IET Generation, Transmission & Distribution

WILEY

The Institution of Engineering and Technology

ORIGINAL RESEARCH OPEN ACCESS

Selection for Flexibility Areas Using Probabilistic Machine Learning Under Measurement Uncertainty

Demetris Chrysostomou | Jose Luis Rueda Torres | Jochen Lorenz Cremer

Department of Electrical Sustainable Energy, Delft University of Technology, South Holland, The Netherlands

Correspondence: Demetris Chrysostomou (D.Chrysostomou@tudelft.nl)

Received: 19 November 2025 | **Revised:** 13 April 2026 | **Accepted:** 14 May 2026

ABSTRACT

Coordination between transmission system operators (TSOs) and distribution system operators (DSOs) can support TSOs in using the distribution system (DS) flexibility while ensuring feasible operation. Flexibility areas (FAs) can support TSO-DSO coordination, aggregating the total feasible flexibility within the DS. However, existing real-time estimation approaches do not consider the limited measurements within DS. This paper proposes a Bayesian neural network (BNN) to estimate the operating conditions that bound the operational flexibility, including epistemic and aleatoric uncertainties. These uncertainties stem from the limited real-time measurements in DSs and the measurement noise. TSOs can select a threshold that confirms a probability of safety, considering uncertainty margins. The paper also provides FA estimation in DS topologies with two points of common coupling (PCC) with the transmission system. Case studies in the CIGRE and Oberrhein networks compare the proposed BNNs to baseline statistic-based approaches for forecast and measurement uncertainty in FAs. The case studies show the proposed FA estimation under various safety margins and systems with 2-PCC. Case studies also assess various measurement noise levels and evaluate model performance for different DS topologies.

1 | Introduction

Flexibility from service providers (FSPs) in DSs can ensure a resilient and efficient transmission system (TS) and DS operation. FSPs can modify their active and reactive power net injection upon request by the TSOs or DSOs to avoid or address TS or DS issues, such as congestion.

TSOs and DSOs need to coordinate to enable TSOs utilize DS flexibility while also considering the DS constraints. TSO-DSO coordination is inherently complex due to the coupling of TS and DSs, the nonlinear and non-convex nature of AC power flow constraints, and the presence of multiple participants and uncertainties. TSO-DSO coordination is inherently complex due to the coupling of TS and DSs, the nonlinear and non-convex nature of AC power flow constraints, and the presence of multiple participants and uncertainties. TSO-DSO coordination approaches can be diversified to TSO-managed, DSO-managed, and hybrid managed [1]. The most common, DSO-managed approaches [1], can utilize aggregated DS FAs.

DSOs can inform TSOs about the total feasible flexibility available within DSs using FAs, without requiring the exchange of details on DS topology, and operating conditions. A flexibility combination is feasible if the resulting operating condition fulfils all power flow constraints in the DS, and line loading and bus voltages

Abbreviations: BNN, Bayesian neural network; DE, deep ensembles; DFC, density of feasible combinations; DS, distribution system; DSOs, distribution system operators; DSSE, distribution system state estimation; FAs, flexibility areas; FNN, feed-forward neural networks; FSPs, flexibility service providers; MCD, Monte Carlo dropout; MOO, multi-objective optimization; MV, medium voltage; NLL, negative log likelihood; Ob, complete Oberrhein network; Ob0, Oberrhein separated network with substation 0; Ob1, Oberrhein separated network with substation 1; OC, operating condition; OPF, optimal power flow; PCC, point of common coupling; PF, power flow; RMSE, root mean squared; RTI, real-time interface; TS, transmission system; TSOs, transmission system operators; VI, variational inference.

do not exceed rated values. By summarizing the capabilities of FSPs within DS, FAs enable TSOs to access the flexibility needed for balancing supply and demand, managing congestion, and enhancing system stability at the transmission level while ensuring DS feasibility. This TSO-DSO coordination approach simplifies the DSO-TSO interaction, ensures data privacy, and reduces the complexity of operational coordination while still using the potential of DS flexibility to support grid reliability and efficiency. FAs provide the TSOs with information on the controllable space in the active and reactive power plane from DS resources, independent from the specific application and enabling preventive and corrective actions. In contrast, platforms for specific applications, such as congestion management, may focus on active power and primarily for corrective actions. In settings with multiple DSOs, each DSO estimates and communicates the local FA to the TSO. The TSO treats each FA as an aggregated controllable feasible region and identifies the optimal setpoints within these FAs, satisfying the TS needs.

Existing FA estimation approaches mainly require the operating condition (OC) of the DS as input, as [2–6]. However, DSs can have limited real-time observability with measurement units on limited network components [7–10], for example, 5%–40% in medium voltage (MV) DS [11]. Thus, DSOs might not know the exact physical feasibility margins from all DS constraints due to this limited observability. These margins are important for DSOs to evaluate the feasibility of flexibility combinations. Although some existing approaches address forecasting uncertainty in day-ahead FA estimations [12–14], existing FA approaches neglect the limited observability in real-time FA estimation.

This paper proposes a new approach for FA estimation, using probabilistic Bayesian neural networks (BNNs) to assess the risk for non-feasible flexibility due to limited DS observability and DS constraint approximation error. BNNs extend conventional neural networks by including uncertainty estimation in the predictions [15, 16], capturing inherent uncertainty from the available measurements and the uncertainty in model parameters. These probabilistic estimates indicate the model prediction confidence and problem uncertainty, which can be important for robustness in tasks involving critical infrastructure, such as the power grid. The main tractable approaches in BNNs include variational inference (VI) [17], Monte Carlo dropout (MCD) [18] and deep ensembles (DE) [19]. The proposed approach enables the TSOs to select the safety percentage in the resulting area. A lower safety returns a larger FA for TSOs to select an operating point, whereas a higher safety better secures feasibility on the selected operating point. Nevertheless, the proposed approach also includes the density of feasible combinations for each operating point and the given safety. Thus, for a low-safety FA estimation, a higher-density operating point provides more reliability in finding alternative flexibility combinations.

FA estimation approaches apply steady-state simulations, typically using power flows (PFs) [5, 14] or optimal power flows (OPF) [3, 4, 13]. PF-based approaches typically apply Monte-Carlo sampling for a predetermined number of combinations. OPF-based approaches apply different multi-objective optimization (MOO) techniques. These approaches typically rely on an initial DS forecast or approximated state to evaluate the feasibility of different flexibility combinations. To deal with uncertainties from renewable sources in aggregated flexibility area estimation, [12] applied a chance-constrained method, and [13, 20] applied robust optimization. Refs. [14, 21] included the probability distributions of forecast errors to determine the probability of feasibility, and [22] focused on forecast uncertainty generation. The approaches of [12–14, 20–22] consider the forecast uncertainties rather than real-time observability uncertainty and map these uncertainties to the network constraints using the power-flow relationships. The proposed approach approximates the network constraint margin uncertainty using BNNs, considering the real-time measurements and limited DS observability.

FAs enable TSOs to account and utilize feasible DS flexibility. To further improve the information encompassed in FAs and support TSOs in the flexibility selection, studies also include costs in estimated areas [5, 23], or reliability metrics [22, 24]. The density of feasible combinations of the proposed approach can provide a reliability metric on the feasible combinations, but also the room for cost optimization when more combinations are feasible [25].

Existing FA estimation approaches rely on TS-DS interconnections with a single PCC [3–5, 13, 14]. However, DS ring topologies with two PCCs are increasingly adopted and can improve the DS losses [26]. The main challenges for estimating FA of TS-DS connected with multiple PCCs are the dependencies between the PCCs [23] and the complexity of representing these dependencies. This paper provides an approach to estimate and represent network flexibility with two PCCs. Hence, enabling the adoption of FA estimation in systems with two PCCs.

Recent works in energy management and DS operation support decision-making under uncertainty, particularly in the presence of high renewable penetration and distributed energy resources [27]. Resilience-oriented approaches emphasize the need to account for uncertainty, disturbances and multi-objective trade-offs between economic performance and system security. The uncertainty characterization in the proposed FA estimation approach supports optimization algorithms to incorporate uncertainty-aware feasibility constraints. This aligns with emerging trends in energy management systems, where data-driven models are integrated into optimization and control loops to enhance system flexibility and resilience [28].

Distribution system state estimation (DSSE) algorithms consider the challenge of limited observability in DSs by estimating the system's state using available measurements and pseudo-measurements. These algorithms primarily estimate nodal voltage magnitudes and angles [8, 29–31], though some also estimate nodal active and reactive powers [32] or additional metrics as line loading [33]. DSSE approaches applied Bayesian models for non-Gaussian pseudo-measurement uncertainties [32], to fuse measurements with varying sampling rates [29], or to generate data samples to train deep learning-based DSSE models [31, 34]. Bayesian models for DSSE often model uncertainties but typically provide a single expected state estimate [29, 32]. These DSSE approaches generally focus on estimating voltage magnitudes and angles. The proposed FA estimation approach applies BNNs to estimate the probabilistic DS constraint margins, accounting for the available measurements to provide an FA that is constrained by the safety margins selected by the TSO. The proposed approach also includes real-time interface (RTI)

measurement considerations. RTI is a platform in the Netherlands that mandates DER with a capacity higher than 1 MW to provide real-time measurements to system operators. Thus, unlike these DSSE approaches, the proposed BNN focuses on the constrained variables needed to evaluate the feasibility of flexibility shifts. The proposed FA estimation approach directly utilizes the available measurements to aggregate the feasible flexibility at the PCCs without requiring DSSE.

This paper develops an approach for FA estimation with BNNs to consider the estimation and DS observability uncertainties. This paper analyses the performance of BNN approaches considering the impact of network sizes, assumptions about noise levels, and data distributions for the power system task. The proposed FA approach also includes an algorithm for networks with two PCCs. Therefore, the main contributions are:

1. FA estimation approach that considers the uncertainty for the DS constraint margins due to limited DS observability and estimator uncertainty.
2. BNN model structures to estimate FAs considering real-time uncertainties and data distributional changes.
3. Approach approximating and representing FAs in networks with two-PCCs.

The case studies use the CIGRE 15-bus MV network and Oberrhein 70-, 109- and 179-bus MV networks. The case studies compare the proposed BNN model to alternative approaches for FA uncertainty, and demonstrate the proposed approach for FA estimation in DS with one and two PCCs.

