TASK-DRIVEN SELECTION AND SIZING OF SIMPLE CYCLE RESERVOIRS

Anonymous authors
Paper under double-blind review

ABSTRACT

Sparse deterministic reservoirs can reduce deployment cost relative to random echo-state networks (ESNs), but it is unclear when a simple cycle is sufficient for a new forecasting task. We separate this architecture-selection problem from the subsequent choice of cycle size. For any fixed reservoir, optimal squared prediction risk decomposes into irreducible noise and projection error onto the reservoir feature span. For a linear simple cycle represented over a declared finite history, delays congruent modulo the cycle size share state coordinates, yielding an exact covariance-weighted representation loss and a model-conditional minimum-cost certificate. A conservative training-only screen based on this analysis recommends cycles on 23 of 72 benchmark configurations, with no observed recommendation exceeding a validation-selected nonlinear random ESN by more than the prespecified 2% held-out NRMSE margin. The certified regime is narrow, however: a reservoir-free diagnostic developed on controlled tasks fails to transfer reliably to 83 configurations from five real datasets. Importantly, this failure is primarily one of predicting cycle suitability rather than of the architecture itself: a cheaper cycle remains within 2% of the random ESN on 50/83 configurations. Separately, once the cycle family is chosen, a near-best validation rule reduces deployed cycle dimension by 42.0% on 24 held-out configurations while satisfying the 2% practical-equivalence criterion throughout. These results distinguish what can be certified from what must still be validated directly.

1 INTRODUCTION

Reservoir computing (RC) uses a fixed recurrent dynamical system to transform an input sequence into a state representation, followed by a trained readout, typically linear or ridge regression (Jaeger, 2001; Lukoševičius, 2012). Because the recurrent weights are not trained, performance depends strongly on the reservoir architecture, dynamics, and size.

Echo-state networks (ESNs) usually use randomly initialized recurrent graphs and can vary across seeds and hyperparameter choices (Racca & Magri, 2021; Takagaki et al., 2026). Deterministic reservoirs instead use fixed sparse topologies. A simple directed cycle has only N recurrent edges for N states, yet minimum-complexity and cycle reservoirs can achieve strong predictive performance (Rodan & Tiňo, 2011; Li et al., 2024; Martinuzzi, 2025). This motivates our central question: when is a simple cycle sufficient for a forecasting task, and when should it be rejected in favor of a richer reservoir?

We distinguish this architecture decision from a second, easier problem: once the cycle family has been chosen, how large should the cycle be? Architecture selection asks whether a cycle can meet a prescribed predictive tolerance relative to a nonlinear random ESN at lower deployment cost. Cycle sizing asks for the smallest cycle that remains near the best validation performance within that family. Our architecture screen may therefore return ABSTAIN; the sizing rule never serves as evidence that the cycle is the right architecture.

The theory gives a narrow regime in which architecture sufficiency can be analyzed directly. Once its recurrent and input weights are fixed, a reservoir defines a feature space for the input history. Under squared loss, optimal risk is irreducible noise plus projection error onto that space, so two non-nested reservoirs have no task-independent performance ordering. For a linear simple cycle, finite size has a more concrete effect: after N shifts, information wraps around the ring, so delays

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

congruent modulo N share state coordinates. Under a finite linear-history model, this aliasing yields an exact covariance-weighted representation loss and a model-conditional minimum-cost certificate.

We then test how far this structure can be used in practice. A conservative training-only screen combines linear-history adequacy, the analytic certificate, and a small nonlinear-ESN challenge. On a broad benchmark it is selective but has no observed unsafe recommendation. A simpler reservoir-free diagnostic, however, fails prospectively across new real datasets, and richer post-hoc task descriptors do not rescue cross-dataset prediction. However, the cycles themselves remain competitive much more often than these diagnostics can predict. Thus the main difficulty outside the certified regime is not whether cycles can perform well, but whether their suitability can be inferred cheaply before directly validating competing architectures.

Our contributions are:

1. We characterize fixed-reservoir prediction as projection onto a feature space and derive an exact covariance-weighted representation loss for finite-history linear cycles, making modulo-N delay aliasing explicit.
2. We turn this result into a conservative training-only architecture screen that either recommends a minimum-cost cycle under stated assumptions or abstains.
3. We show prospectively that reservoir-free task diagnostics do not reliably predict cycle suitability across real datasets, even though cycles themselves frequently remain competitive on the accuracy–cost frontier.
4. Conditional on using a cycle, we give a coarse-to-local validation rule that reduces deployed state dimension while preserving near-best predictive accuracy.

2 Related Work

Deterministic reservoir topology. Minimum-complexity ESNs replace random recurrent graphs by sparse deterministic structures (Rodan & Tiňo, 2011). Simple cycles have universal approximation results under suitable constructions (Li et al., 2024), but universality is an expressivity statement rather than a finite-sample performance ordering. Martinuzzi (Martinuzzi, 2025) shows that several deterministic minimal topologies can outperform random ESNs on chaotic attractor reconstruction. Our contribution is different: we do not propose deterministic topology itself, but ask when a specific cycle can be recommended for a new forecasting task and when the method should abstain. Our lack of cycle recommendations on the tested chaotic forecasting tasks therefore does not contradict attractor-reconstruction results obtained under different objectives and settings.

Capacity, dimension, and pruning. Reservoir size is only an indirect proxy for the representation actually used by a task. Prior work studies reservoir dimension (Carroll, 2020) and task-specific pruning (Yadav & Stender, 2025); recent dynamics-aware pruning additionally emphasizes matched random-pruning and freshly initialized smaller-reservoir baselines (Laudari & Adhikari, 2026). These approaches ask which components of an existing reservoir can be removed. We instead ask a complementary design question: what is the minimum cost of a chosen architecture at a declared predictive tolerance? The distinction matters because pruning a parent reservoir and initializing a smaller reservoir are not equivalent interventions.

Robust ESN selection. Random ESNs can vary substantially across seeds, motivating robust hyperparameter validation and stability-aware design (Racca & Magri, 2021; Takagaki et al., 2026). Our nonlinear guard uses the same reliability motivation but a different decision target: it is not a general ESN optimizer, but a training-only challenger used to reject a cycle recommendation when a material nonlinear advantage cannot be ruled out.

3 A General Limit and a Relative Complexity

We first state a limitation that applies to any fixed reservoir, not only to cycles. Let Z_t be all information available when the forecast is made (for example, a finite input history), and let Y_t be the

target. The conditional mean $m(Z_t) = \mathbb{E}[Y_t \mid Z_t]$ is the optimal squared-loss predictor if the entire history Z_t can be used. Here $\mathbb{E}$ denotes expectation and P_Z denotes the probability distribution of Z_t.

