

Article

Ensemble Extreme Learning Machines for Uncertainty Quantification in Ordinary Differential Equations

Sajad Ahmad Sheikh ¹  and Lateef Ahmad Wani ^{2,*} ¹ Department of Mathematics, University of Kashmir, South Campus, Anantnag 192101, Jammu and Kashmir, India; sajadsheikh@uok.edu.in² Department of Mathematics and Statistics, College of Science, King Faisal University, P.O. Box 400, Al-Ahsa 31982, Saudi Arabia

* Correspondence: lwani@kfu.edu.sa

Abstract

This paper proposes an uncertainty-quantified surrogate framework for ordinary differential equations (ODEs) based on ensembles of extreme learning machines (ELMs). The method constructs $M = 100$ independently randomized ELM surrogates for a given ODE trajectory, uses the ensemble mean as the predictor, and defines a pointwise ensemble spread as a preliminary uncertainty measure. To obtain statistically valid prediction intervals, a split-conformal calibration procedure is applied to the ensemble spread, yielding finite-sample marginal coverage under exchangeability while preserving computational efficiency, since each ELM is trained via a single ridge-regression solve. Theoretical results establish exact satisfaction of the prescribed initial condition, stability of the ensemble mean and variance with respect to perturbations in the training data, and almost-sure convergence of the empirical ensemble variance to the corresponding random-feature prediction variance. The hidden-layer sampling hypothesis is made explicit: the experiments use bounded hyperbolic-tangent features with independent uniform draws as the default and independent normal draws in sensitivity tests, both of which satisfy the finite-moment assumptions required by the convergence theorem. Comparisons with capacity-matched Bayesian random-feature neural surrogates and Monte Carlo dropout clarify differences in uncertainty representation. Numerical experiments on exponential, logistic, and damped oscillator dynamics demonstrate accurate reconstruction and calibrated uncertainty quantification in sparse and noisy regimes. Additional ablation studies quantify the effect of the denominator safeguard, calibration-sample size, ensemble size, training time, noise level, and hidden-parameter distribution.

Keywords: extreme learning machine; uncertainty quantification; ordinary differential equations; conformal prediction; Bayesian neural networks; Monte Carlo dropout; random features



Academic Editors: Nikola Anđelić and Vedran Mrzljak

Received: 23 June 2026

Revised: 8 July 2026

Accepted: 9 July 2026

Published: 14 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

MSC: 68T07; 65L05; 62G20; 34A12

1. Introduction

Uncertainty quantification (UQ) has become a central theme in scientific machine learning because accurate point predictions are insufficient when a learned surrogate is used for scientific inference, control, or forecasting. In differential-equation settings, the reliability of a learned trajectory is often as important as the trajectory itself, especially when data are sparse, noisy, or only partially informative. Recent work has therefore

emphasized predictive uncertainty, calibration, and model misspecification in scientific machine learning, physics-informed learning, and universal differential equations; see, for instance, the broad UQ framework of Psaros et al. [1], the comparative study of Schmid et al. [2], and the recent analysis of total uncertainty in physics-informed reconstruction by De Florio et al. [3]. At the same time, Bayesian neural networks, Monte Carlo dropout, and deep ensembles remain widely used tools for uncertainty-aware neural modeling [4–8].

Extreme learning machines (ELMs), introduced by Huang, Zhu, and Siew [9], provide a computationally economical random-feature paradigm. In a single-hidden-layer feed-forward network, the hidden parameters are sampled randomly and then frozen, while the output layer is learned analytically. This yields a training procedure that is much cheaper than iterative gradient-based optimization. The basic theory and regression properties of ELMs are well established [9,10], and randomized-feature versions of ELM ideas have recently been used successfully for numerical differential equations and scientific computing, including collocation for PDEs [11,12] and Bayesian physics-informed ELM variants for noisy PDE problems [13]. In parallel, Guignard, Amato, and Kanevski [14] developed a careful study of variability and interval estimation for ELM regression in generic supervised settings.

The present work differs from the general ELM regression uncertainty analysis of Guignard et al. [14] in four concrete ways. First, the target object here is an ODE trajectory represented through an initial-condition-preserving trial form, rather than a generic regression map. Second, the hidden-layer randomization is used as a random-feature ensemble for trajectory reconstruction, while the output weights are still obtained by closed-form ridge solves. Third, raw ensemble spread is treated only as a local nonconformity scale; calibrated coverage is supplied by a split-conformal step. Fourth, the theoretical statements are formulated around ODE trajectory bands, exact initial-condition enforcement, response stability of the ELM smoother, random-feature variance convergence, and pointwise marginal conformal coverage. These ingredients make the method closer to a calibrated numerical surrogate than to a general-purpose ELM regression interval.

The proposed ensemble also differs from deep ensembles [6]. A deep ensemble trains several independently initialized deep networks by iterative optimization and typically interprets the diversity of the trained networks as epistemic uncertainty. The present construction uses a single-hidden-layer random-feature architecture; only the output layer is trained, each member is obtained from one ridge solve, and the ensemble spread is passed through a distribution-free calibration layer. Thus the uncertainty band is tied to randomized-feature variability plus conformal calibration, rather than to optimizer-dependent diversity among separately trained deep models. This distinction is important in sparse ODE reconstruction, where a lightweight ensemble can be generated with little computational cost and then calibrated on held-out trajectory values.

The problem considered in this paper is novel in a precise sense. The existing ELM-UQ literature studies either generic supervised regression variability [14] or Bayesian output-layer formulations for PDEs [13], while ODE and scientific-machine-learning UQ studies are dominated by Bayesian, variational, dropout, and physics-informed neural approaches [1–3,15]. The combination developed here—initial-condition-preserving ELM trajectory reconstruction, independently randomized hidden layers, spread-normalized conformal scores, and a finite-sample marginal coverage theorem—appears to have been formalized only partially in the existing literature. It is this combination of computational speed, ODE-specific reconstruction, and calibrated uncertainty that motivates the present work.

The main methodological idea is intentionally simple. We represent the unknown ODE trajectory by a trial function that satisfies the initial condition exactly. We then fit $M = 100$ independently randomized ELMs to training data generated by an ODE solution or a trusted numerical solver. The ensemble mean becomes the point predictor. The empirical spread of the ensemble quantifies how much the random hidden-layer realization changes the reconstructed trajectory. Because raw ensemble spread alone does not guarantee calibrated uncertainty, we normalize calibration residuals by the ensemble spread and use split conformal prediction [16,17] to obtain finite-sample valid predictive bands under exchangeability. This construction is particularly attractive in ODE applications because the ELM members are fast to train and trivially parallelizable and the conformal adjustment is model-agnostic.

The contributions of the paper are fourfold. The first contribution is methodological: we formulate an ensemble ELM architecture for ODE trajectory reconstruction in which every member satisfies the initial condition exactly and is trained by a closed-form ridge solve. The second contribution is theoretical: we prove a linear smoother representation, derive stability bounds for both the ensemble mean and ensemble spread, establish almost-sure convergence of the empirical spread to the random-feature variance, and obtain a finite-sample marginal coverage theorem for a conformalized ensemble interval. The third contribution is comparative: we place the proposed interval next to Bayesian and dropout-based uncertainty mechanisms, proving that exact Bayesian predictive coverage requires a correctly specified probabilistic model, while Monte Carlo dropout intervals quantify mask randomness and require additional calibration or noise modeling for coverage in biased or noisy settings. The fourth contribution is applied: we carry out a reproducible ODE benchmark study based on sparse and noisy observations of exponential, logistic, and damped-oscillator dynamics, together with calibration, sharpness, ablation, noise-sensitivity, and timing diagnostics.