2 | Limited Observability in Steady State Flexibility Area Estimation

2.1 | Problem Introduction

Let a DS state described by $X(t) \in \mathrm{R}^{n\times 2}, V(t) \in \mathrm{R}^{n}, L(t) \in \mathrm{R}^{k}$ at time t. The matrix $X(t)$ is the active and reactive power injected at each of the n nodes, $V(t)$ is the voltage magnitude at each n, and $L(t)$ is the loading at each of the k lines and transformers. The FSPs can linearly modify $X(t)$ by increasing or reducing the injection on a subset of the DS nodes $\hat{n}$. However, the impact on the network component loading and voltage is non-linear. Thus, a combination of flexibility shifts from FSPs at t is $U(t,\tau) \in \Omega^{U}(t,\tau) \subset \mathrm{R}^{\hat{n}\times 2}$, where $\Omega^{U}(t,\tau)$ is the set of all flexibility combinations at t, and τ is the duration until the new state is achieved. Applying $U(t,\tau)$ would result in a new state after τ as:

$$X(t+\tau) = X(t) + B \cdot U(t,\tau), \tag{1}$$

$$V(t+\tau) = V(t) + F^{v}(X(t), V(t), U(t,\tau)), \tag{2}$$

$$L(t+\tau) = L(t) + F^{l}(X(t), L(t), U(t,\tau)), \tag{3}$$

where $B \in \{0,1\}^{n\times\hat{n}}$ is a binary matrix with 1 on each FSP's corresponding node and 0 otherwise, for example, if only the fifth FSP is connected on the first node, the first row of B will be 1 only on the fifth element. $F^{v}(\cdot)$ and $F^{l}(\cdot)$ are the non-linear functions of flexibility impacts on network component voltage and loading, respectively.

A network includes voltage constraints $c_v^{\min}, c_v^{\max}$, and loading constraints $c_l^{\max}$. The objective of a steady-state FA estimation algorithm is to describe the feasible area (considering DS constraints) of active and reactive power exchange at π PCCs between the TSO and DSO, given all available flexibility combinations $\Omega^{U}(t,\tau)$. A single PCC's active and reactive power exchange is $x_{\mathrm{PCC}}(t)$. As shown in [10, 25], an additional dimension for the density of feasible combinations (DFC) can further demonstrate the reliability in achieving a PCC operating point. Reliability in terms of a plethora of alternatives to achieve said operating point. Thus, the FA objective is to identify $Y(t+\tau) \subset \mathrm{R}^{\pi\times 3}$, as:

$$Y(t+\tau) = \Big\{ y(t+\tau) = F^{A}(X(t), L(t), V(t), U(t,\tau))|, U(t,\tau) \in \Omega^{U}(t,\tau), c_v^{\min} \le V(t+\tau) \le c_v^{\max}, L(t+\tau) \le c_l^{\max} \Big\}, \tag{4}$$

where $y(t+\tau) \in \mathrm{R}^{\pi\times 3}$ describes the active power, reactive power and DFC [25] for each PCC operating point. Function $F^{A}(\cdot)$ applies each $U(t,\tau) \in \Omega^{U}(t,\tau)$ on (1)–(3) and aggregates the feasible combinations for each $x_{\mathrm{PCC}}(t+\tau)$ as [25].

2.2 | FA Uncertainty

Existing approaches mainly apply (1)–(3) by solving the power flow relationships. However, DSOs do not deterministically know $X(t), V(t), L(t)$ to estimate the FAs. Depending on the intended FA estimation speed, the $X(t), V(t), L(t)$ source for uncertainty can be:

1. In the day-ahead estimation of FAs, DSOs forecast the DS operating conditions, which can differ in real-time.
2. In the real-time estimation of FAs, DSOs only measure a subset of the network components.

Addressing the day-ahead FA estimation forecast uncertainty, existing approaches [12–14] include the standard deviation in DER injection forecasts for $X(t)$ to provide robust or chance constrained FAs. However, for limited observability in real-time, existing approaches neglect the resulting FA error [10]. This research addresses the impact of real-time limited DS observability on the FA estimation task.

FA estimation uncertainty can be split into aleatoric and epistemic. Aleatoric uncertainty is inherent in data and cannot be reduced, for example, measurements and pseudo-measurements have noise that causes uncertainty. Epistemic uncertainty relates to the model's structure, parameters or assumptions due to limited data. Different data distributions and patterns can be absent in the model development, however appear in model deployment. Existing FA estimation approaches for forecasts focus on aleatoric uncertainty as they do not include the uncertainty in prediction model parameters but rather the uncertainty associated with the data. Considering this uncertainty, a safety margin for the DS loading and voltage constraints can alleviate any potential impact from the miscalculation of the initial DS conditions.

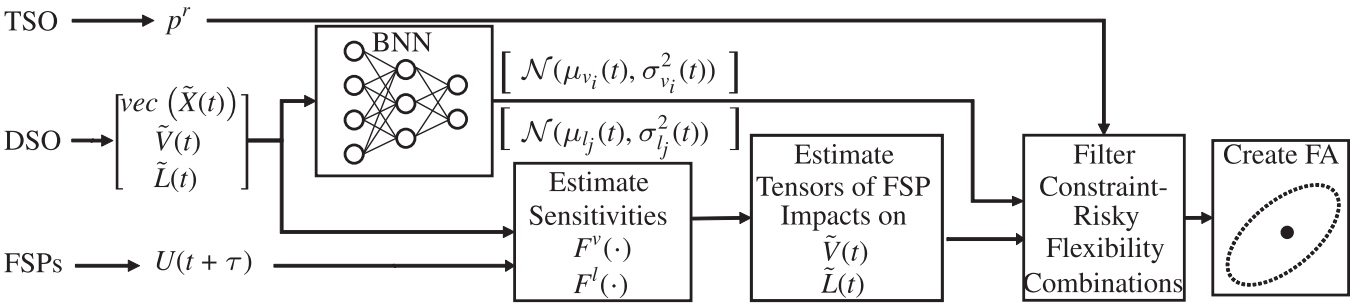


FIGURE 1 | Overview of proposed FA estimation process.

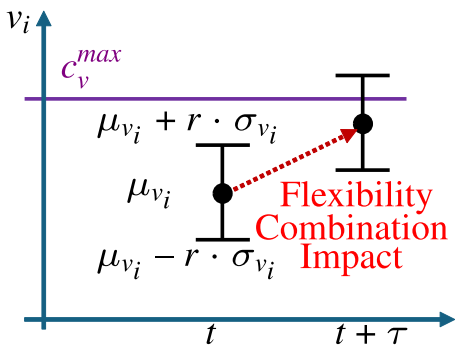


FIGURE 2 | Feasibility evaluation example.

2.3 | Proposed FA Estimation Approach

Figure 1 shows the proposed FA estimation. The DSO provides the measured and pseudo-measured DS variables. The TSO provides the percentage p^r for safety margin, that is, the minimum confidence that the flexibility combinations are feasible considering the real-time DS operation uncertainty. FSPs provide their offered flexibility, to obtain the $U(t,\tau)$. The proposed approach estimates $F^v(\cdot), F^l(\cdot)$ and the tensors of flexibility impacts on the constrained voltage and loading variables using the modified TensorConvolution+ algorithm of [25]. This algorithm estimates $F^v(\cdot), F^l(\cdot)$ and the tensors of flexibility impacts by performing power flow estimations, and modified convolution operations to combine flexibility shifts. Therefore, the functions $F^v(\cdot), F^l(\cdot)$ are non-analytical and are further examined in [25]. The BNN estimates the distributions of the constrained DS voltage and loading variables using the available measurements and pseudo-measurements. Using these distributions, the p^r, and the estimated flexibility impact of each FSP combination, the proposed approach evaluates the feasibility of each flexibility combination. The approach then aggregates the filtered feasible combinations as in [25] and returns the flexibility area. Thus, the proposed approach adds minor computational burden to [25], mainly for the BNN estimation. Figure 2 illustrates how the considered uncertainty and accepted safety by the TSO can influence the flexibility considered feasible. For the TSO-selected safety and uncertainty levels in Figure 2, the impact from the flexibility combination could result in over-voltage. Thus, in this example, the proposed algorithm would consider the combination non-feasible.

The proposed approach first assumes a local approximation of voltage and loading sensitivity to a flexibility action. If two DS states $\{\hat{X}(t), \hat{V}(t), \hat{L}(t)\}, \{\tilde{X}(t), \tilde{V}(t), \tilde{L}(t)\}$ are within a small ball of radius ϵ, then $\forall U(t,\tau) \in \Omega^U(t,\tau)$:

$$F^v(\hat{X}(t), \hat{V}(t), U(t,\tau)) \approx F^v(\tilde{X}(t), \tilde{V}(t), U(t,\tau)) \tag{5}$$

$$F^l(\hat{X}(t), \hat{L}(t), U(t,\tau)) \approx F^l(\tilde{X}(t), \tilde{L}(t), U(t,\tau)) \tag{6}$$

Thus, this local approximation assumption represents the sensitivity functions as similar for close DS states. Hence, with error $e \leq \epsilon$ between an estimated $\{\tilde{X}(t), \tilde{V}(t), \tilde{L}(t)\}$, and a real $\{\hat{X}(t), \hat{V}(t), \hat{L}(t)\}$, the voltage and loading sensitivities of DS components to flexibility shifts are assumed similar.

Let Ω^b be the set of DS buses excluding the PCC buses, Ω^λ the set of DS lines and transformers, that is, $|\Omega^\lambda| = k$. Let $\bar{X}(t), \bar{V}(t), \bar{L}(t)$ be the expected network state values considering the measured and pseudo-measured variables, that is, measured variables have lower noise. Considering the limited observability, any estimated values for voltage and loading $\bar{V}(t), \bar{L}(t)$ will have some mismatch $\varepsilon_v^e, \varepsilon_l^e$ from the actual voltage and loading values $V(t), L(t)$. The proposed probabilistic estimator represents this imperfect estimation and task's stochasticity by estimating the distribution for each bus voltage $v_i(t) \forall i \Omega^b$, and line or transformer loading $l_j(t) \forall j \in \Omega^\lambda$, given the measured and pseudo-measured values. Thus, the proposed probabilistic model estimates a normal distribution for each $v_i(t), l_j(t)$ as $\mathcal{N}(\mu_{v_i}(t), \sigma^2_{v_i}(t)), \forall i \in \Omega^b, \mathcal{N}(\mu_{l_j}(t), \sigma^2_{l_j}(t)), \forall j \in \Omega^\lambda$, where $\mu_{v_i}(t), \sigma^2_{v_i}(t)$ are the model's expected value and standard deviation for the voltage magnitude of bus i at t, and $\mu_{l_j}(t), \sigma^2_{l_j}(t)$ are the model's expected value and standard deviation for the loading of line or transformer j at t. Using these standard deviations, the proposed approach applies the quantile function $H(\cdot)$ (inverse of cumulative distribution function) to select a safety level for the network constraints. The quantile function takes a probability and a distribution as inputs and returns the margin r. The probability of a sample being inside the margin r equals the input probability p^r. For example, for $p = 84\%$, the $H(\cdot)$ of a normal distribution will return $r = 1$, for the margins $\mu_{v_i}(t) \pm r \cdot \sigma_{v_i}(t)$. Hence, 84% of samples from the input distribution would be within the $\mu_{v_i}(t) \pm \sigma_{v_i}(t)$. Through $H(\cdot)$, the TSOs can select the safety percentage for the FA, $p^r\%$, being the probability of the actual network state being within the estimated margins, modifying (4), (2) and (3) to:

$$Y(t+\tau) = \{y(t+\tau) = F^A(\bar{X}(t), \bar{L}(t), \bar{V}(t), U(t,\tau)|, U(t,\tau)$$
$$\in \Omega^U(t,\tau), c_v^{\min} \leq \mu_{v_i}(t+\tau) - r \cdot \sigma_{v_i}(t) \forall i \in \Omega^b,$$

