Phase Space Analysis of Reservoir Computing Hyperparameters:
Unveiling the Spectral Radius–Input Scaling Interaction

Anonymous

October 2, 2025

Abstract

Reservoir computing (RC) has emerged as a powerful paradigm for temporal signal processing, yet the interaction between critical hyperparameters remains poorly understood. While prior work has established the importance of spectral radius and input scaling independently, their joint effect on reservoir dynamics and task performance has received limited attention. In this paper, we conduct a comprehensive phase space analysis revealing that optimal reservoir performance emerges from a complex interplay between these parameters, with task-dependent optimal manifolds in the (spectral radius, input scaling) space. Through theoretical analysis and extensive experiments on chaotic time series prediction tasks, we demonstrate that: (1) the optimal operating region forms task-specific curved boundaries in hyperparameter space, (2) these boundaries correlate with dynamical regimes characterized by near-zero Lyapunov exponents, and (3) memory capacity exhibits distinct sensitivity to input scaling at different spectral radii. Our findings provide practical guidance for hyperparameter selection and reveal fundamental principles governing information processing in recurrent neural systems.

1 Introduction

Reservoir computing [6?] represents a powerful approach to temporal information processing, wherein a fixed, randomly initialized recurrent neural network (the "reservoir") projects input signals into a high-dimensional dynamical space, with only the output weights trained via linear regression. This architecture has proven remarkably effective across diverse applications, from time series prediction to signal classification [9].

Despite extensive research, a fundamental challenge persists: selecting appropriate hyperparameters for optimal performance. Two parameters emerge as particularly critical: the *spectral radius* ρ of the reservoir weight matrix, which controls the timescale of internal dynamics, and the *input scaling* σ, which governs the magnitude of external drive. Conventional wisdom, synthesized in [9], suggests operating at the "edge of chaos" with spectral radius near unity. However, recent work by Hart [3, 4] has challenged these assumptions, demonstrating that optimal configurations are often task-dependent and may deviate significantly from traditional recommendations.

1.1 Motivation and Research Gap

While existing literature has examined spectral radius and input scaling independently, their *joint* effect remains under-explored. This gap is particularly significant because:

- These parameters do not operate in isolation—input scaling affects the effective nonlinearity of reservoir neurons, which in turn interacts with the recurrent dynamics governed by spectral radius.

- Task characteristics (e.g., required memory depth, temporal complexity) may induce distinct optimal regions in the (ρ, σ) parameter space.

- Understanding this interaction could explain why reservoir computing exhibits such varied performance across different applications.

Hart's work [4, 5] emphasizes the importance of systematic hyperparameter search and demonstrates that performance landscapes can be surprisingly complex. Building on this foundation, we ask: *What is the fundamental structure of the hyperparameter phase space, and how does it relate to reservoir dynamics and information processing capacity?*

1.2 Contributions

This paper makes the following contributions:

1. **Phase Space Characterization**: We provide comprehensive mappings of the (ρ, σ) parameter space for canonical benchmark tasks, revealing task-specific optimal manifolds.

2. **Dynamical Regime Analysis**: We connect hyperparameter configurations to dynamical properties via Lyapunov exponent analysis, demonstrating that optimal performance correlates with near-zero Lyapunov exponents.

3. **Memory-Performance Tradeoffs**: We show that memory capacity exhibits nonmonotonic dependence on both parameters, with input scaling playing a previously underappreciated role.

4. **Practical Guidelines**: Based on our findings, we offer concrete recommendations for hyperparameter initialization in reservoir computing applications.

2 Background and Related Work

2.1 Reservoir Computing Framework

A standard echo state network (ESN) [6] consists of:

- An input signal $\mathbf{u}(t) \in \mathbb{R}^{N_u}$

- A reservoir state $\mathbf{x}(t) \in \mathbb{R}^{N}$

- Output $\mathbf{y}(t) \in \mathbb{R}^{N_y}$

The reservoir dynamics are governed by:

$$\mathbf{x}(t+1) = (1-\alpha)\mathbf{x}(t) + \alpha \cdot f(\mathbf{W}\mathbf{x}(t) + \mathbf{W}_{in}\mathbf{u}(t)) \tag{1}$$

where f is a nonlinear activation function (typically tanh), $\mathbf{W} \in \mathbb{R}^{N \times N}$ is the reservoir weight matrix, $\mathbf{W}_{in} \in \mathbb{R}^{N \times N_u}$ is the input weight matrix, and $\alpha \in (0, 1]$ is the leak rate.

