Hidden Data Recovery and Forecasting via Next-Generation Reservoir Computing With Multiscale Delay Selection

Artem Badarin, Andrey Andreev, and Alexander Hramov

***Abstract*—Reconstructing hidden dynamics and forecasting nonlinear time series remain central challenges in machine learning and nonlinear system modeling. Next-generation reservoir computing (NG-RC) provides an efficient framework for these tasks, yet its standard formulation with uniform delays can be suboptimal for signals with multiple characteristic time scales. We propose NG-RC with multiscale delay selection (NGRC-MDS), in which irregular delay vectors are constructed directly from the temporal structure of the signal. The method combines empirical mode decomposition (EMD) with kernel density estimation (KDE) to extract dominant temporal scales and form compact, interpretable delay sets. Experiments on the adaptive Kuramoto network, the Lorenz system, and the Mackey–Glass system show that the proposed approach consistently matches or exceeds optimized NG-RC accuracy while requiring minimal tuning. In particular, it improves hidden-state recovery in multiscale dynamical systems and remains effective for both short-term and long-term forecasting of delayed chaotic dynamics. These results show that multiscale delay selection provides a practical extension of NG-RC for nonlinear systems with complex temporal organization.**

***Index Terms*—Delay embedding, hidden data recovery, multiscale delay selection, next-generation reservoir computing (NG-RC), reservoir computing (RC), time-series forecasting.**

I. Introduction

RECOVERING hidden or missing data remains a key challenge in many domains, including climatology, neuroscience, and medicine [1], [2], [3]. In recent years, machine learning approaches have become the dominant tool for addressing this problem, often demonstrating higher accuracy compared to traditional, including statistical, methods. Among these approaches, recurrent neural networks (RNNs) and their simplified variant, reservoir computing (RC), have attracted particular attention due to their computational efficiency in modeling and predicting dynamical systems [4]. This efficiency is due to the fact that only the output layer is trained, while the rest of the network remains fixed.

Received 15 August 2025; revised 9 April 2026; accepted 15 June 2026. This work was supported by Russian Science Foundation under Grant 24-68-00030. *(Corresponding author: Artem Badarin.)*

The authors are with the Research Institute of Applied Artificial Intelligence and Digital Solutions, Plekhanov Russian University of Economics, 115054 Moscow, Russia, and also with Saratov State University, 410012 Saratov, Russia (e-mail: Badarin.A.A@mail.ru; Andreevandrei1993@gmail.com; HramovAE@gmail.com).

Digital Object Identifier 10.1109/TNNLS.2026.3706608

The theoretical properties of RC models with linear output were studied in detail [5], showing that such models are universal approximators of causal transformations of stochastic time series. This supports their applicability to recovery problems where the structure of the input data may not be fully known or controlled [6].

Furthermore, theoretical support is provided by [7], which demonstrates that RC with a linear reservoir and linear output is equivalent to a vector autoregressive (VAR) model. When the output layer includes polynomial terms, the architecture becomes equivalent to a nonlinear VAR (NVAR) model, which corresponds to a truncated discrete Volterra series. Thus, RC can be interpreted as a method for constructing a fixed feature space based on time delays and their polynomial combinations.

Building on this equivalence, the next-generation reservoir computing (NG-RC) model was proposed, which eliminates the reservoir entirely [8]. Instead of evolving internal states, the model directly transforms the input signal into a delay-vector augmented with all monomials up to a specified polynomial order. This significantly simplifies the architecture, reduces the number of hyperparameters to essentially the choice of delays and the polynomial order, and ensures that each feature is interpretable, as it corresponds to a specific polynomial term of the input.

The use of polynomial features in the delay vector is supported by a solid theoretical foundation. As shown in [9], polynomial functionals form the basis of the Volterra decomposition and can yield high predictive accuracy in regression tasks involving dynamical systems. While universal approximation theoretically requires an infinite number of polynomial terms, in practice, low-order expansions are often sufficient to achieve high accuracy [8]. This makes polynomial representations a natural and effective choice for time-series reconstruction.

Since NG-RC constructs its feature space directly from delayed observations and their polynomial combinations, the selection of time delays becomes a central modeling decision rather than a secondary hyperparameter. More broadly, delay-coordinate reconstruction provides the classical framework for recovering hidden dynamics from partial observations. Beginning with Takens' embedding theorem [10] and its subsequent extensions to more general invariant sets [11], and multivariate observation functions [12], this line of work established sufficient conditions for the existence of faithful reconstructions but did not prescribe a universal procedure for selecting delays from finite noisy data.

This gap has motivated a substantial body of work on delay-selection methods. Early and still widely used approaches rely on uniform embeddings, in which a single lag is determined using statistical or information-theoretic criteria such as the autocorrelation function or the first minimum of mutual information [13], and subsequent coordinates are formed as integer multiples of this lag. However, later studies demonstrated that the uniform restriction is not universally justified. In particular, Nichkawde [14] showed that the first minimum of mutual information provides a well-motivated criterion only for selecting the first delayed coordinate, whereas each subsequent coordinate should be chosen so as to contribute information that is not redundant with respect to those already included.

In practice, different delay-selection methods optimize fundamentally different aspects of reconstruction quality, including statistical independence between delay coordinates, geometrical unfolding of the attractor, preservation of topological invariants, and predictive performance, which gives rise to the coexistence of both uniform and nonuniform embedding strategies [15]. The problem becomes especially consequential for multiscale dynamics, where a single fixed lag may fail to capture the temporal organization of the signal at all relevant scales. Krämeret al. [16] addressed this by formulating delay selection as a sequential construction of informative coordinates, combining the continuity statistic with the L-statistic to evaluate irrelevance, noise amplification, and geometric complexity. At the same time, Pan and Duraisamy [17] connected delay structure to the spectral properties of the signal, showing that for linear time-delayed models, the minimum number of delays required for exact recovery is determined by the sparsity of the Fourier spectrum, while both the number of delays and the sampling interval affect the numerical conditioning of the model. Their analysis suggests that the choice of delays should account for the temporal structure of the observed signal.

For signals with multiple characteristic time scales, the uniform delay assumption becomes particularly restrictive: a single fixed lag may fail to capture the temporal organization of the signal at all relevant scales, limiting the expressiveness of the polynomial feature space. This observation motivates the central goal of this work: to develop a principled method for selecting nonuniform delays in NG-RC that explicitly accounts for the dominant temporal scales of the target signal. To this end, we propose to identify these scales directly from the structure of the observed data and distribute delays accordingly, ensuring that both short- and long-range dynamics are represented without an excessive number of coordinates. As demonstrated in the experiments below, this structure-informed selection yields a substantial improvement in reconstruction accuracy compared to uniform delays in standard NG-RC. Moreover, the key parameters of the proposed method—namely, the amplitude threshold and the number of delays per time scale—remain physically interpretable and robust across different dynamical regimes, requiring no per-instance tuning.

II. Methods

A. Benchmark Systems Under Study

Benchmarking was performed for reconstruction and forecasting tasks, with forecasting considered in both one-step-ahead and long-term regimes. Reconstruction performance was evaluated on two systems: a high-dimensional network of coupled phase oscillators with adaptive interactions [3] and the classical Lorenz system [8]. The Mackey–Glass system was used as a standard benchmark for one-step-ahead prediction in nonlinear dynamics with delay. This design enables a task-specific assessment of the proposed approach across systems with different dimensionality and dynamical structure.

1) Adaptive Kuramoto Network: We model the first system using a network of Kuramoto phase oscillators, consisting of $N = 300$ oscillators. The dynamics of the system are described by the following equation [3]:

$$\dot{\phi}_i(t) = \omega_i + \lambda \sum_{j \neq i} w_{ij}(t) \sin\left(\phi_j - \phi_i\right) \tag{1}$$

where ϕ_i is the phase of the ith oscillator, ω_i is its natural frequency, $\lambda = 1$ and w_{ij} represent the time-dependent connection weight between oscillators i and j. We update the weights based on the following adaptive rule:

$$\dot{w}_{ij}(t) = p_{ij}(t) - \left(\sum_{k \neq i} p_{ik}(t)\right) w_{ij}(t) \tag{2}$$

where $P_{ij}(t)$ is the average phase correlation between oscillators i and j over a characteristic memory time $T_m = 15$

$$P_{ij}(t) = \frac{1}{T_m} \left| \int_{t-T_m}^{t} \exp^{i(\phi_i(t') - \phi_j(t'))} dt' \right|. \tag{3}$$

We solve the system numerically using the fourth-order Runge–Kutta method for the differential components and apply quadrature rules for the integral components, with a time step $\delta t = 0.01$ [18]. The initial conditions of ϕ_i and the natural frequency ω_i are generated randomly in a uniform distribution $[-\pi, \pi]$. Each oscillator initially interacts with $K = 20$ randomly selected neighboring nodes, with $w_{ij} = 1/K$. We first solve the system without adaptation, using a fixed coupling strength $w_{ij} = 1/K$, and introduce adaptation at $t = 2T_m$. It should be noted that this system was introduced in Ref. [19] and thoroughly investigated in works [20], [21], including studies on forecasting [22] and data recovery [3].

To describe the macroscopic dynamics of the first adaptive Kuramoto network, we use the concept of macroscopic signals [23]. We consider several macroscopic signals from different parts of the network. For this, we randomly categorize the oscillators into $M = 6$ groups (S_j, $j = 1, \ldots, M$) with an equal number of elements (i.e., $M_{\text{group}} = N/M = 50$ oscillators in each group). For all groups, the following condition is imposed on each pair of groups: $S_i \cap S_j = \varnothing$, where $i \neq j$.

The macroscopic signals for each group are defined as

$$X_j(t) = \frac{1}{M_{\text{group}}} \sum_{i \in S_j} \sin\left[\phi_i(t)\right]. \tag{4}$$

We note that one of the M macroscopic signals is randomly selected as the target, while the remaining $M-1$ signals are used as inputs, following [3].

2) Lorenz System: We use the classical Lorenz system [24] as a benchmark example, following the same parameters and integration methods as in [8] to ensure comparability.

The Lorenz system is defined as

$$\dot{x} = \sigma (y - x), \quad \dot{y} = x(\rho - z) - y, \quad \dot{z} = xy - \beta z \tag{5}$$

with parameters $\sigma = 10$, $\rho = 28$, and $\beta = 8/3$. We primarily focus on reconstructing the z-component of the Lorenz system from the x and y components, as it is known to be the most challenging to recover due to its nonlinear dependence on both inputs and is commonly used as a benchmark [8]. For completeness, we also evaluate the reconstruction of the x and y variables from the remaining two.

3) Mackey–Glass System: We use the Mackey–Glass system [25] as a benchmark for both short-term and long-term forecasting. It is a standard model for nonlinear delayed chaotic dynamics.

The Mackey–Glass system is defined as

$$\dot{x}(t) = \frac{\beta x(t-\tau)}{1 + x(t-\tau)^n} - \gamma x(t) \tag{6}$$

with the standard parameters $\tau = 17$, $\beta = 0.2$, $\gamma = 0.1$, and $n = 10$. Numerical integration was performed with a step $h = 0.2$ and initial condition $x_0 = 1.2$.

To ensure comparability with previous studies, the model settings were chosen based on experimental protocols reported in earlier works on chaotic time-series prediction [26], [27], [28]. These studies differ in the lengths of the training and testing segments, as well as in the integration step used to generate the time series.

To obtain a robust performance estimate, we considered ten different realizations constructed from one long chaotic trajectory using shifted windows. For each realization, the input sequence consisted of the signal values at time t, and the target sequence consisted of the values at time $t+1$. Since chaotic trajectories are highly sensitive to the selected region of the attractor, evaluating several train/test splits provides a more reliable estimate of prediction performance.