A fixed reservoir architecture A with N states transforms Z_t into the feature vector $\phi_{A,N}(Z_t)$. A linear readout can form linear combinations of these features and a constant. We denote the resulting closed feature space by $\mathcal{S}_{A,N}$. It is a subspace of $L^2(P_Z)$, the space of square-integrable functions of Z; the superscript "2" refers to the squared norm $\|f\|^2_{L^2(P_Z)} = \mathbb{E}[f(Z)^2]$. Finally, $\Pi_{\mathcal{S}} m$ denotes the orthogonal projection of m onto $\mathcal{S}$: the representable function in $\mathcal{S}$ closest to m in mean squared distance.

Proposition 1 (Projection risk and universal dominance)**.** *Under squared loss, the best population risk achievable by the fixed reservoir is*

$$\mathcal{R}^*_{A,N} = \mathbb{E}[\mathrm{Var}(Y_t \mid Z_t)] + \left\| m - \Pi_{\mathcal{S}_{A,N}} m \right\|^2_{L^2(P_Z)}. \tag{1}$$

For two fixed feature spaces $\mathcal{S}_1$ and $\mathcal{S}_2$, $\mathcal{R}^*_1 \le \mathcal{R}^*_2$ for every square-integrable regression function m if and only if $\mathcal{S}_2 \subseteq \mathcal{S}_1$.

The two terms in Eq. equation 1 have different meanings. The first is irreducible uncertainty in Y_t after observing Z_t. The second is representation error: information needed by the regression function but absent from the reservoir feature span. A larger state vector need not have smaller representation error; what matters is whether its span contains the task-relevant direction.

The dominance statement follows directly. If $\mathcal{S}_2 \subseteq \mathcal{S}_1$, every predictor available to reservoir 2 is also available to reservoir 1. If the inclusion fails, there is a target direction represented by reservoir 2 but not by reservoir 1, providing a counterexample to universal dominance. Thus a finite simple cycle and a random $\tanh$ ESN generally have no task-independent ordering. This does not say that either architecture is poor; it says that a claim of superiority needs a task distribution or an explicit comparison. The formal proof is in Appendix B.

This result motivates a cost-at-tolerance quantity that is explicitly architecture relative.

Definition 2 (Architecture-relative predictive complexity)**.** Let $\mathcal{C}(A, N)$ denote a declared implementation cost, such as online multiplies, recurrent edges, state dimension, or measured latency, and let $\mathcal{R}^*_{\mathrm{oracle}} = \mathbb{E}[\mathrm{Var}(Y_t \mid Z_t)]$. For tolerance $\epsilon \ge 0$, define

$$\mathfrak{C}_A(\epsilon) = \inf_N \left\{ \mathcal{C}(A, N) : \mathcal{R}^*_{A,N} - \mathcal{R}^*_{\mathrm{oracle}} \le \epsilon \right\}, \tag{2}$$

with value $+\infty$ if the feasible set is empty.

Thus $\mathfrak{C}_A(\epsilon)$ is the cheapest implementation in family A whose excess population risk is at most ϵ; on our finite candidate grids the infimum is a minimum. If no candidate meets the tolerance, the architecture is not certified. We measure cost primarily by online multiplications, with state dimension and recurrent-edge count as supporting measures. The same task can therefore be cheap for one feature family and expensive for another. In the linear-cycle specialization below, covariance weights errors according to the input histories that actually occur.

4 Problem Setup

At time t, the model receives $u_t \in \mathbb{R}^d$, updates an N-dimensional reservoir state x_t, and uses a trained readout to forecast the target h steps ahead. The cycle studied analytically and in the broad screening benchmark is linear:

$$x_t = g P_N x_{t-1} + W_{\mathrm{in}} u_t, \tag{3}$$

where P_N shifts each state to its successor around the ring, $g \in (0, 1)$ is the cycle gain, and W_{in} injects the input. Only the affine readout

$$\widehat{y}_{t+h} = W_{\mathrm{out}} x_t + b \tag{4}$$

is trained, using ridge regression. The nonlinear comparator is a sparse random recurrent ESN with a $\tanh$ activation. Thus the analytic model and deployed cycle are the same linear architecture.

For a discrete size grid $\mathcal{N}$, the within-cycle objective is not the exact validation minimizer at any cost. We seek the smallest cycle whose validation NRMSE e_N lies within relative tolerance τ of the best candidate:

$$N_\tau = \min \left\{ N \in \mathcal{N} : e_N \le (1 + \tau) e_{\min} \right\}, \qquad e_{\min} = \min_{N' \in \mathcal{N}} e_{N'}, \tag{5}$$

with $\tau = 0.01$ in the sizing protocol. Architecture screening uses the distinct 2% NRMSE tolerance described below. The linear-history oracle in the theory is a population reference and is not the random ESN comparator used in experiments.

5 Cycle Size Controls Temporal Resolution

5.1 Why delays collide in a cycle

After washout, the current state contains geometrically weighted past inputs:

$$x_t = \sum_{k=0}^{\infty} g^k P_N^k W_{\text{in}} u_{t-k}. \tag{6}$$

Because one trip around the ring takes N steps, $P_N^N = I$. With scalar input injected at one reference node, coordinate j therefore contains

$$x_j(t) = \sum_{m=0}^{\infty} g^{j+mN} u_{t-j-mN}. \tag{7}$$

Thus delays $j, j+N, j+2N, \ldots$ are not separately visible to the readout; they appear as one weighted mixture. Increasing N separates more delays before wraparound. This is the mechanism by which an undersized cycle can lose predictive information. Once all task-relevant delay combinations are already represented, however, extra states need not improve prediction. The formal statement and proof are in Appendix C.

5.2 The task-aligned loss used by the certificate

To quantify which collisions matter, collect the most recent L inputs in $z_t = [u_t^\top, \ldots, u_{t-L+1}^\top]^\top$ and suppose that the predictable part of the target is linear in this declared history:

$$y_{t+h} = H^\top z_t + \varepsilon_t, \quad \mathbb{E}[\varepsilon_t \mid z_t] = 0, \quad \mathbb{E}[\varepsilon_t^2] = \sigma_\varepsilon^2. \tag{8}$$

Here H identifies the target-relevant combination of delays, while $\Sigma = \text{Cov}(z_t)$ describes which histories occur. If the declared finite-history cycle state is $x_t = F_N z_t$, its representation loss is

$$E_N = H^\top \Sigma H - c_N^\top G_N^\dagger c_N, \quad G_N = F_N \Sigma F_N^\top, \quad c_N = F_N \Sigma H, \tag{9}$$

where $G_N^\dagger$ is the Moore–Penrose pseudoinverse. Ordinary least squares gives minimum population MSE $\sigma_\varepsilon^2 + E_N$, with $E_N \geq 0$. The loss is zero precisely when the cycle features contain the target-relevant covariance-weighted direction. This is why covariance rank or Hankel rank alone is insufficient: rank counts available directions but does not determine whether the target uses the missing ones. The full theorem and proof are in Appendix D.

