

UNIVERSIDADE FEDERAL DO PARANÁ

JOÃO VITOR VIEIRA FLAUZINO

A FRAMEWORK TO QUANTIFY DISORDER IN TEMPORAL DATA

CURITIBA

2025

JOÃO VITOR VIEIRA FLAUZINO

A FRAMEWORK TO QUANTIFY DISORDER IN TEMPORAL DATA

Dissertação apresentada ao *Programa de Pós-graduação em Física* do *Setor de Ciências Exatas* da *Universidade Federal do Paraná*, como parte dos requisitos necessários à obtenção do grau de Mestre em Física.

Supervisor: Prof. Dr. Sergio Roberto Lopes.

CURITIBA

2025

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)
UNIVERSIDADE FEDERAL DO PARANÁ
SISTEMA DE BIBLIOTECAS – BIBLIOTECA DE CIÊNCIA E TECNOLOGIA

Flauzino, João Vitor Vieira

A framework to quantify disorder in temporal data / João Vitor Vieira
Flauzino. – Curitiba, 2025.
1 recurso on-line : PDF.

Dissertação (Mestrado) - Universidade Federal do Paraná, Setor de
Exatas, Programa de Pós-Graduação em Física.

Orientador: Sergio Roberto Lopes

1. Desordem. 2. Entropia (Teoria da informação). 3. Comportamento
caótico nos sistemas. 4. Ruído – Medição. I. Universidade Federal do Paraná.
II. Programa de Pós-Graduação em Física. III. Lopes, Sergio Roberto. IV.
Título.

Bibliotecário: Douglas Lenon da Silva CRB-9/1892

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação FÍSICA da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de **JOÃO VITOR VIEIRA FLAUZINO**, intitulada: **"A framework to quantify disorder in temporal data"**, sob orientação do Prof. Dr. SERGIO ROBERTO LOPES, que após terem inquirido o aluno e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 18 de Julho de 2025.

Assinatura Eletrônica

20/07/2025 08:19:15.0

SERGIO ROBERTO LOPES

Presidente da Banca Examinadora

Assinatura Eletrônica

18/07/2025 17:25:28.0

CESAR MANCHEIN

Avaliador Externo (UNIVERSIDADE DO ESTADO DE SANTA CATARINA)

Assinatura Eletrônica

23/07/2025 07:35:21.0

THIAGO DE LIMA PRADO

Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

AGRADECIMENTOS

Agradeço aos meus pais, Ilson e Maria Diva, pelo amor incondicional e por todo apoio. Sou grato à Maya, minha finada filha de quatro patas, afinal as principais ideias desse trabalho vieram enquanto passeava com ela.

Aos docentes do IFMT - Campus Fronteira Oeste e da UFPR que contribuíram com minha formação. Em especial, ao meu orientador, Prof. Dr. Sergio Roberto Lopes, pela orientação e dedicação. Bem como aos professores Thiago de Lima Prado, Jürgen Kurths e Norbert Marwan por participarem ativamente da minha pesquisa e ao professor Ricardo Luiz Viana por me apoiar ao longo da formação.

Ao Programa de Pós-Graduação em Física e à Universidade Federal do Paraná pela oportunidade e suporte. Ao CNPq e CAPES pelo apoio financeiro.

Por fim, sou grato ao tempo e à Jesus Cristo.

*“Umas coisas nascem de outras,
enroscam-se,
desatam-se,
confundem-se,
perdem-se,
e o tempo vai andando sem se perder a si. ”*
— (Machado de Assis, "Esaú e Jacó")

RESUMO

A desordem é onipresente em dados de sistemas físicos—seja decorrente de ruído de medição ou de variabilidade intrínseca ao sistema—e constitui uma propriedade informacional fundamental. Quantificar essa desordem é desafiador porque as séries temporais costumam ser curtas e exibem dinâmicas híbridas de ordem e desordem, flutuações (des)correlacionadas significativas e assinaturas que podem ser ambíguas, por exemplo, confundindo caos com estocasticidade. Neste trabalho, introduzimos um arcabouço formal que quantifica a desordem ao explorar a ação do grupo simétrico no espaço de recorrência, o que gera classes de equivalência de microestados de recorrência equiprováveis. Ao maximizar a entropia de informação resultante, derivamos uma métrica robusta que discrimina entre dinâmicas caóticas, processos estocásticos correlacionados e ruído não correlacionado—mesmo quando os conjuntos de dados são reduzidos. Nosso método também caracteriza a estrutura de correlação do ruído que perturba sistemas determinísticos. Quando aplicado a registros paleoclimáticos, os mínimos no quantificador de desordem coincidem com grandes transições do Cenozóico, marcando épocas dominadas por mecanismos de forçamento específicos. Testes estatísticos extensivos em modelos paradigmáticos e em dados empíricos demonstram que esse quantificador é uma ferramenta valiosa para avaliar a natureza subjacente de fenômenos complexos.

Palavras-chave: Desordem; recorrências; grupo simétrico; entropia da informação; caos; ruído.

ABSTRACT

Disorder is ubiquitous in data from physical systems—whether arising from measurement noise or inherent system variability—and represents a fundamental informational property. Quantifying this disorder is challenging because time series are often short and exhibit hybrid order–disorder dynamics, significant (un)correlated fluctuations, and signatures that can be ambiguous, e.g. conflating chaos with stochasticity. In this work, we introduce a formal framework that quantifies disorder by exploiting the action of the symmetric group on recurrence space, which induces equivalence classes of equiprobable recurrence microstates. By maximizing the resulting information entropy, we derive a robust metric that discriminates among chaotic dynamics, correlated stochastic processes, and uncorrelated noise—even when datasets are small. Our method also characterizes the correlation structure of noise perturbing deterministic systems. When applied to paleoclimate records, minima in the disorder quantifier coincide with major Cenozoic transitions, marking epochs dominated by specific forcing mechanisms. Extensive statistical testing on paradigmatic models and empirical data demonstrates that this quantifier is a valuable tool for assessing the underlying nature of complex phenomena.

Keywords: Disorder; recurrences; symmetric group; information entropy; chaos; noise.

LIST OF FIGURES

5.1	(a) The mean of the disorder quantifier $\langle \Xi \rangle$ with its corresponding standard deviation shaded on each curve as a function of the length of the time series K for the logistic map, WGN, PGN and RGN. (b) The p -values of Welch's t-test $[A, B]$ comparing the mean of disorder for the model A and model B as a function of the time series length [63].	49
5.2	Boxplots with overlaid violin plots showing the distribution of the disorder quantifier Ξ for each discrete dynamical system under parameter values p_1 and p_2 [63].	51
5.3	Boxplots with violin overlays showing the distribution of the disorder quantifier Ξ for each continuous dynamical system under parameter values p_1 (blue) and p_2 (red) [63].	52
5.4	Boxplots with violin plots showing the values of the proposed quantifier Ξ for each stochastic process and its parameter values p_1 and p_2 [63].	52
5.5	The mean of the disorder quantifier $\langle \Xi \rangle$ for the autoregressive model. The disorder decreases as the parameter ϕ increases [63].	53
5.6	(a) Mean disorder quantifier $\langle \Xi \rangle$ for the noisy logistic model $y_t(\phi) = x_t + \eta_t(\phi)$, plotted against the variance ratio σ_η^2/σ_x^2 . Each curve corresponds to a different AR correlation parameter ϕ . (b) Corresponding p -values from Welch's t -test comparing the mean disorder between pairs of models as a function of σ_η^2/σ_x^2 [63].	54
5.7	Boxplots with violin plots showing the values of the proposed disorder quantifier for critical cases [63].	55
5.8	Boxplots with violin plots showing the values of the Determinism (DET) for critical cases [63].	56
5.9	Boxplots with violin plots showing the values of the Permutation Entropy (PE) and Multivariate Permutation Entropy (MvPE) for critical cases [63].	56
5.10	Boxplots with violin plots showing the values of the Statistical Complexity (SC) for critical cases [63].	57

5.11	Boxplots with violin plots showing the values of the structurality (Δ) and dynamical complexity (DC) for critical cases [63].	58
6.1	Adapted figure [21] showing the carbon data (CENOGRID), the DET measure, and the spectrogram.	64
6.2	The disorder quantifier Ξ applied to benthic foraminiferal carbon isotope ($\delta^{13}\text{C}$) data as a function of geological age (in million years ago, Ma). The quantifier was computed using a sliding window of length $K = 200$ (2 Ma), with a timescale $N = 4$ (0.04 Ma). The background depicts the mean disorder levels of White Gaussian Noise (WGN), Pink Gaussian Noise (PGN), and Red Gaussian Noise (RGN), with $K = 200$. A linear fit highlights the overall trend. Vertical dotted lines denote stage boundaries.	65

LIST OF TABLES

5.1 Summary of the comparison between the proposed quantifier Ξ and similar
traditional methods. 59

CONTENTS

1	Introduction	13
2	Some prior approaches to measuring complexity	16
2.1	Determinism (DET)	16
2.2	Permutation entropy (PE)	19
2.3	Statistical complexity (SC)	20
2.4	Structurality (Δ)	22
2.5	Dynamical complexity (DC)	23
3	Benchmark models and their challenges	25
3.1	Discrete dynamical systems	25
3.1.1	Logistic map	25
3.1.2	Henon map	25
3.1.3	Ikeda map	26
3.1.4	Towel map	26
3.1.5	Standard map	27
3.2	Continuous dynamical systems	27
3.2.1	Double-gyre system	27
3.2.2	Coupled Rössler system	28
3.2.3	Mackey-Glass system	28
3.2.4	Lorenz system	29
3.2.5	Chua's circuit	29
3.3	Stochastic processes	30
3.3.1	White Gaussian noise (WGN)	30
3.3.2	Autoregressive model (AR)	30
3.3.3	Noisy logistic model	30
3.3.4	Gaussian-noisy logistic model (GNL)	31
3.3.5	Uniform-noisy logistic model (UNL)	31
3.3.6	Power law noise	31