For long-term forecasting, the trajectory was generated by the fourth-order Runge–Kutta method with integration step $\delta t = 0.01$ and resampled to the effective step $\delta t_{\text{eff}} = 0.2$. The training set contained 10^4 samples. Evaluation was performed on 100 test segments of length 5×10^4, separated by gaps of 1000 samples. During training, the model was fit to predict the next sample from previous observations; during testing, forecasting was performed recursively, with each predicted value used as input for the subsequent step.

B. Next-Generation Reservoir Computing

For reconstruction and forecasting, we employ the NG-RC model proposed in [8]. NG-RC is a data-driven version of classical RC for time-series analysis and tasks such as signal recovery and forecasting. The key idea is to construct an augmented state vector from delayed input signals and their nonlinear combinations.

Let us now construct the augmented state vector. At time i, we have N_{in} input signals, denoted as $\mathbf{X}_i = [X_{i1}, X_{i2}, \ldots, X_{iN_{\text{in}}}]^T$. To account for the temporal dynamics, we introduce k state delays, each separated by a time interval Δt. The linear part of the augmented state vector is then formed by concatenating these delayed inputs as follows:

$$\mathbf{V}_{\text{lin}} = \mathbf{X}_i \oplus \mathbf{X}_{i-\Delta t} \oplus \mathbf{X}_{i-2\Delta t} \oplus \cdots \oplus \mathbf{X}_{i-k\Delta t} \tag{7}$$

where k represents the number of delays and Δt is the delay time.

The nonlinear component of the augmented state vector is defined as a nonlinear function that can be derived from the linear feature vector using polynomials. The original paper [8] proposes the following formulation for the nonlinear part of the augmented state vector:

$$\mathbf{V}_{\text{nonlin}}^{(p)} = \underbrace{\mathbf{V}_{\text{lin}} \lceil\otimes\rceil \mathbf{V}_{\text{lin}} \lceil\otimes\rceil \cdots \lceil\otimes\rceil \mathbf{V}_{\text{lin}}}_{p \text{ times}} \tag{8}$$

where p is the polynomial order that specifies the number of V_{lin} vectors generated by the polynomial expansion, and the operator $\lceil\otimes\rceil$ computes all unique pairwise products of the elements in these vectors. This enables the model to capture higher order interactions between the input signals.

Finally, the complete feature vector is formed by combining the linear and nonlinear features, along with a bias term

$$\Psi_i = 1 \oplus \mathbf{V}_{\text{lin}} \oplus \mathbf{V}_{\text{nonlin}}^{(p)}. \tag{9}$$

To reconstruct the target signal X_i^{target} at time i, we apply a simple linear output layer

$$\mathbf{X}_i^{\text{rec}} = \mathbf{W}_{\text{out}} \Psi_i \tag{10}$$

where $\mathbf{W}_{\text{out}}$ are the output weights.

The output weights W_{out} are determined by solving the ridge regression problem

$$\mathbf{W}_{\text{out}} = \mathbf{X}^{\text{target}} \Psi^T \left(\Psi \Psi^T + \alpha \mathbf{I}\right)^{-1} \tag{11}$$

where $\mathbf{I}$ is the identity matrix and α is the regularization parameter. For the Kuramoto network and the Lorenz system, α was set according to the corresponding reference studies used for comparison, yielding $\alpha = 10^{-6}$ and $\alpha = 0.05$, respectively [3], [8]. For the Mackey–Glass system, we used $\alpha = 10^{-5}$, as the comparison involved several studies with different setups, and this value was chosen as a representative weak regularization level.

C. Analysis of the Delay Representation

To better understand why different delay-selection strategies lead to different reconstruction performance, we analyzed the quality of the resulting feature representation using trustworthiness and feature–target mutual information.

Trustworthiness quantifies how well local neighborhood relations are preserved in the constructed representation [29], [30]

$$\begin{aligned} \mathrm{Tr} = 1 &- \frac{2}{N_s\, n_{\text{nb}} (2N_s - 3n_{\text{nb}} - 1)} \\ &\times \sum_{i=1}^{N_s} \sum_{j \in \mathcal{N}_i^{n_{\text{nb}}}} \max(0, r(i,j) - n_{\text{nb}}) \end{aligned} \tag{12}$$

where N_s is the number of samples, n_{nb} is the number of nearest neighbors, $\mathcal{N}_i^{n_{nb}}$ is the set of n_{nb} nearest neighbors of sample i in the constructed representation, and $r(i, j)$ is the rank of sample j when all samples are ordered according to their distance from sample i in the original space. Higher values of Tr indicate better preservation of local geometric structure. In our study, trustworthiness was computed using *scikit-learn* [31].

To quantify how informative the constructed features are with respect to the target signal, we computed the average mutual information between each feature and the target

$$\bar{I}\left(\Psi; X^{\text{target}}\right) = \frac{1}{D_{\text{eff}}} \sum_{m=1}^{D_{\text{eff}}} I\left(\Psi_m; X^{\text{target}}\right) \tag{13}$$

where Ψ_m is the mth component of the feature representation, D_{eff} is the number of nonconstant features, and $I(\Psi_m; X^{\text{target}})$ is the mutual information between this feature and the target signal. Higher values of $\bar{I}(\Psi; X^{\text{target}})$ indicate that, on average, the constructed features contain more information relevant to the target. Mutual information was estimated using *scikit-learn*, which is based on a nonparametric k-nearest-neighbor estimator [31], [32].

Thus, trustworthiness characterizes the geometric quality of the delay representation, whereas feature–target mutual-information quantifies its relevance for reconstruction.

D. Quality of Recovery

To evaluate the accuracy of hidden data recovery, we calculated the reconstruction error as follows:

$$\epsilon = \frac{\sum_i \left(X_i^{rec} - X_i^{\text{target}}\right)^2}{\sum_i \left(X_i^{\text{target}} - \overline{X^{\text{target}}}\right)^2} \tag{14}$$

where X^{rec} and X^{target} are the reconstructed and target signals, respectively.

For the one-step-ahead prediction task, we used RMSE and NRMSE, as these metrics are commonly reported in previous studies on chaotic time-series prediction [26], [27], [28]. They were defined as

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(X_i^{pred} - X_i^{\text{target}}\right)^2} \tag{15}$$

$$\text{NRMSE} = \frac{\text{RMSE}}{\sigma\left(X^{\text{target}}\right)} \tag{16}$$

where X^{pred} is the predicted signal, X^{target} is the target signal, N is the number of samples, and $\sigma(X^{\text{target}})$ is the standard deviation of the target signal.

For the long-term forecasting task, we additionally used the valid prediction time (VPT), which quantifies the time interval over which the predicted and true trajectories remain close under a fixed error threshold. Following [33], for $\eta = 0.9$ it was defined as:

$$n_{\text{VPT}} = \min\left\{ n : \frac{\|x_n - r_n\|}{\sqrt{\frac{1}{N}\sum_{k=0}^{N}\|x_k\|^2}} > \eta \right\} \tag{17}$$

$$\text{VPT} = \frac{n_{\text{VPT}}\,\delta t}{T_\Lambda} \tag{18}$$

where x_n and r_n denote the true and predicted states, respectively, $T_\Lambda = 1/\Lambda$ is the Lyapunov time, Λ is the largest Lyapunov exponent of the original dynamical system, and δt is the sampling step. Thus, VPT measures the forecast horizon in Lyapunov units.

E. Noise-Robustness Analysis

To assess the robustness of the proposed method to observational noise, we additionally analyzed the reconstruction of the Lorenz system in the presence of additive noise. Two noise types were considered: white Gaussian noise and flicker noise with power spectral density proportional to $1/f$.

For each realization, additive noise was applied to the observed input signals in both the training and test sets, as well as to the training target signal

$$\begin{aligned}\tilde{\mathbf{X}}(t) &= \mathbf{X}(t) + \boldsymbol{\eta}(t)\\ \tilde{X}^{\text{target}}(t) &= X^{\text{target}}(t) + \eta^{\text{target}}(t)\end{aligned} \tag{19}$$

where $\boldsymbol{\eta}(t)$ denotes the noise process, generated independently for each input and target time series. Reconstruction accuracy was evaluated with respect to the clean test target signal.

The noise level was controlled by the signal-to-noise ratio (SNR), varied from 0 to 100 dB with a step of 5 dB. For each SNR value, the noise variance was set according to

$$\sigma_\eta^2 = \frac{\sigma_x^2}{10^{\text{SNR}/10}} \tag{20}$$

where σ_x^2 is the variance of the clean signal.

The results were averaged over 10 realizations of the Lorenz system. We compared NG-RC with multiscale delay selection (NGRC-MDS) with two NG-RC baselines using regular delays, with the delay step selected either by the first minimum of the average mutual information or by the first minimum of the autocorrelation function. In both NG-RC baselines, the number of delays was fixed to 10 according to the noise-free reconstruction results. For NGRC-MDS, we used $n = 5$ and considered two threshold values: $\rho = -30$ dB and $\rho = -5$ dB.

III. Effect of Delay Parameters on the Reconstruction in NG-RC

We analyze the influence of delay-related parameters in NG-RC, focusing on how the reconstruction performance varies with time delay and number of delays. We also explore how these parameters relate to the internal temporal structure of the system.

A. Adaptive Kuramoto Network

We analyzed the influence of NG-RC parameters on signal reconstruction quality, taking into account factors such as the time delay Δt, the number of delays k, and the degree of nonlinearity [see Fig. 1(a) and (b)]. Fig. 1(a) shows how the reconstruction error depends on Δt and k for $p = 1$. We observed that the minimum reconstruction error (14) is achieved under the condition

$$\Delta t \times k = T_{\text{opt}} \tag{21}$$

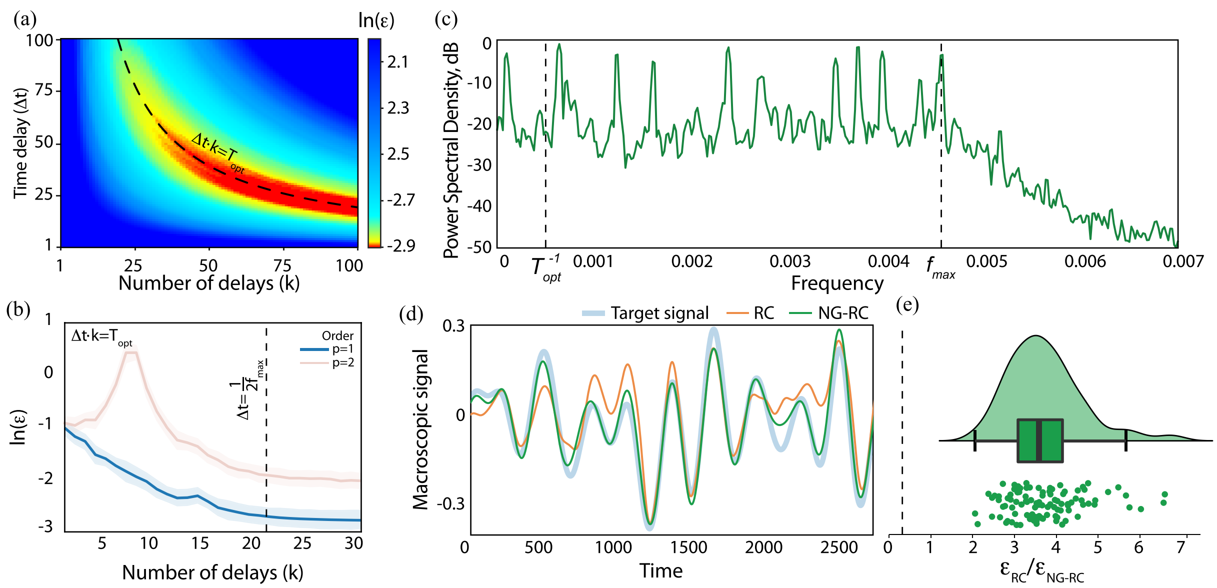

Fig. 1. Effect of delay parameters on NG-RC reconstruction performance for the adaptive Kuramoto network. (a) Logarithmic reconstruction error $\ln(\epsilon)$ dependent on time delay Δt and number of delays k ($p = 1$), averaged over 10 randomly formed sets of target macroscopic signals. The dashed line indicates the optimal memory depth: $\Delta t \times k = T_{opt}$. (b) Dependence of reconstruction error on the number of delays for linear ($p = 1$) and quadratic ($p = 2$) NG-RC, with memory depth fixed at $T_{\max}$. Shaded areas represent standard deviation over repetitions. The vertical dashed line corresponds to the Nyquist sampling limit: $\Delta t = 1/(2f_{max})$. (c) Averaged power spectral density of 10 target macroscopic signals. Vertical lines indicate the maximum frequency $f_{\max}$ and the frequency corresponding to the optimal memory depth T_{opt}^{-1}. (d) Example of signal reconstruction: The target signal (black), classical RC output (orange), and NG-RC output (green). (e) Boxplot of reconstruction error ratios $\epsilon_{\text{RC}}/\epsilon_{\text{NG-RC}}$ for 100 randomly formed macrosignals. Each point represents one reconstruction. Ratios above 1 indicate superior performance of NG-RC.