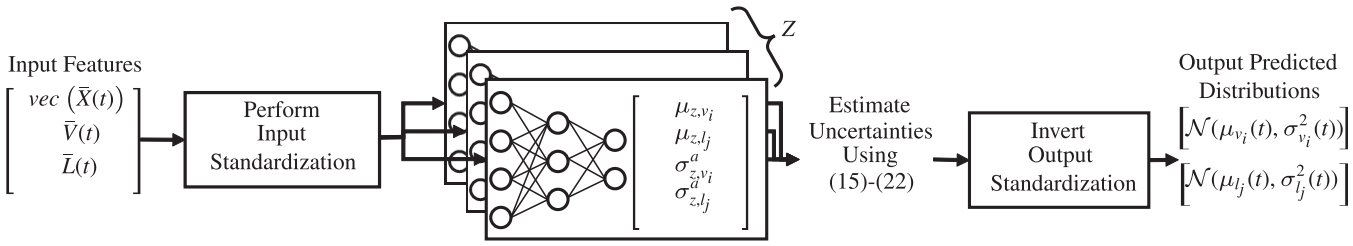


FIGURE 3 | Overview of proposed BNN model output estimation process.

$$\mu_{v_i}(t+\tau) + r \cdot \sigma_{v_i}(t) \le c_v^{\max} \forall i \in \Omega^b, \mu_{l_j}(t+\tau) + r \cdot \sigma_{l_j}(t) \le c_l^{\max} \forall j \in \Omega^\lambda\}, \tag{7}$$

$$r = H(p^r), \tag{8}$$

$$\mu_{v_i}(t+\tau) = \mu_{v_i}(t) + F^v(\bar{X}(t), \bar{V}(t), U(t,\tau)), \forall i \in \Omega^b, \tag{9}$$

$$\mu_{l_j}(t+\tau) = \mu_{l_j}(t) + F^l(\bar{X}(t), \bar{L}(t), U(t,\tau)), \forall j \in \Omega^\lambda. \tag{10}$$

2.4 | Two-PCC FA Estimation

Considering two PCCs in FA estimation would require aggregating feasible points in four dimensions, the active and reactive power per PCC. Visualizing these four dimensions can be challenging, while this increased dimensionality can worsen the implementation complexity or computational burden for the FA estimation. The proposed approach considers that the TSO can select and decouple the flexibility for active and reactive power. Alternatively, estimating two single-PCC FAs, one for each PCC can represent the problem while neglecting dependencies between the PCC operating points.

The proposed approach modifies $y(t+\tau)$ of (7) selecting the first two dimensions as (i) active powers of the two PCCs and (ii) reactive powers of the two PCCs. The third dimension is the density of feasible combinations. This approach initially runs power flows to estimate $F^v(\cdot), F^l(\cdot)$ as in Figure 1. However, the sensitivity matrices' (and impact Tensors') first two dimensions are the TSO's selection of active or reactive powers. If the TSO considers active and reactive power flexibility from the two PCCs, then the DSO can perform two simulations, selecting two of the four dimensions for each simulation. However, this selection will decouple the two sets of dimensions.

Exceeding two PCCs would increase the number of dimensions needed to store the coupled aggregated shifts on each PCC. More importantly, an increased number of PCCs would complicate the decoupling and selection of dimensions to visualize the FAs. Thus, this research is focused on meshed and radial DSs with one or two PCC.

3 | Bayesian Neural Network Based Approach

Figure 3 illustrates the proposed BNN model process in estimating the bus voltage magnitude and the line and transformer loading distributions. The BNN model estimates the distributions per datapoint Z times, each time with different model parameters to approximate the epistemic uncertainty. For each parameter setting $z \in Z$ and input features $\chi = [\text{vec}(\bar{X}), \bar{V}, \bar{L}] \in \mathrm{R}^{3 \cdot n + k}$, the BNN estimates the output means μ_{z,v_i}, μ_{z,l_j} and aleatoric standard deviations $\sigma^a_{z,v_i}, \sigma^a_{z,l_j}$ for bus voltages and line or transformer loading. The loading variables have different scales and variability from the voltage and power injection variables. Hence, the proposed approach first performs standard scaling on all input features.

The proposed supervised BNN uses a training dataset D with $|D|$ data points. During training, the model processes the dataset D in batches. The model uses these batches to compute predictions, compare them to the true labels using a loss function, and update its parameters via backpropagation. Each datapoint $d \in D$ includes input features χ_d and targets $\phi_d = [v_i \forall i \in \Omega^b, l_j \forall j \in \Omega^\lambda] \in \mathrm{R}^{|\Omega^b| + |\Omega^\lambda|}$. vec$(\cdot)$ vectorizes an input matrix and $|\cdot|$ is a set cardinality.

Conventional supervised feed-forward neural networks (FNNs) return point predictions for the above targets ϕ_d given the input features χ_d, and trained parameters W. These FNNs approximate the underlying input-output relationship using the available training data. However, these point predictions are also impacted by the inherent noise within the training data and the data distributions, which might differ from the real-life applications. Therefore, understanding what the models do not know can be critical for ML [35].

BNNs consider the aleatoric and epistemic uncertainties in their estimations. To consider these uncertainties, BNNs do not make point predictions but rather predict distributions of possible outputs for the given inputs:

$$p(v_i|\chi, D) = \int p(v_i|\chi, W) p(W|D) dW, \forall i \in \Omega^b, \tag{11}$$

$$p(l_j|\chi, D) = \int p(l_j|\chi, W) p(W|D) dW, \forall j \in \Omega^\lambda, \tag{12}$$

where $p(\cdot)$ is a probability distribution, $p(v_i|\chi, W), p(l_j|\chi, W)$ are the likelihoods of the outputs given the inputs and model parameters, $p(W|D)$ is the posterior distribution over the model parameters. The integral represents the marginalization over all possible model parameters W. The v_i, l_j in (11) and (12) are the values from the dataset targets, split to return one normal distribution per target.

3.1 | Epistemic Uncertainty

To consider the epistemic uncertainty, BNNs do not have a single setting of parameters but rather use different settings of parameters weighted by the posterior probabilities [16], the $p(W|D)$ term of (11) and (12). As (11) and (12) can be computationally challenging, main approaches such as VI [17], MCD [18] and DE [19] approximate $p(v_i|\chi, D)$, $p(l_j|\chi, D)$. VI uses a variational posterior $q(W|D)$ to approximate the true posterior $p(W|D)$, thus optimizing the parameters of the distribution $q(W|D)$. MCD applies dropout during training and inference time. During inference, the BNN model estimates the output distributions for the same inputs multiple times, with the activated dropout resulting in different neurons being deactivated each time. Therefore, for Z estimations for the same inputs, Z different parameter settings are used, approximating (11) and (12) with:

$$p(v_i|\chi, D) \approx \frac{1}{Z}\sum_{z=1}^{Z} p(v_i|\chi, W_z), \quad W_z \sim q(W|D), \forall i \in \Omega^b, \tag{13}$$

$$p(l_j|\chi, D) \approx \frac{1}{Z}\sum_{z=1}^{Z} p(l_j|\chi, W_z), \quad W_z \sim q(W|D), \forall j \in \Omega^\lambda. \tag{14}$$

To provide Z different parameter settings as in MCD, DE train Z independent FNNs, resulting in Z different parameter settings, applying (13) and (14). With (13) and (14) each pass utilizes different settings, resulting in different estimates to showcase the model's parameter uncertainty.

3.2 | Aleatoric Uncertainty

To consider and estimate aleatoric uncertainty, BNNs include standard deviation $\sigma^a_{z,v_i}, \sigma^a_{z,l_j}$ for each predicted output in the $p(v_i|\chi, W), p(l_j|\chi, W)$ terms of (11), (12), (13) and (14). The proposed BNN for FA estimation considers aleatoric uncertainty as heteroscedastic [35], that is, dependent on each data point rather than being constant for all data, for example, varying environmental factors can impact measurements and pseudo-measurements differently between prediction instances.

3.3 | BNN Model Output Distribution Estimation

The mean value, and the epistemic, aleatoric and total standard deviation per output are:

$$\mu_{v_i} = \frac{1}{Z}\sum_{z=1}^{Z} \mu_{z,v_i}, \quad \forall i \in \Omega^b, \tag{15}$$

$$\sigma_{v_i} = \sqrt{(\sigma^a_{v_i})^2 + (\sigma^e_{v_i})^2}, \tag{16}$$

$$\sigma^a_{v_i} = \sqrt{\frac{1}{Z}\sum_{z=1}^{Z}(\sigma^a_{z,v_i})^2}, \quad \forall i \in \Omega^b, \tag{17}$$

$$\sigma^e_{v_i} = \sqrt{\frac{1}{Z}\sum_{z=1}^{Z}(\mu_{z,v_i} - \mu_{v_i})^2}, \quad \forall i \in \Omega^b, \tag{18}$$

$$\mu_{l_j} = \frac{1}{Z}\sum_{z=1}^{Z} \mu_{z,l_j}, \quad \forall j \in \Omega^\lambda, \tag{19}$$

$$\sigma_{l_j} = \sqrt{(\sigma^a_{l_j})^2 + (\sigma^e_{l_j})^2}, \tag{20}$$

$$\sigma^a_{l_j} = \sqrt{\frac{1}{Z}\sum_{z=1}^{Z}(\sigma^a_{z,l_j})^2}, \quad \forall j \in \Omega^\lambda, \tag{21}$$

$$\sigma^e_{l_j} = \sqrt{\frac{1}{Z}\sum_{z=1}^{Z}(\mu_{z,l_j} - \mu_{l_j})^2}, \quad \forall j \in \Omega^\lambda. \tag{22}$$

(16), (17), (18), (20), (21) and (22), utilize the additive property of variance for independent estimations, before taking its square root to estimate the standard deviation. Equations (17) and (21) estimate the aleatoric variances for each target as the average variance from the Z BNN output standard deviations [35]. Equations (18) and (22) estimate the epistemic variances for each target as the average variance of the mean predictions for each target distribution [35]. Equations (16) and (20) add aleatoric and epistemic variances to obtain the total variance [35].

