



OPEN ACCESS

EDITED BY

Ying Zhang,
Oklahoma State University, United States

REVIEWED BY

Zipeng Liang,
Hong Kong Polytechnic University, Hong Kong SAR, China
Haoyong Chen,
South China University of Technology, China
Yun Liu,
South China University of Technology, China
Ge Chen,
Purdue University, United States
Haris M. Khalid,
University of Dubai, United Arab Emirates

*CORRESPONDENCE

Huaying Zhang,
✉ zhytgyx@163.com
Qian Liu,
✉ liuqian365@hnu.edu.cn

RECEIVED 16 July 2024

ACCEPTED 17 September 2024

PUBLISHED 30 September 2024

CITATION

Liang X, Zhang H, Liu Q, Liu Z and Liu H (2024) A support vector regression-based interval power flow prediction method for distribution networks with DGs integration. *Front. Energy Res.* 12:1465604. doi: 10.3389/fenrg.2024.1465604

COPYRIGHT

© 2024 Liang, Zhang, Liu, Liu and Liu. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A support vector regression-based interval power flow prediction method for distribution networks with DGs integration

Xiaorui Liang¹, Huaying Zhang^{1*}, Qian Liu^{2*}, Zijun Liu¹ and Huicong Liu¹

¹New Smart City High-Quality Power Supply Joint Laboratory of China Southern Power Grid, Shenzhen Power Supply Co., Ltd., Shenzhen, Guangdong, China, ²College of Electrical and Information Engineering, Hunan University, Changsha, Hunan, China

In distribution networks with distributed generators (DGs), power generation and load demand exhibit increased randomness and volatility, and the line parameters also suffer more frequent fluctuations, which may result in significant state shifts. Existing model-driven methods face challenges in efficiently solving uncertain power flow, especially as the size of the system increases, making it difficult to meet the demand for rapid power flow analysis. To address these issues, this paper proposes an SVR-based interval power flow (IPF) prediction method for distribution networks with DGs integration. The method utilizes intervals to describe system uncertainty and employs Support Vector Regression (SVR) for model training. The input feature vector consists of the intervals of active power generation, load demand, and line parameters, while the output feature vector represents the intervals of voltage or line transmission power. Ultimately, the SVR-based IPF prediction model is established, capturing the linear mapping relationship between input data and output IPF variables. Simulation results demonstrate that the proposed method exhibits high prediction accuracy, strong adaptability, and optimal computation efficiency, meeting the requirements for rapid and real-time power flow analysis while considering the uncertainty in distribution networks with DGs integration.

KEYWORDS

data-driven method, interval power flow, support vector regression, distribution network, distributed generators

1 Introduction

1.1 Motivation

In the context of widespread integration of distributed generators (DGs) such as wind and photovoltaic (PV) power into distribution networks, power generation exhibits uncertainty due to the inherent volatility and randomness of wind and solar. In addition, load demand and line parameters also exhibit uncertainty which is caused by user consumption behaviors and environmental factors, respectively. These issues caused the power flow state in the system to undergo rapid and intricate changes. Considering these uncertainties, uncertain power flow (PF) methods are proposed by researchers. However,

most existing uncertain PF methods are model-driven. As the system scale increases, the model complexity grows, leading to a significant reduction in computational efficiency, which fails to meet the requirements for rapid assessment of system states in distribution networks. Improving the computational efficiency of uncertain PF analysis can provide assurance for real-time monitoring and dispatching of distribution systems, ensuring stable and efficient operation. There is an urgent need for efficient and rapid methods for uncertain PF analysis in distribution networks that can effectively address system uncertainty.

1.2 Focus and potential

This paper focuses on addressing the computational efficiency issues of uncertain PF, primarily in two aspects: describing system uncertainty using intervals and employing data-driven methods for PF prediction, enabling real-time interval power flow (IPF) calculations in distribution systems. The potential of this research lies in its ability to significantly enhance the real-time monitoring and operational capabilities of distribution networks with integrated DGs. By addressing the limitations of existing model-driven uncertain PF methods, the proposed approach could lead to more efficient PF analysis, particularly in the face of the increasing penetration of renewable energy sources (RES). This has offered a scalable solution for real-time PF analysis in increasingly complex and uncertain environments.

1.3 Preceding research

Commonly used methods for handling uncertainty currently include robust, probabilistic, and interval methods. Among them, the robust method is mainly used for optimization (Zheng et al., 2024), such as energy management under the uncertainty of renewable energy generation and electric vehicles (EVs) (Tan et al., 2024). When calculating power flow, the probabilistic method and interval algorithm are more frequently employed, which are called probabilistic power flow (PPF) and interval power flow (IPF). IPF has the advantages of simple modelling and high security compared with PPF. Existing IPF methods primarily consist of iterative approaches (Mori and Yuihara, 1999; Barboza et al., 2004) and optimization techniques (Zhang et al., 2017; 2018; 2023). For iterative approaches, the Interval Newton iteration was first employed. To avoid solving the equations in the Interval Newton method, the Krawczyk method was introduced. The interval problem was broken down into multiple sub-intervals, and each sub-interval was solved iteratively using the Krawczyk method (Mori and Yuihara, 1999). The Interval Newton iteration framework was combined with the Krawczyk operator in (Barboza et al., 2004), enhancing convergence performance. The introduction of the Affine Algorithm (AA) (Vaccaro et al., 2010) increased the efficiency and accuracy of solving interval nonlinear equation systems. The convergence of the Krawczyk-Moore iteration was enhanced by introducing AA, and the correlation issues of interval computation were addressed. Optimization methods, which avoid iteration and convergence

problems, have gained widespread attention in recent years. The optimization model for the IPF solution was constructed by converting intervals into affine forms (Zhang et al., 2017), improving the efficiency of solving IPF. An optimization scenario method (OSM) was improved to solve IPF (Zhang et al., 2018), directly obtaining the range of power flow variables through the optimization models. In IPF analysis for distribution networks, the rise of AA has led to a trend of combining it with the Distflow model, including solving the affine Distflow model using forward-backward substitution (Cheng et al., 2023; Lyu et al., 2023) and directly establishing AA-based IPF optimization models (Leng et al., 2020; Cao et al., 2024). However, existing uncertainty analysis based on physical models suffers from the drawback of increased computational complexity, resulting in lengthy processing times, making it challenging to meet the power grid's demand for swift power flow computations.

Due to the advancements in computer and digital communication technologies, data acquisition in power systems has made significant progress. The deployment of Wide Area Measurement Systems (WAMS) has enabled the reliable collection of high-precision, wide-area synchronized electrical quantities, including voltage, current, phase angles, et al. This progress has fostered the development of data-driven power flow analysis methods, providing a solution to the issue of low efficiency in traditional model-driven power flow analysis (Fu et al., 2024). A data-driven linear PF model incorporating the support vector regression (SVR) and ridge regression (RR) algorithms was proposed in (Li et al., 2023). Similarly, a linear regression model was solved by RR to suppress the effect of data collinearity in (Chen, Y. et al., 2022). In distribution networks, the single-phase PF model is often considered. For instance, a data-driven single-phase linear PF model was introduced in (Xing et al., 2021). A data-driven convex model for hybrid AC/DC microgrids operation involving bi-directional converters was proposed in (Liang et al., 2023). Nevertheless, distribution power systems (DPSs) are generally unbalanced and it is still necessary to study linear three-phase distribution PF models. A data-driven-aided linear three-phase PF model for DPSs considering the imbalance was constructed in (Liu, Y. et al., 2022), and a data-driven piecewise linearization for distribution three-phase stochastic power flow was proposed in (Chen, J. et al., 2022), mitigating the errors of model-based PF linearization approaches. To overcome the challenge of obtaining accurate results with linear model-based data-driven methods, an approach with high adaptability to the nonlinearity of PF was proposed based on the thought of Koopman operator theory (Guo et al., 2022). What's more, a risk-free method was proposed in (Dong et al., 2022) to accelerate AC power flow with machine learning-based initiation, reducing the PF computation time. To tackle the challenges of the hidden measurement noise in the data-driven PF linearization, the problem was transformed into a regression model where the structure of the PF equations was exploited (Liu et al., 2020). Besides, the local load fluctuation suppression and its interaction with distribution system should also be addressed which brings the exact necessity towards the power flow prediction (Khalid et al., 2022; Rehman et al., 2024). Also, here the role of ancillary services and renewable energy integration should also be addressed towards covering the intermittency (Musleh et al., 2019; Sun et al., 2020). In some

cases, the database may not possess the envisioned completeness and appropriateness. There is a trend that combines the physical model-driven and data-driven. This can make up for the issues arising from incomplete data (Xing et al., 2022; Liu et al., 2021). A hybrid physical model-driven and data-driven approach for linearizing the power flow model was proposed in (Tan et al., 2020), and the linearized errors are obtained by the partial least squares regression-based data-driven approach. In the condition of lack of data, physical model parameters are introduced to assist the data-driven training process (Shao et al., 2023), and a highly scalable data-driven algorithm for stochastic AC-OPF that has extremely low sample requirements was presented in (Mezghani et al., 2020). To enhance the performance and generalization ability of the data-driven model, a physics-guided neural network was proposed to solve the PF problem by encoding different granularity of Kirchhoff's laws, and system topology into the rebuilt PF model (Hu et al., 2021). The fusion of robust principles with data-driven approaches has also enhanced the precision of data-driven methods. The worst-case errors were probabilistically constrained through distributionally robust chance-constrained programming (Liu, Y. et al., 2022; Chen et al., 2020). It also allows guaranteeing the linearization accuracy for a chosen operating point. In addition, a more comprehensive summary and discussion of existing data-driven PF linearization was presented in (Jia and Hug, 2023). For data-driven methods, support vector machine (SVM) is widely used due to its strong robustness and generalization ability, particularly excelling in scenarios with small samples and high dimensionality. Addressed to the N-k1-k2 cascading outages, the researchers employ SVM for classifier training, enabling the fast, reliable, and robust computation of active and reactive power flows (Xue and Liu, 2021). The SVM is utilized for optimal power flow with small-signal stability constraints in (Liu, J. et al., 2022), achieving high computational efficiency and economic benefits.

Although data-driven PF methods have made significant advancements, combining data-driven approaches with uncertainty still presents challenges. On the one hand, data-driven methods require a large amount of real or simulated data, which is what uncertain PF lacks. Historical data is difficult to obtain, and generating simulated data often incurs higher costs compared to deterministic PF. On the other hand, effectively integrating uncertainty into data-driven models is a challenge, as these uncertainties are often high-dimensional, increasing the complexity of modeling. In response, interval approaches offer the advantages of simple modeling and high simulation accuracy, while SVR can handle high-dimensional data, making it suited to the requirements. Therefore, this paper adopts interval modeling to represent uncertainties and selects SVR as the data-driven approach.

1.4 Contribution

This paper is dedicated to improving the computational efficiency of IPF in distribution networks to achieve real-time analyses, providing essential support for the rapid response of uncertain distribution systems with DGs integration. To this end, a method for IPF prediction in distribution networks based on SVR is proposed by combining data-driven methods with interval approaches. Accordingly, the research makes the following contributions.