Normalize the loss by the linear-history oracle risk,

$$R_N = E_N / \sigma_\varepsilon^2, \tag{10}$$

and let U_N be a simultaneous upper confidence envelope estimated from training data. An NRMSE tolerance δ corresponds to relative-MSE tolerance $\tau_\delta = (1+\delta)^2 - 1$. The theory-derived selection rule is

$$\widehat{N}_{\text{cert}} \in \arg\min_{N \in \mathcal{N}} \{\mathcal{C}(N) : U_N \leq (1+\delta)^2 - 1\}, \tag{11}$$

with ABSTAIN when the feasible set is empty. For the fixed $\delta = 0.02$ margin, the threshold is 0.0404. Appendix E states the population, finite-sample-conditional, and cost corollaries formally.

Role in the algorithm. The main procedure uses the theory in exactly three ways: Eq. equation 7 explains what size changes; Eq. equation 10 supplies the task-aligned quantity to estimate; and Eq. equation 11 returns a minimum-cost candidate or abstains. The remaining projection identities establish these claims but are not extra algorithmic steps.

Scope. The model-conditional certificate concerns population representation error under an optimal linear readout. A stationary cycle also contains an older-input tail, so the finite map $x_t = F_N z_t$ is exact only after reset/truncation or when that tail is explicitly included or controlled. On observational data, our finite-lag calculation is therefore a guarded plug-in approximation. The certificate also does not compare the cycle with an arbitrary nonlinear ESN or account for finite-sample ridge estimation. Those gaps motivate the adequacy check and nonlinear challenger in the next section.

6 Selective Screening and Practical Sizing

We use two separate procedures. The first asks whether the training data provide enough evidence to recommend a cycle over a broader architecture search. The second is used only after the cycle family has been chosen and selects a small near-best value of N. Their inputs and saved outputs are summarized in Appendix Table 1.

6.1 Training-only cycle-suitability rule

The analytic certificate is exact only under the finite linear-history model, so we guard it with two empirical checks. All thresholds are fixed before outer validation or test evaluation.

1. **Linear-history adequacy.** Fit the finite-history linear model on an internal training block and evaluate it on a later internal block. If $R^2 < 0.50$, return ABSTAIN.
2. **Model-conditional certificate.** For $N \in \{8, 16, 24, 32, 48, 64, 96, 128\}$, estimate R_N and an approximate simultaneous 95% upper envelope U_N. With $\delta = 0.02$, choose the cheapest N satisfying $U_N \leq 0.0404$; if none exists, return ABSTAIN.
3. **Nonlinear guard.** On a reserved final 20% of the training interval, compare the provisional cycle with fixed-size random-$\tanh$ ESNs using three prespecified seeds. Recommend the cycle only if one-sided paired confidence bounds do not show a material ESN advantage; otherwise return ABSTAIN.

The nonlinear guard is deliberately small: it is a rejection check, not a full ESN optimizer. After the decision is saved, the held-out audit uses a stronger ESN selected over all candidate sizes and ten seeds. Exact envelope construction, covariance estimation, seeds, and confidence-bound details are given in Appendix G.1.

6.2 Accuracy-constrained cycle sizing

Once a cycle has been selected, we choose the smallest queried size whose validation NRMSE is within 1% of the best queried value. On the grid

$$\mathcal{N} = \{8, 16, 24, 32, 48, 64, 96, 128, 192, 256\}, \tag{12}$$

we first evaluate $\{8, 16, 32, 64, 128, 256\}$. We then query any missing immediate neighbors of both the coarse-grid minimizer and the smallest coarse-grid size within 1% of that minimum, recompute the best observed validation error, and return the smallest queried size satisfying Eq. equation 5. The primary objective is smaller deployment size; query reduction is secondary.

6.3 Reservoir-free diagnostic transfer

Because the guarded screen is deliberately selective, we also test whether training-data properties alone can identify additional tasks suited to a cycle. Controlled experiments suggested a simple nonlinear-gain descriptor: the fractional reduction in internal-validation MSE obtained by a fixed nonlinear lag predictor relative to a linear lag predictor. We fix the rule "recommend a cycle when nonlinear gain ≤ 0.10" before applying it to the real-data benchmark. This diagnostic is separate from the analytic certificate and has no theoretical guarantee.

7 Experimental Protocol

The experiments test three questions: whether near-best validation sizing can reduce cycle size, whether the guarded architecture screen can recommend cycles without material held-out loss, and

whether a cheaper task-only diagnostic transfers to unseen real datasets. We call a series–horizon pair a configuration; configurations from the same generator or dataset form a group. Dataset frequencies and detailed grids are reported in Appendix A.

Held-out cycle sizing. We developed the 1% near-best tolerance and coarse-to-local policy on 36 ETTh1 and Exchange configurations, then fixed both before evaluating 24 configurations from ETTm1, Electricity, Traffic, and Weather. The full size grid is evaluated only after the proposed size is recorded. Practical equivalence to the exact validation-best cycle is assessed by paired moving-block bootstrap on test residuals using a prespecified 2% relative-NRMSE margin; exact horizons, folds, history length, ridge grid, and bootstrap settings are in Appendix G.2.

Broad architecture-screening benchmark. We evaluate the guarded screen on 72 forecasting configurations in 33 groups spanning linear FIR/AR/VAR processes, NARMA tasks, Lorenz–63, Rössler, Mackey–Glass, and six real forecasting datasets (Lai et al., 2018; Zhou et al., 2021; Wu et al., 2021). The screen either recommends a cycle or abstains. Only after that decision is saved do we select a stronger random-$\tanh$ ESN comparator over the full audit grid. The exact candidate sizes, seeds, hyperparameters, and family counts are given in Appendices G.1 and I.

Prospective real-data transfer. We next evaluate the fixed nonlinear-gain rule on 83 configurations from five dataset groups: ETTh1, Exchange, Weather, Traffic, and Metro (Zhou et al., 2021; Lai et al., 2018; Wu et al., 2021; Hogue, 2019). These datasets were not used to construct the nonlinear-gain rule. For each configuration, the training-only diagnostic decision and validation-selected cycle and random-ESN specifications are saved before test evaluation. A recommendation is unsafe when $\mathrm{NRMSE}_{\mathrm{cyc}} > 1.02\,\mathrm{NRMSE}_{\mathrm{ESN}}$. Persistence, seasonal-naive, and direct-lag ridge baselines assess absolute forecasting performance.

After this prospective evaluation, we perform an explicitly exploratory audit using predictive-state and Hankel summaries, linear adequacy, nonlinear gain, cycle-specific representation proxies, seasonality, memory, and chronological instability. Logistic regression and random forests are evaluated by leave-one-dataset-out prediction. Because these models are fitted after the 83 outcomes are known, they are not treated as a second prospective validation.