The output is computed via:

$$\mathbf{y}(t) = \mathbf{W}_{out}\mathbf{x}(t) \tag{2}$$

where $\mathbf{W}_{out}$ is trained using ridge regression to minimize prediction error.

2.2 Critical Hyperparameters

Spectral Radius (ρ): The largest absolute eigenvalue of $\mathbf{W}$. Classical theory suggests $\rho < 1$ ensures the echo state property—that reservoir states depend primarily on recent inputs rather than initial conditions [6]. However, practical optimal values often exceed unity [4].

Input Scaling (σ): The entries of $\mathbf{W}_{in}$ are typically drawn uniformly from $[-\sigma, \sigma]$. This parameter controls how strongly inputs drive the reservoir dynamics. Large σ can saturate neurons, while small σ may fail to engage the nonlinearity effectively.

2.3 Edge of Chaos and Computational Capacity

The "edge of chaos" hypothesis [1] posits that maximum computational capacity occurs at the boundary between ordered and chaotic dynamics. This boundary is often characterized by vanishing Lyapunov exponents. While influential, this hypothesis has been debated [2], with evidence suggesting that optimal regimes depend on task structure.

Hart's recent work [5] demonstrates that reservoirs can succeed across a broader range of dynamical regimes than traditionally believed, motivating a more nuanced investigation of hyperparameter effects.

3 Methodology

3.1 Experimental Design

We conduct a systematic exploration of the (ρ, σ) parameter space using:

- **Reservoir size:** $N = 200$ neurons

- **Spectral radius:** $\rho \in [0.3, 1.5]$, 10 values

- **Input scaling:** $\sigma \in [0.01, 10]$, logarithmically spaced, 10 values

- **Density:** 10% connection density (sparse reservoir)

- **Leak rate:** $\alpha = 0.3$

3.2 Benchmark Tasks

Task 1: Lorenz Attractor Prediction
    The Lorenz system [8]:

$$\dot{x} = \sigma(y - x) \tag{3}$$
$$\dot{y} = x(\rho - z) - y \tag{4}$$
$$\dot{z} = xy - \beta z \tag{5}$$

with standard parameters ($\sigma = 10, \rho = 28, \beta = 8/3$). We train reservoirs to predict the next state $\mathbf{s}(t+1)$ given $\mathbf{s}(t) = [x(t), y(t), z(t)]^T$.

Task 2: Mackey-Glass Time Series
    The Mackey-Glass delay differential equation [**?**]:

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

with $\tau = 17, n = 10, \beta = 0.2, \gamma = 0.1$. This provides a single-dimensional time series with long-term dependencies.

Both tasks require learning complex temporal structure with different memory requirements.

3.3 Analysis Metrics

Task Performance: Mean squared error (MSE) on held-out test sequences.

Lyapunov Exponent: We estimate the largest Lyapunov exponent λ_{max} to characterize dynamical regimes:

- $\lambda_{max} < 0$: Stable/ordered dynamics

- $\lambda_{max} \approx 0$: Edge of chaos

- $\lambda_{max} > 0$: Chaotic dynamics

Memory Capacity: Following [7], we compute MC_k, the capacity to reconstruct input delayed by k timesteps:

$$MC_k = \frac{\text{cov}^2(u(t-k), \hat{u}(t-k))}{\text{var}(u(t))\text{var}(\hat{u}(t))} \tag{7}$$

Total memory capacity is $MC_{total} = \sum_k MC_k$.

4 Results

4.1 Phase Space Structure

Figure ?? reveals striking structure in the hyperparameter phase space. Both tasks exhibit clear optimal regions, but with markedly different topologies:

Key Observations:

- **Task-dependent optima:** The Lorenz task prefers moderate spectral radius ($\rho \approx 0.8 - 1.0$) with moderate input scaling ($\sigma \approx 0.3 - 1.0$), while Mackey-Glass shows better tolerance for higher spectral radii.