Firstly, an IPF model for distribution networks based on the OSM is established considering system uncertainty as intervals. In addition to the uncertainty of power generation and load demand, the uncertainty of line parameters is also considered in this model. Due to environmental variations, the parameters of network lines exhibit a certain level of uncertainty. This consideration improves the accuracy of the model.

Secondly, an IPF prediction model is constructed using SVR based on the interval dataset generated by simulation. Different from traditional data-driven models, this model is a multi-output model that separately outputs the upper and lower bounds of the power flow results. This interval result fully considers various uncertainties in the distribution system, as the model is trained with these uncertainties incorporated.

Thirdly, the established SVR-based IPF prediction approach has been demonstrated to have high prediction accuracy and computational efficiency. The effectiveness of this approach is validated through studies on both IEEE 33bw and IEEE 69 cases. The IEEE 33bw case is primarily used to evaluate the model's accuracy, while the IEEE 69 case is mainly used to analyze the model's computational efficiency.

The IPF model for distribution networks is introduced in Section 2. The training and prediction algorithm through SVR is introduced in Section 3. The procedure of the method is introduced in Section 4. The case studies are conducted in Section 5, and conclusions in Section 6.

2 Construction of IPF model for distribution networks

2.1 Distflow formulation

The relaxed Distflow model for the radial distribution network is expressed as Equations 1–4. Before constructing the model, it is customary to assume that the transmission lines do not involve parallel grounding branches and to specify that the direction of current and power flow from node i to node j is positive.

$$v_j = v_i - 2(r_{ij}P_{ij} + x_{ij}Q_{ij}) + (r_{ij}^2 + x_{ij}^2)l_{ij}, \forall (i, j) \in \mathbf{B} \quad (1)$$

$$P_{ij}^2 + Q_{ij}^2 \leq l_{ij}v_i \Leftrightarrow \left\| \begin{array}{c} 2P_{ij} \\ 2Q_{ij} \\ l_{ij} - v_i \end{array} \right\|_2 \leq l_{ij} + v_i \quad (2)$$

$$\sum_{k:j \rightarrow k} P_{jk} - \sum_{i:i \rightarrow j} (P_{ij} - r_{ij}l_{ij}) = p_j, \forall j \in \mathbf{D} \quad (3)$$

$$\sum_{k:j \rightarrow k} Q_{jk} - \sum_{i:i \rightarrow j} (Q_{ij} - x_{ij}l_{ij}) = q_j, \forall j \in \mathbf{D} \quad (4)$$

The model is the branch power flow model after convex relaxation, where Equation 1 is the voltage equation, Equation 2 is the power equation at the sending end of the branch, Equations 3, 4 are the power balance equation. \mathbf{B} and \mathbf{D} are the set of branches and nodes. We set that $l_{ij} = |I_{ij}|^2$ and $v_i = |V_i|^2$, where V_i is the voltage vector of node i , and I_{ij} is the current vector flowing through branch (i, j) . r_{ij} is the resistance and x_{ij} is the reactance of transmission line. P_{ij} and Q_{ij} are the active and reactive line transmission power from node i to node j , respectively. Note that more than one upstream and downstream branch is connected to

node j . p_j and q_j are the injection active and reactive power of node j , respectively, which are equal to the power generation minus the load demand, i.e., $p_j = p_j^G - p_j^L$.

$$\begin{aligned} & \min (\max) x_i \\ & \text{s.t.} \begin{cases} \mathbf{h}(\mathbf{x}) = \boldsymbol{\xi} \\ \underline{\mathbf{h}} \leq \boldsymbol{\xi} \leq \bar{\mathbf{h}} \end{cases} \end{aligned} \quad (9)$$

2.2 Modelling of IPF based on distflow

In active distribution networks with DGs integration, the output of distributed generators and flexible loads both exhibit a certain degree of uncertainty, which has a significant impact on the safe and stable operation of the distribution networks. Therefore, it is essential to consider these uncertainties. In this paper, the interval approach is utilized to describe uncertainties, ensuring the security of system operation. Additionally, the network parameters, including line resistance and reactance, may experience variations due to external environmental factors. To make the model more practical, the uncertainties of these parameters are considered simultaneously during modelling.

In the interval approach, the active power generation and load demand, as well as line parameters are represented in interval form, and the interval results for variables such as voltage and line transmission power can be obtained. Representing the interval form in $\hat{\chi}$, where $\hat{\chi} = [\underline{\chi}, \bar{\chi}]$, the IPF model based on Distflow for distribution networks can be expressed as Equations 5–8.

$$\hat{v}_j = \hat{v}_i - 2(\hat{r}_{ij}\hat{P}_{ij} + \hat{x}_{ij}\hat{Q}_{ij}) + (\hat{r}_{ij}^2 + \hat{x}_{ij}^2)\hat{l}_{ij}, \quad \forall (i, j) \in \mathbf{B} \quad (5)$$

$$\left\| \begin{matrix} 2\hat{P}_{ij} \\ 2\hat{Q}_{ij} \\ \hat{l}_{ij} - \hat{v}_i \end{matrix} \right\|_2 \leq \hat{l}_{ij} + \hat{v}_i, \quad \forall (i, j) \in \mathbf{B} \quad (6)$$

$$\sum_{k:j \rightarrow k} \hat{P}_{jk} - \sum_{i:i \rightarrow j} (\hat{P}_{ij} - \hat{r}_{ij}\hat{l}_{ij}) = \hat{p}_j, \quad \forall j \in \mathbf{D} \quad (7)$$

$$\sum_{k:j \rightarrow k} \hat{Q}_{jk} - \sum_{i:i \rightarrow j} (\hat{Q}_{ij} - \hat{x}_{ij}\hat{l}_{ij}) = \hat{q}_j, \quad \forall j \in \mathbf{D} \quad (8)$$

where $\hat{p}_j = \hat{p}_j^G - \hat{p}_j^L$, $\hat{q}_j = q_j^G - q_j^L$. \hat{p}_j^G and q_j^G are the active and reactive power generation, respectively. \hat{p}_j^L and q_j^L are the active and reactive load demand, respectively.

The IPF model based on Distflow can draw inspiration from the principles of OSM for its solution. In this approach, the interval uncertainties of the IPF model are regarded as variables that vary in their interval bounds, and the desired variables are set as the objective functions. Thus, it involves transforming the resolution of a set of interval nonlinear equations into variable optimization problems. The core of OSM is based on the Extreme Value Theorem through which we can get two points of conclusions. We simplify Equations 5–8 as $\mathbf{h}(\mathbf{x}) = [\underline{\mathbf{h}}, \bar{\mathbf{h}}]$ where $[\underline{\mathbf{h}}, \bar{\mathbf{h}}]$ are interval input data and \mathbf{x} are the variables of the IPF model. The first point is that there is a fixed \mathbf{x} corresponding to an arbitrary scenario $\boldsymbol{\xi} \in [\underline{\mathbf{h}}, \bar{\mathbf{h}}]$ in the power flow calculation. The second point is that there exists a special scenario $\xi_i^{\min} (\xi_i^{\max})$ for each single variable x_i making x_i minimum (maximum) for all scenarios $\boldsymbol{\xi} \in [\underline{\mathbf{h}}, \bar{\mathbf{h}}]$. The minimum and maximum are denoted as x_i^{\min} and x_i^{\max} , and the interval $[x_i^{\min}, x_i^{\max}]$ is the solution of x_i under the input data $[\underline{\mathbf{h}}, \bar{\mathbf{h}}]$.

From the two points of conclusions, the solution for IPF model is reduced to find ξ_i^{\min} and ξ_i^{\max} for each variable x_i by establishing the minimum and maximum optimization models Equation 9 of power flow variables.

Taking the variable v_i in distribution networks, for example, solving the IPF model Equations 5–8 can be transformed into solving the optimization model Equation 10, and the model can be solved through commercial solvers such as CPLEX.

$$\begin{aligned} & \min (\max) v_i, \quad \forall i \in \mathbf{D} \\ & \left\{ \begin{aligned} & \left\| \begin{matrix} 2P_{ij} \\ 2Q_{ij} \\ l_{ij} - v_i \end{matrix} \right\|_2 \leq l_{ij} + v_i, \quad \forall (i, j) \in \mathbf{B} \\ & v_j = v_i - 2(r_{ij}P_{ij} + x_{ij}Q_{ij}) + (r_{ij}^2 + x_{ij}^2)l_{ij}, \quad \forall (i, j) \in \mathbf{B} \\ & \sum_{k:j \rightarrow k} P_{jk} - \sum_{i:i \rightarrow j} (P_{ij} - r_{ij}l_{ij}) = p_j, \quad \forall j \in \mathbf{D} \\ & \sum_{k:j \rightarrow k} Q_{jk} - \sum_{i:i \rightarrow j} (Q_{ij} - x_{ij}l_{ij}) = q_j, \quad \forall j \in \mathbf{D} \\ & q_j^G - \bar{q}_j^L \leq q_j \leq q_j^G - \underline{q}_j^L, \quad \forall j \in \mathbf{D} \\ & \underline{p}_j^G - \bar{p}_j^L \leq p_j \leq \bar{p}_j^G - \underline{p}_j^L, \quad \forall j \in \mathbf{D} \\ & \underline{x}_{ij}(l_{ij}) \leq x_{ij}(l_{ij}) \leq \bar{x}_{ij}(\bar{l}_{ij}), \quad \forall (i, j) \in \mathbf{B} \end{aligned} \right. \end{aligned} \quad (10)$$

It can be succinctly described as searching for a specific scenario $\xi_i^{\min} (\xi_i^{\max})$ among all uncertain scenarios of the distribution network, which can minimize (maximize) the voltage magnitude $|V_i|$ at node i , so as to obtain the voltage interval $[V_i^{\min}, V_i^{\max}]$. Naturally, the objective function v_i of Equation 10 can also be replaced with active power transmission P_{ij} or reactive power transmission Q_{ij} .

3 IPF prediction method for distribution networks based on SVR

As the system scale increases, the efficiency of model-driven IPF analysis significantly decreases, which does not meet the current demands for rapid PF computations in distribution networks. Therefore, the data-driven approach has garnered attention for achieving faster IPF computations. The SVR has been opted for in this research due to its advantages of handling high-dimensional data, which is aligned with the characteristics of IPF analysis.

3.1 Construction of eigenvectors in IPF

In the typical SVR framework, the model is designed for single-output problems. However, in the context of IPF models, situations may arise where some nodes attain their maximum values while others reach their minimum values within the same input scenario since both input data and output variables are represented as intervals. Therefore, the SVR model for IPF is fundamentally a multiple-output problem. Corresponding to the same input scenario, the situation where different nodes attain either maximum or minimum values may vary. In such cases, training the SVR model based on the specific input and a singular minimum (or maximum) output would lead to a significant decrease in model accuracy. Based on this, the feature vectors in IPF model are established.

The well-constructed feature vectors are crucial prerequisites for ensuring the effectiveness of data-driven model learning. In the

analysis of extensive historical state data for distribution networks with DGs, it is essential to determine the input and output features for the IPF analysis at first. Given that the primary characteristic of distribution networks with DGs is the uncertainty of renewable power generation and load demand, which significantly impacts IPF analysis results, the sequence of renewable power generation and load demand for the distribution system is selected as the input eigenvector of the SVR model, and the sequence of node voltages and active line transmission power, which is indicative of power flow results, is selected as the output feature vector.