The paper is organized as follows. Section 2 introduces the ODE reconstruction setting, the data model, the hidden-layer sampling distribution, and the ensemble ELM architecture. Section 3 defines the conformalized spread-based interval and records the training and calibration procedure. Section 4 contains the main theoretical results and their proofs, including the comparison principles for Bayesian and dropout baselines. Section 5 reports the numerical experiments and discusses the observed calibration-sharpness trade-off, ablation results, noise sensitivity, and computational scaling. Section 6 concludes the paper.

2. Preliminaries

Consider a scalar initial-value problem

$$\begin{cases} u'(t) = f(t, u(t)), & t \in [0, T], \\ u(0) = u_0, \end{cases} \quad (1)$$

where $f : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ is continuous and the solution u is assumed unique on $[0, T]$. The scalar formulation is adopted for clarity of proofs; vector-valued systems are handled componentwise or by block-design matrices, and the same arguments remain valid with only notational modifications.

We assume that one has access to reference trajectory values at design times $0 = t_1 < t_2 < \dots < t_n \leq T$. These values may come from a trusted numerical solver, a high-fidelity simulator, or sufficiently accurate measurements. Denote the response vector by

$$\mathbf{y} = (y_1, \dots, y_n)^\top, \quad y_i \approx u(t_i),$$

with the punctuation included here and below to make the displayed formulae part of the surrounding sentences. When noisy observations are used, one may write $y_i = u(t_i) + \varepsilon_i$ with mean-zero noise ε_i . Since the focus of the present paper is confidence bands for the reconstructed trajectory itself, the clean-state setting is emphasized in the theory; noisy-response coverage can be handled by adding an independent noise term or by calibrating against observed responses instead of latent states.

The defining feature of the surrogate is that the initial condition is enforced exactly through the trial form

$$\hat{u}(t) = u_0 + t g(t), \quad (2)$$

where g is represented by a single-hidden-layer randomized neural network. Thus, regardless of how g is fitted, one always has $\hat{u}(0) = u_0$.

For a hidden width $m \in \mathbb{N}$ and an activation function $\sigma : \mathbb{R} \rightarrow \mathbb{R}$, define a random feature map

$$\phi_\omega(t) = (\sigma(a_1 t + b_1), \dots, \sigma(a_m t + b_m))^\top,$$

where $\omega = ((a_1, b_1), \dots, (a_m, b_m))$ is sampled from a distribution Π on $\mathbb{R}^{2m}$. The k th ELM member uses an independent draw ω_k and models

$$\hat{u}_k(t) = u_0 + h_k(t)^\top \hat{\beta}_k, \quad h_k(t) = t \phi_{\omega_k}(t) \in \mathbb{R}^m. \quad (3)$$

In all numerical experiments the physical time is first scaled to $[0, 1]$ before random features are formed. The default hidden-layer distribution is the product law

$$a_j \stackrel{\text{i.i.d.}}{\sim} \text{Unif}[-12, 12], \quad b_j \stackrel{\text{i.i.d.}}{\sim} \text{Unif}[-3, 3], \quad j = 1, \dots, m,$$

with independent draws across hidden units and across ensemble members. A normal-sampling sensitivity test uses

$$a_j \stackrel{\text{i.i.d.}}{\sim} N(0, 3^2), \quad b_j \stackrel{\text{i.i.d.}}{\sim} N(0, 3^2).$$

The selection criteria for Π are as follows: independent sampling across ensemble members, non-degenerate hidden features on the scaled time interval, and finite fourth moments of the resulting pointwise predictions after ridge fitting. For the bounded activation $\sigma = \tanh$ used here, both the uniform and normal choices satisfy these criteria automatically; the verification is given in Lemma 2.

With the design matrix

$$H_k = \begin{pmatrix} h_k(t_1)^\top \\ \vdots \\ h_k(t_n)^\top \end{pmatrix} \in \mathbb{R}^{n \times m},$$

and centered response vector

$$\mathbf{y}^\circ = \mathbf{y} - u_0 \mathbf{1}_n,$$

we compute the ridge-regression output weights by

$$\hat{\beta}_k = (H_k^\top H_k + \lambda I_m)^{-1} H_k^\top \mathbf{y}^\circ, \quad \lambda > 0. \quad (4)$$

The ridge parameter λ is included for stability and to avoid singularity when $m > n$ or the sampled features are nearly collinear.

The ensemble contains $M = 100$ independent members throughout the main applied study, although the theory is stated for a generic positive integer M . The ensemble mean and empirical spread are defined pointwise by

$$\bar{u}_M(t) = \frac{1}{M} \sum_{k=1}^M \hat{u}_k(t), \quad s_M^2(t) = \frac{1}{M-1} \sum_{k=1}^M (\hat{u}_k(t) - \bar{u}_M(t))^2. \quad (5)$$

The quantity $s_M(t)$ is the primitive uncertainty score produced by the randomized ensemble. It measures the sensitivity of the reconstruction to the hidden-layer randomization and, as shown below, converges to the random-feature prediction variance when M grows.

3. Conformalized Ensemble-ELM Uncertainty Bands

A raw interval of the form

$$\bar{u}_M(t) \pm z_{1-\alpha/2} s_M(t),$$

is easy to compute but has no universal coverage guarantee. In particular, the empirical spread only reflects the variability induced by the random feature draw; it may fail to account for model bias, heteroscedasticity, or local approximation error. To obtain a coverage statement that does not rely on a correct probabilistic model, we calibrate the spread by split conformal prediction.

We split the available design points into a training set $\mathcal{I}_{\text{tr}}$ and a calibration set $\mathcal{I}_{\text{cal}}$, with cardinalities n_{tr} and n_{cal} , respectively. All M ELM members are trained using only the training set. Then, for each calibration point $i \in \mathcal{I}_{\text{cal}}$, we compute the nonconformity score

$$R_i = \frac{|y_i - \bar{u}_M(t_i)|}{s_M(t_i) + \gamma}, \quad (6)$$

where $\gamma > 0$ is a small denominator safeguard. Let $R_{(1)} \leq \dots \leq R_{(n_{\text{cal}})}$ be the order statistics of the calibration scores, and define

$$q_{1-\alpha} = R_{(\lceil (n_{\text{cal}}+1)(1-\alpha) \rceil)}. \quad (7)$$

The proposed pointwise uncertainty band is then

$$\mathcal{C}_\alpha(t) = [\bar{u}_M(t) - q_{1-\alpha}(s_M(t) + \gamma), \bar{u}_M(t) + q_{1-\alpha}(s_M(t) + \gamma)]. \quad (8)$$

The width adapts locally through $s_M(t)$, while the conformal factor $q_{1-\alpha}$ corrects global under- or over-dispersion of the raw ensemble uncertainty.

The value of γ is a numerical regularization parameter for small-spread regions. It prevents a calibration residual from being divided by an empirical ensemble standard deviation that is close to zero. The experiments use $\gamma = 5 \times 10^{-3}$ in the response units of each benchmark, and Section 5.4 reports a sensitivity analysis over $\gamma \in \{0, 10^{-4}, 10^{-3}, 5 \times 10^{-3}, 10^{-2}\}$ and $n_{\text{cal}} \in \{10, 15, 20, 30\}$.

The training and calibration steps are therefore as follows. First, sample $M = 100$ independent hidden-layer parameter sets and fit the corresponding ELM output weights by (4). Second, evaluate the ensemble mean and spread at the calibration inputs. Third, form the normalized residuals (6) and compute the empirical conformal quantile (7). Finally, report the interval (8) on any desired time grid. The procedure is non-iterative at the ELM level, parallel across ensemble members, and agnostic to the particular ODE generating the trajectory data.

The conformal theorem in Section 3 is a pointwise marginal statement under exchangeability of the calibration pair and the future pair after the training split has been fixed. For deterministic high-fidelity ODE values sampled at randomized design points, this can be read as coverage over the randomized design-and-split mechanism. For observational time series with serially correlated measurement errors, the exchangeability condition becomes a modeling assumption that should be replaced by a block, trajectory-level, or other dependence-aware conformal construction. The present experiments use independent observation noise and random train/calibration splits, so the numerical study probes the stated pointwise calibration mechanism rather than simultaneous-in-time coverage of an entire dependent path.

