Model-Free Privacy Preserving Power Flow Analysis in Distribution Networks

Dong Liu, Juan S. Giraldo, *Senior Member, IEEE*, Peter Palensky, *Senior Member, IEEE*, and Pedro P. Vergara, *Senior Member, IEEE*

***Abstract*—With the increasing availability of smart meter (SM) data and the frequent lack of accurate network topology information, model-free power flow (PF) calculation has gained traction, often leveraging artificial neural networks (ANNs). However, training such models typically requires large volumes of SM data, raising significant privacy concerns for households in distribution networks. To address this challenge, we propose a privacy-preserving PF calculation framework that incorporates two local privacy-enhancing mechanisms: a Local Randomisation Strategy (LRS) and a Zero-Knowledge Proof (ZKP)-based data collection strategy. The LRS provides irreversible transformation of power data, ensuring strong privacy protection while preserving data utility. In parallel, the ZKP-based strategy enables secure and trustworthy voltage data collection, allowing smart meters to interact with distribution system operators without disclosing actual voltage magnitudes. To address performance degradation caused by seasonal variations in load profiles, we further integrate an incremental learning strategy into the online application. Extensive evaluations across three datasets demonstrate that the proposed framework can efficiently collect one month of SM data within one hour while maintaining most voltage magnitude estimation errors lower than 0.01 p.u. under varying measurement noise and seasonal conditions.**

***Index Terms*—Distribution network, power flow, privacy-preserving, zero-knowledge proof, local randomization.**

I. Introduction

POWER flow (PF) calculation is an essential tool for the operation and planning of distribution networks (DNs). Traditional PF calculation requires the parameters of the network topology, which is not normally available, especially in low-voltage distribution networks (LVDNs) [1]. To address this issue, model-free PF approaches have been proposed based on deep learning models such as artificial neural networks (ANNs) [2], [3]. By training on extensive smart meter (SM) data (e.g., active power, reactive power, and voltage magnitude) that encompass a wide range of operational scenarios, ANNs achieve high estimation accuracy for voltage magnitude, angle and missed measurements [4]. However, distribution system operators (DSOs) may not have access to the time series SM data of all customer data in LVNDs due to the law and privacy concerns, especially in Europe [5]. For instance, DSOs in the Netherlands are typically limited to accessing aggregated load data, such as monthly or quarterly totals, or high-resolution aggregated data without SM identifiers to protect consumer privacy [6]. SM data ownership resides with the end-users. DSOs may access personal SM data only through explicit user consent, typically formalised via a service agreement.

Moreover, the collection and analysis of SM data may expose sensitive personal information, such as the brand of electronic devices (e.g., laptop, refrigerator), household appliance usage patterns, and even financial status [7], [8]. Additionally, detailed load profiles can reveal user behavior patterns, including typical meal times, coffee preparation routines, and whether the user is working from home. Privacy concerns extend further to social network inference and mobility patterns. For example, by analysing electric vehicle (EV) charging behaviors–especially when combined with spatiotemporal data–it may be possible to infer a user's workplace, habitual travel patterns, or even close social ties (e.g., frequent overnight charging near the same external household) [9]. These privacy risks significantly hinder the willingness of users to share SM data, thereby limiting the scalability and practical deployment of model-free PF methods.

Traditional encryption algorithms, like symmetric encryption, can safeguard SM data during transmission [10], [11]. However, decrypted data is essential for complex calculations like optimization problems and PF calculations. This decryption exposes household privacy to potential cyber threats at the control centre. Existing privacy-preserving approaches for measurement sharing in DNs could be roughly divided into equipment-based and algorithm-based categories. Equipment-based approaches, such as using house energy storage systems (ESSs) to alter shared data, can introduce variations that obscure real distributions of SM data [12]. Specifically, by charging or discharging the ESSs, users can reshape the load curve of devices like a Laptop. Hence, the extracted pattern no longer matches the actual usage, reducing appliance identification risk. Nevertheless, the high installation cost of ESSs constrains its applicability for privacy protection and is mainly used for energy storage. Moreover, new algorithms

Received 3 March 2025; revised 27 May 2025; accepted 21 July 2025. Date of publication 28 July 2025; date of current version 23 October 2025. This work was supported by the China Scholarship Council under Grant 202206130017. Paper no. TSG-00371-2025. *(Corresponding author: Pedro P. Vergara.)*

Dong Liu, Peter Palensky, and Pedro P. Vergara are with the Intelligent Electrical Power Grids Group, Delft University of Technology, Delft 2628 CD, The Netherlands (e-mail: p.p.vergarabarrios@tudelft.nl).

Juan S. Giraldo is with the Techno-Economic Energy Transition Studies Group, Netherlands Organisation for Applied Scientific Research, The Hague 2509 JE, The Netherlands.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TSG.2025.3593249.

Digital Object Identifier 10.1109/TSG.2025.3593249

or indicators are needed to trade off data privacy and utility, enabling the altered data to be compatible with existing algorithms at the DSO.

Algorithm-based approaches integrate homomorphic encryption (HE), data aggregation and randomization. A fully HE approach supports homomorphic addition and multiplication, i.e., the decrypted results obtained based on encrypted data is the same as the results obtained based on unencrypted data. This characteristic of HE enables SM to share the encrypted data with the DSO to participate in the calculation and receive the encrypted results. A privacy protection strategy based on partial HE was proposed in [13] to ensure the privacy of the updated parameters during alternating optimization in optimal power flow (OPF) calculation. HE was combined with an aggregation strategy to ensure household privacy in [14], which does not result an intense calculation burden for SM. To ensure privacy in energy management systems, distributed optimization models with HE were proposed and their convergence was verified in [15]. However, the traditional HE algorithm is computationally intensive and only supports homomorphic addition and multiplication, which does not guarantee the accuracy and convergence of the classical PF model. In addition, a Newton method-based PF analysis is conducted in a distributed environment, where secure multiparty computation was utilized to ensure privacy, while this approach requires details of the network topology [16].

Noise disturbance is a straightforward and effective strategy for data randomization, making them widely utilized in data privacy protection. Differential Privacy (DP), a leading noise disturbance strategy, is commonly used to enhance the privacy of data [17]. A Laplace noise disturbance strategy is employed to enhance the privacy level and accuracy of distributed dispatch problems [18]. In [19], a local randomization approach was introduced to independently ensure the privacy of users and noise disturbance was employed to enhance the privacy level. Two low-cost and lightweight DP strategies were introduced to reduce the expense of ESSs-based privacy-preserving approaches in [20]. To trade off the privacy level and accuracy loss induced by injected noise, the results in [21] show that the OPF model with DP strategy should be further remodelled. Instead of injecting noise into network parameters and inputs, OPF variables are modelled as a function of random noises, guaranteeing the feasibility and privacy level of the model [22]. Nevertheless, DP is applied to ensure privacy in OPF and distributed optimization problems in the above papers, and its application in PF calculations is quite rare. This is primarily due to the differences between the two tasks: OPF focuses on decision-making, where moderate data perturbations typically preserve solution optimality. In contrast, PF aims to accurately estimate voltage based on real-time power data, making it more sensitive to input distortions, especially for the model-free PF calculation. As a result, directly applying DP in PF calculation can lead to convergence issues or intolerable estimation errors.

DP strategies can also be integrated into ANNs to ensure the privacy of training datasets, e.g., backwards gradient with noises [17]. Nevertheless, several limitations hinder their application for PF calculation and are summarized as 1) The generalization of ANN is inherently limited, posing a challenge in achieving a similar accuracy when an ANN is trained on original data and data with noise [23]. 2) Adding noise (e.g., Laplace or Gaussian noise) to SM data might cause two different data points to be the same value, affecting the accuracy of PF calculation. Ensuring consistent monotonicity before and after data encryption is essential to maintain the accuracy of ANN-based PF calculation. 3) DP requires that adjacent datasets differ by at most one sample. While OPF methods can satisfy this requirement by focusing solely on critical loads, PF calculations must ensure privacy protection for every SM data point, not just critical loads, due to their direct impact on state estimation accuracy [24]. In summary, the privacy-preserving strategy for model-free PF calculation should trade off privacy level, data utility and accuracy.

Zero-knowledge proof (ZKP) is widely used in scenarios where privacy and confidentiality are paramount, such as authentication protocols and privacy-preserving blockchain technologies [25]. ZKP enables the verification of transactions without revealing transaction details, enhancing privacy and scalability in the blockchain. A ZKP-based energy trading approach integrates the ZKP into the blockchain to ensure fair operation without cheating participants [26]. To protect privacy in the hybrid energy dispatch, ZKP was adopted to ensure the accuracy of the profit variation, aiming to fairly and economically distribute relocated profit in [27]. A privacy-preserving strategy was proposed based on blockchain and ZKP in [28] for safely authentication the connected vehicle in a distributed environment other than central authority. Besides, variants of ZKP extend its usability by eliminating the need for interaction between the prover and verifier, such as Zero-knowledge succinct non-interactive arguments of knowledge. Compared to above approaches, ZKP offers a potential solution for privacy preservation in model-free PF calculation. It allows SMs to prove data validity without revealing the actual value, with minimal computation and no need for mutual communication or trusted third parties. Nevertheless, the application of ZKP in privacy-preserving and dataset-sharing in DNs is still in its early stages.

On the other hand, the statistical distribution of SM data varies across seasons due to differences in the use of electrical devices and energy consumption behaviours, such as the increased use of heating pumps during winter and the charging patterns of ESSs. These seasonal variations impact the accuracy of the ANN-based PF analysis. Re-training the ANN with updated data to maintain accuracy introduces substantial computational overhead, and the accuracy of the updated ANN on old datasets may significantly decrease. Incremental learning (IL) strategies are proposed to mitigate catastrophic forgetting during ANN updating by preserving knowledge acquired from old data while learning new information from the latest datasets. IL strategies can be broadly categorized into three clusters: Architecture-based, Regularization-based, and Rehearsal-based [29], [30]. Nevertheless, IL strategies have been extensively applied in classification tasks, and their

application in enhancing the generalization of ANN-based PF approaches in DNs remains relatively underexplored.