Cost and uncertainty. Deployment cost is measured primarily by online multiplications, with state dimension and recurrent-edge count reported as supporting measures. For the broad benchmark, uncertainty is reported both across recommended configurations and after aggregation by the 33 predeclared groups; a group is unsafe if any recommended configuration in that group violates the 2% margin.

8 Results

8.1 Held-out cycle sizing reduces deployment size

Across all 24 held-out configurations, the sizing rule recovers the exhaustive full-grid 1%-near-best validation size. The selected cycles have mean size $N = 43$, compared with $N = 90$ for the exact validation minimizer, corresponding to a mean per-configuration size reduction of 42.0%. Mean test NRMSE is 0.35106 for the proposed selector and 0.35059 for the exact validation minimizer.

For every configuration, the one-sided 95% upper bound on relative NRMSE degradation remains below the prespecified 2% margin; the largest is 1.75%. The rule also evaluates 17.5% fewer candidate sizes on average. Thus its main benefit is a smaller deployed reservoir rather than a large reduction in search cost. Full sizing summaries and the post-hoc comparison with the unattainable test-optimal size are in Appendix H.

8.2 Broad screening is conservative but selective

Across 72 benchmark configurations, the guarded screen recommends a cycle on 23 (31.9% coverage) and abstains on 49. None of the 23 recommendations exceeds the prespecified 2% held-out NRMSE margin relative to the validation-selected nonlinear random ESN. Among recommended

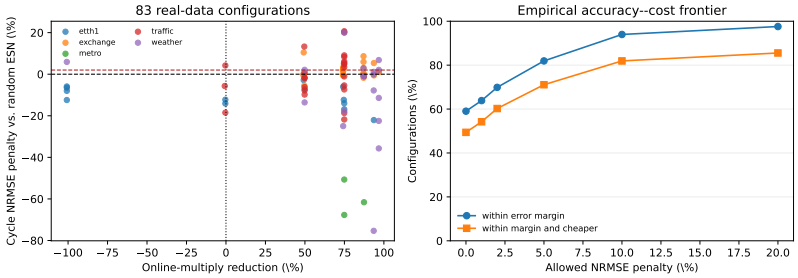

Figure 1: Accuracy–cost outcomes for 83 real-data configurations. Points with lower NRMSE and lower online-multiplication cost than the validation-selected random-$\tanh$ ESN represent strict cycle Pareto dominance. Test outcomes are used only for evaluation, not selection.

configurations, the mean relative test-NRMSE difference is -0.84%, while online multiplications fall by 86.4% on average. Relative to the validation-selected cycle, the analytically selected cycle reduces online multiplications by 38.7% on average.

The abstentions show that the method is conservative rather than a general cycle detector. Only 12 of the 49 abstentions correspond post hoc to cases in which the validation-selected cycle is more than 2% worse than the random ESN. Abstention therefore means insufficient evidence for a cycle recommendation, not evidence that a random ESN is necessary. Family-level results are given in Appendix I and Table 6.

Zero observed violations among 23 recommendations still leaves substantial statistical uncertainty: the exact one-sided 95% upper bound on the violation rate is 12.2% at the configuration level and 12.7% after grouping. We therefore do not infer a population failure rate below 5% from this benchmark.

8.3 Task descriptors fail to predict cycle suitability across datasets

The prospective real-data experiment does not support the controlled-task nonlinear-gain rule as a reliable selector. At the fixed threshold of 0.10, it recommends cycles on 69 of 83 configurations, of which 19 violate the 2% NRMSE margin. The observed configuration-level unsafe fraction is 27.5%, with a one-sided 95% upper bound of 37.7%; violations occur in three of the five dataset groups.

The exploratory audit does not identify a stronger reservoir-free rule. Across the tested descriptors and classifiers, the best leave-one-dataset-out AUC is 0.570, and zero observed violations can be obtained post hoc only at 1/83 coverage. Richer task summaries therefore do not provide useful cross-dataset discrimination in this benchmark.

Crucially, failure of the diagnostic is not failure of the cycle architecture. Cycles achieve test NRMSE no worse than the random ESN on 49/83 configurations and are both more accurate and cheaper on 41/83. With a 2% tolerance, 50/83 configurations have a cheaper cycle within the margin, with 73.9% lower online-multiplication cost on average; 24 of these also outperform both naive baselines. The central difficulty is therefore predicting cycle suitability from task descriptors, not finding settings in which cycles are competitive.

9 Discussion

The experiments separate two questions that are easy to conflate. First, can cycle sufficiency be established without fitting a broad set of competing reservoirs? Our answer is yes, but only in a restricted regime. The projection analysis and delay-aliasing calculation give a model-conditional certificate when a finite linear-history description is adequate, and the guarded screen converts that result into a deliberately conservative recommendation rule. Its abstentions are therefore part of the method rather than failures of optimization.

Second, can simple task descriptors extend this recommendation to settings outside the certified regime? The prospective transfer experiment suggests that this is substantially harder. A diagnostic

that appeared useful on controlled tasks failed across new real datasets, and adding richer post-hoc descriptors did not yield useful leave-one-dataset-out discrimination. Yet direct evaluation shows that cycles themselves remain competitive in many of those same configurations. This distinction matters: the evidence does not say that simple cycles are rarely useful; it says that their usefulness is difficult to infer from reservoir-free task summaries alone.

The practical implication is a three-part workflow. Use the analytic screen when its assumptions are defensible; outside that regime, validate competing architectures directly rather than trusting a task-only heuristic; and, once a cycle is chosen, size it by an accuracy-constrained validation rule rather than defaulting to the exact validation minimizer. This keeps architecture selection and deployment sizing separate while exploiting the favorable cost of cycles when the data support them.

10 Limitations

The model-conditional certificate assumes a correctly specified finite linear-history representation, and the plug-in confidence envelope used in practice has no theorem-level finite-sample guarantee. The 83 prospective configurations come from only five dependent dataset groups, so they provide limited evidence about transfer to new domains. The broader diagnostic audit is exploratory and post hoc. Our comparisons also exclude trained sequence models and several other reservoir families. Finally, multiply counts are proxies for latency or energy, and competitiveness with a random ESN does not by itself imply improvement over a naive forecast.

11 Conclusion

Simple cycles can be inexpensive and competitive, but deciding when to use them is harder than sizing them once chosen. We identify a narrow regime in which cycle sufficiency can be analyzed through task-aligned representation loss and converted into a conservative model-conditional screen. Outside that regime, the task-only diagnostics tested here do not transfer reliably, so the accuracy–cost frontier should be validated directly. Conditional on choosing a cycle, near-best validation sizing can then reduce deployment cost with little loss in predictive accuracy.

Reproducibility Statement

The anonymous supplementary archive contains executable code for the analytic screen, controlled experiments, prospective real-data transfer study, exploratory diagnostic audit, and cycle-sizing pipeline. It includes fixed protocols and manifests, saved pre-test decisions, preprocessing code, result tables, environment requirements, and automated integrity and leakage checks. Random seeds, chronological splits, candidate grids, and practical margins are recorded in the supplied artifacts. Raw third-party datasets are not redistributed; required filenames, schemas, and preprocessing steps are documented. The README distinguishes prospective from post-hoc analyses and provides smoke-test and full-run commands.