where T_{opt} is the optimal memory depth of the model, indicating how far back in time the model can look to reconstruct the current state. This relationship is indicated by the black dashed line as shown in Fig. 1(a).

Fig. 1(b) illustrates how the reconstruction error varies with the number of delays when $\Delta t \times k = T_{\text{opt}}$ is held constant. Increasing the number of delays initially leads to a pronounced decrease in error, and the error reaches its minimum at a certain optimal value k_{opt}. Beyond this point, further increases in the number of delays do not improve performance. A similar trend holds for the NG-RC with quadratic nonlinearity ($p = 2$), although for this nonlinear reservoir, the minimum reconstruction error of the macroscopic signal is notably higher than that of the linear NG-RC.

We examine how the model's memory depth relates to the spectrum of oscillation periods. The optimal depth, T_{opt}, lies within the band defined by the two dominant low-frequency peaks [Fig. 1(c)], matching the slowest characteristic periods. This memory depth is sufficient to capture the key dynamics, whereas extending it further adds mostly redundant information and decreases the model's ability to generalize.

We determine the optimal number of delays, k_{opt}, by using the condition $\Delta t \times k = T_{\text{memory}}$ [see Fig. 1(b) and (c)]. Our analysis shows that the optimal value of k_{opt} occurs when the delay duration

$$\Delta t = \frac{1}{2f_{\max}} \tag{22}$$

which is consistent with the Nyquist–Shannon sampling theorem.

This conclusion arises from examining the relationship between the delay duration and the system's dynamics. As we increase the number of delays while simultaneously reducing the delay duration along the line $\Delta t \times k = T_{\text{opt}}$, the reconstruction error decreases. However, this decrease is limited: once Δt reaches $1/2f_{\max}$, the error stabilizes, and further reductions in delay duration do not lead to significant improvements. This behavior stems from the fact that shorter delay durations enable the NG-RC model to capture high-frequency components critical for accurate reconstruction. Using delays longer than $1/2f_{\max}$ would violate the Nyquist–Shannon theorem, leading to information loss and an increase in reconstruction error.

We compared the errors in reconstructing the macroscopic signal using NG-RC with optimal parameters against the errors produced by an optimized classical RC model from an earlier study [3]. The results are presented in Fig. 1(e) for 100 randomly formed sets of nonoverlapping sets of oscillators. In all tested cases, NG-RC outperforms classical RC in terms of reconstruction accuracy. The average error reduction reaches a factor of 3.2, and in specific cases, the improvement approaches a sevenfold decrease. Details on the parameter optimization procedure for the classical RC model can be found in [3].

Fig. 1(d) provides a further illustration of this performance, depicting both the original macroscopic signal (gray line) and the NG-RC reconstructed signal (green line). The result presented was obtained using the following NG-RC parameters: a delay duration $\Delta t = 100$, $k = 21$ delays, and a polynomial order $p = 1$. As evident in the figure, the reconstructed signal faithfully replicates the temporal dynamics of the original, demonstrating phase coherence between the true and reconstructed waveforms. Nevertheless, a more granular analysis reveals subtle errors in amplitude reconstruction, particularly

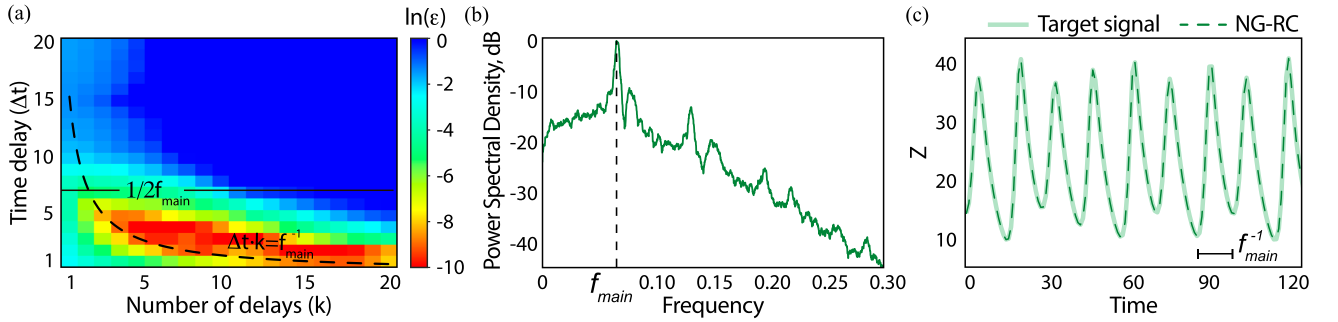

Fig. 2. Effect of delay parameters on NG-RC reconstruction performance for the Lorenz system. (a) Logarithmic reconstruction error, $\ln(\varepsilon)$, as a function of Δt and k for $p = 2$. The dashed line indicates the condition $\Delta t \times k = f_{\text{main}}^{-1}$, where f_{main} denotes the dominant spectral component. The horizontal solid line corresponds to $\Delta t = 1/2 f_{\text{main}}$, representing the Nyquist–Shannon limit required to adequately resolve the dominant frequency. (b) Averaged power spectral density of the target z-component. The vertical line marks the dominant frequency f_{main}. (c) Example of NG-RC reconstruction (dashed) compared to the target signal (solid) for $\Delta t = 4$, $k = 5$ and $p = 2$.

in the vicinity of local extrema. In these regions, the reconstructed signal exhibits a tendency to either underestimate or overestimate the true value.

Overall, these findings demonstrate that NG-RC with optimally selected parameters substantially outperforms the classical RC model in reconstructing macroscopic signals, resulting in lower errors and more accurately capturing the essential features of the underlying dynamics.

B. Lorenz Systems

To investigate how the intrinsic time scales of the system relate to the choice of delay parameters in NG-RC, we analyzed the Lorenz system using quadratic nonlinearity ($p = 2$), following the methodology introduced in the original NG-RC study [8]. We analyzed the influence of NG-RC parameters on signal reconstruction quality [see Fig. 2(a)].

Unlike systems with well-defined oscillation frequencies, the Lorenz attractor exhibits quasiperiodic dynamics. Therefore, we estimated only the dominant spectral component f_{main} from the power spectral density of the target signal, as shown in Fig. 2(b). Following the approach used in the Kuramoto case, we compared this time scale with the total memory depth of the model, defined as $\Delta t \times k$. Fig. 2(a) shows the reconstruction error as a function of the time delay Δt and the number of delays k, with the dashed line representing the condition $\Delta t \times k = f_{\text{main}}^{-1}$. The horizontal solid line corresponds to $\Delta t = 1/2 f_{\text{main}}$, representing the Nyquist–Shannon limit required to adequately resolve the dominant frequency.

Although the point of minimum reconstruction error does not lie exactly on the line $\Delta t \times k = f_{\text{main}}^{-1}$, it is located very close to it. This suggests that aligning the model's memory depth with the characteristic time scale of the signal leads to near-optimal performance. Moreover, the results clearly show that increasing the delay duration Δt beyond the Nyquist limit ($\Delta t > 1/2 f_{\text{main}}$) leads to a sharp increase in reconstruction error, emphasizing the importance of preserving sufficient temporal resolution.

Fig. 2(c) presents an example reconstruction using the optimal parameter combination. The NG-RC model successfully captures the main structure and variability of the signal, achieving a low reconstruction error.

These results confirm that, even in systems with quasiperiodic dynamics, the characteristic time scales estimated directly from the signal provide a meaningful basis for choosing delay parameters. This further highlights the strong connection between the temporal properties of the target signal and the configuration of the NG-RC model.

IV. NG-RC With Multiscale Delay Selection

We propose a modification of the NG-RC model that constructs delay vectors based on the intrinsic time scales of the target signal. Unlike standard approaches with uniform delays, our method analyzes the signal's temporal structure to extract dominant periods and generates an irregular set of delays that reflects its multiscale dynamics. This allows the model to capture both fast and slow components more effectively while relying on a small number of intuitive and robust parameters.

A. Irregular Delay-Vector Construction

To identify the dominant time scales, we first perform empirical mode decomposition (EMD, [34], [35]), which represents the signal as a sum of L intrinsic mode functions (IMFs; see Step I in Fig. 3) and a residual trend $r(t)$

$$X^{target}(t) = \sum_{k=1}^{L} \text{IMF}_l(t) + r(t) \tag{23}$$

where each $\text{IMF}_l(t)$ corresponds to a narrowband component oscillating within a well-defined frequency range. Unlike traditional methods based on projection onto fixed bases (e.g., Fourier), EMD adaptively decomposes the signal using its local extrema structure, enabling scale separation without imposing harmonic assumptions. This property is particularly beneficial for analyzing nonlinear and nonstationary signals whose oscillatory components exhibit temporal asymmetry and frequency modulation, making them poorly approximated by sums of harmonic signals. The number of retained IMFs controls the range of temporal scales included in the analysis. Since successive EMD modes generally correspond to progressively slower oscillations [34], increasing their number allows longer characteristic time scales to be taken into

