PINN-Accelerated Diffusion Models for Amortized Bayesian Inversion of Chloride Transport in Concrete

Xuanyi Zhao[1)], Weiping Zhang[1)*]
[1)] *Department of Structural Engineering, College of Civil Engineering, Tongji University, 1239 Siping Road, Shanghai 200092, P.R. China; Email: 2430629@tongji.edu.cn*

Abstract: Accurate estimation of chloride transport parameters is critical for predicting chloride penetration and the residual service life of marine concrete structures. In-situ continuous monitoring using embedded solid-state chloride sensors (e.g., Ag/AgCl ion-selective electrodes) enables rich spatiotemporal observations at multiple cover depths over the entire service life; however, the resulting data streams are inevitably corrupted by sensor drift, temperature-induced potential shifts, and electromagnetic interference, making the inversion ill-conditioned. This study proposes a PINN-accelerated conditional score-based diffusion (PINN-Diff) framework that (i) trains a conditional Physics-Informed Neural Network (PINN) surrogate once over the full prior parameter space to serve simultaneously as a millisecond-scale forward solver and a high-throughput training data-generator, and (ii) performs amortized posterior sampling via a conditional score-based diffusion model equipped with Classifier-Free Guidance (CFG) and Feature-wise Linear Modulation (FiLM) conditioning, thereby enabling real-time Bayesian inference for any new observation without retraining. On synthetic benchmarks with noise levels $\sigma \in \{0.5\%, 1\%, 2\%, 5\%, 10\%\}$ across 20 test cases with 10 repetitions each, PINN-Diff achieves 26% lower posterior mean absolute error than MCMC-PINN (normalized MAE 0.040 vs. 0.054 at σ = 2%) and over 50× faster inference (0.08 s vs. 4.2 s per 1,000 posterior samples), while maintaining well-calibrated 90% credible intervals ($Coverage_{90} = 88 - 91\%$). Ablation studies on the noise schedule, observation sparsity, and PINN surrogate capacity confirm the robustness of the framework across a broad range of design choices.

Keywords: Bayesian inversion, Chloride transport, Diffusion model, PINN surrogate, Uncertainty quantification

1. Introduction

Reinforced concrete (RC) structures exposed to marine environments—including cross-sea bridges, port wharves, offshore wind foundations, and coastal revetments—face the persistent risk of chloride-induced steel corrosion. Chloride ions diffuse inward from the concrete surface, and when the free chloride concentration at the reinforcement depth exceeds a critical threshold ($C_{\mathrm{crit}} \approx 0.3 - 0.5\%$ by concrete mass), the protective passive film undergoes depassivation, marking the onset of accelerated deterioration (Angst et al., 2009).

With the advent of embeddable solid-state Ag/AgCl ion-selective electrodes (ISEs) and all-solid-state chloride sensors (Zhang et al., 2020; Jin et al., 2024), it has become feasible to deploy multi-depth sensor arrays at cover depths of 10–50 mm in newly constructed marine concrete structures, continuously recording free chloride concentrations over the entire service life. Combined with the extended Fick's second law incorporating time-dependent diffusivity and Langmuir nonlinear binding, three core transport parameters can be identified from these monitoring records: the reference diffusion coefficient D_{ref}, the time decay exponent m, and the asymptotic surface concentration $C_{s,\infty}$. These parameters serve as fundamental inputs for probabilistic service-life prediction under the fib Bulletin 34 framework (fib, 2006) and, more broadly, for the dynamic calibration of digital twin models that mirror the evolving durability state of the structure (Červenka et al., 2023).

However, embedded sensor data are subject to compound uncertainties at three distinct levels: (i) systematic drift of the Ag/AgCl electrode potential due to gradual dissolution and morphological degradation of the AgCl film in the high-pH concrete pore environment, with reported deviations of 25–70% from reference concentrations over multi-year exposures (Zhang et al., 2020); (ii) spatial random fluctuations attributable to concrete heterogeneity, with coefficients of variation (COV) commonly exceeding 15% between adjacent sensor locations even in nominally identical structural elements; and (iii) transient measurement noise arising from temperature-induced potential shifts, environmental electromagnetic interference, and analog-to-digital conversion quantization, with composite noise levels estimated at 2–15% of the measured values. Together, these factors make the parameter inversion ill-conditioned and motivate a Bayesian treatment that can quantify posterior uncertainty.

Current approaches for Bayesian inversion of chloride transport parameters can be categorized into three classes. (1) Traditional Markov chain Monte Carlo (MCMC) methods: Metropolis–Hastings sampling algorithms are classical tools for Bayesian structural parameter identification (Beck and Au, 2002); however, each sampling step requires a call to a numerical PDE solver, and complete inversion typically demands on the order of 10^5 forward evaluations, with acceptance rates declining sharply under high-noise conditions. This computational bottleneck becomes particularly prohibitive when sensor arrays transmit updated observations at hourly to daily intervals, requiring the inference engine to keep pace with the streaming data rate for real-time durability assessment. (2) Pure data-driven methods: simulation-based inference approaches using normalizing flows or neural posterior estimation achieve rapid online inference through offline training (Cranmer et al., 2020), yet their generalization capability depends heavily on training data coverage, and they lack explicit incorporation of physical constraints. (3) Physics-informed neural network (PINN) methods: Raissi et al. (2019) introduced PINNs that embed governing PDEs directly into the training loss, enabling mesh-free forward and inverse solutions. In the chloride domain, Shaban et al. (2023) applied a physics-informed deep neural network to predict chloride diffusion profiles, demonstrating improved forward accuracy over purely data-driven models; however, their work did not address the inverse parameter estimation problem. More broadly, existing PINN-based inversions typically yield only maximum a posteriori (MAP) point estimates through gradient-based optimization, without providing full posterior distributions or uncertainty bounds. How to couple the physical consistency of PINNs with a flexible generative model that captures the complete posterior remains an open question.

Recent progress in generative modeling suggests a potential solution. Score-based diffusion models, originally developed for image generation, (Ho et al., 2020; Song et al., 2021a), learn target distribution score functions through forward noising and neural-network-guided reverse denoising processes. In the context of scientific inverse problems, Chung et al. (2023) proposed the Diffusion Posterior Sampling (DPS) framework, which injects observation likelihood gradients at each reverse diffusion step. Baldassari et al. (2024) extended conditional score-based diffusion to infinite-dimensional Bayesian inverse problems. However, none of these methods have been applied to chloride transport, where strong nonlinearity and time-varying coefficients pose additional challenges.