To address this research gap, this paper introduces a privacy-preserving PF calculation framework, composed of two local strategies and an IL strategy. The proposed framework is particularly suitable for scenarios where SM data access is constrained by user consent or regulatory frameworks, such as those aligned with the General Data Protection Regulation (GDPR) [31]. The main contributions of this paper can be summarized as follows:

- A local privacy-preserving strategy is introduced to share active and reactive power data with DSOs for PF calculation. The proposed method employs modified sigmoid and randomized functions to irreversibly randomize power data without relying on third-party entities or DP frameworks, ensuring data confidentiality at the SM side.
- A ZKP-based data collection mechanism is proposed to enable the secure acquisition of voltage magnitude datasets for model training. This approach allows SM to verify the correctness and utility of the data while preserving the privacy of actual voltage values.
- To enable PF analysis based on privacy-transformed power data, a data-driven approach is required–one that can effectively learn the relationship between the transformed SM data and voltage magnitude. Therefore, a model-free PF analysis framework is adopted to accommodate the nonlinearity and uncertainty introduced by the data transformation process.
- To mitigate the accuracy degradation caused by seasonal variations in load profiles, an IL strategy is employed to periodically update the ANN model. The updates are triggered based on a Wasserstein Distance (WD)-based indicator, allowing the model to adapt dynamically to temporal changes and maintain PF calculation accuracy.

The remainder of this paper is organised as follows: Section II provides an overview of ANN-based power flow calculations, ZKP, and the IL strategy. Section III details the proposed framework, including the local privacy-preserving strategy, ZKP-based data collection process, ANN architecture, and the IL updating mechanism of the trained ANN. Section IV presents the case studies and discusses the parameter settings. Finally, Section V offers the concluding remarks and the limitations for application.

II. Preliminaries

A. Model-Free Power Flow Model

Power flow calculation aims to calculate the unknown variables of each bus in the distribution networks, including reactive power, active power, voltage magnitude and phase angle, depending on the type of bus. The relation among these four variables is formulated as expressions (1) and (2). To calculate the PF in a DN with N buses and no distributed generators, there are $2N$ unknown variables–typically the voltage magnitudes and phase angles at each bus. To determine the unknown values, it is necessary to formulate and solve a corresponding set of $2N$ nonlinear equations that represent the power balance at each bus.

$$P_n = V_n \sum_{m=1}^{N} V_m (G_{nm} \cos\theta_{nm} + B_{nm} \sin\theta_{nm}) \tag{1}$$

$$Q_n = V_n \sum_{m=1}^{N} V_m (G_{nm} \sin\theta_{nm} - B_{nm} \cos\theta_{nm}) \tag{2}$$

where G_{nm} and B_{nm} are the real and imaginary parts of the element in the bus admittance matrix Y_{bus} at position (n, m), which represent the electrical coupling between node n and node m based on the network topology and parameters. P_n, Q_n and V_n are the injected active power, injected reactive power and voltage magnitude at node n, respectively. $\theta_{nm} = \theta_n - \theta_m$ is the phase angle difference between nodes n and m.

The Newton-Raphson method is a widely adopted iterative technique for solving PF equations. However, it requires detailed knowledge of DN topology, including line connections, cable lengths, and impedance parameters. Such information is normally unavailable or incomplete in practical DNs [32]. To overcome this limitation, model-free PF approaches have been introduced, which aim to learn the nonlinear mapping between measured inputs and unknown variables directly from historical data without relying on explicit network models [33].

By leveraging the nonlinear mapping capabilities of ANN, voltage magnitude and phase angle could be estimated from active and reactive power. According to (1) and (2), the inputs to the ANN could be the active power P_n and reactive power Q_n at each node. The outputs are the voltage magnitude V_n and phase angle θ_n. The estimation of voltage magnitude and angle are formulated as (3) and (4).

$$V = F_{\psi 1}(P_1, P_2, \ldots, P_N, Q_1, Q_2, \ldots, Q_N) \tag{3}$$

$$\theta = F_{\psi 2}(P_1, P_2, \ldots, P_N, Q_1, Q_2, \ldots, Q_N) \tag{4}$$

$$[V_1, \ldots, V_N, \theta_i, \ldots, \theta_N] = F_{\psi 3}(P_1, \ldots, P_N, Q_1, \ldots, Q_N) \tag{5}$$

where $F_{\psi 1}$, $F_{\psi 2}$ and $F_{\psi 3}$ are the ANNs mapping the given SM data to voltage magnitude, angle and both, respectively.

To train an ANN-based PF model, a loss function commonly used is the mean squared error (MSE) $\mathcal{L}_{\text{MSE}}$ between the estimated and actual values, which are formulated as follows:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N} \sum_{n=1}^{N} \left[\alpha_V \left(V_n - \widetilde{V}_n\right)^2 + \alpha_\theta \left(\theta_n - \widetilde{\theta}_n\right)^2 \right] \tag{6}$$

where $\widetilde{V}_n$ and $\widetilde{\theta}_n$ are the estimated voltage magnitude and angle from ANNs. α_V and α_θ are the weights for the voltage magnitude and phase angles, introduced to ensure balance between terms in the objective function.

The ANN-based approach shows several advantages over traditional methods, including faster computation times, robustness to the error in SM data, and the ability to model complex, nonlinear relationships for DNs with unavailable topology [4].

B. Zero Knowledge Proof

ZKP is a cryptographic protocol that allows one participant (named *prover*) to convince another participant (named *verifier*) that a given statement is true without revealing private information beyond the fact that the statement is true. The statement could be data values, a private calculation, a relationship, etc. A ZKP strategy must satisfy three fundamental properties: completeness, soundness, and zero knowledge. Completeness ensures that if the statement is true and the prover follows the protocol, the verifier will definitely trust the statement. Soundness guarantees that if the statement is false, no participant can convince the verifier, i.e., reject the statement. Zero-knowledge ensures that the verifier learns nothing except that the statement is true; specifically, the verifier gains no knowledge that could help reconstruct the prover secret.

Pedersen Commitment (PC) is a typical cryptographic commitment scheme, which is employed in privacy-preserving protocols and ZKP applications to enhance privacy and security. In ZKP scheme, PC allows the prover to commit to a value while keeping it hidden and later reveal it without altering the commitment. PC scheme is developed based on discrete logarithms to ensure that it is infeasible to change the committed value once it has been set, thus providing binding and hiding properties. Specifically, in ZKP schemes for range or equality proofs, a prover adopts PC to commit to value x_0 and then construct proofs demonstrating knowledge of these values or properties about them (i.e., proving that a committed value falls within a pre-set range or is the pre-set values) without revealing the actual values. PC based on the elliptic curve is formulated as follows:

$$C_{g,h}(x_0, r) = g^{x_0} \cdot h^r \tag{7}$$

where g and h are randomly sampled from the pre-set elliptic curve, respectively. r is a random value within $[1, p]$, aiming to mask the true value. p is a large prime number. The general process of PC-based ZKP is described as:

1) Generate parameters p, g and h.
2) Prover generates an random value r, and then create and send the commit $C_{g,h}(x_0, r)$.
3) Verifier generates a challenge b and sends it to the prover.
4) Prover masks the random value r using (8) and function $mod\{\cdot\}$,[1] and the obtained s is send to the verifier.

$$s = mod\{r + x_0 \cdot (1 - b), p\} \tag{8}$$

5) Verifier checks the result using the private verification function f^*.

$$f^*\big(C_{g,h}(s, r), p, g, h, b, s\big) = \begin{cases} 1, \text{ if TRUE} \\ 0, \text{ if FALSE} \end{cases} \tag{9}$$

Note that PC is homomorphic. This feature facilitates efficient ZKP constructions. This homomorphism enables the verifier to conduct addition directly on commitments, expressed as (10). The formulation and detail of the ZKP process can be found in [27].

$$C_{g,h}(s1, r1) + C_{g,h}(s2, r2) = C_{g,h}(s1 + s2, r1 + r2) \tag{10}$$

[1] *mod* $\{\cdot\}$ is a function that returns the remainder of a division between two numbers.

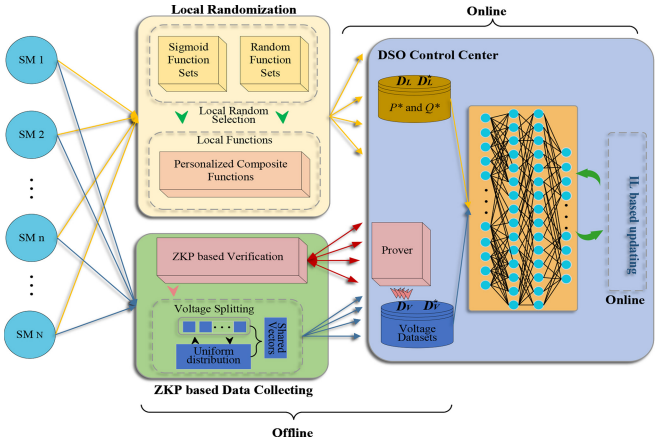


Fig. 1. Framework of ANN-based PF calculation with local privacy-preserving strategy: orange, blue, and red lines represent load transfer, voltage transfer, and the communication channel, respectively.