3.1.1 Construction of input eigenvector adapted to variations in source-grid-load

The uncertainty of source, grid, and load is represented in interval form for the IPF model. Therefore, the values in the input eigenvector should be intervals distinguishing from conventional eigenvectors. However, directly using interval values for training poses challenges such as computational complexity, model misfit, and difficulty in interpreting learning patterns. To address these issues, it is necessary to identify relevant parameters that can characterize interval features, such as interval midpoints and interval radii, to replace interval values during training. The midpoint of the interval is the operating point of generator, which reflects the randomness of generator output. The interval radius can reflect the fluctuation degree of uncertain data. Therefore, the interval midpoint of source and radius of source-grid-load data is used to construct the input eigenvector instead of interval values.

Take the renewable active power generation $\hat{p}_j^G = [p_j^G, \bar{p}_j^G]$ as an example, the relationships Equations 11, 12 exist in the interval.

$$p_{-j}^G = p_{0,j}^G - \Delta p_j^G, \bar{p}_j^G = p_{0,j}^G + \Delta p_j^G \quad (11)$$

$$\Delta p_j^G = \sigma \cdot p_{0,j}^G \quad (12)$$

where $p_{0,j}^G$ represents the interval midpoint, Δp_j^G is the interval radius, σ is the fluctuation coefficient. The $p_{0,j}^G$ and Δp_j^G can characterize the features of the renewable active power generation interval. For a certain distribution network, the value of the input eigenvector can be changed by changing the midpoint $p_{0,j}^G$ or the fluctuation coefficient σ . Besides, the active and reactive load demands, and line parameters follow the similar principle.

The eigenvector for source includes the sequence of renewable active power generation $\Delta p^G = \{\Delta p_1^G, \Delta p_2^G, \dots, \Delta p_M^G\}$ and $p_0^G = \{p_{0,1}^G, p_{0,2}^G, \dots, p_{0,M}^G\}$, which for load includes the sequences of active and reactive load demand $\Delta p^L = \{\Delta p_1^L, \Delta p_2^L, \dots, \Delta p_D^L\}$, $\Delta q^L = \{\Delta q_1^L, \Delta q_2^L, \dots, \Delta q_D^L\}$, and which for grid includes the sequences of line parameters $\Delta r = \{\Delta r_1, \Delta r_2, \dots, \Delta r_B\}$, $\Delta x = \{\Delta x_1, \Delta x_2, \dots, \Delta x_B\}$. According to this, the input eigenvector adaptable to variations in source-grid-load can be formulated as follows:

$$X_m = [\Delta p^G, p_0^G, \Delta p^L, \Delta q^L, \Delta r, \Delta x] \quad (13)$$

where M is the number of DGs, D is the number of nodes, B is the number of branches.

3.1.2 Construction of output feature vector

When conducting PF analysis, it is essential to consider the output features that can reflect power system quality and

stability. In power flow results, node voltage or line transmission power can be used to evaluate system stability. Therefore, the node voltage is selected as output features in this paper. In IPF model, node voltages are represented as interval values, so that the output features of the SVR training model are essentially intervals. However, training the model directly with interval values as the output vector may lead to issues such as model complexity and low interpretability. To address the issues, it is preferable to choose upper and lower bounds that characterize interval features as the output feature vector. This involves establishing the SVR model with two output nodes. According to this, the output feature vector in IPF can be constructed as Equation 14.

$$Y_{out1} = V_{min}, Y_{out2} = V_{max} \quad (14)$$

Certainly, we can also construct the output feature vector as presented in Equation 15 to obtain the predictive results of line transmission power.

$$Y_{out1} = P_{ij,min}, Y_{out2} = P_{ij,max} \quad (15)$$

3.2 Modelling of SVR-based IPF prediction

Support Vector Machine (SVM) is a binary classification algorithm, and its fundamental model is a linear classifier that maximizes the margin in the feature space. The objective of SVM learning is to find a hyperplane that separates the samples, guided by the principle of maximizing the margin. This ultimately translates into solving a convex quadratic programming problem. The variant of SVM used in this research for IPF prediction is SVR, specifically designed for solving regression problems. The principle of SVR is presented in Figure 1. SVR can be categorized into three types according to the linear separability of the training data, including Linear Hard ϵ -SVR, Linear ϵ -SVR, and ϵ -SVR.

The original data for IPF analysis is considered linearly non-separable. Therefore, this paper selects the ϵ -SVR model to explore the connection between the input and output of the

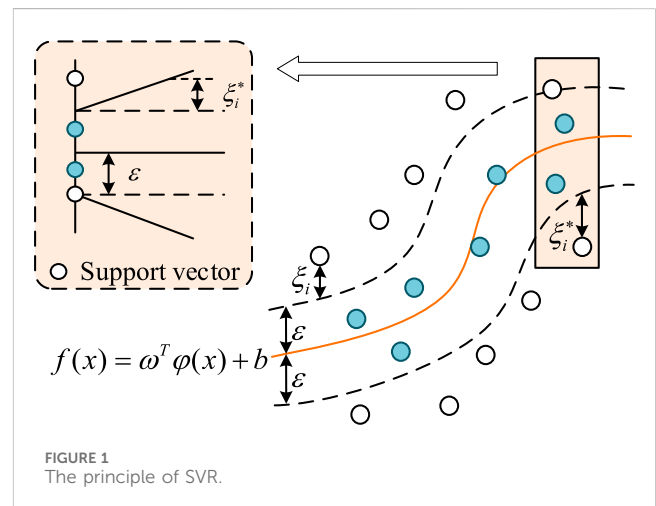


FIGURE 1 The principle of SVR.

IPF for distribution systems. Based on the constructed feature vectors in IPF, the ϵ -SVR model for IPF prediction is established as follows.

According to the description in 3.1, the training data set of the model can be obtained as $T = \{(X_{in}, Y_{out1}, Y_{out2})_1, (X_{in}, Y_{out1}, Y_{out2})_2, \dots, (X_{in}, Y_{out1}, Y_{out2})_N\}$ $X_{in} \in R^d$. Then divide the training data set into two groups $T_1 = \{(X_{in}, Y_{out1})_1, (X_{in}, Y_{out1})_2, \dots, (X_{in}, Y_{out1})_N\}$ and $T_2 = \{(X_{in}, Y_{out2})_1, (X_{in}, Y_{out2})_2, \dots, (X_{in}, Y_{out2})_N\}$, and two SVR training models Equations 16, 17 can be built for the minimum and maximum outputs depending on each group of training data.

$$\min_{\omega_l, b_l} \frac{1}{2} \|\omega_l\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \tag{16}$$

$$s.t. |(\omega_l \cdot X_{in}) + b_l - Y_{out1}| \leq \epsilon + \xi,$$

$$\xi_i, \xi_i^* \geq 0, i = 1, 2, \dots, N$$

$$\min_{\omega_u, b_u} \frac{1}{2} \|\omega_u\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \tag{17}$$

$$s.t. |(\omega_u \cdot X_{in}) + b_u - Y_{out2}| \leq \epsilon + \xi,$$

$$\xi_i, \xi_i^* \geq 0, i = 1, 2, \dots, N$$

where ω_l and ω_u are the normal vectors, b_l and b_u are constants, ξ_i, ξ_i^* are the slack variables, C is the penalty factor, and $C > 0$. ϵ represents the distance swept by the hyperplane across the regions on either side, and the “ ϵ -band” includes all training points of each training data set.

3.3 Solving of SVR-based IPF prediction model

The SVR training models are solved in this section. To reduce the complexity of solving, the models Equations 16, 17 can be

transformed into Equations 18, 19 through applying the Lagrangian function and choosing an appropriate kernel function $K(x, x')$.

$$\min_{\alpha_l^* \in R^{2N}} \sum_{i,j=1}^N (\alpha_{l,i}^* - \alpha_{l,j}^*) (\alpha_{l,j}^* - \alpha_{l,i}^*) K(X_{in,i}, X_{in,j})$$

$$+ \epsilon \sum_{i=1}^N (\alpha_{l,i}^* + \alpha_{l,i}^*) - \sum_{i=1}^N Y_{out1,i} (\alpha_{l,i}^* - \alpha_{l,i}^*),$$

$$s.t. \sum_{i=1}^N (\alpha_{l,i}^* - \alpha_{l,i}^*) = 0,$$

$$0 \leq \alpha_{l,i}^*, \alpha_{l,i}^* \leq C, i = 1, 2, \dots, N \tag{18}$$

$$\min_{\alpha_u^* \in R^{2N}} \sum_{i,j=1}^N (\alpha_{u,i}^* - \alpha_{u,j}^*) (\alpha_{u,j}^* - \alpha_{u,i}^*) K(X_{in,i}, X_{in,j})$$

$$+ \epsilon \sum_{i=1}^N (\alpha_{u,i}^* + \alpha_{u,i}^*) - \sum_{i=1}^N Y_{out2,i} (\alpha_{u,i}^* - \alpha_{u,i}^*),$$

$$s.t. \sum_{i=1}^N (\alpha_{u,i}^* - \alpha_{u,i}^*) = 0,$$

$$0 \leq \alpha_{u,i}^*, \alpha_{u,i}^* \leq C, i = 1, 2, \dots, N \tag{19}$$

where $\alpha_{l,i}, \alpha_{l,i}^*, \alpha_{u,i}, \alpha_{u,i}^*$ are the Lagrange multipliers corresponding to the inequality constraints. The optimization problems can be solved by commercial solvers. The optimal solutions are attained as $\bar{\alpha}_l = (\bar{\alpha}_{l,1}, \bar{\alpha}_{l,1}^*, \dots, \bar{\alpha}_{l,N}, \bar{\alpha}_{l,N}^*)^T$ $\bar{\alpha}_u = (\bar{\alpha}_{u,1}, \bar{\alpha}_{u,1}^*, \dots, \bar{\alpha}_{u,N}, \bar{\alpha}_{u,N}^*)^T$, respectively. Then the decision functions are constructed as (20) and (21), and the corresponding \bar{b}_l and \bar{b}_u can be calculated by Equations 22, 23, respectively. It is noted that \bar{b} is calculated differently depending on $\bar{\alpha}_j$ or $\bar{\alpha}_k$.

$$f_{\min}(x) = \sum_{i=1}^N (\bar{\alpha}_{l,i}^* - \bar{\alpha}_{l,i}) K(X_{in,i}, x) + \bar{b}_l \tag{20}$$

$$f_{\max}(x) = \sum_{i=1}^N (\bar{\alpha}_{u,i}^* - \bar{\alpha}_{u,i}) K(X_{in,i}, x) + \bar{b}_u \tag{21}$$

