Developments in the Built Environment 26 (2026) 100931



Contents lists available at ScienceDirect

Developments in the Built Environment

journal homepage: www.sciencedirect.com/journal/developments-in-the-built-environment





# Data-driven federated hierarchical Bayesian neural network framework for region-specific prediction of chloride content in bridge decks

Chi-Ho Jeon [a], Gitae Roh [b,c,*], Chang-Su Shim [b], In Ho Cho [c], Ki-Yeol Kim [b], Ji-hun Shin [b], Ki-Tae Park [a]

[a] Department of Structural Engineering Research, Korea Institute of Civil Engineering and Building Technology, Goyang, Republic of Korea
[b] Department of Civil Engineering, Chung-Ang University, Seoul, Republic of Korea
[c] Dept. Civil, Construction and Environmental Engineering, Iowa State University, Iowa, USA

ARTICLE INFO

*Keywords:*
Federated learning
Hierarchical Bayesian neural network
Chloride content
Bridge deck
Data driven
Region heterogeneity

ABSTRACT

Chloride ingress induces reinforcement corrosion and reduces bridge durability. Deicing-salt application during winter further accelerates deterioration, making reliable long-term prediction essential for maintenance planning. However, two limitations hinder data-driven approaches: chloride data are primarily obtained via destructive testing, yielding scarce long-term observations, and security constraints restrict predictive modeling across agencies. To address these limitations, we propose a federated hierarchical Bayesian neural network framework, in which a federated scheme, aligned with the small number of agencies managing bridges, allows each district to share only BNN–ANN model parameters rather than raw data, while a hierarchical structure separates region-specific patterns to capture inter-regional heterogeneity. A global Bayesian neural network captures deck-level aleatoric and epistemic uncertainties and common temporal features, while region-specific exponential trend parameters are retained locally. The proposed federated framework showed robust predictive performance, and prediction intervals narrowed in data-scarce service-year ranges, confirming effective knowledge transfer to data-poor regions.

## 1. Introduction

Chloride ingress is a critical factor in the deterioration of reinforced concrete bridge structures, particularly bridge decks that are directly exposed to environmental conditions. Among the various deteriorating agents, chloride is widely recognized as a dominant cause of reinforcement corrosion and service life reduction (Cady and Weyers, 1983; Kassir and Ghosn, 2002). Chloride-induced corrosion causes cracking, delamination, and spalling, which degrade structural serviceability and load-carrying capacity. The deterioration process is generally categorized into corrosion initiation and propagation stages, forming a theoretical basis for long-term durability assessment (Tuutti, 1982). Accordingly, durability design provisions incorporate environmental exposure conditions such as coastal locations and classify exposure grades to determine appropriate concrete strength and cover thickness (CEN, 2013; AASHTO, 2020; Ministry of Land and Infrastructure and Transport (MOLIT), 2021). However, the effects of incorporating durability design are inhibited as the increased application of deicing salt during winter accelerates chloride accumulation in bridge decks, intensifying the need for a reliable remaining-service-life prediction.

Mechanism-based probabilistic approaches have been proposed to evaluate chloride-related deterioration and structural reliability. Stewart and Rosowsky (1998) highlighted the importance of reliability-based service life assessment under uncertainty, whereas Bertolini et al. (2013) systematically reviewed chloride transport and corrosion mechanisms. Song et al. (2025) proposed a deterioration grading framework for bridge decks by discretizing the corrosion progression into defined phases and inspection stages, thereby facilitating structured data management.

Recently, data-driven approaches have been introduced to model various aspects of chloride-related deterioration. At the material level, Liu et al. (2021) developed an artificial neural network (ANN)-based predictive model for the chloride diffusion coefficient of concrete using a database of 653 test results encompassing 13 influential parameters including water-to-binder ratio, curing age, and supplementary cementitious material content, demonstrating that neural networks can

* Corresponding author. Department of Structural Engineering Research, Korea Institute of Civil Engineering and Building Technology, Goyang, Republic of Korea.
*E-mail address:* shrlxo2001@cau.ac.kr (G. Roh).

https://doi.org/10.1016/j.dibe.2026.100931
Received 23 February 2026; Received in revised form 22 April 2026; Accepted 23 April 2026
Available online 30 April 2026
2666-1659/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

effectively capture the complex nonlinear relationships governing chloride transport properties. Ashraf et al. (2022) investigated the synergistic effects of bentonite clay and silica fume on the mechanical and durability performance of sustainable concrete, employing machine learning techniques to predict durability-related properties under varying supplementary cementitious material compositions. Golafshani et al. (2023) proposed a machine learning approach for predicting the chloride migration coefficient using a comprehensive dataset of over 1000 experimental results, in which XGBoost and support vector machine algorithms exhibited strong predictive accuracy. These studies primarily target material-level transport properties derived from laboratory specimens under controlled conditions, and have demonstrated that machine learning can serve as an effective tool for characterizing the chloride resistance of concrete.

At the structural level, data-driven approaches have also been applied to predict chloride behavior under field exposure conditions. Roh et al. (2025) proposed an inspection data-driven deterioration-prediction model for bridge decks that emphasized the importance of chloride-related features in remaining service life estimation. Wu et al. (2023) applied deep learning to predict the time-dependent chloride penetration depth in concrete exposed to coastal environments. Abdellatief et al. (2025) developed a convolutional neural network (CNN) framework trained on 284 samples with 11 material and environmental features to predict the surface chloride concentration in marine tidal zones. Ahmad et al. (2021) applied novel machine learning techniques to predict surface chloride concentration in concrete containing waste materials. Tran et al. (Tran and Nguyen, 2023) proposed a physics-informed deep neural network that integrates Fick's diffusion law into the loss function, enabling the simulation of chloride transport behavior while reducing dependence on extensive training data. Although these studies have demonstrated remarkable predictive capabilities, they share a common prerequisite: a large-scale centralized dataset must be available for model training. This centralized paradigm poses significant limitations when predicting the cumulative chloride content in bridge decks as a function of service years, because such field inspection data are inherently distributed across multiple management agencies and regions.

In practice, the construction of such centralized datasets is severely constrained by both data acquisition limitations and institutional barriers. Conventional centralized machine learning approaches require raw data generated by multiple agencies or sites to be aggregated on a single server, introducing challenges related to data ownership, security, and institutional constraints (Sculley et al., 2015; Konečný et al., 2016; Zhao et al., 2018). Moreover, the chloride content of bridge decks is primarily obtained via destructive core sampling, which limits spatial and temporal coverage. Nondestructive techniques such as half-cell potential and electrical resistance measurements only provide indirect indicators rather than direct chloride quantification (Elsener, 2001; Polder et al., 2000). Consequently, long-term time-series chloride data remain scarce, and the systematic construction of large-scale training datasets is often restricted by data-sharing policies (Wang et al., 2023). Alhakeem (2020) developed deterioration prediction models for concrete bridge decks using machine learning techniques based on the U.S. National Bridge Inventory database, demonstrating that data-driven bridge deck condition assessment is effective but remains constrained by data availability and distributional heterogeneity across regions. These constraints underscore the need for alternative learning paradigms that can leverage distributed inspection data without requiring centralized data aggregation.

In addition to the data-sharing constraints, deterioration prediction is inherently accompanied by substantial uncertainty arising from measurement noise, environmental variability, and limited data availability. To explicitly quantify such uncertainty, probabilistic learning frameworks based on Bayesian neural networks (BNNs) have recently attracted considerable attention. Kamel et al. (2025) proposed a framework combining generative adversarial networks for data augmentation with BNNs to predict chloride and sulfate ingress into concrete, achieving reliable uncertainty quantification with $R^2$ values of 0.95. Lei et al. (2026) introduced an active learning–enhanced BNN framework for predicting the life-cycle performance and reliability of prestressed concrete beams in coastal environments, demonstrating the effectiveness of Bayesian approaches in jointly quantifying aleatoric and epistemic uncertainties. Wang et al. (2025) developed an improved probabilistic model for pavement deterioration prediction by coupling a BNN with a cuckoo search algorithm, confirming that BNN-based frameworks provide more reliable predictions with reduced uncertainty compared with deterministic alternatives. These studies collectively demonstrate that BNNs are well suited for deterioration forecasting where observational data are inherently noisy and limited. However, they operate within centralized learning paradigms and do not incorporate hierarchical structures to account for the spatial heterogeneity of environmental exposure across administrative regions. Addressing both the distributed data constraints and the need for uncertainty quantification within a single framework remains an open challenge.

Federated learning (FL) provides a distributed training paradigm that enables collaborative model learning without sharing raw data (McMahan et al., 2017; Kairouz and McMahan, 2021; Yang et al., 2019), offering a potential solution to the data-sharing constraints identified above. Recent studies have applied FL to structural health monitoring (SHM) applications, primarily in the context of vibration-based damage detection using acceleration signals. Abdulkarem et al. (2023) proposed a personalized FL framework combined with tensor data fusion for damage detection in bridge structures, achieving 94–97% accuracy on a cable-stayed bridge dataset without centralizing sensor data. Cheema et al. (2025) developed a clustered FL approach for population-based SHM of bridges, leveraging principal angle–based clustering to account for heterogeneous data distributions across structures. Pham et al. (2025) applied a CNN-based federated model for vibration-based damage identification in submerged structure–foundation systems. Yang et al. (2025) introduced a privacy-preserving FL framework for SHM with application to miter gates. In a methodological context, Yurochkin et al. (2019) proposed a Bayesian nonparametric approach for federated aggregation of neural network weights; however, their framework was designed for general classification tasks and has not been extended to probabilistic deterioration prediction under hierarchical spatial structures. Overall, existing FL applications in SHM have predominantly focused on short-term, sensor-driven structural damage detection. The application of FL to long-term chemical deterioration prediction—such as the cumulative chloride ingress in bridge decks over decades of service—combined with Bayesian uncertainty quantification under hierarchical spatial structures remains largely unexplored. This gap is particularly significant given that chloride-related inspection data are not consistently mandated, and chloride concentration is strongly influenced by regional factors such as coastal proximity and deicing-salt application. Bridges located in environmentally similar regions can therefore be grouped for joint model training, and knowledge derived from data-rich bridges can be transferred to data-scarce bridges, enabling more reliable condition assessments and long-term predictions.

In this study, a federated learning (FL) approach was proposed to address the scarcity of chloride ingress data on individual bridges, together with the security and sharing constraints imposed on such data across managing agencies, aligned with the small number of agencies managing bridges in practice. At the same time, because chloride accumulation is directly governed by region-dependent factors such as coastal proximity and deicing-salt usage, a hierarchical structure was adopted so that training can be carried out at the administrative district level to account for inter-regional heterogeneity (Fig. 1). In addition, to reflect both the aleatoric and epistemic characteristics inherent to chloride content and the unbalanced nature of its collection across a bridge deck, a Bayesian approach was incorporated to explicitly represent predictive uncertainty.

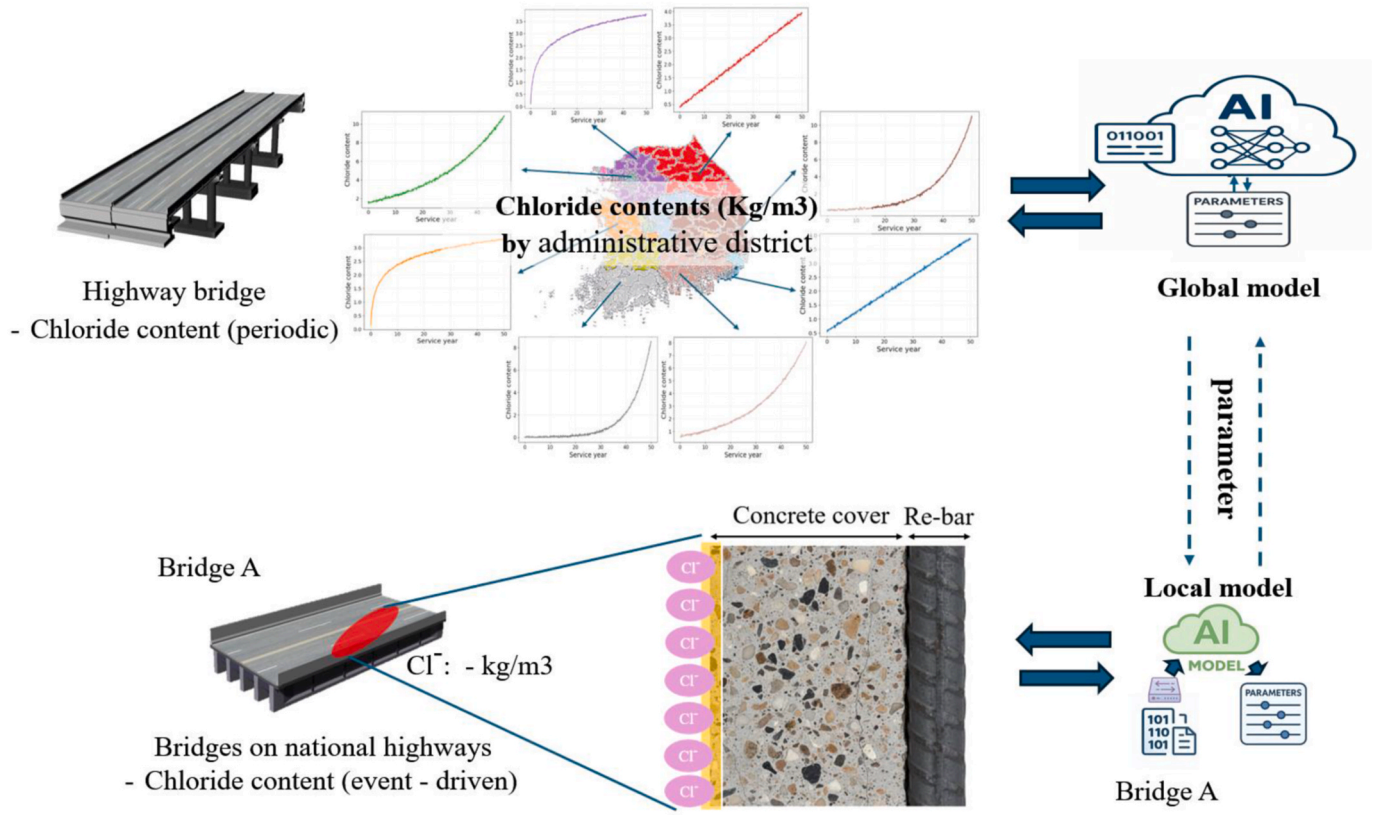



**Fig. 1.** Federated learning (FL) model for the prediction of chloride content on a bridge deck.

Within this framework, each administrative district acts as a client that trains a local artificial neural network (ANN), while a global Bayesian neural network (BNN) serves as the central model, and the two exchange only model parameters rather than raw inspection data. Specifically, the means of the weights and biases of the TimeNet component of the BNN are transmitted to each client as the initial parameters of the local ANN, which is then fine-tuned on its own regional dataset to capture region-specific characteristics. The fine-tuned local parameters are subsequently returned to the central server and approximated as normal distributions consistent with the BNN weight structure. During the integration of these local updates, a weighting scheme is further introduced to control the contribution of each local distribution, thereby preventing contamination of the global model by individual clients and maintaining the probabilistic consistency of the federated BNN.

## 2. Dataset

### 2.1. Overview

Periodic or condition-based inspections are typically conducted to evaluate bridge conditions and the results are archived as bridge condition inspection data. In this study, the chloride-content data were extracted from these reports, in conjunction with the corresponding service years and administrative districts. The dataset was classified into 17 administrative districts and separately organized for the global and local models within the FL framework. The global dataset included chloride data from 1632 bridges, whereas the local dataset contained measurements from 348 bridges. Fig. 2 shows the composition of the dataset based on the administrative districts. All measurements were obtained from the bridge decks without distinguishing between specific deck locations or surfaces, representing the chloride conditions of the overall deck.

### 2.2. Dataset for model learning

#### 2.2.1. Global model dataset

The dataset utilized to train the global model in the FL framework was constructed by randomly selecting bridges from 17 administrative districts, resulting in regional variability in the number of bridges and their service years. The distribution of bridges by district and service years, which were categorized into 5-year bins, is shown in Fig. 3(a). The inspection data within the first 10 years of service are limited, whereas the number of observations increases thereafter, with a noticeable concentration around the 10-year mark in most regions. Fig. 3(b)–(e) show the quantile-based service-year distributions for the four regions with the largest datasets. Chloride dispersion is most pronounced between 10 and 25 years of service and decreased beyond this range.

#### 2.2.2. Local model dataset

The local model dataset was constructed using bridges that do not overlap with those in the dataset for the global model. No data were available for three administrative districts (loc7, loc8, and loc16); therefore, these were excluded from the distribution of bridges by district and service years that is shown in Fig. 4(a). Fig. 4(b)–(e) show the quantile-based chloride distributions according to service years for the four districts with the largest number of observations. As in the global model dataset, loc9 contains the most data, and observations within the first 10 years of service remain limited. The chloride dispersion is most pronounced between 10 and 25 years of service, exhibiting a pattern similar to that in the global model dataset. Overall, the local dataset was organized to maintain a comparable distributional structure and ensure consistency within the FL framework.