AI Use Statement

Generative-AI tools assisted with polishing the writing of this paper in terms of flow and English grammar, as well as drafting the structure of the paper. They also assisted with generating synthetic datasets, and generating Python scripts to run simulations. All mathematical arguments, experimental designs, code, generated data, and reported results were independently checked by the authors.

References

Thomas L. Carroll. Dimension of reservoir computers. *Chaos*, 30:013102, 2020.

John Hogue. Metro interstate traffic volume. UCI Machine Learning Repository, 2019.

Herbert Jaeger. The "echo state" approach to analysing and training recurrent neural networks. Technical Report 148, German National Research Center for Information Technology, 2001.

Guokun Lai, Wei-Cheng Chang, Yiming Yang, and Hanxiao Liu. Modeling long- and short-term temporal patterns with deep neural networks. In *Proceedings of the 41st International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 95–104, 2018. doi: 10.1145/3209978.3210006.

Sudip Laudari and Puspa Raj Adhikari. Rethinking reservoir pruning: A dynamical perspective for echo state networks. *arXiv preprint arXiv:2608.04593*, 2026.

Bowen Li, Richard S. Fong, and Peter Tiňo. Simple cycle reservoirs are universal. *Journal of Machine Learning Research*, 25:1–28, 2024.

Mantas Lukoševičius. A practical guide to applying echo state networks. In *Neural Networks: Tricks of the Trade*, pp. 659–686. Springer, 2012.

Francesco Martinuzzi. Minimal deterministic echo state networks outperform random reservoirs in learning chaotic dynamics. *Chaos*, 35:091108, 2025. doi: 10.1063/5.0288751.

Alberto Racca and Luca Magri. Robust optimization and validation of echo state networks for learning chaotic dynamics. *Neural Networks*, 142:252–268, 2021.

Ali Rodan and Peter Tiňo. Minimum complexity echo state network. *IEEE Transactions on Neural Networks*, 22(1):131–144, 2011.

Shunsuke Takagaki, Ryunosuke Fukuzaki, Koki Tateishi, Toshikazu Noda, and Hiroyasu Ando. Robust design of echo state networks for soft sensor applications based on risk-aware optimization and stability testing. In *International Conference on Neural Information Processing*, volume 2297 of *Communications in Computer and Information Science*, pp. 31–45. Springer, 2026.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In *Advances in Neural Information Processing Systems*, volume 34, pp. 22419–22430, 2021.

Yadav and Stender. Task-specific node pruning enhances computational efficiency of reservoir computing networks. *Chaos*, 35:083123, 2025.

Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 11106–11115, 2021.

A Decision and Dataset Reference

This appendix summarizes the two decision problems studied in the paper and the real-world datasets used in the empirical evaluations. The architecture-screening problem asks whether the available training data provide sufficient evidence to recommend a simple cycle reservoir at all, whereas the sizing problem assumes that the cycle architecture has already been chosen and selects the smallest cycle whose validation performance remains near the best observed value. These decisions use different data and should therefore be interpreted separately. Table 1 summarizes this distinction.

The real-data experiments cover several commonly used multivariate forecasting datasets with substantially different sampling frequencies, numbers of variables, and application domains. All experiments preserve temporal order and use chronological splits rather than random shuffling. Table 2 provides a compact reference for the datasets and their native sampling frequencies.

Table 1: The two decisions studied in this paper. Neither procedure accesses test outcomes before its decision is saved.

Decision	Question	Selection data	Output
Architecture screen	Is there sufficient evidence to use a cycle?	Training and reserved training-only guard interval	Cycle with N, or ABSTAIN
Within-cycle sizing	What is the smallest near-best cycle?	Training and validation	Selected N

Table 2: Real-world forecasting datasets. Frequencies describe the original collections; experiments use chronological splits without shuffling.

Name	What is measured	Native frequency
ETTh1	Transformer load and oil-temperature variables	Hourly
ETTm1	The same transformer family at finer resolution	15 minutes
Electricity	Electricity demand from 321 client meters	Hourly
Traffic	Road occupancy from 862 freeway sensors	Hourly
Weather	Multivariate meteorological observations	10 minutes
Exchange	Exchange rates for eight currencies/countries	Daily
Metro	Interstate traffic volume with calendar/weather covariates	Hourly

B Proof of Proposition 1

Write $Y = m(Z) + \eta$, where $\mathbb{E}[\eta \mid Z] = 0$. For every $f \in \mathcal{S}$, orthogonality gives

$$\mathbb{E}[(Y - f(Z))^2] = \mathbb{E}[\eta^2] + \mathbb{E}[(m(Z) - f(Z))^2] \tag{13}$$

$$= \mathbb{E}[\mathrm{Var}(Y \mid Z)] + \|m - f\|^2_{L^2(P_Z)}. \tag{14}$$

Minimizing over the closed subspace $\mathcal{S}$ proves Eq. equation 1. If $\mathcal{S}_2 \subseteq \mathcal{S}_1$, projection onto $\mathcal{S}_1$ cannot have greater error than projection onto $\mathcal{S}_2$. Conversely, if $\mathcal{S}_2 \nsubseteq \mathcal{S}_1$, choose $m \in \mathcal{S}_2 \setminus \mathcal{S}_1$. Because $\mathcal{S}_1$ is closed, $\|m - \Pi_{\mathcal{S}_1} m\|_2 > 0$, whereas $\|m - \Pi_{\mathcal{S}_2} m\|_2 = 0$. Thus reservoir 1 cannot dominate reservoir 2 for all tasks. Interchanging the indices shows that non-nested spaces admit tasks favoring either reservoir.

C Formal Delay-Aliasing Statement

Proposition 3 (Modulo-N delay aliasing in a simple cycle)**.** *Assume scalar input is injected with unit weight at a single reference node of the linear cycle in Eq. equation 3, and consider the stationary post-washout solution. Index state coordinates by $j \in \{0, \dots, N-1\}$. Then Eq. equation 7 holds, up to the chosen orientation of the cycle. Hence two delays occupy the same state coordinate if and only if they are congruent modulo N.*

Proof. Because $P_N^N = I$, the term associated with delay k in Eq. equation 6 occupies the coordinate reached after k cyclic shifts. Writing $k = j + mN$ groups exactly the delays returning to coordinate j, which gives Eq. equation 7. □

D Exact Linear-Reservoir Theorem and Proof

Theorem 4 (Exact excess risk of a linear reservoir)**.** *Under Eq. equation 8, let $x_t = Fz_t$, $\Sigma = \mathrm{Cov}(z_t)$, $G = F\Sigma F^\top$, and $c = F\Sigma H$. The minimum population MSE over all linear readouts is*

$$\mathcal{R}_F^* = \sigma_\varepsilon^2 + \mathcal{E}_F, \qquad \mathcal{E}_F = H^\top \big[\Sigma - \Sigma F^\top (F\Sigma F^\top)^\dagger F\Sigma\big] H. \tag{15}$$

Equivalently, $\mathcal{E}_F = H^\top \Sigma H - c^\top G^\dagger c \geq 0$. Moreover, $\mathcal{E}_F = 0$ if and only if $\Sigma^{1/2} H$ lies in the row space of $F\Sigma^{1/2}$.