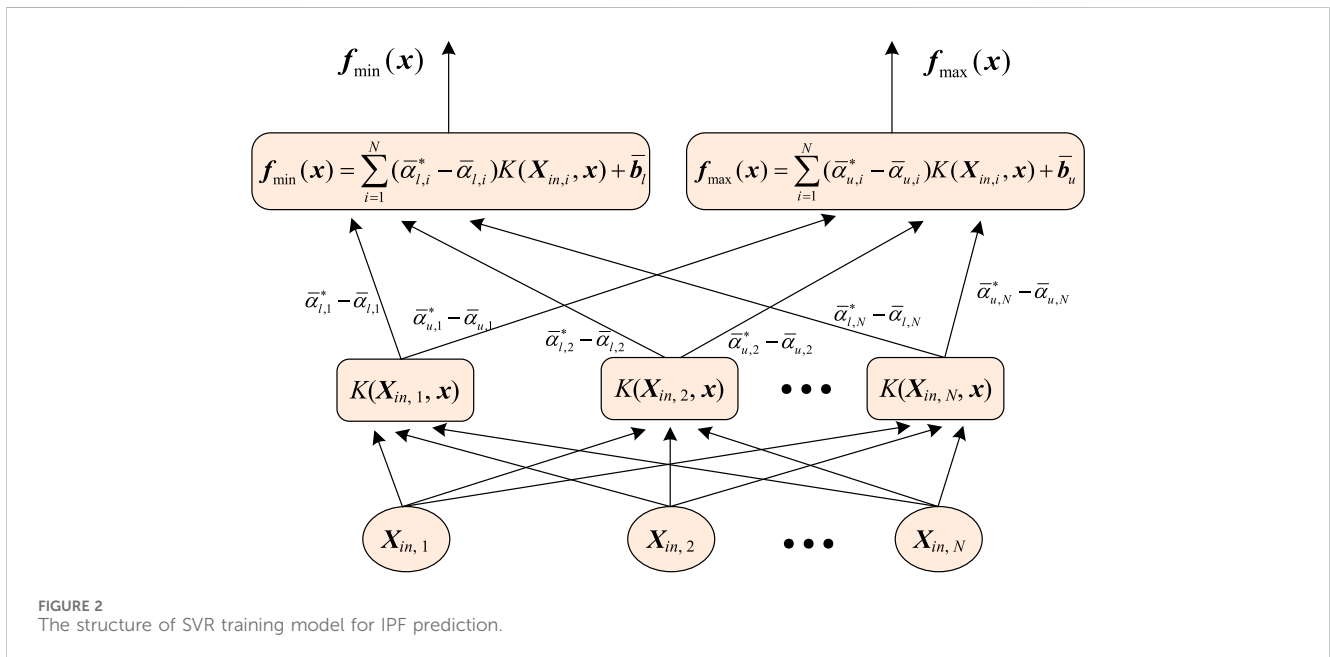


FIGURE 2 The structure of SVR training model for IPF prediction.

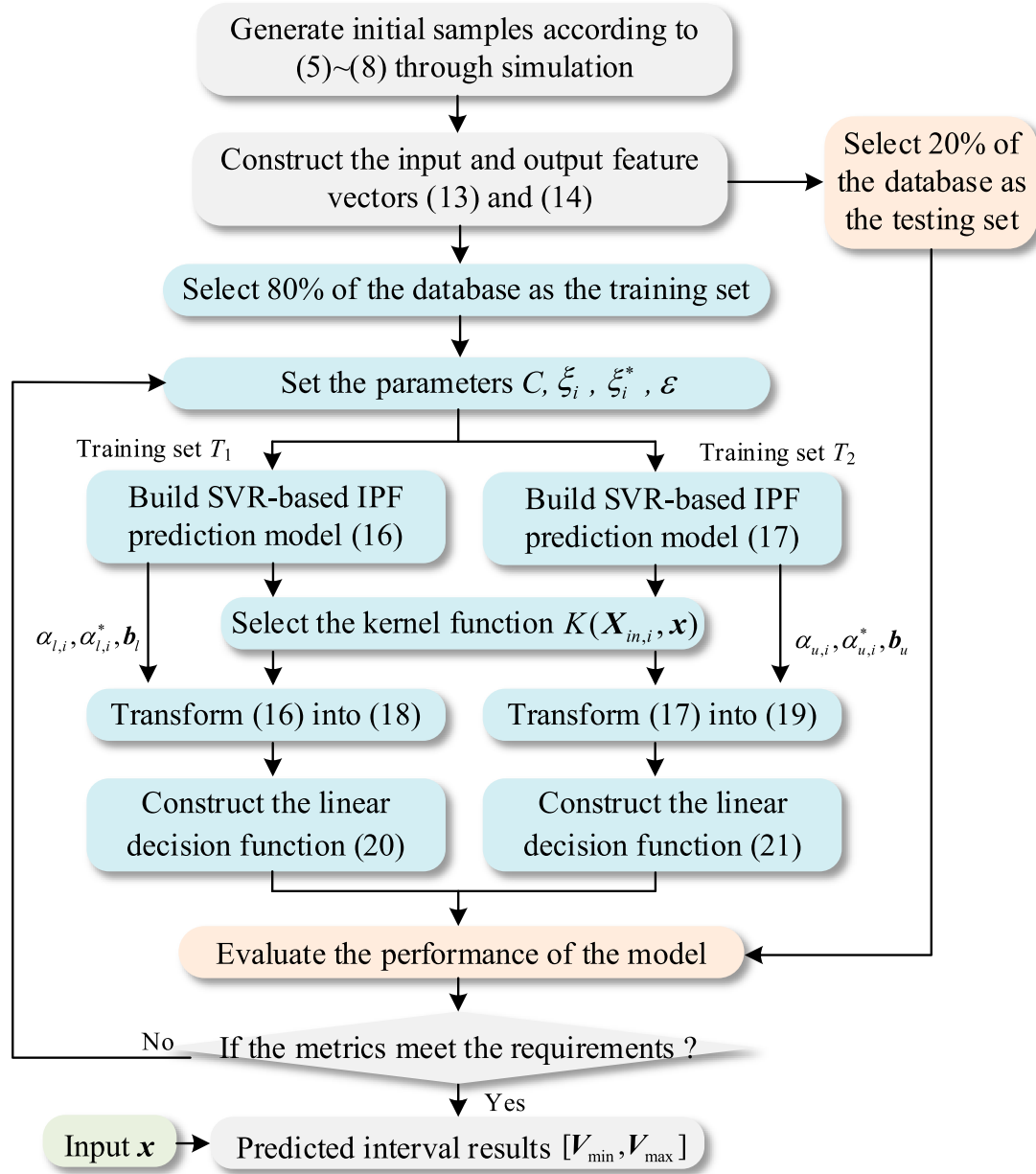


FIGURE 3 The flow chart of SVR-based IPF prediction method.

$$\bar{b}_l = Y_{out1,j} - \sum_{i=1}^N (\bar{\alpha}_{l,i}^* - \bar{\alpha}_{l,i}) K(X_{in,i}, X_{in,j}) + \epsilon$$

$$\bar{b}_l = Y_{out1,k} - \sum_{i=1}^N (\bar{\alpha}_{l,i}^* - \bar{\alpha}_{l,i}) K(X_{in,i}, X_{in,k}) - \epsilon \tag{22}$$

$$\bar{b}_u = Y_{out2,j} - \sum_{i=1}^N (\bar{\alpha}_{u,i}^* - \bar{\alpha}_{u,i}) K(X_{in,i}, X_{in,j}) + \epsilon$$

$$\bar{b}_u = Y_{out2,k} - \sum_{i=1}^N (\bar{\alpha}_{u,i}^* - \bar{\alpha}_{u,i}) K(X_{in,i}, X_{in,k}) - \epsilon \tag{23}$$

The structure of SVR training model for IPF prediction can be depicted as shown in Figure 2.

According to Figure 2, the minimum and maximum values of the power flow results, that is the interval results $[V_{min}, V_{max}]$ or $[P_{ij,min}, P_{ij,max}]$ for IPF prediction in the distribution network, are obtained based on the corresponding linear mapping relationships for any given input eigenvector.

4 The procedure of SVR-based IPF prediction method

In the SVR-based IPF prediction method for distribution networks, the first step involves establishing an IPF model for generating the initial sample database through simulation. The

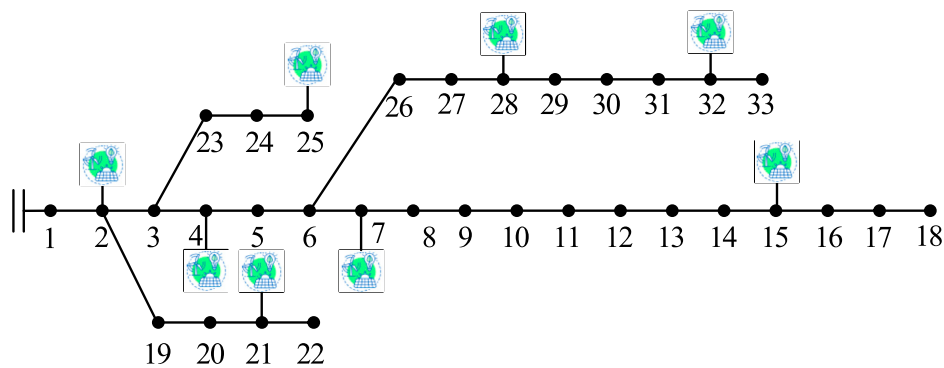


FIGURE 4
The topology of enhanced IEEE 33bw distribution network.

database includes intervals for node injections of active and reactive power, intervals for line parameter fluctuations, and corresponding intervals for node voltage. The data are then processed to construct the input and output feature vectors. Subsequently, the SVR training model is established, and the formulation of linear decision function can be determined by solving the model. Finally, when given a specific input eigenvector, the interval results for power flow variables can be predicted according to the decision function. The detailed procedure of SVR-based IPF prediction is expressed as follows and the flow chart is presented in Figure 3.

- Step1. Data Generating. Generate a diverse set of initial samples according to the established IPF model Equations 5–8 through simulation, where each set comprises the interval of active and reactive node power injection fluctuations, the interval of line parameter fluctuations, and the associated node voltage intervals.
- Step2. Data Preprocessing. Select and extract features from the initial samples, and construct the input and output feature vectors Equations 13, 14 for each set of data in the database according to Section 3.1. To ensure the accuracy of model training, normalize the input data. Then select 80% of the database as the training set and 20% of the database as the testing set.
- Step3. IPF Prediction Model Construction. Set the parameters C , ξ_i , ξ_i^* , and ε . Then, obtain T_1 and T_2 from the training dataset, and build SVR-based IPF prediction models Equations 16, 17 depending on T_1 and T_2 , respectively. Meanwhile, experiment with different kernel functions $K(X_{in,i}, \mathbf{x})$ and select the one that yields the best results.
- Step4. IPF Prediction Model Solving. Transform the constructed models Equations 16, 17 into Equations 18, 19, and the parameters $\alpha_{l,i}$, $\alpha'_{l,i}$, $\alpha_{u,i}$, $\alpha'_{u,i}$, \mathbf{b}_l , and \mathbf{b}_u can be obtained by solving Equations 18, 19. Then construct the linear decision functions Equation 20, 21 for predicting the minimum and maximum value of power flow variables, respectively.
- Step5. Model Evaluation. Based on the testing dataset, evaluate the performance of the model using appropriate metrics, such as mean absolute error (MAE) and root mean square error (RMSE). Then determine if the metrics meet the

requirements. If the metrics meet the expectations, proceed to step6; otherwise, adjust the parameters C , ξ_i , ξ_i^* , ε and return to step 3.