4. Main Results

We now present the main results. The first lemma records the exact initial-condition property and the linear smoother form of each ensemble member. The next theorem gives stability estimates for the ensemble mean and spread. We then establish almost sure convergence of the empirical spread and the finite-sample coverage theorem for the conformalized interval. The section ends with two comparison propositions clarifying how the proposed uncertainty mechanism differs from ideal Bayesian prediction and Monte Carlo dropout.

Lemma 1. *For each ensemble member $k \in \{1, \dots, M\}$, the predictor (3)–(4) satisfies $\hat{u}_k(0) = u_0$. Moreover,*

$$\hat{u}_k(t) = u_0 + \ell_k(t)^\top \mathbf{y}^\circ, \quad \ell_k(t) = H_k(H_k^\top H_k + \lambda I_m)^{-1} h_k(t) \in \mathbb{R}^n. \quad (9)$$

Hence each ELM member is a linear smoother in the responses.

Proof. Since $h_k(t) = t\phi_{\omega_k}(t)$, we have $h_k(0) = \mathbf{0}$, and therefore from (3)

$$\hat{u}_k(0) = u_0 + h_k(0)^\top \hat{\beta}_k = u_0.$$

This proves the initial-condition statement.

For the linear smoother representation, substitute (4) into (3):

$$\hat{u}_k(t) = u_0 + h_k(t)^\top (H_k^\top H_k + \lambda I_m)^{-1} H_k^\top \mathbf{y}^\circ.$$

The right-hand side is a scalar. Using the identity $a^\top Bc = (B^\top a)^\top c$ and the symmetry of $(H_k^\top H_k + \lambda I_m)^{-1}$, we obtain

$$\hat{u}_k(t) = u_0 + (H_k(H_k^\top H_k + \lambda I_m)^{-1} h_k(t))^\top \mathbf{y}^\circ = u_0 + \ell_k(t)^\top \mathbf{y}^\circ,$$

which is exactly (9). The dependence on the responses is therefore linear. This completes the proof. $\square$

Theorem 1. *Fix $t \in [0, T]$. For the k th ensemble member define*

$$C_k(t) = \frac{\|h_k(t)\|_2 \|H_k\|_2}{\lambda}.$$

Then for any two response vectors $\mathbf{y}, \tilde{\mathbf{y}} \in \mathbb{R}^n$ with centered versions $\mathbf{y}^\circ, \tilde{\mathbf{y}}^\circ$, one has

$$|\hat{u}_k(t; \mathbf{y}) - \hat{u}_k(t; \tilde{\mathbf{y}})| \leq C_k(t) \|\mathbf{y}^\circ - \tilde{\mathbf{y}}^\circ\|_2. \quad (10)$$

Consequently,

$$|\bar{u}_M(t; \mathbf{y}) - \bar{u}_M(t; \tilde{\mathbf{y}})| \leq \left(\frac{1}{M} \sum_{k=1}^M C_k(t) \right) \|\mathbf{y}^\circ - \tilde{\mathbf{y}}^\circ\|_2, \quad (11)$$

and

$$|s_M(t; \mathbf{y}) - s_M(t; \tilde{\mathbf{y}})| \leq \left(\frac{\sum_{k=1}^M C_k(t)^2}{M-1} \right)^{1/2} \|\mathbf{y}^\circ - \tilde{\mathbf{y}}^\circ\|_2. \quad (12)$$

Proof. Let

$$A_k = (H_k^\top H_k + \lambda I_m)^{-1} H_k^\top.$$

Using (3), we write

$$\hat{u}_k(t; \mathbf{y}) - \hat{u}_k(t; \tilde{\mathbf{y}}) = h_k(t)^\top A_k (\mathbf{y}^\circ - \tilde{\mathbf{y}}^\circ).$$

Therefore,

$$|\hat{u}_k(t; \mathbf{y}) - \hat{u}_k(t; \tilde{\mathbf{y}})| \leq \|h_k(t)\|_2 \|A_k\|_2 \|\mathbf{y}^\circ - \tilde{\mathbf{y}}^\circ\|_2.$$

Now

$$\|A_k\|_2 = \|(H_k^\top H_k + \lambda I_m)^{-1} H_k^\top\|_2 \leq \|(H_k^\top H_k + \lambda I_m)^{-1}\|_2 \|H_k\|_2.$$

Since the smallest eigenvalue of $H_k^\top H_k + \lambda I_m$ is at least λ , we have

$$\|(H_k^\top H_k + \lambda I_m)^{-1}\|_2 \leq \frac{1}{\lambda}.$$

Combining the last three displays proves (10).

To obtain (11), average (10) over $k = 1, \dots, M$ and use the following triangle inequality:

$$|\bar{u}_M(t; \mathbf{y}) - \bar{u}_M(t; \tilde{\mathbf{y}})| \leq \frac{1}{M} \sum_{k=1}^M |\hat{u}_k(t; \mathbf{y}) - \hat{u}_k(t; \tilde{\mathbf{y}})| \leq \left(\frac{1}{M} \sum_{k=1}^M C_k(t) \right) \|\mathbf{y}^\circ - \tilde{\mathbf{y}}^\circ\|_2.$$

For the spread, let

$$\mathbf{v} = (\hat{u}_1(t; \mathbf{y}), \dots, \hat{u}_M(t; \mathbf{y}))^\top, \quad \tilde{\mathbf{v}} = (\hat{u}_1(t; \tilde{\mathbf{y}}), \dots, \hat{u}_M(t; \tilde{\mathbf{y}}))^\top.$$

Let $P = I_M - \frac{1}{M} \mathbf{1}_M \mathbf{1}_M^\top$ be the centering matrix. Then

$$s_M(t; \mathbf{y}) = \frac{\|P\mathbf{v}\|_2}{\sqrt{M-1}}, \quad s_M(t; \tilde{\mathbf{y}}) = \frac{\|P\tilde{\mathbf{v}}\|_2}{\sqrt{M-1}}.$$

Hence,

$$|s_M(t; \mathbf{y}) - s_M(t; \tilde{\mathbf{y}})| \leq \frac{1}{\sqrt{M-1}} \|P(\mathbf{v} - \tilde{\mathbf{v}})\|_2 \leq \frac{1}{\sqrt{M-1}} \|\mathbf{v} - \tilde{\mathbf{v}}\|_2,$$

because $\|P\|_2 \leq 1$. Using (10) componentwise gives

$$\|\mathbf{v} - \tilde{\mathbf{v}}\|_2^2 = \sum_{k=1}^M |\hat{u}_k(t; \mathbf{y}) - \hat{u}_k(t; \tilde{\mathbf{y}})|^2 \leq \left(\sum_{k=1}^M C_k(t)^2 \right) \|\mathbf{y}^\circ - \tilde{\mathbf{y}}^\circ\|_2^2.$$

Taking square roots yields (12). The proof is complete. $\square$

Lemma 2 (Finite-moment verification for the hidden-layer sampling). *Fix a compact interval $[0, T]$, a finite training vector $\mathbf{y}$, a ridge parameter $\lambda > 0$, and the bounded activation $\sigma = \tanh$.*

Then, for every fixed $t \in [0, T]$, the random prediction $\hat{u}_1(t)$ generated by (3)–(4) has finite fourth moment under both the uniform and normal hidden-parameter laws specified in Section 2.