3.4 | Model Training Loss Function

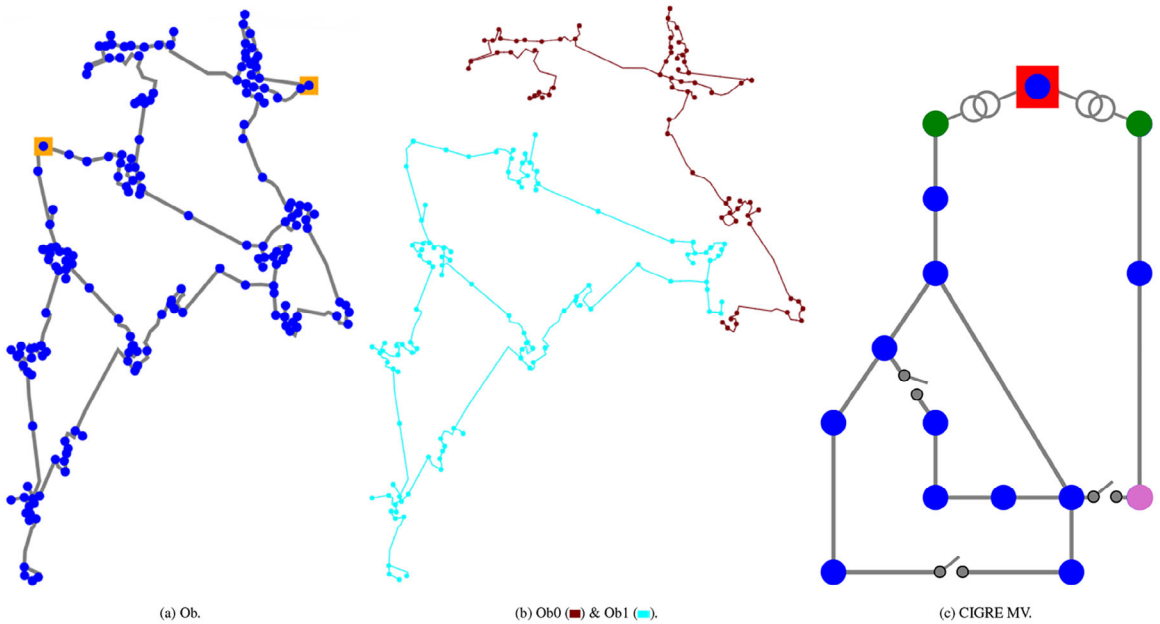
As the BNN estimates a distribution per output, the negative log-likelihood (NLL) function measures how well the estimated distributions represent the actual distribution generating the data [19, 36, 37]. NLL considers the confidence in the aleatoric uncertainty, and the error between the predicted mean and the target values as:

$$L_{\text{NLL}}(\theta) = \frac{1}{|D|}\sum_{d=1}^{|D|}\left(\frac{\phi_d - \mu_{\chi_d}}{2\cdot(\sigma^a_{\chi_d})^2} + \frac{1}{2}\log(\sigma^a_{\chi_d})^2\right) \tag{23}$$

The BNN outputs, $\mu_{\chi_d} = [\mu_{z,v_i}\forall i \in \Omega^b, \mu_{z,l_j}\forall j \in \Omega^\lambda]^T, \sigma^a_{\chi_d} = [\sigma^a_{z,v_i}\forall i \in \Omega^b, \sigma^a_{z,l_j}\forall j \in \Omega^\lambda]^T$ include voltage and loading variables that differ in scale and deviation. For VI, the loss can also include a regularization term on the Kullback–Leibler divergence. Different scales and variances for a subset of outputs can impact the loss function to focus on the subset that produces a higher numerical loss, for example, a 0.5 p.u. voltage mismatch is severe, whereas a 0.5% loading mismatch is minor. Hence, the proposed approach performs standard scaling in the target outputs during training, to ensure (23) is not impacted by the different output units. This scaling results in standardized (15)–(22). Thus, the proposed BNN output estimation (Figure 3) performs inverse standardization on the outputs to obtain p.u. voltage and % loading scaling.

4 | Case Studies

The case studies used the CIGRE MV network, the networks from the separated Oberhein substations 0 (Ob0), 1 (Ob1) and the connected Oberrhein network (Ob) as visualized in Figure 4. All case studies considered the PCC buses, the HV-MV transformers, and the buses below the HV-MV transformers as observable (v, I, P and Q measurements). In the CIGRE network, the wind turbine with 1.5 MW capacity would require RTI measurements. For Oberrhein networks, all buses connected to three lines or more were considered observable.



(a) Ob. (b) Ob0 (■) & Ob1 (■). (c) CIGRE MV.

FIGURE 4 | Test network lines (—), buses (●), HV-MV transformer stations (■), transformers (⊘), external grid (■), CIGRE DS observable buses (●), RTI observable buses (●) and switches (⊘). Panel (b) shows how Ob separates into Ob0 and Ob1.

The case study in Section 4.1 included the CIGRE network of Figure 4 in a meshed (closed switches) and radial topology (open switches) with four FSPs. For each FSP, Section 4.1 considered all shifts $[\Delta P, \Delta Q]$ with steps 0.1 MW constrained by the FSP's initial output. For each shift, 100 different DS operating conditions were sampled by changing the power injection $X(t)$ of all non-observable loads and generators with 0 mean and $\sigma^{\text{pm}} \in \{1\%, 2\%, 5\%, 10\%, 20\%, 50\%\}$ standard deviation, for pseudo-measurement errors.

The case study in Section 4.2 compares the proposed BNN model's performance to modified uncertainty evaluations from FA estimation literature. Appendix A includes the comparative analysis performed on BNN structures: VI, MCD and DE models. The MCD model had the highest consistency under different test set distributions. The training, validation and test datasets were on synthetic data generated using CIGRE, Ob0, Ob1 and Ob networks as summarised in Appendix Table A.1. The MCD model was trained for 500 epochs, with a learning rate of 0.0001.

For the baseline design, existing forecast-based evaluations map the uncertainty of nodal injections to the network constraints using the power flow relationships [12, 14, 21, 22]. For example, [12] assumed a mathematical linear relationship between power injection forecast errors and deviations in voltage magnitudes and angles. However, the formula in [12] is rather complex, includes participation factors, and excludes the loading variables. The approach in [14] runs multiple power flows for different forecasts to approximate the probability of feasibility. This case study approximates the constraint uncertainty mapping using power flows and data statistics. This statistical approximation considers a forecast DS operating condition per hour for each of the Figure 4 systems. For each hour forecast, the statistical approximation estimates 100 'actual' operating conditions with σ^{pm}, and $\psi^{\text{ld}}, \psi^{\text{pv}}$ the load, and PV correlation, respectively. The average standard deviation from the 100 samples returns the $\sigma_{v_i} \forall i \in \Omega^b, \sigma_{l_j} \forall j \in \Omega^\lambda$ for each forecast condition. For each network, the approximated $\sigma_{v_i} \forall i \in \Omega^b, \sigma_{l_j} \forall j \in \Omega^\lambda$ used datasets with $\sigma^{\text{pm}} = 20\%, \psi^{\text{ld}} = 60\%, \psi^{\text{pv}} = 70\%$, like the BNN training datasets. The case study in Section 4.2 also considers a second baseline, where the measured buses included $\sigma_\eta^m, \sigma_\eta^m$ power and loading measurement standard deviation, instead of σ^{pm}, to create a measurement-based statistical approach. The baseline test datasets for each network vary the $\sigma^{\text{pm}} \in \{10\%, 20\%, 50\%\}$ and $\sigma_\eta^m = \sigma_l^m \in \{1\%, 2\%, 5\%\}$ as the BNN test datasets. For each baseline model's test set evaluation, the estimated mean resulted from the power flow solutions, whereas the standard deviations were the training set hourly standard deviations.

The evaluation metrics for the BNNs and the baselines were (i) the prediction interval coverage probability [38–42] for 95% (PICP_{95}):

$$\text{PICP}_{95} = \frac{1}{|\tilde{D}|} \sum_{d=1}^{|\tilde{D}|} \begin{cases} 1 & \text{if } \phi_d \in \left[\mu_{\chi_d} \pm 1.96 \cdot \sigma_{\chi_d}\right] \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

where $\tilde{D}$ are the validation and test sets, and $\sigma_{\chi_d} = [\sigma_{v_i} \forall i \in \Omega^b, \sigma_{l_j} \forall j \in \Omega^\lambda]^T$. The PICP_{95} metric shows how well the model confidence reflects the actual test data, that is, PICP_{95} should be as close to 0.95, whereas $\text{PICP}_{95} > 0.95$ shows under-confident model, and $\text{PICP}_{95} < 0.95$ overconfident model. An overconfident model estimates too narrow uncertainty intervals. Additional evaluation metrics are: (ii) the root mean squared error (RMSE) between the mean predictions and the actual values [19] (should be as close to zero as possible) (iii) the negative log-likelihood [16, 19, 36] (should have the lowest value).

The case study in Section 4.3 estimates safety-constrained FAs in the OB1 with, $c_v^{\max} = 1.05, c_v^{\min} = 0.95$, and five FSPs. The active and reactive power discretization steps were 0.05 MW, 0.1 MVAR, respectively. The $\sigma_{v_i} = 0.01$ p.u.$\forall i \in \Omega^b$,

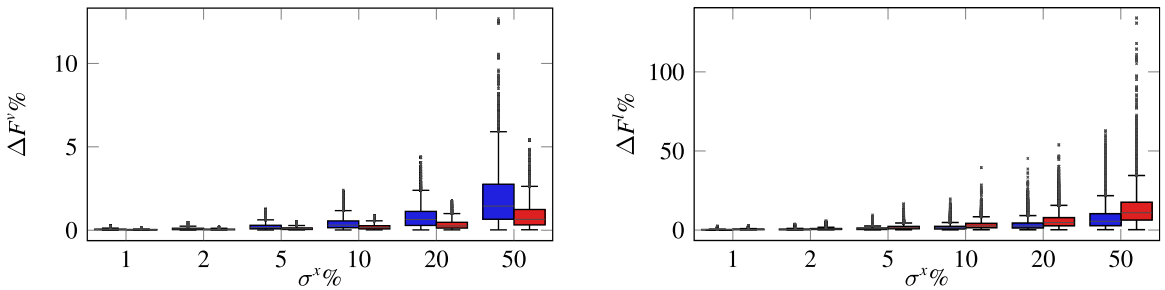


FIGURE 5 | Voltage and loading sensitivity deviations for operating conditions with different pseudo-measurement standard deviation for the CIGRE DS radial (■) and mesh (■) topologies.

$\sigma_{l_j} = 4\%$ for all the lines and 2% for the measured transformers. The FSP flexibilities were any power setpoint within each FSP's nominal apparent power.

The case study in Section 4.4 estimates safety-constrained FAs in the OB with two PCCs, with seven FSPs. The active and reactive power discretization steps were 0.03 MW, 0.03 MVAR, respectively. The uncertainties were the same as Section 4.3. Due to the initial DS conditions being outside the $0.95 - 1.05$ p.u. limits, the constraints were $c_v^{\min} = 0.9$ p.u. and $c_v^{\max} = 1.1$ p.u. The FSP flexibilities for the active power scenarios were any setpoints below the initial active power. For reactive powers, the flexibilities were between -150% and 150% from the initial reactive powers for the loads and between -50% and 50% from the initial active power outputs for the generators (as the initial setpoints had 0 reactive power).

4.1 | Sensitivity Local Approximation

This case study analyses the impact of different pseudo-measurement deviation levels on the sensitivity between DS operating conditions, for the validity of the assumption in Section 2.3. The FA estimation algorithm approximates the sensitivity of network buses $F^v(\cdot)$ and lines $F^l(\cdot)$ to FSP shifts using measured and pseudo-measured values $\{\hat{X}(t), \hat{V}(t), \hat{L}(t), U(t,\tau)\}$ instead of the actual, unobserved values $\{\tilde{X}(t), \tilde{V}(t), \tilde{L}(t), U(t,\tau)\}$, assuming F^v, F^l to remain approximately constant for pseudo-measured operating conditions close to the real operating conditions. Figure 5 compares the absolute percentage difference in sensitivities between actual and pseudo-measured conditions $(\Delta F^v\%, \Delta F^l\%)$.