3.3.7	Nonlinear colored noise	32
3.3.8	Bounded random walk	33
3.3.9	Fractional Gaussian noise	33
3.3.10	Langevin equation	33
4	A formalism for disorder quantification in data	35
4.1	Describing the disorder condition in data	35
4.2	Codifying sequences of the data	36
4.3	Verifying consequences of the disorder condition	38
4.4	Measuring data conformity to the disorder condition	41
4.5	Implementing the framework algorithm	43
5	Numerical validations using benchmark models	46
5.1	Methodological considerations on the disorder quantification	47
5.2	Statistical tests for the quantifier evaluation	47
5.3	The disorder quantifier and the data size	48
5.4	The quantifier and the classification of different model regimes	50
5.4.1	Disorder in discrete deterministic systems	50
5.4.2	Disorder in continuous deterministic systems	51
5.4.3	Disorder in stochastic processes	51
5.5	Disorder and past dependence	53
5.6	Disorder in data from deterministic-stochastic mixtures	53
5.7	Comparisons with prior measures	54
5.7.1	Testing determinism (DET)	55
5.7.2	Testing permutation entropy (PE)	56
5.7.3	Testing statistical complexity (SC)	57
5.7.4	Testing structurality (Δ) and dynamical complexity (DC)	57
5.7.5	Summarizing the comparative tests	58
6	Real-world application to understand Earth's dynamics	60
6.1	The Cenozoic dataset	60
6.1.1	Data acquisition and structure	60
6.1.2	Physical interpretation and geological stages	61
6.1.3	Scientific relevance and prior insights	63
6.2	Disorder in the Cenozoic dataset	64
6.2.1	Dominant triggers and disorder minima	65
6.2.2	Theoretical and practical implications	66
7	Conclusions and perspectives	67
A	Appendix: Published articles	69

A.1	Quantifying disorder in data (<i>Physical Review Letters</i>)	69
A.2	Analytical results in calculating the entropy of recurrence microstates (<i>Physica A</i>)	70
A.3	Recurrence microstates for machine learning classification (<i>Chaos</i>)	70
Bibliography		72

Introduction

The analysis of physical systems is deeply rooted in the interpretation of empirical data. Yet, a recurring challenge arises when such data originates from systems governed by a complex interplay of order and disorder—where deterministic laws and stochastic fluctuations coexist [1]. Observed across disciplines from climatology to biophysics, these systems raise fundamental questions about how to characterize their underlying dynamics [2]. Paradoxically, fluctuations themselves often carry rich information about system behavior, particularly when stochastic components exhibit temporal correlations [3, 4].

For example, paleoclimate time-series data may reflect fluctuations stemming from deterministic responses to orbital cycles or stochastic resonance triggered by environmental noise. Disentangling these origins extends beyond theoretical inquiry; it holds critical implications for forecasting system behavior, characterizing signals, identifying noise sources, and detecting shifts in a system’s state [5, 6, 7, 8]. However, this task is hindered by limited temporal resolution, small datasets, experimental artifacts, and the lack of universal tools to rigorously quantify properties such as disorder.

Traditional approaches, such as entropy-based metrics (e.g., Shannon entropy [9], permutation entropy [10]) and complexity measures [11, 12, 13, 14], aim to address these challenges by encoding data into symbolic sequences or probability distributions. Rooted in information theory, these methods leverage entropy frameworks to quantify unpredictability and irregularity. Shannon’s foundational work on information entropy [9], initially developed for communication systems, was later adapted to dynamical systems through metrics like approximate entropy [15] and sample entropy. Subsequent advancements, such as permutation entropy [10], introduced symbolic encoding to map time series into ordinal patterns, enabling statistical characterization of temporal structures. These methods were refined to distinguish deterministic chaos (e.g., Lorenz systems) from time-correlated stochastic processes (e.g., fractional Brownian motion) or uncorrelated noise [16].

The core premise involves transforming temporal data into discrete symbolic sequences or partitioned phase-space regions, constructing probability distributions that reflect dynamical behavior. For instance, symbolic false nearest neighbors [17] and Kaplan’s test

for determinism [18] use embedding techniques to classify trajectories in reconstructed phase spaces, differentiating low-dimensional chaos from high-dimensional stochasticity. Recent innovations, such as direct complexity [14] and stochasticity quantifiers [19], incorporate multiscale and graph-theoretic frameworks to address nonstationary and high-dimensional data. These tools have been applied across domains, from detecting epileptic seizures in EEGs [20] to identifying climatic predictability [21].

Despite their successes, these methods face three key limitations. First, they often conflate disorder with unrelated dynamical features like nonlinearity and nonstationarity. Second, their performance degrades with high-dimensional or short datasets [22, 23], as sparse sampling biases probability distributions. Third, many rely on heuristic parameter choices (e.g., partition size, embedding dimension) rather than first-principles criteria [24]. Symbolic encoding techniques, for example, require ad hoc definitions of embedding dimensions or binning thresholds, while recurrence-based analyses depend on phase-space proximity thresholds. Such subjective parameterization introduces biases and undermines reproducibility. Additionally, in high-dimensional systems like turbulent flows or gene regulatory networks, traditional tools oversimplify complex interactions into scalar metrics, further limiting their utility.

We present a novel methodology grounded in recurrence analysis to rigorously quantify disorder within arbitrary numerical datasets. Central to this approach is the demonstration that entropy maximization constitutes an indispensable methodological strategy for achieving optimal estimation of disorder, thereby eliminating reliance on ad hoc parameterization. The proposed framework enables precise discrimination between three fundamental dynamical regimes: (i) deterministic dynamics, (ii) temporally correlated stochastic processes, and (iii) uncorrelated random noise, through systematic evaluation of disorder magnitude. Furthermore, the method permits identification and characterization of the fundamental properties of system-corrupting noise, including its correlation structure and intensity.

To validate the methodology’s practical utility, we apply it to high-resolution paleoclimate proxy records [21], successfully detecting dynamical transitions in Earth’s Quaternary climate system. This empirical application demonstrates the framework’s capacity to resolve regime shifts and decode noise signatures in real-world systems, establishing its broader relevance for data-driven investigations across scientific domains.

Helpful discussions with Prof. Jürgen Kurths contribute significantly to the formulation of the numerical validation strategy presented herein. Insightful comments and guidance from Prof. Norbert Marwan on the real-world application are hereby acknowledged. The computational algorithm derived from the framework proposed herein represents the main practical outcome of this research, and the valuable contributions of Prof. Thiago de Lima Prado to its development are gratefully acknowledged.

The remainder of this dissertation is organized as follows:

Chapter 2: Some prior approaches to measuring complexity

Critical review of existing complexity metrics (statistical complexity, permutation entropy, recurrence measures) with analysis of their theoretical foundations and practical limitations in disorder quantification.

Chapter 3: Benchmark models and their challenges

Presentation of paradigmatic dynamical systems and stochastic processes serving as testbeds: deterministic chaos, stochastic processes, and hybrid models.

Chapter 4: A formalism for disorder quantification

Derivation of a theoretical framework based on information theory, group theory, and recurrence analysis. Proof of entropy-based quantifier uniqueness under the ideal disorder condition. Detailed algorithmic architecture to directly measure disorder from data based on the proposed framework.

Chapter 5: Numerical validations using benchmark models

Quantitative assessment through statistical tests. Systematic comparison with traditional methods using synthetic benchmarks.

Chapter 6: Real-world application to understand Earth's dynamics

Application to proxy records revealing disorder signatures in Cenozoic climate transitions.

Chapter 7: Conclusions and Perspectives

Synthesis of methodological advances, discussion of limitations, and roadmap for further applications.

Some prior approaches to measuring complexity

In this chapter, we present a comparative analysis between our proposed approach and six well-established methods for characterizing the complexity and disorder of time series: the determinism derived from recurrence quantification analysis (DET) [25], the permutation entropy (PE) [26], the multivariate permutation entropy (MvPE) [27], the statistical complexity (SC) [16], structurality (Δ) [28], and dynamical complexity (DC) [28].

The implementation of the methods was carried out using the Julia language, specifically leveraging the `ComplexityMeasures.jl` [29] and `RecurrenceAnalysis.jl` [30] packages, except for Δ and DC, implemented directly. These tools provide robust and efficient computational frameworks for calculating the respective complexity measures, ensuring reproducibility and methodological consistency across all analyses.

The objective of this comparative study is to evaluate the sensitivity, robustness, and discriminative power of each method when applied to the benchmark datasets, thereby highlighting the strengths and potential limitations of existing approaches relative to our proposed quantifier.

2.1 Determinism (DET)

The determinism (DET) measure, derived from recurrence quantification analysis (RQA), evaluates the predictability and structural regularity of time series based on the proportion of recurrent points that form diagonal lines in the recurrence plot [25]. A high DET value indicates that the system exhibits deterministic and predictable patterns, whereas low values reflect stochastic or highly disordered dynamics.

In the context of this study, we expect DET values to approach 1 for systems characterized by strong deterministic behavior. Conversely, as the system's dynamics become

more disordered or unpredictable, the DET measure is expected to decrease significantly. This sensitivity to the underlying temporal structure makes DET a valuable tool for distinguishing between different regimes of dynamical systems.

Determinism (DET) quantifies the proportion of recurrence points forming diagonal structures in a Recurrence Plot (RP), serving as a crucial indicator of deterministic dynamics in time series analysis. Its computation involves a multi-step procedure. First, given a dataset $\mathfrak{D} = (\mathbf{x}_i)_{i=1}^K$, we construct the recurrence plot by defining the binary matrix

$$R_{i,j} = \Theta(\varepsilon - \|\mathbf{x}_i - \mathbf{x}_j\|), \quad i, j = 1, \dots, K,$$

where ε is a chosen distance threshold and $\Theta(\cdot)$ is the Heaviside step function. A value $R_{i,j} = 1$ indicates that states \mathbf{x}_i and \mathbf{x}_j are within distance ε , i.e. a recurrence has occurred.

Next, we identify diagonal lines of consecutive ones parallel to the main diagonal (the Line of Identity, LOI). A diagonal line of length l at offset Δ satisfies

$$R_{k,k+\Delta} = 1 \quad \text{for } k = i, i+1, \dots, i+l-1,$$

where Δ is the fixed column offset and i is the starting row index. To avoid counting small, spurious recurrences arising from noise, we impose a minimum diagonal length l_{\min} , typically $l_{\min} \geq 2$.

Once all diagonal lines of length $l \geq l_{\min}$ are located, we compile a histogram $P(l)$, which records the number of diagonal lines of each length l . Let L_{\max} denote the maximum diagonal length observed. The DET measure is then defined as

$$\text{DET} = \frac{\sum_{l=l_{\min}}^{L_{\max}} l [P(l)]}{\sum_{\substack{i,j=1 \\ i \neq j}}^K R_{i,j}}.$$

Here, the numerator $\sum_l l P(l)$ counts all recurrence points that belong to diagonal structures of length at least l_{\min} , while the denominator $\sum_{i \neq j} R_{i,j}$ counts the total number of off-diagonal recurrence points. By construction, DET ranges from 0 (no diagonal lines of length $\geq l_{\min}$) to 1 (all recurrences lie on diagonals).