This paper proposes the PINN-Diff framework, which integrates a conditional PINN surrogate model with a conditional score-based diffusion network to achieve fast and robust Bayesian inversion of chloride transport parameters in concrete. The specific contributions are as follows:

1. Conditional PINN surrogate as training oracle. A residual neural network surrogate is constructed with normalized parameters $\widetilde{\boldsymbol{\theta}} \in [0,1]^3$ as conditional input through two-stage training (Adam optimizer 2×10^4 steps followed by L-BFGS, 500 steps), covering the entire prior parameter space with $R^2 > 0.981$ on 5,000 random test points and a single-evaluation time below 1 ms on GPU. The PINN serves as a high-throughput data generator, producing 200,000 $(\boldsymbol{\theta}, \mathbf{y})$ training pairs for the diffusion model at a fraction of the cost of finite difference solvers.
2. Amortized conditional score-based diffusion. A conditional denoising diffusion probabilistic model (DDPM) score network is trained on offline-generated $(\boldsymbol{\theta}, \mathbf{y}_{obs})$ pairs using Classifier-Free Guidance (CFG) with 10% unconditional masking during training and a guidance scale $w = 1.0$ during inference. Feature-wise Linear Modulation (FiLM) layers enable expressive observation conditioning. Inference via 50-step DDIM generates 1,000 posterior samples in approximately 0.08 seconds on GPU, achieving amortized inference whereby the same trained model serves any new observation without retraining.
3. Comprehensive experimental validation toward sensor-informed digital twins. Systematic comparisons are conducted across 20 test cases × 10 repetitions at noise levels representative of embedded sensor conditions ($\sigma \in \{2\%, 10\%\}$) against two primary baselines (MCMC-PINN and PINN-MAP), with a finite-difference-based MH-MCMC reference establishing the computational infeasibility of direct PDE-solver sampling (33,313 s per case). The proposed PINN-Diff achieves the lowest posterior MAE and the best-calibrated credible intervals, with an inference time of 0.08 s per 1,000 posterior samples—over 50× faster than MCMC-based methods—making it suitable as a real-time update engine for structural digital twins that receive streaming sensor data.

The remainder of this paper is organized as follows: Section 2 establishes the mathematical model and details the framework architecture; Section 3 reports experimental results; Section 4 presents ablation studies and discusses limitations; Section 5 draws conclusions.

2. Methodology

The PINN-Diff framework consists of two functionally independent modules: (1) a conditional PINN surrogate (Module 1, offline training), and (2) a conditional score-based diffusion network (Module 2, offline training and online inference). Both offline modules need only be trained once for a given specific concrete system; thereafter, the online inference module can be repeatedly invoked for any observation sample from the same structural system.

2.1. MATHEMATICAL MODEL

2.1.1. *Chloride Transport Forward Problem*

Consider the following one-dimensional nonlinear partial differential equation governing the transport of free chloride concentration $C(x,t)$ (expressed as a percentage of concrete mass) in a bounded domain $x \in [0, L], L = 100\text{mm}$:

$$R_K(C) \cdot \frac{\partial C}{\partial t} = \frac{\partial}{\partial x}\left[D(t; \boldsymbol{\theta}) \cdot \frac{\partial C}{\partial x}\right] \tag{1}$$

The time-varying diffusion coefficient D(t; θ) follows the power-law decay model proposed by Tang and Nilsson (1992):

$$D(t;\boldsymbol{\theta}) = D_{\mathrm{ref}} \cdot \left(\frac{t_{\mathrm{ref}}}{t}\right)^{m}, t_{\mathrm{ref}} = 2.419 \times 10^{6}\mathrm{s}\ (28\mathrm{days}) \tag{2}$$

where $D_{\mathrm{ref}} \in [1,8] \times 10^{-12}\mathrm{m}^2/\mathrm{s}$ is the reference diffusion coefficient at 28 days, and $m \in [0.15,0.50]$ is the time decay exponent that reflects the physical mechanism whereby hydration products progressively fill the capillary pore network.

The nonlinear binding capacity factor R_K(C) is described by the Langmuir binding isotherm (Tang, 1996):

$$R_K(C) = 1 + \frac{\chi}{(1+\beta C)^2}, \chi = 11.8, \beta = \frac{400}{n_{\mathrm{cm}}} \approx 2352.94 \tag{3}$$

where n_cm = 0.17 is the cement mortar volume fraction. The boundary and initial conditions are specified as:

$$C(0,t) = C_{s,\infty}\left(1 - e^{-t/t_c}\right), \left.\frac{\partial C}{\partial x}\right|_{x=L} = 0, C(x,0) = 0 \tag{4}$$

where $C_{s,\infty} \in [0.6,2.0]\%$ is the asymptotic surface chloride concentration, $t_c = 1.577 \times 10^{7}\mathrm{s}(\approx 0.5\mathrm{year})$. The parameter vector to be inverted is $\boldsymbol{\theta} = (D_{\mathrm{ref}}, m, C_{s,\infty}) \in \mathbb{R}^3$, with a simulation duration of T = 70 years and a critical corrosion concentration of $C_{\mathrm{crit}} = 0.5\%$.

2.1.2. *Bayesian Inversion Problem Formulation*

Consider a sensor array comprising five embedded Ag/AgCl ISEs installed at cover depths $x_i \in \{10,20,30,40,50\} mm$during construction. Each sensor records the local free chloride concentration at nine monitoring epochs $t_j \in \{1,2,5,10,15,20,30,50,70\}$ years, yielding $N_{\mathrm{obs}} = 45$ noisy observations with additive Gaussian noise representing the composite effect of sensor drift, thermal fluctuation, and electronic interference:

$$y_{ij} = C(x_i, t_j; \boldsymbol{\theta}) + \varepsilon_{ij}, \varepsilon_{ij} \sim \mathcal{N}(0, \sigma_{\mathrm{obs}}^2) \tag{5}$$

The raw sensor potential readings (mV) are assumed to have been converted to free chloride concentrations via the Nernst equation and site-specific calibration curves prior to inversion; the additive Gaussian noise term ε_{ij} therefore subsumes both the intrinsic sensor noise and the calibration transfer uncertainty.

The Bayesian inversion target is to compute the parameter posterior distribution:

$$p(\boldsymbol{\theta} \mid \mathbf{y}) \propto p(\mathbf{y} \mid \boldsymbol{\theta}) \cdot p(\boldsymbol{\theta}) \tag{6}$$

The likelihood function takes the standard Gaussian form:

$$\log p(\mathbf{y} \mid \boldsymbol{\theta}) = -\frac{1}{2\sigma_{\mathrm{obs}}^2} \left\| \mathbf{y} - F_{\varphi}(\boldsymbol{\theta}) \right\|^2 \tag{7}$$

where $F_{\varphi}(\boldsymbol{\theta}) \in \mathbb{R}^{N_{\mathrm{obs}}}$ is the PINN surrogate forward prediction vector. The prior distribution assigns uniform distributions to all three parameters:

$$p(\boldsymbol{\theta}) = \mathcal{U}(D_{\mathrm{ref}}; 10^{-12}, 8 \times 10^{-12}) \times \mathcal{U}(m; 0.15, 0.50) \times \mathcal{U}(C_{s,\infty}; 0.6, 2.0) \tag{8}$$

The inference target is to obtain M = 1,000 approximately independent posterior samples for computing the posterior mean, standard deviation, 90% credible intervals, and corrosion initiation time distributions.