The simulations only included $F^v(\tilde{X}(t), \tilde{V}(t), U(t,\tau)) > 0.0005$ p.u., $F^l(\tilde{X}(t), \tilde{L}(t), U(t,\tau)) > 0.5\%$, to avoid instabilities from low-sensitive components to FSP shifts.

The results had average line loading sensitivites $F^l(\tilde{X}(t), \tilde{L}(t), U(t,\tau))$ between $[-16.52\%, 9.23\%]$ and average voltage sensitivities $F^l(\tilde{X}(t), \tilde{L}(t), U(t,\tau))$ between $[-0.017, 0.019]$ p.u. for the radial network. For the mesh network, the average loading and voltage sensitivities were $[-4.32\%, 5.13\%]$, and $[-0.009, 0.008]$ p.u., respectively. From the results in Figure 5, the sensitivities remain approximately consistent, with an average deviation less than 5% for loading and less than 1% for voltage with pseudo-measurements with 20% or less standard deviation. For 50% deviation, the loading sensitivities deviate by an average of $\approx 11\%$ for loading in the meshed network.

ΔF^l is higher in the meshed network, whereas ΔF^v is higher in the radial network. Addressing the first observation, in radial networks, the load flow paths are relatively limited; in mesh networks, the power injection deviations can redistribute the load flow paths. Addressing the second observation, the limited paths in radial networks could mean that a different power injection on a bus would impact all upstream or downstream bus voltages.

4.2 | BNN Comparison With Baselines

Figure 6 compares the PICP_{95}, RMSE and NLL between the baselines and the proposed BNN of Section 3 for varying test set noise deviations and networks. Table 1 includes the PICP_{95}, RMSE and NLL values with the best performing model highlighted for each noise and network scenario. The results indicate that the network structure and noise levels impact the BNN less than the baselines. In PICP_{95}, the BNN outperforms the baselines with an average $\text{PICP}_{95}v = 0.91$, and $\text{PICP}_{95}l = 0.94$ compared to the measurement-based $\text{PICP}_{95}v = 0.73$, and $\text{PICP}_{95}l = 0.78$, and forecast-based $\text{PICP}_{95}v = 0.67$, and $\text{PICP}_{95}l = 0.77$. The baseline models consistently underestimate the uncertainty, with values significantly lower than 0.95, except in the lower-noise scenarios for loading variables. In these scenarios, $0.9 \leq \text{PICP}_{95}l \leq 0.95$. However, this means that the uncertainty estimation for loading can be accurate if the noise is half of the one considered during estimation. The baseline exclusion of epistemic uncertainty in the estimations contributes to this significant underperformance compared to the BNN. As Figure 7 shows, the epistemic uncertainty can be a substantial part of the total uncertainty. Figure 6 shows that the BNN drastically improves the voltage and loading RMSE and NLL compared to the baselines. Further, BNN is less impacted by the different noise levels and networks than the baselines. The BNN has, on average, more than four times improved voltage RMSE and more than three times improved loading RMSE compared to the baselines.

Figure 7 shows example MCD predictions on the vertical axis with measurement inputs on the horizontal axis. The uncertainties of Figure 7 have a 95% confidence margin (two standard deviations). To improve the visualization clarity, 90% of the datapoints were removed in Figure 7 in densely populated regions. The figures show that epistemic uncertainty depends on the dataset's number of examples. In the regions with more concentrated

TABLE 1 | Baseline and MCD BNN model performances.

Dataset	Forecast-based statistical baseline						Measurement-based statistical baseline						MCD BNN					
	$PICP_{95}$		RMSE		NLL		$PICP_{95}$		RMSE		NLL		$PICP_{95}$		RMSE		NLL	
	v	*l*	*v*	*l*	*v*	*l*	*v*	*l*	*v*	*l*	*v*	*l*	*v*	*l*	*v*	*l*	*v*	*l*
Te1.C	0.62	0.93	0.077	15.7	−0.64	3.6	0.82	**0.95**	0.032	9.9	−2.26	3.13	**0.99**	1	**0.012**	**5.3**	**−3.28**	**2.73**
Te2.C	0.59	0.81	0.086	20	−0.18	4.1	0.77	0.87	0.037	12.6	−1.92	3.46	**0.99**	**0.99**	**0.011**	**5.5**	**−3.27**	**2.77**
Te3.C	0.50	0.56	0.122	37.3	2.44	7.87	0.64	0.69	0.055	23.2	0.56	5.97	**0.92**	**0.94**	**0.016**	**7.8**	**−2.93**	**3.04**
Te1.OB0	0.79	**0.95**	0.08	17.5	−1.08	3.47	0.81	**0.95**	0.07	16.3	−1.27	3.41	**1**	0.99	**0.013**	**5**	**−2.87**	**2.91**
Te2.OB0	0.72	0.81	0.091	22.3	−0.67	4.03	0.73	0.8	0.079	20.8	−0.84	3.98	**0.99**	**0.98**	**0.014**	**5.3**	**−2.89**	**2.92**
Te3.OB0	0.58	0.54	0.122	37.1	1.5	7.92	0.58	0.53	0.108	34.8	1.47	8	**0.84**	**0.87**	**0.03**	**9**	**−2.23**	**3.41**
Te1.OB1	0.8	0.9	0.108	18.3	−0.69	3.5	0.83	0.9	0.087	15.9	−1.04	3.34	**0.89**	**0.94**	**0.027**	**5.9**	**−2.6**	**2.74**
Te2.OB1	0.74	0.8	0.112	20.6	−0.47	3.88	0.77	0.79	0.091	17.9	−0.8	3.75	**0.92**	**0.96**	**0.016**	**4.3**	**−2.81**	**2.63**
Te3.OB1	0.6	0.58	0.139	30.9	0.95	6.56	0.61	0.57	0.115	26.9	0.74	6.53	**0.79**	**0.91**	**0.02**	**5**	**−2.24**	**2.82**
Te1.OB	0.79	0.91	0.102	19.9	−0.8	3.53	0.82	0.91	0.083	17.5	−1.12	3.4	**0.85**	**0.92**	**0.032**	**7.9**	**−2.45**	**2.98**
Te2.OB	0.73	0.81	0.105	22	−0.57	3.93	0.75	0.8	0.086	19.3	−0.86	3.82	**0.95**	**0.97**	**0.015**	**4.7**	**−3.06**	**2.73**
Te3.OB	0.59	0.58	0.136	34.3	1.3	7.2	0.6	0.57	0.113	30.2	1.16	7.23	**0.76**	**0.9**	**0.019**	**6**	**−2.3**	**3.01**
Average	**0.67**	**0.77**	**0.107**	**24.7**	**0.09**	**4.97**	**0.73**	**0.78**	**0.08**	**20.4**	**−0.52**	**4.67**	**0.91**	**0.94**	**0.019**	**6**	**−2.74**	**2.89**

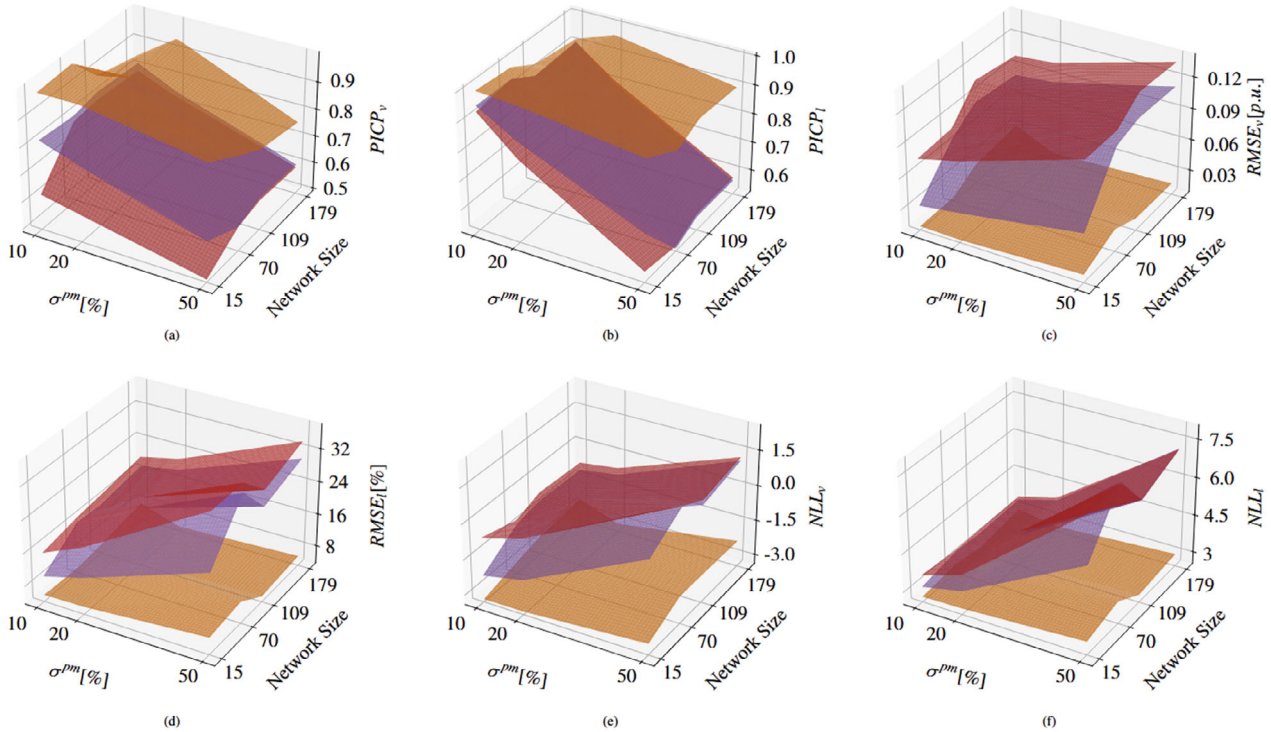


FIGURE 6 | Comparison of forecast-based statistical baseline (), measurement-based statistical baseline () and MCD BNN () models under varying pseudo-measurement (and measurement) noise and network size. The network size corresponds to the number of buses.