- Step6. Interval Power Flow Prediction. Give the independent and specific input eigenvector \mathbf{X} , so that obtain the corresponding minimum voltage value V_{\min} and maximum voltage value V_{\max} through substituting \mathbf{X} into the decision functions and De-normalization. Finally, the predicted interval results $[V_{\min}, V_{\max}]$ can be yielded.

5 Case studies

The performance of the proposed SVR-based IPF prediction method is tested on IEEE 33bw and IEEE 69 distribution networks on an Intel(R) Core(TM) i5 PC, 2.50 GHz processor with 8 GB RAM. The algorithm is implemented in MATLAB. The IEEE 33bw case is primarily used to validate the accuracy of the established IPF prediction model and its adaptability to various system fluctuations. Meanwhile, the IEEE 69 case is employed to verify the efficiency and real-time capability of the proposed algorithm in predicting IPF of the distribution network.

5.1 IEEE 33bw case study

The IEEE 33bw case is illustrated in Figure 4, and the case has been enhanced to include eight distributed renewable energy sources. All parameters are valued according to the per unit (p.u.) system of analysis, with 10 MVA chosen as the basic power of the test case. The detailed original power generation data for these eight DGs are presented in Table 1. The voltage limits of all buses except the slack bus are constrained to $[0.9, 1.1]$.

5.1.1 Evaluation of the model

The active power generation fluctuation ranges of DGs are assumed to be $\pm 20\%$ of the original data, which is also assumed on the active and reactive load demand, that is the set the fluctuation coefficient $\sigma_1 = 0.2$. Meanwhile, considering the slight fluctuations in distribution network line parameters under both internal and

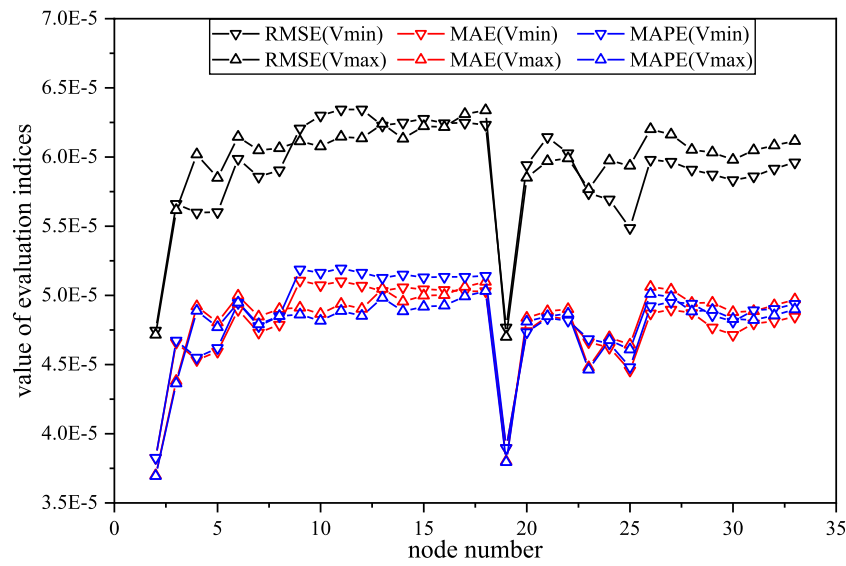


FIGURE 5 The evaluation results of RMSE, MAE and MAPE for the SVR-based IPF prediction model.

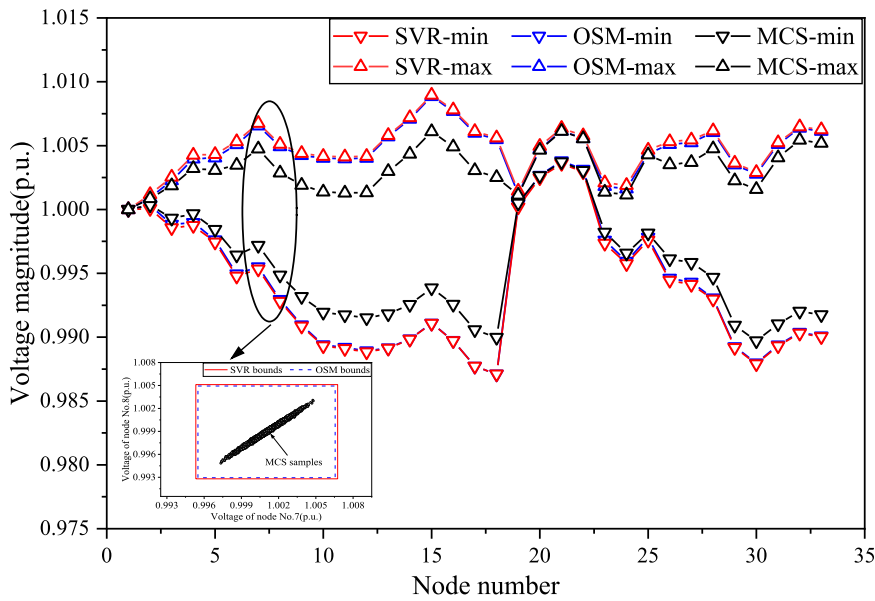


FIGURE 6 The voltage interval results obtained by SVR, OSM and MCS for IEEE 33bw case in Scenario 1.

external conditions, assume the fluctuation range is $\pm 10\%$ of r_{ij} and x_{ij} , that is $\sigma_2 = 0.1$. Within these ranges, 1,500 sets of initial data are randomly generated through simulation for the distribution network, where 1,200 sets are training sets, and 300 sets are testing sets. In the model training process, the parameters of SVR model are set to be $C = 5,000$, $\epsilon = 0.0001$, $\xi_i = 1$. The kernel function is selected as $K(x_i, x_j) = x_i^T x_j$.

To evaluate the model's performance comprehensively and objectively, the indices of mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE) and

R2 for the testing sets are calculated in this paper. They are defined as Equations 24–27. Using these metrics together helps to avoid biases introduced by a single metric, enhancing the robustness of the evaluation.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \tag{24}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \tag{25}$$

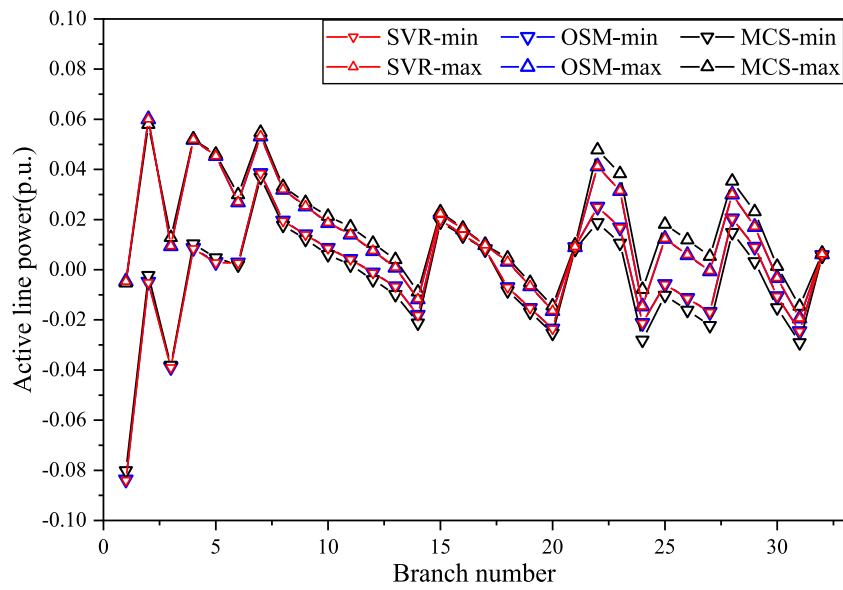


FIGURE 7 The active line power interval results obtained by SVR, OSM and MCS for IEEE 33bw case in Scenario 1.

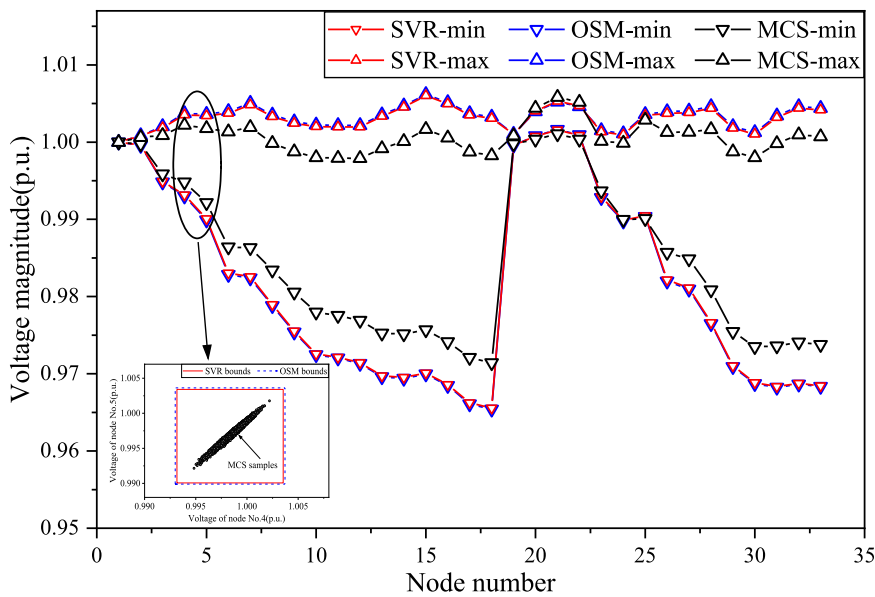


FIGURE 8 The voltage interval results obtained by SVR, OSM and MCS for IEEE 33bw case in Scenario 2.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (26)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (27)$$

where n is the number of samples, y_i is the observed values, \hat{y}_i is the corresponding model-predicted value, and \bar{y} is the mean of

the observed values. The evaluation results of the indices for the SVR-based IPF prediction model in IEEE 33bw case are presented in Figure 5.