The determination of an appropriate distance threshold ε constitutes a critical methodological step in Recurrence Plot (RP) construction. The literature provides extensive discussion regarding the optimal selection of this parameter. Some primary strategies are:

1. Recurrence Rate Optimization: Calibrates the threshold ε to achieve a target

recurrence density (typically $\text{RR}_{\text{target}} \in [5\%, 20\%]$) by solving:

$$\frac{1}{N^2} \sum_{i,j=1}^N \Theta(\varepsilon - \|\mathbf{x}_i - \mathbf{x}_j\|) = \text{RR}_{\text{target}} \quad (2.1)$$

where $\Theta(\cdot)$ is the Heaviside function, ensuring cross-system comparability [31].

2. Phase Space Topology Methods: Defines recurrence through local attractor density invariants, generating scale-adaptive recurrence plots that capture essential dynamics while minimizing tangential motion artifacts [32, 33].
3. Noise-Adaptive Thresholds: Determines ε proportionally to the system's noise floor:

$$\varepsilon = k \cdot \hat{\sigma}_{\text{noise}}, \quad k \in [3, 5] \quad (2.2)$$

with $\hat{\sigma}_{\text{noise}}$ denoting the estimated noise standard deviation [34].

4. Maximum Entropy Optimization: Maximizes the Shannon entropy of recurrence matrix microstates (submatrices of size $m \times m$):

$$S = - \sum_{i=1}^{\mathcal{N}} p_i \ln p_i, \quad p_i = \frac{n_i}{\sum_k n_k} \quad (2.3)$$

where \mathcal{N} is the number of unique microstates and n_i their occurrence frequency [35].

In summary, DET captures the degree to which the system's trajectory revisits similar states in a deterministic sequence, as evidenced by diagonal lines in the RP. Higher values of DET imply more regular, predictable dynamics, while lower values indicate increased randomness or stochasticity.

In this work, we follow the approach implemented by Westerhold et al. [21]. Specifically, we adopt the Julia function `GlobalRecurrenceRate(0.10)`, which automatically adjusts the threshold to achieve a fixed recurrence rate of 10%. This choice ensures consistency with established practices and facilitates comparability with prior analyses of complex geophysical datasets.

Additional tools of Recurrence Quantification Analysis (RQA), including laminarity and diagonal line entropy, can be combined to provide further interpretations. Nevertheless, DET has frequently been applied to empirical data comparable to that examined here, supporting interpretations like “less predictable” or “more stochastic”, which align with the concept of disorder in data.

2.2 Permutation entropy (PE)

The Permutation Entropy (PE) and its multivariate extension (MvPE) are widely recognized as natural and effective quantifiers of complexity in time series analysis [26, 27]. Both measures are designed to capture the diversity of ordinal patterns present in a dataset, with higher entropy values reflecting greater complexity and randomness in the underlying dynamics.

In general, the expected behavior of these quantifiers is to increase as the complexity of the system grows. For highly regular or deterministic systems, PE and MvPE attain lower values, while in stochastic or chaotic systems with more intricate temporal structures, these measures approach their theoretical maximums.

The calculation of PE requires the specification of two parameters: the embedding dimension m , which defines the length of ordinal patterns, and the embedding delay τ , which sets the spacing between the components of these patterns. In this study, we adopt the parameter values established by Rosso et al. [16], namely $m = 6$ and $\tau = 1$. This choice ensures compatibility with a significant body of prior research and facilitates comparison with results reported in the literature. Additionally, when dealing with multivariate systems, we compute PE using only the x -variable, as originally proposed for univariate data analysis.

Given a univariate time series $(x_t)_{t=1}^K$, we first embed the data in an m -dimensional space using a time delay τ . Concretely, for each time index $t = 1, 2, \dots, K - (m - 1)\tau$, we form the delay vector

$$\mathbf{X}_t = (x_t, x_{t+\tau}, x_{t+2\tau}, \dots, x_{t+(m-1)\tau}).$$

Next, for each vector \mathbf{X}_t , we identify the ordinal pattern—namely, the unique permutation $\pi = (r_0, r_1, \dots, r_{m-1})$ of the indices $(0, 1, \dots, m - 1)$ that satisfies

$$x_{t+r_0\tau} \leq x_{t+r_1\tau} \leq \dots \leq x_{t+r_{m-1}\tau}.$$

When two components are equal, we break ties by their original temporal order, so that if $x_{t+a} = x_{t+b}$ and $a < b$, then $r_a < r_b$. There are $m!$ possible permutations in total; we denote this set by Π . For each permutation $\pi_k \in \Pi$, we estimate its empirical probability by counting how many embedded vectors \mathbf{X}_t realize that pattern:

$$p(\pi_k) = \frac{1}{K - (m - 1)\tau} \sum_{t=1}^{K-(m-1)\tau} \mathbf{1}_{\pi_k}(\mathbf{X}_t),$$

where $\mathbf{1}_{\pi_k}(\mathbf{X}_t)$ is the indicator function that equals 1 if \mathbf{X}_t has ordinal pattern π_k and 0

otherwise. Finally, the permutation entropy of order m is defined as

$$H(m) = - \sum_{k=1}^{m!} p(\pi_k) \log_2 p(\pi_k),$$

using the convention $0 \log_2 0 \equiv 0$. For scale-invariance, one often normalizes

$$H_{\text{norm}}(m) = \frac{H(m)}{\log_2(m!)},$$

so that $H_{\text{norm}}(m)$ lies in $[0, 1]$. In practice, one selects the embedding dimension in the range $3 \leq m \leq 7$, provided that $K \gg m!$, and typically uses $\tau = 1$ unless the autocorrelation suggests a larger delay.

Theoretical properties of permutation entropy include: (1) $H(m) \approx 0$ for strongly regular or periodic signals; (2) $H(m) \approx \log_2(m!)$ when the process is effectively random (e.g., white noise); (3) invariance under any monotonically increasing transformation of the data; and (4) robustness to observational noise.

The Multivariate Permutation Entropy (MvPE) extends the PE framework to multivariate time series, accounting for the joint ordinal structures across multiple dimensions [27]. A well-established guideline in the application of MvPE is to set the length of the ordinal patterns m equal to the dimension D of the dataset, thereby ensuring that the complexity of the multivariate interdependencies is adequately captured. Following this approach, we set $m = D$ and maintain the embedding delay at $\tau = 1$ throughout our analysis.

The consistent application of these parameter choices across all models enables a fair and meaningful comparison of PE and MvPE with other complexity measures. The resulting analyses highlight the sensitivity of these quantifiers to different dynamical regimes and their effectiveness in detecting variations in system complexity.

2.3 Statistical complexity (SC)

The Statistical Complexity (SC) is a quantifier designed to complement the Permutation Entropy (PE) by providing a more nuanced characterization of the underlying dynamics of time series [16]. While PE primarily measures the degree of disorder within a system, SC integrates both disorder and structure, enabling the differentiation between random and chaotic behaviors that might otherwise exhibit similar entropy values.

The expected behavior of SC is well-established: it tends to decrease for purely stochastic signals, such as white noise, due to their maximal disorder but minimal structural complexity. Conversely, SC increases for deterministic chaotic systems, where the presence of structured yet unpredictable dynamics leads to higher complexity values. This sensitivity makes SC particularly useful in distinguishing between different types of complex systems

that may be indistinguishable through entropy measures alone.

The computation of SC fundamentally depends on the evaluation of the PE, as originally formulated by Rosso et al. [16]. Specifically, SC combines the normalized PE with a disequilibrium term that quantifies the divergence between the actual probability distribution of ordinal patterns and the uniform distribution. Consequently, the parameter choices for SC are inherently linked to those selected for PE.

Statistical Complexity, denoted by C , is defined as the product of a normalized entropy measure H and a disequilibrium measure Q :

$$C = Q \cdot H,$$

where H is typically the normalized Permutation Entropy.

To compute C , we first obtain the normalized Permutation Entropy H_{norm} . Given the ordinal pattern probability distribution

$$P = \{p(\pi_k)\}_{k=1}^{m!},$$

the unnormalized permutation entropy is

$$H(m) = - \sum_{k=1}^{m!} p(\pi_k) \log_2 p(\pi_k).$$

We then normalize by $\log_2(m!)$:

$$H_{\text{norm}} = \frac{H(m)}{\log_2(m!)} = \frac{- \sum_{k=1}^{m!} p(\pi_k) \log_2 p(\pi_k)}{\log_2(m!)}.$$

Next, we define the reference (equilibrium) distribution P_e to be the uniform distribution over all $m!$ ordinal patterns:

$$P_e = \{p_e(\pi_k) = 1/m!\}_{k=1}^{m!}.$$

The disequilibrium Q is then computed via the Jensen–Shannon divergence $J(P, P_e)$. Explicitly,

$$J(P, P_e) = S\left(\frac{P + P_e}{2}\right) - \frac{S(P)}{2} - \frac{S(P_e)}{2},$$

where $S(\cdot)$ denotes the Shannon entropy (base 2) of a distribution. Concretely,

$$S(P) = - \sum_{k=1}^{m!} p(\pi_k) \log_2 p(\pi_k), \quad S(P_e) = - \sum_{k=1}^{m!} \frac{1}{m!} \log_2 \frac{1}{m!} = \log_2(m!).$$

We normalize $J(P, P_e)$ by a constant Q_0 , chosen so that the maximum possible Jensen–Shannon

divergence (between P and P_e) equals 1. Specifically,

$$Q_0 = -\frac{1}{2} \left[\frac{m!+1}{m!} \log_2(m!+1) - 2 \log_2(2m!) + \log_2(m!) \right]^{-1}.$$

Thus,

$$Q = Q_0 J(P, P_e).$$

Finally, the Statistical Complexity is given by

$$C = H_{\text{norm}} \cdot Q,$$

which jointly measures both the normalized uncertainty (through H_{norm}) and the departure from equilibrium (through Q).

Key properties of C include:

- *Boundedness*: $0 \leq C \leq C_{\text{max}}$, where C_{max} depends on m .
- *Invariance*: C is invariant under affine transformations of the original time series.
- *Sensitivity*: C detects both purely stochastic and deterministic structures.
- *Extremes*:
 - $C \approx 0$ for perfectly ordered (e.g., periodic) or completely random (e.g., white noise) systems.
 - $C > 0$ for complex systems exhibiting correlated patterns.