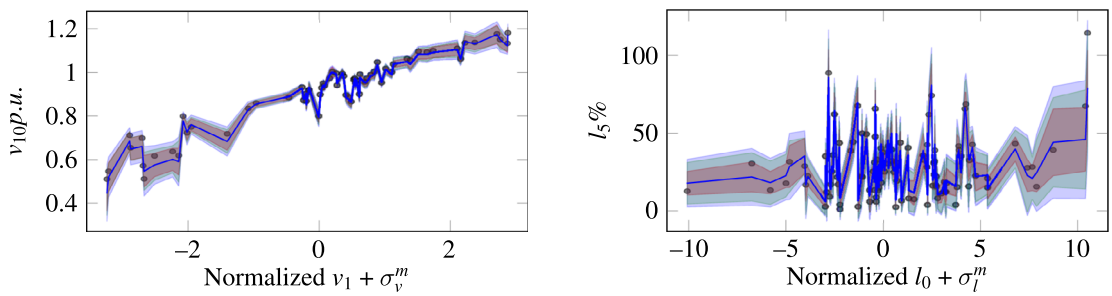


FIGURE 7 | Voltage and loading predictions by MCD model for CIGRE DS with mean (), epistemic uncertainty (), aleatoric uncertainty (), total uncertainty (), and true data ().

datapoints (closer to 0 in the horizontal axes), the epistemic uncertainty is smaller than the less-populated regions. This difference highlights that more training examples can reduce the model uncertainty in the parameter space and the epistemic uncertainty. Figure 7 also illustrates that the aleatoric uncertainty varies between datapoints.

Considering the computational burdens, the baselines would not require additional estimations as the FA estimation would solve the power flows to estimate the sensitivities in Figure 1. However, the MCD BNN would add a negligible prediction duration, as Table A.2 shows (between 0.09–0.3).

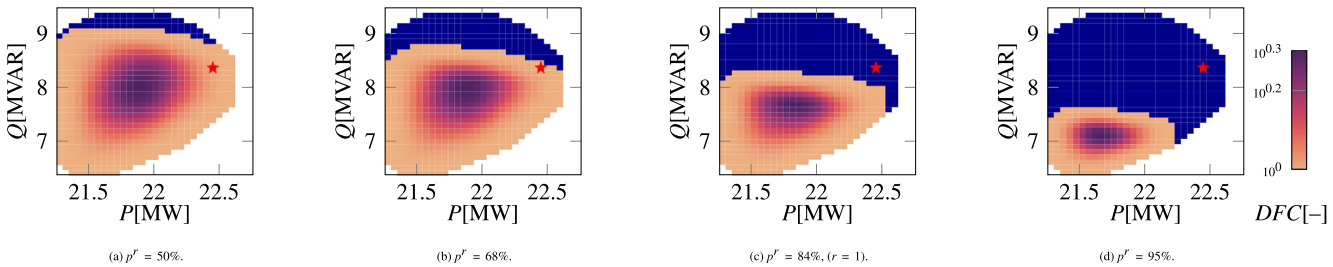
4.3 | Measurement Uncertainty-Controlled FA Estimation

This section shows the proposed FA estimation process of Section 2.3. Using the BNN mean prediction and uncertainty, the TSO can select a safety margin on the FAs with the p^r input as in Figure 1. Figure 8 shows the proposed approach for FA estimation with different safety levels, $p^r = 50\%, 68\%$ ($\pm$ the estimated standard deviation), 84% ($r = 1$) and 95% ($\pm 2\times$ the estimated standard deviation). DFC is the density of feasible combinations for each FA point. The increasing safety margin further restricts the FA feasible space. The uncertainty levels in Figure 8c,d shows that even the initial DS operating condition can be out of the selected safety margins.

The FA estimation durations using the Google Colab's A100 GPU required 15.6, 16.2, 17.1 and 18.7s for Figure 8a–d, respectively. The difference in computational time as the margins increase is due to the increased number of network buses and lines that can reach the network constraints.

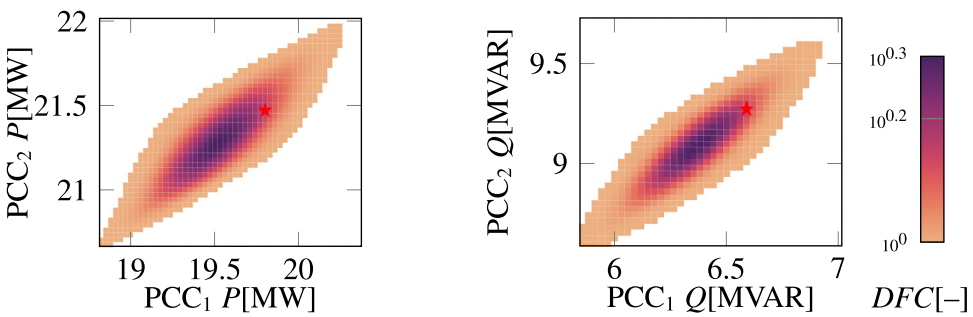
4.4 | Two-PCC FA Estimation

This case study shows the proposed FA estimation approach for DS with two PCCs as explained in Section 2.4. Figure 9



(a) p^r = 50%. (b) p^r = 68%. (c) p^r = 84%, (r = 1). (d) p^r = 95%.

FIGURE 8 | Safety-Constrained FAs of OB1 5 FSPs with different safety probabilities p^r for TSO selection. Feasible shifts ( ). Not feasible FSP shift combinations ( ). Initial operating point ( ).



(a) Active Powers. (b) Reactive Powers.

FIGURE 9 | Safety-constrained FAs of OB with seven FSPs for active power flexibility support in (a) and reactive power flexibility support in (b). Feasible shifts ( ). Initial operating point ( ).

shows the resulting FAs for the two PCCs when considering only active power in Figure 9a or only reactive power flexibility in Figure 9b. The diagonal shapes in Figure 9 show the correlation between the two PCCs' power exchange, for example, a flexibility shift to decrease the active power at PCC_1 can also reduce the active power at PCC_2 for the majority of flexibility combinations. However, some combinations (low DFC) can also shift each PCC's power unequally, for example, decreasing the active power in Figure 9a of PCC_1 by 0.5 MW and maintaining the active power of PCC_2 at 21.5 MW.

Using these FAs, the TSO can select any feasible active or reactive power setpoint for the correlated PCCs. The FA estimation required 11.1 s for Figure 9a and 7.5 s for Figure 9b with the A100 GPU.

4.5 | Discussion

Through the proposed approach, TSOs can consider real-time uncertainties when accounting for the available flexibility, unlike prior FA approaches. These real-time uncertainties are due to limited measurements, input noise, and model parameters. Using BNNs, the estimated uncertainties can generalize better to different noise levels and networks compared to the statistical baselines extended from [12, 14, 21, 22].

The results of Section 4.2 show the proposed BNN structure improves the uncertainty estimation in all metrics compared to the statistical baselines. The baselines underestimate the uncertainty in the majority of tests ($PICP_{95} < 0.95$). The measurement and pseudo-measurement noise increase impacts the baselines significantly more than the MCD BNN in all metrics. In terms of computational burden, MCD requires low training times (0.8–1.2 h) and prediction duration 0.1–0.3). The computational burden added to [25] for FA estimation is mainly the BNN prediction duration, with FA estimations requiring between 7.5–18.7. Prior approaches for uncertainty consideration in FA estimation offered simplicity, but showed limited performance compared to the proposed BNN structure. This also highlights the need for more advanced, possibly learning-based approaches for uncertainty quantification.

The TSO's selection for safety probabilities can impact the size of the FA, as well as the risks of selecting operating conditions that would be non-feasible for the DS constraints. A p^F of 95% could be highly restrictive, as it would require all voltage and loading conditions in the DS to have a margin of two times the uncertainty's standard deviation. A selection of $p^F = 68\%$ could provide sufficient safety margin, while also permitting a broader FA.

The two-PCC results in Section 4.4 show the algorithm's capability for fast estimation, and the interdependency between the two PCCs, as the areas show diagonal rotation, for example, shifts increasing PCC_1P also increase PCC_2P. Nevertheless, decoupling the active and reactive powers neglects the active-reactive power interdependency, where active power shifts can cause reactive power deviations and the opposite.

The results of Appendix A indicate BNN approaches can perform well for the power-system task, maintaining a relatively consistent $PICP_{95}$, RMSE, and NLL when trained in different networks. Between VI, MCD and DE BNNs, the MCD has, on average, the best l $PICP_{95}$, RMSE (for v and l), and NLL (for v and l).

The proposed approach was evaluated in synthetic data. Nevertheless, the case studies used a realistic percentage of measurement units [11], varied operating conditions, network topology and structure, and possible measurement and pseudo-measurement noise. These variations were selected to reflect the possible impact of deployment in real life under various scenarios. Nevertheless, retraining is required under topological changes that were absent during training. Nevertheless, the proposed MCD model's low training time requirements would enable DSOs to re-train the model with new data collections.

The proposed approach can be extended to support co-optimization with economic objectives. The estimated FA can be embedded within optimization frameworks, such as congestion management, operational planning and flexibility procurement mechanisms. Furthermore, the approach can be extended to multi-energy systems by adding variables in the input space, representing heat, gas or storage dynamics. In such settings, the approach can estimate coupled FAs across energy sources.

The sensitivity local approximation may cause incorrect FA estimation for loading approximations in cases with high pseudomeasurement noise (e.g., 50% noise), mainly in meshed DS topologies. To mitigate such cases, data-driven approaches could be utilized to provide pseudomeasurements based on similar prior network states [43]. The local sensitivity approximation could be ineffective after topological changes since the flow of energy would follow alternative paths. Correct anticipation of the network topology is crucial for all FA estimation algorithms to evaluate the feasibility of flexibility combinations.

A limitation of the proposed approach is the underconfidence in test sets with lower noise than the training sets ($PICP_{95} > 0.95$). To mitigate this limitation in case of low pseudo-measurement and measurement noise, DSOs can utilize real-life data to train another BNN MCD in ≈ 1 h. The demonstrated BNN latency showed minimal impact on the FA estimation duration. The underlying FA estimation, utilizing the structure of [25] would similarly have consistent speed under varying network structures, but may require high memory usage under increasing number of FSPs, for the tensor operations.

5 | Conclusions

System operators can use the proposed approach to select safety levels in real-time FA estimations, considering the limited observability in DS. Aiming for real-time operation, the proposed approach is fast and considers the uncertainty from limited real-time DS measurements. In addition, with the proposed approach, TSOs can select active and reactive power flexibility aggregation in DS with two-PCCs. The proposed BNN structure improves the accuracy of the estimated uncertainty and RMSE compared to the baselines. The proposed BNN improves the RMSE compared to baselines by an average $\approx 4\times$ for voltage and $\approx 3\times$ for loading. The average MCD BNN test uncertainty using $PICP_{95}$ shows slight overconfidence with 0.91 for voltage and 0.94 for loading, unlike the baselines that show significant overconfidence.

Future work includes evaluating the BNN performance in real data from distribution networks and studying the impact of alternative measurement locations in DS. Future work could explore how to improve the network measurements for FAs, considering recent approaches on optimal measurement placement and observability levels for DSSE [44, 45]. Such work could provide information on how to reduce the uncertainty in FA, whereas this paper proposes a method to anticipate and consider the uncertainty in FAs.

Supporting Information

Additional supporting information may be found in the online version of the article on the publisher's website.

Author Contributions

Demetris Chrysostomou: conceptualization, methodology, data curation, software, validation, formal analysis, investigation, visualization, writing – original draft. **Jose Luis Rueda Torres:** conceptualization, methodology, writing – original draft, supervision, project administration. **Jochen Lorenz Cremer:** conceptualization, methodology, writing – original draft, supervision, project administration.