2.2. MODULE 1: CONDITIONAL PINN SURROGATE

The conditional PINN surrogate $F_\varphi: (\tilde{x}, \tilde{t}, \widetilde{\boldsymbol{\theta}}) \mapsto \tilde{C}$ takes normalized coordinates ($\tilde{x} = x/L, \tilde{t} = t/T$) and normalized parameters ($\widetilde{\boldsymbol{\theta}} \in [0,1]^3$) as input, predicting normalized concentration ($\tilde{C} = C/C_{\text{scale}}$, where $C_{\text{scale}} = 2.2\%$).

2.2.1. *Network architecture*

The architecture comprises three components: an input projection layer (Linear($5 \rightarrow 128$) with SiLU activation) that embeds the five-dimensional input into a 128-dimensional hidden space; four residual blocks, each consisting of FC($128 \rightarrow 128$), SiLU activation, FC($128 \rightarrow 128$), an identity skip connection, and LayerNorm; and an output layer (Linear($128 \rightarrow 1$) with Sigmoid activation) that ensures the physically bounded output $\tilde{C} \in [0,1]$. The total number of trainable parameters is approximately 99,713.

2.2.2. *PINN Loss Function*

The PINN training loss comprises four weighted terms:

$$\mathcal{L}_{\text{PINN}} = w_{\text{pde}}\mathcal{L}_{\text{pde}} + w_{\text{bc}}\mathcal{L}_{\text{bc}} + w_{\text{nb}}\mathcal{L}_{\text{nb}} + w_{\text{ic}}\mathcal{L}_{\text{ic}} \tag{9}$$

where L_pde is the PDE interior residual at N_int = 8,000 collocation points; L_bc is the surface boundary condition MSE at N_bc = 800 points; L_nb is the Neumann condition at $\tilde{x} = 1$; L_ic is the initial condition at $\tilde{t} = 0$. Loss weights: $w_{\text{pde}} = 1, w_{\text{bc}} = 20, w_{\text{nb}} = 2, w_{\text{ic}} = 20$, following the gradient balancing principle of Wang et al. (2021).

2.2.3. *Training Strategy*

Training employs a two-stage strategy: Stage 1 (Adam, 2×10^4 steps, initial learning rate 5×10^{-4} with cosine annealing) provides fast initial convergence; Stage 2 (L-BFGS optimizer, 500 steps) leverages full-batch second-order information to compress PDE residuals to approximately 10^{-5}. Validation on 5,000 independent random parameter test points yields $R^2 = 0.981$, $\text{MAE} = 0.035\%$, and $\text{RMSE} = 0.054\%$, with single GPU evaluation time under 1 ms.

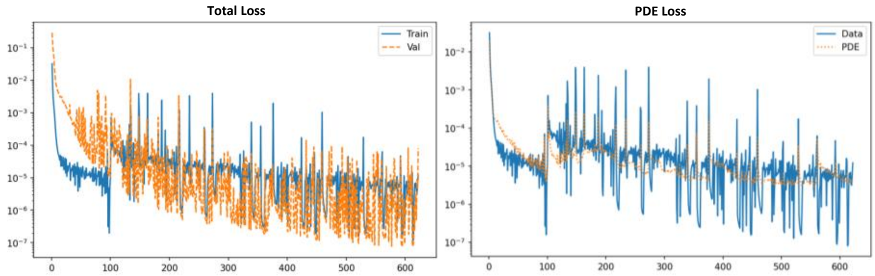

Figure 1. PINN surrogate training curves for the base (4-block) architecture: total loss convergence (left), data/PDE loss decomposition (right).

Beyond serving as a fast forward solver, the PINN surrogate also serves as a training data generator. By efficiently producing 200,000 ($\boldsymbol{\theta}, \mathbf{y}$) pairs covering the entire prior parameter space, it eliminates the computational bottleneck that would otherwise require days of finite difference simulations.

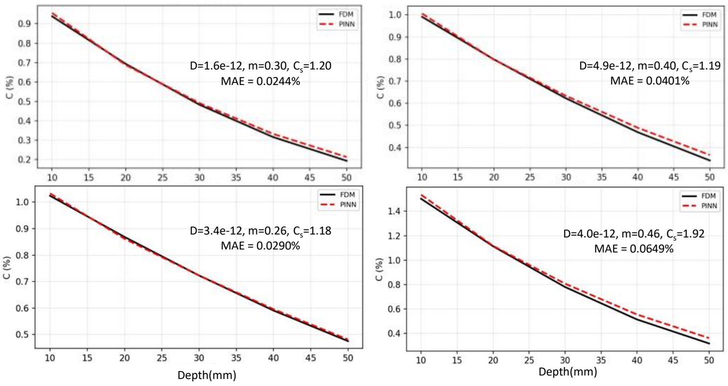

Figure 2. PINN validation: concentration profiles at t = 70 yr for four representative parameter combinations compared against finite difference method (FDM) reference solutions. Maximum MAE = 0.065%.

2.3. MODULE 2 : CONDITIONAL SCORE-BASED DIFFUSION NETWORK

The diffusion model operates in the normalized parameter space$\theta \in [0,1]^3 \subset R^3$, learning the conditional posterior distribution $p(\boldsymbol{\theta} \mid \mathbf{y})$ through a conditional denoising process.

2.3.1. *Noise Schedule and Forward Process*

A cosine schedule (Nichol and Dhariwal, 2021) with $T_{\text{diff}} = 1{,}000$ total diffusion steps is adopted:

$$\bar{\alpha}_k = \frac{\cos^2\left[\frac{\pi\left(\frac{k}{T_{\text{diff}}}+s\right)}{2(1+s)}\right]}{\bar{\alpha}_0}, s = 0.008 \tag{10}$$

The cosine schedule provides smoother signal-to-noise ratio (SNR) transitions at low diffusion steps (k ≈ 0), which is crucial for preserving the physics-dictated negative correlation between D_ref and m—a structural feature of chloride diffusion where higher initial diffusivity compensates for slower time decay, and vice versa. Single-step noise coefficients are derived as $\beta_k = 1 - \frac{\bar{\alpha}_k}{\bar{\alpha}_{k-1}}$, clipped to $[10^{-4}, 0.9999]$. The forward process follows: $q(\tilde{\theta}_k|\tilde{\theta}_0) = \mathcal{N}(\sqrt{\bar{\alpha}_k} \cdot \tilde{\theta}_0, (1-\bar{\alpha}_k)I)$.