It can be observed from Figure 5 that the value of the evaluation indices is ideal. For the lower and upper bounds of each node voltage, the RMSE evaluation results are within the range $[4.7 \times 10^{-5}, 6.4 \times 10^{-5}]$, the MAE evaluation results are within the range $[3.6 \times 10^{-5}, 5.2 \times 10^{-5}]$, and the MAPE evaluation results are within the range $[3.6 \times 10^{-5}, 5.3 \times 10^{-5}]$, all of which

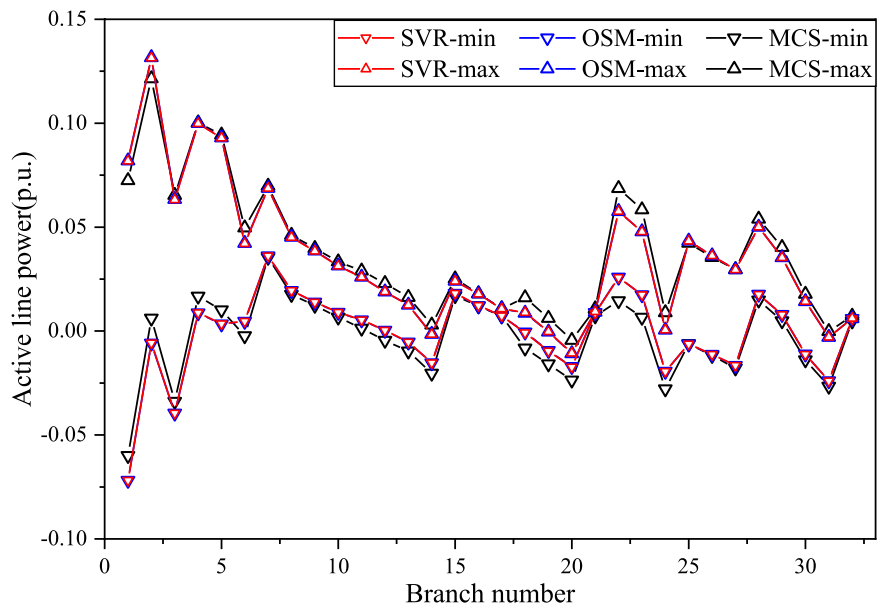


FIGURE 9 The active line power interval results obtained by SVR, OSM and MCS for IEEE 33bw case in Scenario 2.

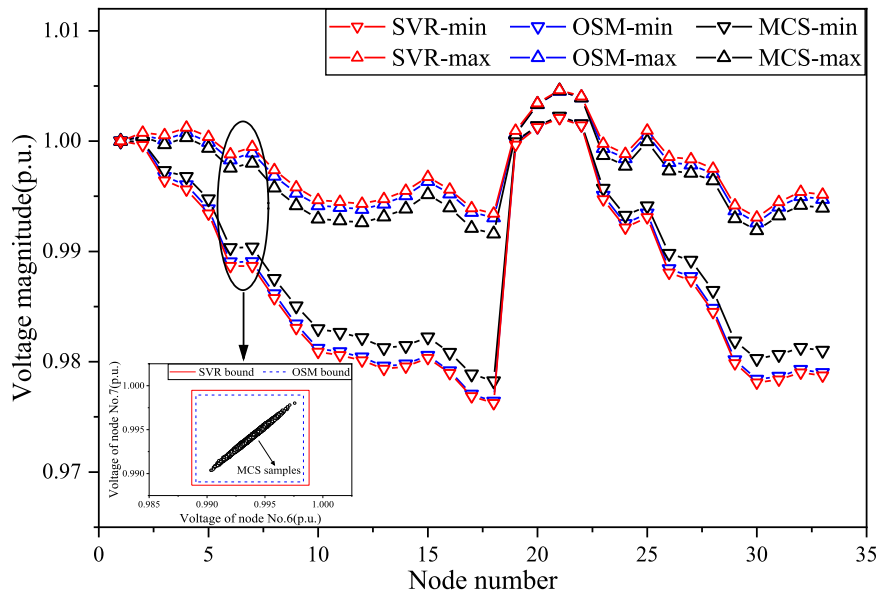


FIGURE 10 The voltage interval results obtained by SVR, OSM and MCS for IEEE 33bw case in Scenario 3.

are relatively small. Besides, the value of R2 can reach above 0.95 for both lower and upper bounds of each node voltage. These support the notion that the model's predicted values exhibit minimal deviation from the true values, indicating a strong fit of the model to the testing sets, which confirms the superior performance of the established SVR-based IPF prediction model.

5.1.2 Comparison with the OSM and MCS

To validate the accuracy and adaptability of the SVR-based IPF prediction model, three scenarios were designed to conduct the proposed method compared with the OSM (Zhang et al., 2017) and MCS. The forward-backward substitution is employed in MCS for solving general distribution network power flow. The three operating scenarios are described as follows.

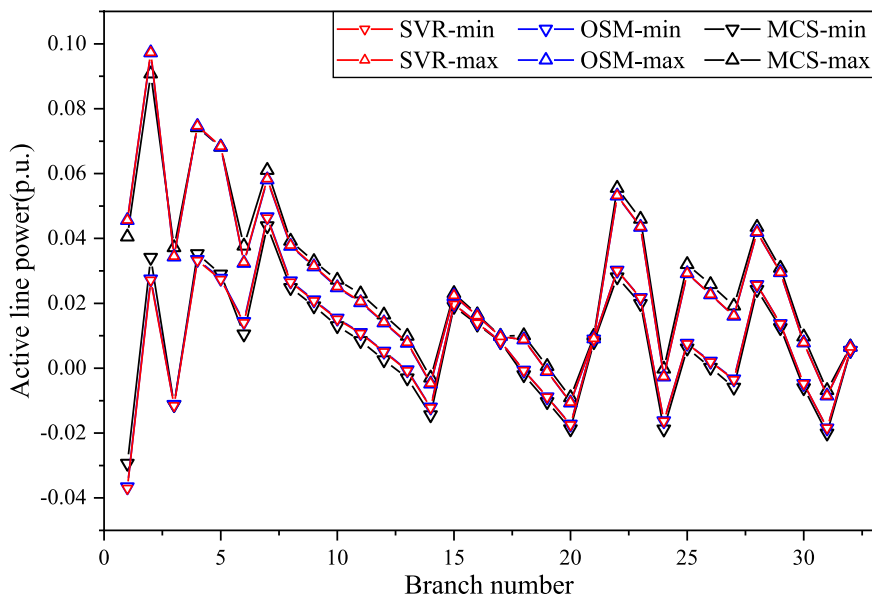


FIGURE 11 The active line power interval results obtained by SVR, OSM and MCS for IEEE 33bw case in Scenario 3.

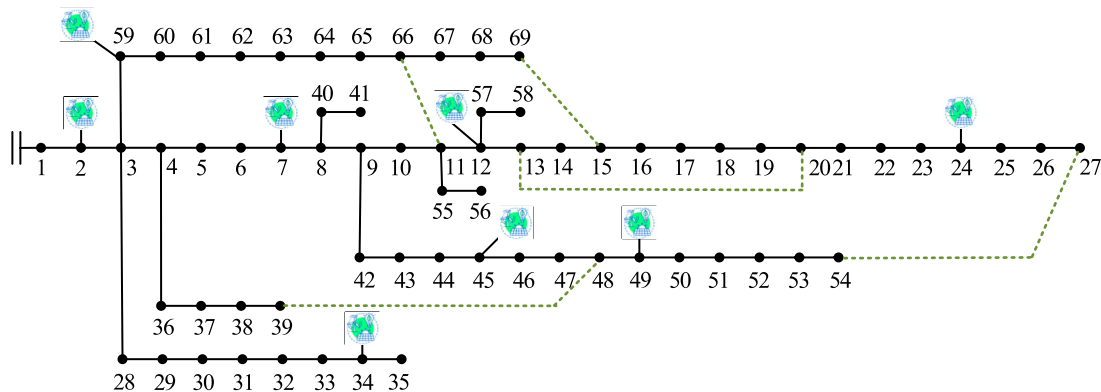


FIGURE 12 The topology of enhanced IEEE 69 distribution network.

Scenario 1: The same operating points, and the different fluctuation ranges;

Scenario 2: The different operating points, and the same fluctuation ranges;

Scenario 3: The different operating points, and the different fluctuation ranges,

where the settings of these scenarios are changed based on the training data. The operating points represent the original active power generation p_0^G , and the fluctuation ranges are set by changing the fluctuation coefficients σ_1 and σ_2 . They represent the randomness and volatility of uncertain data in distribution networks.

5.1.2.1 The simulation under scenario 1

In Scenario 1, the original active power generation data was the same as that in Table 1, and the fluctuation coefficients were set as

$\sigma_1 = 0.1, \sigma_2 = 0.05$. Thus, a new set of input eigenvector $X_{in,I}$ was introduced. The parameters of SVR model were set to be $C = 5,000, \epsilon = 0.0001, \xi_i = 1$, and the MCS was conducted with a sample size of 10,000 to ensure a high accuracy level. The simulation results under this scenario are demonstrated as follows.

The voltage interval results obtained by the SVR, OSM and MCS for IEEE 33bw case are presented in Figure 6, and the active line transmission power interval results are presented in Figure 7. Additionally, in Figure 6, the voltage interval boundary values of node No. 7 and No. 8 obtained by SVR and OSM are compared with the results of MCS sampling for a more intuitive presentation. It can be observed from Figure 6 that the voltage interval results obtained by SVR are very close to those acquired by the OSM, and the voltage interval range obtained by SVR and OSM is larger than that obtained by MCS. This is to be expected, because the initial data for SVR

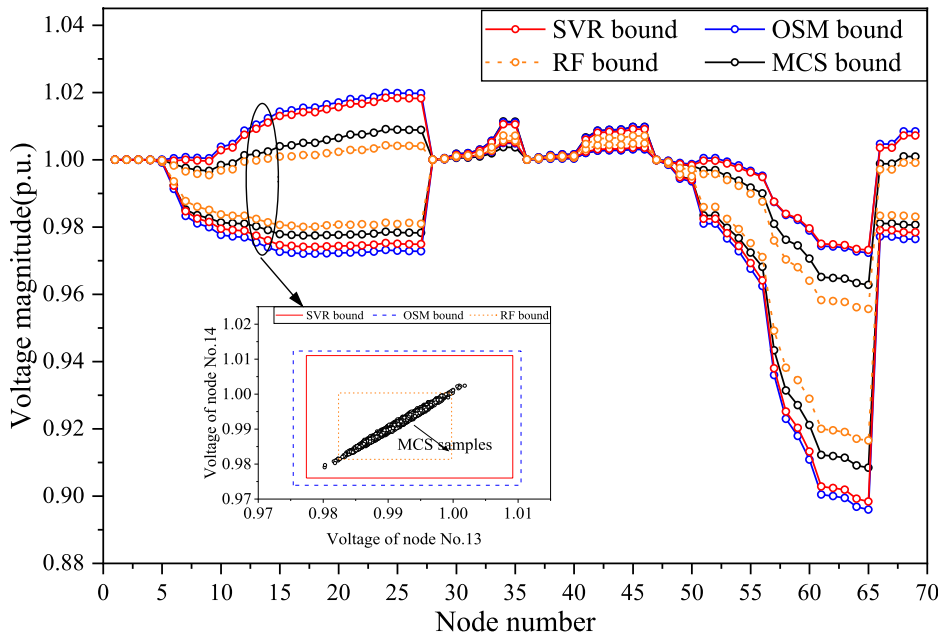


FIGURE 13
The voltage interval results obtained by SVR, RF, OSM and MCS for IEEE 69 case.