Proof. For $\sigma = \tanh$, each hidden feature satisfies $|\sigma(a_j t + b_j)| \leq 1$. Therefore $\|h_1(t)\|_2 \leq T\sqrt{m}$ and $\|H_1\|_2 \leq T\sqrt{nm}$ for every realization of the hidden parameters. From (4),

$$\|\hat{\beta}_1\|_2 \leq \|(H_1^\top H_1 + \lambda I_m)^{-1}\|_2 \|H_1\|_2 \|\mathbf{y}^\circ\|_2 \leq \frac{T\sqrt{nm}}{\lambda} \|\mathbf{y}^\circ\|_2.$$

Consequently,

$$|\hat{u}_1(t)| \leq |u_0| + \|h_1(t)\|_2 \|\hat{\beta}_1\|_2 \leq |u_0| + \frac{T^2 m \sqrt{n}}{\lambda} \|\mathbf{y}^\circ\|_2.$$

The final bound is deterministic and finite. Hence $\mathbb{E}(|\hat{u}_1(t)|^4 | \mathcal{D}_{\text{tr}}) < \infty$. The argument uses only the boundedness of the activation and the ridge parameter, so it applies equally to the stated uniform and normal laws for the hidden weights and biases. $\square$

Theorem 2. Fix $t \in [0, T]$ and condition on the training data. Suppose that the ensemble members $\hat{u}_1(t), \hat{u}_2(t), \dots$ are independent and identically distributed through the random hidden-layer draws, and that

$$\mathbb{E}[|\hat{u}_1(t)|^4 | \mathcal{D}_{\text{tr}}] < \infty.$$

Lemma 2 verifies this condition for the uniform and normal hidden-layer sampling laws used in the numerical experiments. Define

$$\mu(t) = \mathbb{E}[\hat{u}_1(t) | \mathcal{D}_{\text{tr}}], \quad v(t) = \text{Var}(\hat{u}_1(t) | \mathcal{D}_{\text{tr}}).$$

Then, as $M \rightarrow \infty$,

$$\bar{u}_M(t) \rightarrow \mu(t) \quad \text{almost surely,} \quad (13)$$

and

$$s_M^2(t) \rightarrow v(t) \quad \text{almost surely.} \quad (14)$$

In particular, the empirical ensemble spread converges almost surely to the random-feature predictive standard deviation.

Proof. Conditional on the training data, the random variables $\hat{u}_1(t), \hat{u}_2(t), \dots$ are i.i.d. with finite first moment, because the assumed finite fourth moment implies finite first and second moments. By the strong law of large numbers,

$$\frac{1}{M} \sum_{k=1}^M \hat{u}_k(t) \rightarrow \mathbb{E}[\hat{u}_1(t) | \mathcal{D}_{\text{tr}}] = \mu(t) \quad \text{almost surely.}$$

This is exactly (13).

Again, by the strong law of large numbers applied to the square, we have

$$\frac{1}{M} \sum_{k=1}^M \hat{u}_k(t)^2 \rightarrow \mathbb{E}[\hat{u}_1(t)^2 | \mathcal{D}_{\text{tr}}] \quad \text{almost surely.}$$

Recall the identity

$$s_M^2(t) = \frac{M}{M-1} \left(\frac{1}{M} \sum_{k=1}^M \hat{u}_k(t)^2 - \left(\frac{1}{M} \sum_{k=1}^M \hat{u}_k(t) \right)^2 \right).$$

Since $M/(M-1) \rightarrow 1$, the first average converges almost surely to $\mathbb{E}[\hat{u}_1(t)^2 \mid \mathcal{D}_{\text{tr}}]$, and the second term converges almost surely to $\mu(t)^2$. Therefore,

$$s_M^2(t) \rightarrow \mathbb{E}[\hat{u}_1(t)^2 \mid \mathcal{D}_{\text{tr}}] - \mu(t)^2 = \text{Var}(\hat{u}_1(t) \mid \mathcal{D}_{\text{tr}}) = v(t) \quad \text{almost surely.}$$

This proves (14). The final statement follows by taking square roots, since the square-root map is continuous on $[0, \infty)$. $\square$

Theorem 3. Let $\mathcal{D}_{\text{tr}}$ be fixed and train the ensemble predictor using only the training split. Suppose that the calibration pairs

$$\{(t_i, y_i) : i \in \mathcal{I}_{\text{cal}}\}$$

and a future test pair (t_*, y_*) are exchangeable conditional on $\mathcal{D}_{\text{tr}}$. Define the calibration scores by (6) and the interval by (8). Then the conformalized ensemble interval satisfies the finite-sample marginal coverage guarantee

$$\mathbb{P}\{y_* \in \mathcal{C}_\alpha(t_*) \mid \mathcal{D}_{\text{tr}}\} \geq 1 - \alpha. \quad (15)$$

Consequently,

$$\mathbb{P}\{y_* \in \mathcal{C}_\alpha(t_*)\} \geq 1 - \alpha.$$

Proof. Conditional on the training data, the fitted predictor $t \mapsto \bar{u}_M(t)$ and the spread function $t \mapsto s_M(t)$ are deterministic because they depend only on the training split and on the sampled ensemble members. Therefore, the calibration scores

$$R_i = \frac{|y_i - \bar{u}_M(t_i)|}{s_M(t_i) + \gamma}, \quad i \in \mathcal{I}_{\text{cal}},$$

and the test score

$$R_* = \frac{|y_* - \bar{u}_M(t_*)|}{s_M(t_*) + \gamma}$$

are exchangeable conditional on $\mathcal{D}_{\text{tr}}$, because they are obtained by applying the same measurable map to exchangeable input-response pairs.

Let $N = n_{\text{cal}} + 1$ denote the total number of scores after adjoining R_* . By exchangeability, the rank of R_* among the N scores is uniformly distributed on $\{1, 2, \dots, N\}$ conditional on $\mathcal{D}_{\text{tr}}$. The quantile definition (7) uses the index

$$r = \lceil (n_{\text{cal}} + 1)(1 - \alpha) \rceil = \lceil N(1 - \alpha) \rceil.$$

Hence

$$\mathbb{P}\{R_* \leq q_{1-\alpha} \mid \mathcal{D}_{\text{tr}}\} \geq \frac{r}{N} \geq 1 - \alpha.$$

By the definition of R_* , the event $R_* \leq q_{1-\alpha}$ is equivalent to

$$|y_* - \bar{u}_M(t_*)| \leq q_{1-\alpha}(s_M(t_*) + \gamma),$$

which is in turn equivalent to $y_* \in \mathcal{C}_\alpha(t_*)$. Therefore

$$\mathbb{P}\{y_* \in \mathcal{C}_\alpha(t_*) \mid \mathcal{D}_{\text{tr}}\} \geq 1 - \alpha.$$

Taking expectations with respect to the training split proves the unconditional statement as well. $\square$

Remark 1. Theorem 3 is formulated for future responses y_* . In ODE surrogate applications where the calibration labels are high-fidelity state values, the same theorem provides coverage for the latent trajectory. If the available labels are noisy observations, the theorem gives valid predictive bands for the observation process; a separate estimate of the noise level can then be used to convert predictive intervals into state intervals when appropriate.

Proposition 1. Consider a Bayesian neural surrogate in an ideal linear-Gaussian regime with deterministic feature map $\psi : [0, T] \rightarrow \mathbb{R}^p$, latent coefficient vector $\beta \sim N(0, \tau^2 I_p)$, and observation model

$$y_i = \psi(t_i)^\top \beta + \varepsilon_i, \quad \varepsilon_i \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma^2).$$

Let $X \in \mathbb{R}^{n \times p}$ have rows $\psi(t_i)^\top$, and let

$$\Sigma_{\text{post}} = \left(\tau^{-2} I_p + \sigma^{-2} X^\top X \right)^{-1}, \quad \mu_{\text{post}} = \sigma^{-2} \Sigma_{\text{post}} X^\top \mathbf{y}.$$

Then, for a new input t_* , the posterior predictive law of the future response is Gaussian with mean $\psi(t_*)^\top \mu_{\text{post}}$ and variance

$$\psi(t_*)^\top \Sigma_{\text{post}} \psi(t_*) + \sigma^2.$$

Therefore the interval

$$\psi(t_*)^\top \mu_{\text{post}} \pm z_{1-\alpha/2} \sqrt{\psi(t_*)^\top \Sigma_{\text{post}} \psi(t_*) + \sigma^2} \quad (16)$$

has exact posterior predictive coverage $1 - \alpha$ under the assumed probabilistic model.