Interpreting C alongside H_{norm} in the $C \times H$ plane yields a complexity–entropy diagram, where the lower boundary corresponds to $C = 0$ (either perfect order or maximal randomness), and the upper boundary $C = C_{\text{max}}(H)$ represents maximal structural complexity for a given entropy. Intermediate values of C indicate processes that combine both disorder and nontrivial correlations.

In this study, we adopt the same embedding dimension $m = 6$ and embedding delay $\tau = 1$ as employed in the PE analysis, ensuring consistency across quantifiers and facilitating direct comparisons. By leveraging this complementary relationship between SC and PE, the analysis provides a comprehensive assessment of system complexity, capturing both the degree of randomness and the presence of underlying structure within the datasets.

2.4 Structurality (Δ)

The structurality measure, denoted by Δ , quantifies the spatial organization of dynamical trajectories in phase space through the analysis of Poincaré sections. A lack of

discernible structure in the dynamics manifests as a large Δ , while well-ordered dynamics yield $\Delta \approx 0$.

To compute structurality, a Poincaré section of the system is discretized into an $N_b \times N_b$ grid of boxes. Defining the indicator variable v_{ij} as 1 if the box (i, j) contains at least one crossing point and 0 otherwise, the structurality is given by:

$$\Delta = \sum_{i,j=1}^{N_b} \frac{v_{ij}}{N_b^2}.$$

Firstly, we need to construct the frame in which the Poincaré section is analysed. A way to do this is to consider the maximum M_i and minimum m_i possible values along the axis i of the chosen frame and then choose the number of boxes N_b of length $l_b = \|M_i - m_i\|/N_b$ as a function of the number N_p of points in the Poincaré section:

$$N_b \geq 10 \log_{10} N_p$$

This measure is primarily grounded in deterministic concepts, such as trajectories and phase space representations, and inherently depends on several methodological choices, including the definition of the Poincaré section, the selection of an appropriate grid resolution, and the reference frame. Moreover, it was originally validated using long time series, as the number of crossing points in the Poincaré section typically satisfies $N_p > 10^3$ [28].

2.5 Dynamical complexity (DC)

The dynamical complexity (DC) is a composite metric developed to quantify simultaneously the unpredictability and the spatial disorder of a dynamical system. Formally, it is defined as the sum of the permutation entropy and the structurality Δ .

In this formulation, the permutation entropy provides a measure of temporal unpredictability by evaluating the diversity of ordinal patterns in the time series, while the structurality captures the spatial dispersion of trajectories in phase space through the analysis of Poincaré sections.

It is expected that a value of $DC = 0$ characterizes a fully predictable and perfectly ordered system, as typically observed in purely periodic dynamics. Conversely, $DC = 1$ represents processes that are nonpredictable but have organized dynamics, while $DC = 2$ indicates maximal dynamical complexity, exemplified by uniform white noise that is both temporally unpredictable and spatially dispersed.

By combining these two complementary aspects—temporal unpredictability and spatial structure—DC offers a comprehensive characterization of dynamical regimes. For example, chaotic maps may display high permutation entropy yet maintain a relatively confined

structure in phase space, resulting in low structurality. In such cases, relying solely on a single measure may lead to misclassification, whereas DC can correctly capture the coexistence of temporal disorder and spatial organization.

Benchmark models and their challenges

In this chapter, we introduce a set of benchmark models—ranging from paradigmatic deterministic systems to stochastic processes—against which different complexity-quantification methods can be evaluated. For each system, we specify two values of the control parameter, p_1 and p_2 , chosen so that p_2 is known to produce a higher level of dynamical disorder than p_1 .

3.1 Discrete dynamical systems

3.1.1 Logistic map

The Logistic map [36, 37] is one of the simplest and most extensively studied models in chaos theory. Despite its algebraic simplicity, it exhibits a rich spectrum of dynamical behaviors, transitioning from fixed points to periodic oscillations, and ultimately to fully developed chaos as the parameter increases. This makes it an ideal, pedagogical and practical model for exploring bifurcations and chaos. For our purposes, we choose $p_1 = 3.9$ and $p_2 = 4.0$. At $p_1 = 3.9$, the system is already in a chaotic regime. By contrast, $p_2 = 4.0$ places the system at the edge of maximal chaos, where sensitivity to initial conditions is extreme and the attractor becomes fully developed. This range allows us to test the capacity of our disorder quantifier to distinguish subtle changes within chaotic regimes.

$$x_{n+1} = p_{1,2}x_n(1 - x_n). \quad (3.1)$$

3.1.2 Henon map

The Henon map [37, 38] is a two-dimensional discrete-time dynamical system, well known for its chaotic attractor and complex geometric structure. It is frequently used to

study attractors, fractal dimensions, and Lyapunov exponents. The coupling between x and y introduces nonlinearity and memory effects, making the system particularly challenging for time-series analysis and phase space reconstruction. We select $p_1 = 1.1$ and $p_2 = 1.4$ as control parameter values. At $p_1 = 1.1$, the system dynamics are moderately chaotic, with some underlying regularities. At $p_2 = 1.4$, the map exhibits fully developed chaos with a prominent chaotic attractor. These distinct regimes provide a suitable benchmark for testing the robustness of the proposed disorder quantification methods.

$$\begin{cases} x_{n+1} = 1 - p_{1,2}x_n^2 + y_n, \\ y_{n+1} = 0.3x_n. \end{cases} \quad (3.2)$$

3.1.3 Ikeda map

The Ikeda map [37, 39] models the dynamics of a laser beam in an optical cavity, with its behavior governed by a nonlinear dependence on the instantaneous state through the phase term t_n . This system is widely studied in optics and nonlinear dynamics due to its ability to produce complex attractors and chaotic regimes as parameters vary. The nonlinearity induced by the phase feedback mechanism makes the system highly sensitive to parameter changes and initial conditions, posing significant challenges for accurate prediction and analysis. We use $p_1 = 7.1$ and $p_2 = 7.3$ as the values of the control parameter. At $p_1 = 7.1$, the system is on the threshold of chaotic behavior, while $p_2 = 7.3$ leads to more pronounced chaotic dynamics and a filled phase space. These parameter choices allow us to test the sensitivity of our disorder quantifier to phase-induced nonlinear complexity.

$$\begin{cases} x_{n+1} = 0.84 + 0.9(x_n \cos t_n - y_n \sin t_n), \\ y_{n+1} = 0.9(x_n \sin t_n + y_n \cos t_n), \\ t_n = 0.4 - \frac{p_{1,2}}{1+x_n^2+y_n^2}. \end{cases} \quad (3.3)$$

3.1.4 Towel map

The Towel map [40, 40] is a higher-dimensional discrete dynamical system derived from generalizations of the Logistic map, incorporating coupling between multiple variables and nonlinear feedback mechanisms. Its complexity stems from the interaction among the three variables, creating a high-dimensional chaotic attractor that is difficult to analyze using standard time-series methods. This model is relevant for studying multi-variable chaos, hyperchaos, and intricate bifurcation structures. We set $p_1 = 3.3$ and $p_2 = 3.8$. At $p_1 = 3.3$, the system displays moderately complex dynamics, while at $p_2 = 3.8$, the system exhibits hyperchaotic behavior (chaos across multiple dimensions). These parameter values are chosen to highlight the effectiveness of the proposed disorder quantifier in distinguishing

different degrees of multidimensional chaos.

$$\begin{cases} x_{n+1} = p_{1,2}x_n(1 - x_n) - 0.05(y_n + 0.35)(1 - 2z_n), \\ y_{n+1} = 0.1[(y_n + 0.35)(1 - 2z_n) - 1](1 - 1.9x_n), \\ z_{n+1} = 3.78z_n(1 - z_n) + 0.2y_n. \end{cases} \quad (3.4)$$

3.1.5 Standard map

The Standard map [41, 42] is a fundamental model for studying Hamiltonian chaos and transport in conservative systems. It represents a kicked rotator and captures essential features of phase space mixing and the transition from regular to chaotic motion as the nonlinearity parameter increases. One of the key challenges in analyzing this system is the coexistence of regular and chaotic regions in phase space, known as a mixed phase space, which complicates the detection and quantification of disorder. We select $p_1 = 1.0$ and $p_2 = 2.0$ for the perturbation strength. At $p_1 = 1.0$, the system shows a mixture of regular islands and chaotic seas, while at $p_2 = 2.0$, the chaotic component dominates the phase space. This controlled transition makes the Standard map particularly useful for evaluating the capacity of our disorder quantifier to detect changes in systems with conservative dynamics.

$$\begin{cases} \theta_{n+1} = (\theta_n + \rho_{n+1}) \mod 2\pi, \\ \rho_{n+1} = \rho_n + p_{1,2} \sin \theta_n \mod 2\pi. \end{cases} \quad (3.5)$$

3.2 Continuous dynamical systems

3.2.1 Double-gyre system

The Double-Gyre system [28, 43] is a standard model in fluid dynamics and oceanography, widely used to investigate transport phenomena, mixing processes, and the formation of Lagrangian coherent structures. Its dynamics result from nonlinear interactions between spatial variables (x, y) and auxiliary variables (u, v) , leading to intricate flow patterns that are highly sensitive to parameter variations. These features make it an excellent benchmark for testing recurrence-based analyses and entropy quantifiers. We set $p_1 = 0.25$ and $p_2 = 0.10$, following previous studies [28]. The lower value $p_2 = 0.10$ generates more regular and predictable trajectories, whereas the higher value $p_1 = 0.25$ enhances nonlinearity and increases flow complexity, producing chaotic behavior filling up its phase space. These distinct dynamical regimes provide a valuable contrast for assessing the sensitivity of the proposed disorder quantifier.

$$\begin{cases} \dot{x} = p_{1,2}\pi \sin [\pi(ux^2 + x - u)] \sin \pi y, \\ \dot{y} = p_{1,2}\pi \cos [\pi(ux^2 + x - u)] \cos \pi y, \\ \dot{u} = v, \\ \dot{v} = -\left(\frac{\pi}{5}\right)^2 u. \end{cases} \quad (3.6)$$

3.2.2 Coupled Rössler system

The coupled Rössler system [28, 44] is a prototypical model for exploring synchronization phenomena and the emergence of complex collective dynamics in coupled nonlinear oscillators. It captures how interactions between subsystems influence the global behavior, including the transition between synchronized and desynchronized states. This system's sensitivity to parameter changes, particularly in the coupling strength, poses significant challenges for detecting subtle transitions in disorder. We consider $p_1 = 0.25$ and $p_2 = 0.40$ for the coupling parameter. At p_1 , the system tends toward partial synchronization or weak interaction, leading to relatively regular dynamics. In contrast, the stronger coupling at p_2 induces more complex interdependence between the oscillators. These contrasting regimes are ideal for evaluating the responsiveness of the proposed disorder quantifier to coupling-induced complexity.