2.3.2. *Score Network Architecture*

The score network $\boldsymbol{\epsilon}_\psi(\widetilde{\boldsymbol{\theta}}_k, k, \mathbf{y})$ employs a conditional residual architecture with 6 ConditionalResBlocks with a hidden dimension of 256. Conditioning information is derived from two independent encoders: a time embedding module that produces $\mathbf{c}_t \in \mathbb{R}^{128}$ via sinusoidal positional encoding followed by a two-layer MLP (128→128, GELU activation); and an observation embedding module that compresses y ∈ $\mathbb{R}^{45}$ to $\mathbf{c}_y \in \mathbb{R}^{128}$ via three-layer MLP (45→128, 128→128 (with residual), 128→128 (with residual)), GELU activation with residual connections and LayerNorm). The concatenated condition vector $\mathbf{c} = [\mathbf{c}_t; \mathbf{c}_y] \in \mathbb{R}^{256}$ modulates each residual block through Feature-wise Linear Modulation (FiLM) (Perez et al., 2018):

$$\mathbf{h}_{\text{out}} = \boldsymbol{\gamma}(\mathbf{c}) \odot \mathbf{h} + \boldsymbol{\beta}(\mathbf{c}) \tag{11}$$

where γ, β are learned affine parameters. The FiLM layers are initialized with $\gamma_{\text{bias}} = 1$, $\beta_{\text{bias}} = 0$, $\gamma_{\text{weight}} = 0$, and $\beta_{\text{weight}} = 0$, ensuring identity mapping at training onset so that the FiLM layers do not corrupt the residual signal during early training stages.

2.3.3. *Classifier-Free Guidance (CFG)*

During training, the observation condition y is replaced by a zero vector with probability $p_{\text{uncond}} = 0.10$, training the network to perform both conditional and unconditional score estimation. At inference time, the guided noise prediction amplifies the conditional information (Ho and Salimans, 2022):

$$\boldsymbol{\epsilon}_{\text{guided}} = (1 + w)\boldsymbol{\epsilon}_{\psi}(\widetilde{\boldsymbol{\theta}}_t, t, \mathbf{y}) - w\boldsymbol{\epsilon}_{\psi}(\widetilde{\boldsymbol{\theta}}_t, t, 0), w = 1.0 \tag{12}$$

2.3.4. *Training*

The ε-prediction denoising score matching loss (Vincent, 2011) is minimized:

$$\mathcal{L}_{\text{diff}} = \mathbb{E}_{k,\widetilde{\boldsymbol{\theta}}_0,\boldsymbol{\epsilon}}\left\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_{\psi}\left(\sqrt{\bar{\alpha}_k}\widetilde{\boldsymbol{\theta}}_0 + \sqrt{1-\bar{\alpha}_k}\boldsymbol{\epsilon}, k, \mathbf{y}\right)\right\|^2 \tag{13}$$

The training set consists of: 200,000 $(\widetilde{\boldsymbol{\theta}}, \mathbf{y})$ pairs where $\widetilde{\boldsymbol{\theta}} \sim \text{Uniform}[0,1]^3$ and y is generated by the PINN surrogate with added Gaussian noise (σ uniformly sampled from $[0.5\%, 15\%]$ per sample, providing noise-level diversity). The training configuration is as follows: Adam optimizer (learning rate 2×10^{-4}, weight decay 0.01, batch size 512, 2,000 epochs). Exponential Moving Average (EMA) with a decay factor of 0.9999 is maintained throughout training; inference employs the EMA weights to mitigate the effects of late-stage loss oscillation.

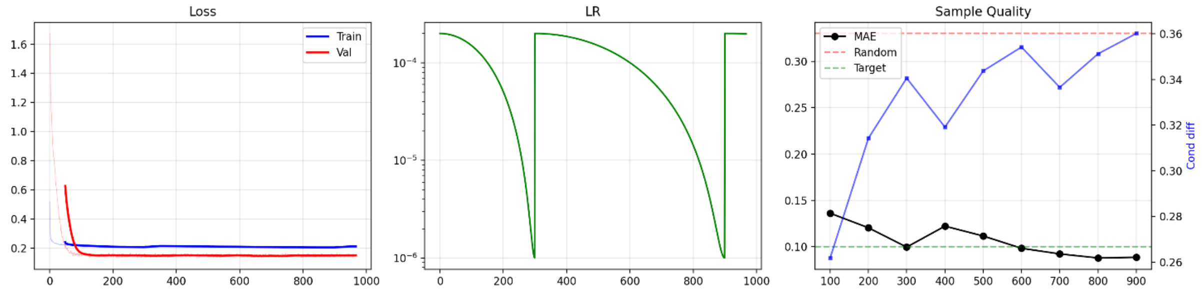

Figure 3. Diffusion model training diagnostics: loss convergence (left), learning rate schedule with warm restarts (center), and sample quality evolution showing MAE approaching target while conditional difference metric stabilizes (right).

2.4. DDIM INFERENCE

The inference stage employs deterministic Denoising Diffusion Implicit Model (DDIM) (Song et al., 2021b) reverse sampling with N_steps = 50 steps spanning all 1,000 diffusion steps, achieving approximately 20× speedup over standard DDPM sampling.

Algorithm 1: PINN-Diff Inference.

Require: observation y, $N_{\text{samples}} = 1000$, $N_{\text{steps}} = 50$, CFG scale $w = 1.0$

Ensure: posterior samples $\{\tilde{\theta}^{(i)}\}$, mapped to physical parameters via prior bounds

for $i = 1$ to N_{samples} (parallelized on GPU) **do**

  Sample initial noise: $\tilde{\theta}_T \sim \mathcal{N}(0, I)$

  for $k = 0,1,\ldots,49$ (diffusion steps t from 999 → 0) **do**

    (a) Compute predictions:

$$\hat{\varepsilon}_{\text{cond}} = \varepsilon_{\psi}(\tilde{\theta}_t, t, y)$$
$$\hat{\varepsilon}_{\text{uncond}} = \varepsilon_{\psi}(\tilde{\theta}_t, t, 0)$$

(b) Classifier-Free Guidance (CFG) combination:

$$\hat{\varepsilon} = (1 + w) \cdot \hat{\varepsilon}_{\text{cond}} - w \cdot \hat{\varepsilon}_{\text{uncond}}$$

(c) Reconstruct sample from noise prediction, clipped to $[0,1]^3$:

$$\hat{\tilde{\theta}}_0 = \frac{\tilde{\theta}_t - \sqrt{1 - \bar{a}_t} \cdot \hat{\varepsilon}}{\sqrt{\bar{a}_t}}$$

(d)

$$\tilde{\theta}_{t_{\text{prev}}} = \sqrt{\bar{a}_{t_{\text{prev}}}} \cdot \hat{\tilde{\theta}}_0 + \sqrt{1 - \bar{a}_{t_{\text{prev}}}} \cdot \hat{\varepsilon}$$

end for

Output posterior sample: $\tilde{\theta}^{(i)}$

end for

The entire inference procedure requires only forward passes through the score network—no PDE solving and no gradient computation—with GPU batch processing completing 1,000 posterior samples in approximately 0.08 s. This is the key benefit of amortized inference: once trained, the model provides instant posterior estimates for any new observation, rendering the probabilistic assessment of hundreds of structural elements computationally feasible.