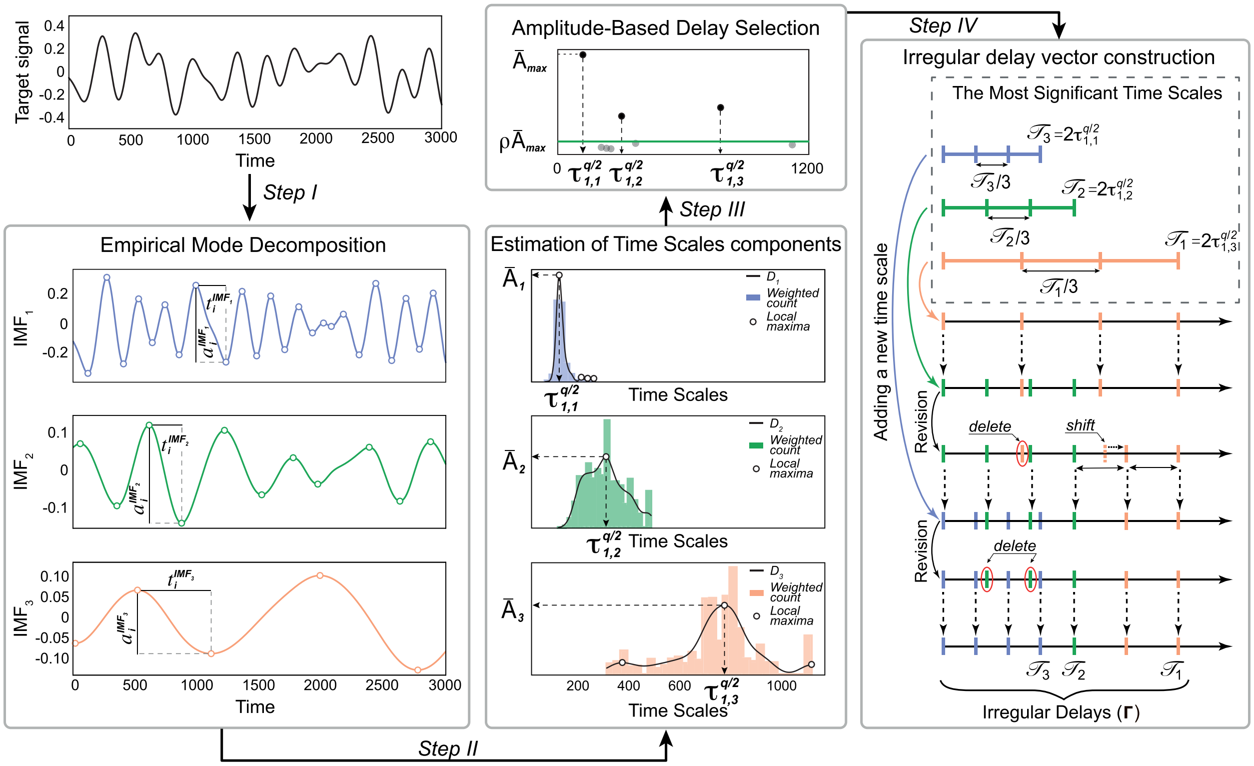

Fig. 3. Schematic overview of the proposed method for constructing an irregular delay vector. Step I: The target signal is decomposed into IMFs using EMD. Step II: For each IMF$_l$, quasi-half-periods $t_i^{\mathrm{IMF}_l}$ and amplitudes $a_i^{\mathrm{IMF}_l}$ are extracted. Each time scale is assigned a weight $w_i^{\mathrm{IMF}_l} = a_i^{\mathrm{IMF}_l}$. KDE is applied to the set of $t_i^{IMF_l}$ to obtain a continuous density function $D_l(t^{\mathrm{IMF}_l})$ over time scales. Step III: Dominant time scales $\tau_{i,l}^{q/2}$ are selected by amplitude thresholding ρ and scaled to obtain full periods $\mathcal{T}$. Step IV: The final irregular delays $\mathbf{\Gamma}$ are constructed from $\mathcal{T}$ using the procedure described in Fig. 4, with a resolution parameter $n = 3$. The algorithm ensures that each period $\mathcal{T}_i \in \mathcal{T}$ is covered with resolution no greater than $\mathcal{T}_i/n$ while striving for uniform coverage and reducing redundancy.

account and thus increases the effective memory depth of the multiscale delay vector. These time scales should, however, be interpreted only approximately, as EMD provides an adaptive rather than exact spectral decomposition. As shown below, the reconstruction error saturates after inclusion of only a few modes, indicating that the dominant temporal organization of the signal is captured already at low IMF numbers.

Next, for each mode, we identify all extrema $t_{i,\mathrm{extr}}^{\mathrm{IMF}_l}$ and compute the quasi-half-period $t_i^{\mathrm{IMF}_l}$ and the corresponding amplitude $a_i^{\mathrm{IMF}_l}$ as

$$\begin{aligned} t_i^{\mathrm{IMF}_l} &= t_{i+1,\mathrm{extr}}^{\mathrm{IMF}_l} - t_{i,\mathrm{extr}}^{\mathrm{IMF}_l} \\ a_i^{\mathrm{IMF}_l} &= \left|\mathrm{IMF}_l\left(t_{i+1,\mathrm{extr}}^{\mathrm{IMF}_l}\right) - \mathrm{IMF}_l\left(t_{i,\mathrm{extr}}^{\mathrm{IMF}_l}\right)\right|. \end{aligned} \tag{24}$$

This formulation enables us to robustly characterize multiscale temporal structures in the presence of waveform distortion or frequency variability.

As the next step, we quantitatively describe the temporal organization of the signal (see Step II in Fig. 3). This is achieved by constructing a distribution of quasi-half-periods $t_i^{\mathrm{IMF}_l}$ extracted from each lth mode. Each quasi-half-period is assigned a weight $w_i^{\mathrm{IMF}_k} = a_i^{\mathrm{IMF}_k}$, thereby incorporating the amplitude of the corresponding oscillation.

We then apply kernel density estimation (KDE) with a Gaussian kernel to the set of weighted quasi-half periods, using the bandwidth selected according to Scott's rule [36]. This choice provides data-adaptive smoothing that accounts for both the sample size and the dispersion of the extracted quasi-half periods, which may differ substantially across IMFs. As a result, we obtain a continuous density function $D_l(t^{\mathrm{IMF}_l}; w_1^{\mathrm{IMF}_l}, \ldots, w_{N_l}^{\mathrm{IMF}_l})$ over time scales, where each $w_i^{\mathrm{IMF}_l}$ is the weight associated with the ith of the N_l extracted quasi-half periods. For brevity, we omit the explicit notation of weights in the arguments of D_l in the remainder of the text. Thus, D_l should be interpreted as an amplitude-weighted occurrence density over time scales.

To enable consistent thresholding and accurate comparison across modes, the estimated KDE distributions are normalized such that their peak value corresponds to the average oscillation amplitude of the respective mode which is calculated as

$$\bar{A}_l = \frac{1}{N_l}\sum_{i=1}^{N_l} a_i^{\mathrm{IMF}_l}. \tag{25}$$

This normalization ensures that modes with higher average amplitudes proportionally influence the identification and selection of dominant temporal scales.

After normalization, the positions of local maxima in each D_l are identified and merged into a unified set of dominant time scales, denoted as $\tau_{i,l}^{q/2}$ (see Step III in Fig. 3). For each dominant time scale, we evaluate its normalized density value $D_l(\tau_{i,l}^{q/2})$, taken from the corresponding KDE function. The dominant time scales are then filtered by applying a

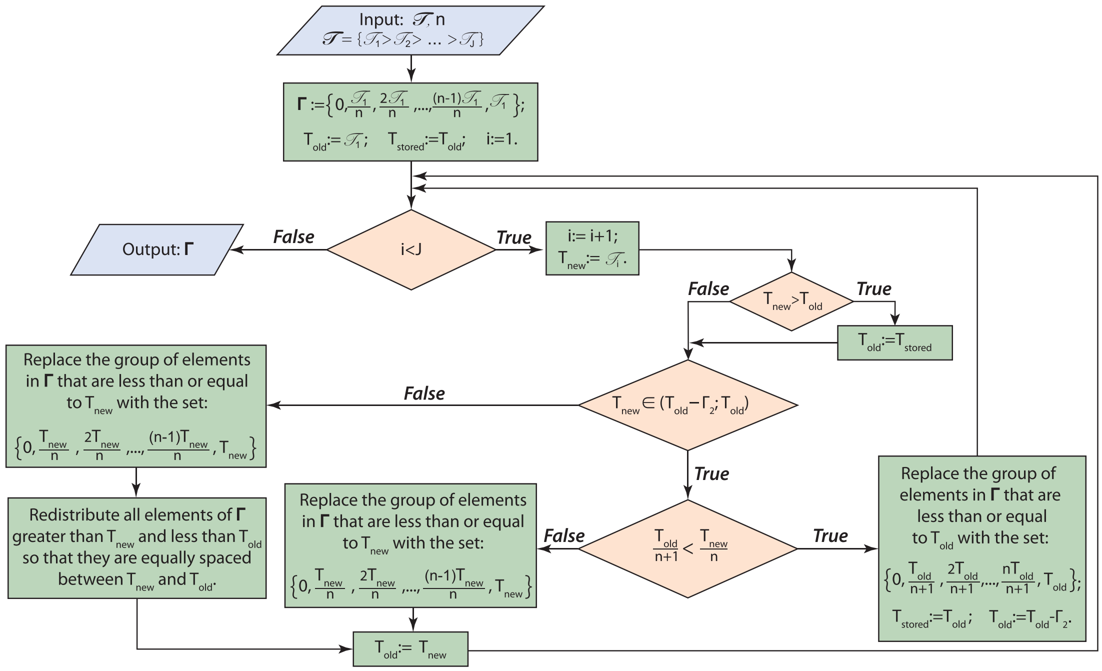

Fig. 4. Block diagram of the algorithm for constructing the irregular delays $\mathbf{\Gamma}$ from a set of dominant time scales $\mathcal{T}$. Each time scale $\mathcal{T}_i$ is sequentially inserted with resolution n, ensuring coverage with a step size no greater than $\mathcal{T}_i/n$ while avoiding redundancy through replacement and redistribution steps.

global amplitude threshold relative to the maximum average amplitude across all modes

$$\tau_{i,l}^{q/2} \in \mathcal{T}^{q/2} \quad \text{if} \quad D_l\left(\tau_{i,l}^{q/2}\right) \geq \rho \cdot \bar{A}_{\max} \tag{26}$$

where $\bar{A}_{\max} = \max_l \bar{A}_l$ and $\mathcal{T}^{q/2}$ are the final set of the most significant quasi-half periods, and $\rho \in [0, 1]$ is a tunable threshold. This step ensures that only time scales with sufficiently high amplitude significance are used for constructing the multiscale delay vector.

In the final step, the selected quasi-half-periods $\mathcal{T}^{q/2}$ are scaled by a factor of two to form a set of dominant full periods $\mathcal{T}$ (The set $\mathcal{T}$ was additionally constrained so that at least ten cycles of the longest selected period were present within the training sequence, providing a minimal statistical basis for reliable estimation of the corresponding temporal scale.). Based on this set and a resolution parameter n, which defines the number of delays per period, the final delay vector $\mathbf{\Gamma}$ is constructed using the procedure described in Fig. 4 and illustrated as an example in Step IV of Fig. 3. The algorithm ensures that each period $\mathcal{T}_i \in \mathcal{T}$ is covered with resolution no greater than $\mathcal{T}_i/n$ while striving for uniform coverage and reducing redundancy.

The linear part of the augmented state vector is formed analogously to (7), but using the irregular delays $\mathbf{\Gamma}$

$$\mathbf{V}_{\text{lin}}^{\text{irr}} = \mathbf{X}_{i-\Gamma_1} \oplus \mathbf{X}_{i-\Gamma_2} \oplus \mathbf{X}_{i-\Gamma_3} \oplus \cdots \oplus \mathbf{X}_{i-\Gamma_{N_\Gamma}} \tag{27}$$

where N_Γ is the number of delays in the set $\mathbf{\Gamma}$.

A nonlinear transformation is then applied to $V_{\text{lin}}^{\text{irr}}$ following the same procedure as in the regular-delay model [see (8)], resulting in the final augmented state vector Ψ.

B. Application to the Adaptive Kuramoto Network

We first analyzed the influence of the number of retained IMF modes on NGRC-MDS performance for the adaptive Kuramoto network. For this analysis, the parameters were fixed at $\rho = -30$ dB and $n = 10$, so that the extracted dominant time scales were covered with sufficiently high resolution. As shown in Fig. 5(a), the reconstruction error decreases substantially when the number of selected IMF modes increases from 1 to 3 and then saturates. This indicates that the dominant temporal scales relevant for reconstruction are already captured by a small number of EMD components.