Proof. Since both the prior and the likelihood are Gaussian and the observation model is linear in β , the posterior distribution of β given $(X, \mathbf{y})$ is Gaussian with the displayed mean and covariance. This is a standard conjugate computation obtained by completing the square in the exponent of the joint density.

Now let $y_* = \psi(t_*)^\top \beta + \varepsilon_*$ with $\varepsilon_* \sim N(0, \sigma^2)$ independent of all observed data and independent of β conditional on the posterior. Since an affine transformation of a Gaussian vector is Gaussian, $\psi(t_*)^\top \beta \mid X, \mathbf{y}$ is normal with mean $\psi(t_*)^\top \mu_{\text{post}}$ and variance $\psi(t_*)^\top \Sigma_{\text{post}} \psi(t_*)$. Adding the independent Gaussian noise ε_* yields the predictive variance

$$\psi(t_*)^\top \Sigma_{\text{post}} \psi(t_*) + \sigma^2.$$

Hence,

$$\frac{y_* - \psi(t_*)^\top \mu_{\text{post}}}{\sqrt{\psi(t_*)^\top \Sigma_{\text{post}} \psi(t_*) + \sigma^2}} \mid X, \mathbf{y} \sim N(0, 1).$$

It follows immediately that the two-sided interval (16) contains y_* with probability $1 - \alpha$ conditional on X and $\mathbf{y}$. This is the claimed exact posterior predictive coverage. $\square$

Proposition 2. Let a dropout surrogate at a fixed input t be represented by

$$F_d(t) = \sum_{j=1}^p \frac{d_j}{p_0} w_j h_j(t), \quad (17)$$

where $d_1, \dots, d_p$ are independent Bernoulli (p_0) random variables with $0 < p_0 \leq 1$, and $w_j, h_j(t)$ are deterministic. Then

$$\mathbb{E}[F_d(t)] = \sum_{j=1}^p w_j h_j(t), \quad (18)$$

and

$$\text{Var}(F_d(t)) = \frac{1-p_0}{p_0} \sum_{j=1}^p w_j^2 h_j(t)^2. \quad (19)$$

If the true response is $Y(t) = u(t) + \varepsilon(t)$ with mean-zero noise variance $\sigma_\varepsilon^2(t)$ and structural bias

$$b(t) = u(t) - \mathbb{E}[F_d(t)],$$

then

$$\mathbb{E}[(Y(t) - F_d(t))^2] = \text{Var}(F_d(t)) + b(t)^2 + \sigma_\varepsilon^2(t). \quad (20)$$

Consequently, intervals based only on the Monte Carlo estimate of (19) may under-cover whenever $b(t)^2 + \sigma_\varepsilon^2(t)$ is not negligible.

Proof. Using linearity of expectation and the fact that $\mathbb{E}[d_j] = p_0$, we obtain

$$\mathbb{E}[F_d(t)] = \sum_{j=1}^p \frac{\mathbb{E}[d_j]}{p_0} w_j h_j(t) = \sum_{j=1}^p w_j h_j(t),$$

which proves (18).

For the variance, independence of the d_j implies that the cross-covariances vanish. Since $\text{Var}(d_j) = p_0(1 - p_0)$, we have

$$\text{Var}(F_d(t)) = \sum_{j=1}^p \text{Var}\left(\frac{d_j}{p_0} w_j h_j(t)\right) = \sum_{j=1}^p \frac{w_j^2 h_j(t)^2}{p_0^2} \text{Var}(d_j) = \frac{1-p_0}{p_0} \sum_{j=1}^p w_j^2 h_j(t)^2,$$

which is (19).

To prove (20), write

$$Y(t) - F_d(t) = (u(t) - \mathbb{E}[F_d(t)]) + \varepsilon(t) - (F_d(t) - \mathbb{E}[F_d(t)]).$$

The first term is the deterministic bias $b(t)$. The second term has mean zero and variance $\sigma_\varepsilon^2(t)$. The third term has mean zero and variance $\text{Var}(F_d(t))$. Assuming the observation noise is independent of the dropout mask, the cross terms vanish after taking expectations. Therefore

$$\mathbb{E}[(Y(t) - F_d(t))^2] = b(t)^2 + \sigma_\varepsilon^2(t) + \text{Var}(F_d(t)).$$

This proves (20). Since a naive dropout interval uses only the square root of (19), it ignores the bias and observation-noise contributions. When these missing terms are appreciable, the interval can be narrower than required for nominal coverage, and under-coverage follows. $\square$

Remark 2. The comparison between Theorem 3, Proposition 1, and Proposition 2 is conceptually important. The conformalized ensemble-ELM band achieves finite-sample marginal coverage without requiring a correctly specified likelihood or prior. The Bayesian interval is exact only inside a fully specified probabilistic model. The dropout interval, unless explicitly supplemented by an observation-noise model or an external calibration step, quantifies only mask-induced randomness and therefore does not by itself provide a coverage guarantee.

5. Numerical Simulations and Discussion

This section reports an executed numerical study. The aim is threefold: to verify that the conformalized ensemble ELM produces reliable pointwise trajectory bands in sparse-data ODE reconstruction; to compare its calibration and sharpness against capacity-matched Bayesian and dropout baselines; and to quantify the sensitivity to calibration regularization, calibration-sample size, ensemble size, training time, noise level, and hidden-layer sampling law.

5.1. Experimental Protocol

We considered three benchmark dynamics: exponential decay, logistic growth, and an underdamped oscillator. For each problem, we generated 60 sparse observation times on the corresponding interval, forced the initial time $t = 0$ into the training set, and used a 20/20/20 split for training, calibration, and hold-out observations. The reported accuracy and coverage metrics were computed on a dense latent-state test grid of 400 equally spaced points. Independent Gaussian observation noise was added to the sparse samples, with standard deviations equal to 3%, 3.5%, and 4% of the standard deviation of the exact trajectory for the exponential, logistic, and oscillatory benchmarks, respectively.

These three noise levels were chosen to provide a moderate-noise regime with slightly increasing difficulty across the three examples: the exponential problem is monotone and smooth, the logistic problem contains a nonlinear transition region, and the damped oscillator is phase-sensitive. Higher-noise behavior is evaluated separately in Section 5.6, where the noise level is varied over 2%, 4%, 8%, and 12% of the trajectory standard deviation.

The ensemble ELM used $M = 100$ members, hidden width $m = 200$, hyperbolic tangent activation, ridge parameter $\lambda = 10^{-4}$, the hidden-parameter law described in Section 2, and the trial form $\hat{u}(t) = u_0 + t g(t)$ so that the initial condition was satisfied exactly for every ensemble member. The split-conformal score used the normalized residual in (6) with $\gamma = 5 \times 10^{-3}$. The aggregate numbers below are the averages over three independent random splits.

The two baselines were made capacity-matched to the ELM hidden layer. The Bayesian neural surrogate used the same one-hidden-layer $m = 200$ tanh random-feature architecture with a Gaussian output-layer prior $\beta \sim N(0, \tau^2 I_m)$, $\tau = 1$, and a Gaussian likelihood using the observation-noise level specified above. Its predictive mean and variance were computed from the closed-form Bayesian linear posterior. The Monte Carlo dropout baseline used the same hidden width $m = 200$, the same tanh features, ridge-trained output weights, keep probability $p_0 = 0.9$, and 100 stochastic dropout forward passes. All three methods used the same initial-condition trial form. The supplied Python 3.14.6 script reproduces the figures and stores the numerical metrics used in the tables and plots.