Let $X = FZ$ and consider a linear readout $w^\top X$. Using $Y = H^\top Z + \varepsilon$ with $\mathbb{E}[\varepsilon \mid Z] = 0$,

$$\mathbb{E}[(Y - w^\top X)^2] = \sigma_\varepsilon^2 + \mathbb{E}[(H^\top Z - w^\top FZ)^2] \tag{16}$$

$$= \sigma_\varepsilon^2 + H^\top \Sigma H - 2w^\top F\Sigma H + w^\top F\Sigma F^\top w. \tag{17}$$

Writing $G = F\Sigma F^\top$ and $c = F\Sigma H$, let $A = F\Sigma^{1/2}$ and $q = \Sigma^{1/2}H$. Then $G = AA^\top$ and $c = Aq$, so $c \in \mathrm{range}(A) = \mathrm{range}(AA^\top) = \mathrm{range}(G)$. Hence the minimum of the quadratic form is attained by any least-squares solution $w^* = G^\dagger c + v$ with $v \in \ker(G)$ and has value

$$\sigma_\varepsilon^2 + H^\top \Sigma H - c^\top G^\dagger c. \tag{18}$$

Substituting G and c yields Eq. equation 15. If $\Sigma^{1/2}$ denotes the positive-semidefinite square root, then

$$\mathcal{E}_F = \left\| \left(I - A^\top (AA^\top)^\dagger A\right) \Sigma^{1/2} H \right\|_2^2. \tag{19}$$

The matrix in parentheses is the orthogonal projector onto the complement of the row space of A, proving nonnegativity and the stated zero-loss condition.

Scope of the finite-history representation. The proof above is exact for the stated condition $X = FZ$. A stationary cycle in Eq. equation 6 generally depends on the entire input past, so replacing that past by a finite lag vector is exact only after reset/truncation or when the omitted tail is identically zero. The IID calculation in Appendix F retains the full geometric tail and is exact under its own assumptions. In the general observational-data procedure, the finite-lag covariance calculation is therefore a plug-in approximation and is guarded empirically rather than asserted as an unconditional theorem.

E Certificate and Cost Corollaries

Corollary 5 (Cycle sufficiency at an NRMSE tolerance)**.** *Under Theorem 4, suppose the cycle and linear-history oracle use the same positive NRMSE normalization. For $\delta > 0$, define*

$$\tau_\delta = (1+\delta)^2 - 1. \tag{20}$$

If $R_N \le \tau_\delta$, then $\sqrt{\mathcal{R}^*_{F_N}} \le (1+\delta)\sqrt{\mathcal{R}^*_{\mathrm{lin}}}$. For $\delta = 0.02$, $\tau_\delta = 0.0404$.

Corollary 6 (Finite-sample certified minimum-cost cycle)**.** *Suppose U_N is a simultaneous $(1-\alpha)$ upper confidence envelope for R_N over the finite candidate set $\mathcal{N}$. If the feasible set in Eq. equation 11 is nonempty, then with probability at least $1-\alpha$ the selected cycle satisfies Corollary 5. If the set is empty, the certified decision is* ABSTAIN*. For increasing cost, the optimizer is the smallest certified N.*

Corollary 7 (Architecture-relative complexity of the cycle)**.** *Define*

$$\mathfrak{C}^{\mathrm{rel}}_{\mathrm{cycle}}(\tau) = \inf_N \{\mathcal{C}(N) : R_N \le \tau\}. \tag{21}$$

Then $\mathfrak{C}^{\mathrm{rel}}_{\mathrm{cycle}}(\tau) = \mathfrak{C}_{\mathrm{cycle}}(\tau\sigma_\varepsilon^2)$. For $\mathcal{C}(N) = N$, the optimum is the smallest feasible cycle.

Proof of Corollary 5. By Theorem 4, $\mathcal{R}^*_{F_N}/\mathcal{R}^*_{\mathrm{lin}} = 1 + R_N$. If $R_N \le (1+\delta)^2 - 1$, then $\mathcal{R}^*_{F_N}/\mathcal{R}^*_{\mathrm{lin}} \le (1+\delta)^2$; taking square roots gives the claim.

Proof of Corollary 6. On the simultaneous-coverage event, every feasible candidate satisfies $R_N \le U_N \le \tau_\delta$. Corollary 5 therefore applies to the selected candidate. The event has probability at least $1-\alpha$. If the feasible set is empty, no stated guarantee follows, so abstention is required.

Proof of Corollary 7. Because $R_N = E_N/\sigma_\varepsilon^2$, the constraint $R_N \le \tau$ is equivalent to the absolute excess-risk constraint $E_N \le \tau\sigma_\varepsilon^2$. Substitution into Definition 2 yields $\mathfrak{C}^{\mathrm{rel}}_{\mathrm{cycle}}(\tau) = \mathfrak{C}_{\mathrm{cycle}}(\tau\sigma_\varepsilon^2)$. If $\mathcal{C}(N)$ is increasing, minimizing cost is equivalent to choosing the smallest feasible N.

F Detailed Linear Sizing Theory

This appendix collects additional derivations for the linear sizing analysis.

F.1 IID CLOSED FORM

Assume scalar zero-mean IID input with $\mathrm{Var}(u_t) = 1$ and an FIR target

$$y_{t+1} = \sum_{k=0}^{K-1} h_k u_{t-k} + \varepsilon_t, \tag{22}$$

with $h_k = 0$ for $k \geq K$. Equation equation 7 groups lags by residue class modulo N. Define

$$v_j = \frac{g^{2j}}{1 - g^{2N}}, \qquad c_j = \sum_m h_{j+mN} g^{j+mN}. \tag{23}$$

The captured predictive signal is $C_N = \sum_j c_j^2 / v_j$, and

$$R_N = \frac{\|h\|_2^2 - C_N}{\sigma_\varepsilon^2}. \tag{24}$$

This gives the population analytic sizing curve used in the controlled experiments.

F.2 PROJECTION FORM

If $\Sigma = LL^\top$, let $q = L^\top H$ and $A_N = F_N L$. If P_N denotes the orthogonal projector onto the row space of A_N, then

$$E_N = \|(I - P_N)q\|_2^2. \tag{25}$$

Thus representational sufficiency depends on alignment with the task-relevant direction, not on rank alone.