C. Incremental Learning

IL strategies aim to mitigate catastrophic forgetting during ANN updating by preserving knowledge acquired from previous datasets while learning new information from the latest datasets. Compared to traditional class incremental learning, the IL strategy for PF analysis must handle complex ANN architectures due to high-dimensional output vectors. Architecture-based IL strategy is a widely used approach to preserving knowledge in continual learning scenarios. Within this category, freezing the weights of specific layers or reducing the learning rate are common and effective techniques [30], as illustrated in (12) and (13).

$$\psi^{k-1} = \left[W^{k-1}, B^{k-1}\right] \tag{11}$$

$$\psi_{\mathbf{l}}^{k} = \psi_{\mathbf{l}}^{k-1}, \qquad \forall l \in \mathcal{L} \tag{12}$$

$$\eta^k = \zeta.\eta^{k-1} \tag{13}$$

where $\boldsymbol{l}$ and $\mathcal{L}$ denote the index and set of frozen layers. η is the learning rate. ψ^{k-1} and η^{k-1} are the parameters and learning rate obtained at the end of the $(k-1)$-th update, respectively. ζ is a positive constant satisfying $0 < \zeta < 1$.

III. Privacy Preserving Power Flow Analysis

The proposed model-free PF calculation framework with local privacy-preserving is illustrated in Fig. 1.[2] We assume that a reliable bidirectional communication channel exists between SMs and the DSO [34]. These channels support: (1) offline collection and store of training dataset, and (2) real-time periodic updates of randomised power data for online PF analysis (e.g., 15-minute intervals). In the offline stage, randomized power and voltage magnitude profiles are

[2] The source code associated with this paper is available online at [https://github.com/distributionnetworksTUDelft/Model-Free-Privacy-Preserving-Power-Flow-Analysis-in-Distribution-Networks].

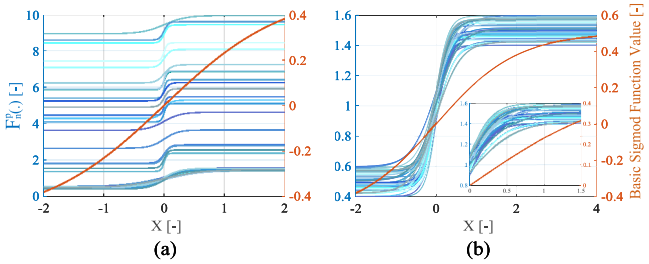


Fig. 2. Nonlinear transformation function without constraints in (a) and with constraints in (b).

collected by the DSO, which serves as the training dataset for the ANN. The privacy of the two datasets is ensured by the proposed *Local Randomization Strategy* (LRS) and the *ZKP-based data collection strategy*, respectively. In the online stage, only randomized power data are shared with the DSO for estimating the voltage magnitude via an ANN. Moreover, an architecture-based IL strategy is employed to update the ANN model upon receiving datasets corresponding to different seasonal variations.

A. Local Randomization Strategy

The local randomization strategy for the active and reactive SM data must ensure the monotonicity, irreversibility and non-negativity of the data to maintain privacy levels and structural accuracy. Motivated by the randomization strategy in DP and the principle that addition operations are irreversible, we propose to locally randomize the active power P and reactive power Q by two personalized composite functions $F_n^p(x)$ and $F_n^q(x)$, which is depicted in Fig. 1. The constructed composite functions consist of a modified sigmoid function and a constant function, represented by expression (14)-(19).

$f_n^p(x)$ and $f_n^q(x)$ are the modified sigmoid functions used to introduce nonlinearity into the active and reactive power data transformation, respectively. The parameter a_n^p represents the unique scaling factor for customer n when randomizing active power, while a_n^q is the corresponding scaling factor for reactive power. Both a_n^p and a_n^q are specific to each user.

$$f_n^p(x) = \frac{1}{1+e^{-a_n^p x}} - 0.5, \qquad \forall n \in \mathcal{N} \tag{14}$$
$$g_n^p(x) = c_n^p, \qquad \forall n \in \mathcal{N} \tag{15}$$
$$F_n^p(x) = f_n^p(x) + g_n^p(x), \qquad \forall n \in \mathcal{N} \tag{16}$$
$$f_n^q(x) = \frac{1}{1+e^{-a_n^q x}} - 0.5, \qquad \forall n \in \mathcal{N} \tag{17}$$
$$g_n^q(x) = c_n^q, \qquad \forall n \in \mathcal{N} \tag{18}$$
$$F_n^q(x) = f_n^q(x) + g_n^q(x), \qquad \forall n \in \mathcal{N} \tag{19}$$

$g_n^p(x)$ and $g_n^q(x)$ represent constant functions, where c_n^p and c_n^q are the locally offset noises for active and reactive power, respectively. Similar to the scaling factors, these constant noises are unique for each user and differ between active and reactive power. $F_n^p(x)$ and $F_n^q(x)$ are the composite functions combining the logistic transformation and the constant noise for active and reactive power, respectively. This strategy ensures that the randomization transforms the power data according to the unique parameters of each user to weaken the inherent relationship between active and reactive power data. An illustrative example for $F_n^p(x)$ and $F_n^q(x)$ is depicted in Fig. 2 (a).

The utility of the transformed power data P^* should be preserved to ensure both the convergence and high accuracy of the ANN-based PF calculation. To achieve privacy protection by obfuscating the data, i.e., ensuring that the transformed data from different users remain indistinguishable, while simultaneously maintaining data utility, constraints (20) and (21) are imposed on the parameters a_n^q and c_n^q. These constraints balance the trade-off between privacy and accuracy, preventing excessive distortion of the original data. An illustrative example demonstrating the structure of the composite functions under these constraints is shown in Fig. 2 (b), indicating the constrainted transformation function will make the SM data different from the original values but close to each other.

$$\underline{a} \le a_n^p, a_n^q \le \overline{a}, \qquad \forall n \in \mathcal{N} \tag{20}$$
$$\underline{c} \le c_n^p, c_n^q \le \overline{c}, \qquad \forall n \in \mathcal{N} \tag{21}$$

where $\underline{a}, \overline{a}, \underline{c}$ and $\overline{c}$ are the upper limits for the parameters in expression (16) and (19), respectively. As illustrated in Fig. 2 (b), the constrained transformation formula effectively concentrates the data distribution, mitigating potential convergence challenges for the ANN.

The $\underline{a}, \overline{a}, \underline{c}$ and $\overline{c}$ can be determined by assessing the sensitivity of ANN to these parameters and subsequently defining the parameter range. The upper and lower limits of the parameters are also set based on the correlation between the initial data and the transformed data. This approach ensures an optimal balance between data privacy and data utility.

The proposed LRS establishes a robust framework for maintaining data utility while safeguarding sensitive information. Each user (i.e., each SM) in the preset region selects their parameters a_n^p, a_n^q, c_n^p, and c_n^p individually and does not share them with other users and the DSO. This individualized selection process significantly enhances data security, as it becomes more challenging to deduce the original data from the encrypted data without knowing the specific parameters used in the randomization process. The received randomized power data are stored D_L at the DSO centre.

B. ZKP-Based Data Collection

The framework of the proposed ZKP-based data collecting strategy is depicted in Fig. 3. In this scheme, the DSO acts as the prover, receiving uniformly distributed vectors $\widetilde{V}$, and proves to the verifier that it knows the index location of the split data. The SM serve as the verifier, sharing only uniformly distributed vectors $\widetilde{V}$ with split voltage magnitudes and verifying the correctness of the DSO's claim.

The SM randomly splits the voltage magnitudes and inserts them into uniformly distributed vectors $\widetilde{V}$, which are then shared with the DSO. An illustrative example is depicted in Fig. 4. For instance, each voltage magnitude is split into two parts, and vector $\widetilde{V}$ is obtained by (22)-(24). The upper and

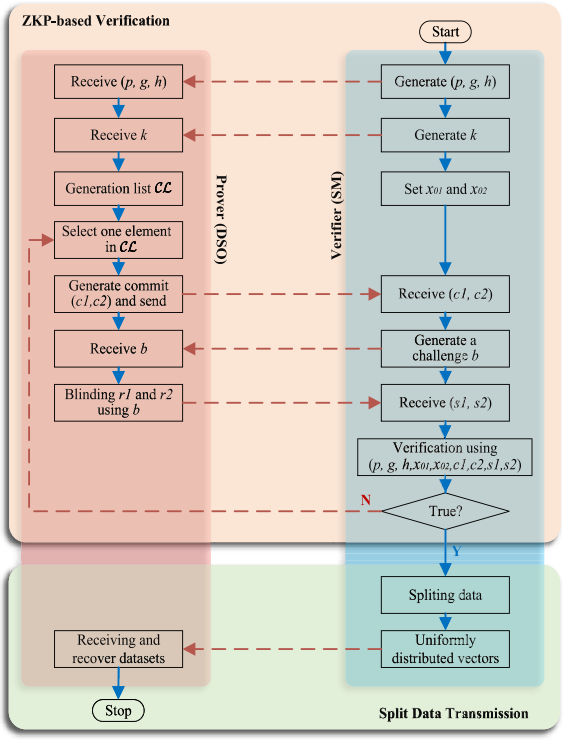


Fig. 3. Framework of the ZKP-based data collection process: red lines indicate communication paths and blue lines illustrate the process steps.

lower limits for $\widetilde{V}$ are set as $\overline{\beta} \cdot \max\{v^*_{n,t,1}, v^*_{n,t,2}\}$ and $\underline{\beta} \cdot \min\{v^*_{n,t,1}, v^*_{n,t,2}\}$.

$$v_{n,t} = v^*_{n,t,1} + v^*_{n,t,2} \tag{22}$$

$$v^*_{n,t,1} = f_r\{\underline{\gamma} \cdot v_{m,n}, \overline{\gamma} \cdot v_{m,n}\} \tag{23}$$

$$\widetilde{V} = \left[v_1, \ldots, v^*_{n,t,2}, \ldots, v^*_{n,t,1}, \ldots, v_I\right] \tag{24}$$

where $\underline{\beta}$ and $\overline{\beta}$ are scalar parameters, with $\overline{\beta} > 1$ and $\underline{\beta} < 1$, used to adjust the scaling range. The parameters $\underline{\gamma}$ and $\overline{\gamma}$ are selected to be close in value (e.g., 0.49 and 0.51) to ensure that the split data points are close to the randomly generated points in $\widetilde{V}$.