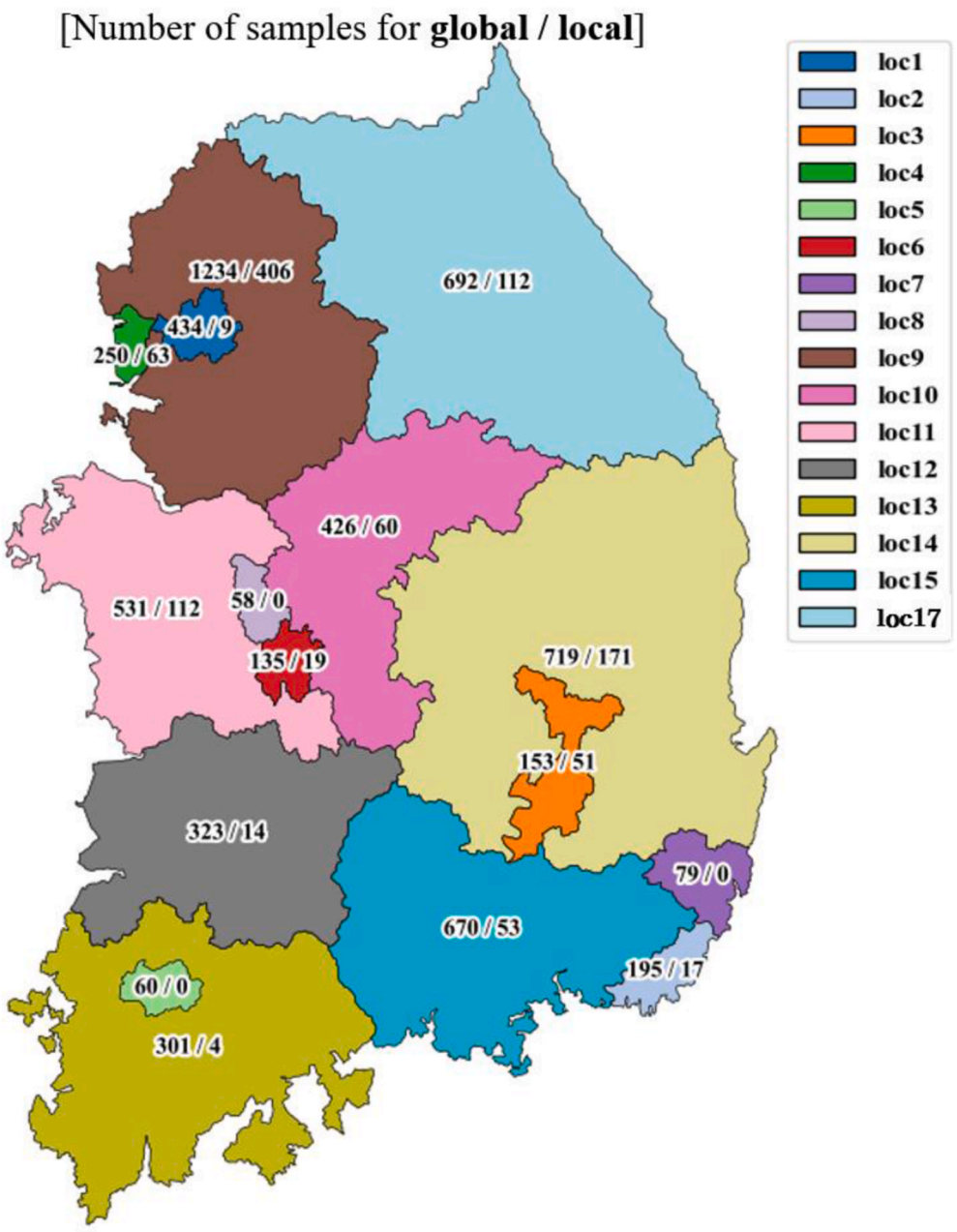



**Fig. 2.** Dataset compositions for global and local models for different administrative regions.

### 2.3. Data preprocessing and quantile-based reconstruction of chloride content

Stepwise data preprocessing and quantile-based distribution reconstruction procedures were applied to transform the chloride-content data, which were unevenly distributed across administrative districts and individual bridges, into a format suitable for the FL-based prediction framework. The procedure was designed to reflect the physical characteristics of chloride ingress and practical maintenance considerations. As illustrated in Fig. 5, the overall data processing workflow comprises five stages. The detailed methodology for each step is described in Sections 2.3.1–2.3.5.

#### 2.3.1. Data sanitization and region-wise service-year filtering

First, data-type standardization and removal of missing values were performed on the raw data to ensure consistency and data integrity. The raw data are defined as follows:

$$\mathscr{D}_{\text{raw}} = \{(r_i, b_i, t_i, y_i)\}_{i=1}^{N}, \tag{1}$$

where $r$, $b$, $t$, and $y$ denote the administrative district, bridge identifier, number of service years, and chloride content, respectively, and $N$ is the total number of samples.

Observations containing missing values for any of the aforementioned four variables were excluded from the analysis. Subsequently, to reflect the regional bridge management histories and distributional characteristics, an upper bound of service years, $T_r^{\max}$, was defined for each administrative district $r$. Observations exceeding this threshold were removed. This step was implemented to prevent a small number of unrealistically long service-year records within certain districts from distorting the overall chloride distribution. Bridges with service years exceeding 40–50 years were subjected to this filtering criterion. The resulting data structure is expressed as follows:

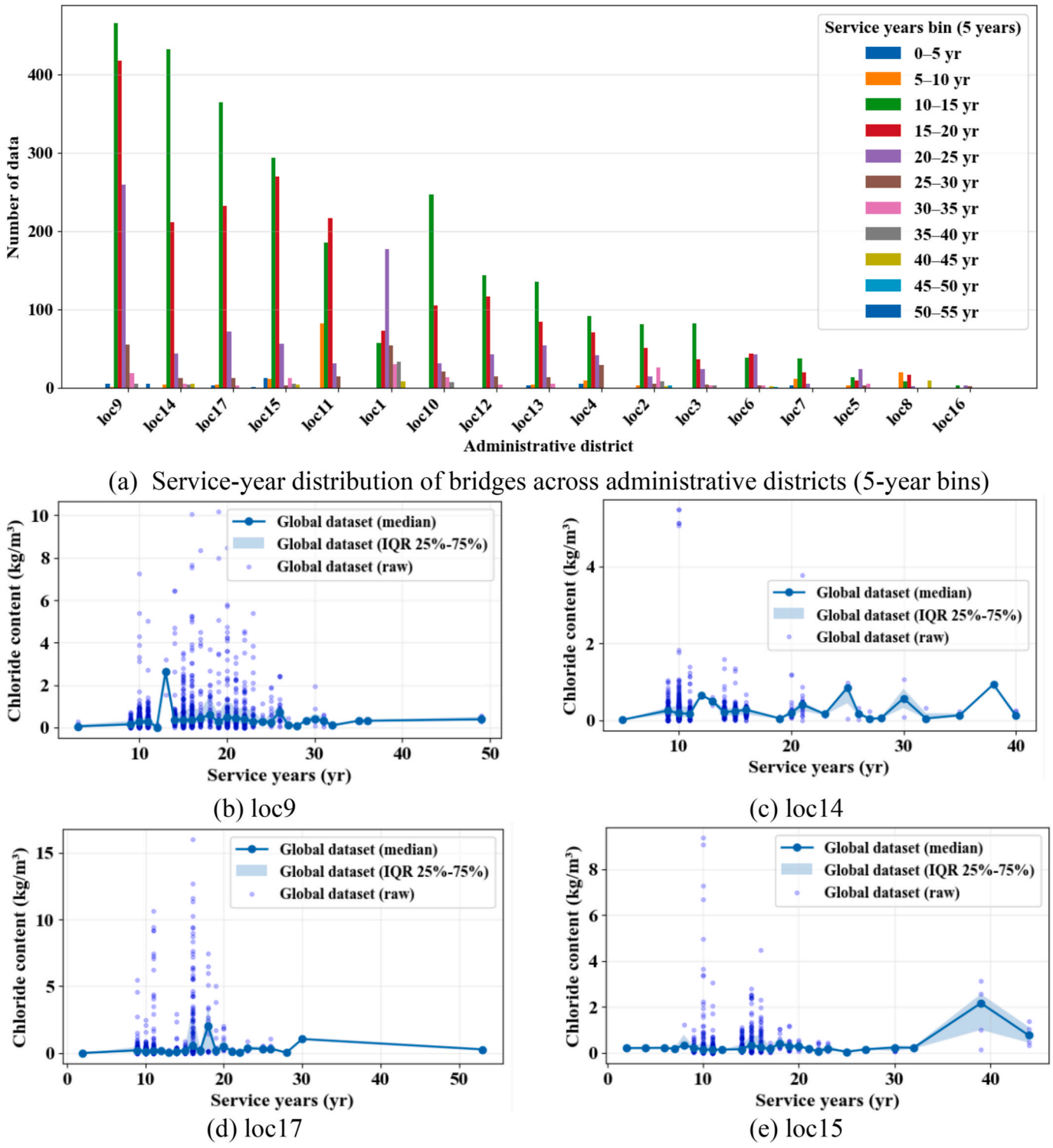



**Fig. 3.** Dataset distribution for the global model based on region and service years.

$$\mathscr{D}_1 = \{(r,b,t,y) \in \mathscr{D}_{\text{raw}} \mid t \le T_r^{\max}\}. \quad (2)$$

*2.3.2. Baseline anchoring at initial service year*

Significant deterioration is generally not expected within the first 10 years after construction, and periodic inspections typically begin only after this point. This is because drainage systems function properly during this initial period and no visually identifiable damage occurs, which reduces the need for inspection. As a result, chloride observations within the 0–10 year interval are structurally absent from the dataset. If a data-driven model is trained without compensating for this data gap, the interval prior to the first observation must be estimated through extrapolation in the absence of observed values, which can distort the accumulation trend in the early service-year range. Because such early-stage distortion propagates cumulatively to the estimation of chloride trends in subsequent service years, a preprocessing step that prevents this effect in advance is required.

To address this issue, the intrinsic chloride level inherent in hardened concrete immediately after curing was set as the baseline at the

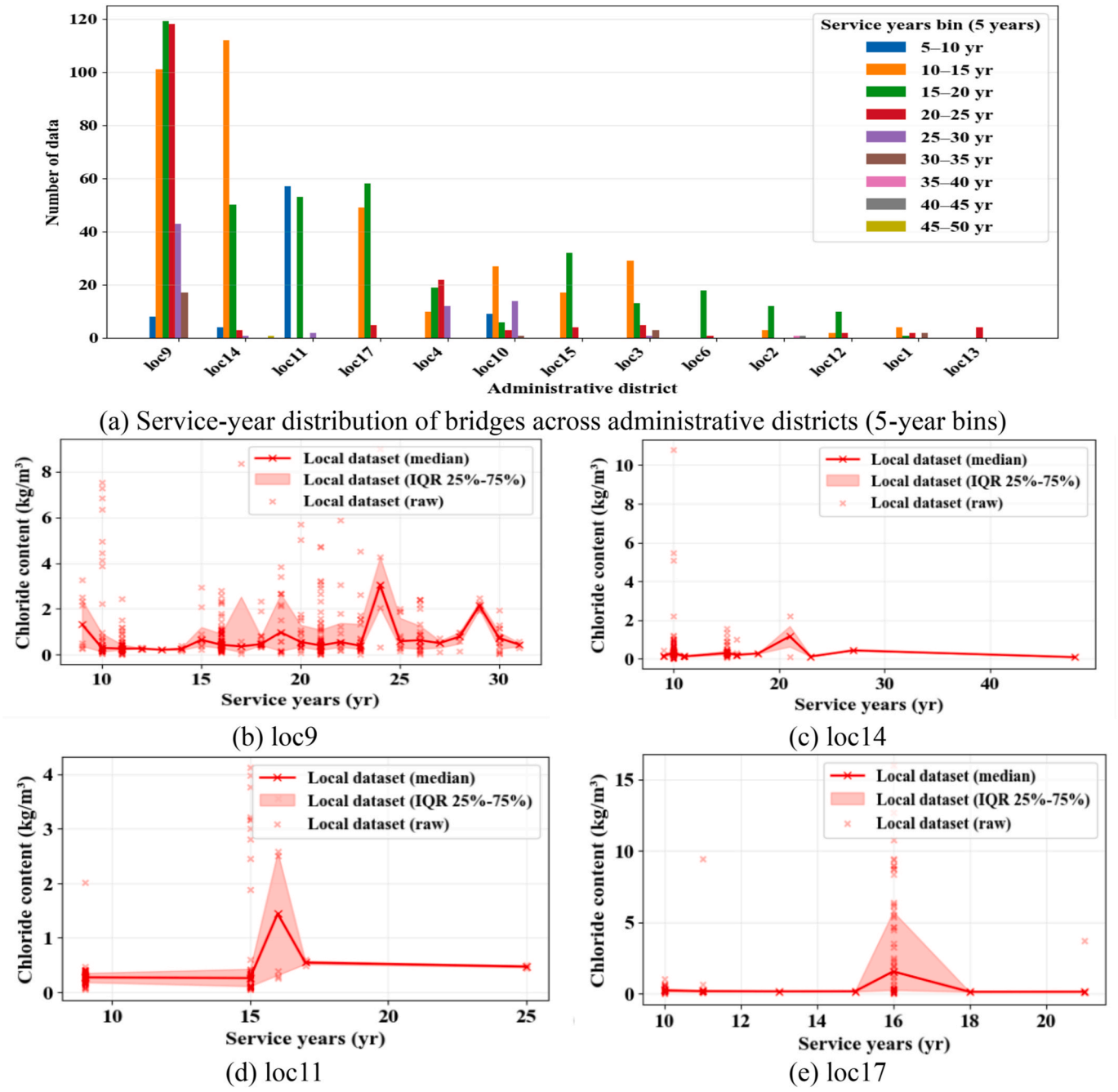



**Fig. 4.** Dataset distribution for the local model based on region and service years.

reference time point. Chloride is a cumulative deterioration factor, and its content during the early service years is physically expected to remain close to zero, with the intrinsic material-level chloride content of hardened concrete typically taking values below 0.25 kg/m$^3$. Reflecting this characteristic, a baseline chloride value (0–0.25 kg/m$^3$) was introduced at the reference time point for each administrative district–bridge combination and was defined as the initial value.

### *2.3.3. Filtering of low-chloride observations based on deck exposure characteristics*

Chloride behavior in bridge decks is influenced by coastal exposure and the use of deicing salts and generally increases as service years progress. However, because chloride data are collected across the entire deck surface, observations are also obtained from locations where drainage functions effectively or where, due to structural characteristics, elevated chloride levels are unlikely to occur. Since the present study aims to address the cumulative behavior of chloride content over time, such low-chloride observations do not represent this cumulative behavior. Moreover, when a substantial number of low-chloride observations are included in the dataset, data-driven training may exhibit a misleading trend in which chloride content appears to decrease or remain unchanged over time, which adversely affects the overall behavior captured in subsequent deterioration assessments.

To remove this bias and correctly capture the cumulative behavior of chloride content, observations below the same threshold (kg/m$^3$) used in the baseline anchoring step were excluded from the dataset. To ensure a conservative evaluation and to capture meaningful deterioration progression, all such observations were removed regardless of whether they corresponded to later service-year time points. This procedure is expressed as follows.

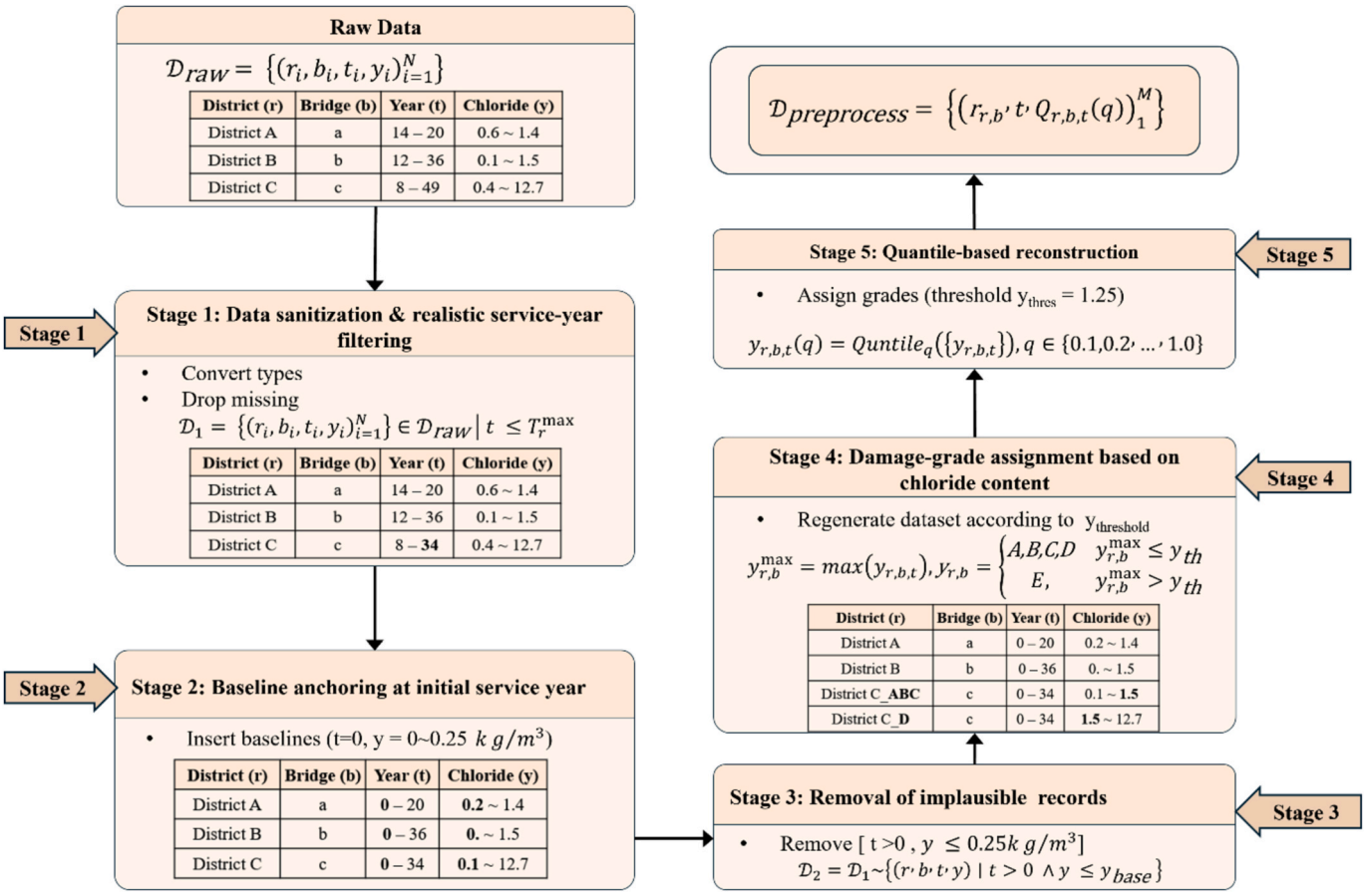



**Fig. 5.** Data preprocessing workflow for model training.

$$\mathscr{D}_2 = \mathscr{D}_1 \setminus \{(r, b, t, y) | t > 0 \wedge y \le y_{\text{base}} (= 0.25\ kg/m^3)\}. \quad (3)$$

### 2.3.4. Damage-grade assignment based on chloride content

Based on the preprocessed dataset, the maximum chloride content was computed for each bridge and classified into damage grades using a threshold of $y_{th} = 1.2$ kg/m$^3$, in accordance with the Korean chloride grading criteria (Ministry of Land and Infrastructure and Transport (MOLIT), 2021), as listed in Table 1. Because chloride ingress is cumulative and can produce extreme values in a small number of cases, a threshold-based classification was applied to control outliers and stabilize the learning process.