F.3 FINITE-SAMPLE SAFE ANALYTIC SELECTOR

The controlled one-shot experiments estimate H, Σ, and the linear-oracle residual risk from training data and use coefficient draws to form a simultaneous upper envelope U_N for the normalized excess risk $R_N = E_N / \mathcal{R}^*_{\mathrm{lin}}$. For a declared NRMSE tolerance δ, the certified selector is

$$\widehat{N}_{\mathrm{safe}} \in \arg\min_N \{\mathcal{C}(N) : U_N \leq (1+\delta)^2 - 1\}, \tag{26}$$

and returns ABSTAIN when the feasible set is empty. The broad benchmark uses $\delta = 0.02$, hence a relative-MSE threshold of 0.0404. This selector is exact only under the linear finite-history assumptions; the broad screening procedure therefore adds the separate linear-adequacy and nonlinear-guard gates.

G IMPLEMENTATION DETAILS FOR SCREENING AND SIZING

G.1 ARCHITECTURE-SCREEN IMPLEMENTATION

For $N \in \{8, 16, 24, 32, 48, 64, 96, 128\}$, the normalized representation loss is estimated with 1,000 Gaussian coefficient draws and a HAC covariance estimate; lag covariance is estimated by plug-in. The nonlinear guard uses random-$\tanh$ ESNs of fixed size $N = 64$ with seeds $\{0, 1, 2\}$. For each seed, we form the paired loss $e^2_{\mathrm{cyc}} - (1.02)^2 e^2_{\mathrm{esn}}$ and a Bonferroni-adjusted one-sided 95% Newey–West upper bound with maximum lag 24. The post-decision audit searches $N \in \{8, 16, 24, 32, 48, 64, 96, 128\}$ over ten seeds, with average recurrent degree six, spectral rescaling to 0.7, input scale 0.1, and ridge grid $\{10^{-8}, 10^{-6}, 10^{-4}, 10^{-2}, 1\}$.

Broad benchmark composition. The 72 configurations comprise 36 linear controlled tasks from FIR, AR, and VAR processes; six NARMA tasks; 12 deterministic-chaos tasks from Lorenz–63, Rössler, and Mackey–Glass trajectories; and 18 configurations from the six real datasets in Table 2. Real-data horizons are $\{1, 6, 24\}$ and chaotic-system horizons are $\{1, 5, 10, 25\}$.

G.2 Held-out sizing implementation

The held-out study uses horizons $h \in \{1, 24\}$, chronological folds $\{0.5, 0.65, 0.8\}$, history length 96, cycle gain $g = 0.7$, and ridge parameters $\{10^{-6}, 10^{-4}, 10^{-2}, 1\}$. Practical equivalence is evaluated with 2,000 paired moving-block bootstrap resamples of test residuals using block length 24. The primary margin is a 2% relative increase in NRMSE; an absolute-NRMSE margin of 0.005 is retained only as sensitivity analysis.

H Additional Held-Out Sizing Results

Table 3: Overall held-out sizing summary. Percentage size reduction is computed per configuration before averaging.

Method	Mean N	Mean test NRMSE	Mean edges	Sizes queried
Near-best selector	43.0	0.3511	43.0	8.25
Exact validation minimizer	90.0	**0.3506**	90.0	10.00

Post-hoc oracle diagnostic. Relative to the unattainable test-optimal size, the selected cycle is within 2% of the minimum test NRMSE on 91.7% of configurations. The median absolute test regret is 1.30×10^{-4}. The two exceptions show validation–test shift; this comparison is diagnostic only and is never used for model selection.

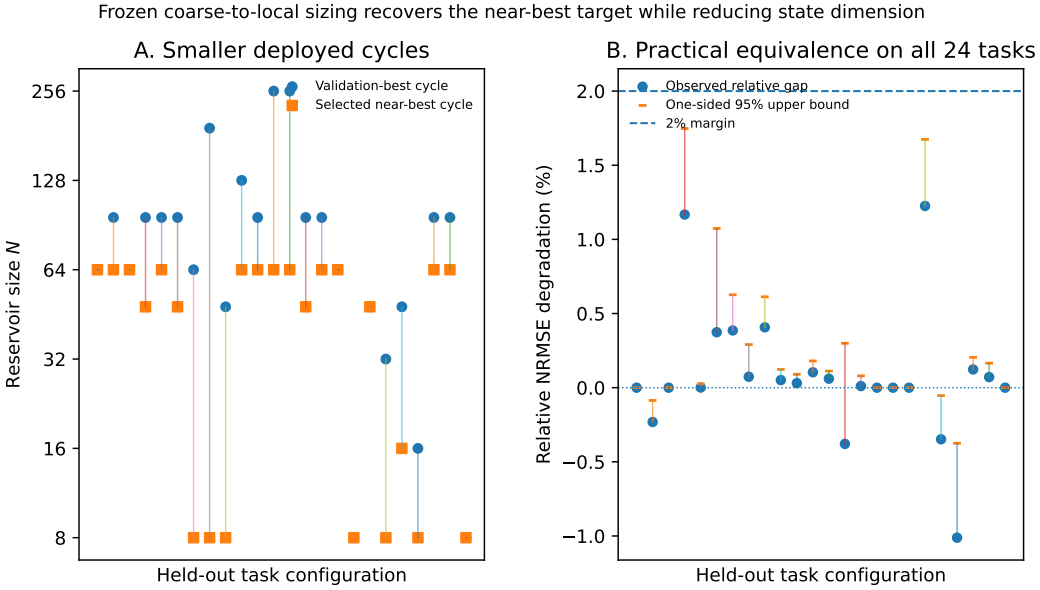

Figure 2: Held-out sizing over 24 configurations. Left: selected near-best cycle size versus the exact validation minimizer. Right: observed relative test-NRMSE gap and one-sided 95% moving-block-bootstrap upper bound; all upper bounds remain below the prespecified 2% practical-equivalence margin.

Table 4 shows that the average size reduction is not driven by a single dataset. The strongest reduction occurs on ETTm1, while all four datasets satisfy the 2% relative-equivalence criterion under the one-sided 95% bound on every task.

Table 4: Held-out sizing by dataset. "Size red." is the mean per-task fractional reduction relative to the exact validation minimizer.

Dataset	Tasks	Mean selected N	Mean val-best N	Size red.
Electricity	6	58.7	85.3	27.8%
ETTm1	6	36.0	130.7	70.8%
Traffic	6	49.3	94.7	26.4%
Weather	6	28.0	49.3	43.1%

I Screening Flow and Comparison Budgets

This appendix provides additional detail on how the architecture-screening decision is constructed and how its computational burden compares with direct reservoir fitting. The screening procedure is intentionally sequential: inexpensive training-only checks are applied first, and more costly comparisons are used only when the earlier conditions are satisfied. The accompanying figures and tables summarize this decision flow and report the corresponding comparison budgets, separating the cost of making a recommendation from the later held-out evaluation used to assess that recommendation.