We then examined how reconstruction accuracy depends on the NGRC-MDS parameters ρ and n for different numbers of retained IMFs. The corresponding error maps are shown in Fig. 5(b). For both IMF selections, NGRC-MDS achieves near-optimal reconstruction accuracy over a broad region of the parameter space, indicating good robustness. Increasing n initially improves performance by providing denser coverage of the extracted time scales, whereas further increases yield only moderate gains. At the same time, if ρ becomes too high, some relevant time scales fall below the threshold and are excluded from the delay set, which leads to a deterioration in reconstruction accuracy. The maps also show that using four IMF modes slightly broadens the high-accuracy region compared to three modes Fig. 5(b).

Finally, we compared NGRC-MDS with regular-delay NG-RC baselines as a function of the number of delays [Fig. 5(c)]. For the NG-RC model, the number of delays is defined by k, and for each value of k, the optimal delay step, denoted as $\Delta t^*(k)$, is empirically determined by minimizing the reconstruction error according to $\Delta t^*(k) = \arg\min_{\Delta t} \varepsilon(k, \Delta t)$. This optimization procedure ensures that NG-RC operates under its best-performing configuration for each specified number

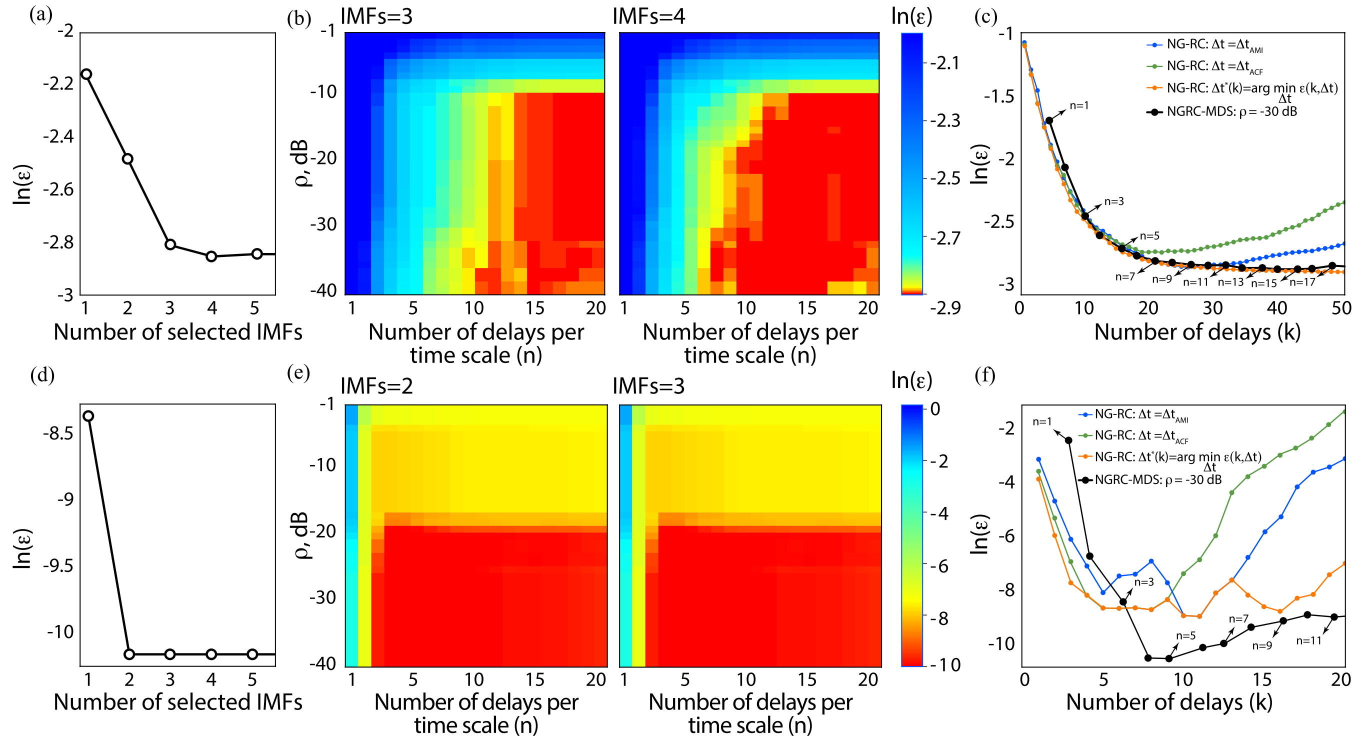

Fig. 5. Influence of NGRC-MDS parameters and retained IMF modes on reconstruction performance for the adaptive Kuramoto network (a)–(c) and the Lorenz system (d)–(f). (a) and (d) Average logarithmic reconstruction error $\ln(\varepsilon)$ as a function of the number of selected IMF modes for $\rho = -30$ dB and $n = 10$. (b) and (e) Reconstruction error $\ln(\varepsilon)$ as a function of the threshold ρ and the number of delays per time scale n, shown for different numbers of retained IMF modes. (c) and (f) Comparison of the average logarithmic reconstruction error for NG-RC with $\Delta t = \Delta t_{\mathrm{AMI}}$, NG-RC with $\Delta t = \Delta t_{\mathrm{ACF}}$, NG-RC with optimized delay step $\Delta t^*(k) = \arg\min_{\Delta t} \varepsilon(k, \Delta t)$, and NGRC-MDS with $\rho = -30$ dB. For NGRC-MDS, the number of retained IMF modes was fixed to 4 in (c) and 2 in (f); the values of n are indicated near the corresponding points.

of delays. We also included two signal-driven baselines with fixed regular delays, where the delay step was selected from the first minimum of the autocorrelation function, Δt_{ACF}, or from the first minimum of the average mutual information, Δt_{AMI}. On the contrary, for NGRC-MDS the number of delays is determined automatically for each parameter pair (ρ, n) and may vary across realizations. To ensure comparability, each point on the NGRC-MDS curve represents the average number of delays and the corresponding average reconstruction error over ten randomly generated macroscopic signals.

As shown in Fig. 5(c), NGRC-MDS remains competitive with the fully optimized NG-RC model and clearly outperforms the regular-delay baselines based on ACF and AMI when the number of delays increases. This result confirms that nonuniform delay sets aligned with the intrinsic multiscale structure of the target dynamics provide a robust alternative to regular embeddings.

C. Application to the Lorenz System

We performed the same analysis for the Lorenz system [see Fig. 5(d)–(f)]. First, we examined how the number of retained IMF modes affects the reconstruction accuracy of NGRC-MDS. For this analysis, the parameters were fixed at $\rho = -30$ dB and $n = 10$, so that the extracted dominant time scales were covered with sufficiently high resolution. As shown in Fig. 5(d), the reconstruction error decreases sharply when the number of selected IMF modes increases from 1 to 2 and then remains nearly unchanged. This indicates that only a small number of EMD components is sufficient to capture the dominant temporal structure relevant for reconstruction.

We then analyzed the reconstruction error across the NGRC-MDS parameter space for two and three retained IMF modes [Fig. 5(e)]. Similar to the Kuramoto case, high accuracy is achieved over a broad range of parameter values. Increasing n initially improves reconstruction accuracy by providing denser coverage of the extracted time scales. At the same time, when ρ becomes too high, some relevant time scales are excluded from the multiscale delay set, which leads to a deterioration in reconstruction accuracy. The comparison of the two maps shows that increasing the number of retained modes from two to three produces only minor changes in the high-accuracy region, consistent with the saturation observed in Fig. 5(d).

Fig. 5(f) presents the same comparison with regular-delay NG-RC baselines as in the Kuramoto case, including the optimized configuration with $\Delta t^*(k)$, as well as fixed-delay baselines based on Δt_{ACF} and Δt_{AMI}. As shown in Fig. 5(f), NGRC-MDS demonstrates near-optimal reconstruction accuracy and outperforms the regular-delay baselines based on both ACF and AMI over a broad range of delay numbers. The minimum reconstruction error is reached at intermediate values of n, whereas further increases in the number of delays lead to a gradual degradation in performance. This effect is likely related to the limited size of the training set, which contains only 500 data points (including warm-up, which is equal to

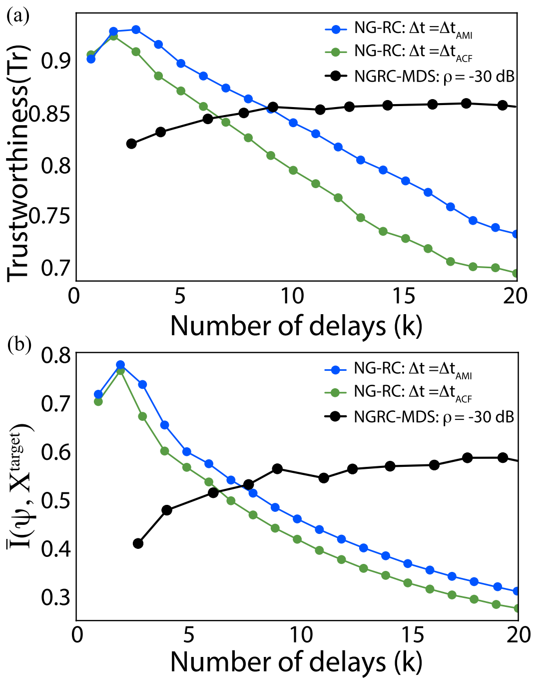

Fig. 6. Analysis of the feature representation quality for the Lorenz system. (a) Trustworthiness of the constructed feature representation as a function of the number of delays for NG-RC with the delay step selected by the first minimum of average mutual information, NG-RC with the delay step selected by the first minimum of the autocorrelation function, and NGRC-MDS with a threshold of $\rho = -30$ dB and IMFs = 2. (b) Average feature–target mutual information, $\bar{I}(\Psi; X^{\text{target}})$, as a function of the number of delays for the same models.

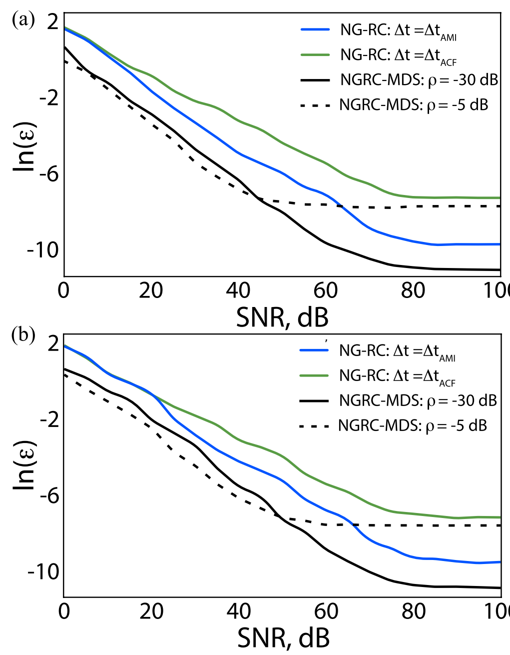

Fig. 7. Robustness of reconstruction to additive noise in the Lorenz system. (a) Average logarithmic reconstruction error, $\ln(\varepsilon)$, versus SNR for additive white Gaussian noise. (b) Average logarithmic reconstruction error, $\ln(\varepsilon)$, versus SNR for additive flicker noise. Curves are shown for NG-RC with ten delays and delay step Δt_{AMI}, NG-RC with ten delays and delay step Δt_{ACF}, and NGRC-MDS with thresholds $\rho = -30$ dB and $\rho = -5$ dB.