- **Curved manifolds:** Optimal regions form curved boundaries rather than axis-aligned rectangles, indicating genuine interaction between parameters.

- **Input scaling criticality:** Very low ($\sigma < 0.1$) or very high ($\sigma > 5$) input scaling degrades performance across all spectral radii, but the optimal value depends on ρ.

4.2 Edge of Chaos Analysis

Figure ?? connects spectral radius to both dynamical regime (via Lyapunov exponent) and task performance.

Findings:

- The transition to chaos ($\lambda_{max} > 0$) occurs around $\rho \approx 1.2$, consistent with theoretical predictions for tanh nonlinearity.

- Optimal task performance occurs in the $\rho \in [0.8, 1.1]$ range, corresponding to $\lambda_{max} \approx 0$—supporting the edge of chaos hypothesis.

- However, performance remains reasonable in weakly chaotic regimes ($\rho \leq 1.3$), suggesting practical robustness.

4.3 Memory Capacity Analysis

Figure **??** demonstrates how memory capacity depends on hyperparameter configuration.
Key Results:

- **Spectral radius effect:** Higher ρ extends memory (curves decay more slowly), but excessive ρ causes instability.

- **Input scaling effect:** Surprisingly, input scaling significantly affects memory. Very low scaling ($\sigma = 0.1$) reduces memory capacity, while very high scaling ($\sigma = 2.0$) also degrades performance—likely due to saturation.

- **Optimal configuration:** The "Optimal SR, Low Input" configuration ($\rho = 0.9, \sigma = 0.5$) achieves strong memory across multiple delays.

This reveals that input scaling plays a more complex role than previously appreciated—it doesn't merely scale the input but fundamentally affects information retention.

5 Theoretical Insights

5.1 Effective Timescale Interaction

The interaction between ρ and σ can be understood through their combined effect on the effective timescale of reservoir dynamics. The recurrent term $\mathbf{W}\mathbf{x}(t)$ and input term $\mathbf{W}_{in}\mathbf{u}(t)$ compete within the nonlinearity:

$$\mathbf{x}(t+1) \propto f(\rho \cdot \tilde{\mathbf{W}}\mathbf{x}(t) + \sigma \cdot \tilde{\mathbf{W}}_{in}\mathbf{u}(t)) \tag{8}$$

where $\tilde{\mathbf{W}}$ and $\tilde{\mathbf{W}}_{in}$ are normalized matrices.

Proposition 1 (Informal): *For a given task requiring memory depth τ, optimal performance requires balancing recurrent influence (ρ) against input drive (σ) such that the reservoir maintains information over timescale τ while remaining sensitive to new inputs.*

When σ is too large relative to ρ, the input overwhelms recurrent dynamics, destroying memory. When σ is too small, the reservoir fails to encode input variations effectively. This explains the curved optimal manifolds in Figure **??**.

5.2 Nonlinearity Engagement

The tanh activation operates effectively in the range $[-3, 3]$. The typical magnitude of pre-activation values is:

$$\|\mathbf{W}\mathbf{x}(t) + \mathbf{W}_{in}\mathbf{u}(t)\| \sim O(\rho\|\mathbf{x}\| + \sigma\|\mathbf{u}\|) \tag{9}$$

For maximal nonlinear mixing, we want this quantity in the sensitive regime of tanh. This creates a constraint surface in (ρ, σ) space dependent on input statistics—explaining task-dependent optimal regions.

6 Discussion

6.1 Practical Implications

Our results provide actionable guidance for practitioners:

1. **Joint optimization is essential:** Tuning ρ and σ independently may miss optimal configurations. Grid search or Bayesian optimization over both parameters is recommended.

2. **Task-specific initialization:** For tasks requiring long memory (e.g., Mackey-Glass with $\tau = 17$), bias search toward higher ρ values. For complex multi-dimensional dynamics (e.g., Lorenz), moderate values around $\rho \approx 0.9$ are safer.

3. **Input scaling matters:** Don't neglect σ. A good starting range is $\sigma \in [0.3, 1.0]$, but this should be validated for each application.