The sum of any two data points within these vectors $\widetilde{V}$ remains nearly similar, making it difficult to infer the actual voltage values and thereby enhancing privacy. Furthermore, the SM does not provide any information to the DSO on how to extract the original voltage values. Thus, the DSO needs to verify the correct index of each split voltage in the uniformly distributed vectors $\widetilde{V}$ without revealing the true values (i.e., convincing SM that DSO know the true indexes).

As shown in Fig. 3, the SMs first sets the true indexes of the split voltages locally and only shares the dimension I of the uniformly distributed vectors with the DSO. Then, the DSO generates a list $\mathcal{CL}$ containing all possible combinations of indices for the true split voltages, with only one being the correct solution. The DSO selects one potential solution and generates two PCs $c1$ and $c2$ using (7), corresponding to the chosen indices. These PCs are then sent to the SM. Afterwards, the SM sends a challenge b to the DSO, prompting the DSO to compute s_1 and s_2 using (8) and the received b. The DSO then

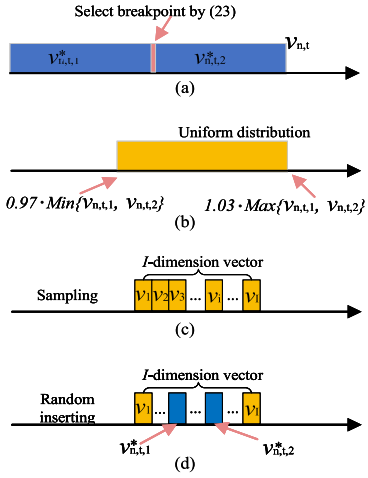


Fig. 4. Illustrative example of uniformly distributed vectors construction.

sends s_1 and s_2 back to the SM. SM independently decrypts them using (25) and obtain r_1 and r_2, respectively,

$$r = mod\{(s - (1 - b) \cdot x_0), p\} \tag{25}$$

The SM generates the two commitments $c1^*$ and $c2^*$ for the true indexes and compares them with the received commitments $c1$ and $c2$. If they are equal, the SM inform the DSO to receive the time-series dataset and terminates the interaction with this SM. The DSO recovers the true voltage datasets with the true indexes and storage in D_V. Each SM will conduct the above process before sharing the split time-series voltage datasets.

The proposed LRS and ZKP-based data collection strategy serve distinct purposes and periods. The LRS is employed for two scenarios: (1) offline collection of randomized power data used to train the ANN-based PF model, and (2) online sharing of real-time randomized power data (e.g., every 15 minutes), which is then used by the trained ANN to estimate voltage magnitudes. In contrast, the ZKP-based data collection strategy is primarily designed for secure acquisition of target data (i.e., voltage magnitudes) during the initial training of the ANN or periodic updating by IL. Thus, the ZKP-based data collection strategy is only activated when updates to the ANNs are required, other than being activated every 15 minutes.

C. ANN-Based Power Flow Calculation

Multilayer Perceptron (MLP) is employed to conduct ANN-based power flow analysis. The training dataset D_L contains as input data the randomized active power P and reactive power Q collected using the proposed *LRS* strategy in Section III-A, while the output dataset contains the voltage magnitudes D_V obtained using the ZKP-based collection strategy in Section III-B. Consequently, the dimension of the output layer is N. The activation function is represented by $\sigma_1 \ldots \sigma_5$. Since we aim to use ANN to estimate voltage magnitude, we set $\alpha_\theta = 0$ and $\alpha_V = 1$ in expression (6). The loss function is defined as expression (26).

$$\mathcal{L}_{MLP} = \frac{1}{N}\sum_{n=1}^{N}\left[\left(V_n - \widetilde{V}_n\right)^2\right] \tag{26}$$

Algorithm 1: Model-Free Privacy Preserving PF Analysis in Distribution Networks

Input: D_L, D_V, D_f, l, η, tt^*,M_1
ANN-2 ← Train ANN using D_V and D_L.
D_f ← *LRS* (P,Q) at each SM
Estimate V using ANN-2 and D_f
if $tt \leq tt^*$ **then**
  D_L^* ← *LRS* (P,Q) at each SM
  D_V^* ← *ZKP(V)* parallelly
  Freeze the last l layers of ANN-2.
  for $i = 1$ *to* M_1 **do**
    Train ANN-2 with η
    $\mathcal{L}_{RMLP}$ ← calculate loss using Eq. (26)
    Backpropagate $\mathcal{L}_{RMLP}$
    ANN-2 ← Update ANN-2
  end
  return ANN-2
end
Return step 2

$$\tilde{V}_n = \sigma_i\Big(W_{i-1}\sigma_{i-1}(\cdots(W_1\sigma_1(D_n)+b_1) + \cdots + b_{i-1})\Big) + b_i \tag{27}$$

where W_i and b_i represent the weights and biases of the hidden layers and will be optimized during the training process. $\tilde{V}_n$ is the estimated voltage magnitude.

The ANN structure is not fixed, allowing the DSO to tailor it to specific network requirements. The proposed privacy-pervasive and the IL strategy can be integrated into existing or future NNs used at the DSO.

D. Online Application

Given datasets D_L and D_V, an ANN is trained and stored at the DSO. When the DSO receives transformed data D_f from SMs that employ *LRS*, the trained ANN is utilized for PF analysis. To mitigate the accuracy loss induced by the seasonal variations, the DSO must periodically update the weights of the ANN. Algorithm 1 provides a detailed description of the complete online application and updating process of the ANN, illustrating how the model evolves to maintain robust performance and accuracy in PF calculations.

The Wasserstein distance (WD) is employed as an indicator to trigger ANN updates. WD quantifies the cost of transforming one distribution into another, making it a suitable metric for assessing distributional differences. The WD values range from 0 to 1, with 0 indicating high similarity and 1 indicating significant dissimilarity. Specifically, the training voltage datasets are divided into N_v chronologically ordered sub-datasets (e.g., ten-day intervals). The WD value W_i between each sub-dataset and the new dataset is calculated. The WD-based indicator, denoted as tt, shown in equation (28).

$$tt = \sum_{i=1}^{N_v} W_i \tag{28}$$

When the indicator tt surpasses a pre-set threshold, tt^*, the ANN updates are triggered. Firstly, multiple weeks of new SM data are collected, denoted as D_L^* and D_V^*. Then, the architecture-based IL strategy, as described in Eq. (12), is used to update the ANN instead of re-training a new model, i.e., trading off the accuracy and computation cost. The IL strategy requires the specification of a pre-set parameter l and a maximum epoch limit $M1$.

IV. Experimental Results

In this section, we verify the feasibility and accuracy of the proposed local privacy-preserving strategies on four ANNs and three LVDNs in The Netherlands denoted as LV-52, LV-64, and LV-95 [32]. A power factor of 0.95 is assumed for all cases and is used in the power flow calculations to generate the corresponding voltage datasets. Three datasets are utilized for the respective LVDNs:

- Dataset I: Time-series load profiles for individual households with a 15-minute resolution in the LV-52 network, selected and scaled from [35].
- Dataset II: Time-series load profiles with a 1-hour resolution, collected from London, UK, used in the LV-64 network.
- Dataset III: Time-series load profiles with a 1-hour resolution, collected from the Netherlands, applied to customers in the LV-95 network.

A. Local Data Randomization

Results on the non-linear transformation of the active power dataset are depicted in Fig. 5, illustrating both the distribution (the second row) and correlation (the third row) before and after the transformation. P^* represents the transformed active power. The first column of results pertains to the original dataset, while the subsequent five columns relate to the transformed data. The upper and lower limits of parameter a_n^p and a_n^q are set as [0, 0.2], [1, 1.5], [3, 8], [8, 10], [0, 50], and for parameter c_n^p and c_n^q, the limits are set as [0, 0.5], [0.5, 1], [0.9, 1], [1.2, 5], [0, 50], represented by Set I to VI, respectively. The first row displays the data distribution. The second row presents the frequency distribution histogram of each sample and the probability density curve. The distribution of the initial dataset aligns closely with a Beta distribution, which contrasts sharply with the distribution observed in the transformed data. This significant change in distribution enhances the privacy-preserving properties of the data transformation. The third row shows the correlation between the transformed data and the original data, indicating the potential for inferring the original data from the transformed data. When the upper limits of parameters (a_n^p, a_n^q c_n^p and c_n^q) increases, the correlation between the transformed datasets becomes less distinct, making it challenging to trace the transformed dataset back to the original dataset, as illustrated in columns 3 to 6. However, a larger upper limit also leads to a broader distribution (i.e., in the first row). The expanded distribution range causes the data to be dispersed across different intervals, deviating from the normalized distribution. Thus, the upper limits for parameters a_n^p and c_n^q are set to [3, 8] and [0.9, 1.1], respectively.