$k\Delta t$ for NG-RC and $\max(\mathbf{\Gamma})$ for NGRC-MDS), analogously to [8]. However, the reconstruction error remains lower than that of NG-RC with regular delays, and close to or below that of the fully optimized NG-RC model.

To better understand why NGRC-MDS provides more accurate reconstruction, we additionally analyzed the quality of the resulting feature representation for the Lorenz system (Fig. 6). In this comparison, the regular-delay NG-RC baselines were evaluated over different numbers of delays, whereas NGRC-MDS was constructed using $\rho = -30$ dB, two IMFs, and different n. Fig. 6(a) shows the trustworthiness of the constructed feature space as a function of the number of delays. For the regular-delay NG-RC baselines, trustworthiness gradually decreases as the number of delays increases, indicating that the local neighborhood relations present in the observed dynamics are preserved less accurately in the resulting representation. On the contrary, for NGRC-MDS the trustworthiness remains high and nearly constant over a broad range of delay numbers. This indicates that the multiscale delay vector preserves the local structure of the observed dynamics more faithfully.

A similar tendency is observed for the average feature–target mutual information shown in Fig. 6(b). For the ACF- and AMI-based regular embeddings, this quantity decreases as the number of delays increases, indicating that newly added delayed coordinates contribute progressively less useful information for reconstructing the target variable. On the contrary, NGRC-MDS yields consistently higher values of $\bar{I}(\Psi; X^{\text{target}})$, showing that its constructed features remain more informative with respect to the target signal. This suggests that the multiscale delay-selection procedure not only preserves the local geometry more effectively but also places the coordinates at temporal positions that better capture the informative dependencies between the observed variables and the reconstruction target.

Taken together, these results show that the advantage of NGRC-MDS is not limited to lower reconstruction error. The method also yields a feature representation that is both geometrically more consistent and more informative for the recovery task. This explains why multiscale delays aligned with the dominant temporal scales of the system are better suited for reconstruction than uniform embeddings based on a single fixed delay step.

D. Robustness to Additive Noise

We additionally evaluated the robustness of the proposed approach to additive noise using the Lorenz system (Fig. 7). The comparison included NGRC-MDS and two regular-delay NG-RC baselines with delay steps selected from the first minimum of the average mutual information, Δt_{AMI}, and the autocorrelation function, Δt_{ACF}. For the regular-delay baselines, the number of delays was fixed to 10 according to the noise-free reconstruction analysis. For NGRC-MDS, we used $n = 5$ and two threshold values, $\rho = -30$ dB and $\rho = -5$ dB.

Fig. 7 shows the average logarithmic reconstruction error as a function of the SNR for additive white Gaussian noise and additive flicker noise. In both cases, the error decreases with increasing SNR. At low SNR, the best performance is achieved for $\rho = -5$ dB, whereas at high SNR the lowest error is obtained for $\rho = -30$ dB. Thus, under strong noise a higher threshold is beneficial because it suppresses weak noise-contaminated oscillatory components, while under weak noise a lower threshold is preferable because it preserves more informative temporal scales.

Compared with the regular-delay baselines, NGRC-MDS yields lower reconstruction error across the full SNR range for both noise types. This indicates that the proposed multiscale delay selection remains effective under both uncorrelated and temporally correlated noise.

V. Application to Forecasting of the Mackey–Glass System

To further evaluate the generalizability of the proposed approach, we applied NGRC-MDS to forecasting of the Mackey–Glass system. In contrast to the previous sections, which focused on hidden-state reconstruction, this part addresses prediction in a delayed dynamical system. The Mackey–Glass equation is infinite-dimensional due to the delayed feedback term and exhibits complex chaotic behavior, which makes it a standard benchmark for nonlinear time-series forecasting.

A. Short-Term Forecasting

We first considered the short-term one-step-ahead forecasting regime in order to compare NGRC-MDS with existing forecasting approaches under standard benchmark conditions.

Table I illustrates the comparison of NGRC-MDS with the regular-delay NG-RC model and previously published approaches under experimental settings adopted from recent studies on chaotic time-series prediction [26], [27], [28]. In all three benchmark settings, NGRC-MDS yields lower prediction error than NG-RC. Moreover, for sufficiently large values of n, the proposed method achieves the best RMSE and NRMSE values among all compared approaches, including recurrent, convolutional, Transformer-based, and reservoir-computing models reported in these studies. Thus, multiscale delay selection remains effective not only relative to the regular-delay NG-RC baseline but also in comparison with a broad set of recent forecasting methods.

Table II summarizes the dependence of prediction accuracy and computational cost on the training-set size. As the number of training samples increases, both NG-RC and NGRC-MDS improve their prediction accuracy, but NGRC-MDS consistently remains more accurate. The relative improvement in NRMSE, measured as $\epsilon_{\text{rel}} = \text{NRMSE}_{\text{NG-RC}}/\text{NRMSE}_{\text{NGRC-MDS}}$, increases from 12.10 for $N_{\text{train}} = 10^3$ to 35.60 for $N_{\text{train}} = 10^5$. The improvement is also accompanied by a reduction in computational cost. At $N_{\text{train}} = 10^5$, the delay-selection time is reduced by 67.2% (4.2878 versus 1.4060 s), and the model training time by 30.0% (0.4577 versus 0.3202 s).

Overall, the Mackey–Glass results show that the proposed multiscale delay-selection strategy generalizes beyond the hidden-state reconstruction task considered above. Even for one-step-ahead prediction in a delayed infinite-dimensional chaotic system, NGRC-MDS preserves its advantage over the regular-delay baseline and remains competitive with recently published forecasting models.

TABLE I
Benchmarking on the Mackey–Glass One-Step Forecasting

	Model	RMSE	NRMSE
Fu (2022) [26]	DTIGNet	6.05×10^{-4}	—
	TI-LSTM	1.29×10^{-3}	—
	T-Inception	1.67×10^{-3}	—
	DGRU	1.17×10^{-3}	—
	CNN	2.41×10^{-3}	—
	Transformer	2.07×10^{-3}	—
	NG-RC ($k = 10$)	2.94×10^{-4}	1.30×10^{-3}
	NGRC-MDS ($n = 5$)	1.55×10^{-4}	6.87×10^{-4}
	NGRC-MDS ($n = 10$)	$\mathbf{3.30 \times 10^{-5}}$ *	$\mathbf{1.44 \times 10^{-4}}$ *
	NGRC-MDS ($n = 15$)	$\mathbf{2.30 \times 10^{-5}}$ **	$\mathbf{1.01 \times 10^{-4}}$ **
	NGRC-MDS ($n = 20$)	$\mathbf{2.20 \times 10^{-5}}$ ***	$\mathbf{9.90 \times 10^{-5}}$ ***
Jiang (2025) [27]	TDR	—	1.17×10^{-2}
	DDR	—	9.27×10^{-3}
	DTDR	—	9.17×10^{-3}
	DATDR	—	5.26×10^{-3}
	LD-TDR	—	5.79×10^{-3}
	LS-TDR	—	7.23×10^{-3}
	NG-RC ($k = 11$)	6.97×10^{-4}	2.80×10^{-3}
	NGRC-MDS ($n = 5$)	3.94×10^{-4}	1.59×10^{-3}
	NGRC-MDS ($n = 10$)	$\mathbf{8.20 \times 10^{-5}}$ *	$\mathbf{3.28 \times 10^{-4}}$ *
	NGRC-MDS ($n = 15$)	$\mathbf{4.80 \times 10^{-5}}$ **	$\mathbf{1.94 \times 10^{-4}}$ **
	NGRC-MDS ($n = 20$)	$\mathbf{4.50 \times 10^{-5}}$ ***	$\mathbf{1.80 \times 10^{-4}}$ ***
Liu (2026) [28]	Multi-stage Chua's RC	—	2.39×10^{-3}
	NG-RC ($k = 21$)	3.44×10^{-3}	1.51×10^{-2}
	NGRC-MDS ($n = 5$)	1.01×10^{-3}	4.44×10^{-3}
	NGRC-MDS ($n = 10$)	$\mathbf{1.55 \times 10^{-4}}$ *	$\mathbf{6.80 \times 10^{-4}}$ *
	NGRC-MDS ($n = 15$)	$\mathbf{8.30 \times 10^{-5}}$ **	$\mathbf{3.65 \times 10^{-4}}$ **
	NGRC-MDS ($n = 20$)	$\mathbf{7.50 \times 10^{-5}}$ ***	$\mathbf{3.30 \times 10^{-4}}$ ***

For NG-RC, the delay step was determined by AMI and the optimal number of delays was selected. For NGRC-MDS, $\rho = -30$ dB and three IMF modes were used. The three best results within each article block are marked by *, **, and ***.

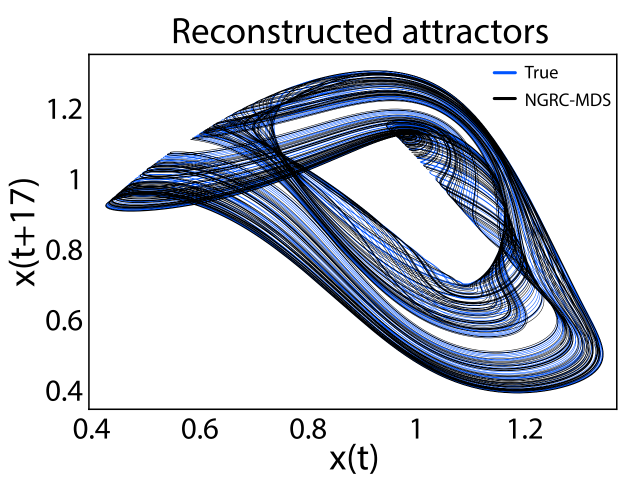

Fig. 8. Comparison of reconstructed attractors of the Mackey–Glass system in the $(x(t), x(t+17))$ plane for the true trajectory (blue) and the NGRC-MDS prediction (black).

B. Long-Term Forecasting

For long-term forecasting, NGRC-MDS was evaluated using polynomial order $p = 3$ and multiscale delay-selection parameters $n = 5$, $\rho = -30$ dB, and IMFs = 3. This choice was made because the inclusion of cubic terms yielded more accurate and more stable forecasts than the lower order configuration

TABLE II
COMPARISON OF ACCURACY AND COMPUTATION TIME OF NGRC-MDS AND NG-RC ACROSS TRAINING-SET SIZES

	NGRC-MDS							NG-RC				
N_{train}	k	t_{delays}	t_{train}	t_{full}	RMSE	NRMSE	ϵ_{rel}	k	t_{delays}	t_{train}	t_{full}	t_{search}
10^3	20.3	0.0071	0.0113	0.0184	3.81×10^{-4}	1.67×10^{-3}	12.10	11	0.0355	0.0017	0.0373	0.0448
10^4	25.7	0.0492	0.0457	0.0949	1.21×10^{-4}	5.33×10^{-4}	24.49	23	0.3844	0.0351	0.4195	0.6332
10^5	27.2	1.4060	0.3202	1.7262	8.80×10^{-5}	3.84×10^{-4}	35.60	32	4.2878	0.4577	4.7455	9.1713

Here, k denotes the mean number of delays for NGRC-MDS ($\rho = -30, n = 10$, IMFs=3) and the optimal number of delays for NG-RC. For NG-RC, $\Delta t = \Delta t_{\text{AMI}}$, where Δt_{AMI} was obtained from the first minimum of the average mutual information using a nearest-neighbor estimator. t_{delays} is the delay-selection time, t_{train} is the model training time, $t_{full} = t_{delays} + t_{train}$, t_{search} is the cumulative time required to identify k_{optimal}, including delay estimation and training of all models with the number of delays from 1 to k_{optimal}, and $\epsilon_{rel} = \text{NRMSE}_{\text{NG-RC}}/\text{NRMSE}_{\text{NGRC-MDS}}$. All time values are reported in seconds. The computations were performed on an Apple M1 Max processor.