The classified damage grade was subsequently redefined in the chloride distribution reconstruction stage by combining it with a corresponding administrative district identifier. The data preprocessing procedure is expressed as follows.

**Table 1**
Bridge rating criteria based on chloride content (Ministry of Land and Infrastructure and Transport (MOLIT), 2021).

| Grade | Condition | Description |
|---|---|---|
| A | Excellent | Very low risk of steel reinforcement corrosion. ($Cl^- < 0.15\ kg/m^3$) |
| B | Good | Minor chloride ingress with negligible impact on durability. ($0.15\ kg/m^3 < Cl^- < 0.3\ kg/m^3$) |
| C | Moderate | Potential for chloride-induced deterioration. ($0.3\ kg/m^3 < Cl^- < 0.6\ kg/m^3$) |
| D | Poor | High chloride concentration with elevated risk of reinforcement corrosion. ($0.6\ kg/m^3 < Cl^- < 1.2\ kg/m^3$) |
| E | Severe | Posing a significant risk of severe loss of structural capacity. ($Cl^- > 1.2\ kg/m^3$) |

$$y_{r,b}^{max} = \max_t \left(y_{r,b,t}\right) \quad (4)$$

$$y_{r,b} = \begin{cases} \text{A, B, C, D} & y_{r,b}^{max} \le y_{\text{th}} \\ E, & y_{r,b}^{max} > y_{\text{th}} \end{cases} \quad (5)$$

### 2.3.5. Quantile-based reconstruction of chloride distributions

In the actually collected dataset, the number of observations varies considerably across bridges and service years. If such data are used in their raw form, bridges with a larger number of observations occupy a disproportionate share of the overall dataset, undermining the equity of contributions across bridges. In addition, chloride content exhibits substantial variability even under similar conditions due to differences in sampling locations and measurement times, and therefore the use of individual observations as they are does not adequately reflect such distributional characteristics. To address these issues, a quantile-based distributional representation was adopted in place of individual observations. For each combination of administrative district, bridge, and service year, the chloride content quantiles are defined as shown in Eq. (6).

$$Q_{r,b,t}(q) = \text{Quantile}_q\left(\left\{y_{r,b,t}\right\}\right), q \in \{0.1, 0.2, \ldots, 1.0\}. \quad (6)$$

By reconstructing each bridge with the same number of quantile values (10), the design ensures that all bridges contribute equally to the dataset regardless of their individual observation density. When a bridge is summarized using a single statistic such as the mean or the maximum, characteristics of the chloride content distribution—such as its range, central tendency, and asymmetry—are lost, although such distributional information is essential for the subsequent uncertainty quantification

stage. The quantile-based reconstruction is an approach that simultaneously achieves equity in bridge-level summarization and preservation of distributional information. Through this procedure, the chloride content data are reconstructed into a quantile-based dataset, as expressed in Eq. (7).

$$\mathscr{D}_{\text{quant}} = \left\{ \left( r_{r,b}, t, Q_{r,b,t}(q) \right) \right\} \tag{7}$$

This quantile-based representation mitigates the information loss associated with mean-based analysis and preserves the distributional range and asymmetry of chloride-induced deterioration. When integrated with the subsequently described H-BNN structure, it enables joint quantification of regional environmental variability and uncertainty resulting from data scarcity.

The structure of the final processed dataset is expressed as shown in Eq. (8). The final dataset comprises the administrative district identifier, service years, and quantile-reconstructed chloride-content values.

$$\mathscr{D}_{\text{preprocess}} = \left\{ \left( r_{r,b}, t, Q_{r,b,t}(q) \right) \right\}_1^M \tag{8}$$

where $M$ denotes the total number of samples after preprocessing.

## 3. Methodology

### 3.1. Federated learning model framework

Before defining the assessment and prediction model for chloride data, it is necessary to first analyze the characteristics inherent in the data. Chloride content involves numerical uncertainty in the data itself within a bridge, while at the same time exhibiting consistent patterns and behaviors driven by regional factors such as snowfall, precipitation, and proximity to the coast. Reflecting these characteristics, a Hierarchical-FL-BNN structure is proposed, which integrates a hierarchical architecture for learning region-specific patterns with a Bayesian approach for quantifying data uncertainty. The overall architecture of the proposed model is shown in Fig. 6.

Within the framework, each administrative district is defined as a client, and the chloride data collected from each district are preserved within the client to reflect the characteristics of the bridge maintenance environment, in which data exchange is restricted between different managing agencies. The training procedure consists of three steps. First, a hierarchical BNN is trained as the global model using the data from all clients. Second, representative parameters extracted from the posterior distribution of the trained BNN are used as the initial values of the local ANN at each client, and fine-tuning is then performed using the data of the corresponding administrative district. Third, the TimeNet weights of the ANNs trained at each client are collected by the server, which aggregates them element-wise and fuses the result with the posterior distribution of the global BNN. In this process, the information shared with the server is limited to the region-agnostic TimeNet weights, while the region-specific functional trend parameters of each administrative district are preserved within the client. This separation is intended to strengthen the temporal deterioration progression patterns that are generalizable across regions through collaborative learning, while preserving region-specific deterioration characteristics without information loss caused by averaging.

The federated learning structure is designed to be generalizable to a multi-round federated configuration, and as the first instantiation of this design, this paper presents a single-round instantiation. This is because a sufficient amount of service-year data must be obtained at once in order to meaningfully estimate the weight distribution (mean and standard deviation) of the client-side ANN, and when data accumulating along the time axis are divided into rounds, the statistical reliability of the distribution computed at each round is degraded. Accordingly, in this study, the framework is instantiated in a single-round form by utilizing the available service-year data collectively.

At the server-side fusion stage, two types of Gaussian information are combined. The first is the marginal Gaussian of the TimeNet weights extracted from the posterior distribution of the global BNN, which reflects the epistemic uncertainty formed during the global training stage. The second is the Gaussian computed by aggregating the ANN fine-

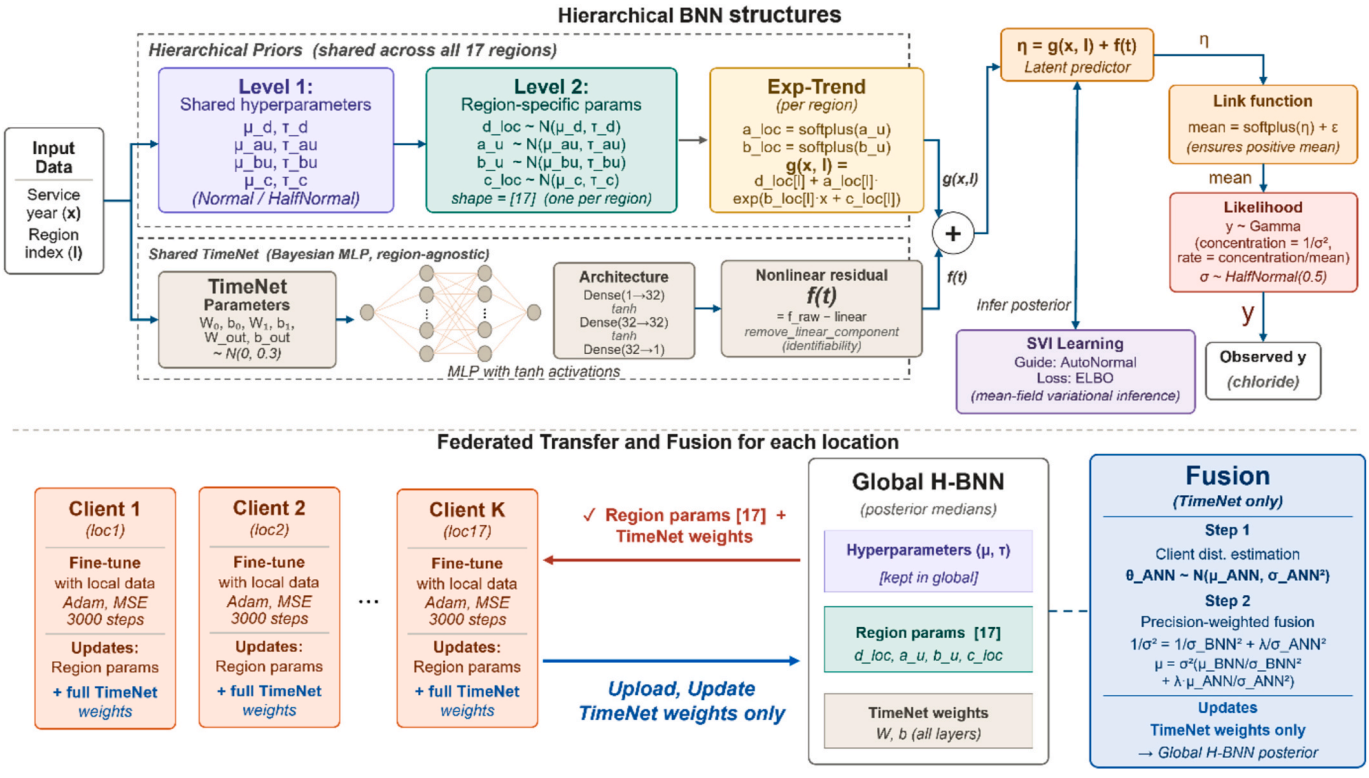



**Fig. 6.** Overall architecture of the proposed hierarchical federated BNN framework.

tuning results from each client in an element-wise manner, which quantifies the region heterogeneity through the spread of weights across administrative districts. The two distributions are combined in a precision-weighted manner based on the principle of conjugate Gaussian update, where the BNN posterior distribution plays the role of a prior and the aggregated ANN result plays the role of auxiliary information (quasi-likelihood) distilled from the client data. The specific formulation and statistical assumptions of this procedure are described in detail in the following section. Table 2 presents the differences from FedAvg, a representative federated learning model.

### 3.2. Hierarchical federated Bayesian neural network (BNN) for the global model

To preserve region-specific characteristics as inherent parameters and model the temporal evolution of chloride content at the administrative district level, an integrated predictive framework combining a hierarchical structure, an exponential time trend, and a probabilistic neural network (TimeNet) is proposed. An exponential function was incorporated into the mean structure to represent the cumulative and monotonically increasing behavior of chloride over time. The framework was further designed to capture the regional heterogeneity and nonlinear residual behavior, enabling a comprehensive representation of chloride deterioration pattern.

#### 3.2.1. Hierarchical exponential trend

Consider an observation $y_i$ representing the chloride content measured at time $t_i$ for location $l$, where all observations take positive real values. For each observation, a latent predictor was defined, which is expressed as

$$\eta_i = d_\ell + a_\ell \exp(b_\ell x_i + c_\ell) + f_\theta(t_i), \tag{9}$$

where $d_l$ denotes the location-specific baseline level, $a_l$, $b_l$, and $c_l$ represent location-specific exponential growth coefficients, $x_i$ corresponds to the service years elapsed since the reference time (with $x_i \geq 0$), and $f_\theta(t_i)$ denotes a nonlinear residual term modeled by a neural network.

To ensure that the exponential trend increases monotonically with time, reparameterization was applied, as expressed in the following equation.

$$a_\ell = \text{softplus}(a_{u,\ell}) + \varepsilon,\ b_\ell = \text{softplus}(b_{u,\ell}) + \varepsilon \tag{10}$$

Through this reparameterization, the constraints $a_l \geq 0$ and $b_l \geq 0$ were strictly enforced, ensuring that chloride content does not decrease over time, thereby reflecting its physically non-decreasing cumulative behavior. To prevent numerical divergence of the exponential term, $b_l x_i + c_l$ was clipped within a predefined range. The location-specific parameters follow hierarchical prior distributions, as expressed subsequently.

$$d_\ell \sim \mathcal{N}(\mu_d, \tau_d), a_{u,\ell} \sim \mathcal{N}(\mu_{au}, \tau_{au}), b_{u,\ell} \sim \mathcal{N}(\mu_{bu}, \tau_{bu}), c_\ell \sim \mathcal{N}(\mu_c, \tau_c) \tag{11}$$

The hyperparameters $\mu_d$ and $\tau_d$ were assigned weakly informative prior distributions. This hierarchical structure facilitates systematic information sharing across locations, enabling data-scarce sites to leverage information from the overall dataset, thereby yielding a more stable parameter estimation.

#### 3.2.2. TimeNet-based BNN term

The nonlinear residual term $f_\theta(t_i)$ was modeled using a multilayer perceptron (TimeNet) architecture. TimeNet comprises two hidden layers with tanh activation functions. The network weights $\theta$ are modeled as random variables and assumed to follow a Gaussian prior distribution.

$$\theta \sim \mathcal{N}(0, \sigma_w^2) \tag{12}$$

The constant and first-order linear terms with respect to time were removed from the TimeNet output to mitigate the identifiability issue between the exponential trend and TimeNet components. Specifically, the best-fitting linear function of time was estimated and subtracted from the TimeNet output, enabling the neural network to learn the nonlinear residual component. This procedure clarifies the functional separation between the exponential trend and neural network components.

#### 3.2.3. Gamma likelihood and mean structure

The chloride content is strictly non-negative and exhibits an asymmetric distribution whose variance may increase over time. Considering the geometric characteristics of bridge decks, cantilever regions occupy a smaller proportion than traffic lanes, and this spatial heterogeneity influences the empirical distribution. Accordingly, observation $y_i$ is assumed to follow a gamma distribution, as expressed in Eq. (13), where $\mu_i$ denotes the conditional mean of $y_i$. To ensure positivity, a softplus transformation was applied to the latent predictor $\eta_i$, and $\mu_i$ was defined as shown in Eq. (14). To control dispersion, a variance parameter $\sigma$ was introduced and formulated as in Eq. (15), such that $\mathbb{E}[y_i] = \mu_i$ under the specified gamma parameterization. Moreover, the observations were truncated such that all values were constrained to be greater than or equal to a small positive constant $\varepsilon$ to satisfy the support condition of the gamma distribution.

$$y_i \sim \text{Gamma}(\alpha_i, \beta_i), \tag{13}$$

$$\mu_i = \text{softplus}(\eta_i) + \varepsilon_\mu \tag{14}$$

$$\alpha_i = \frac{1}{\sigma^2 + \varepsilon},\ \beta_i = \frac{\alpha_i}{\mu_i} \tag{15}$$

#### 3.2.4. Variational inference with mini-batch stochastic variational inference

Because the closed-form posterior distributions for all weights are intractable, variational inference (VI) with an AutoNormal guide was employed. An independent normal approximation was assumed for all latent variables, and the parameters were estimated by maximizing the evidence lower bound (ELBO). To improve computational efficiency for largescale datasets, mini-batch subsampling was applied to the observation likelihood. At each iteration, a subset of observations was sampled to approximate the log-likelihood. The hierarchical parameters and TimeNet weights were treated as global latent variables, and the variational parameters were updated using a stochastic gradient descent

**Table 2**
Comparison between FedAvg and the proposed FL-Hierarichical-BNN framework.

| Aspect | FedAvg | Proposed FL-BNN | Differences |
|---|---|---|---|
| Global model | Deterministic NN | Hierarchical BNN | Captures epistemic and aleatoric uncertainty |
| Information shared | All local weights | TimeNet weights only | Region-specific parameters preserved locally |
| Aggregation | Weighted mean | Element-wise Gaussian (μ, σ) | Cross-client variance retained as region heterogeneity |
| Use of prior knowledge | None | BNN posterior as Bayesian prior | Existing global knowledge integrated, not replaced |
| Fusion rule | Arithmetic averaging | Precision-weighted conjugate update | Bayesian-consistent integration of two information sources |
| Round configuration | Multi-round | Single-round instantiation (Multi-round capable) | Ensures statistical reliability of weight distribution estimation |
| Region heterogeneity | Averaged out | Quantified and preserved | Region-specific deterioration patterns are not lost |

with the Adam optimizer.

#### 3.2.5. Posterior predictive inference

The posterior predictive distributions were obtained by repeated sampling from the posterior region. For each draw, the hierarchical parameters and TimeNet weights were sampled from their posterior distributions and new observations were accordingly generated. The predictive means and PI were computed from these samples. The predictive performance at each location was evaluated using the root mean square error (RMSE), mean absolute error (MAE), and coverage probability of the PI.

### 3.3. Artificial neural network (ANN) for the local model

In the local model stage, the parameters learned from the hierarchical exponential trend BNN were used to construct a deterministic model at the administrative district level, which was subsequently fine-tuned using regional data. The model architecture and size were identical to those of the global model to ensure structural consistency. This procedure comprised three steps: (1) extraction of the BNN parameters, (2) initialization of the ANN, and (3) fine-tuning using the local model dataset. The configuration of the model is shown in Fig. 7.

#### 3.3.1. Extraction of representative parameters from the BNN

After the BNN training was completed, representative values of the posterior distributions for all neuron weights were extracted from the AutoNormal guide. In this study, the posterior median of each weight was used to construct a deterministic parameter set. In Eq. (16), the resulting $\hat{\theta}$ represents the BNN after the probabilistic uncertainty is removed, which can be interpreted as the corresponding deterministic ANN model.

$$\begin{aligned}\hat{\theta} &= \mathrm{median}\left(q_\phi(\theta)\right)\\ \left(\theta = \left\{d_{\swarrow}, a_{u,\swarrow}, b_{u,\swarrow}, c_{\swarrow}, W_{\mathrm{TimeNet}}, b_{\mathrm{TimeNet}}\right\}\right)\end{aligned} \tag{16}$$

#### 3.3.2. Construction of an ANN with exponential trend structure

The extracted parameters were used as the initial weights of an ANN with a similar architecture as that of the BNN, which can be expressed as shown in Eq. (17). The predictive output of the ANN model is defined by Eq. (18).

$$\begin{aligned}\left(d_{\swarrow}, a_{u,\swarrow}, b_{u,\swarrow}, c_{\swarrow}\right)_{\mathrm{ANN}} &\leftarrow \left(d_{\swarrow}, a_{u,\swarrow}, b_{u,\swarrow}, c_{\swarrow}\right)_{\mathrm{BNN}},\\ \left(W_{\mathrm{TimeNet}}, b_{\mathrm{TimeNet}}\right)_{\mathrm{ANN}} &\leftarrow \left(W_{\mathrm{TimeNet}}, b_{\mathrm{TimeNet}}\right)_{\mathrm{BNN}}.\end{aligned} \tag{17}$$

$$\eta_i = d_{\swarrow(i)} + \mathrm{softplus}\left(a_{u,\swarrow(i)}\right)\exp\left(\mathrm{softplus}\left(b_{u,\swarrow(i)}\right)x_i + c_{\swarrow(i)}\right) + f(t_i), \tag{18}$$

where $x_i$ denotes the cumulative time variable measured from the reference year and $f(t_i)$ represents the nonlinear residual function that retains the TimeNet structure learned from the BNN.