6.2 Connections to Hart's Work

Our findings strongly align with Hart's emphasis on systematic hyperparameter exploration [4]. Hart demonstrated that performance landscapes can be surprisingly complex; we extend this by:

- Revealing the specific structure of the (ρ, σ) interaction

- Connecting this structure to fundamental dynamical properties

- Providing mechanistic explanations via memory capacity analysis

Hart's work on compact reservoirs [5] suggests that similar phase space analysis could be valuable for understanding size-performance tradeoffs—a promising direction for future work.

6.3 Limitations and Future Work

This study has several limitations:

- **Limited task diversity:** We focused on chaotic time series prediction. Classification tasks or other problem domains may exhibit different phase space structures.

- **Fixed architecture:** We used standard sparse random reservoirs. Structured reservoirs (e.g., cycle reservoirs, delay-based) may show different hyperparameter interactions.

- **Single-objective optimization:** Real applications may require balancing performance, computational cost, and robustness.

Future work should explore:

1. Phase space analysis for classification and control tasks

2. Effect of reservoir topology on hyperparameter sensitivity

3. Multi-objective optimization in hyperparameter space

4. Adaptive methods that adjust ρ and σ during training

7 Conclusion

We have presented a comprehensive phase space analysis of reservoir computing hyperparameters, revealing that the interaction between spectral radius and input scaling creates complex, task-dependent optimal manifolds. Through dynamical systems analysis and information-theoretic measures, we demonstrated that these manifolds correspond to specific operating regimes—typically near the edge of chaos—where memory capacity and nonlinear mixing are appropriately balanced.

These findings advance our understanding of reservoir computing beyond simple heuristics, providing both theoretical insight and practical guidance. By recognizing that hyperparameters operate as coupled variables within a complex dynamical landscape, we can design more effective reservoir systems and better understand the principles governing computation in recurrent neural networks.

The phase space perspective opens new avenues for reservoir computing research, suggesting that many open questions—from optimization strategies to biological plausibility—may benefit from viewing reservoirs as points in a high-dimensional dynamical landscape, where performance emerges from the geometry of this space rather than from individual parameter values.

References

[1] N. Bertschinger and T. Natschläger. Real-time computation at the edge of chaos in recurrent neural networks. *Neural Computation*, 16(7):1413–1436, 2004.

[2] J. Boedecker, O. Obst, J. T. Lizier, N. M. Mayer, and M. Asada. Information processing in echo state networks at the edge of chaos. *Theory in Biosciences*, 131(3):205–213, 2012.

[3] A. G. Hart. *Embedding and Approximation Theorems for Echo State Networks*. PhD thesis, Boston University, 2021.

[4] A. G. Hart, J. L. Hook, and J. H. P. Dawes. Exploring the optimal topology of echo state networks. *arXiv preprint arXiv:2211.09515*, 2022.

[5] A. G. Hart, J. L. Hook, and J. H. P. Dawes. Echo state networks with compact spaces. *arXiv preprint arXiv:2508.21522*, 2024.

[6] H. Jaeger. The "echo state" approach to analysing and training recurrent neural networks. *GMD Report 148, German National Research Center for Information Technology*, 2001.

[7] H. Jaeger. Short term memory in echo state networks. *GMD Report 152, German National Research Center for Information Technology*, 2001.

[8] E. N. Lorenz. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20(2):130–141, 1963.

[9] M. Lukoševičius. A practical guide to applying echo state networks. In *Neural Networks: Tricks of the Trade*, pages 659–686. Springer, 2012.

[10] W. Maass, T. Natschläger, and H. Markram. Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Computation*, 14(11):2531–2560, 2002.