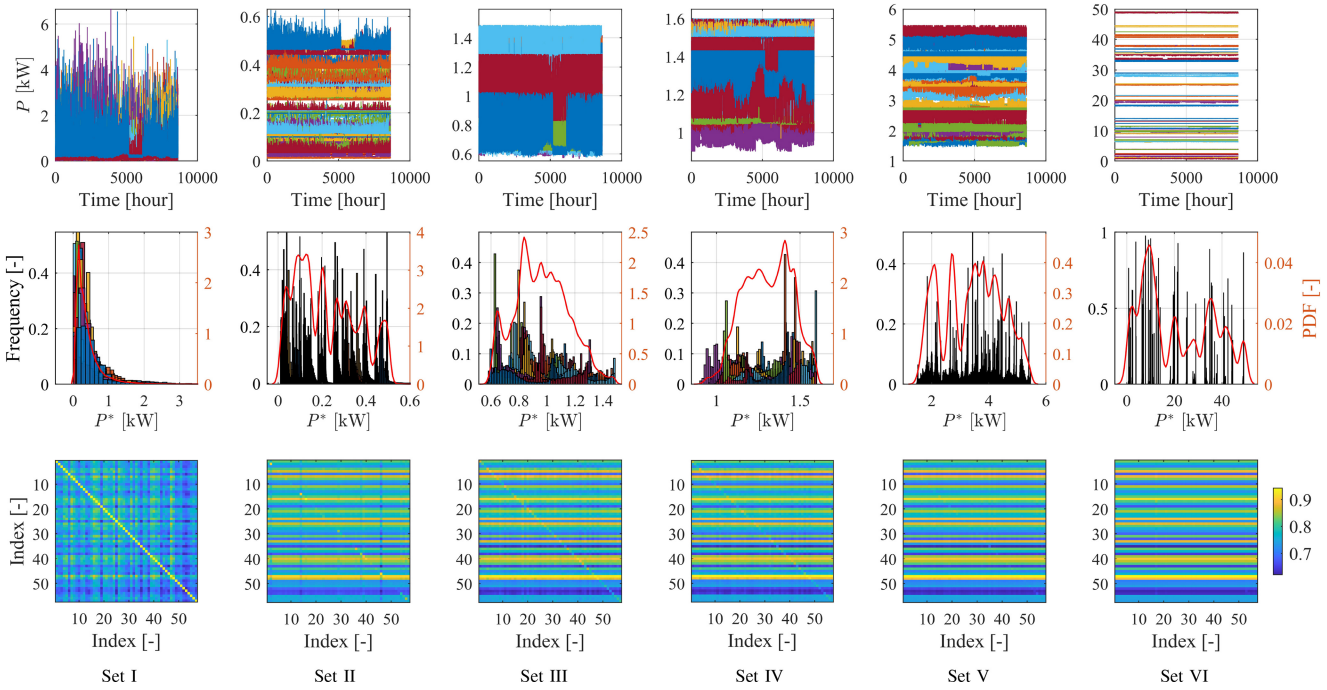


Fig. 5. Distribution of original P and transformed dataset P^* (in the first and second row) and their correlations (in the third row).

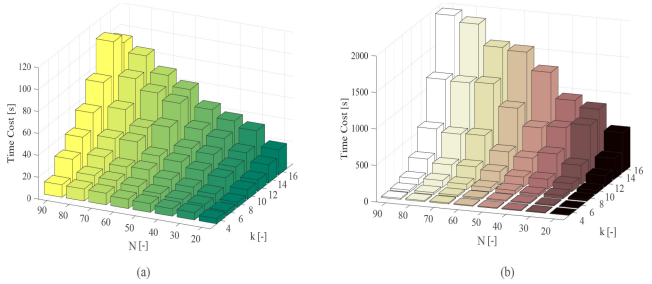


Fig. 6. Time cost of ZKP-based data collection: (a) with data split into 2 sub-parts, and (b) with data split into 4 sub-parts.

B. ZKP-Based Data Collection Efficiency

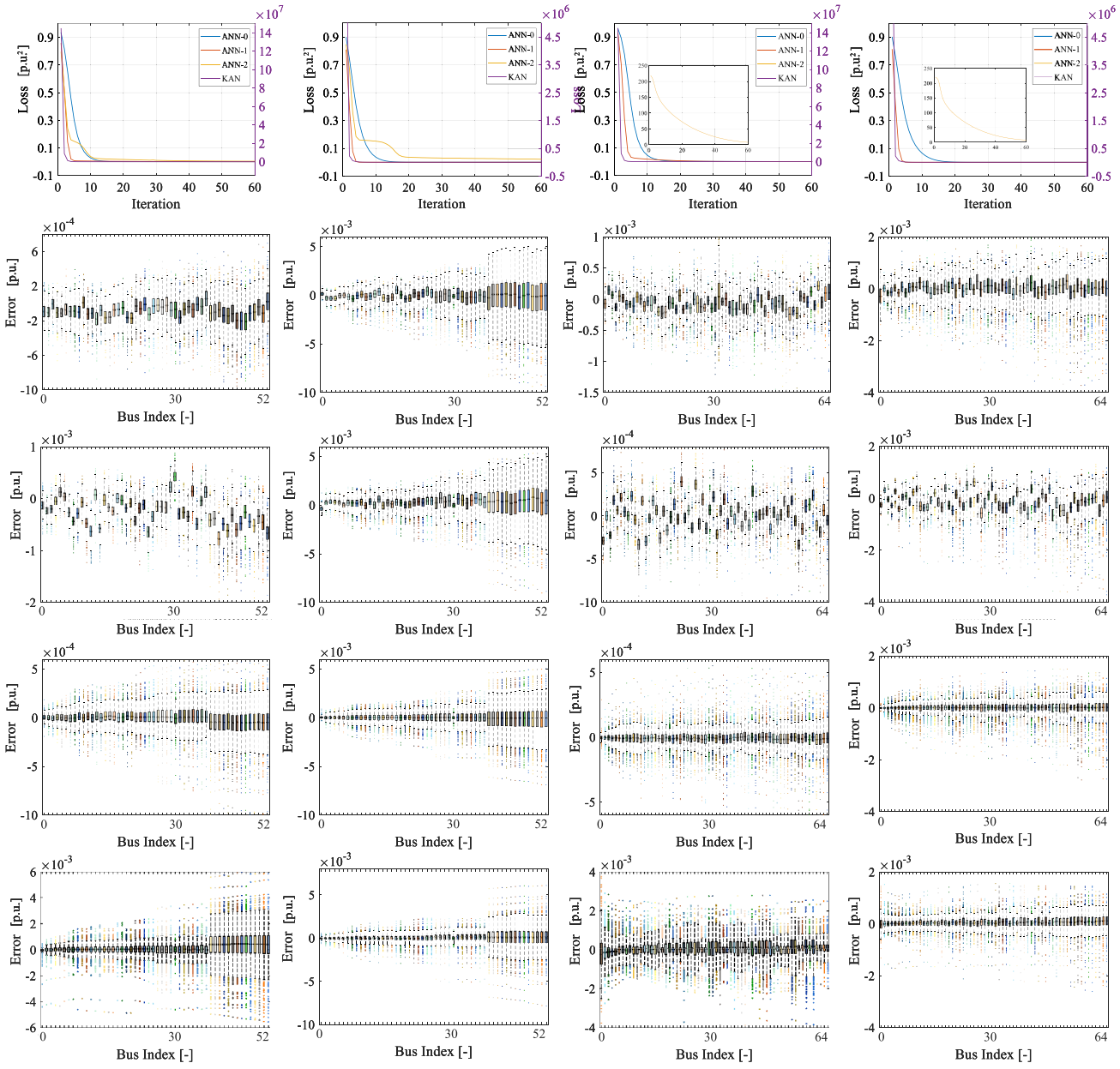
In this case study, we employ the *Secp256k1* elliptic curve for cryptographic operations, which is renowned for its security properties and widespread use in blockchain systems. Communication between the DSO and the SMs is facilitated using ZeroMQ (*zmq* package) with a PAIR socket configuration, ensuring reliable, bi-directional message exchange. All cases are conducted on one laptop with a preset port ID. The impact of parameters I and N on the data collection efficiency is analysed. The maximum values for I and N are set at 16 and 90, respectively. The dimension of the shared dataset is 2800 (i.e., a one-month dataset with 15 min resolution). The time cost is used as an indicator to show the efficiency of the proposed ZKP-based collection strategy, as shown in Fig. 6.

In Fig. 6, the overall time cost increases with both the dimension I of the uniformly distributed vector $\tilde{V}_n$ and the number of SM sharing data with the DSO. The time cost is used to reveal the communication burden. A larger vector dimension I introduces more potential insertion positions for the split data, thereby expanding the search space for the prover (i.e., the DSO), leading to more communication between prover and verifier to verify the true values. Moreover, increasing the number of SM also leads to more individual interactions with the DSO, further contributing to the overall time cost. As shown by the comparison between subfigures (a) and (b) in Fig. 6, the time cost increases more significantly with the dimension of the uniformly distributed vector $\tilde{V}_n$ than with the number of SMs. For example, when both I and N are at their maximum values, the time required for transmitting 2-split data is 97 seconds, whereas it rises to 1930 seconds for 4-split data, illustrating the exponential growth in overhead with higher-dimensional input vectors.

Nonetheless, the proposed ZKP-based strategy remains practical, as it can securely collect data from up to 90 users in under 40 minutes without compromising user privacy. It is also worth noting that the above experiments were conducted sequentially on a personal laptop. In a parallelised deployment environment, such as on a server with concurrent processing capabilities, the total data collection time could be significantly reduced.

C. Accuracy Evaluation

The ANN outlined in [36] is employed as the baseline neural network for PF calculations, referred to as ANN-0. For comparative purposes, a five-layer fully connected ANN and a six-layer ANN with a scalar are also utilized, designated as ANN-1 and ANN-2, respectively. The learning rate is set to 5.0×10^{-6}, the maximum number of epochs is 1500, and the batch size is 25. The hidden layer comprises 512 neurons, and the activation function tanh is used. The AdamW optimizer is employed to train the ANN. Besides, a three-layer Kolmogorov-Arnold Network (KAN) [37], with an equal number of neurons, is included as a test case. Given the



(a) LV-52 with D_0 (b) LV-52 with D_0^* (c) LV-64 with D_0 (d) LV-64 with D_0^*

Fig. 7. The first row presents the convergence curve; the second to fifth rows display the estimation errors for ANN-0, ANN-1, ANN-2, and KAN, respectively.

robust capabilities of KAN, the learning rate is set to 0.001, with a smaller batch size of 5, and the maximum number of epochs is set to 150. Both the untransformed dataset D_0 and the transformed dataset D_0^* are used to train these four ANNs across the three networks. The SM dataset comprises 2,880 samples (i.e., 96 samples per day over 30 days), with 80% of the data allocated for training and the remaining 20% for testing. Besides, the voltage magnitude datasets were also normalised to ensure consistent feature scaling across all output data, aiming to improve the training stability and prediction accuracy of the ANN-based PF calculation.The experiment results are illustrated in Fig. 7.