Through this initialization strategy, the ANN does not start from random weights, but instead incorporates a long-term exponential trend and temporal patterns that were previously learned from the entire global dataset.

#### 3.3.3. Local fine-tuning of ANN

The ANN was further trained on an individual local dataset corresponding to a given administrative district through additional backpropagation. At this stage, a deterministic objective function based on the mean squared error (MSE) (Eq. (19)) was employed rather than a probabilistic likelihood model.

$$\mathscr{L}_{\mathrm{data}} = \frac{1}{N}\sum_{i=1}^{N}\left(\eta_i - y_i\right)^2 \tag{19}$$

Moreover, to prevent excessive deviations of the parameters from the physically informed trend structure learned by the BNN, a Maximum A Posteriori-type loss function was formulated by incorporating a weak Gaussian prior-based regularization term. Optimization was performed using the Adam optimizer with mini-batch training.

$$\mathscr{L} = \mathscr{L}_{\mathrm{data}} + \lambda\left(\|d_{\swarrow}\|^2 + \|a_{u,\swarrow}\|^2 + \|b_{u,\swarrow}\|^2 + \|c_{\swarrow}\|^2\right) \tag{20}$$

Here, $\lambda$ was set to a sufficiently small prior weight to ensure that the data-fitting term remained dominant in the overall objective.

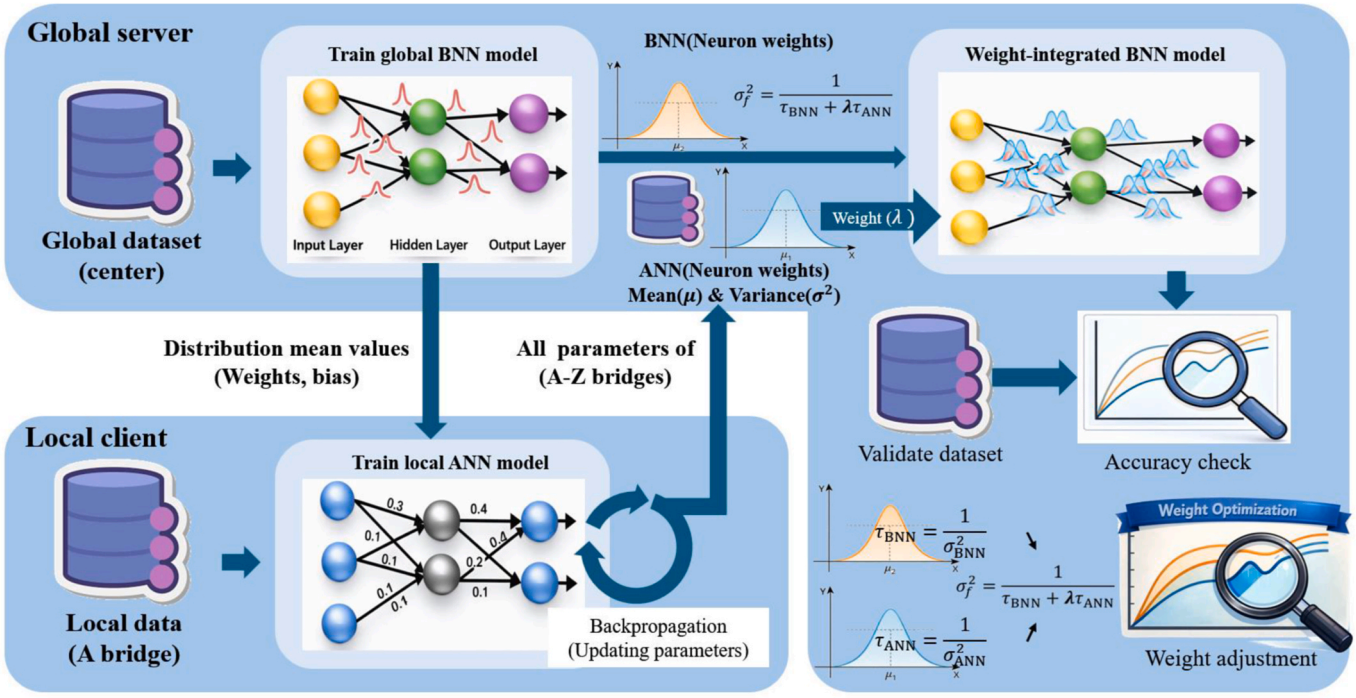



**Fig. 7.** BNN–ANN parameter sharing and fusion scheme in the hierarchical federated BNN framework.

### 3.4. Hybrid parameter-based BNN

#### 3.4.1. Extraction of TimeNet parameter distributions from the ANN

After training the ANN for each administrative district, the learned weights were collected from the local ANN models. Element-wise mean (Eq. (21)) and standard deviation (Eq. (22)) were computed for these weights.

$$\mu_{\mathrm{ANN}} = \frac{1}{K}\sum_{k=1}^{K} w_k \tag{21}$$

$$\sigma_{\mathrm{ANN}} = \sqrt{\frac{1}{K-1}\sum_{k=1}^{K}(w_k - \mu_{\mathrm{ANN}})^2} \tag{22}$$

Here, $K$ denotes the number of available ANN models for the corresponding site.

The parameters of the exponential trend ANN models trained at the administrative district level were stored separately for each region. In this study, only the weights corresponding to the TimeNet component were extracted and aggregated, and their distributions were approximated using a normal distribution.

#### 3.4.2. Fusion of weight distributions and comparative analysis

A precision-weighted Gaussian fusion method was applied to integrate the posterior distribution of the BNN with the weight distribution estimated from the ANN ensemble. Assuming that the two distributions represented distinct information sources of the same weight, a product-of-experts-based Gaussian fusion was performed under a conditional independence assumption. The corresponding parameters were combined as shown in Eqs. (23)–(26), yielding an updated distribution. The initial weighting parameter $\lambda$ was set to 1 to regulate the contribution of the ANN information.

$$\tau_{\mathrm{BNN}} = \frac{1}{\sigma^2_{\mathrm{BNN}}}, \tau_{\mathrm{ANN}} = \frac{1}{\sigma^2_{\mathrm{ANN}}} \tag{23}$$

$$\sigma_f^2 = \frac{1}{\tau_{\mathrm{BNN}} + \lambda\tau_{\mathrm{ANN}}} \tag{24}$$

$$\mu_f = \sigma_f^2(\tau_{\mathrm{BNN}}\mu_{\mathrm{BNN}} + \lambda\tau_{\mathrm{ANN}}\mu_{\mathrm{ANN}}) \tag{25}$$

$$w^{\mathrm{FUSED}} \sim \mathcal{N}\left(\mu_f, \sigma_f^2\right) \tag{26}$$

Here, $\mu_f$ denotes the mean of the fused weight distribution, $\sigma_f^2$ represents the variance of the fused weight distribution, and $\lambda$ is a weighting parameter that controls the relative confidence assigned to the ANN distribution during fusion.

#### 3.4.3. Evaluation strategy

To compare the predictive performance on a per-location basis, the RMSE, MAE, and 90% PI were calculated. This evaluation quantitatively assessed the influence of replacing the TimeNet posterior distribution on the predictive accuracy and uncertainty representation.

## 4. Implementation details and results

All computational procedures in this study were conducted on an NVIDIA A100 GPU, and the JAX (Bradbury et al., 2018) and NumPyro (Phan et al., 2019) frameworks were used for probabilistic inference. To ensure computational efficiency and GPU compatibility, 64-bit precision was disabled and all major tensors were processed using the single-precision floating-point format.

### 4.1. Hierarchical BNN (global model)

#### 4.1.1. Data scaling and training split

The input variable, that is, service years, was converted into a continuous time variable (“year”) and standardized using the mean and standard deviation of the training data. The output variable, that is, chloride content, was modeled directly on the raw scale. The train–test split was performed at the administrative district level. Within each district, the data were divided into 10 quantile bins based on service years, and 20% of the observations in each bin were randomly assigned to the test set. To ensure statistical stability, minimum sample size constraints were imposed: at least five samples per administrative district and two samples per quantile bin.

#### 4.1.2. Model hyperparameter settings

TimeNet is designed to process univariate input series; it has two hidden layers with 32 neurons each, tanh activation functions, and a one-dimensional output. All weights and biases were modeled as random variables with normal prior distributions. During the training, these parameters were sampled once and shared across the entire dataset. Model training was performed using stochastic VI with the AutoNormal guide as the variational distribution. The Adam optimizer was used with a learning rate of $1.0 \times 10^{-4}$. The hyperparameters of the model are listed in Table 3. Trace ELBO, a VI-based objective function, was used as the loss function.

#### 4.1.3. Ablation study on the preprocessing strategy

Before training with the preprocessed dataset, an ablation study was conducted on the loc9 ABCD and loc9 E administrative district–grade combinations to quantitatively verify the individual contribution of each of the three preprocessing steps proposed in Section 2.3. The following four model variants were trained and compared.

- M0 (Full): the baseline model with all three preprocessing steps applied (the proposed model in this paper).
- M1 (No baseline anchoring): the step in Section 2.3.2 is removed; that is, the kg/m$^3$ anchor at the reference time point is not introduced.
- M2 (No low-chloride filtering): the step in Section 2.3.3 is removed; that is, observations with kg/m$^3$ are included in the training data as they are.
- M3 (No quantile reconstruction): the step in Section 2.3.5 is removed; that is, raw observations are used as they are without quantile reconstruction.

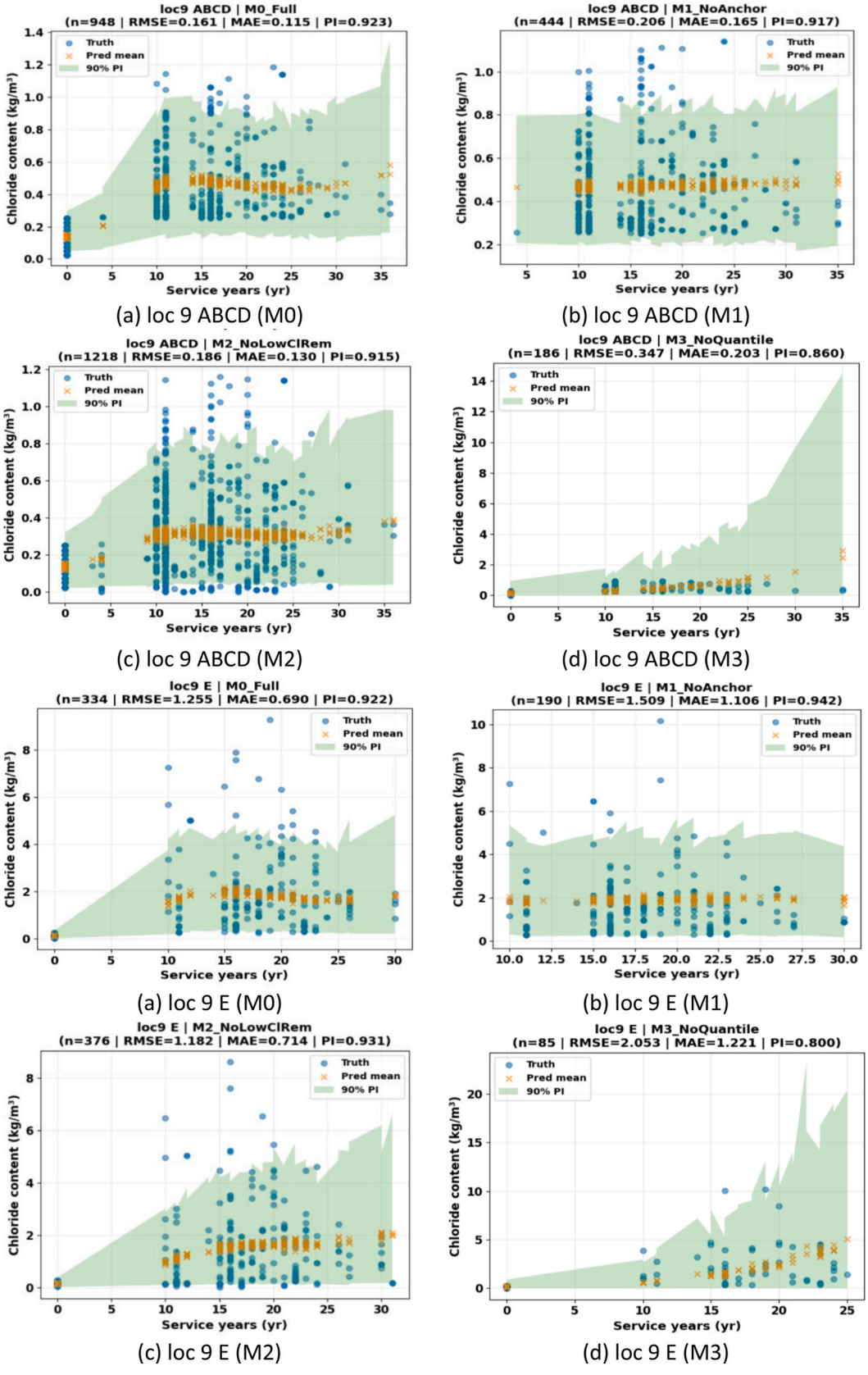
All four models use the same model architecture and hyperparameters. The evaluation metrics are RMSE, MAE, 90% PI coverage, and mean PI width. The ablation study results are summarized in Fig. 8 and Table 4.

Overall, the M0 model showed the best performance in both regions in terms of MAE, PI width, and (for loc9 ABCD) RMSE, while PI coverage also stably achieved the 90% target level in both regions. The effect of each preprocessing step exhibited a consistent tendency across the two regions.

First, M1 showed the most pronounced MAE deterioration in both regions. In loc9 ABCD, RMSE increased by 28% and MAE by 43%, while

**Table 3**
Hyperparameters for the Bayesian neural network.

| Item | Description |
|---|---|
| Optimization algorithm | Adam |
| Learning rate | $1 \times 10^{-4}$ |
| Maximum training steps | 100,000 |
| Mini-batch size | 2048 |
| Loss function | Trace ELBO |




(a) loc 9 ABCD (M0) (b) loc 9 ABCD (M1)

(c) loc 9 ABCD (M2) (d) loc 9 ABCD (M3)

(a) loc 9 E (M0) (b) loc 9 E (M1)

(c) loc 9 E (M2) (d) loc 9 E (M3)

**Fig. 8.** Ablation study for preprocessed dataset.

**Table 4**
Ablation study results for the loc9 ABCD and loc9 E regions.

| Region | Model | $n_{train}$ | $n_{test}$ | RMSE | MAE | 90% PI cov. | PI width |
|---|---|---|---|---|---|---|---|
| loc9 ABCD | M0 (Full) | 3792 | 948 | 0.161 | 0.115 | 0.923 | 0.493 |
| | M1 (No anchoring) | 1776 | 444 | 0.206 | 0.165 | 0.917 | 0.624 |
| | M2 (No low-Cl filtering | 4872 | 1218 | 0.186 | 0.130 | 0.915 | 0.587 |
| | M3 (No quantile) | 752 | 186 | 0.347 | 0.203 | 0.860 | 1.657 |
| loc9 E | M0 (Full) | 1336 | 334 | 1.255 | 0.690 | 0.922 | 2.513 |
| | M1 (No anchoring) | 760 | 190 | 1.509 | 1.106 | 0.942 | 4.704 |
| | M2 (No low-Cl filtering) | 1504 | 376 | 1.182 | 0.714 | 0.931 | 2.807 |
| | M3 (No quantile) | 350 | 85 | 2.053 | 1.221 | 0.800 | 6.972 |

in loc9 E, RMSE increased by 20% and MAE by 60% (0.690 → 1.106). When the anchor is removed, the training data are reduced to approximately half their original size and the information in the 0–10 year interval is completely lost, causing the model to estimate the early interval through extrapolation. The resulting instability in the early parameter estimation is also confirmed by the increase in PI width. In particular, for loc9 E, the fact that the PI width nearly doubled even though the PI coverage was over-covered at 0.942 indicates that the model fails to provide reliable and sharp predictions in the absence of anchoring.

M2, in which low-chloride observations are not removed, showed an overall performance degradation in both regions despite using a larger amount of training data than M0. In loc9 ABCD, RMSE increased by 16% and MAE by 13%. In loc9 E, RMSE decreased by approximately 6%, but MAE increased slightly and the PI width also expanded from 2.513 to 2.807. The partial improvement in RMSE is interpreted as the result of raw observations in the high-chloride region contributing to the point prediction of some high-end samples, but in terms of MAE—which is more robust to outliers—and uncertainty calibration, M0 was consistently superior in both regions. This indicates that low-chloride observations do not represent the cumulative behavior and weaken the overall trend captured by the model by flattening its temporal pattern.

M3, in which the quantile reconstruction is removed, caused the most significant performance degradation in both regions. In loc9 ABCD, RMSE more than doubled, MAE increased by 77%, and the PI width expanded to 3.36 times its original value, while PI coverage dropped to 0.860, falling short of the 90% target. The same pattern appeared even more extremely in loc9 E, where RMSE increased to 1.64 times, MAE by 77%, and the PI width to 2.77 times, and PI coverage dropped to 0.800, failing to meet the nominal confidence level despite the substantial expansion of the PI width. The quantile reconstruction performs two roles simultaneously. First, it improves the stability of the training data by generating 10 quantile values for each (bridge, service-year) combination. In this ablation, the training data of M3 amounted to only 752 samples in loc9 ABCD and 350 samples in loc9 E. Second, it preserves the range, central tendency, and asymmetry of the chloride content distribution that are difficult to encode through raw observation representations. In the absence of such distributional information, the H-BNN fails to properly calibrate the predictive uncertainty, and this effect is particularly pronounced in the high-chloride region (loc9 E) where the distributional variability is large.