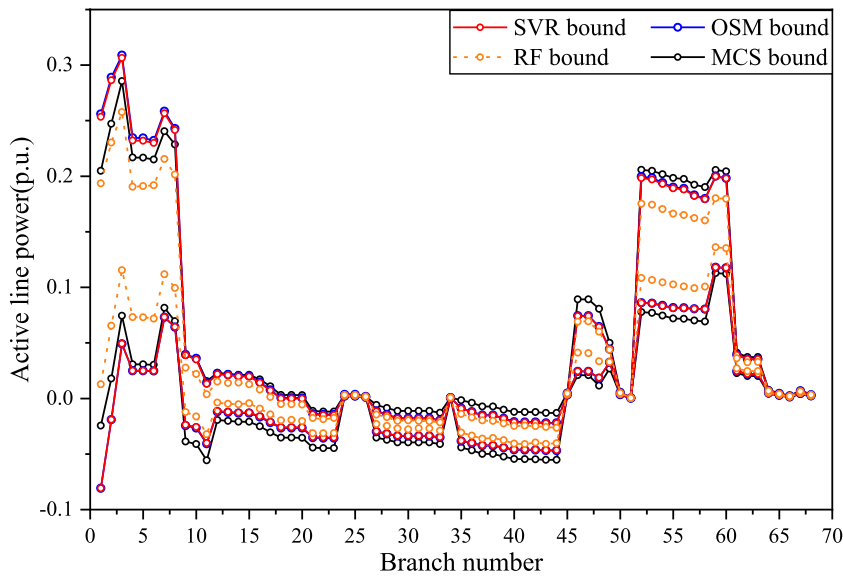


FIGURE 14
The active line power interval results obtained by SVR, RF, OSM and MCS for IEEE 69 case.

model training is generated through OSM, and the OSM takes into consideration of the extreme scenarios that are ignored by the MCS method. It can be seen from Figure 7 that the interval ranges of active line transmission power obtained by the three methods are relatively close. This is because the line transmission power is related to power generation, load demand, and line parameters, and the

Distflow model for the distribution network is linear, so that the active line power results obtained by different methods are close under the same interval input values. The simulation results indicate that the established SVR-based IPF prediction method possesses high predictive accuracy and performs a strong adaptability to different fluctuations.

TABLE 1 The original power generation data of DGs for IEEE 33bw case (p.u.).

Bus number	Active power generation P_G	Reactive power generation Q_G
2	0.0800	0.0600
4	0.0570	0.0500
7	0.0510	0.0440
15	0.0420	0.0380
21	0.0380	0.0340
25	0.0600	0.0500
28	0.0400	0.0320
32	0.0490	0.0450

5.1.2.2 The simulation under scenario 2

In Scenario 2, the original active power generation data was listed in Table 2, and the fluctuation coefficients were set as $\sigma_1 = 0.2$, $\sigma_2 = 0.1$. Thus, a new set of input eigenvector $X_{in,II}$ was introduced. The parameters of SVR model and the sample size of MCS remain the same as (I).

The interval bound results obtained by the SVR, OSM, and MCS for the voltage magnitudes of nodes and the active transmission power of branches for IEEE 33bw case in Scenario 2 are presented in Figures 8, 9, respectively. Similarly, the voltage interval boundaries of nodes No. 4 and No. 5 are selected in Figure 8 for comparison with the MCS sampling results. The SVR is observed to have acquired a similar voltage range to OSM, which is wider than that of MCS. The active line power interval bounds acquired by SVR are close to that obtained by OSM and MCS. These results show that the proposed method also has high precision under scenario 2, which proves that the SVR-based IPF prediction model can adapt to different operating points.

5.1.2.3 The simulation under scenario 3

In Scenario 3, the original active power generation data was the same as that in Table 2, and the fluctuation coefficients were set as $\sigma_1 = 0.1$, $\sigma_2 = 0.05$. Thus, a new set of input eigenvector $X_{in,III}$ was introduced. The parameters of SVR model and the sample size of MCS remain the same as (I). The simulation results under this scenario are demonstrated as follows.

Scenario 3 was set up to verify the accuracy of the proposed algorithm when the operating points and fluctuation ranges change simultaneously. The interval bounds of voltage and active line power obtained by SVR, OSM, and MCS in Scenario 3 are presented in Figures 10, 11. Besides, the voltage boundaries of nodes No. 6 and No. 7 obtained by SVR and OSM are also compared with the MCS sampling results in Figure 10. It can be observed that the voltage ranges obtained by SVR and OSM are still very close, which are more conservative than those obtained by MCS. The active line power ranges acquired by the three methods remain close. The simulation results are expected. Furthermore, compared to scenario 2, the voltage and active line power interval ranges obtained by the three methods are both smaller. This is because the fluctuation ranges are reduced while the operating points remain still. The simulation in scenario 3 validates that the proposed algorithm can

TABLE 2 The original power generation data of DGs for IEEE 33bw case in Scenario 2 (p.u.).

Bus number	Active power generation P_G	Reactive power generation Q_G
2	0.0720	0.0600
4	0.0545	0.0500
7	0.0490	0.0440
15	0.0355	0.0380
21	0.0320	0.0340
25	0.0515	0.0500
28	0.0335	0.0320
32	0.0405	0.0450

maintain high prediction accuracy under different operating points and fluctuation ranges.

In summary, based on simulations under different scenarios, the proposed SVR-based IPF prediction model can adapt to various operational states and environmental fluctuations. In different operating scenarios, this method achieves prediction accuracy comparable to the OSM which is model-driven. Besides, the SVR method provides a more conservative interval range than MCS, which ensures distribution system security under high-dimensional uncertainty. It demonstrates high computational accuracy and strong adaptability of the proposed approach.

5.2 IEEE 69 case study

The IEEE 69 case is applied to validate the efficiency of the proposed SVR-based IPF prediction method. The distribution network is enhanced to include eight DGs. The topology of enhanced IEEE 69 case is presented in Figure 12 and the original active and reactive power generation of DGs are shown in Table 3. All parameters are valued in p.u., and the base power is set to 10 MVA. The voltage limits of all nodes except the slack bus are constrained to $[0.9, 1.1]$.

In IEEE 69 case, 500 sets of training data were generated under the condition of fluctuations with $\sigma_1 = 0.2$, $\sigma_2 = 0.1$. The model training parameters were set as $C = 5,000$, $\epsilon = 0.0001$, $\xi_i = 1$, and the kernel function is selected as $K(x_i, x_j) = x_i^T x_j$. After the model was trained, the predictions were conducted under Scenario 3 as defined in Section 5.1.2. To further validate the model's applicability, the generator operating points were randomly selected within $\pm 30\%$ of the original active power generation data. What's more, the fluctuation coefficients for power generation and load demand were set to $\sigma_1 = 0.3$, and the fluctuation coefficient for line parameters was set to $\sigma_2 = 0.15$, which aims to assess the model's adaptability under expanded fluctuation ranges. This case was carried out with SVR, OSM, and MCS as well.

To further demonstrate the advantage of the proposed SVR-based IPF method, this case additionally incorporated the Random Forest (RF) method for interval power flow prediction. To balance both prediction accuracy and efficiency, the parameters for training the RF model were set as follows: the number of decision trees was

TABLE 3 The original power generation data of DGs for IEEE 69 case (p.u.).

Bus number	Active power generation P_G	Reactive power generation Q_G
2	0.0500	0.0400
7	0.0370	0.0300
12	0.0410	0.0340
24	0.0320	0.0280
34	0.0280	0.0240
45	0.0400	0.0400
49	0.0300	0.0220
59	0.0390	0.0350

set to 100, the minimum leaf size was set to 5, and the model was configured as a regression model. Besides, the system uncertainty parameters were consistent with those described above. Considering the above all, the simulation results are presented in Figures 13, 14.

Figures 13, 14 show the voltage interval ranges and active line transmission power ranges, respectively. The voltage bounds of nodes No. 13 and No. 14 obtained by SVR, RF and OSM are depicted in Figure 13 compared with the MCS samples. It can be observed that under large-scale fluctuations, the voltage ranges obtained by the SVR and OSM are relatively close, and the error precision is determined to be 0.001 upon calculations. The error is mainly attributed to the insufficient size of the training dataset, which can be mitigated by increasing the number of training samples. However, the two methods yield very close active transmission power ranges with high prediction accuracy. Furthermore, compared to the MCS, SVR and OSM obtain wider voltage ranges, as explained in 5.1.2. This case study validates the adaptability of the proposed method to different networks and their ability to handle large fluctuation ranges.

Comparing the SVR method proposed in this paper with the RF method, the SVR method achieves higher prediction accuracy. It is evident that the prediction error using the RF method is relatively large, with an error precision of only 0.01, which shows a significant deviation from the interval results obtained by the OSM. Additionally, the interval obtained by the RF method is narrower, possibly because the predictions of the trees in the model are more concentrated and less flexible in handling extreme cases. Meanwhile, for the RF method, improving prediction accuracy requires increasing the number of decision trees, but this comes at the cost of increased computation time. Through multiple experiments, the accuracy gain from adding more decision trees was found to be negligible.

To validate the efficiency of the proposed method, the computation time of the algorithm compared to RF and OSM is shown in Table 4. The computation time includes the total time for solving voltage and line transmission power. It is noticeable that the online computation speed of the SVR-based IPF prediction method and RF prediction method is significantly faster than the OSM. Meanwhile, the online computation speed of SVR is also faster than that of RF. Besides, the offline training of SVR and RF requires more time compared to the OSM computation, and the offline training times of SVR and RF are comparable. However, the

TABLE 4 The computation time of SVR prediction method, RF prediction method, and OSM for IEEE 69 case.

	Online computation time/s	Offline training time/s
SVR prediction	0.0162	501.58
RF prediction	2.18	553.36
OSM	317.55	—

training and OSM both require significant amounts of time, which increases as the number of system nodes grows. In contrast, the online computation time of SVR is minimally affected by the system scale, and predictions are conducted based on the trained results in practical applications, so that the online computation time is more crucial.

The comparison demonstrates that the proposed SVR approach achieves a significant improvement in computational efficiency over model-driven approaches and is more suitable for large-scale systems. Additionally, the SVR approach has advantages over the RF method in both computational accuracy and efficiency, demonstrating that it is more suitable for IPF analysis compared to other data-driven methods. In summary, the proposed SVR approach is more suitable for rapid and real-time PF analysis of distribution networks with DGs.

6 Conclusion

To address the uncertainty in PF and overcome the efficiency challenges faced by traditional model-driven methods, an SVR-based IPF prediction method for PF analysis in distribution networks is proposed through combining data-driven methods with interval theory. This method considers uncertainty as intervals and employs SVR for model training. The training data is generated through simulation of the established IPF model for distribution network including the intervals of node power injections, line parameters, and the minimum/maximum PF variables. Then the input and output feature vectors for IPF are constructed and the multi-output SVR-based IPF prediction model is established based on the training dataset. To assess the performance of the proposed method, several simulations are conducted both on IEEE 33bw case and IEEE 69 case.