5.2. Representative Trajectory Reconstructions

Figure 1 shows representative reconstructions for the three ODE classes using the conformalized ensemble ELM. The method tracks the exact trajectory closely for the monotone examples and remains calibrated for the oscillatory trajectory, where sparse observations make phase reconstruction more demanding. The uncertainty band widens in the nonlinear transition region of the logistic problem and around phase-sensitive portions of the damped oscillator.

For the oscillatory problem, Figure 2 shows the qualitative difference among the three uncertainty mechanisms. The Bayesian random-feature interval is sharp but under-covers when the sparse data induce phase bias. The dropout interval is very wide in this capacity-matched implementation because mask-induced variability in oscillatory extrapolation is

large. The conformalized ELM produces an intermediate band with coverage close to the nominal level and a much lower Winkler score.

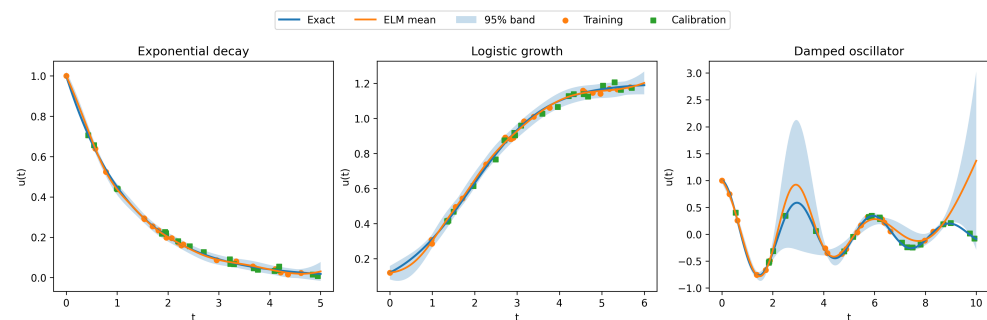


Figure 1. Representative conformalized ensemble-ELM reconstructions for the exponential, logistic, and damped-oscillator benchmarks. Circles denote training observations and squares denote calibration observations.

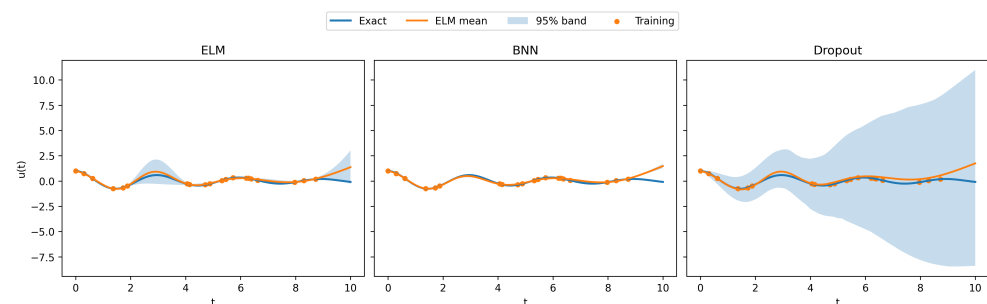


Figure 2. Method comparison on a representative damped-oscillator reconstruction. The conformalized ensemble ELM balances point accuracy and calibrated interval width more effectively than the two native baselines on this phase-sensitive benchmark.

5.3. Aggregate Calibration and Sharpness

The full metric summary is reported in Table 1. Coverage probability (CP) is computed on the dense latent-state test grid at nominal level 0.95; AW is the average interval width; RMSE is the root mean squared error of the predictive mean; and the Winkler score rewards narrow intervals while penalizing missed coverage.

Table 1. Average performance over three independent random splits. Coverage probability (CP) is reported on the dense latent-state test grid at nominal level 0.95. The table is resized to the text width to avoid boundary overflow.

Problem	Method	CP	AW	RMSE	Winkler
Damped oscillator	BNN	0.361 ± 0.078	0.094 ± 0.016	0.183 ± 0.116	2.813 ± 1.223
	Dropout	0.987 ± 0.022	9.204 ± 1.191	0.304 ± 0.127	9.231 ± 1.161
	ELM	0.952 ± 0.024	0.418 ± 0.181	0.179 ± 0.112	0.480 ± 0.241
Exponential decay	BNN	1.000 ± 0.000	0.085 ± 0.036	0.009 ± 0.002	0.085 ± 0.036
	Dropout	0.976 ± 0.036	0.334 ± 0.052	0.012 ± 0.003	0.342 ± 0.065
	ELM	0.987 ± 0.022	0.048 ± 0.005	0.009 ± 0.002	0.049 ± 0.006
Logistic growth	BNN	1.000 ± 0.000	0.079 ± 0.021	0.013 ± 0.003	0.079 ± 0.021
	Dropout	0.950 ± 0.025	0.366 ± 0.085	0.018 ± 0.004	0.370 ± 0.082
	ELM	0.983 ± 0.029	0.072 ± 0.015	0.013 ± 0.003	0.073 ± 0.014

The conformalized ELM achieved CP values 0.987 ± 0.022 , 0.983 ± 0.029 , and 0.952 ± 0.024 on the exponential, logistic, and damped-oscillator benchmarks, respectively. The first two values are mildly conservative, while the oscillatory value is close to the target. The Bayesian random-feature baseline is highly effective on the two smooth

monotone examples, achieving CP equal to 1.000 with competitive widths, but it under-covers on the oscillatory problem because the posterior variance is sharp relative to phase bias. The dropout baseline reaches or exceeds the nominal target in these experiments, while its widths are much larger, especially in the damped oscillator case. These results are consistent with the theory: model-based intervals can be sharp when the model captures the trajectory well, whereas conformal calibration directly corrects the empirical residual scale.

The coverage and sharpness trends are summarized visually in Figures 3 and 4. Figure 3 shows the separation in empirical coverage. Figure 4 shows that the ELM intervals provide the lowest Winkler score on the exponential and damped-oscillator benchmarks and remain competitive on the logistic benchmark.

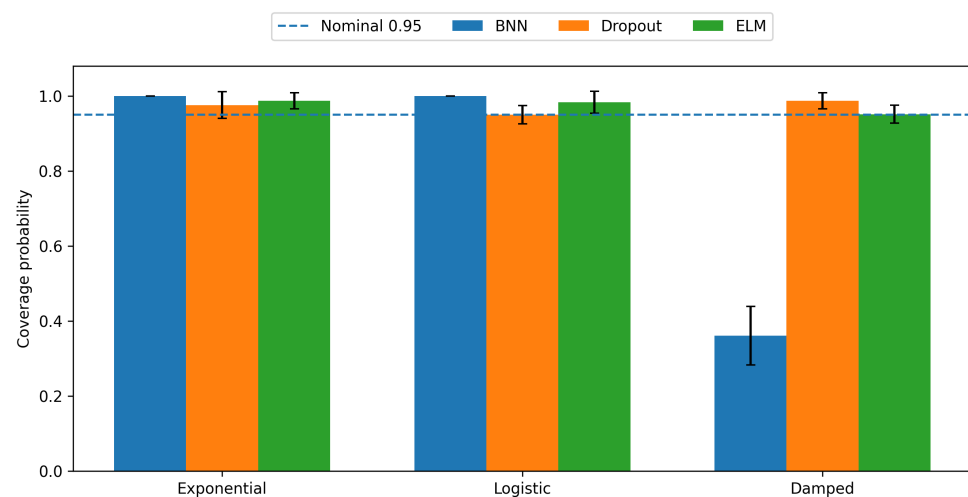


Figure 3. Empirical coverage probability on the dense latent-state test grid. The dashed horizontal line marks the nominal 95% target.