#### 4.1.4. Posterior prediction and evaluation

After training with preprocessed dataset, 300 posterior predictive samples were generated to compute the predictive means and the corresponding PIs. Model performance was evaluated separately for the training and test datasets using RMSE, MAE, and the coverage probability of 90% PI. Figs. 9 and 10 show the training and test results, respectively, for the four administrative districts with the largest number of observations.

The model evaluation showed that the average RMSE, MAE, and 90% PI coverage across locations were 0.937, 0.283, and 0.906, respectively, for the training data. For the test data, the RMSE, MAE, and 90% PI coverage were 0.989, 0.294, and 0.897, respectively, indicating only limited performance degradation. These results suggest that, when the service-year distributions of the training and evaluation datasets are similarly structured, the model maintains a stable predictive accuracy and well-calibrated uncertainty.

Analysis across service-year intervals showed that the PI width increased in later service-year ranges, which corresponded to regions with relatively sparse observations and consequently higher estimated variance. In the intervals where the empirical chloride distribution became more dispersed, both the predictive mean and interval width increased concurrently. This pattern reflects the interaction between the data dispersion and exponential trend structure, leading to amplified uncertainty in later service periods.

A comparison of the performance across regions indicated that the predictive mean tended to align more closely with the global trend in data-scarce regions, whereas region-specific temporal patterns were more distinctly captured in data-rich regions. This behavior reflects the partial pooling induced by the hierarchical prior structure, in which region-specific parameters are governed by shared hyperparameters. Consequently, the inter-regional predictive variance was not excessively inflated, and regional characteristics were reasonably preserved.

### 4.2. Local model

#### 4.2.1. Overview of the ANN model

A separate local model dataset was used for fine-tuning of the ANN, and the preprocessing procedure applied to the global model dataset was retained. The input variable (service years) was standardized using the same transformation adopted during the BNN training. The ANN architecture was identical to that of the BNN to ensure structural consistency. In Eq. (20), the regularization weight was set to $\lambda = 10^{-3}$, enabling the data-fitting term to dominate the learning process. ANN fine-tuning was performed using the Adam optimizer with a learning rate of $1.0 \times 10^{-3}$ and a mini-batch size of 256, training for a maximum of 3000 steps with the MSE as the loss function.

#### 4.2.2. Evaluation and analysis

Training was conducted separately for each administrative district. The results for the four regions with the largest number of observations before and after tuning the BNN-initialized weights within the ANN model are illustrated in Fig. 11.

Prior to parameter fine-tuning, the global BNN model stably reproduced the long-term temporal trend of the chloride content with respect to service years, even when evaluated using a local model dataset. However, the limited incorporation of region-specific environmental and operational characteristics resulted in systematic deviations in the absolute prediction levels. This reflects the hierarchical probabilistic structure of the global model, which emphasizes the shared temporal trends and overall uncertainty. Although robust in capturing time-dependent patterns, it has a limited ability to correct region-specific bias.

After region-level local fine-tuning, personalization was achieved based on local observational data while preserving the structural trend learned by the global model. Systematic errors observed within specific service-year intervals were reduced, and the predictive mean shifted toward the observed values. This reflects the complementary interaction

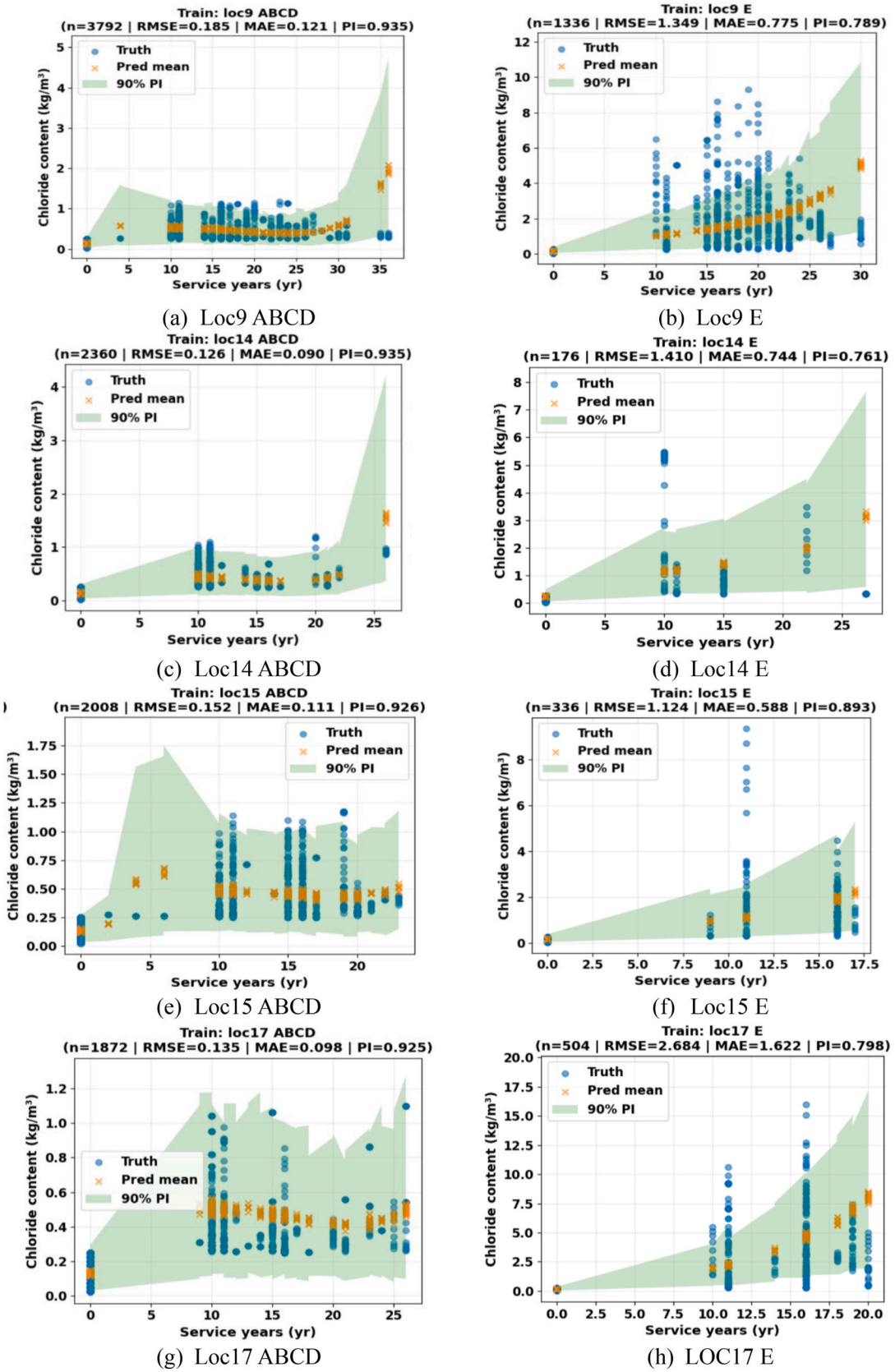



**Fig. 9.** Training data results for the top-four regions.

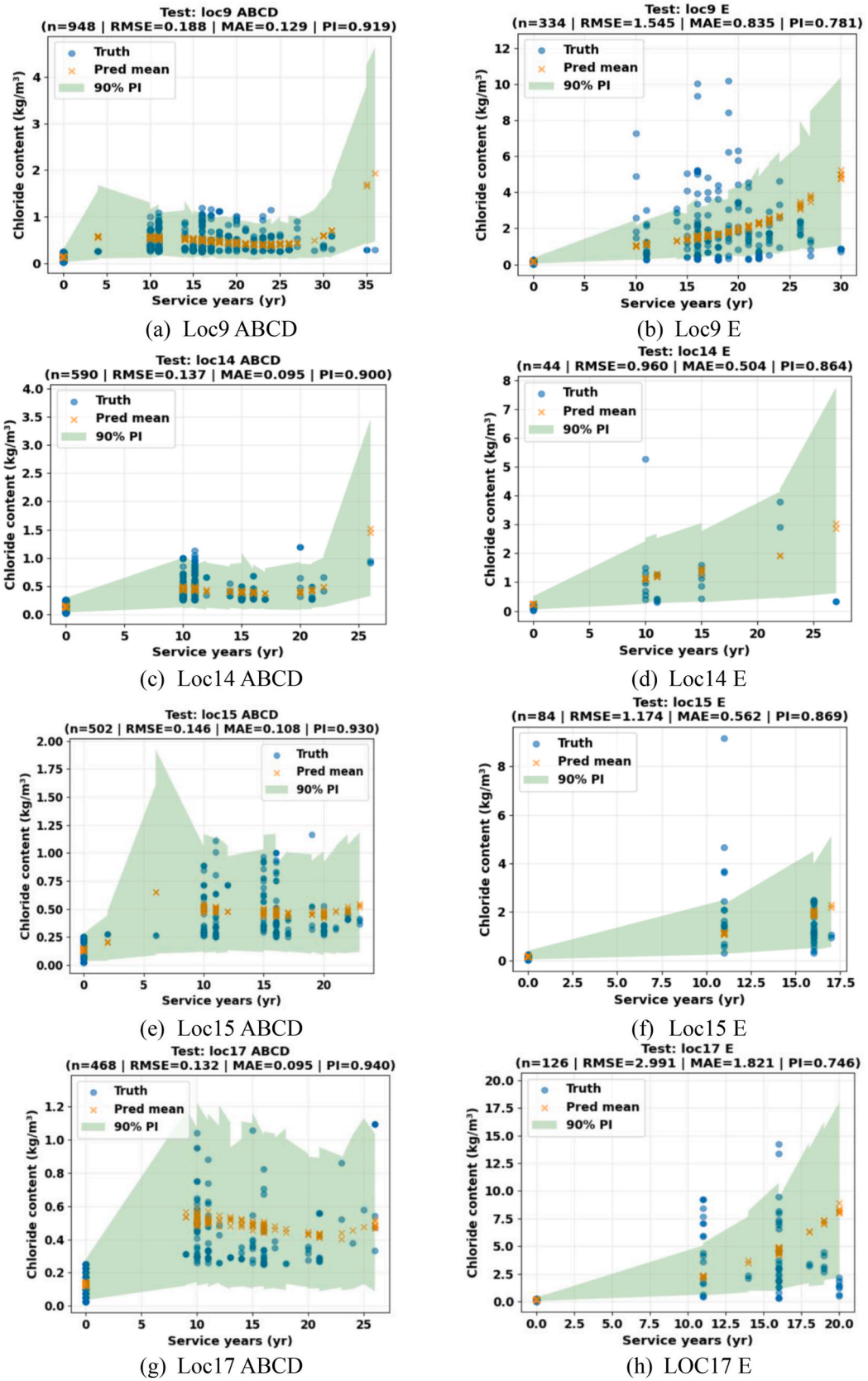
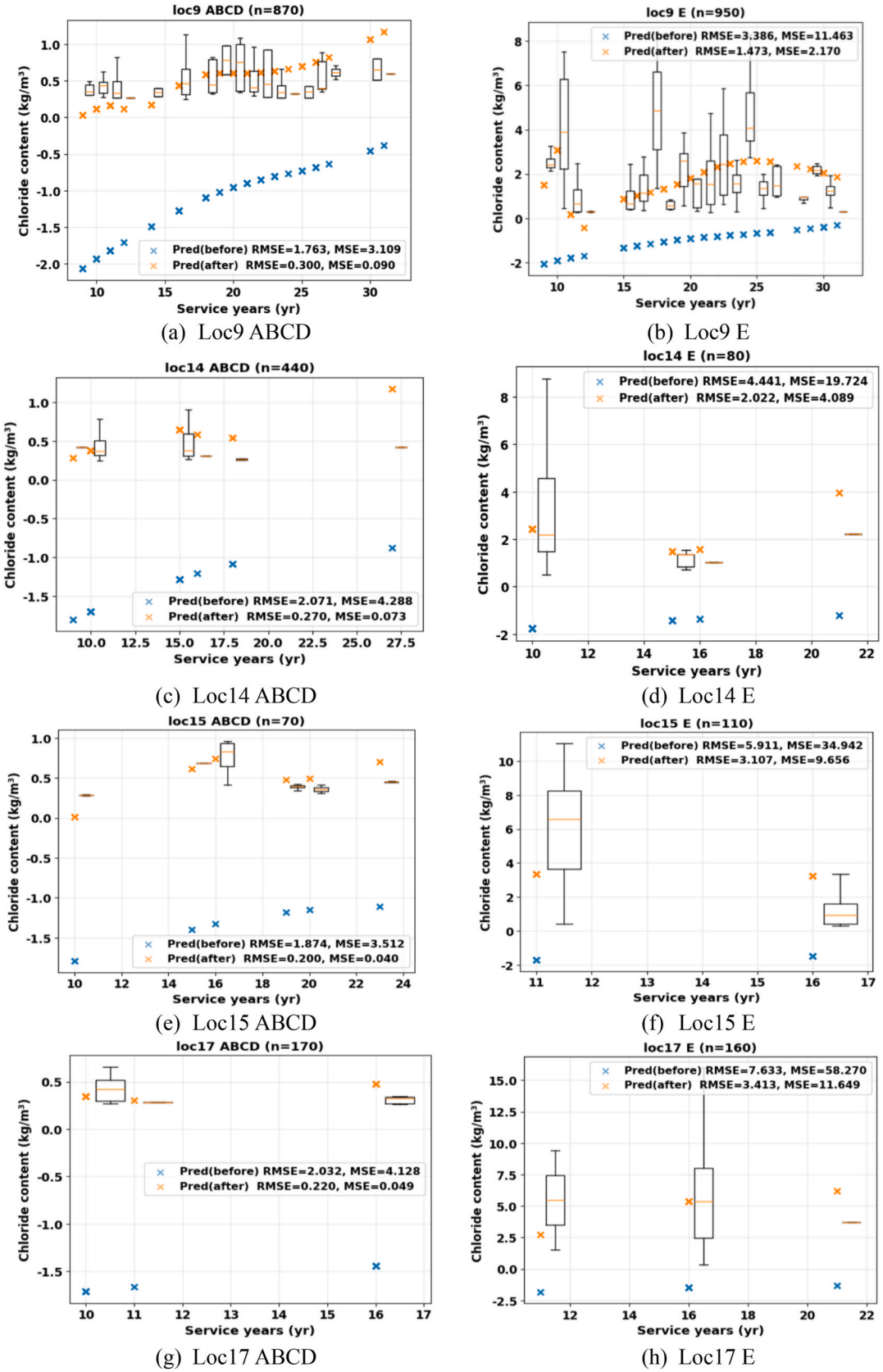



**Fig. 10.** Test data results for the top-four region.




(a) Loc9 ABCD (b) Loc9 E

(c) Loc14 ABCD (d) Loc14 E

(e) Loc15 ABCD (f) Loc15 E

(g) Loc17 ABCD (h) Loc17 E

**Fig. 11.** Comparison of the predictions of the ANN model before and after BNN weight tuning.

between the generalized structural prior of the global model and parameter adaptation driven by the local model dataset. These findings demonstrate the effectiveness of the FL-based personalization strategy in deterioration-prediction problems with heterogeneous spatial distributions and data imbalance, preserving global generalization while improving local data fidelity.

### 4.3. Parameter integration between global and local models

Two predictions were compared for each administrative district: the BNN-initialized ANN prediction (before fine-tuning) and ANN prediction after local fine-tuning. Performance was evaluated on the raw scale (chloride content) based on the MSE, RMSE, and MAE values. This comparison quantitatively assesses the extent to which the transfer of structural information from the BNN improves the regional data fidelity. During the ANN training, the parameters were tuned via backpropagation at the individual bridge level. Accordingly, the tuned parameters from all the bridges within the local dataset were aggregated and approximated as normal distributions. The resulting ANN-derived parameter distributions were then reintegrated into the original global model structure and evaluated using the global dataset test data. A comparison between the original global BNN and model constructed using the ANN-derived parameter distributions is shown in Fig. 12.

As expected, the BNN, whose posterior distributions were inferred from a large dataset, demonstrated superior predictive accuracy. Conversely, the performance of the local models varied according to the service-year distribution and number of observations within specific intervals, resulting in differences in both accuracy and predictive variance. The local model parameters were transmitted to the central server to incorporate information from the local dataset into the central model. The weight distributions of the local and central models were then combined to update the central model. As shown in Fig. 13, the two normal distributions were fused to reconstruct the updated central weight distribution.

The region-wise predictive performance of the BNN model constructed using the fused weight distribution and that of the original global BNN model are compared in Fig. 14. This analysis was conducted on four representative administrative districts with sufficient data (loc9, loc14, loc15, and loc17), separately for the ABCD and E grade subsets, and the variations in RMSE, MAE, and PI coverage across service-year intervals are also shown within each region.

The analysis showed that the fused model preserved the performance of the baseline BNN in most regions while achieving additional improvement in some regions. In the data-rich ABCD grade, the performance difference between before and after fusion remained generally marginal, which indicates that the global BNN had already secured stable predictive performance in these regions, leaving little room for further improvement. Nevertheless, clear improvements due to fusion were observed in some service-year intervals; in particular, in the (AASHTO, 2020; Liu et al., 2021) and (Liu et al., 2021; Abdellatief et al., 2025) intervals of loc17 ABCD, the RMSE decreased by approximately 17–39%, indicating a pronounced fusion effect. In addition, in the 0–10 year service-year interval of loc9 ABCD and loc15 ABCD, the fused model effectively reduced the predictive dispersion (Fig. 14(a) and (e)). In the E grade, the fusion effect varied across service-year intervals depending on the amount and distributional characteristics of the data in each region, which suggests that the region-specific information extracted from the ANN ensemble contributes to the global model to different degrees across regions in data-scarce environments.

These results demonstrate that the fusion approach, which integrates the global structural information held by the BNN with the region-specific characteristics learned from the local ANN at the distribution level, stably maintains region-wise predictive performance and provides additional improvement in some regions compared to using a single global model alone.

The same analysis was also conducted on regions with relatively scarce data, and the results are presented in Fig. 15. In both regions, the improvement due to fusion was more pronounced compared to that observed in the data-rich representative regions, and the effect was clearly observed particularly in the service-year intervals with a small number of samples. In the case of loc7 ABCD, both RMSE and MAE decreased by approximately 79–82% in the (AASHTO, 2020; Liu et al., 2021) and (Abdellatief et al., 2025; Zhao et al., 2018) intervals, where the number of samples was very limited (5 and 4, respectively); this indicates that, in an environment where the global BNN alone could not achieve sufficient learning, the fusion effectively complemented the region-specific information from the ANN ensemble. In loc4 ABCD as well, a fusion effect was confirmed, with RMSE decreasing by approximately 16% in the [0,5] interval, which has the largest number of samples ($n = 108$). These results demonstrate that fusion functions as a mechanism that complements the deficiency of the global BNN with region-specific information in regions or service-year intervals where the data are scarce.