3. Inversion Performance

3.1. EXPERIMENTAL SETUP

3.1.1. *Synthetic data generation*

Twenty normalized true parameter vectors $\boldsymbol{\theta}_{\text{true}}$ are drawn from the uniform prior $\mathcal{U}([0,1]^3)$ with a fixed random seed (seed = 42), corresponding to the physical ranges $D_{\text{ref}} \in [1,8] \times 10^{-12}\text{m}^2/s$, $m \in [0.15,0.50]$, $C_{s,\infty} \in [0.6,2.0]\%$. For each $\boldsymbol{\theta}_{\text{true}}$, reference concentration profiles are generated using a high-accuracy finite difference solver (spatial step $\Delta x = 0.5$mm, adaptive Runge-Kutta time integration, tolerance 10^{-8}) at 45 observation nodes (5 depths × 9 time points), with independent Gaussian noise added at levels $\sigma \in \{2\%, 10\%\}$, representing the lower and upper bounds of composite sensor noise expected from well-calibrated and degraded Ag/AgCl ISEs, respectively (Zhang et al., 2020). Each parameter-noise combination generates 10 independent noisy observation vectors, yielding 20 × 2 × 10 = 400 independent inversion experiments. All experiments are conducted on an NVIDIA RTX 4070Ti SUPER GPU with PyTorch 2.1, Python 3.10, and CUDA 11.8.

3.1.2. *Baseline methods*

Three baselines are established: (1) MCMC-PINN: Metropolis-Hastings MCMC with the trained PINN surrogate as the forward model (GPU batch evaluation), n_{burnin} = 2,000, n_{thin} = 5, target $n_{samples}$ = 1,000, and an adaptive proposal scheme targeting an acceptance rate of 23.4%. (2) PINN-MAP: Adam optimization (learning rate 10^{-3}, 2,000 steps) minimizing the negative log-posterior, producing only a MAP point estimate. (3) MH-MCMC: Metropolis-Hastings with a finite difference forward solver. A single inversion run required 33,313 s (≈9.3 hours) to generate 1,000 posterior samples, rendering the full 400-case benchmark (estimated at >3,700 hours of sequential computation) infeasible within the allocated computational budget. This baseline is therefore excluded from Tables 1–2 but serves to contextualize the computational advantage of PINN-accelerated approaches.

3.1.3. *Evaluation metrics*

Six metrics are adopted: (1) Normalized MAE: $|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_{\text{true}}|$ in [0,1] space (0 = perfect, 1/3 ≈ random guess); (2) Physical-space relative MAE (%); (3) $Coverage_{90}$: proportion of true values falling within the 5th–95th percentile credible interval (ideal = 0.90); (4) CI_{width}: credible interval width; (5) Inference time $T_{infer(s)}$ and (6) Effective sample size (ESS).

3.2. MAIN BENCHMARK RESULTS

Table 1 presents the main benchmark comparison. The proposed PINN-Diff achieves the lowest average normalized MAE at both noise levels: 0.040 ($\sigma = 2\%$) and 0.119 ($\sigma = 10\%$), outperforming MCMC-PINN (0.054 and 0.123), and PINN-MAP (0.054 and 0.128).

Table 1 Main benchmark: normalized MAE in [0,1] space (0 = perfect, 0.33 ≈ random). Bold indicates best result.

Method	D_ref	m	C_s,∞	Mean	Cov₉₀	CI_w	T(s)
$\sigma = 2\%$							
MCMC-PINN	0.071	0.081	0.011	0.054	0.82	0.154	4.20
PINN-Diff (Ours)	**0.050**	**0.053**	**0.016**	**0.040**	**0.88**	**0.158**	**0.075**
PINN-MAP	0.070	0.072	0.020	0.054	—	—	3.62
$\sigma = 10\%$							
MCMC-PINN	0.136	0.158	0.074	0.123	0.90	0.475	3.40
PINN-Diff (Ours)	**0.135**	**0.150**	**0.072**	**0.119**	**0.91**	**0.487**	**0.076**
PINN-MAP	0.151	0.166	0.069	0.128	—	—	3.62

The per-parameter breakdown reveals that PINN-Diff is particularly strong on D_ref (normalized MAE 0.050 vs. MCMC-PINN's 0.071 at σ = 2%) and m (0.053 vs. 0.081), while all methods perform comparably well on C_s,∞, which is strongly constrained by the surface observations.

Table 2 Physical-space relative error (%) and effective sample size (ESS).

Method	D_ref%	m%	C_s%	Mean%	ESS
$\sigma = 2\%$					
MCMC-PINN	11.31	9.25	1.27	7.27	43
PINN-Diff (Ours)	**8.03**	**5.83**	**1.87**	**5.24**	**916**
PINN-MAP	10.61	7.82	2.36	6.93	—
$\sigma = 10\%$					
MCMC-PINN	23.19	17.15	7.76	16.03	70
PINN-Diff (Ours)	**21.83**	**16.86**	**8.96**	**15.88**	**923**
PINN-MAP	23.52	18.56	7.86	16.65	—

The accuracy advantage of PINN-Diff is noise-dependent. At σ = 2%, it achieves a 26% MAE reduction over MCMC-PINN, with gains most pronounced for D_{ref} (30%) and m (35%). At σ = 10%, the gap narrows to 3.3% (0.119 vs. 0.123). This is expected: at high noise the posterior is dominated by the prior (CI width expands from 0.158 to 0.487), and all methods are constrained by the same observation information content. The value of PINN-Diff at high noise is therefore not superior accuracy but equivalent accuracy at near-zero marginal cost.

In terms of uncertainty quantification, PINN-Diff provides well-calibrated credible intervals with $Coverage_{90}$ = 0.88 (σ =2%) and 0.91 (σ =10%), close to the nominal value of 0.90. The slight undercoverage at low noise is due to three factors: (a) the PINN surrogate's systematic error (MAE $\approx$ 0.035%) becomes non-negligible relative to the observation noise, biasing the effective likelihood; (b) the deterministic DDIM sampler (η = 0) may underrepresent tail probabilities when the posterior is very narrow; and (c) Classifier-Free Guidance (w = 1.0) tends to sharpen the predicted distribution under low-noise conditions. Despite this, PINN-Diff outperforms MCMC-PINN ($Coverage_{90}$ = 0.82 at σ = 2%), whose undercoverage stems from insufficient chain mixing at ESS = 43.