TABLE III
COMPARISON OF LONG-TERM FORECASTING PERFORMANCE ON THE MACKEY–GLASS SYSTEM

Method	VPT (Lyapunov times)
NGRC-MDS	**11.3**
Hybrid RC-NGRC [39]	5
Infinite-dimensional NG-RC [40]	8.3
Asymmetrically connected reservoir networks [41]	8
Physical reservoir based on synthetic active particles [42]	5.4
Reservoir computing with adaptive EI balance [43]	10

using only linear and quadratic terms. The predicted trajectories reproduced the Mackey–Glass attractor geometry well (Fig. 8). To quantify the similarity of the attractor structure, we additionally estimated the correlation dimension. The mean correlation dimension was 1.9553 for the true trajectories and 1.9547 for the predicted trajectories, corresponding to a relative error of about 0.03% [37]. A paired Student's t-test showed no significant difference between the true and predicted estimates ($t = 0.172$, $df = 99$, $p = 0.864$). To evaluate the duration of accurate forecasting, we used the VPT defined by (17). For the Mackey–Glass system, the conversion to Lyapunov units was performed using the largest Lyapunov exponent $\lambda_1 = 0.007$ reported in [38]. The mean VPT was 11.285 Lyapunov times.

For comparison with recent studies, the corresponding long-term forecasting results are summarized in Table III. The mean VPT obtained for NGRC-MDS, 11.3 Lyapunov times, lies above the values reported for the other considered approaches and is approximately 13% higher than the 10 Lyapunov times reported by Srinivasan et al. [43], which is the closest result in the compared set. This comparison, however, should be treated with caution, since the cited works differ somewhat in the precise definition of VPT, sampling step, training length, and evaluation protocol.

VI. DISCUSSION

We proposed an approach to hidden data recovery and chaotic time-series forecasting based on NG-RC with multiscale construction of the delay vector from the target signal. Its key idea is that the temporal organization of the target signal can be used to guide delay selection, since the signal inherently contains information about the characteristic time scales of the underlying dynamics, expressed in its oscillatory patterns. Accordingly, the delay vector should be constructed so as to represent the principal temporal scales present in the signal. This makes it possible to align the feature space with the internal structure of the dynamics and to account for multiscale components of the process. By contrast, the standard approach with uniform delays uses a single time step between successive lags, which, in the presence of multiple characteristic scales, inevitably leads to a compromise between detailed representation of fast variations and nonredundant representation of slow components. In this context, a nonuniform delay vector aligned with the dominant temporal scales of the signal appears to be a more natural solution.

This reasoning is consistent with the work of Pan and Duraisamy [17], who showed, in the context of linear time-delayed models of nonlinear dynamics, that the delay structure required for accurate signal recovery is determined by the spectral content of the signal. For systems with multiscale dynamics, this implies that the delay vector should be aligned with several dominant temporal scales of the target signal. In the proposed approach, each such scale is represented with a prescribed resolution, that is, by at least n points per period. Thus, the choice of delays is naturally related to the redundancy–irrelevance tradeoff, since excessively close lags increase representational redundancy, whereas overly sparse coverage of the temporal scales leads to the loss of dynamically relevant information.

The idea of explicitly accounting for multiscale temporal structure has also been actively developed in recent time-series forecasting models [26], [27], [44], [45]. In particular, Fu et al. [26] implement it in the DTIGNet model through an improved temporal-inception module, in which features are extracted at different temporal scales and then complemented by a gated recurrent unit block for modeling temporal dependencies. Similarly, Bao et al. [44] account for multiscale structure at the architectural level by combining a sparse multiscale expert network based on temporal convolutional networks with the Informer model for capturing long-term dependencies. In the work of Cao et al. [45], the same idea is implemented in the MSPatch model by partitioning the input sequence into fragments of different sizes followed by integration of features extracted at different scales. A related approach has also been developed in RC. In particular, Jiang et al. [27] proposed the LD-TDR and LS-TDR architectures, in which different parts of the reservoir are designed to capture long-term memory and local short-term dynamics, respectively. Taken together, these studies indicate that explicitly accounting for multiscale dynamics can substantially improve forecasting performance.

In our case, the same general idea is transferred to the level of delay-vector construction, which appears particularly natural for NG-RC, since the lag structure directly determines the construction of the feature space.

Recent studies in NG-RC further emphasize the importance of delay design for dynamical inference and forecasting. In particular, recent work has shown that delayed inputs can provide an effective basis for nonlinear feature construction in NG-RC and support dynamical inference tasks such as attractor reconstruction, bifurcation mapping, and phase recovery [46]. At the same time, the numerical stability of NG-RC has been shown to depend strongly on the conditioning of the polynomial feature matrix, while excessively small lags can lead to illconditioning and unstable long-term dynamics [33]. These observations indicate that delay selection plays a dual role: it is important both for capturing the temporal structure of the signal and for maintaining numerical stability. From this perspective, the proposed nonuniform multiscale delay construction is well suited to systems with multiscale dynamics.

The proposed method was tested under diverse conditions, including systems with different dimensionality and dynamical organization, different training-set lengths, and both uncorrelated and correlated observational noise. In the adaptive Kuramoto network, where optimized regular-delay NG-RC already outperformed the classical reservoir model from [3], NGRC-MDS achieved comparable reconstruction accuracy to the fully optimized NG-RC configuration while avoiding exhaustive search over regular delays. In the classical Lorenz benchmark, where the z-variable is reconstructed from x and y, NGRC-MDS outperformed the regular-delay NG-RC baselines and remained close to, or below, the fully optimized NG-RC model over a broad range of delay numbers. This advantage is consistent with the representation analysis: the proposed delay construction better preserves the local structure of the observed dynamics, as indicated by higher trustworthiness, and yields features that are, on average, more informative with respect to the target signal, as reflected by the higher mean feature–target mutual information. At the same time, this mutual-information estimate is computed independently for each feature and does not control for possible redundancy shared across coordinates; therefore, it should be interpreted only as a partial proxy for feature relevance.

The noise-robustness analysis further showed that NGRC-MDS provides lower reconstruction error than regular-delay baselines over the full range of both white and flicker noise levels considered. This result is notable because additive noise inevitably affects the EMD decomposition itself: at low SNR, a substantial fraction of noise energy is absorbed by the first IMFs and may therefore enter the resulting delay set. Our results indicate that this effect can be partly controlled by the threshold parameter ρ: under strong noise, a higher threshold suppresses weak noise-dominated components, whereas under weak noise a lower threshold retains more informative temporal scales. A related limitation of classical EMD is mode mixing, i.e., leakage of oscillations between neighboring IMFs. In the present framework, however, this effect is mitigated because delay selection is based on dominant scales aggregated across retained IMFs and weighted by their amplitudes, making the final delay set less sensitive to exact IMF boundaries. More robust ensemble-based EMD variants may further improve stability, but in our preliminary tests they did not provide a substantial practical advantage relative to the standard EMD used here.

The Mackey–Glass results further underline the practical effectiveness of NGRC-MDS. In the one-step-ahead forecasting setting, we compared it not only with our regular-delay NG-RC baseline but also with several representative modern forecasting models reported in [26], [27], and [28], including Transformer-based architectures, GRU- and LSTM-based recurrent networks, CNN-based models, and advanced RC approaches. Across all three benchmark protocols, NGRC-MDS consistently achieved the lowest reported RMSE and NRMSE values for sufficiently large n. In particular, under the experimental settings adopted from [26], the RMSE was reduced from 6.05×10^{-4} for DTIGNet, a multiscale temporal-inception model, to 2.20×10^{-5}. Under the settings of [27], the NRMSE was reduced from 5.26×10^{-3} for DATDR, a depth asynchronous time-delay reservoir with interlayer delay operators for enhanced short-term memory, to 1.80×10^{-4}. Under the settings of [28], the NRMSE was reduced from 2.39×10^{-3} for a multistage Chua's RC model to 3.30×10^{-4}. Overall, these comparisons indicate that the superiority of NGRC-MDS extends beyond standard RC baselines and remains clear relative to a broad range of state-of-the-art deep learning models, including Transformer, recurrent, and convolutional architectures.

In the long-term forecasting regime, repeated autonomous runs from different trajectory segments showed that the inclusion of third-order nonlinear terms yielded the most accurate and stable predictions. This is likely because the Mackey–Glass equation combines delayed feedback with strong nonlinear interactions, making accurate autonomous continuation particularly demanding in terms of nonlinear feature representation. The reconstructed attractor also remained very close to the true one: the relative difference in the mean correlation dimension was only about 0.03%. The resulting mean VPT was about 11.3 Lyapunov times. As summarized in Table III, this value lies near the upper end of the range reported in recent studies [39], [40], [41], [42], [43], while the predicted dynamics preserve the geometric structure of the underlying attractor.

A. Limitations and Future Work

The proposed method has several limitations. First, in its current form, it is restricted to a 1-D target signal. Extension to multivariate targets is possible in principle, but requires a separate methodological study. Second, the multiscale delay construction relies on statistical regularities of the signal, including EMD-derived scales and their density estimation, and therefore requires sufficiently long training sequences. For short data segments, both the decomposition and the resulting delay estimates may become less reliable. Another issue is the potential instability of long-term forecasting caused by poor conditioning of the feature matrix when delays become too closely spaced, which is consistent with recent analyses of instability and regularization in NG-RC [33], [47]. Our preliminary results indicate that such instability can be substantially reduced by additional regularization, in particular by injecting small noise into the training data [47], although this stabilization was not used in the results reported here. Finally, increasing the polynomial order rapidly increases the number of features, and the current framework remains inherently offline, since delays and features are extracted from a fixed training segment rather than updated during operation. These limitations motivate future work on multivariate extensions, data-driven selection of nonlinear order and compact feature subsets, and adaptation of the method to online learning.

VII. Conclusion

In this work, we showed that uniform delays can limit NG-RC performance for signals with multiple characteristic time scales, since a single regular lag cannot efficiently represent both fast and slow dynamics. To address this limitation, we proposed a multiscale delay-selection approach for NG-RC, in which irregular delays are constructed from the dominant temporal scales of the target signal.

In the adaptive Kuramoto network and the Lorenz system, the proposed method achieved reconstruction accuracy comparable to or better than optimized NG-RC with regular delays while showing good robustness across parameter settings. The analysis of the constructed feature spaces further indicated that the proposed delay selection better preserves local structure and yields more informative features for recovery.

The Mackey–Glass experiments showed that the same framework is also effective for forecasting. In the short-term regime, it outperformed regular-delay NG-RC and showed lower prediction errors than several recent forecasting approaches. In the long-term regime, it provided long VPT s while accurately preserving attractor geometry. Overall, these results show that multiscale delay selection provides a practical and effective extension of NG-RC for both hidden data recovery and forecasting in nonlinear systems with complex temporal organization.

References

[1] N. Bochow, A. Poltronieri, M. Rypdal, and N. Boers, "Reconstructing historical climate fields with deep learning," *Sci. Adv.*, vol. 11, no. 14, pp. –8, 2025, Art. no. eadp055.

[2] É. Plésiat, R. J. H. Dunn, M. G. Donat, and C. Kadow, "Artificial intelligence reveals past climate extremes by reconstructing historical records," *Nature Commun.*, vol. 15, no. 1, p. 9191, Oct. 2024.