$$\begin{cases} \dot{x}_1 = -y_1 - z_1, \\ \dot{y}_1 = x_1 + 0.492y_1, \\ \dot{z}_1 = 2 + z_1(x_1 - 4) + p_{1,2}(z_2 - z_1), \\ \dot{x}_2 = -y_2 - z_2, \\ \dot{y}_2 = x_2 + 0.480y_2, \\ \dot{z}_2 = 2 + z_2(x_2 - 4) + p_{1,2}(z_1 - z_2). \end{cases} \quad (3.7)$$

3.2.3 Mackey-Glass system

The Mackey-Glass system [28, 45] is a canonical time-delay differential equation, originally developed to model physiological processes such as blood cell regulation. Its dynamics are fundamentally governed by the time-delay parameter, which induces a transition from periodic to chaotic behavior as it increases. The infinite-dimensional nature of time-delay systems introduces computational challenges, particularly in state-space reconstruction and in handling delayed terms accurately. For our analysis, we set $p_1 = 1.8$ and $p_2 = 6.0$ as the values for the delay parameter. When $\tau = p_1 = 1.8$, the system typically exhibits periodic or mildly chaotic behavior, while $\tau = p_2 = 6.0$ leads to highly irregular and chaotic dynamics. This controlled variation in complexity makes the Mackey-Glass model essential for testing the effectiveness of our framework in detecting

disorder in delayed systems.

$$\dot{x} = 2 \cdot \frac{x_{\tau=p_{1,2}}}{1 + x_{\tau=p_{1,2}}^{10}} - x. \quad (3.8)$$

3.2.4 Lorenz system

The Lorenz system [41, 46] is one of the most iconic models of deterministic chaos, originally proposed to study atmospheric convection. It has become a foundational paradigm for exploring phenomena such as bifurcations, chaotic attractors, and the limits of predictability in chaotic systems. Due to its sensitive dependence on initial conditions and parameters, small variations can lead to dramatically different trajectories, presenting challenges in long-term prediction and also in disorder quantification. We use $p_1 = 24$ and $p_2 = 32$ as the values for a parameter that directly affects the system's energy input. At $p_1 = 24$, the system typically resides in a more regular periodic regime, whereas at $p_2 = 32$, it exhibits robust chaotic behavior characterized by the well-known Lorenz attractor. These well-documented transitions make the Lorenz system an ideal benchmark for validating the sensitivity and reliability of the disorder quantifier.

$$\begin{cases} \dot{x} = 10(y - x), \\ \dot{y} = x(p_{1,2} - z) - y, \\ \dot{z} = xy - \frac{8}{3}z. \end{cases} \quad (3.9)$$

3.2.5 Chua's circuit

Chua's Circuit [37, 47] is a paradigmatic example of a simple electronic circuit capable of exhibiting chaotic oscillations. It is widely used both in theoretical studies and experimental validations of chaos theory, as well as in practical applications such as secure communications and signal processing. The presence of the piecewise-linear nonlinearity $h(x)$ introduces sharp transitions and complex bifurcation structures, posing analytical and numerical challenges. For this system, we set $p_1 = 33$ and $p_2 = 25.58$ as the values of the parameter influencing the circuit's dynamics. The value $p_1 = 33$ tends to promote more regular oscillatory behavior, while $p_2 = 25.58$ induces chaotic oscillations with increased complexity. These contrasting regimes allow us to assess the capability of our disorder quantifier to detect changes in the system's dynamic behavior, especially in the presence of discontinuous nonlinearities.

$$\begin{cases} \dot{x} = 15.6(y - x - h(x)), \\ \dot{y} = x - y + z, \\ \dot{z} = -p_{1,2}y, \\ h(x) = \frac{-5}{7}x + \frac{1}{2} \left(\frac{-8}{7} - \frac{-5}{7} \right) (|x + 1| - |x - 1|). \end{cases} \quad (3.10)$$

3.3 Stochastic processes

3.3.1 White Gaussian noise (WGN)

White Gaussian noise [48, 49] represents the simplest form of stochastic signal, characterized by complete absence of temporal correlation and a constant power spectral density across all frequencies. It is extensively used as a benchmark for randomness in signal processing, statistical physics, and control theory. The main challenge for quantifiers in dealing with WGN is recognizing its maximal stochasticity while distinguishing it from complex deterministic signals with irregular but structured behavior. Using a standard normal distribution ensures comparability with other stochastic processes and facilitates statistical interpretation.

$$x_t \sim \mathcal{N}(0, 1), \quad \text{independent and identically distributed (i.i.d.).} \quad (3.11)$$

3.3.2 Autoregressive model (AR)

The autoregressive (AR) [49, 50] model introduces memory into a stochastic process through the parameter ϕ , which governs the strength of dependence on previous values. As ϕ approaches 1, the process exhibits stronger persistence and slower decay of autocorrelations. Conversely, when $\phi = 0$, the process reduces to white Gaussian noise. This model is widely used in econometrics, signal processing, and climatology for modeling time-dependent stochastic phenomena. The challenge for disorder quantification lies in detecting subtle changes in correlation induced by varying ϕ . For benchmarking purposes, ϕ can be selected at intermediate values (e.g., $\phi = 0.5$) to represent moderate correlation, contrasting with $\phi = 0$ for the uncorrelated case.

$$\eta_t(\phi) = \phi\eta_{t-1} + w_t, \quad w_t \sim \mathcal{N}(0, 1), \quad 0 \leq \phi < 1. \quad (3.12)$$

3.3.3 Noisy logistic model

A version of the logistic map but corrupted by some level of correlated noise.

$$y_t(\phi) = x_t + \eta_t(\phi), \quad (3.13)$$

where x_t is the chaotic logistic map, and $\eta_t(\phi)$ is the autoregressive noise defined previously, with a specified variance ratio:

$$\frac{\sigma_\eta^2}{\sigma_x^2}, \quad (3.14)$$

in which σ_x^2 and σ_η^2 are the variance of the logistic map x_t and AR-noise $\eta_t(\phi)$. The studied parameters of this model ranges are $\phi \in \{0.00, 0.25, 0.50, 0.75, 0.90\}$ and $\frac{\sigma_\eta^2}{\sigma_x^2} \in [0, 1]$.

In particular, this model is especially important for the present work because it enables the study of signals composed of both deterministic chaos and a stochastic component, serving as a paradigmatic example of data typically encountered in complex phenomena. Furthermore, it allows for the investigation not only of the underlying deterministic dynamics but also of the characteristics of the noise processes corrupting the system.

3.3.4 Gaussian-noisy logistic model (GNL)

For comparative tests, we consider a particular configuration of the noisy logistic model (Subsection 3.3.3) with a fixed variance ratio $\sigma_\eta^2/\sigma_x^2 = 0.3$, and set $\phi = p_1 = 0.50$ and $\phi = p_2 = 0.25$. In this configuration, we refer to the process as the Gaussian-noisy logistic (GNL) model.

3.3.5 Uniform-noisy logistic model (UNL)

Time series s_t generated from the logistic map

$$y_{t+1} = 3.99 y_t(1 - y_t),$$

with uniform white noise u_t added such that

$$s_t = (1 - p_{1,2}) y_t + p_{1,2} u_t,$$

where $p_1 = 0.45$ and $p_2 = 0.55$. This configuration is referred to as the uniform-noisy logistic (UNL) model [28].

3.3.6 Power law noise

Power law noise [48, 51], also known as $1/f^\alpha$ noise, is ubiquitous in both natural and artificial systems, ranging from electronic circuits to biological processes and climate dynamics. It is characterized by long-range correlations in the frequency domain, where the spectral exponent α determines the degree of correlation: larger values of α imply stronger low-frequency components and more persistent behavior, whereas smaller α tends towards white noise.

In our analysis, we selected $\alpha = p_2 = 0$, representing ideal white noise with no temporal correlation, commonly referred to as White Gaussian Noise (WGN). This choice provides a baseline of maximal stochastic disorder with minimal memory, serving as a fundamental reference point for the performance of disorder quantifiers.

Additionally, we considered $\alpha = 1$, referred to as Pink Gaussian Noise (PGN) or $1/f$ noise, characterized by a power spectrum inversely proportional to frequency. This leads to long-range temporal correlations, which are prevalent in many natural systems including electronic devices, biological rhythms, and financial markets. The structured decay of spectral power makes PGN an important benchmark for evaluating the sensitivity of disorder measures to the presence of memory and correlation in stochastic processes.

To further extend the range of correlation structures, we included $\alpha = p_1 = 2$, corresponding to a Brownian or random walk-like process, also known as Red Gaussian Noise (RGN). This type of noise exhibits a power spectrum that decays quadratically with frequency, emphasizing strong low-frequency components and resulting in a highly correlated stochastic signal. RGN is frequently used to model cumulative stochastic processes such as particle diffusion, climate variability, and financial time series. Compared to pink noise, RGN exhibits even stronger temporal dependencies, posing a significant challenge for disorder quantifiers to accurately assess the degree of disorder and to differentiate it from deterministic systems with memory.

These stochastic signals were generated using the `SignalAnalysis.jl` package to ensure reproducibility and adherence to standard spectral properties. A fundamental challenge in this context is the differentiation between structured stochasticity and deterministic chaos, as both can exhibit broadband frequency spectra. The broad variation in spectral properties across WGN, PGN, and RGN allows us to rigorously evaluate the capacity of disorder measures to detect changes in the correlation structure of stochastic signals.