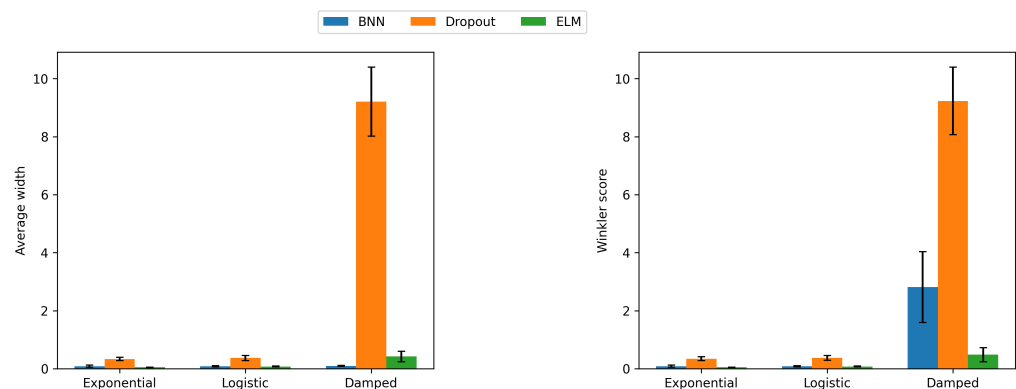


Figure 4. Average interval width and average Winkler score. Lower Winkler values indicate a better balance between sharpness and calibration.

5.4. Denominator and Calibration-Size Ablation

To separate the influence of the denominator safeguard from the finite calibration split, we repeated the logistic-growth experiment for $\gamma \in \{0, 10^{-4}, 10^{-3}, 5 \times 10^{-3}, 10^{-2}\}$ and $n_{\text{cal}} \in \{10, 15, 20, 30\}$. Figure 5 reports the resulting coverage and average width. With only $n_{\text{cal}} = 10$ calibration points, using $\gamma = 0$ yielded CP 0.925 and AW 0.068, while $\gamma = 5 \times 10^{-3}$ yielded CP 1.000 and AW 0.051. For $n_{\text{cal}} = 20$, the same change reduced AW from 0.350 to 0.083 while preserving CP at 0.995–1.000. Thus the finite calibration quantile is the main source of discreteness and mild conservativeness, while the denominator safeguard

regularizes small-spread calibration scores and prevents unstable interval inflation. Values between 10^{-3} and 10^{-2} produced similar conclusions in this benchmark.

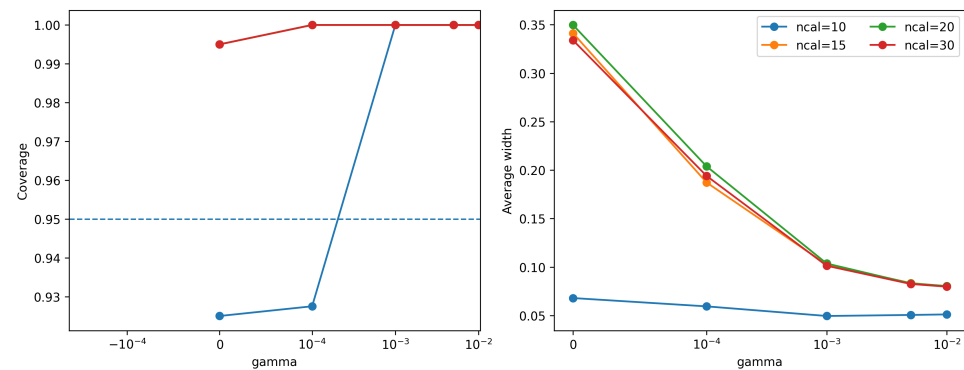


Figure 5. Ablation over the denominator safeguard γ and calibration size n_{cal} for logistic growth. The left panel reports empirical coverage and the right panel reports average width.

5.5. Ensemble-Size Sensitivity and Training Time

To examine the role of the ensemble size itself, we repeated the logistic experiment for

$$M \in \{5, 10, 20, 40, 60, 80, 100, 150\}.$$

Figure 6 shows that the accuracy and interval metrics stabilize once the ensemble reaches moderate size. In the executed Python implementation, the measured ELM training-and-prediction time increased approximately linearly from 0.0048 s at $M = 5$ to 0.097 s at $M = 150$. The time at the main value $M = 100$ was 0.063 s for this split. This confirms that the additional ensemble members mainly add embarrassingly parallel ridge solves, while the calibration and evaluation steps remain inexpensive.

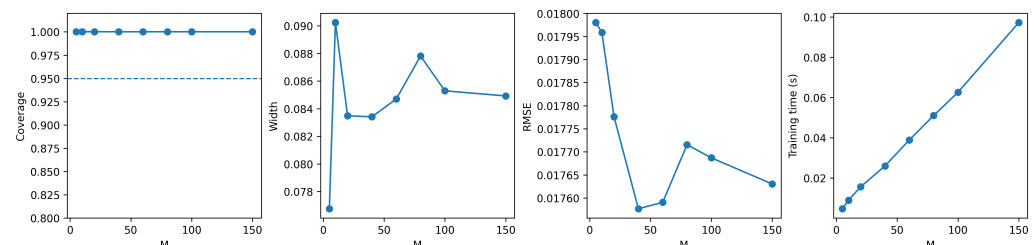


Figure 6. Sensitivity of the conformalized ensemble ELM to the ensemble size M on a representative logistic-growth split. The fourth panel reports the measured training-and-prediction time in the supplied Python implementation.

5.6. Noise Level and Hidden-Distribution Sensitivity

Figure 7 tests higher noise levels and the hidden-parameter distribution on the damped oscillator. The noise standard deviation was set to 2%, 4%, 8%, and 12% of the trajectory standard deviation, and the hidden features were sampled either from the default uniform law or from the normal law stated in Section 2. Uniform sampling produced mean CP values 0.983, 0.994, 1.000, and 0.998 across the four noise levels, while normal sampling produced 0.953, 0.941, 0.984, and 0.999. The average widths increased with noise, as expected from calibration against noisier observations. This experiment verifies that the hidden-layer moment hypothesis used in Theorem 2 is compatible with the actual uniform and normal sampling schemes and that the method remains calibrated in substantially noisier regimes.

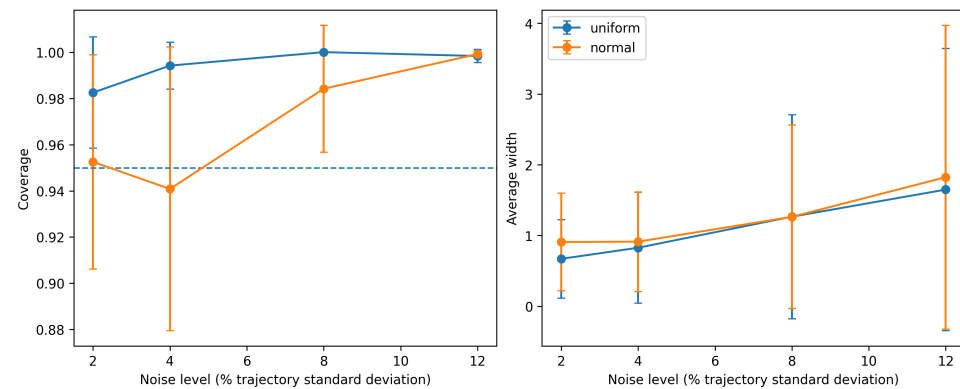


Figure 7. Sensitivity to observation noise and hidden-layer sampling distribution on the damped oscillator. Uniform and normal hidden-parameter laws both preserve near-nominal or conservative coverage over the tested noise range.

