generation and energy storage: A case study on the modified IEEE 33-bus system. *Energy Eng.* 2025;122(4):1337–58. [[CrossRef](#)].

7. Fu C, Wang X, Li Z, Liu X, Zhang X, Zhao Z. Optimal evaluation of photovoltaic consumption schemes in distribution networks based on BASS model for photovoltaic installed capacity prediction. *Energy Eng.* 2025;122(5):1805–21. [[CrossRef](#)].
8. Georgilakis PS, Hatziaargyriou ND. Optimal distributed generation placement in power distribution networks: Models, methods, and future research. *IEEE Trans Power Syst.* 2013;28(3):3420–8. [[CrossRef](#)].
9. Dorji P, Lachowicz S, Bass O. Optimization of size and siting of distributed generation in unbalanced distribution systems: A literature review. *Electr Power Syst Res.* 2025;249:112039. [[CrossRef](#)].
10. Hung DQ, Mithulanathan N, Bansal RC. Integration of PV and BES units in commercial distribution systems considering energy loss and voltage stability. *Appl Energy.* 2014;113:1162–70. [[CrossRef](#)].
11. Fotopoulou M, Tsekouras GJ, Rakopoulos DC, Kanellos FD, Kontargyri VT. Investment planning optimization for energy communities with photovoltaic and storage. *IEEE Access.* 2026;14:98275–94. [[CrossRef](#)].
12. Pushkarna M, Ashfaq H, Singh R, Kumar R. A new analytical method for optimal sizing and siting of type-IV DG in an unbalanced distribution system considering power loss minimization. *J Electr Eng Technol.* 2022;17(5):2579–90. [[CrossRef](#)].
13. Pushkarna M, Ashfaq H, Singh R, Kumar R. An optimal placement and sizing of type-IV DG with reactive power support using UPQC in an unbalanced distribution system using particle swarm optimization. *Energy Syst.* 2024;15(1):353–70. [[CrossRef](#)].
14. Maji S, Kayal P. A simplified multi-objective planning approach for allocation of distributed PV generators in unbalanced power distribution systems. *Renew Energy Focus.* 2024;48:100541. [[CrossRef](#)].
15. Dong A, Lee SK. Research on a distributed photovoltaic two-level planning method based on the SCMPSO algorithm. *Energies.* 2024;17(13):3251. [[CrossRef](#)].
16. Varathan G, Belwin Edward J. A review of uncertainty management approaches for active distribution system planning. *Renew Sustain Energy Rev.* 2024;205:114808. [[CrossRef](#)].
17. Liu Z, Wen F, Ledwich G. Optimal siting and sizing of distributed generators in distribution systems considering uncertainties. *IEEE Trans Power Deliv.* 2011;26(4):2541–51. [[CrossRef](#)].
18. Kheirkhah AR, Meschini Almeida CF, Kagan N, Leite JB. Optimal probabilistic allocation of photovoltaic distributed generation: Proposing a scenario-based stochastic programming model. *Energies.* 2023;16(21):7261. [[CrossRef](#)].
19. Khasanov M, Kamel S, Halim Houssein E, Rahmann C, Hashim FA. Optimal allocation strategy of photovoltaic and wind turbine-based distributed generation units in radial distribution networks considering uncertainty. *Neural Comput Appl.* 2023;35(3):2883–908. [[CrossRef](#)].
20. Ayaz MS, Malekpour M, Azizipanah-Abarghooee R, Karimi M, Terzija V. Probabilistic photovoltaic generation and load demand uncertainties modelling for active distribution networks hosting capacity calculations. *Int J Electr Power Energy Syst.* 2024;159:110016. [[CrossRef](#)].
21. Lazo J, Watts D. Stochastic model for active distribution networks planning: An analysis of the combination of active network management schemes. *Renew Sustain Energy Rev.* 2024;191:114156. [[CrossRef](#)].
22. Barutcu IC. Chance-constrained optimization of photovoltaic system allocation considering power loss, voltage level, and line current. *Int Trans Electr Energy Syst.* 2023;2023:5587904. [[CrossRef](#)].
23. Zhu J, Huang Y, Lu S, Shen M, Yuan Y. Incorporating local uncertainty management into distribution system planning: An adaptive robust optimization approach. *Appl Energy.* 2024;363:123103. [[CrossRef](#)].
24. Prabawa P, Choi DH. Distributionally robust PV planning and curtailment considering cyber attacks on electric vehicle charging under PV/load uncertainties. *Energy Rep.* 2024;11:3436–49. [[CrossRef](#)].
25. Vaccaro A, Cañizares CA. An affine arithmetic-based framework for uncertain power flow and optimal power flow studies. *IEEE Trans Power Syst.* 2017;32(1):274–88. [[CrossRef](#)].
26. Li Q, Sun H, Liu M, Li J, Ji W, Wang K. Complex affine arithmetic based uncertain sensitivity analysis of voltage fluctuations in active distribution networks. *Front Energy Res.* 2024;12:1374986. [[CrossRef](#)].
27. Gao C, Luo Q, Chen P, Xu Z, Weng X, Liu X, et al. Generation of typical scenarios for distribution networks in planning stage considering photovoltaic and load growth characteristics. *Front Energy Res.* 2025;13:1498656. [[CrossRef](#)].

28. Li J, Wang S, Li Y, Liu H. Photovoltaic generation scenario reduction method based on improved K-means clustering. In: *Proceedings of the International Conference on Information Control, Electrical Engineering and Rail Transit*. Singapore: Springer; 2023. p. 1–16. [[CrossRef](#)].
29. Antoniak CE. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann Stat*. 1974;2(6):1152–74. [[CrossRef](#)].
30. Granell R, Axon CJ, Wallom DCH. Clustering disaggregated load profiles using a Dirichlet process mixture model. *Energy Convers Manag*. 2015;92:507–16. [[CrossRef](#)].
31. Frimane Â, Soubdhan T, Bright JM, Aggour M. Nonparametric Bayesian-based recognition of solar irradiance conditions: Application to the generation of high temporal resolution synthetic solar irradiance data. *Sol Energy*. 2019;182:462–79. [[CrossRef](#)].
32. Dai L, You D, Yin X. Data driven robust energy and reserve dispatch based on a nonparametric Dirichlet process Gaussian mixture model. *Energies*. 2020;13(18):4642. [[CrossRef](#)].
33. Huang Y, Chen Y, Shao Z, Chen F, Shao Y, Zhang Y, et al. Risk assessment of distribution network based on Dirichlet process mixture model and the cumulant method. *Inventions*. 2026;11(2):42. [[CrossRef](#)].
34. Comba J, Stolfi J. Affine arithmetic and its applications to computer graphics. In: *Proceedings of the Brazilian Symposium on Computer Graphics and Image*; 1993 Oct 19–22; Recife, Brazil. p. 9–18.
35. Adusumilli BS, Raj V, Boddeti KK. Modified affine arithmetic-based power flow analysis with uncertainty. *Electr Power Compon Syst*. 2018;46(6):728–37. [[CrossRef](#)].
36. Pfenninger S, Staffell I. Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data. *Energy*. 2016;114:1251–65. [[CrossRef](#)].
37. Wu H, Kan X. Hourly electric power load and transmission data at the provincial level in China. *Zenodo*. 2023. [[CrossRef](#)].