Acknowledgements

This research is part of the research program 'MegaMind - Measuring, Gathering, Mining and Integrating Data for Self-Management in the Edge of the Electricity System,' (partly) financed by the Dutch Research Council (NWO) through the Perspectief program under number P19-25.

Funding

This research is part of the research program "MegaMind - Measuring, Gathering, Mining and Integrating Data for Self-Management in the Edge of the Electricity System," funded by the Dutch Research Council (NWO) through the Perspectief program (grant number P19-2).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are openly available in [46].

References

1. A. G. Givisiez, K. Petrou, and L. F. Ochoa, "A Review on TSO-DSO Coordination Models and Solution Techniques," *Electric Power Systems Research* 189 (2020): 106659.

2. M. Bolfek and T. Capuder, "An Analysis of Optimal Power Flow Based Formulations Regarding DSO-TSO Flexibility Provision," *International Journal of Electrical Power & Energy Systems* 131 (2021): 106935.

3. F. Capitanescu, "TSO-DSO Interaction: Active Distribution Network Power Chart for TSO Ancillary Services Provision," *Electric Power Systems Research* 163 (2018): 226–230.

4. N. Savvopoulos, C. Y. Evrenosoglu, T. Konstantinou, T. Demiray, and N. Hatziargyriou, "Contribution of Residential PV and BESS to the Operational Flexibility at the TSO-DSO Interface," in *International Conference on Smart Energy Systems and Technologies (SEST)* (IEEE, 2021), 1–6.

5. M. Heleno, R. Soares, J. Sumaili, R. J. Bessa, et al., "Estimation of the Flexibility Range in the Transmission-Distribution Boundary," in *IEEE Eindhoven PowerTech* (IEEE, 2015), 1–6.

6. G. Prionistis, C. Vournas, and M. Vrakopoulou, "A Fast Method to Approximate the Flexibility Region of an Active Distribution Network in *PQ* Space," in *IEEE Belgrade PowerTech* (IEEE, 2023), 1–6.

7. B. Azimian, R. S. Biswas, S. Moshtagh, A. Pal, L. Tong, and G. Dasarathy, "State and Topology Estimation for Unobservable Distribution Systems Using Deep Neural Networks," *IEEE Transactions on Instrumentation and Measurement* 71 (2022): 1–15.

8. B. Habib, E. Isufi, W. van Breda, A. Jongepier, and J. L. Cremer, "Deep Statistical Solver for Distribution System State Estimation," *IEEE Transactions on Power Systems* 39, no. 2 (2024): 3178–3189.

9. A. Primadianto and C. N. Lu, "A Review on Distribution System State Estimation," *IEEE Transactions on Power Systems* 32, no. 5 (2017): 3875–3883.

10. D. Chrysostomou, J. L. Rueda Torres, and J. L. Cremer, "Exploring Operational Flexibility of Active Distribution Networks With Low Observability," in *IEEE Belgrade PowerTech* (IEEE, 2023), 1–6.

11. M. Q. Tran, "Voltage Monitoring and Control in the Active Distribution Networks," (PhD diss., Delft University of Technology, 2024).

12. T. Chen, Y. Song, D. J. Hill, and A. Y. Lam, "Enhancing Flexibility at the Transmission-Distribution Interface With Power Flow Routers," *IEEE Transactions on Power Systems* 37, no. 4 (2021): 2794–2806.

13. M. Kalantar-Neyestanaki, F. Sossan, M. Bozorg, and R. Cherkaoui, "Characterizing the Reserve Provision Capability Area of Active Distribution Networks: A Linear Robust Optimization Method," *IEEE Transactions on Smart Grid* 11, no. 3 (2020): 1864–1875.

14. D. M. Gonzalez, J. Hachenberger, J. Hinker, F. Rewald, et al., "Determination of the Time-Dependent Flexibility of Active Distribution Networks to Control Their TSO-DSO Interconnection Power Flow," in *Power Systems Computation Conference (PSCC)* (IEEE, 2018), 1–7.

15. E. Goan and C. Fookes, "Bayesian Neural Networks: An Introduction and Survey," in *Case Studies in Applied Bayesian Data Science: CIRM Jean-Morlet Chair, Fall 2018* (Springer, 2020), 45–87.

16. A. G. Wilson and P. Izmailov, "Bayesian Deep Learning and a Probabilistic Perspective of Generalization," *Advances in Neural Information Processing Systems* 33 (2020): 4697–4708.

17. C. Zhang, J. Bütepage, H. Kjellström, and S. Mandt, "Advances in Variational Inference," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41, no. 8 (2018): 2008–2026.

18. Y. Gal and Z. Ghahramani, "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning," in *International Conference on Machine Learning* (PMLR, 2016), 1050–1059.

19. B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and Scalable Predictive Uncertainty Estimation Using Deep Ensembles," *Advances in Neural Information Processing Systems* 30 (2017): 6402–6413.

20. N. Savvopoulos, N. Hatziargyriou, and H. Laaksonen, "A Holistic Approach to the Efficient Estimation of Operational Flexibility From Distributed Resources," *IEEE Open Access Journal of Power and Energy* 11 (2024): 230–241.

21. S. Ge, Z. Xu, H. Liu, C. Gu, and F. Li, "Flexibility Evaluation of Active Distribution Networks Considering Probabilistic Characteristics of Uncertain Variables," *IET Generation, Transmission & Distribution* 13, no. 14 (2019): 3077–3087.

22. N. Majumdar, P. Kengkat, R. Yermekbayev, and L. Hofmann, "Reliability Parameterised Distribution Grid Flexibility Aggregation Considering Renewable Uncertainties," in *2023 58th International Universities Power Engineering Conference (UPEC)* (IEEE, 2023), 1–6.

23. J. Silva, J. Sumaili, R. J. Bessa, et al., "Estimating the Active and Reactive Power Flexibility Area at the TSO-DSO Interface," *IEEE Transactions on Power Systems* 33, no. 5 (2018): 4741–4750.

24. A. Churkin, W. Kong, J. N. M. Gutierrez, P. Mancarella, et al., "Assessing Distribution Network Flexibility via Reliability-Based *pq* Area Segmentation," in *IEEE Belgrade PowerTech* (IEEE, 2023), 1–6.

25. D. Chrysostomou, J. L. R. Torres, and J. L. Cremer, "Tensor Convolution-Based Aggregated Flexibility Estimation in Active Distribution Systems," *IEEE Transactions on Smart Grid* 15, no. 6 (2024): 5621–5633.

26. R. J. De Groot, J. Morren, and J. G. Slootweg, "Closed-Ring Operation of Medium Voltage Distribution Grids: Theory Meets Practice," in *23rd International Conference and Exhibition on Electricity Distribution* (CIRED, 2015), 1–5.

27. F. Erdinç, A. Çiçek, O. Erdinç, R. Yumurtacı, M. Z. Oskouei, and B. Mohammadi-Ivatloo, "Decision-Making Framework for Power System With RES Including Responsive Demand, ESSS, EV Aggregator and Dynamic Line Rating as Multiple Flexibility Resources," *Electric Power Systems Research* 204 (2022): 107702, https://www.sciencedirect.com/science/article/pii/S0378779621006830.

28. O. Kırat, A. Çiçek, and T. Yerlikaya, "A New Artificial Intelligence-Based System for Optimal Electricity Arbitrage of a Second-Life Battery Station in Day-Ahead Markets," *Applied Sciences* 14, no. 21 (2024): 10032, https://www.mdpi.com/2076-3417/14/21/10032.

29. J. A. D. Massignan, J. B. A. London, M. Bessani, C. D. Maciel, R. Z. Fannucchi, and V. Miranda, "Bayesian Inference Approach for Information Fusion in Distribution System State Estimation," *IEEE Transactions on Smart Grid* 13, no. 1 (2022): 699–709.

30. U. Kuhar, M. Pantoš, G. Kosec, and A. Švigelj, "The Impact of Model and Measurement Uncertainties on a State Estimation in Three-Phase Distribution Networks," *IEEE Transactions on Smart Grid* 10, no. 3 (2019): 2850–2859.

31. K. R. Mestav, J. Luengo-Rozas, and L. Tong, "State Estimation for Unobservable Distribution Systems via Deep Neural Networks," in *2018 IEEE Power and Energy Society General Meeting (PESGM)* (IEEE, 2018), 1–5.

32. P. A. Pegoraro, A. Angioni, M. Pau, A. Monti, C. Muscas, F. Ponci, and S. Sulis, "Bayesian Approach for Distribution System State Estimation With Non-Gaussian Uncertainty Models," *IEEE Transactions on Instrumentation and Measurement* 66, no. 11 (2017): 2957–2966.

33. M. Vanin, T. Van Acker, R. D'hulst, and D. Van Hertem, "Exact Modeling of Non-Gaussian Measurement Uncertainty in Distribution System State Estimation," *IEEE Transactions on Instrumentation and Measurement* 72 (2023): 1–11.

34. K. R. Mestav, J. Luengo-Rozas, and L. Tong, "Bayesian State Estimation for Unobservable Distribution Systems via Deep Learning," *IEEE Transactions on Power Systems* 34, no. 6 (2019): 4910–4920.

35. A. Kendall and Y. Gal, "What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?" *Advances in Neural Information Processing Systems* 30 (2017): 5574–5584.

36. S. K. Lind, Z. Xiong, P. E. Forssén, and V. Krüger, "Uncertainty Quantification Metrics for Deep Regression," *Pattern Recognition Letters* 186 (2024): 1–7.

37. Y. Lu, Y. He, G. Wang, C. Fang, and Z. Sun, "Multi-Target Neurallss Regression With Learned Confidence Space," in *2024 International Joint Conference on Neural Networks (IJCNN)* (IEEE, 2024), 1–8.

38. L. Ø. Bentsen, N. D. Warakagoda, R. Stenbro, and P. Engelstad, "Relative Evaluation of Probabilistic Methods for Spatio-Temporal Wind Forecasting," *Journal of Cleaner Production* 434 (2024): 140209.

39. N. Tagasovska and D. Lopez-Paz, "Single-Model Uncertainties for Deep Learning," *Advances in Neural Information Processing Systems* 32 (2019): 6417–6428.

40. P. Weng, Y. Tian, Y. Liu, and Y. Zheng, "Time-Series Generative Adversarial Networks for Flood Forecasting," *Journal of Hydrology* 622 (2023): 129718.

41. Q. Zhu, Y. Xu, Q. Lin, Z. Ming, and K. C. Tan, "Clustering-Based Short-Term Wind Speed Interval Prediction With Multi-Objective Ensemble Learning," *IEEE Transactions on Emerging Topics in Computational Intelligence* 9, no. 1 (2025): 1–12.

42. J. Wang, S. Wang, B. Zeng, and H. Lu, "A Novel Ensemble Probabilistic Forecasting System for Uncertainty in Wind Speed," *Applied Energy* 313 (2022): 118776.

43. S. Afrasiabi, S. Allahmoradi, and X. Liang, "Pseudo-Measurement Models in Distribution Networks: A Review," *IET Smart Energy Systems* 1, no. 1 (2025): 56–72.