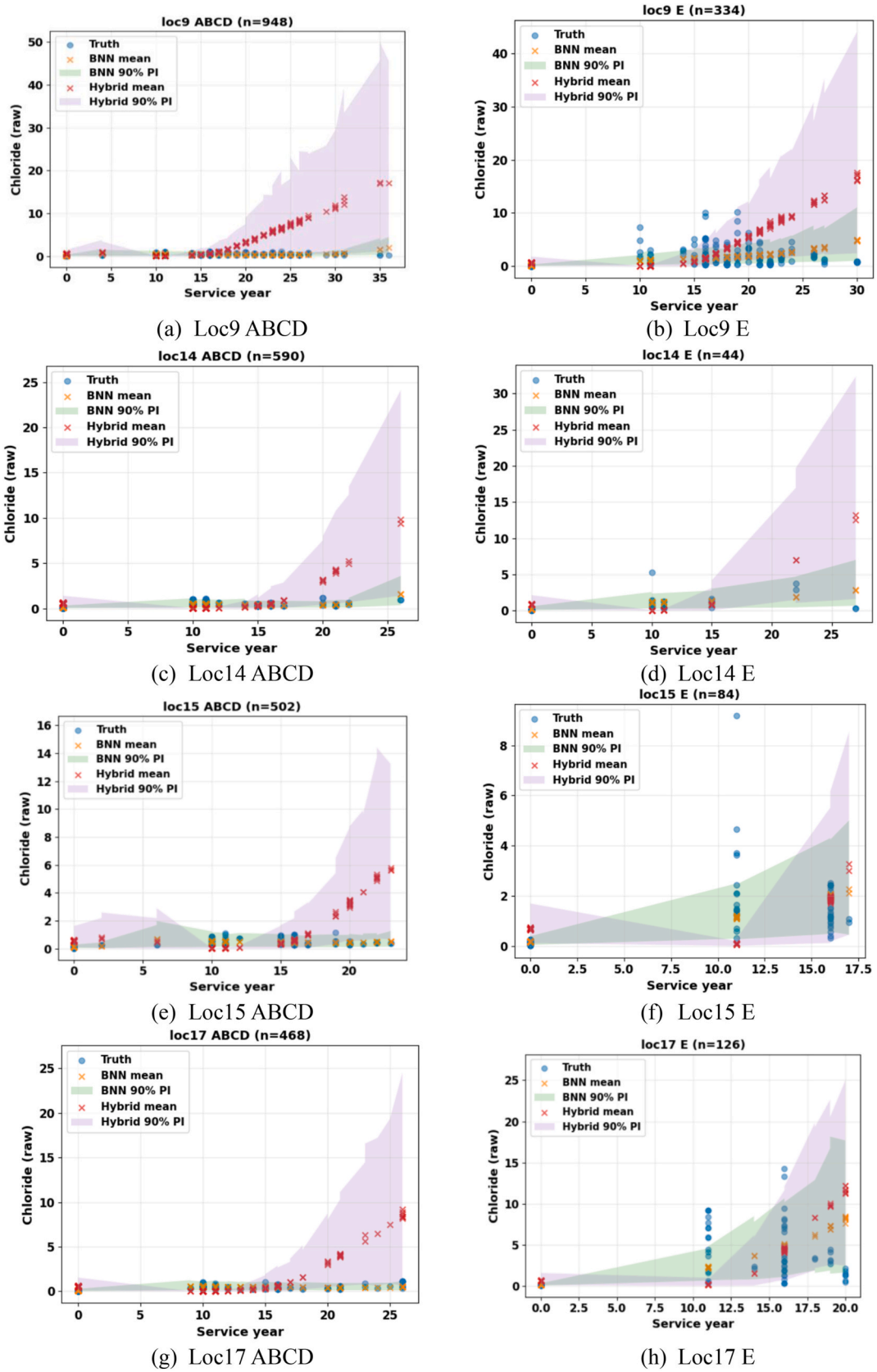
### 4.4. Adjusting the weight of distribution of the local model

#### 4.4.1. Effect of weight ($\lambda$) on predictive accuracy

A precision-weighted fusion technique was applied to integrate the posterior distributions of the time-dependent parameters from the BNN with the normal distributions estimated from the ANN ensemble. The parameter $\lambda$ functioned as a weighting coefficient that controls the relative confidence assigned to the ANN-based distribution; therefore, larger values of $\lambda$ increase the contribution of the ANN posterior to the fused distribution. To investigate the sensitivity of the predictive accuracy to this parameter, $\lambda$ was varied over the range {0.01, 0.03, 0.1, 0.3, 1.0, 3.0, 10.0}. For each value of $\lambda$, four categories of performance metrics were evaluated: (i) global accuracy was quantified by the RMSE and MAE between the predicted mean and observed values; (ii) calibration was assessed using the 90% PI coverage, defined as the proportion of observations contained within the 90% prediction interval; (iii) sharpness was measured by the mean PI width, representing the average width of the 90% PI; and (iv) location-wise performance was characterized by $\Delta\text{RMSE} = \text{RMSE}_{\text{fused}} - \text{RMSE}_{\text{baseline}}$, in which negative values indicate performance improvement over the baseline. This procedure enabled the simultaneous assessment of overall global performance and region-specific fusion effects. Figs. 16 and 17 illustrate how variations in $\lambda$ influence model accuracy, uncertainty characteristics, and regional performance within the precision-weighted fusion framework.

Fig. 16 (a) shows the global prediction accuracy as a function of the fusion weight $\lambda$. For $\lambda \leq 1$, both RMSE and MAE remain nearly constant, indicating stable predictive performance within this range. However, when $\lambda \geq 3$, the error metrics exhibit a gradual increase. This trend suggests that the ANN posterior does not outperform the BNN posterior in terms of mean predictive accuracy. Consequently, assigning excessive weight to the ANN-derived information may degrade the overall predictive performance of the fused model. Fig. 16 (b) illustrates the average PI width (sharpness) as a function of $\lambda$. As $\lambda$ increases, the PI width decreases, reflecting the reduction in posterior variance due to the increased precision contributed by the ANN component in the fusion process. In this context, $\lambda$ acts as an uncertainty shrinkage factor: larger values of $\lambda$ lead to narrower PIs and consequently reduced predictive uncertainty.

Fig. 17 (a) shows the variation of the 90% PI coverage with respect to $\lambda$. For $\lambda \leq 1$, the coverage remains relatively stable at approximately 0.895–0.897. However, as $\lambda$ increases beyond 1, the coverage gradually decreases, remaining below the target value of 0.90 across all $\lambda$ values and reaching its lowest level at $\lambda = 10$. This behavior indicates that increasing the contribution of the ANN-based precision progressively narrows the PIs, leading to systematic under-coverage. Therefore, assigning excessive weight to the ANN posterior results in overconfident predictions and an underestimation of uncertainty. The regional




(a) Loc9 ABCD (b) Loc9 E

(c) Loc14 ABCD (d) Loc14 E

(e) Loc15 ABCD (f) Loc15 E

(g) Loc17 ABCD (h) Loc17 E

**Fig. 12.** Comparison of accuracy between the global BNN and the locally fine-tuned ANN.

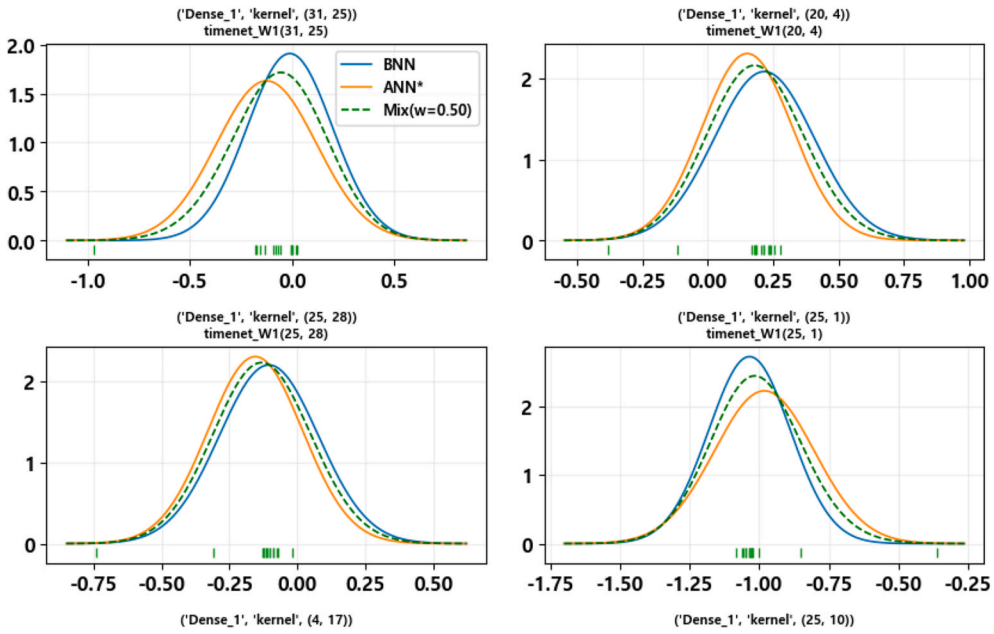



**Fig. 13.** Integration of the normal distributions of BNN and ANN model weights.

variation in performance is illustrated in the ΔRMSE heatmap shown in Fig. 17 (b). In some regions, the application of fusion resulted in a reduction in the RMSE, indicating performance improvement. However, in other regions, performance degradation is observed. Notably, as $\lambda$ increases, the number of regions exhibiting performance deterioration also increases. This trend suggests that the ANN-based posterior is not uniformly beneficial across all regions, and assigning excessive weight to the ANN-derived information may adversely influence the regional predictive performance.

Fig. 18 shows the RMSE–$\lambda$ curves for the regions where the performance-improvement effect was the most pronounced. In most regions, the RMSE is minimal within the range $\lambda \approx 0.01$–$1.0$, whereas performance degradation is observed as $\lambda$ increases beyond this range. This pattern suggests the existence of region-specific optimal $\lambda$ values and indicates that an adaptive fusion strategy at the regional level may be more effective than a single global $\lambda$ in improving overall model performance.

Overall, the analysis demonstrates that, although the ANN posterior has limitations when used as a standalone replacement model, it can serve as a complementary information source when combined with the BNN posterior through precision-weighted fusion, yielding performance improvements in certain regions. Furthermore, excessively large values of $\lambda$ can simultaneously lead to underestimation of uncertainty and degradation in predictive accuracy. For the present dataset, the range $0.1 \leq \lambda \leq 1$ was observed to provide the most stable balance between accuracy and calibration.

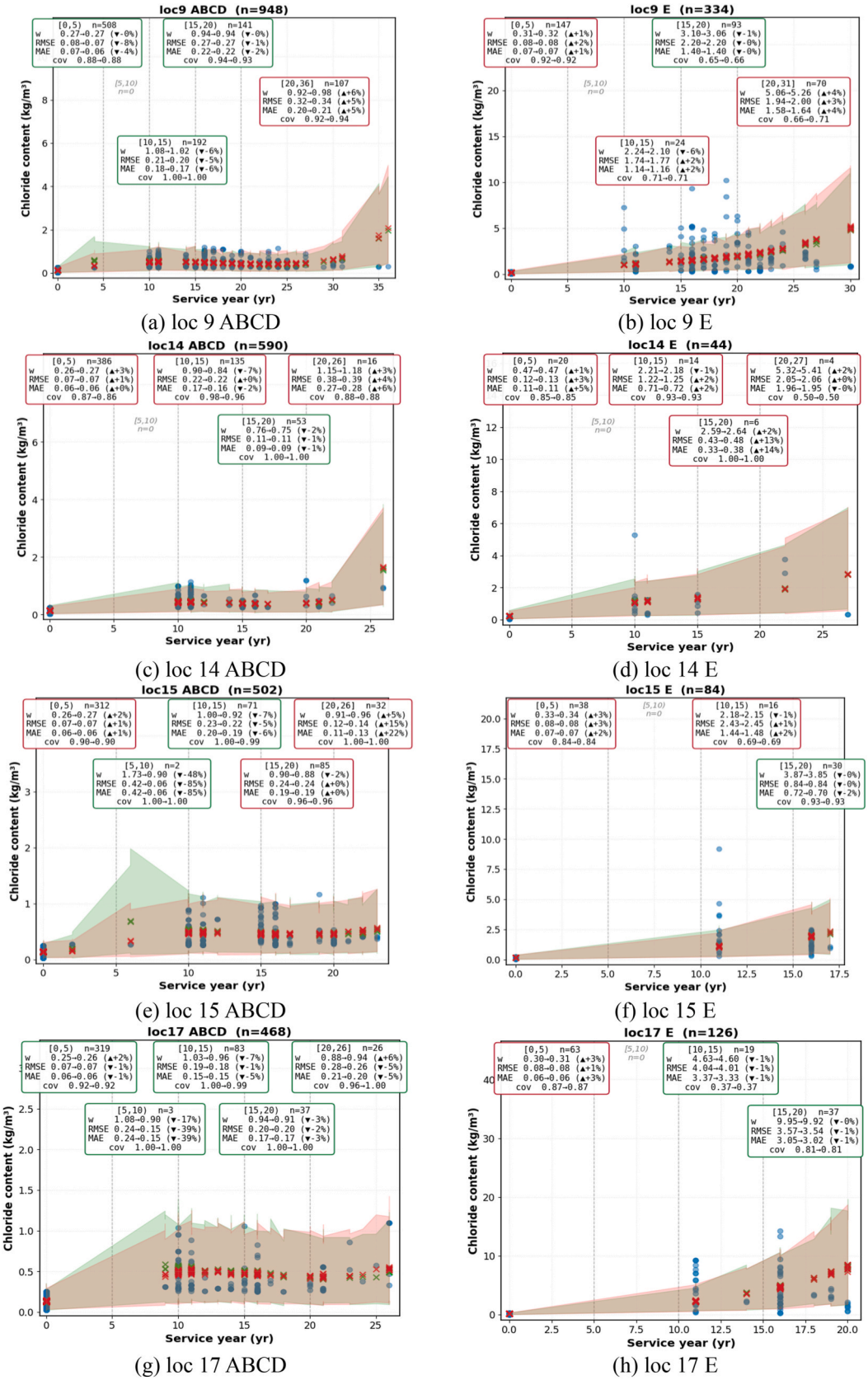
#### 4.4.2. *Calibration analysis of the fused predictive distribution*

The sensitivity analysis in Section 4.4.1 evaluated the influence of fusion on global predictive accuracy and on the reliability at a single quantile through RMSE, MAE, and 90% PI coverage. However, the extent to which fusion preserves the probabilistic reliability of the entire predictive distribution of the BNN is difficult to confirm sufficiently using single-quantile-based metrics alone. To this end, two distribution-level calibration metrics are introduced in this section to additionally verify the fusion results. The first is the expected calibration error (ECE), which quantifies the average difference between the nominal quantiles claimed by the model and the empirical frequencies of observations actually contained within those quantiles. The second is the continuous ranked probability score (CRPS), a proper scoring rule that measures the distance between the entire predictive distribution and the actual observations. For both metrics, values closer to zero indicate better calibration.

The calibration evaluation was performed on two groups of regions. The first group consists of the four representative administrative districts (loc9, loc14, loc15, and loc17), and the second group consists of the four regions (loc10 E, loc4 E, loc1 ABCD, and loc12 ABCD) in which a clear improvement in calibration due to fusion was observed, which were selected to jointly analyze cases where fusion complements the deficiencies of the global BNN. For each region, the ECE and CRPS values of the model at different λ values are summarized in Table 5, and the calibration plots at the λ value showing the best calibration within the four administrative districts are presented together in Fig. 19.

The analysis revealed three consistent trends. First, the REPLACE model, in which the TimeNet posterior distribution of the BNN is completely replaced with the ANN ensemble, showed a substantial deterioration in both CRPS and ECE compared to the BASELINE in all regions. The increase in ECE was particularly pronounced, with a deterioration on the order of 4 to 9 times observed in most regions, which indicates that the direct replacement of the global TimeNet distribution undermines the probabilistic structure of the predictive distribution. Second, in the four data-rich representative administrative districts, the FUSED model maintained CRPS and ECE values that were essentially identical to those of the BASELINE regardless of the variation in λ, with the difference generally remaining within a few percent. This quantitatively demonstrates that the precision-weighted Gaussian fusion preserves the probabilistic reliability of the global BNN while integrating the information from the ANN side. Third, in the fusion-improved regions, a clear calibration improvement compared to the BASELINE was observed at appropriate λ values. In particular, in loc1 ABCD, the ECE decreased from 0.0417 to 0.0226 at $\lambda = 10$, corresponding to a reduction of approximately 46%, and in loc10 E and loc4 E, where the number of samples is limited, simultaneous improvements in both CRPS and ECE were observed within a small λ range (0.01–0.1). This indicates that the ANN ensemble retained region-specific information that had not been sufficiently captured by the global BNN, and that the precision-weighted fusion integrated such information into the global distribution without undermining its probabilistic structure.