The first row in Fig. 7 demonstrates that all ANNs converge to a value near zero at similar speeds when trained on both the original and transformed datasets. This suggests that the *LRS* does not significantly impact the convergence speed. The four ANNs are trained to the preset maximum number of epochs, and their estimation errors, measured by MSE, are depicted in the second through fifth rows. The results presented in the second and third rows reveal that the error distributions remain within comparable intervals. For example, the second and fourth columns in these rows exhibit error distributions primarily concentrated within the ranges of [–0.005,0.005] and [–0.002,0.002] *p.u.*, respectively. This suggests that a deeper NN structure does not significantly contribute to improving the accuracy of PF calculation, thereby highlighting the limitations of this design (i.e., increasing the number of hidden layers of the ANN) for further performance enhancement. The error

distributions in the second and third rows versus the fourth row show that errors in the fourth row are more tightly clustered around zero. This suggests that applying a standard scaler for feature transformation slightly reduces the estimation bias in voltage magnitudes. Furthermore, the fifth row reveals that KAN, despite having fewer layers and fewer training iterations, can achieve accuracy comparable to that of ANN-2 in certain cases, such as with transformed data in the LV-52 and LV-64 networks. Comparing the first with the second columns and the third with the fourth columns, while ANN-0 and ANN-1 perform well on the original data, their accuracy diminishes on the transformed data. In contrast, ANN-2 maintains high accuracy on both the transformed and original datasets, highlighting the effectiveness of the scalar. The performance of the four ANNs under the LV-95 network with the dataset from the Netherlands is similar to their performance under the LV-64 network with the dataset from London (as shown in the third and fourth columns in Fig. 7).

On the other hand, the feasibility and accuracy of the trained ANN-based PF calculations are significantly impacted by network topology changes. This limitation stems from the inherent generalisation of ANNs, which are fundamentally dependent on their training data distribution and architectural presets. For instance, the ANN trained for the LV-52 network cannot be directly applied to perform accurate PF analysis for the LV-64 topology without retraining. Consequently, NN-based approaches typically require either complete retraining or targeted fine-tuning to mitigate the performance degradation caused by network topology changes.

D. Robustness Evaluation

Considering the random error brought by SMs, three kinds of Gaussian error (i.e., $v_e \sim \mathcal{N}(\mu, \sigma)$) were generated and added to the simulation voltage magnitude data according to the accuracy requirements for SMs [38]. The mean μ of the Gaussian distribution was set as 0, and three times the standard deviation 3σ of the Gaussian distribution was set as 0.2%, 0.5%, and 1%. The performance of ANN-2 under LV-64 with transformed data with measurement error is verified.

The convergence speed of the ANN-2 is depicted in Fig. 8 (a). While ANN-2 demonstrates the ability to converge to a lower error value when trained on data with inherent errors, while the convergence process is relatively slow. To assess the robustness of ANN-2, four datasets with varying error levels 0%, 0.2%, 0.5%, and 1% were employed, with the ANN-2 initially trained on a dataset containing a 0.2% error. The results, shown in Fig. 8 (b), reveal that the error distributions across all test scenarios are strikingly consistent, falling within a narrow range of –0.002 to 0.001 *p.u.*. Furthermore, Table I summarizes the testing errors of ANN-2 across 16 different scenarios, which are consistent with the results in Fig. 8. These results indicate that the impact of measurement errors is almost negligible. This underscores the advantage of ANN-based PF calculations over model-based methods: ANN-based approaches exhibit enhanced robustness to errors and inaccurate datasets, leading to more reliable and accurate results.

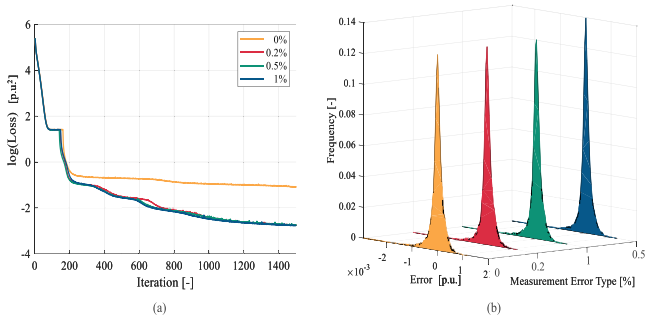


Fig. 8. Convergence curve of ANN-2 under varying error levels in (a) and probability density distribution of the estimation error in (b).

TABLE I
TEST ERROR (10^{-4}) OF ANN-2 UNDER MULTIPLE SCENARIOS

Test Data		0%	0.2%	0.5%	1%
Train Data	0%	−0.0697±2.6	-0.0531±2.4	-0.0837±2.6	-0.0374±2.5
	0.2%	-0.0945±2.7	0.0466±2.7	-0.0576±2.6	0.0166±2.4
	0.5%	-0.0238±2.6	-0.0228±2.6	-0.0201±2.7	-0.1166±2.4
	1%	-0.0628±2.6	-0.1306±2.5	-0.0694±2.6	0.0097±2.9

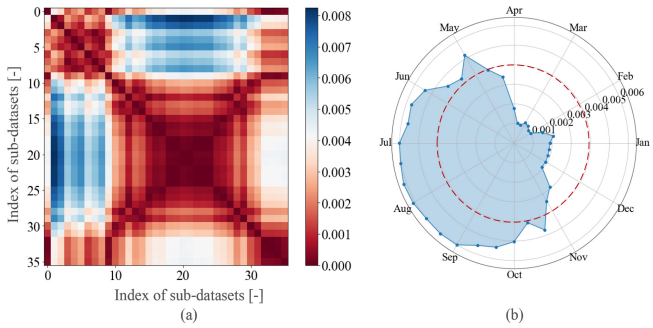


Fig. 9. WD among 10-day datasets in LV-95 in (a) and the WD-based indicator in (b).

E. IL-Based Updating

The one-year voltage dataset from LV-95 was divided into 36 datasets, each covering a 10-day period. The WD between these datasets was calculated and visualized in Fig. 9 (a). For instance, using the first 10 datasets as the training data, the WD-based indicator t was obtained and displayed in Fig. 9 (b). Fig. 9 reveals significant differences between the distribution of summer and winter datasets, reflecting seasonal variations. The threshold tt^* can be determined by DSO based on historical data or practical experience (e.g., as indicated by the red dashed circle in Fig. 9 (b)).

In this paper, the parameters of the last three layers of ANN-2 are frozen as a benchmark, meaning $\boldsymbol{l}$ is set as [4, 5, 6]. The remaining 3 layers introduced in Section III-C are updated during the incremental learning period. To assess the feasibility of the employed IL-based updating strategy for ANN-2, a series of experiments were conducted on the LV-95 network across 12 different scenarios, varying the number of frozen layers (ranging from the 2nd to the 6th layer) and adjusting the learning rate (from 5.0×10^{-5} to 5.0×10^{-7}). A three-month dataset in summer (24 samples per day for 90 days) was utilized for initial training. An additional 20-day dataset from winter was employed for IL-based updating,

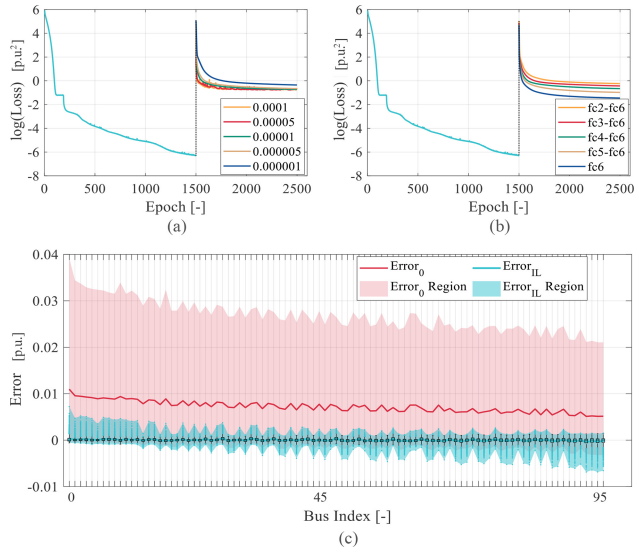


Fig. 10. Loss curve of ANN-2 with IL under multiple learning rates in (a), multiple frozen layers in (b) and distribution of the estimation error in (c).

TABLE II
TEST MAXIMUM/MEAN ERROR (10^{-4}) OF ANN-2 UNDER MULTIPLE SCENARIOS

Frozen Layer		fc2-fc6	fc4-fc6	fc6
Learning Rate	1×10^{-6}	144/0.42	89.1/0.04	60.2/0.16
	5×10^{-6}	103/0.05	72.1/0.01	56.1/0.21
	1×10^{-5}	92.5/0.15	71.0/0.09	54.9/0.25
	5×10^{-5}	87.3/0.34	65.9/0.48	66.1/0.53
	1×10^{-4}	83.9/1.97	85.0/0.98	73.2/1.91

followed by a 60-day dataset used for testing. The maximum number of epochs for updating was set at 1,000. The outcomes of these experiments are detailed in Fig. 10 and Table II, which summarize the maximum and mean estimation errors.