[3] A. Badarin, A. Andreev, V. Klinshov, V. Antipov, and A. E. Hramov, "Hidden data recovery using reservoir computing: Adaptive network model and experimental brain signals," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 34, no. 10, Oct. 2024, Art. no. 103121.

[4] M. Yan, C. Huang, P. Bienstman, P. Tino, W. Lin, and J. Sun, "Emerging opportunities and challenges for the future of reservoir computing," *Nature Commun.*, vol. 15, no. 1, p. 2056, Mar. 2024.

[5] L. Gonon and J.-P. Ortega, "Reservoir computing universality with stochastic inputs," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 1, pp. 100–112, Jan. 2020.

[6] A. V. Andreev et al., "Stochastic cloning of dynamical systems with hidden variables," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 112, no. 1, Jul. 2025, Art. no. 015303.

[7] E. Bollt, "On explaining the surprising success of reservoir computing forecaster of chaos? The universal machine learning dynamical system with contrast to VAR and DMD," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 31, no. 1, Jan. 2021, Art. no. 013108.

[8] D. J. Gauthier, E. Bollt, A. Griffith, and W. A. S. Barbosa, "Next generation reservoir computing," *Nature Commun.*, vol. 12, no. 1, pp. 1–8, Sep. 2021.

[9] M. O. Franz and B. Schölkopf, "A unifying view of Wiener and Volterra theory and polynomial kernel regression," *Neural Comput.*, vol. 18, no. 12, pp. 3097–3118, Dec. 2006.

[10] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence*. Cham, Switzerland: Springer, 1981, pp. 366–381.

[11] T. Sauer, J. A. Yorke, and M. Casdagli, "Embedology," *J. Stat. Phys.*, vol. 65, nos. 3–4, pp. 579–616, Nov. 1991.

[12] E. R. Deyle and G. Sugihara, "Generalized theorems for nonlinear state space reconstruction," *PLoS ONE*, vol. 6, no. 3, pp. e1829–5, 2011.

[13] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," *Phys. Rev. A, Gen. Phys.*, vol. 33, no. 2, p. 1134, 1986.

[14] C. Nichkawde, "Optimal state-space reconstruction using derivatives on projected manifold," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 87, no. 2, Feb. 2013, Art. no. 022905.

[15] E. Tan, S. Algar, D. Corrêa, M. Small, T. Stemler, and D. Walker, "Selecting embedding delays: An overview of embedding techniques and a new method using persistent homology," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 33, no. 3, Mar. 2023, Art. no. 032101.

[16] K.-H. Krämer, G. Datseris, J. Kurths, I. Z. Kiss, J. L. Ocampo-Espindola, and N. Marwan, "A unified and automated approach to attractor reconstruction," *New J. Phys.*, vol. 23, no. 3, 2021, Art. no. 033017.

[17] S. Pan and K. Duraisamy, "On the structure of time-delay embedding in linear models of non-linear dynamical systems," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 30, no. 7, Jul. 2020, Art. no. 073135.

[18] A. Filiz, "A fourth-order robust numerical method for integro-differential equations," *Asian J. Fuzzy Appl. Math.*, vol. 1, no. 1, pp. 28–33, 2013.

[19] R. Gutiérrez, A. Amann, S. Assenza, J. Gómez-Gardenes, V. Latora, and S. Boccaletti, "Emerging meso-and macroscales from synchronization of adaptive networks," *Phys. Rev. Lett.*, vol. 107, no. 23, 2011, Art. no. 234103.

[20] V. V. Makarov et al., "Emergence of a multilayer structure in adaptive networks of phase oscillators," *Chaos, Solitons Fractals*, vol. 84, pp. 23–30, Mar. 2016.

[21] E. Pitsik et al., "Inter-layer competition in adaptive multiplex network," *New J. Phys.*, vol. 20, no. 7, 2018, Art. no. 075004.

[22] A. V. Andreev, A. A. Badarin, V. A. Maximenko, and A. E. Hramov, "Forecasting macroscopic dynamics in adaptive Kuramoto network using reservoir computing," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 32, no. 10, 2022, Art. no. 103126.

[23] V. A. Maksimenko et al., "Macroscopic and microscopic spectral properties of brain networks during local and global synchronization," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 96, no. 1, Jul. 2017, Art. no. 012316.

[24] E. N. Lorenz, "Deterministic nonperiodic flow," *J. Atmos. Sci.*, vol. 20, no. 2, pp. 130–141, Mar. 1963.

[25] M. C. Mackey and L. Glass, "Oscillation and chaos in physiological control systems," *Science*, vol. 197, no. 4300, pp. 287–289, Jul. 1977.

[26] K. Fu, H. Li, and P. Deng, "Chaotic time series prediction using DTIGNet based on improved temporal-inception and GRU," *Chaos, Solitons Fractals*, vol. 159, Jun. 2022, Art. no. 112183.

[27] D. Jiang, Y. Zeng, M. You, and G. Wang, "Long-short term time-delay reservoir for nonlinear time series forecasting task," *Appl. Soft Comput.*, vol. 188, Feb. 2025, Art. no. 114418.

[28] X. Liu, B. Liu, and G. Guan, "High-accuracy prediction of chaotic time series using multi-stage coupled Chua's circuit-based reservoir computing," *Neurocomputing*, vol. 672, Apr. 2026, Art. no. 132867.

[29] J. Venna and S. Kaski, "Neighborhood preservation in nonlinear projection methods: An experimental study," in *Proc. Int. Conf. Artif. neural Netw.*, 2001, pp. 485–491.

[30] L. van der Maaten, "Learning a parametric embedding by preserving local structure," in *Proc. Artif. Intell. Statist.*, 2009, pp. 384–391.

[31] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Mar. 2011.

[32] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 69, no. 6, Jun. 2004, Art. no. 066138.

[33] E. R. dos Santos and E. Bollt, "On the emergence of numerical instabilities in next generation reservoir computing," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 35, no. 12, 2025, Art. no. 123102.

[34] N. E. Huang et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London. Ser. A: Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.

[35] D. Laszuk. (2017). *Python Implementation of Empirical Mode Decomposition Algorithm*. [Online]. Available: https://github.com/laszukdawid/PyEMD

[36] D. W. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization*. Hoboken, NJ, USA: Wiley, 2015.

[37] P. Grassberger and I. Procaccia, "Measuring the strangeness of strange attractors," *Phys. D, Nonlinear Phenomena*, vol. 9, nos. 1–2, pp. 189–208, Oct. 1983.

[38] J. Doyne Farmer, "Chaotic attractors of an infinite-dimensional dynamical system," *Phys. D, Nonlinear Phenomena*, vol. 4, no. 3, pp. 366–393, Mar. 1982.

[39] R. Chepuri, D. Amzalag, T. M. Antonsen, and M. Girvan, "Hybridizing traditional and next-generation reservoir computing to accurately and efficiently forecast dynamical systems," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 34, no. 6, Jun. 2024, Art. no. 063114.

[40] L. Grigoryeva, H. L. J. Ting, and J.-P. Ortega, "Infinite-dimensional next-generation reservoir computing," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 111, no. 3, Mar. 2025, Art. no. 035305.

[41] S. K. Rathor, M. Ziegler, and J. Schumacher, "Asymmetrically connected reservoir networks learn better," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 111, no. 1, Jan. 2025, Art. no. 015307.

[42] X. Wang and F. Cichos, "Harnessing synthetic active particles for physical reservoir computing," *Nature Commun.*, vol. 15, no. 1, p. 774, Jan. 2024.

[43] K. Srinivasan, D. Plenz, and M. Girvan, "Boosting reservoir computing with brain-inspired adaptive control of E-I balance," *Nature Commun.*, vol. 16, no. 1, p. 10212, Nov. 2025.

[44] W. Bao, Y. Cao, Y. Yang, and S. Wen, "Long short-term financial time series forecasting based on residual multiscale TCN sparse expert network and informer," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 36, no. 10, pp. 19200–19209, Oct. 2025.

[45] Y. Cao, Z. Tian, W. Guo, and X. Liu, "MSPatch: A multi-scale patch mixing framework for multivariate time series forecasting," *Expert Syst. Appl.*, vol. 273, May 2025, Art. no. 126849.

[46] R. Cestnik and E. A. Martens, "Next-generation reservoir computing for dynamical inference," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 36, no. 1, Jan. 2026, Art. no. 013115.

[47] Y. Zhang, E. Roque dos Santos, H. Zhang, and S. P. Cornelius, "How more data can hurt: Instability and regularization in next-generation reservoir computing," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 35, no. 7, Jul. 2025, Art. no. 073102.

Artem Badarin was born in March 1993. He received the B.S. degree in radiophysics, the M.S. degree in applied mathematics and physics, and the Ph.D. degree in radiophysics from Saratov State University, Saratov, Russia, in 2015, 2017, and 2020, respectively.

From 2015 to 2021, he was a Junior Researcher at Saratov State Technical University, Saratov, and Innopolis University, Innopolis, Russia. From 2021 to 2025, he was a Senior Researcher with the Baltic Center for Neurotechnology and Artificial Intelligence, Immanuel Kant Baltic Federal University, Kaliningrad, Russia. Since 2025, he has been a Leading Researcher at the Research Institute of Applied Artificial Intelligence and Digital Solutions, Plekhanov Russian University of Economics, Moscow, Russia. His research interests include mathematical modeling, neuroscience, neural networks, and nonlinear dynamics.

Dr. Badarin has received a number of fellowships for young scientists in physics during the whole period of study and has been repeatedly awarded diplomas of different levels.

Andrey Andreev was born in December 1993. He received the B.S. degree in radiophysics, the M.S. degree in applied mathematics and physics, and the Ph.D. degree in radiophysics from Saratov State University, Saratov, Russia, in 2015, 2017, and 2020, respectively.

From 2015 to 2021, he was a Junior Researcher at Saratov State Technical University, Saratov, and Innopolis University, Innopolis, Russia. From 2021 to 2025, he was a Senior Researcher with the Baltic Center for Neurotechnology and Artificial Intelligence, Immanuel Kant Baltic Federal University, Kaliningrad, Russia. Since 2025, he has been a Leading Researcher at the Research Institute of Applied Artificial Intelligence and Digital Solutions, Plekhanov Russian University of Economics, Moscow, Russia. His research interests include artificial intelligence, mathematical modeling, neuroscience, neural networks, and nonlinear dynamics.

Alexander Hramov was born in Saratov, Russia, in September 1974. He received the M.D. and Ph.D. degrees in electronic engineering from Saratov State University, Saratov, in 1996 and 1999, respectively.

From 1999 to 2014, he held positions as a Researcher, an Associate Professor, and a Full Professor with Saratov State University. From 2014 to 2018, he was a Leading Researcher with the Science and Educational Center of Artificial Intelligence Systems and Neurotechnology, and the Head of the Department of Automation, Control, and Mechatronics, Saratov State Technical University, Saratov. From 2019 to 2021, he was a Professor and the Head of the Laboratory of Neuroscience and Cognitive Technology of Innopolis University, Kazan, Russia. From 2021 to 2025, he was the Head of the Baltic Center for Neurotechnology and Artificial Intelligence, Immanuel Kant Baltic Federal University, Kaliningrad, Russia. He is currently the Head of the Research Institute of Applied Artificial Intelligence and Digital Solutions, Plekhanov Russian University of Economics, Moscow, Russia. His research interests include complex systems and networks theory, methods of brain diagnostics and neuroimaging data processing, development of AI methods, and applied research in digital medicine, neurotechnologies, and education.

Dr. Hramov is a Corresponding Member of Russian Academy of Science in 2025.