The calibration plots at the λ values that yielded the best calibration in each region are presented in Fig. 20. While the BASELINE and FUSED




(a) loc 9 ABCD (b) loc 9 E

(c) loc 14 ABCD (d) loc 14 E

(e) loc 15 ABCD (f) loc 15 E

(g) loc 17 ABCD (h) loc 17 E

**Fig. 14.** Comparison between the baseline BNN model and fused BNN–ANN model.

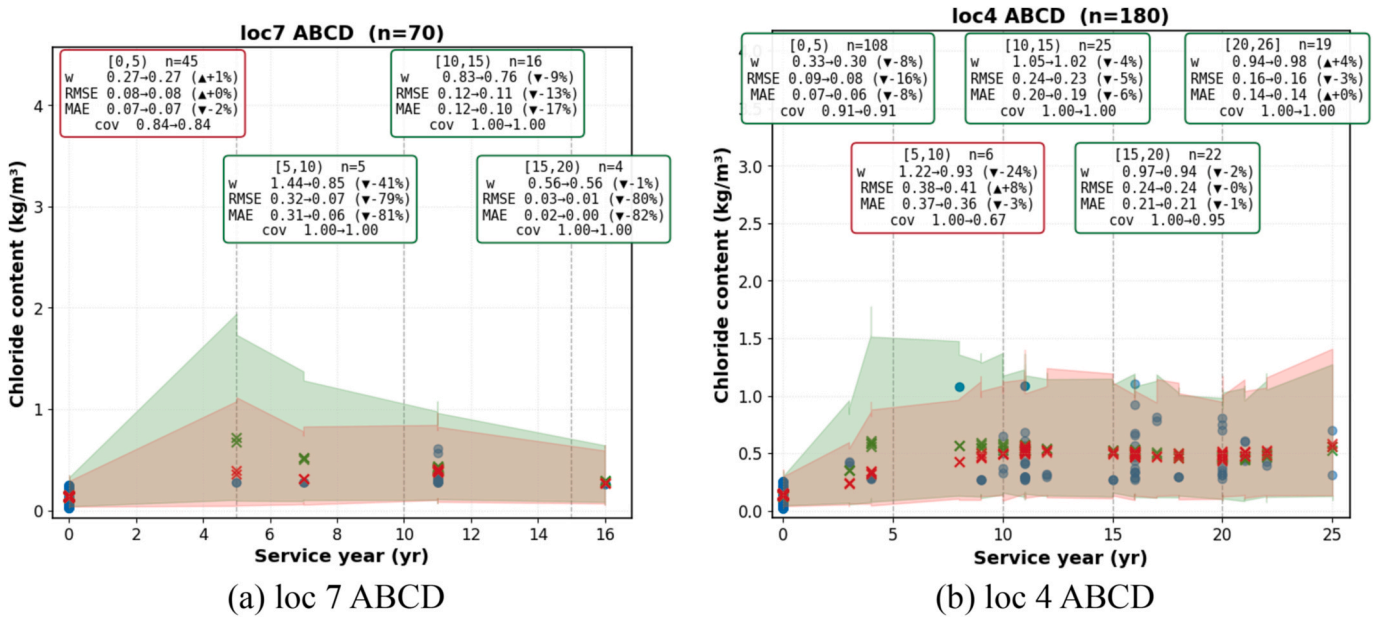



**Fig. 15.** Comparison between the baseline BNN model and the fused BNN–ANN model in data-scarce regions.

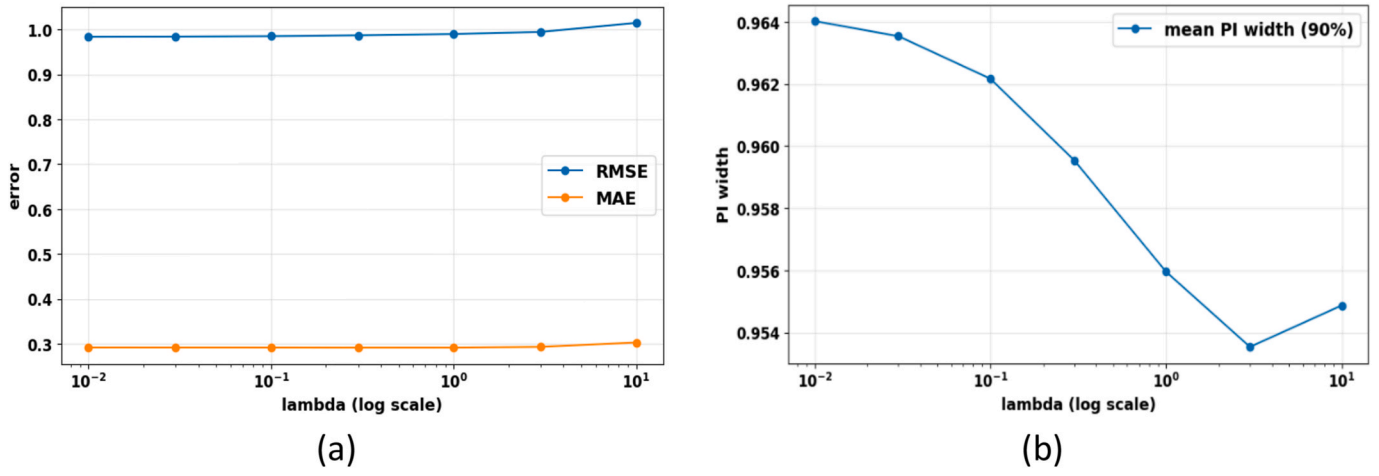



**Fig. 16.** (a) Global prediction accuracy versus fusion weight λ, (b) Prediction interval coverage (90% PI) as a function of fusion weight λ

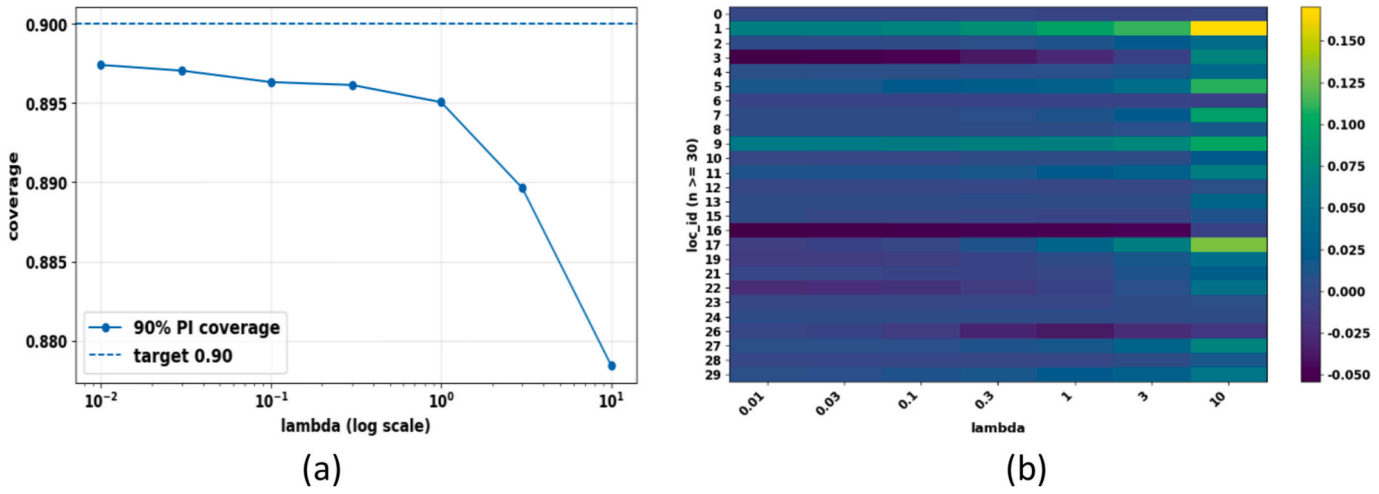



**Fig. 17.** (a) Mean prediction interval width versus fusion weight λ, (b) Per-location RMSE difference between baseline BNN and fused Model across λ values.

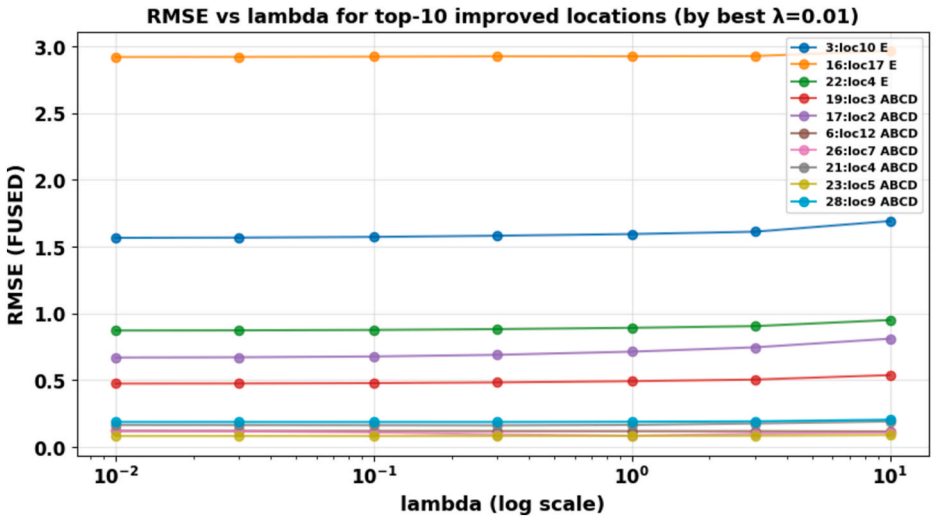



**Fig. 18.** RMSE sensitivity to fusion weight λ for top-10 improved locations.

**Table 5**
Calibration metrics (CRPS/ECE) at different λ values for the representative administrative districts and the fusion-improved regions.

| Region | BASELINE (BNN) | REPLACE (ANN) | FUSED (0.01) | FUSED (0.1) | FUSED (1) | FUSED (10) |
|---|---|---|---|---|---|---|
| loc9 ABCD | 0.0857/0.0181 | 0.6581/0.1986 | 0.0850/0.0202 | 0.0847/0.0209 | 0.0844/0.0220 | 0.0894/0.0177 |
| loc9 E | 0.5997/0.0519 | 1.4666/0.2414 | 0.6028/0.0548 | 0.6044/0.0559 | 0.6105/0.0580 | 0.6283/0.0773 |
| loc14 ABCD | 0.0673/0.0329 | 0.3359/0.2003 | 0.0675/0.0349 | 0.0676/0.0364 | 0.0678/0.0393 | 0.0764/0.0489 |
| loc14 E | 0.3494/0.1483 | 1.0360/0.1902 | 0.3560/0.1531 | 0.3559/0.1471 | 0.3570/0.1459 | 0.3735/0.1316 |
| loc15 ABCD | 0.0714/0.0286 | 0.3814/0.2024 | 0.0713/0.0285 | 0.0712/0.0280 | 0.0713/0.0298 | 0.0756/0.0255 |
| loc15 E | 0.4192/0.0630 | 0.7531/0.1238 | 0.4194/0.0574 | 0.4205/0.0568 | 0.4217/0.0593 | 0.4396/0.0706 |
| loc17 ABCD | 0.0638/0.0280 | 0.4008/0.2306 | 0.0635/0.0300 | 0.0632/0.0288 | 0.0623/0.0311 | 0.0684/0.0305 |
| loc17 E | 1.2281/0.0584 | 1.7864/0.2135 | 1.2012/0.0657 | 1.2031/0.0657 | 1.2065/0.0662 | 1.2302/0.0807 |
| loc10 E | 0.5190/0.0744 | 1.5530/0.2237 | 0.5035/0.0667 | 0.5065/0.0656 | 0.5144/0.0691 | 0.5590/0.0718 |
| loc4 E | 0.3856/0.1386 | 2.1326/0.2930 | 0.3730/0.1316 | 0.3754/0.1333 | 0.3835/0.1404 | 0.4116/0.1526 |
| loc1 ABCD | 0.0881/0.0417 | 1.0882/0.3053 | 0.0882/0.0383 | 0.0877/0.0353 | 0.0867/0.0297 | 0.0878/0.0226 |
| loc12 ABCD | 0.0601/0.0225 | 0.3530/0.2249 | 0.0595/0.0218 | 0.0592/0.0255 | 0.0587/0.0289 | 0.0579/0.0408 |

curves are located almost identically on, or even closer to, the perfect calibration line, the REPLACE curve clearly deviates from the perfect calibration line in all regions. In particular, in the fusion-improved regions, the FUSED curve is positioned closer to the perfect calibration line than the BASELINE curve, which visually confirms that fusion improves calibration by integrating the information from the ANN side.

## 5. Discussion

### 5.1. Characteristics of the proposed framework

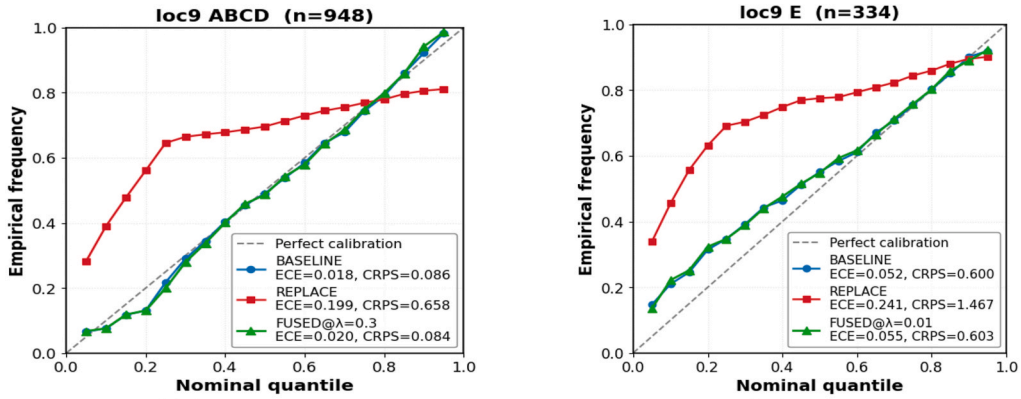
In this study, an FL framework combining a hierarchical BNN with client-wise ANNs was proposed for the service-year-based prediction of region-specific chloride content in bridge decks. The proposed FL structure functions as a mechanism that goes beyond a simple distributed learning scheme and simultaneously achieves three objectives: (i) securing the generalization capability of the global model, (ii) enhancing the adaptability to regional characteristics, and (iii) maintaining structural consistency between the global and local models. The global model stably captured the temporal deterioration patterns generalizable across regions even under heterogeneous environmental conditions, and the systematic error specific to each administrative district was reduced through local fine-tuning. In addition, the distribution-based precision-weighted fusion strategy integrated global and region-specific information in a balanced manner, thereby improving the predictive performance while preventing excessive shrinkage of the posterior distribution.

The objective is to construct a predictive model that explicitly takes into account the regional, numerical, and environmental uncertainties inherently embedded in chloride within bridge decks, while at the same time continuously reflecting the temporal behavior of chloride that accumulates over time. To achieve this objective, a structure was adopted in which the global model retains the generalized temporal patterns of all regions, and the local information of each client is continuously reflected into the global model in the form of region-specific parameters. In other words, the global model functions not as a static single training output, but as a dynamic information repository that becomes progressively robust as the information accumulated by each region is integrated at the parameter level.

### 5.2. Limitations

#### 5.2.1. Limitations of the normal approximation assumption on ANN weights

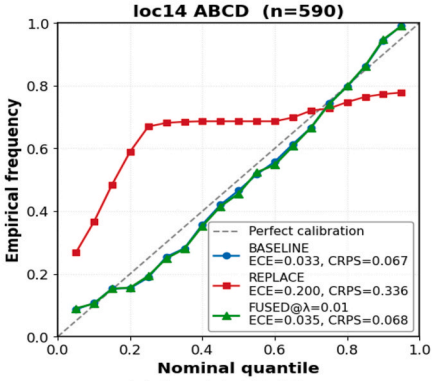
At the fusion stage of the model parameters, the weight distribution of the ANN ensemble is approximated as an element-wise Gaussian. This assumption is a choice made to maintain consistency with the AutoNormal variational guide adopted by the global BNN, but it is not a statistically complete assumption in that the distribution is estimated from a limited sample of 17 administrative districts. In addition, since the data quality and distributional characteristics of each administrative district are not uniform, there is a possibility that the ANN-side distribution is influenced by the data bias of some regions. In this study, the weight distributions on both the ANN and BNN sides were assumed to belong to the Gaussian family and were fused in a precision-weighted manner; however, the possibility cannot be ruled out that the actual weights of the ANN ensemble exhibit non-Gaussian distributional characteristics such as heavy-tailed or multimodal forms. Additional



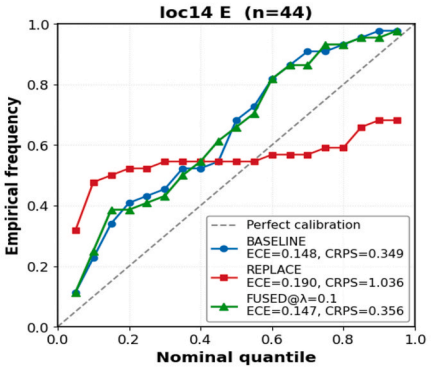

(a) loc 9 ABCD



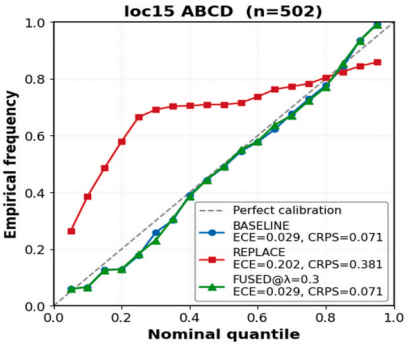

(b) loc 9 E



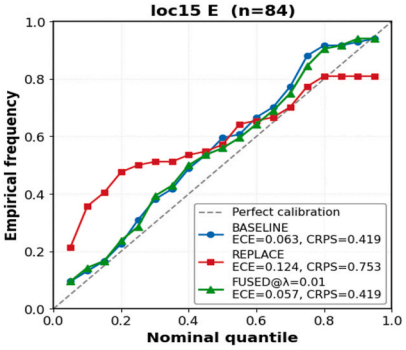

(c) loc 14 ABCD



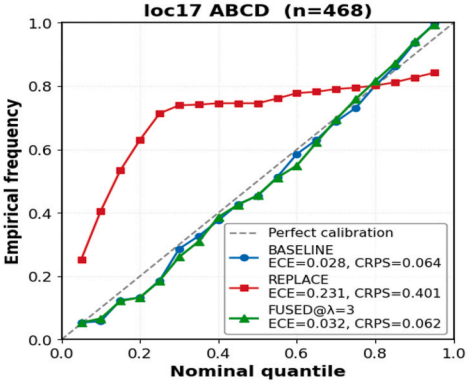

(d) loc 14 E



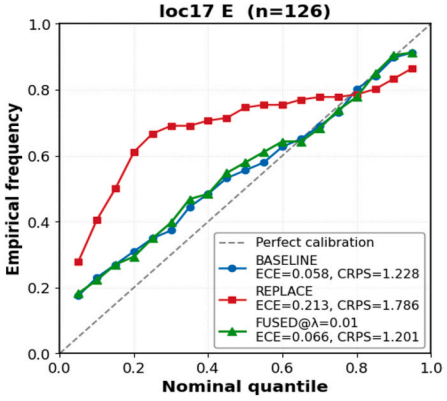

(e) loc 15 ABCD




(f) loc 15 E




(g) loc 17 ABCD




(h) loc 17 E

**Fig. 19.** Calibration plots of the BASELINE, REPLACE, and FUSED models for the four representative administrative districts.