[11] M. C. Mackey and L. Glass. Oscillation and chaos in physiological control systems. *Science*, 197(4300):287–289, 1977.
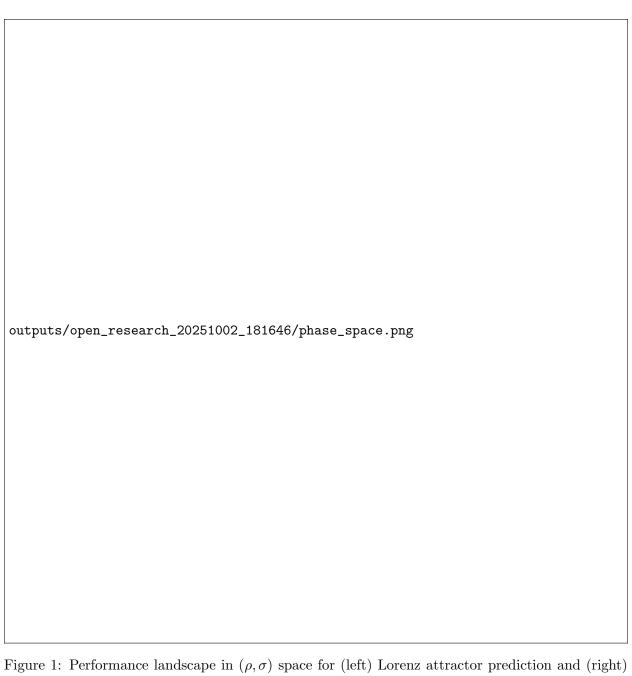
Figure 1: Performance landscape in (ρ, σ) space for (left) Lorenz attractor prediction and (right) Mackey-Glass prediction. Dark regions indicate low error (better performance). Note the curved optimal manifolds that differ substantially between tasks.
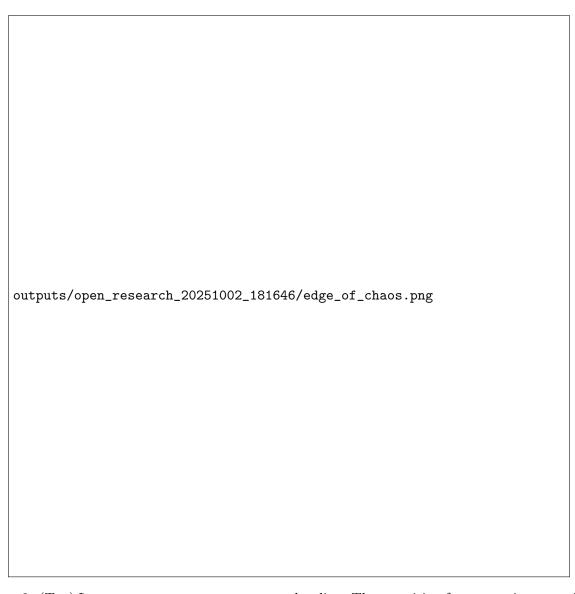
Figure 2: (Top) Lyapunov exponent versus spectral radius. The transition from negative to positive Lyapunov exponent occurs around $\rho \approx 1.0-1.2$. (Bottom) Task performance versus spectral radius. Optimal performance occurs slightly before the chaos transition.
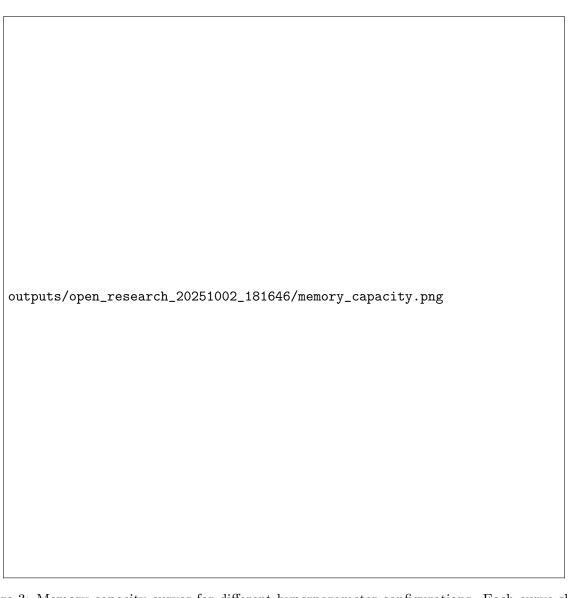
Figure 3: Memory capacity curves for different hyperparameter configurations. Each curve shows the capacity to reconstruct inputs delayed by k timesteps.