3.3.7 Nonlinear colored noise

The nonlinear colored noise model [52] introduces memory and nonlinearity into an otherwise random sequence, thereby generating complex stochastic behavior. It is a stochastic process specifically designed to challenge the distinction of deterministic and stochastic signals via the “noise titration” technique. This process is relevant for modeling phenomena where stochastic fluctuations are modulated by nonlinear feedback. The parameter p controls the strength of the nonlinear interaction between successive noise terms. We select $p_1 = 4$, yielding a strong nonlinear influence and more structured noise, and $p_2 = 0$, corresponding to a purely linear combination of independent random variables, effectively reducing the process to simple scaled noise. The transition from structured to unstructured noise in this model provides a useful test for evaluating disorder quantifiers’

ability to capture nonlinear stochastic dynamics.

$$\begin{cases} x_n = 3y_n + p_{1,2}y_{n-1}(1 - y_n), \\ y_n \sim \mathcal{U}[0, 1] \text{ (uniform white noise)}. \end{cases} \quad (3.15)$$

3.3.8 Bounded random walk

The bounded random walk [53] is a variation of the classical random walk with reflecting boundaries introduced via an exponential confinement term. This model is significant in applications where stochastic processes are constrained within physical or theoretical limits, such as population dynamics or particle motion in bounded media. The parameter p scales the amplitude of the stochastic forcing: low p values lead to nearly deterministic confinement, whereas higher p introduce greater randomness. We select $p_1 = 0.05$, representing weak noise where the confinement dominates, and $p_2 = 0.4$, allowing for stronger stochastic excursions within the bounds. This controlled variation provides a clear scenario to assess how disorder quantifiers respond to noise-induced transitions within bounded stochastic systems.

$$\begin{cases} x_{n+1} = x_n + e^{-15} \left(e^{-3(x_n-100)} - e^{3(x_n-100)} \right) + p_{1,2}\delta_n, \\ \delta_n \sim \mathcal{N}(0, 1) \text{ (white Gaussian noise)}. \end{cases} \quad (3.16)$$

3.3.9 Fractional Gaussian noise

Fractional Gaussian noise (fGn) [54, 55] generalizes white Gaussian noise by introducing long-range dependencies parameterized by the Hurst exponent H . This process is widely used in fields such as hydrology, finance, and physiology to model persistent or anti-persistent behaviors. For $H < 0.5$, the process exhibits anti-persistence, while for $H > 0.5$, it becomes persistent. We select $p_1 = 0.1$, inducing strong anti-persistence with rapid oscillations and low predictability, and $p_2 = 0.5$, corresponding to standard uncorrelated Gaussian white noise. These choices test the sensitivity of disorder measures to changes in memory structure and long-range dependence within Gaussian stochastic processes.

$$\{X_t\}_{t \in \mathbb{Z}} \text{ with Hurst exponent } H = p_{1,2} \in (0, 1). \quad (3.17)$$

3.3.10 Langevin equation

The Langevin equation [56, 57] is a fundamental model in stochastic dynamics, describing systems influenced by both deterministic damping and random forcing. It is widely employed in statistical physics, chemical kinetics, and financial mathematics to model relaxation processes under the influence of noise. The parameter p controls the strength of deterministic relaxation: larger values of p correspond to faster returns to equilibrium,

whereas smaller values allow noise to dominate over longer timescales. For this study, we select $p_1 = 0.1$, which results in longer time correlations in the trajectory, and $p_2 = 1.0$, which induces rapid damping and shorter time correlations. This contrast provides an ideal framework to evaluate the sensitivity of disorder measures to variations in the balance between deterministic and stochastic forces.

$$\begin{cases} \frac{dv}{dt} = -p_{1,2}v + w_t, \\ w_t \sim \mathcal{N}(0, 1) \text{ (white Gaussian noise).} \end{cases} \quad (3.18)$$

A formalism for disorder quantification in data

To construct a meaningful quantifier of disorder in data, we begin by introducing an idealized disorder condition. This condition is formally defined and its consequences are explored in a systematic framework. We start by characterizing the disorder condition in terms of the data structure, then develop a method to encode sequences accordingly. Next, we examine the measurable implications that follow from the hypothesis and, finally, propose a procedure to assess how closely empirical data conforms to the ideal disorder scenario.

4.1 Describing the disorder condition in data

The first definition breaks the full dataset into smaller overlapping segments, allowing us to study local patterns within the data. The second definition introduces the idea of “similar events”, where two segments are considered similar if their elements are close enough, based on a chosen level of tolerance.

Definition 4.1.1 (The data and its subsequences). Let $\mathfrak{D} = (\mathbf{x}_i)_{i=1}^K$ be a sequential representation of data elements. Inside \mathfrak{D} , there are smaller sequences $\mathbf{X}_{(n)}$ of N elements for $0 \leq n \leq K - N$ such that $\mathbf{X}_{(n)} = (\mathbf{x}_i)_{i=n+1}^{n+N}$.

Definition 4.1.2 (Similar events). Let $\tilde{\mathbf{x}}_i$ be an element similar to \mathbf{x}_i as

$$\tilde{\mathbf{x}}_i \in \{\mathbf{x} \in \mathfrak{D} : \|\mathbf{x}_i - \mathbf{x}\| < \delta\}, \quad (4.1)$$

in which δ is an arbitrary nonzero element. Also, a subsequence similar to $\mathbf{X}_{(n)}$ is $\tilde{\mathbf{X}}_{(n)} = (\tilde{\mathbf{x}}_i)_{i=n+1}^{n+N}$.

A completely disordered dataset can be regarded as dense, with each event $\mathbf{X}_{(n)}$ statistically independent of all others [28]. Based on the previous definitions, a probabilistic

treatment becomes appropriate—enabled by considering intervals of values, since the probability of exact single values would be zero. Under this background, we now formally state the main hypothesis as:

Hypothesis 4.1.1 (Statistical independence). The disorder condition implies independence of all elements of the data, such that the probability of occurrence $p(\tilde{\mathbf{X}}_{(n)})$ is

$$p(\tilde{\mathbf{X}}_{(n)}) \equiv \prod_{i=n+1}^{n+N} p(\tilde{\mathbf{x}}_i). \quad (4.2)$$

A more precise formulation of this hypothesis involves expressing the relationship between a given subsequence $\tilde{\mathbf{X}}_{(n)}$ and other similar subsequences. To this end, we reformulate Hypothesis 4.1.1 to explicitly capture this relationship. By employing permutation operations, we express the invariance of the probability distribution of a sequence under its permutations as follows:

Definition 4.1.3 (Disorder condition). Let S_N be the set of all possible permutations of N elements, *i.e.* the so-called symmetric group from Group Theory [58]. Denote a permutation operation σ such that $\sigma \in S_N$. Then, a direct consequence of Eq. 4.2 is:

$$p(\tilde{\mathbf{X}}_{(n)}) = p(\sigma(\tilde{\mathbf{X}}_{(n)})), \quad \forall \sigma \in S_N. \quad (4.3)$$

This implies that, in a disordered dataset, for any given subsequence $\tilde{\mathbf{X}}_{(n)}$, there exists another subsequence $\tilde{\mathbf{X}}_{(k)}$ composed of the same similar events as $\tilde{\mathbf{X}}_{(n)}$, but permuted in temporal order, with the same probability of occurrence. It is important to note that this formal description of disorder reflects a fundamental physical intuition: disordering a system (such as shuffling a deck) is effectively achieved by permuting its constituent elements.

4.2 Codifying sequences of the data

The characterization of disorder in data raises the fundamental challenge of appropriately measuring similar events, particularly in the absence of a theoretically justified partitioning of the data's elements or sequences. Traditional approaches based on partitioning are typically constrained to low-dimensional datasets or a limited class of dynamical systems [22, 23, 59].

To address this limitation, we adopt a strategy based on the concept of recurrence between events, thereby encoding the relationships among data sequences into a finite and computable set of representations. With this objective, we introduce the definition of a recurrence matrix and examine some of its key properties.

Definition 4.2.1 (Recurrence matrix). Denote the recurrence function as $r_{(i,j)} = \Theta(\varepsilon - \|\mathbf{x}_i - \mathbf{x}_j\|)$, with Θ being the Heaviside function and ε the recurrence threshold. Then, a recurrence matrix $\mathbf{M} \equiv \mathbf{R}(\mathbf{X}_{(n)}, \mathbf{X}_{(m)})$ is:

$$\mathbf{R}(\mathbf{X}_{(n)}, \mathbf{X}_{(m)}) \stackrel{def}{=} \begin{bmatrix} r_{(n+1,m+1)} & \cdots & r_{(n+1,m+N)} \\ \vdots & \ddots & \vdots \\ r_{(n+N,m+1)} & \cdots & r_{(n+N,m+N)} \end{bmatrix}. \quad (4.4)$$

The square matrix constructed for the case $N \ll K$ is referred to as a recurrence microstate of size N [11, 60], and represents parts of the well-known recurrence plot, defined as $\mathbf{R}(\mathfrak{D}, \mathfrak{D})$ [61]. This formulation enables the extraction of self-relational patterns within the data by means of a threshold parameter ε , through direct element-wise comparisons. Notably, it bypasses the need for arbitrary phase space partitioning, thereby avoiding segmentation-dependent biases and preserving the intrinsic structure of the original dataset.

To adopt a formal probabilistic perspective, we establish a connection between the probability of observing a specific recurrence matrix and the probability of encountering similar events. In particular, we invoke Definition 4.1.2 to incorporate similarity-based inputs into the matrix construction. This substitution replaces pointwise comparisons with intervals of admissible values, thereby yielding nonzero probabilities and allowing a probabilistic interpretation of recurrence structures.

In this sense, consider that as the data size K increases, the likelihood of finding a point \mathbf{x}_j in close proximity to another point \mathbf{x}_i also increases. In this regime, the condition $\delta \ll \varepsilon$ becomes increasingly valid. In the asymptotic limit $K \rightarrow \infty$, even vanishingly small distances $\delta \rightarrow 0$ become sufficient for \mathbf{x}_i and \mathbf{x}_j to be considered similar under a given threshold ε . We state these assumptions as:

Hypothesis 4.2.1 (The data size). The dataset is sufficiently large to justify treating δ -close elements as effectively generating the same recurrence matrix, provided that the condition $\delta \ll \varepsilon$ holds:

$$\mathbf{R}(\tilde{\mathbf{X}}_{(n)}, \tilde{\mathbf{X}}_{(m)}) \equiv \mathbf{R}(\mathbf{X}_{(n)}, \mathbf{X}_{(m)}). \quad (4.5)$$

This approximation is valid within the regime where $N \ll K$, under which the resulting recurrence matrices are referred to as *recurrence microstates*.