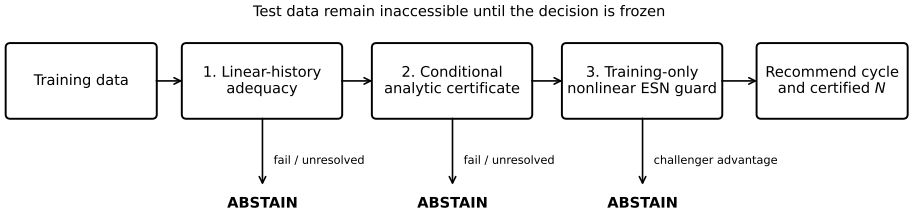

Figure 3: Training-only architecture screening. Test data remain inaccessible until the recommendation or abstention is saved. The analytic certificate is conditional on the linear-history model; the random-$\tanh$ ESN guard protects against a material nonlinear advantage.

Table 5: Comparison budgets. The guard is deliberately cheaper than the post-decision random-ESN audit.

Component	Sizes	Seeds	Tuned quantity	Data used
Cycle certificate	8–128 (8 sizes)	deterministic	N via certificate	training only
Nonlinear guard	fixed $N = 64$	3	ridge	training-only holdout
Random-ESN audit	8–128 (8 sizes)	10	N, seed, ridge	outer validation
Cycle sizing	8–256 (10 sizes)	deterministic	N, ridge	validation

J Broad Benchmark Family Breakdown

The broad screening benchmark combines several qualitatively different forecasting families in order to test the recommendation rule across both settings that closely match its assumptions and settings that deliberately violate them. These include linear stochastic processes, nonlinear synthetic benchmarks, chaotic dynamical systems, and real-world time series. Table 6 breaks down recommendation coverage and observed safety by benchmark family, making it possible to distinguish overall behavior from effects that are concentrated in a particular class of tasks.

Table 6: Family-level broad-benchmark results. Gap is $(e_{\mathrm{cycle}} - e_{\mathrm{random}})/e_{\mathrm{random}}$ on recommended tasks.

Family	Tasks	Rec.	Coverage	Mean gap	Online red.
FIR	16	6	37.5%	-1.51%	90.1%
AR	12	3	25.0%	$+0.07\%$	77.7%
VAR	8	2	25.0%	$+0.17\%$	55.8%
NARMA	6	6	100%	-0.04%	89.8%
Lorenz–63	4	0	0%	–	–
Rössler	4	0	0%	–	–
Mackey–Glass	4	0	0%	–	–
Electricity	3	1	33.3%	-1.57%	92.8%
Weather	3	1	33.3%	$+0.11\%$	89.9%
Traffic	3	1	33.3%	-10.01%	92.8%
ETTh1	3	1	33.3%	$+0.02\%$	92.8%
ETTm1	3	2	66.7%	$+0.43\%$	97.8%
Exchange	3	0	0%	–	–

K Real-Data Analytic Stress Audit

A six-task real-data analytic stress test provides a concrete boundary case for one-shot covariance sizing. The covariance plug-in and covariance-safe selectors are feasible on four of six tasks; the infeasible cases are the two Metro tasks. When feasible, the safe selector tends to be large (mean $N = 264$ among feasible tasks). We therefore do not present covariance-only sizing as a universal deployment method; in the main paper it is used only inside the guarded screening logic where failed certification leads to abstention.

L Prospective Transfer and Diagnostic-Audit Details

The prospective transfer study uses the most recent 4,500 observations from each series, with 150 burn-in points, 2,200 training samples, 600 validation samples, and 600 test samples. History length is 64. Depending on dataset frequency, horizons are ETTh1 $\{1, 24\}$, Exchange $\{1, 5, 20\}$, Weather $\{1, 6, 144\}$, and Traffic/Metro $\{1, 24, 168\}$. Candidate cycle and random-ESN sizes are $\{16, 32, 64, 128\}$; model and ridge choices are made on validation data and saved before test evaluation. The nonlinear-gain decision itself uses training data only.

Table 7 summarizes the prospective transfer results by dataset group. The "Suitable" column is descriptive only: it reports the fraction of configurations for which the cycle test NRMSE is within 2% of the validation-selected random ESN, and is therefore unavailable to the training-only decision rule. The final column indicates whether at least one recommended configuration in that dataset group violates the 2% margin.

Table 7: Prospective transfer results by dataset. "Suitable" is the descriptive fraction of configurations whose cycle test NRMSE is within 2% of the validation-selected random ESN; it is not available to the decision rule.

Dataset	Configs.	Recommended	Suitable	Unsafe dataset?
ETTh1	14	14	100.0%	No
Exchange	24	20	62.5%	Yes
Metro	3	1	100.0%	No
Traffic	24	17	58.3%	Yes
Weather	18	17	66.7%	Yes
All	83	69	69.9%	3/5 groups

For the exploratory diagnostic audit, each model is trained on four dataset groups and evaluated on the fifth, cycling through all five groups. No feature-set or classifier combination attains useful cross-dataset transfer. The best AUC is 0.570 for nonlinear gain alone; the compact diagnostic set (nonlinear gain, linear R^2, and Hankel summaries) attains AUC 0.397 with logistic regression; and the full structural set attains 0.119 with logistic regression and 0.277 with a random forest. These below-chance values are not interpreted as evidence for reversing a score because the direction of

association changes across held-out datasets. The pooled associations of seasonal autocorrelation and seasonal-difference ratio with cycle penalty are modest ($\rho = 0.309$ and -0.304, respectively) and are not prospective selection evidence.

Although the diagnostic audit does not provide a reliable selector, direct post-test comparison shows that cycles remain competitive on many configurations. Table 8 summarizes this descriptive accuracy–cost frontier for several allowed NRMSE penalties. The table is not used for model selection; it characterizes how often a cycle is both within a specified accuracy margin and cheaper than the validation-selected random-$\tanh$ ESN.

Table 8: Descriptive real-data accuracy–cost frontier. "Cheaper" means fewer online multiplies than the validation-selected random-$\tanh$ ESN.

Allowed NRMSE penalty	Within margin	Within and cheaper	Fraction of 83
0%	49	41	49.4%
1%	53	45	54.2%
2%	58	50	60.2%
5%	68	59	71.1%
10%	78	68	81.9%

Counting and interpretation. The 83 rows are configurations, not independent datasets. Their value is coverage across series and horizons under a fixed temporal split, while uncertainty about transfer to unseen domains is still dominated by the five dataset groups. Table 7 should therefore be interpreted at both the configuration and dataset-group levels, rather than as 83 independent observations. Likewise, Table 8 is a post-test characterization of the observed architecture frontier. It supports the descriptive conclusion that cycles are frequently competitive, but it cannot be used to recommend a cycle on a future dataset without fitting and validating candidate architectures.