5.7. Limitations of the Numerical Study

The numerical study remains deliberately controlled. It uses scalar synthetic ODEs with known exact solutions, one-hidden-layer capacity-matched baselines, and three random splits for the aggregate table. The resulting evidence is therefore best interpreted as a focused validation of the coverage mechanism rather than as an exhaustive benchmark against every modern uncertainty-aware neural ODE architecture. In addition, the reported conformal bands are pointwise and marginal, exactly as in Theorem 3; simultaneous-in-time coverage would require a more conservative calibration rule. Finally, the ELM coverage in the monotone experiments is mildly conservative, which is useful for robustness in sparse settings and also motivates future work on adaptive safeguards and locally refined calibration rules to tighten the bands further.

6. Conclusions

This paper has introduced a conformalized ensemble extreme learning machine framework for uncertainty quantification in ordinary differential equations and has complemented the theory by an executed simulation study. The central mechanism is to train independently randomized ELM surrogates, use their ensemble mean as the reconstructed trajectory, and use the ensemble spread as a local uncertainty score. The resulting method preserves the main computational advantage of ELMs, namely, closed-form training of the output layer, while gaining a statistically meaningful uncertainty statement through split-conformal calibration.

The revised formulation makes explicit how the work differs from general ELM regression intervals and deep ensembles, specifies the hidden-layer sampling laws used in the experiments, verifies the finite-moment hypothesis for both uniform and normal sampling, clarifies the exchangeability condition behind the conformal theorem, and reports additional diagnostics for the denominator safeguard, calibration size, noise level, ensemble size, and training time.

The theoretical analysis established that each ensemble member satisfies the initial condition exactly, that the ensemble mean and spread are stable under perturbations of the response data, that the empirical spread converges almost surely to the random-feature predictive variance, and that the conformalized interval achieves finite-sample marginal coverage under exchangeability. The numerical study corroborated these statements in a concrete sparse-data regime. Across the exponential, logistic, and damped-oscillator benchmarks, the conformalized ensemble ELM attained near-nominal or mildly conservative pointwise coverage with competitive RMSE and favorable Winkler scores, especially on

the phase-sensitive oscillatory example where model-based intervals can under-cover or become excessively wide.

For ODE trajectory reconstruction, these results place ensemble ELMs in a useful operating region: they are fast, parallel, mathematically transparent, and calibratable without committing to a fully correct probabilistic model. The most natural next steps are simultaneous-in-time bands, multivariate systems, adaptive random-feature distributions, block conformal calibration for serially dependent observations, and hybrid residual-penalized formulations that combine exact initial-condition enforcement with richer physics-informed structure.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/math14142527/s1>, Supplementary Python code used to generate all figures; generated metric files; and recreated figures corresponding to the results presented in this manuscript.

Author Contributions: Conceptualization, S.A.S. and L.A.W.; Methodology, S.A.S. and L.A.W.; Validation, L.A.W.; Formal analysis, S.A.S. and L.A.W.; Investigation, S.A.S. and L.A.W.; Writing—original draft, S.A.S. and L.A.W.; Writing—review and editing, S.A.S. and L.A.W.; Visualization, L.A.W.; Supervision, L.A.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [Grant No. KFU263757].

Data Availability Statement: The Python code used to generate all figures, together with generated metric files and recreated figures, is provided as Supplementary Material accompanying this manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ODE	Ordinary differential equation
UQ	Uncertainty quantification
ELM	Extreme learning machine
BNN	Bayesian neural network
MC	Monte Carlo
CP	Coverage probability
AW	Average width
RMSE	Root mean squared error

References

1. Psaros, A.F.; Meng, X.; Zou, Z.; Guo, L.; Karniadakis, G.E. Uncertainty Quantification in Scientific Machine Learning: Methods, Metrics, and Comparisons. *J. Comput. Phys.* **2023**, *477*, 111902. [[CrossRef](#)]
2. Schmid, N.; Fernandes del Pozo, D.; Waegeman, W.; Hasenauer, J. Assessment of Uncertainty Quantification in Universal Differential Equations. *Philos. Trans. R. Soc. A* **2025**, *383*, 20240444. [[CrossRef](#)]
3. De Florio, M.; Zou, Z.; Schiavazzi, D.E.; Karniadakis, G.E. Quantification of Total Uncertainty in the Physics-Informed Reconstruction of CVSim-6 Physiology. *Philos. Trans. R. Soc. A* **2025**, *383*, 20240221. [[CrossRef](#)]
4. Neal, R.M. *Bayesian Learning for Neural Networks*; Springer: New York, NY, USA, 1996.
5. Gal, Y.; Ghahramani, Z. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. In Proceedings of the 33rd International Conference on Machine Learning, New York, NY, USA, 19–24 June 2016; pp. 1050–1059.
6. Lakshminarayanan, B.; Pritzel, A.; Blundell, C. Simple and Scalable Predictive Uncertainty Estimation Using Deep Ensembles. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Volume 30, pp. 6402–6413.

7. Ovadia, Y.; Fertig, E.; Ren, J.; Nado, Z.; Sculley, D.; Nowozin, S.; Dillon, J.V.; Lakshminarayanan, B.; Snoek, J. Can You Trust Your Model's Uncertainty? Evaluating Predictive Uncertainty Under Dataset Shift. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019; Volume 32, pp. 13991–14002.
8. Abdar, M.; Pourpanah, F.; Hussain, S.; Rezazadegan, D.; Liu, L.; Ghavamzadeh, M.; Fieguth, P.; Cao, X.; Khosravi, A.; Acharya, U.R.; et al. A Review of Uncertainty Quantification in Deep Learning: Techniques, Applications and Challenges. *Inf. Fusion* **2021**, *76*, 243–297. [\[CrossRef\]](#)
9. Huang, G.B.; Zhu, Q.Y.; Siew, C.K. Extreme Learning Machine: Theory and Applications. *Neurocomputing* **2006**, *70*, 489–501. [\[CrossRef\]](#)
10. Huang, G.B.; Zhou, H.; Ding, X.; Zhang, R. Extreme Learning Machine for Regression and Multiclass Classification. *IEEE Trans. Syst. Man Cybern. Part B* **2012**, *42*, 513–529. [\[CrossRef\]](#)
11. Calabrò, F.; Fabiani, G.; Siettos, C. Extreme Learning Machine Collocation for the Numerical Solution of Elliptic PDEs with Sharp Gradients. *Comput. Methods Appl. Mech. Eng.* **2021**, *387*, 114188. [\[CrossRef\]](#)
12. Wang, Y.; Dong, S. An Extreme Learning Machine-Based Method for Computational PDEs in Higher Dimensions. *Comput. Methods Appl. Mech. Eng.* **2024**, *418*, 116578. [\[CrossRef\]](#)
13. Liu, X.; Yao, W.; Peng, W.; Zhou, W. Bayesian Physics-Informed Extreme Learning Machine for Forward and Inverse PDE Problems with Noisy Data. *Neurocomputing* **2023**, *549*, 126425. [\[CrossRef\]](#)
14. Guignard, F.; Amato, F.; Kanevski, M. Uncertainty Quantification in Extreme Learning Machine: Analytical Developments, Variance Estimates and Confidence Intervals. *Neurocomputing* **2021**, *456*, 436–449. [\[CrossRef\]](#)
15. Lee, J.; Oh, Y.; Kim, S.; Lim, D.Y. Continuum Dropout for Neural Differential Equations. *Proc. AAAI Conf. Artif. Intell.* **2026**, *40*, 22796–22804. [\[CrossRef\]](#)
16. Vovk, V.; Gammerman, A.; Shafer, G. *Algorithmic Learning in a Random World*; Springer: New York, NY, USA, 2005.
17. Lei, J.; G'Sell, M.; Rinaldo, A.; Tibshirani, R.J.; Wasserman, L. Distribution-Free Predictive Inference for Regression. *J. Am. Stat. Assoc.* **2018**, *113*, 1094–1111. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.