Fig. 10 demonstrates that the proposed IL-based updating strategy effectively maintains PF accuracy by fine-tuning the parameters of the unfrozen layers. As illustrated in Fig. 10 (a), the learning rate has a relatively minor effect on the IL-based updates, with the optimal value for this test case determined to be approximately 0.00001, as indicated in Table II. Compared to the learning rate, the depth of the ANN-2 has a greater influence on the update process and the final convergence value, as shown in Fig. 10 (a) and (b), which is consistent with the findings in Table II. Meanwhile, updating more layers (i.e., fewer frozen layers) increases the computational burden. When the learning rate is set to 0.000005 and layers fc4 through fc6 are frozen, the distribution of Error_0 was depicted in Fig. 10 (c). The red region in Fig. 10 (c) shows a noticeable bias in voltage magnitude estimation when tested on a dataset from a different season, meaning that the trained ANN-2 cannot accurately estimate voltage magnitude for the whole year. However, after applying the IL-based updating strategy, this bias and the region of the estimation error are significantly reduced, as evidenced by the light blue region in Fig. 10 (c), indicating the effectiveness of the IL strategy.

TABLE III
INDICES OF HOUSEHOLDS ASSIGNED TO PHASES A, B, AND C IN A THREE-PHASE LV-95

Phase A	Phase B	Phase C
1, 2, 3, 4, 8,	7, 10, 11, 16, 17	5, 12, 13, 18, 21,
9, 14, 15, 19, 20	24, 25, 29, 30, 36,	22, 23, 31, 33, 34,
26, 27, 28, 32, 38	37, 42, 43, 44, 48,	41, 45, 50, 53, 54,
39, 40, 46, 47, 51	49, 55, 56, 59, 60,	57, 61, 65, 66, 70,
52, 58, 62, 63, 64	67, 68, 69, 75, 79,	71, 76, 83, 84, 87,
72, 73, 74, 78, 80	81, 82, 91, 92, 94	88, 93, 95, 35, 77
85, 86, 89, 90,		

F. Performance Assessment in Unbalanced LVDNs

The previous experiments were conducted under the assumption of balanced distribution networks, and also demonstrated that the proposed method can be directly applied to the single-phase management in DNs. This section extends the evaluation to unbalanced three-phase DNs, which more accurately reflect practical operational scenarios. Using the LV-95 network as a case study, the phase connections of individual households are summarised in Table III. A Load-based Unbalance Indicator (LUI_P) quantifies the imbalance in a three-phase DN by measuring the maximum deviation of per-phase active power from the average, expressed as a percentage, as shown next:

$$P_{\text{avg}} = \frac{P_A + P_B + P_C}{3}, \tag{29}$$

$$\text{LUI}_P = \frac{\max_{i\in\{A,B,C\}} |P_i - P_{\text{avg}}|}{P_{\text{avg}}} \times 100\%, \tag{30}$$

where P_A, P_B, and P_C represent the total active power of phases A, B, and C, respectively. P_{avg} is the average active power across the three phases. A higher LUI_P value indicates a more pronounced three-phase load imbalance within the system, whereas a lower value signifies a relatively balanced operating condition.

The proposed local data privacy-preserving strategies are designed for offline training data generation and online randomised power data transmission, offering robust applicability to three-phase DNs. These strategies are decentralised, with computational processes executed locally at each SM without requiring knowledge of the network topology. Consequently, they exhibit resilience to network configuration variations, such as changes in phase connectivity or topological reconfigurations. The feasibility and efficacy demonstrated in prior single-phase scenarios are preserved in three-phase DNs, enabling secure, scalable data sharing while safeguarding user privacy. A new ANN-1 model was trained using data collected from the three-phase LV-95 network, and the results are depicted in Fig. 11.

Fig. 11 (a) and (b) illustrate that, without modifications to the ANN-1 architecture or specific data pre-processing strategies, the performance of ANN-1, originally trained for PF analysis in single-phase DNs, degrades when applied to three-phase DNs. For example, the voltage estimation error range increases from approximately -3×10^{-3} to 2×10^{-3} in the single-phase scenario, to -2×10^{-2} to 1.5×10^{-2} in the three-phase scenario. This behaviour aligns with expectations, as

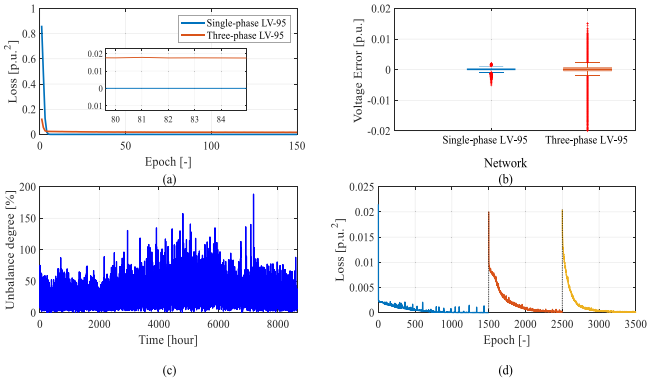


Fig. 11. Simulation results on the three-phase LV-95 network: (a) training loss curve of ANN-1, (b) distribution of voltage estimation errors, (c) network voltage unbalance degree over time, and (d) training loss of ANN-1 with incremental learning under multiple learning rates.

neural networks typically exhibit limited generalisation capabilities. The accuracy deterioration can be mitigated to some extent by either improving the ANN architecture or applying appropriate data preprocessing techniques. Nevertheless, the retrained ANN-1 achieves a voltage estimation error mostly concentrated between -4×10^{-3} and 4×10^{-3} in most cases. Fig. 11 (c) presents the LUI over one year, reflecting dynamic changes in network imbalance levels along with the seasonal variations in the load profile. Such temporal fluctuations also pose challenges for ANN-based PF models. As shown in Fig. 11(d), when the final layer is frozen and the learning rate is kept constant, the employed IL strategy slightly mitigates the adverse effects caused by load seasonal variations and network unbalance, thereby enhancing the long-term application of the model.

V. Conclusion

This paper introduced a privacy-preserving power flow (PF) calculation framework to ensure household privacy in model-free PF calculation. Compared to differential privacy and traditional encryption strategies, the employed local randomisation and zero-knowledge proof (ZKP)-based data collection strategies are structurally simpler and more easily deployable within smart meters (SM). The simulation results indicated that the proposed ZKP-based data collection strategy enables distribution system operators to efficiently and securely collect voltage magnitude, and the proposed local randomisation strategy can effectively change the distribution of the initial data, hindering the true distribution of power datasets. Meanwhile, with the cooperation of artificial neural networks (ANN) in PF analysis, the constructed ANN maintain a high accuracy on both the original dataset, the randomised dataset and the dataset with measurement errors. Moreover, the results underscore the feasibility of deploying an artificial neural network with an incremental learning-based updating strategy in long-term applications, ensuring sustained accuracy and adaptability. Nevertheless, the proposed framework faces two key limitations: the inherent generalization constraints of neural networks to unseen network topology changes, and its dependency on complete SM data inputs. While these factors may impact the accuracy of ANN-based PF accuracy, they do not affect the privacy guarantees of the proposed framework, which remains valid and modular.

References

[1] E. M. S. Duque, J. S. Giraldo, P. P. Vergara, P. H. Nguyen, and H. J. Slootweg, "Tensor power flow formulations for multidimensional analyses in distribution systems," *Int. J. Electr. Power Energy Syst.*, vol. 162, 2024, Art. no. 110275.

[2] L. Guo et al., "Data-driven power flow calculation method: A lifting dimension linear regression approach," *IEEE Trans. Power Syst.*, vol. 37, no. 3, pp. 1798–1808, May 2022.

[3] M. Gao, J. Yu, Z. Yang, and J. Zhao, "Physics embedded graph convolution neural network for power flow calculation considering uncertain injections and topology," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 11, pp. 15467–15478, Nov. 2024.

[4] X. Hu, H. Hu, S. Verma, and Z.-L. Zhang, "Physics-guided deep neural networks for power flow analysis," *IEEE Trans. Power Syst*, vol. 36, no. 3, pp. 2082–2092, May 2021.

[5] K. Huhta, "Smartening up while keeping safe? Advances in smart metering and data protection under EU law," *J. Energy Nat. Resour. Law*, vol. 38, no. 1, pp. 5–22, 2020.

[6] C. Cuijpers and B.-J. Koops, "Smart metering and privacy in Europe: Lessons from the Dutch case," in *European Data Protection: Coming of Age*. Dordrecht, The Netherlands: Springer, 2012, pp. 269–293. [Online]. Available: https://research.tilburguniversity.edu/en/publications/smart-metering-and-privacy-in-europe-lessons-from-the-dutch-case

[7] V. von Loessl, "Smart meter-related data privacy concerns and dynamic electricity tariffs: Evidence from a stated choice experiment," *Energy Policy*, vol. 180, Sep. 2023, Art. no. 113645.

[8] J. Lin, J. Ma, and J. Zhu, "A privacy-preserving federated learning method for probabilistic community-level behind-the-Meter solar generation disaggregation," *IEEE Trans. Smart Grid*, vol. 13, no. 1, pp. 268–279, Jan. 2022.

[9] A. Unterweger, F. Knirsch, D. Engel, D. Musikhina, A. Alyousef, and H. de Meer, "An analysis of privacy preservation in electric vehicle charging," *Energy Inf.*, vol. 5, no. 1, p. 3, 2022.

[10] Z. Cheng, F. Ye, X. Cao, and M.-Y. Chow, "A homomorphic encryption-based private collaborative distributed energy management system," *IEEE Trans. Smart Grid*, vol. 12, no. 6, pp. 5233–5243, Nov. 2021.

[11] W. Chen, Z. Wang, Q. Ge, H. Dong, and G.-P. Liu, "Quantized distributed economic dispatch for Microgrids: Paillier encryption–decryption scheme," *IEEE Trans. Ind. Informat.*, vol. 20, no. 4, pp. 6552–6562, Apr. 2024.

[12] Z. Zheng, T. Wang, A. K. Bashir, M. Alazab, S. Mumtaz, and X. Wang, "A decentralized mechanism based on differential privacy for privacy-preserving computation in smart grid," *IEEE Trans. Comput.*, vol. 71, no. 11, pp. 2915–2926, Nov. 2022.