We are now in a position to examine key properties of this type of encoding. In particular, it allows us to define the probability associated with a recurrence microstate as $p(\mathbf{R}(\tilde{\mathbf{X}}_{(n)}, \tilde{\mathbf{X}}_{(m)}))$. The procedure for computing recurrence microstates consists of counting the frequency of the matrix generated using $\tilde{\mathbf{X}}_{(n)}$ and $\tilde{\mathbf{X}}_{(m)}$ as inputs. Importantly, it also accounts for the reverse configuration, *i.e.*, the matrix obtained by swapping the inputs. This symmetry implies that the probability of a given microstate satisfies the following condition:

Lemma 4.2.1 (Transposition in recurrence microstates). The probability of a recurrence microstate remains unchanged when the sequences are interchanged.

$$p\left(\mathbf{R}\left(\tilde{\mathbf{X}}_{(n)}, \tilde{\mathbf{X}}_{(m)}\right)\right) = p\left(\mathbf{R}\left(\tilde{\mathbf{X}}_{(m)}, \tilde{\mathbf{X}}_{(n)}\right)\right). \quad (4.6)$$

4.3 Verifying consequences of the disorder condition

At this stage, we have established the necessary tools to investigate the implications of the ideal disorder condition, as stated in Definition 4.1.3, within a computable set of symbolic representations derived from the data—namely, the recurrence microstates. Since the disorder condition is formulated in terms of permutations of sequences, it becomes essential to define corresponding operations on recurrence matrices that allow us to formally relate transformations in the sequences to transformations in the microstates they generate.

Definition 4.3.1 (Transposition operator). Let \mathcal{T} denote the transposition operator acting on a matrix. Given a recurrence microstate $\mathbf{R}(\tilde{\mathbf{X}}_{(n)}, \tilde{\mathbf{X}}_{(m)})$, the transposition operator satisfies:

$$\mathcal{T}\mathbf{R}(\tilde{\mathbf{X}}_{(n)}, \tilde{\mathbf{X}}_{(m)}) = \mathbf{R}(\tilde{\mathbf{X}}_{(m)}, \tilde{\mathbf{X}}_{(n)}). \quad (4.7)$$

This behavior of the transposition operator follows directly from Definition 4.4, which specifies the construction of recurrence microstates from pairs of sequences:

$$\mathcal{T} \begin{bmatrix} r_{(n+1,m+1)} & \cdots & r_{(n+1,m+N)} \\ \vdots & \ddots & \vdots \\ r_{(n+N,m+1)} & \cdots & r_{(n+N,m+N)} \end{bmatrix} = \begin{bmatrix} r_{(n+1,m+1)} & \cdots & r_{(n+N,m+1)} \\ \vdots & \ddots & \vdots \\ r_{(n+1,m+N)} & \cdots & r_{(n+N,m+N)} \end{bmatrix} = \mathbf{R}(\tilde{\mathbf{X}}_{(m)}, \tilde{\mathbf{X}}_{(n)}). \quad (4.8)$$

Definition 4.3.2 (Operator of row swaps). Let \mathcal{L}_σ be the operator that swaps rows as described by a given permutation σ . By definition, it satisfies:

$$\mathcal{L}_\sigma \mathbf{R}(\tilde{\mathbf{X}}_{(n)}, \tilde{\mathbf{X}}_{(m)}) = \mathbf{R}(\sigma(\tilde{\mathbf{X}}_{(n)}), \tilde{\mathbf{X}}_{(m)}). \quad (4.9)$$

For example, given the permutation $\sigma = 132$, the operator \mathcal{L}_{132} acts on a matrix \mathbf{M} of order $N = 3$ by exchanging the third and second rows while leaving the first row unchanged.

By doing so, we state the main algebraic result as follows, which establishes equiprobability between recurrence microstates as a consequence of the disorder condition.

Theorem 4.3.1 (Classes of equiprobable microstates). Let $\mathcal{M}_a(\mathbf{M})$ be the class of the recurrence microstate \mathbf{M} generated by a data satisfying the disorder condition (Definition

4.1.3), defined as:

$$\mathcal{M}_a(\mathbf{M}) = \bigcup_{\sigma_i, \sigma_j \in S_N} \{\mathcal{L}_{\sigma_j} \mathcal{T} \mathcal{L}_{\sigma_i} \mathbf{M}, \mathcal{T} \mathcal{L}_{\sigma_j} \mathcal{T} \mathcal{L}_{\sigma_i} \mathbf{M}\}. \quad (4.10)$$

Then:

$$p(\mathbf{M}) = p(\mathbf{M}'), \forall \mathbf{M}' \in \mathcal{M}_a(\mathbf{M}). \quad (4.11)$$

Proof. Let σ_i and σ_j be permutations, $\sigma_i, \sigma_j \in S_N$. Then, Eq. 4.3 implies:

$$p\left(\mathbf{R}\left(\tilde{\mathbf{X}}_{(n)}, \tilde{\mathbf{X}}_{(m)}\right)\right) = p\left(\mathbf{R}\left(\sigma_i(\tilde{\mathbf{X}}_{(n)}), \sigma_j(\tilde{\mathbf{X}}_{(m)})\right)\right), \forall \sigma_i, \sigma_j \in S_N. \quad (4.12)$$

To express this probability invariance under permutations in a matrix form, consider a generic recurrence microstate $\mathbf{M} \equiv \mathbf{R}(\mathbf{X}_{(n)}, \mathbf{X}_{(m)})$ and use the Definitions 4.3.1 and 4.3.2 so that:

$$\begin{cases} \mathbf{R}\left(\sigma_i(\tilde{\mathbf{X}}_{(n)}), \sigma_j(\tilde{\mathbf{X}}_{(m)})\right) = \mathcal{T} \mathcal{L}_{\sigma_j} \mathcal{T} \mathcal{L}_{\sigma_i} \mathbf{M} \\ \mathbf{R}\left(\sigma_j(\tilde{\mathbf{X}}_{(m)}), \sigma_i(\tilde{\mathbf{X}}_{(n)})\right) = \mathcal{L}_{\sigma_j} \mathcal{T} \mathcal{L}_{\sigma_i} \mathbf{M} \end{cases}. \quad (4.13)$$

Considering Eqs. 4.13 and 4.6, we rewrite Eq. 4.12 as

$$p(\mathbf{M}) = p(\mathcal{T} \mathcal{L}_{\sigma_j} \mathcal{T} \mathcal{L}_{\sigma_i} \mathbf{M}) = p(\mathcal{L}_{\sigma_j} \mathcal{T} \mathcal{L}_{\sigma_i} \mathbf{M}), \sigma_i, \sigma_j \in S_N. \quad (4.14)$$

To improve notation, define the set of matrices generated by transpositions and all possible row swaps of \mathbf{M} , which is composed by the kernel of matrix operations used in Equation 4.14:

$$\mathcal{M}_a(\mathbf{M}) = \bigcup_{\sigma_i, \sigma_j \in S_N} \{\mathcal{L}_{\sigma_j} \mathcal{T} \mathcal{L}_{\sigma_i} \mathbf{M}, \mathcal{T} \mathcal{L}_{\sigma_j} \mathcal{T} \mathcal{L}_{\sigma_i} \mathbf{M}\}. \quad (4.15)$$

Equivalently, Equation 4.14 becomes:

$$p(\mathbf{M}) = p(\mathbf{M}'), \forall \mathbf{M}' \in \mathcal{M}_a(\mathbf{M}). \quad (4.16)$$

■

Note that the symmetric group is a well-established structure in Group Theory, and we employed it to characterize disorder as a symmetry—specifically, invariance under permutations. We showed that the disorder condition implies row and transposition operations in recurrence matrices. A key point to address is how we can be certain that equation (8) encompasses all possible row swaps and transpositions of a matrix. A formal justification is that the operations defined in equation (8) form a group induced by the symmetric group S_N . That is, the set of operations $\bigcup_{\sigma_i, \sigma_j \in S_N} \{\mathcal{L}_{\sigma_j} \mathcal{T} \mathcal{L}_{\sigma_i}, \mathcal{T} \mathcal{L}_{\sigma_j} \mathcal{T} \mathcal{L}_{\sigma_i}\}$ constitutes a group. This implies that any row swaps, transpositions, and any combinations of them are already included in this set. Consequently, applying this group of operations to a matrix systematically generates the entire set of matrices related by row swaps and

transpositions, thereby defining a *complete* class. Therefore, since the set of operations in Eq. 4.10 is induced by a group, any composition of operations is included in this set, ensuring the completeness of the disorder condition's implications.

Example 4.3.1 (The trivial case of matrices with order $N = 2$). Consider the case of recurrence microstates with order $N = 2$. The symmetric group S_N in this case, that is, the set of all permutations of two elements, is given by:

$$S_2 = \left\{ \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \right\} \stackrel{def}{=} \{e, \rho\}, \quad (4.17)$$

where the first permutation e is the identity (the neutral element of the group) and the second permutation ρ is the order-reversing permutation. This implies:

$$\begin{cases} e(\tilde{\mathbf{X}}_{(n)}) = e((\tilde{\mathbf{x}}_{n+1}, \tilde{\mathbf{x}}_{n+2})) = (\tilde{\mathbf{x}}_{n+1}, \tilde{\mathbf{x}}_{n+2}) = \tilde{\mathbf{X}}_{(n)} \\ \rho(\tilde{\mathbf{X}}_{(n)}) = \rho((\tilde{\mathbf{x}}_{n+1}, \tilde{\mathbf{x}}_{n+2})) = (\tilde{\mathbf{x}}_{n+2}, \tilde{\mathbf{x}}_{n+1}) \end{cases}. \quad (4.18)$$

With this definition, we can compute the class of each binary matrix according to Theorem 4.3.1. The corresponding class \mathcal{M}_a of a matrix \mathbf{M} of order $N = 2$ is given by Equation 4.10:

$$\begin{aligned} \mathcal{M}_a(\mathbf{M}) = \{ & \mathcal{L}_e \mathcal{T} \mathcal{L}_e \mathbf{M}, \mathcal{L}_e \mathcal{T} \mathcal{L}_\rho \mathbf{M}, \mathcal{L}_\rho \mathcal{T} \mathcal{L}_e \mathbf{M}, \mathcal{L}_\rho \mathcal{T} \mathcal{L}_\rho \mathbf{M}, \\ & \mathcal{T} \mathcal{L}_e \mathcal{T} \mathcal{L}_e \mathbf{M}, \mathcal{T} \mathcal{L}_e \mathcal{T} \mathcal{L}_\rho \mathbf{M}, \mathcal{T} \mathcal{L}_\rho \mathcal{T} \mathcal{L}_e \mathbf{M}, \mathcal{T} \mathcal{L}_\rho \mathcal{T} \mathcal{L}_\rho \mathbf{M} \}. \end{aligned} \quad (4.19)$$