44. A. Salehi, M. Fotuhi-Firuzabad, S. Fattaheian-Dehkordi, M. Gholami, and M. Lehtonen, "Developing an Optimal Framework for PMU Placement Based on Active Distribution System State Estimation Considering Cost-Worth Analysis," *IEEE Access* 11 (2023): 12088–12099.

45. T. C. Xygkis, J. Löfberg, and G. N. Korres, "Investigation of Optimal Phasor Measurement Selection for Distribution System State Estimation Under Various Uncertainties," *IEEE Transactions on Instrumentation and Measurement* 74 (2025): 1–16.

46. D. Chrysostomou, "Data Underlying the PhD Thesis: Learning-Based Distributed Resource Flexibility Estimation for Enhanced TSO-DSO Coordination," 4TU.ResearchData, published 2026, https://doi.org/10.4121/4fdcbda5-e7bf-4b20-bd3c-5f99dce49679.v1.

Appendix A: BNN Comparative Analysis

The comparative analysis included one training, one validation and three test sets for each of the four networks in Figure 4. All datasets followed residential load, PV and wind turbine (where available) hourly profiles and assumed a 40% increase in generation capacity and 20% in load consumption as [8]. For each hour, the datasets sampled different operating conditions for the load consumption, PV and wind turbine generations. The CIGRE dataset also included topological changes between all switch combinations in Figure 4, and three binary features on the status of each switch. The test sets varied in measurement and pseudo-measurement noises (and correlations in CIGRE). The CIGRE network also included an RTI-measured bus (generation with capacity > 1 MVA) with noise double from the respective measurement noise. Table A.1 includes the information for each dataset, σ_v^m is the measurement voltage standard deviation. The BNN model settings were:

1. VI: A variational inference model with four dense layers, one dense variational layer with multivariate normal posterior distribution, with Tanh(·) activation function and $Z = 100$.
2. MCD: A Monte Carlo dropout model with four dense layers, 0.1 dropout, with ReLU(·) activation function and $Z = 100$.
3. DE: A deep ensembles model with 20 FNNs ($Z = 20$) of three dense layers, with ReLU(·) activation function.

All models were trained with a 0.0001 learning rate. VI and MCD were trained for 500 epochs and each DE FNN was trained for 250 epochs. MCD and DE had 512 hidden units for all layers. For VI, the hidden units were 256,128,128,32 for the dense, 32 for the variational layer in CIGRE, and 256,128,128,64 for the dense, and 64 for the variational layer in OB0, OB1 and OB.

In all models, alternative activation functions (ReLU, Tanh), the number of dense layers, and the neurons of each layer were tuned. For VI, a larger imbalance between deterministic and variational layer neurons negatively affected training. Increasing the size of deterministic layers reduced the relative contribution of the variational posterior, while enlarging the variational layer increased the computational overhead, leading to slower convergence. For MCD, the identified number of layers and neurons showcased the best performance, while maintaining fast training. The dropout level for MCD was selected as it balanced predictive performance and uncertainty estimation. For the DE model, the size of the networks was reduced as larger networks did not showcase a clear predictive performance improvement, but increased the training duration.

TABLE A.1 | Dataset settings.

Name	Usage	Net.	Size —	ψ^{ld} %	ψ^{pv} %	$\sigma_\eta^m, \sigma_l^m$ %	σ_v^m %	σ^{pm} %
Tr.C	train	CIGRE	33k	60	70	2	1	20
V.C	validate	CIGRE	7k	60	70	2	1	20
Te1.C	test	CIGRE	7k	60	70	1	0.5	10
Te2.C	test	CIGRE	7k	60	70	2	1	20
Te3.C	test	CIGRE	7k	60	70	5	3	50
Te4.C	test	CIGRE	7k	30	35	1	0.5	10
Te5.C	test	CIGRE	7k	30	35	2	1	20
Te6.C	test	CIGRE	7k	30	35	5	3	50
Tr.OB0	train	OB0	16k	60	70	2	1	20
V.OB0	validate	OB0	3k	60	70	2	1	20
Te1.OB0	test	OB0	3k	60	70	1	0.5	10
Te2.OB0	test	OB0	3k	60	70	2	1	20
Te3.OB0	test	OB0	3k	60	70	5	3	50
Tr.OB1	train	OB1	15k	60	70	2	1	20
V.OB1	validate	OB1	3k	60	70	2	1	20
Te1.OB1	test	OB1	3k	60	70	1	0.5	10
Te2.OB1	test	OB1	3k	60	70	2	1	20
Te3.OB1	test	OB1	3k	60	70	5	3	50
Tr.OB	train	OB	15k	60	70	2	1	20
V.OB	validate	OB	3k	60	70	2	1	20
Te1.OB	test	OB	3k	60	70	1	0.5	10
Te2.OB	test	OB	3k	60	70	2	1	20
Te3.OB	test	OB	3k	60	70	5	3	50

For each network, the models were trained once in the corresponding 'Tr' dataset of Table A.1. Each test set was used to evaluate the models trained on the same network. The validation sets were used to store the model parameters with the best performance during training for each network. Figure A.1 illustrates the model performances for each metric under the varying noise levels and network sizes.

Considering $PICP_{95}$, all models were slightly underconfident (0.95–1) in the low and medium noise scenarios, excluding the MCD which was overconfident in the low noise v for OB1 (0.89) and OB (0.85), and slightly overconfident for OB1 (0.92). In all models, a lower correlation in the test sets reduced $PICP_{95}$ but maintained its levels between 0.91–1 for the low and medium noise. All models were impacted by the high noise scenarios, with MCD maintaining the best performance in most. Considering the voltage RMSE metric, all models showed fluctuations for the different test sets. MCD showed deterioration for lower and higher noise in the larger networks, which can indicate fitting to the noise uncertainty. However, MCD also showed a consistently better RMSE than VI and DE for higher

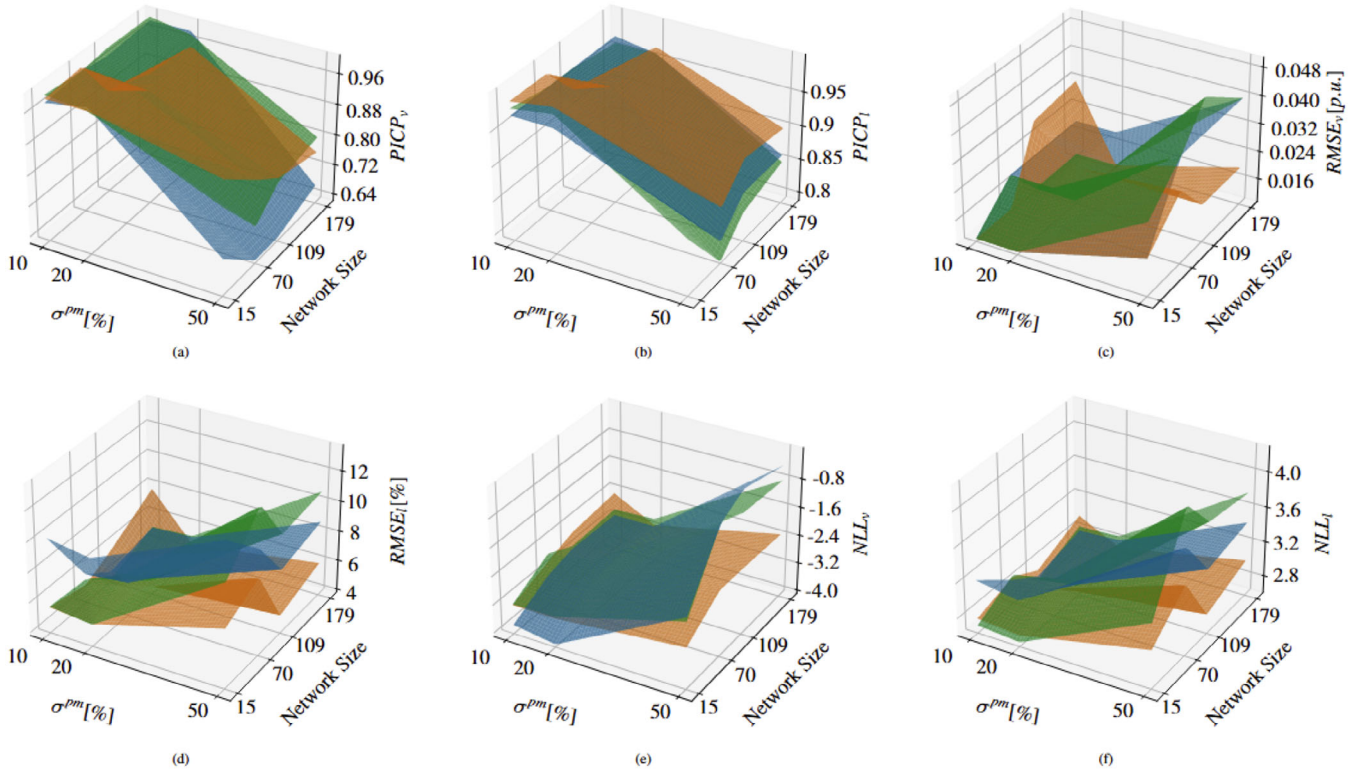


FIGURE A.1 | Comparison of VI (▬), MCD (▬) and DE (▬) models under varying pseudo-measurement (and measurement) noise and network size. The network size corresponds to the number of buses.

TABLE A.2 | Model computational aspects.

Net.	Training duration [h]			Prediction duration [s]		
	VI	MCD	DE	VI	MCD	DE
Cigre	1.8	1	8.1	0.28	0.09	0.01
OB0	10.6	0.8	6.2	0.4	0.12	0.04
OB1	16.6	0.7	7.7	0.36	0.16	0.04
OB	14.7	1.2	9.8	0.46	0.3	0.03

noise levels. Considering the loading RMSE, the VI was highly impacted by the topological changes within the CIGRE datasets, whereas MCD and DE maintained a similar RMSE for all topologies and networks. The MCD mainly outperformed the DE and VI in loading RMSE and was less impacted by the test set noise variations. The NLL metric showed similar levels for all low and medium noise scenarios in all networks and models. However, MCD maintained almost consistent performance in all scenarios with minor noise impacts compared to VI and DE. The VI model was negatively impacted by the different DS topologies in the training dataset and reduced PV and load correlations in the test sets. The MCD and DE did not show significant impacts from the DS topologies. The reduced test set PV and load correlations did not significantly impact the MCD and DE, except for a slight improvement for the l $PICP_{95}$ and a slight deterioration for the l NLL. The results indicate that the MCD model is the most consistent among the three BNNs, especially with increased noise levels. Nevertheless, all BNN models showed low RMSEs, NLL and close to 0.95 $PICP_{95}$.

Table A.2 shows the training and prediction durations per model and network using the Google Colab CPU. The number of input features, neurons and outputs highly impacted the VI training time, with 6–9 times larger durations for OB0, OB1, OB than CIGRE. MCD required significantly less time than DE, as MCD only trained one BNN, whereas DE trained 20 BNNs. The prediction durations were small for all networks, with DE being the fastest. The difference between MCD and DE prediction duration was potentially caused by the larger Z and the MCD having more layers (and parameters).