[13] T. Wu, C. Zhao, and Y.-J. A. Zhang, "Privacy-preserving distributed optimal power flow with partially homomorphic encryption," *IEEE Trans. Smart Grid*, vol. 12, no. 5, pp. 4506–4521, Sep. 2021.

[14] N. Busom, R. Petrlic, F. Sebé, C. Sorge, and M. Valls, "Efficient smart metering based on homomorphic encryption," *Comput. Commun.*, vol. 82, pp. 95–101, May 2016.

[15] Z.-P. Yuan, P. Li, Z.-L. Li, and J. Xia, "A fully distributed privacy-preserving energy management system for networked microgrid cluster based on homomorphic encryption," *IEEE Trans. Smart Grid*, vol. 15, no. 2, pp. 1735–1748, Mar. 2024.

[16] J. von der Heyden et al., "Privacy-preserving power flow analysis via secure multi-party computation," *IEEE Trans. Smart Grid*, vol. 16, no. 1, pp. 344–355, Jan. 2025.

[17] J. Huang, Q. Huang, G. Mou, and C. Wu, "DPWGAN: High-quality load profiles synthesis with differential privacy guarantees," *IEEE Trans. Smart Grid*, vol. 14, no. 4, pp. 3283–3295, Jul. 2023.

[18] L. Sun, D. Ding, H. Dong, and X. Bai, "Privacy-preserving distributed economic dispatch for microgrids based on state decomposition with added noises," *IEEE Trans. Smart Grid*, vol. 15, no. 3, pp. 2424–2433, May 2024.

[19] X. Wang, H. Ishii, L. Du, P. Cheng, and J. Chen, "Privacy-preserving distributed machine learning via local Randomization and ADMM perturbation," *IEEE Trans. Signal Process.*, vol. 68, pp. 4226–4241, 2020.

[20] Z. Zhang, Z. Qin, L. Zhu, J. Weng, and K. Ren, "Cost-friendly differential privacy for smart meters: Exploiting the dual roles of the noise," *IEEE Trans. Smart Grid*, vol. 8, no. 2, pp. 619–626, Mar. 2017.

[21] Z. Yang, P. Cheng, and J. Chen, "Differential-privacy preserving optimal power flow in smart grid," *IET Gener. Transm. Distrib.*, vol. 11, no. 15, pp. 3853–3861, 2017.
[22] V. Dvorkin, F. Fioretto, P. Van Hentenryck, P. Pinson, and J. Kazempour, "Differentially private optimal power flow for distribution grids," *IEEE Trans. Power Syst*, vol. 36, no. 3, pp. 2186–2196, May 2021.
[23] B. Huang and J. Wang, "Applications of physics-informed neural networks in power systems—A review," *IEEE Trans. Power Syst*, vol. 38, no. 1, pp. 572–588, Jan. 2023.
[24] R. Dobbe, Y. Pu, J. Zhu, K. Ramchandran, and C. Tomlin, "Local differential privacy for multi-agent distributed optimal power flow," in *Proc. IEEE PES Innov. Smart Grid Technol. Europe (ISGT-Europe)*, 2020, pp. 265–269.
[25] R. Lavin, X. Liu, H. Mohanty, L. Norman, G. Zaarour, and B. Krishnamachari, "A survey on the applications of zero-knowledge proofs," 2024, *arXiv:2408.00243*.
[26] D. Hou, J. Zhang, S. Huang, Z. Peng, J. Ma, and X. Zhu, "Privacy-preserving energy trading using blockchain and zero knowledge proof," in *Proc. IEEE Int. Conf. Blockchain (Blockchain)*, 2022, pp. 412–418.
[27] X. Zhou, B. Wang, H. Zhao, H. Sun, Q. Guo, and B. Chen, "Incentivized coordinated heat-electricity-gas dispatch: A zero knowledge proof-based solution considering privacy and anti-forgery," *IEEE Trans. Sustain. Energy*, vol. 15, no. 2, pp. 713–725, Apr. 2024.
[28] D. Gabay, K. Akkaya, and M. Cebe, "Privacy-preserving authentication scheme for connected electric vehicles using blockchain and zero knowledge proofs," *IEEE Trans. Veh. Technol.*, vol. 69, no. 6, pp. 5760–5772, Jun. 2020.
[29] G. M. Van de Ven, T. Tuytelaars, and A. S. Tolias, "Three types of incremental learning," *Nat. Mach. Intell.*, vol. 4, no. 12, pp. 1185–1197, 2022.
[30] F. Zenke, B. Poole, and S. Ganguli, "Continual learning through synaptic intelligence," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 3987–3995.
[31] "Regulation (EU) 2016/679 of the European parliament and of the council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (general data protection regulation)," European Union, Brussels, Belgium, document 32016R0679, May 2016. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng.
[32] D. Liu, J. S. Giraldo, P. Palensky, and P. P. Vergara, "Topology identification and parameters estimation of LV distribution networks using open GIS data," *Int. J. Electr. Power Energy Syst.*, vol. 164, Mar. 2025, Art. no. 110395.
[33] N. Lin, S. Orfanoudakis, N. O. Cardenas, J. S. Giraldo, and P. P. Vergara, "PowerFlowNet: Power flow approximation using message passing graph neural networks," *Int. J. Electr. Power Energy Syst.*, vol. 160, Sep. 2024, Art. no. 110112.
[34] Y. Jiang, C.-C. Liu, and Y. Xu, "Smart distribution systems," *Energies*, vol. 9, no. 4, p. 297, Apr. 2016.
[35] K. P. Schneider et al., "Analytic considerations and design basis for the IEEE distribution test feeders," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3181–3188, May 2018.
[36] Z. Kaseb, Y. Xiang, P. Palensky, and P. P. Vergara, "Adaptive activation functions for deep learning-based power flow analysis," in *Proc. IEEE PES Innov. Smart Grid Technol. Europe (ISGT EUROPE)*, 2023, pp. 1–5.
[37] Z. Liu et al., "Kan: Kolmogorov-Arnold networks," 2024, *arXiv:2404.19756*.
[38] S. García, J. M. Mora-Merchán, D. F. Larios, E. Personal, A. Parejo, and C. León, "Phase topology identification in low-voltage distribution networks: A Bayesian approach," *Int. J. Electr. Power Energy Syst.*, vol. 144, Jan. 2023, Art. no. 108525.

Dong Liu received the B.Sc. degree in electrical engineering from the Nanjing University of Science and Technology, Nanjing, China, and the M.Sc. degree in electrical engineering from Hunan University, Changsha, China, in 2022. After that, he joined the Delft University of Technology, The Netherlands, as a Ph.D. Researcher. From April 2025 to July 2025, he was a Visiting Researcher with Liege University, Belgium. His research interests include the application of data-driven approaches, coordinated optimization control, self-taught learning, federate learning, and machine learning methods in distribution networks.

Juan S. Giraldo (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from the Universidad Tecnológica de Pereira, Pereira, Colombia, in 2012, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of Campinas, Campinas, Brazil, in 2015 and 2019, respectively. From October 2019 to May 2021, he was a Postdoctoral Fellow with the Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands. From June 2021 to August 2022, he was a Researcher with the Mathematics of Operations Research Group, University of Twente, Enschede, The Netherlands. He is currently a Researcher with the Techno-Economic Energy Transition Studies Group, Netherlands Organization for Applied Scientific Research (TNO). His current research interests include optimization, planning, and control of energy systems, energy transition pathways, and machine learning applications to energy systems.

Peter Palensky (Senior Member, IEEE) received the M.Sc. degree in electrical engineering and the Ph.D. and Habilitation degrees from the Vienna University of Technology, Austria, in 1997, 2001, and 2015, respectively. He co-founded a Envidatec, a German startup on energy management and analytics, and joined the Lawrence Berkeley National Laboratory, Berkeley, CA, USA, as a Researcher and the University of Pretoria, South Africa, in 2008. In 2009, he became an Appointed Head of Business Unit on Sustainable Building Technologies with the Austrian Institute of Technology (AIT) and later the First Principle Scientist for Complex Energy Systems with AIT. In 2014, he was appointed as a Full Professor of Intelligent Electric Power Grids with TU Delft. His main research fields are energy automation networks, smart grids, and modeling intelligent energy systems. He is active in international committees like ISO or CEN and serves as an IEEE IES AdCom member-at-large in various functions for the IEEE. He is the Editor in Chief of the *IEEE Industrial Electronics Magazine* and an Associate Editor for several other IEEE publications, and regularly organizes IEEE conferences.

Pedro P. Vergara (Senior Member, IEEE) was born in Barranquilla, Colombia, in 1990. He received the B.Sc. degree (with Hons.) in electronic engineering from the Universidad Industrial de Santander, Bucaramanga, Colombia, in 2012, the M.Sc. degree in electrical engineering from the University of Campinas, UNICAMP, Campinas, Brazil, in 2015, and the Ph.D. degree from the University of Campinas, UNICAMP, and the University of Southern Denmark, Denmark, in 2019, funded by the Sao Paulo Research Foundation (FAPESP). In 2019, he joined the Eindhoven University of Technology, TU/e, The Netherlands, as a Postdoctoral Researcher. In 2020, he was appointed as an Assistant Professor with the Intelligent Electrical Power Grids Group, Delft University of Technology, The Netherlands, where he is currently an Associate Professor. His main research interests include the development of algorithms for the control, planning, and operation of electrical distribution systems with high penetration of low-carbon energy resources, such as, electrical vehicles, PV systems, and electric heat pumps using optimization and machine learning approaches. In 2018, he received the Best Presentation Award at the Summer Optimization School organized by the Technical University of Denmark and the Best Paper Award at the 3rd IEEE International Conference on Smart Energy Systems and Technologies, Türkiye, in 2020.