The simulation results show that the proposed method has a good performance. Firstly, the evaluation metrics are calculated to demonstrate the method's high accuracy. Additionally, the proposed method is compared with OSM and MCS in three different scenarios, showcasing robust adaptability across different distribution network cases, operating points, and input data fluctuation ranges. The comparison of interval results obtained by SVR prediction and OSM demonstrates that the SVR approach can achieve prediction accuracy comparable to that of model-driven methods. Meanwhile, the comparative analysis of computation time with the OSM and RF demonstrates that the proposed approach significantly improves computational efficiency compared to model-driven approaches and offers better prediction accuracy and efficiency compared to other data-driven methods. In conclusion, the proposed method exhibits superior computational efficiency and accuracy, meeting the requirements for

handling power flow uncertainty and achieving real-time rapid PF analysis in distribution networks.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

Author contributions

XL: Conceptualization, Methodology, Writing—original draft. HZ: Data curation, Investigation, Writing—review and editing. QL: Methodology, Validation, Writing—original draft, Writing—review and editing. ZL: Software, Writing—review and editing. HL: Investigation, Writing—original draft.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work

References

- Barboza, L. V., Dimuro, G. P., and Reiser, R. S. (2004). "Towards interval analysis of the load uncertainty in power electric systems," in *Proceedings of the international conference on probabilistic methods applied to power systems (ICPMAPS), ames, United States, 12-16 september 2004*, 538–544.
- Cao, Y., Zhou, B., Chung, C. Y., Wu, T., Zheng, L., and Shuai, Z. (2024). A coordinated emergency response scheme for electricity and watershed networks considering spatio-temporal heterogeneity and volatility of rainstorm disasters. *IEEE Trans. Smart Grid.* 15 (4), 3528–3541. doi:10.1109/TSG.2024.3362344
- Chen, J., Li, W., Wu, W., Zhu, T., Wang, Z., and Zhao, C. (2020). "Robust data-driven linearization for distribution three-phase power flow," in *Proceedings of the IEEE 4th conference on energy internet and energy system integration (EI2)*, 1527–1532. Wuhan, China, 30 October–01 November 2020.
- Chen, J., Wu, W., and Roald, L. A. (2022). Data-driven piecewise linearization for distribution three-phase stochastic power flow. *IEEE Trans. Smart Grid.* 13 (2), 1035–1048. doi:10.1109/TSG.2021.3137863
- Chen, Y., Wu, C., and Qi, J. (2022). Data-driven power flow method based on exact linear regression equations. *J. Mod. Power Syst. Clean. Energy.* 10 (3), 800–804. doi:10.35833/MPCE.2020.000738
- Cheng, S., Zuo, X., Yang, K., Wei, Z., and Wang, R. (2023). Improved affine arithmetic-based power flow computation for distribution systems considering uncertainties. *IEEE Syst. J.* 17 (2), 1918–1927. doi:10.1109/JSYST.2022.3176461
- Dong, M., Wiebe, D., and Shi, J. (2022). "An accelerated and risk-free AC power flow method with machine learning based initiation," in *Proceedings of the IEEE electrical power and energy conference (EPEC)*, 103–108. Victoria, Canada, 05–07 December 2022.
- Fu, X., Zhang, C., Xu, Y., Zhang, Y., and Sun, H. (2024). Statistical machine learning for power flow analysis considering the influence of weather factors on photovoltaic power generation. *IEEE Trans. Neural Netw. Learn. Syst.*, 1–15. doi:10.1109/TNNLS.2024.3382763
- Guo, L., Zhang, Y., Li, X., Wang, Z., Liu, Y., Bai, L., et al. (2022). Data-driven power flow calculation method: a lifting dimension linear regression approach. *IEEE Trans. Power Syst.* 37 (3), 1798–1808. doi:10.1109/TPWRS.2021.3112461
- Hu, X., Hu, H., Verma, S., and Zhang, Z. L. (2021). Physics-guided deep neural networks for power flow analysis. *IEEE Trans. Power Syst.* 36 (3), 2082–2092. doi:10.1109/TPWRS.2020.3029557
- Jia, M., and Hug, G. (2023). "Overview of data-driven power flow linearization," in *Proceedings of the IEEE belgrade PowerTech*, 1–6. Belgrade, Serbia, 25–29 June 2023.
- Khalid, H. M., Muyeen, S. M., and Kamwa, I. (2022). An improved decentralized finite-time approach for excitation control of multi-area power systems. *Sustain. Energy Grids Netw.* 31 31, 100692. doi:10.1016/j.segan.2022.100692

was supported by the Science and Technology Project of China Southern Power Grid (090000KK52222133/SZKJXM20222115).

Conflict of interest

Authors XL, HZ, ZL, HL were employed by Shenzhen Power Supply Co., Ltd.

The remaining author declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The authors declare that this study received funding from China Southern Power Grid. The funder had the following involvement in the study: data curation, investigation, the study methodology, data analysis, and writing-review/editing.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Leng, S., Liu, K., Ran, X., Chen, S., and Zhang, X. (2020). An affine arithmetic-based model of interval power flow with the correlated uncertainties in distribution system. *IEEE Access* 8, 60293–60304. doi:10.1109/ACCESS.2020.2982928
- Li, P., Wu, W., Wang, X., and Xu, B. (2023). A data-driven linear optimal power flow model for distribution networks. *IEEE Trans. Power Syst.* 38 (1), 956–959. doi:10.1109/TPWRS.2022.3216161
- Liang, Z., Dong, Z., Li, C., Wu, C., and Chen, H. (2023). A data-driven convex model for hybrid microgrid operation with bidirectional converters. *IEEE Trans. Smart Grid.* 14 (2), 1313–1316. doi:10.1109/TSG.2022.3193030
- Liu, J., Yang, Z., Zhao, J., Yu, J., Tan, B., and Li, W. (2022). Explicit data-driven small-signal stability constrained optimal power flow. *IEEE Trans. Power Syst.* 37 (5), 3726–3737. doi:10.1109/TPWRS.2021.3135657
- Liu, Y., Li, Z., and Zhao, J. (2022). Robust data-driven linear power flow model with probability constrained worst-case errors. *IEEE Trans. Power Syst.* 37 (5), 4113–4116. doi:10.1109/TPWRS.2022.3189543
- Liu, Y., Li, Z., and Zhou, Y. (2022). Data-driven-aided linear three-phase power flow model for distribution power systems. *IEEE Trans. Power Syst.* 37 (4), 2783–2795. doi:10.1109/TPWRS.2021.3130301
- Liu, Y., Wang, Y., Zhang, N., Lu, D., and Kang, C. (2020). A data-driven approach to linearize power flow equations considering measurement noise. *IEEE Trans. Smart Grid.* 11 (3), 2576–2587. doi:10.1109/TSG.2019.2957799
- Liu, Y., Xu, B., Botterud, A., Zhang, N., and Kang, C. (2021). Bounding regression errors in data-driven power grid steady-state models. *IEEE Trans. Power Syst.* 36 (2), 1023–1033. doi:10.1109/TPWRS.2020.3017684
- Lyu, C., Sheng, W., Liu, K., and Dong, X. (2023). Novel affine power flow method for improving accuracy of interval power flow data in cyber physical systems of active distribution networks. *CSEE J. Power Energy Syst.* 9 (5), 1881–1892. doi:10.17775/CSEEJPES.2020.07040
- Mezghani, I., Misra, S., and Deka, D. (2020). Stochastic AC optimal power flow: a data-driven approach. *Electr. Power Syst. Res.* 189, 106567. doi:10.1016/j.eprs.2020.106567
- Mori, H., and Yuihara, A. (1999). "Calculation of multiple power flow solutions with the Krawczyk method," in *Proceedings of the IEEE international symposium on circuits and systems (ISCAS)*, 94–97. Orlando, USA.
- Musleh, A. S., Khalid, H. M., Muyeen, S. M., and Al-Durra, A. (2019). A prediction algorithm to enhance grid resilience toward cyber attacks in WAMCS applications. *IEEE Syst. J.* 13 (1), 710–719. doi:10.1109/JSYST.2017.2741483
- Rehman, A. U., Ullah, Z., Qazi, H. S., Hasani, H. M., and Khalid, H. M. (2024). Reinforcement learning-driven proximal policy optimization-based voltage control for

PV and WT integrated power system. *Renew. Energy*. 227, 120590. doi:10.1016/j.renene.2024.120590

Shao, Z., Zhai, Q., and Guan, X. (2023). Physical-model-aided data-driven linear power flow model: an approach to address missing training data. *IEEE Trans. Power Syst.* 38 (3), 2970–2973. doi:10.1109/TPWRS.2023.3256120

Sun, Y., Zhao, Z., Yang, M., Jia, D., Pei, W., and Xu, B. (2020). Overview of energy storage in renewable energy power fluctuation mitigation. *CSEE J. Power Energy Syst.* 6 (1), 160–173. doi:10.17775/CSEEJPES.2019.01950

Tan, B., Chen, S., Liang, Z., Zheng, X., Zhu, Y., and Chen, H. (2024). An iteration-free hierarchical method for the energy management of multiple-microgrid systems with renewable energy sources and electric vehicles. *Appl. Energy*. 356, 122380. doi:10.1016/j.apenergy.2023.122380

Tan, Y., Chen, Y., Li, Y., and Cao, Y. (2020). Linearizing power flow model: a hybrid physical model-driven and data-driven approach. *IEEE Trans. Power Syst.* 35 (3), 2475–2478. doi:10.1109/TPWRS.2020.2975455

Vaccaro, A., Canizares, C. A., and Villacci, D. (2010). An affine arithmetic-based methodology for reliable power flow analysis in the presence of data uncertainty. *IEEE Trans. Power Syst.* 25 (2), 624–632. doi:10.1109/TPWRS.2009.2032774

Xing, Z., Gong, J., Lao, K. W., and Dai, N. (2021). “Single bus data-driven power estimation based on modified linear power flow model,” in *Proceedings of the 6th international conference on power and renewable energy (ICPRE)* (Shanghai, China), 755–758.

Xing, Z., Lao, K. W., Gao, H., and Dai, N. (2022). “A modified data-driven power flow model for power estimation with incomplete bus data,” in *Proceedings of the 12th international conference on power, energy and electrical engineering (CPEEE)*, 316–320. Shiga, Japan, 25–27 February 2022.

Xue, Y., and Liu, Y. (2021). Intelligent assessment of active and reactive power flow with satisfying accuracy for N-k1-k2 cascading outages. *J. Mod. Power Syst. Clean. Energy*. 9 (5), 986–999. doi:10.35833/MPCE.2020.000312

Zhang, C., Chen, H., Ngan, H., Yang, P., and Hua, D. (2017). A mixed interval power flow analysis under rectangular and polar coordinate system. *IEEE Trans. Power Syst.* 32 (2), 1–1429. doi:10.1109/TPWRS.2016.2583503

Zhang, C., Chen, H., Shi, K., Qiu, M., Hua, D., and Ngan, H. (2018). An interval power flow analysis through optimizing-scenarios method. *IEEE Trans. Smart Grid*. 9 (5), 5217–5226. doi:10.1109/TSG.2017.2684238

Zhang, C., Liu, Q., Zhou, B., Chung, C. Y., Li, J., Zhu, L., et al. (2023). A central limit theorem-based method for DC and AC power flow analysis under interval uncertainty of renewable power generation. *IEEE Trans. Sustain. Energy*. 14 (1), 563–575. doi:10.1109/TSTE.2022.3220567

Zheng, X., Khodayar, M. E., Wang, J., Yue, M., and Zhou, A. (2024). Distributionally robust multistage dispatch with discrete recourse of energy storage systems. *IEEE Trans. Power Syst.*, 1–14. doi:10.1109/TPWRS.2024.3369664