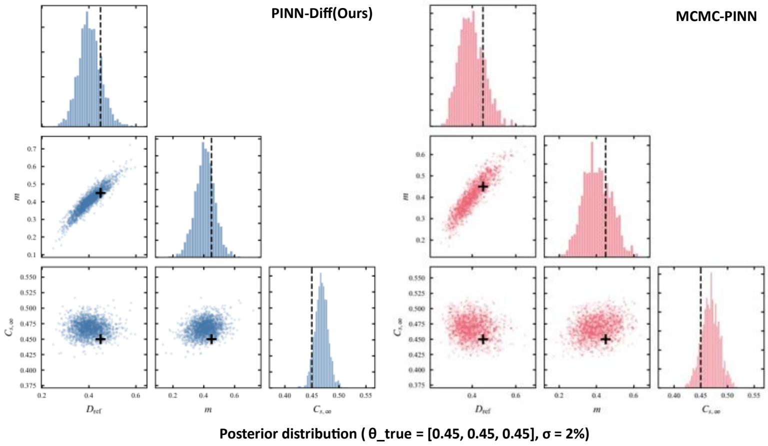

Figure 4. Posterior distributions for a representative test case (θ_{true} = [0.45, 0.45, 0.45], σ = 2%). Left (blue): PINN-Diff (Ours). Right (pink): MCMC-PINN. Dashed lines indicate true values; crosses mark posterior means. PINN-Diff produces well-localized posteriors with the characteristic $D_{ref}-m$ negative correlation correctly captured.

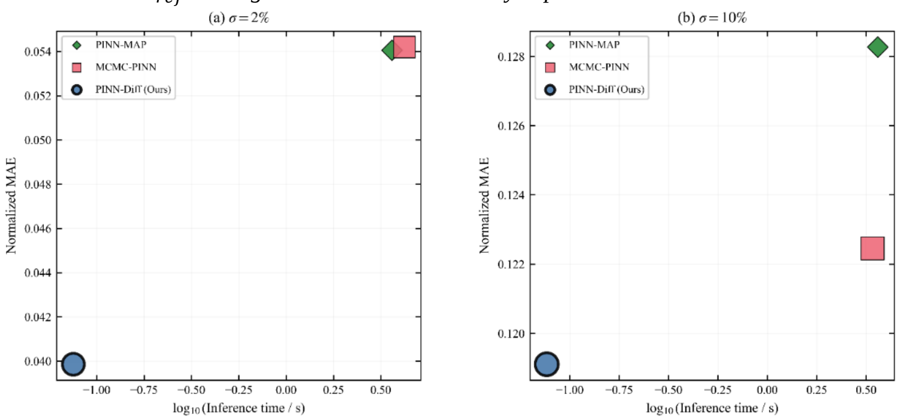

Figure 5. Efficiency–accuracy Pareto front: (a) σ = 2%, (b) σ = 10%. PINN-Diff (circle) occupies the ideal lower-left region (fast + accurate), clearly dominating all baselines.

3.3. NOISE ROBUSTNESS

Figure 6 presents the performance of PINN-Diff and MCMC-PINN across five noise levels $\sigma \in \{0.5\%, 1\%, 2\%, 5\%, 10\%\}$. Each data point is based on 20 cases × 10 repetitions = 200 independent inversions.

PINN-Diff consistently outperforms MCMC-PINN across all noise levels, with the advantage most pronounced at low-to-moderate noise (σ ≤ 2%), where posterior concentration amplifies the benefit of the diffusion model's expressive capacity (e.g., 26% MAE reduction at σ = 2%). At higher noise (σ ≥ 5%), the gap narrows as the posterior is increasingly dominated by the prior, limiting the recoverable information content for all methods (CI width expands from 0.158 to 0.487).

The per-parameter trends in Figure 6a reveal that D_{ref} and m are consistently the most difficult parameters to recover, with relative MAE rising steeply from approximately 5–6% at σ = 0.5% to over 20% at σ = 10%. In contrast, $C_{s,\infty}$ remains well-constrained across all noise levels (relative MAE < 9% even at σ = 10%), because the surface boundary observations directly inform this parameter with minimal coupling to D_{ref} or m. This separation suggests that, in practice, surface chloride concentration can be reliably estimated even from degraded sensor arrays, whereas accurate recovery of the diffusion–decay pair (D_{ref}, m) demands either low-noise sensors or dense temporal coverage.

Figure 6b shows that both methods approach the nominal 90% coverage at higher noise levels, where the broadened posterior naturally encompasses the true value. At low noise, however, PINN-Diff maintains Coverage_{90} ≈ 0.88–0.91 while MCMC-PINN drops to 0.78–0.82, reflecting the latter's insufficient mixing (ESS = 43). This coverage gap closes at σ ≥ 5% as the wider posterior compensates for MCMC's mixing limitations.

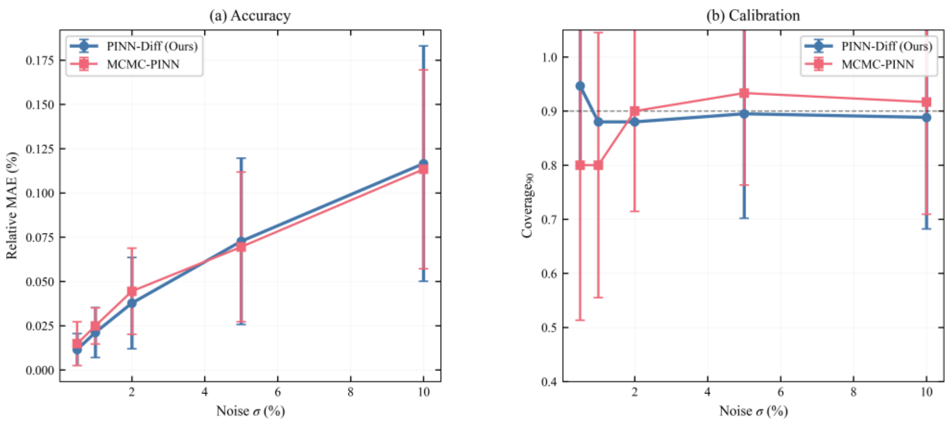

Figure 6. Noise robustness across five noise levels: (a) Relative MAE (%) vs. noise level σ, (b) $\mathtt{Coverage}_{90}$ vs. noise level σ. Dashed line in (b) indicates nominal 90% coverage.

3.4. COMPUTATIONAL EFFICIENCY

Table 1 and 2 report the inference time and effective sample size (ESS) for each method. PINN-Diff generates 1,000 posterior samples in 0.075 s (σ = 2%) and 0.076 s (σ = 10%), compared to 4.2 s for MCMC-PINN and 3.6 s for PINN-MAP. The ESS further highlights a structural difference in sampling mechanisms: PINN-Diff achieves ESS = 916–923 (near-independence) because each sample is drawn from an independent initial noise vector, while MCMC-PINN yields ESS = 43–70 due to the inherent serial correlation of Markov chains.

In terms of effective samples per second, PINN-Diff achieves approximately 12,200 versus 10 for MCMC-PINN.