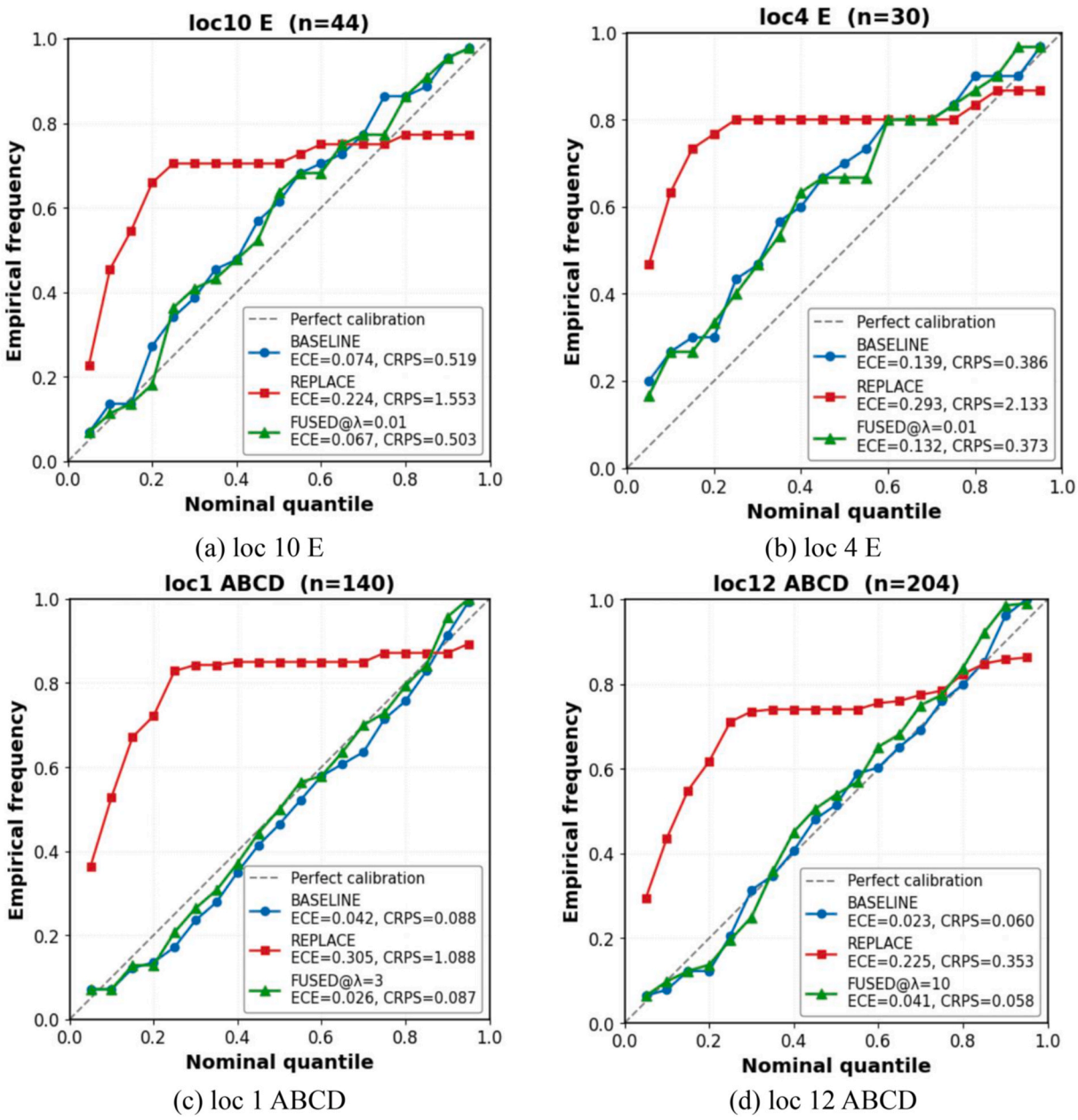



**Fig. 20.** Calibration plots of the BASELINE, REPLACE, and FUSED models for the four regions where fusion yielded a clear improvement in calibration.

quantitative evaluations are required regarding the suitability and statistical reliability of various distribution families in the process of exchanging and fusing model parameters between the ANN and BNN.

*5.2.2. Single-round federated structure*

The framework presented in this paper was designed to be generalizable to a multi-round federated configuration, but only a single-round instantiation was validated in this study. This is because a sufficient amount of service-year data had to be secured at once in order to statistically and meaningfully estimate the weight distribution (mean and standard deviation) of the ANN ensemble, and when data accumulating along the time axis are divided into rounds, the reliability of the distribution estimation at each round is degraded. In addition, the negative influence of the sparsity or distributional bias of some clients on the global model update during the inter-round fusion process may manifest as a behavior similar to data poisoning in a repeated training environment, potentially causing a cumulative degradation of model accuracy. In this study, such risks do not arise within the single-round structure, but for multi-round extension, a robust aggregation strategy is also required.

In addition, the multi-round extension can be divided into two paths depending on how the fusion weighting coefficient λ is configured. When λ is fixed to the same value across all regions, the fusion result retains the form of a single global BNN, which can therefore be used as the starting point for the next round; through this, the framework extends to a multi-round federated learning scheme that progressively improves by cycling through rounds within the same structure. In contrast, as confirmed by the sensitivity analysis in Section 4.4.1, the optimal λ differs across regions, and when region-specific λ values are applied to exploit this property, the fusion results branch by region and the form of a single global BNN is no longer maintained. From this point onward, multi-round iteration within the federated structure no longer holds, and subsequent model improvement must be transitioned into the form of region-level multi-round updates in which each region independently performs updates using its own data.

#### 5.2.3. Limitations of region modeling and input data

The current framework represents region-specific deterioration behavior through exponential trend parameters based on administrative district identifiers. Because the input data have a structure in which such diverse variables are difficult to consider all at once, the framework is performed at the administrative district level, and the update process is also designed in a way that updates only the uncertainty of TimeNet. The reason for this is that, since the differences between data are currently difficult to distinguish, it is considered that adjusting the parameters of the exponential function would cause the range of uncertainty inherent in the function itself to become very wide.

However, since chloride behavior is strongly influenced by more fine-grained environmental factors such as coastal exposure, the use of deicing salts during snowfall, and micro-climate conditions, there is a limitation that environmentally heterogeneous bridges are grouped together within the same administrative district. In addition, the current model structure provides robustness through collaborative learning across clients for the TimeNet component, which represents the time-dependent uncertainty, but it does not provide a progressive update mechanism for the exponential parameters that determine the region-specific deterioration trend.

In addition, structural improvement is also required in terms of input variables and measurement data. Currently, the data use service years and administrative district identifiers as the main inputs, while detailed factors such as exposure environment, coastal proximity, deck location, drainage condition, and material properties are not explicitly included in the model. In this study, the framework was designed such that the influence of these factors is indirectly absorbed through the administrative-district-level hierarchical structure and the uncertainty modeling of the BNN; however, this is an approach that inherently entails substantial uncertainty and acts as a limitation that affects the accuracy of the model. In addition, due to the drainage slope of the deck, relatively high chloride concentrations are observed in the cantilever region, which represents a spatial bias according to the sampling location within a bridge, and the current inspection standards and guidelines lack detailed regulations regarding sampling locations, measurement intervals, and data collection across service-year intervals. The data used in this study are also unevenly distributed across service-year intervals, which results from the fact that chloride measurement is not a mandatory item during periodic inspections. In other words, in addition to a multi-dimensional approach to the variables, it is necessary to establish a more detailed region-wise hierarchical structure and to update the hierarchical function during the federated learning process.

## 6. Conclusion

In this study, an FL-based chloride prediction framework was proposed to address key structural limitations in infrastructure deterioration modeling, particularly data-sharing constraints and the scarcity of long-term time-series data. The framework integrates a H-BNN, an ANN-based transfer learning, and weight-distribution fusion. The chloride content data of the bridge deck exhibit substantial variability owing to differences in sampling location, measurement timing, and environmental conditions. However, long-term observations are often limited, resulting in inherently high uncertainty. To explicitly represent these characteristics, the H-BNN was adopted as the backbone of the global model, enabling the simultaneous representation of aleatoric uncertainty resulting from observational variability and epistemic uncertainty associated with the model structure and data scarcity. The contributions of this study are presented subsequently.

- This study established a data-driven FL framework for chloride assessment and prediction in bridge decks, where chloride data are not systematically managed or long-term time-series observations are insufficient. By leveraging the distributed inspection data, the proposed framework facilitated collaborative model training through parameter exchange without sharing raw data across management agencies or bridge units. Even for data-scarce bridges, it facilitated the quantitative evaluation of chloride content and long-term deterioration trends by transferring knowledge from regions with similar environmental exposure conditions. Accordingly, the proposed system addresses both the data security and scarcity limitations associated with conventional centralized data-driven models.
- To incorporate the spatial heterogeneity of chloride ingress driven by regional environmental factors such as coastal exposure, deicing-salt application, and climate, a hierarchical FL architecture based on administrative districts was proposed. In this structure, a global model is trained on the entire dataset to capture generalized long-term deterioration trends, whereas local models reflect region-specific environmental characteristics and data distributions. Global and regional information were jointly integrated through parameter exchange. This architecture preserves the generalization capability of the global model while enabling regional personalization, thereby providing a structured solution to spatially heterogeneous chloride deterioration problems.
- Given the high dispersion and asymmetric distributional characteristics of the bridge-deck chloride data, a probabilistic learning structure was adopted as the core modeling mechanism. Unlike single-point estimation models, both parameters and predictions were represented as probability distributions, enabling the quantitative characterization of uncertainty in deterioration forecasting. A posterior-level weight fusion strategy was employed to integrate the regional information while preserving the probabilistic structure of the global model. This approach maintains structural consistency in the parameter space between the global and local models, enabling information to be combined without compromising model coherence. Consequently, both the generalization capability of the global model and adaptability of the local models are retained. Through this framework, a probabilistic evaluation system was established that provides not only predictive means but also prediction intervals.

## Data statement

The data are not publicly available due to legal or confidentiality agreements.

## Funding

This research was supported by the BIM-based engineering data standardization for intelligent design and smart construction project (RS-2025-25455930), funded by the Korean Agency for Technology and Standards (KATS) under the Ministry of Trade, Industry and Energy (MOTIE) and administered by the Korea Planning & Evaluation Institute of Industrial Technology (KEIT). This study was also supported by the KAIA Research program on the Development of Marine Bridges Operation using Green Energy, funded by the Ministry of Land, Infrastructure and Transport [Grant number: RS-2024-00401101].

## CRediT authorship contribution statement

**Chi-Ho Jeon:** Data curation, Writing – review & editing. **Gitae Roh:** Conceptualization, Data curation, Funding acquisition, Methodology, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Chang-Su Shim:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing. **In Ho Cho:** Supervision, Writing – review & editing. **Ki-Yeol Kim:** Data curation. **Ji-hun Shin:** Data curation. **Ki-Tae Park:** Funding acquisition.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence

the work reported in this paper.

## Data availability

The data that has been used is confidential.

## References


AASHTO, 2020. AASHTO LRFD Bridge Design Specifications, ninth ed. American Association of State Highway and Transportation Officials, Washington, DC, USA.

Abdellatief, M., Abd-Elmaboud, M.E., Mortagi, M., Saqr, A.M., 2025. A convolutional neural network-based deep learning approach for predicting surface chloride concentration of concrete in marine tidal zones. Sci. Rep. 15, 27611.

Abdulkarem, M., Alwasel, K., Alhassani, Y., 2023. Personalised federated learning framework for damage detection in structural health monitoring. J. Civ. Struct. Health Monit. 13, 295–308.

Ahmad, A., Farooq, F., Ostrowski, K.A., Śliwa-Wieczorek, K., Czarnecki, S., 2021. Application of novel machine learning techniques for predicting the surface chloride concentration in concrete containing waste material. Mater 14, 2297.

Alhakeem, Z.Z., 2020. Deterioration Prediction Models for Condition Assessment of Concrete Bridge Decks Using Machine Learning Techniques. Western Michigan University, Dissertation.

Ashraf, M., Iqbal, M.F., Rauf, M., Ashraf, M.U., Ulhaq, A., Muhammad, H., Liu, Q.F., 2022. Developing a sustainable concrete incorporating bentonite clay and silica fume: mechanical and durability performance. J. Clean. Prod. 337, 130315.

Bertolini, L., Elsener, B., Pedeferri, P., Redaelli, E., Polder, R., 2013. Corrosion of Steel in Concrete: Prevention, Diagnosis, Repair, second ed. Wiley-VCH, Weinheim, Germany.

Bradbury, J., Frostig, R., Hawkins, P., Johnson, M.J., Leary, C., Maclaurin, D., Zhang, Q., 2018. JAX: Composable Transformations of Python+NumPy Programs.

Cady, P.D., Weyers, R.E., 1983. Chloride penetration and the deterioration of concrete bridge decks. Cem. Concr. Aggregates 5, 81–87.

CEN, 2013. EN 206:2013, Concrete – Specification, Performance, Production and Conformity. European Committee for Standardization (CEN), Brussels, Belgium.

Cheema, M.A., Sarwar, M.Z., Cantero, D., Rossi, P.S., 2025. Clustered federated learning for population-based structural health monitoring. IEEE Internet Things J.

Elsener, B., 2001. Half-cell potential mapping to assess repair work on RC structures. Constr. Build. Mater. 15, 133–139.

Golafshani, E.M., Kashani, A., Kim, T., Arashpour, M., 2023. An efficient machine learning approach for predicting concrete chloride resistance using a comprehensive dataset. Sci. Rep. 13, 15325.

Kairouz, P., McMahan, H.B., 2021. Advances and open problems in federated learning, found. Trends Mach. Learn. 14, 1–210.

Kamel, A., Xu, S., Li, G., 2025. AI-based framework for concrete durability assessment using generative adversarial networks and Bayesian neural networks. Constr. Build. Mater. 464, 140160.

Kassir, M.K., Ghosn, M., 2002. Chloride-induced corrosion of reinforced concrete bridge decks. Cement Concr. Res. 32, 139–143.

Konečný, J., McMahan, H.B., Yu, F.X., Richtarik, P., Suresh, A.T., Bacon, D., 2016. Federated Learning: Strategies for Improving Communication Efficiency arXiv preprint arXiv:1610.05492.

Lei, X., Guo, H., Dong, Y., Bastidas-Arteaga, E., 2026. Life-cycle performance prediction and interpretation for coastal and marine prestressed concrete beams using active learning-enhanced Bayesian neural networks. Reliab. Eng. Syst. Saf. 265, 111557.

Liu, Q.F., Iqbal, M.F., Yang, J., Lu, X., Zhang, P., Rauf, M., 2021. Prediction of chloride diffusivity in concrete using artificial neural network: modelling and performance evaluation. Constr. Build. Mater. 268, 121082.

McMahan, H.B., Moore, E., Ramage, D., Hampson, S., Agüera y Arcas, B., 2017. Communication-efficient learning of deep networks from decentralized data. In: Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS). PMLR, pp. 1273–1282.

Ministry of Land, Infrastructure and Transport (MOLIT), 2021. KDS 14 20 40: Durability Design Standard for Concrete Structures. Sejong, Republic of Korea.

Pham, N.L., Ta, Q.B., Kim, J.T., 2025. CNN federated learning for vibration-based damage identification of submerged structure-foundation system. J. Civ. Struct. Health Monit.

Phan, D., Pradhan, N., Jankowiak, M., 2019. Composable Effects for Flexible and Accelerated Probabilistic Programming in Numpyro, 11554 arXiv:1912.

Polder, R., Andrade, C., Elsener, B., Vennesland, Ø., Gulikers, J., Weidert, R., Raupach, M., 2000. Test methods for on site measurement of resistivity of concrete. Mater. Struct. 33, 603–611.

Roh, G., Shim, C., Song, H., 2025. Inspection data-driven machine learning models for predicting the remaining service life of deteriorating bridge decks. Build 15, 2799.

Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Dennison, D., 2015. Hidden technical debt in machine learning systems. Adv. Neural Inf. Process. 28.

Song, H., Kim, K., Shin, J., Roh, G., Shim, C., 2025. Digital twin framework for bridge slab deterioration: from 2D inspection data to predictive 3D maintenance modeling. Build 15, 1979.

Stewart, M.G., Rosowsky, D.V., 1998. Structural safety and serviceability of concrete bridges subject to corrosion. J. Infrastruct. Syst. 4, 146–155.

Tran, V.Q., Nguyen, H.Q., 2023. Physics-informed deep neural network for modeling the chloride diffusion in concrete. Eng. Appl. Artif. Intell. 126, 107024.

Tuutti, K., 1982. Corrosion of Steel in Concrete. Swedish Cement and Concrete Research Institute (CBI), Stockholm, Sweden. Report 4:82.

Wang, J., Liu, Y., Li, S., 2025. Enhanced probabilistic prediction of pavement deterioration using Bayesian neural networks and cuckoo search optimization. Sci. Rep. 15, 92469.

Wang, Y., Tang, P., Liu, K., Cai, J., Ren, R., Lin, J.J., et al., 2023. Characterizing data sharing in civil infrastructure engineering: current practice, future vision, barriers, and promotion strategies. J. Comput. Civ. Eng. 37, 04023001.

Wu, L., Wang, W., Jiang, C., 2023. Deep learning-based prediction for time-dependent chloride penetration in concrete exposed to coastal environment. Heliyon 9, e16869.

Yang, Q., Liu, Y., Chen, T., Tong, Y., 2019. Federated machine learning: concept and applications. ACM Trans. Intell. Syst. Technol. 10, 12.

Yang, Y., Lin, H., Qian, G., Hu, Z., Todd, M.D., 2025. A privacy-preserving framework using federated learning for structural health monitoring with miter gates application. Struct. Health Monit.

Yurochkin, M., Agarwal, M., Ghosh, S., Greenewald, K., Hoang, N., Khazaeni, Y., 2019. Bayesian nonparametric federated learning of neural networks. In: International Conference on Machine Learning. PMLR, pp. 7252–7261.

Zhao, Y., Li, M., Lai, L., Suda, N., Civin, D., Chandra, V., 2018. Federated learning with non-iid data. arXiv preprint arXiv:1806.00582.