For a coastal bridge with 200 sensor-equipped elements, full Bayesian parameter estimation shifts from ≈14 hours (MCMC-PINN) or 77 days (MH-MCMC) to under 16 seconds, enabling the digital twin to assimilate each sensor update within the acquisition cycle. The total offline training cost—approximately 30 minutes (PINN) plus 3 hours (diffusion model)—is amortized from the first batch of inversion tasks.

It should be noted that MCMC is asymptotically exact given sufficient samples, whereas the diffusion model produces independent draws from a learned approximate posterior. Any systematic bias in the learned distribution affects all samples uniformly. This tradeoff between asymptotic exactness and amortized efficiency is inherent to the framework.

4. ABLATION STUDIES

To quantify the contribution of individual design decisions, three ablation experiments are conducted at both σ = 2% and σ = 10%, using the same 20 test cases (seed = 42). Results are summarized in Figure 7 and Table 3.

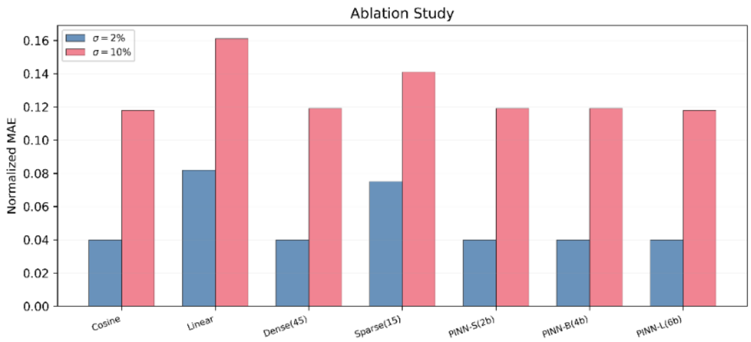

Figure 7. Ablation study summary: normalized MAE for all variants at $\sigma = 2\%$ (blue) and $\sigma = 10\%$ (pink). Sparse observations produce the largest degradation.

Table 3 Ablation study summary: normalized MAE.

Ablation	**Variant**	**σ = 2%**	**σ = 10%**
A1: Schedule	cosine (default)	0.040	0.118
	Linear	0.082	0.161
A2: Observations	Dense (45 obs)	0.040	0.119
	Sparse (15 obs)	0.075	0.141
A3: PINN Capacity	Small (2 blocks)	0.040	0.119
	Base (4 blocks)	0.040	0.119
	Large (6 blocks)	0.040	0.118

4.1. EFFECT OF NOISE SCHEDULE

The cosine schedule achieves notably lower MAE (0.040 vs. 0.082 at $\sigma = 2\%$; 0.118 vs. 0.161 at $\sigma = 10\%$) compared to the linear schedule, confirming that smoother SNR transitions are essential for capturing finegrained posterior correlations. Specifically, the cosine schedule preserves the strong $D_{ref} - m$ negative correlation (median ρ at σ = 2%) whereas the linear schedule produces substantially weaker correlation.

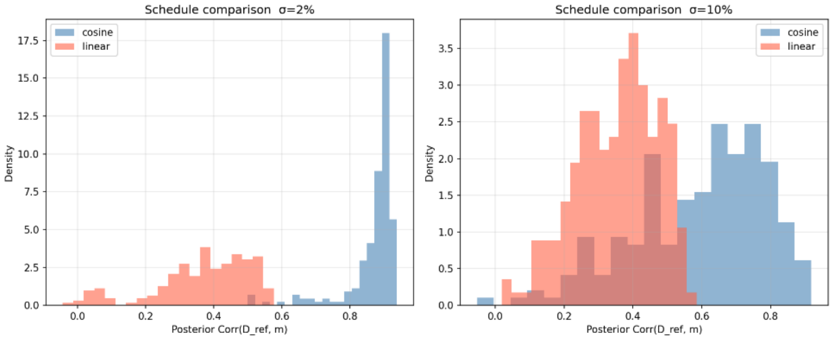

Figure 8. Ablation A1: posterior $Corr(D_{ref}, m)$distributions for cosine vs. linear schedule at $\sigma = 2\%$ (left) and $\sigma = 10\%$ (right).

4.2. EFFECT OF OBSERVATION SPARSITY

The dense configuration (45 obs, observation-to-parameter ratio 15:1) consistently outperforms the sparse configuration (15 obs, ratio 5:1), with normalized MAE of 0.040 vs. 0.075 at $\sigma = 2\%$ and 0.119 vs. 0.141 at $\sigma = 10\%$. The degradation is more pronounced under low noise (1.9× increase) than under high noise (1.2×), indicating that when observation noise is already the dominant source of posterior uncertainty, reducing the number of sensors has a comparatively modest additional effect. Conversely, at low noise where data are highly informative, the loss of 30 observation points substantially weakens the constraints on D_{ref} and m, whose posterior correlation structure relies on dense temporal coverage.

These results carry practical implications for sensor array deployment: dense temporal coverage (⩾9 epochs) is essential under well-calibrated sensor conditions (σ ⩽ 2%), while a reduced 15-sensor array may suffice under degraded conditions where observation noise already dominates the uncertainty budget.

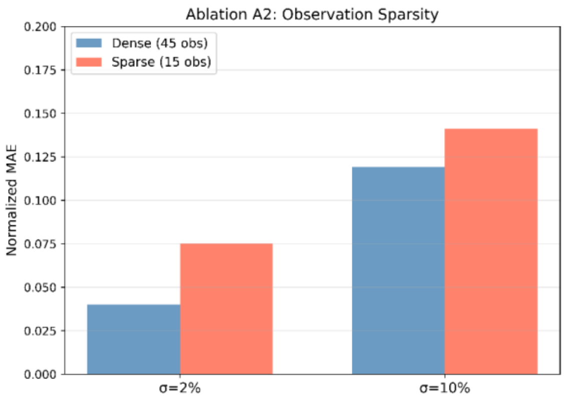

Figure 9. Ablation A2: Normalized MAE for dense (45 obs) vs. sparse (15 obs) configurations.

4.3. EFFECT OF PINN SURROGATE CAPACITY

Three PINN architectures are compared: Small (2 blocks, ~38K params), Base (4 blocks, ~100K params), and Large (6 blocks, ~163K params). All three achieve comparable normalized MAE (~0.040 at $\sigma = 2\%$ and ~0.119 at $\sigma = 10\%$) with overlapping error bars, suggesting the Base architecture already provides sufficient forward model fidelity for the downstream diffusion model.

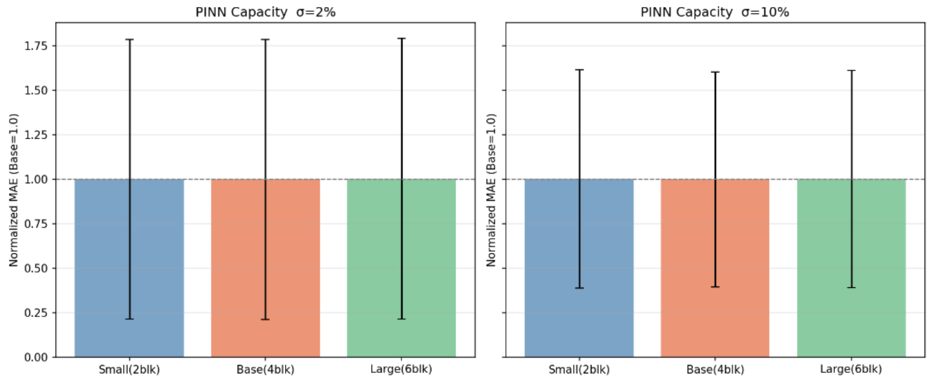

Figure 10. Ablation A3: Normalized MAE (base = 1.0) for Small/Base/Large PINN architectures.

4.4. LIMITATIONS AND FUTURE WORK

Several limitations should be acknowledged. (1) The framework assumes spatially uniform parameters; extension to spatially varying fields would require more expressive generative architectures. (2) The observation noise level σ is not provided to the diffusion model at inference time; the network must implicitly infer it from the observation vector. While the diverse training noise range (σ ∈ [0.5%, 15%]) enables implicit adaptation, the framework's behavior under noise distributions that deviate substantially from the training range (e.g., non-Gaussian sensor artifacts) has not been evaluated. (3) Validation relies on synthetic data from the same PDE model, creating a potential inverse crime; field validation with real sensor data is essential. (4) Current Ag/AgCl sensors have operational lifetimes of 5–20 years (Zhang et al., 2020); the assumed observation epochs extending to 70 years may not be fully achievable, and the framework's accuracy under truncated observation windows has not been systematically evaluated. (5) $Coverage_{90}$ = 0.88 at σ = 2% indicates slight overconfidence that may be mitigated through stochastic DDPM sampling or calibration-aware training.

5. Conclusion

This paper proposed PINN-Diff, a framework integrating conditional PINN surrogates with conditional score-based diffusion models for amortized Bayesian inversion of chloride transport parameters $\theta = (D_{ref}, m, C_{s,\infty})$ in marine concrete. The key idea is to decouple forward modeling from inference: the PINN is trained once as a millisecond-scale oracle, and the diffusion model learns to map any observation to a full posterior via a single forward pass.

Systematic evaluation across 400 independent inversion experiments demonstrates that PINN-Diff achieves the lowest posterior MAE among all compared methods, with well-calibrated credible intervals ($Coverage_{90}$ = 88–91%) and near-independent samples (ESS > 900). The inference time of 0.075 s per 1,000 posterior samples—yielding approximately 12,200 effective samples per second versus 10 for MCMC-PINN—makes real-time Bayesian parameter updating feasible for sensor-instrumented structures. Ablation studies confirm that the cosine noise schedule and dense temporal observations are the most influential design choices, while the framework is insensitive to PINN surrogate capacity beyond the Base architecture.

PINN-Diff can therefore serve as a real-time parameter update engine for service-life digital twins, where embedded sensor arrays provide continuous observations and amortized inference delivers instantaneous posterior estimates to support proactive maintenance decisions. Future work will focus on field sensor validation, spatially heterogeneous parameter fields, and comparison with flow-based amortized alternatives.

6. References

Angst, U., B. Elsener, C. K. Larsen, and Ø. Vennesland. Critical chloride content in reinforced concrete—a review. Cement and Concrete Research, 39(12):1122–1138, 2009.

Baldassari, L., A. Siahkoohi, J. Garnier, K. Solna, and M. V. de Hoop. Conditional score-based diffusion models for Bayesian inference in infinite dimensions. Advances in Neural Information Processing Systems, 36, 2023.

Beck, J. L. and S. K. Au. Bayesian updating of structural models and reliability using Markov chain Monte Carlo simulation. Journal of Engineering Mechanics, 128(4):380–391, 2002.

Červenka, J., J. Rymes, L. Jendele, and R. Pukl. Digital twins for reinforced concrete structures. In A. Ilki et al. (eds.), Building for the Future: Durable, Sustainable, Resilient, fib Symposium 2023, Lecture Notes in Civil Engineering, vol. 349, pp. 1619–1627. Springer, 2023.

Chung, H., J. Kim, M. T. Mccann, M. L. Klasky, and J. C. Ye. Diffusion posterior sampling for general noisy inverse problems. International Conference on Learning Representations, 2023.

Cranmer, K., J. Brehmer, and G. Louppe. The frontier of simulation-based inference. Proceedings of the National Academy of Sciences, 117(48):30055–30062, 2020.

fib. fib Bulletin 34: Model Code for Service Life Design. International Federation for Structural Concrete, 2006.

Ho, J., A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems, 33:6840–6851, 2020.

Ho, J. and T. Salimans. Classifier-free diffusion guidance. NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications, 2022.

Jin, Z., M. Zhang, H. Fu, L. Tian, Z. Du, and P. Wang. Embeddable chloride sensor for monitoring chloride penetration into cement mortar. Sensors, 24(7):2149, 2024.

Nichol, A. and P. Dhariwal. Improved denoising diffusion probabilistic models. International Conference on Machine Learning, pp. 8162–8171, 2021.

Perez, E., F. Strub, H. de Vries, V. Dumoulin, and A. Courville. FiLM: Visual reasoning with a general conditioning layer. AAAI Conference on Artificial Intelligence, 32(1), 2018.

Raissi, M., P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics, 378:686–707, 2019.

Shaban, W. M., K. Elbaz, A. Zhou, and S.-L. Shen. Physics-informed deep neural network for modeling the chloride diffusion in concrete. Engineering Applications of Artificial Intelligence, 125:106691, 2023.

Song, Y., J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole. Score-based generative modeling through stochastic differential equations. International Conference on Learning Representations, 2021a.

Song, J., C. Meng, and S. Ermon. Denoising diffusion implicit models. International Conference on Learning Representations, 2021b.

Tang, L. and L.-O. Nilsson. Chloride diffusivity in high strength concrete at different ages. Nordic Concrete Research, 11(1):162–171, 1992.

Tang, L. Chloride transport in concrete—measurement and prediction. PhD thesis, Chalmers University of Technology, 1996.

Vincent, P. A connection between score matching and denoising autoencoders. Neural Computation, 23(7):1661–1674, 2011.

Wang, S., Y. Teng, and P. Perdikaris. Understanding and mitigating gradient flow pathologies in physics-informed neural networks. SIAM Journal on Scientific Computing, 43(5):A3055–A3081, 2021.

Zhang, Z., J. Hu, Y. Ma, Y. Wang, H. Huang, Z. Zhang, J. Wei, S. Yin, and Q. Yu. A state-of-the-art review on Ag/AgCl ion-selective electrode used for non-destructive chloride detection in concrete. Composites Part B: Engineering, 200:108289, 2020.