Consider, for example, the binary matrix

$$\mathbf{M} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

as a recurrence microstate. Applying some of the operations yields:

$$\begin{cases} \mathcal{L}_\rho \mathcal{T} \mathcal{L}_\rho \mathbf{M} = \mathcal{L}_\rho \mathcal{T} \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} = \mathcal{L}_\rho \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \\ \mathcal{T} \mathcal{L}_\rho \mathcal{T} \mathcal{L}_\rho \mathbf{M} \equiv \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \end{cases}. \quad (4.20)$$

Therefore, the class of this recurrence microstate induced by the disorder condition is:

$$\mathcal{M}_a \left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right) = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}. \quad (4.21)$$

By performing all the operations defined in Equation 4.19 (cf. Theorem 4.3.1) on the

complete set of binary matrices representing all possible recurrence microstates, we obtain the full set of equivalence classes under disorder condition.

$$\begin{array}{ccccccccc}
 \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} & \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} & \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} & \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} & \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} & \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} & \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} & \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix} & \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \\
 \text{(a)} & \text{(b)} & \text{(c)} & \text{(d)} & \text{(e)} & \text{(f)} & \text{(g)} & \text{(h)} & \text{(i)} \\
 \mathcal{M}_1 & \mathcal{M}_2 & & & & \mathcal{M}_3 & & &
 \end{array}$$

$$\begin{array}{ccccccc}
 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} & \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} & \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} & \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} & \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} & \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \\
 \text{(j)} & \text{(k)} & \text{(l)} & \text{(m)} & \text{(n)} & \text{(o)} & \text{(p)} \\
 \mathcal{M}_4 & & \mathcal{M}_5 & & & & \mathcal{M}_6
 \end{array}$$

The main result is that, if the disorder condition is satisfied, the equiprobability of recurrence microstates belonging to the same class, as stated in Equation 4.11, is guaranteed. For the particular case of matrices of order $N = 2$, this equiprobability was already observed numerically [14], since it generates only 16 possible matrices, which facilitates its analysis. It was also explored to distinguish deterministic from stochastic time series; however, a formal explanation, generalization, and optimization of this property measure had so far been lacking.

4.4 Measuring data conformity to the disorder condition

Up to this point, we have imposed an ideal disorder condition on the data in order to establish a relationship among the recurrence microstates derived from it. Under the assumption of complete disorder, we have demonstrated that the recurrence microstates form equivalence classes of matrices that share the same probability. Having established this ideal upper bound for disorder, we now seek to define a quantitative measure of how closely any given dataset approaches this theoretical limit.

We seek to define a function Ξ , bounded such that $0 \leq \Xi \leq 1$, which increases monotonically with the degree of disorder. The function should attain its maximum value, $\Xi = 1$, in the case of perfect disorder—specifically, when the equiprobability condition described by Eq. 4.11 is exactly satisfied.

Let $\Omega(\mathbf{M})$ denote the number of occurrences of a given recurrence microstate $\mathbf{M} \in \mathcal{M}_a$, where \mathcal{M}_a represents an equivalence class of microstates under the ideal disorder condition. The total number of occurrences of all microstates within the same class is then given by $\sum_{\mathbf{M}' \in \mathcal{M}_a} \Omega(\mathbf{M}')$. Accordingly, the probability associated with a specific microstate \mathbf{M} within its class is defined as:

Definition 4.4.1 (Frequentist probability of a microstate). The probability of a microstate within its equivalence class is computed using the relative frequency of its occurrence.

Given a microstate $\mathbf{M} \in \mathcal{M}_a$, this probability is defined as:

$$p(\mathbf{M}) \equiv \frac{\Omega(\mathbf{M})}{\sum_{\mathbf{M}' \in \mathcal{M}_a} \Omega(\mathbf{M}')}, \quad (4.22)$$

where $\Omega(\mathbf{M})$ denotes the number of times \mathbf{M} appears, and the denominator represents the total number of occurrences of all microstates in the class \mathcal{M}_a .

Since we are dealing with a probability distribution in which the ideal disorder condition leads to a uniform distribution, and any deviation from this condition results in a departure from uniformity, it becomes necessary to adopt a function that increases with the degree of equiprobability. A natural choice that satisfies these requirements—being maximal for a uniform distribution and sensitive to deviations from it—is the information entropy [9].

Definition 4.4.2 (Information entropy of a class). The information entropy $\xi_a(\varepsilon)$ associated with the probability distribution of microstates belonging to the same class \mathcal{M}_a is:

$$\xi_a(\varepsilon) \stackrel{\text{def}}{=} - \sum_{\mathbf{M} \in \mathcal{M}_a} p(\mathbf{M}) \ln p(\mathbf{M}). \quad (4.23)$$

Let m_a denote the number of distinct microstates within the equivalence class \mathcal{M}_a of equiprobable recurrence matrices. For a given matrix order N , there exists a finite number A of such contributing classes. Among these, only two classes consist of a single microstate each: the class composed entirely of zeros, and the one composed entirely of ones. As a consequence, their entropies are identically zero and are excluded from the count of effective contributing classes.

Definition 4.4.3 (Mean information entropy of all classes). The total entropy $\xi(\varepsilon)$ is computed by summing the entropy over all nontrivial classes and normalizing the result by the maximum possible entropy amplitude determined by A :

$$\xi(\varepsilon) \stackrel{\text{def}}{=} \frac{1}{A} \sum_{a=1}^A \frac{\xi_a(\varepsilon)}{\ln m_a}. \quad (4.24)$$

Corollary 4.4.1. Let E be the largest possible distance between the elements of a data, $E \equiv \max_{\substack{i,j \\ i \neq j}} \|\mathbf{x}_i - \mathbf{x}_j\|$, $i, j \in [1, K]$. In the limit where the equiprobability condition (Definition 4.1.3) is exactly satisfied, by definition:

$$\begin{aligned} \xi_a(\varepsilon) &= 1, \forall a : 1 \leq a \leq A, \forall \varepsilon : 0 < \varepsilon < E \\ \xi(\varepsilon) &= 1, \forall \varepsilon : 0 < \varepsilon < E \end{aligned} \quad (4.25)$$

Finally, the normalized information entropy $\xi(\varepsilon)$ also increases monotonically with the degree to which the disorder condition is satisfied. As this condition is progressively fulfilled, one obtains $\xi(\varepsilon) \rightarrow 1$. In theory, this convergence occurs independently of the

specific value of the parameter ε . However, in practice, to quantify disorder with minimal error stemming from the estimation of microstate probabilities, we define the disorder quantifier Ξ .

Corollary 4.4.2. *If the ideal disorder condition is satisfied, then the recurrence microstates divided into equiprobable classes (cf. Theorem 4.3.1). Furthermore, the disorder quantifier,*

$$\Xi = \max_{\varepsilon \in (0, E)} \xi(\varepsilon), \quad (4.26)$$

increases monotonically with the degree of disorder and attains its maximum value with minimal error in the estimation of microstate probabilities.

Proof. We quantify the shortfall of each class entropy from its theoretical maximum $\ln m_a$ by defining the relative loss

$$\Delta \xi_a(\varepsilon) = 1 - \frac{\xi_a(\varepsilon)}{\ln m_a},$$

which vanishes when $\xi_a(\varepsilon) = \ln m_a$ and is positive otherwise. Averaging over the A non-trivial classes yields the mean percentage error

$$\Delta \Xi(\varepsilon) = \frac{1}{A} \sum_{a=1}^A \Delta \xi_a(\varepsilon) = 1 - \frac{1}{A} \sum_{a=1}^A \frac{\xi_a(\varepsilon)}{\ln m_a}.$$

We then select the recurrence threshold ε that minimizes this error:

$$\Delta \Xi = \min_{\varepsilon \in (0, E)} \Delta \Xi(\varepsilon) = 1 - \max_{\varepsilon \in (0, E)} \frac{1}{A} \sum_{a=1}^A \frac{\xi_a(\varepsilon)}{\ln m_a}.$$

Since the disorder quantifier is defined by

$$\Xi = \max_{\varepsilon \in (0, E)} \xi(\varepsilon),$$

and $\xi(\varepsilon) = \frac{1}{A} \sum_a \xi_a(\varepsilon) / \ln m_a$, it follows that maximizing $\xi(\varepsilon)$ is equivalent to minimizing $\Delta \Xi(\varepsilon)$. Therefore, Ξ corresponds to the minimal average entropy-loss and thus provides the optimal estimation of disorder. ■

4.5 Implementing the framework algorithm

This section presents the computational procedure for evaluating the disorder quantifier Ξ from a univariate dataset of length K . Implementation notes and optimizations accompany each titled step to ensure efficiency and reproducibility.

Overview

Input: Time series $\{\mathbf{x}_i\}_{i=1}^K$, embedding dimension N , threshold grid size c .

Output: Optimal recurrence threshold ε^* and disorder quantifier Ξ .

Step 1: Identification of Recurrence Microstate Classes

1. Generate all binary matrices of size $N \times N$ satisfying Definition 4.4.
2. Group matrices into equivalence classes \mathcal{M}_a via row-permutation and transposition per Theorem 4.3.1.
3. Store a lookup table mapping each matrix to its class index for constant-time assignment.
4. *Implementation note:* Construct only recurrence-consistent matrices to avoid full enumeration of 2^{N^2} configurations.

Step 2: Generation of Recurrence Microstates

1. Precompute the pairwise distance matrix among series elements.
2. For each threshold ε_j in a uniform grid of size c over $(0, E)$:
 - Slide a window of length N to extract each subsequence $\mathbf{X}_{(n)}$.
 - Build the binary recurrence matrix by comparing within-window distances to ε_j .
 - Assign the matrix to its class via the precomputed lookup table.
3. *Parallelization:* Window and threshold loops are independent and can be parallelized.
4. *Parameter selection:* Choose N to set the analysis timescale (here, $N = 4$).

Step 3: Frequency Counting and Probability Estimation

1. Initialize counters $\Omega(\mathbf{M})$ for each microstate \mathbf{M} .
2. Increment the count for each generated microstate.
3. After processing all subsequences, compute probabilities according to Definition 4.22, ensuring normalization within each class.
4. *Data structures:* Use hash maps or fixed-size arrays for constant-time updates.

Step 4: Disorder Quantifier Evaluation and Threshold Selection

1. For each ε_j , compute $\xi(\varepsilon_j)$ via Equation 4.26.
2. Determine

$$\varepsilon^* = \arg \max_{j=1, \dots, c} \xi(\varepsilon_j), \quad \Xi = \xi(\varepsilon^*).$$

3. *Early stopping:* Terminate if $\xi(\varepsilon)$ stabilizes within tolerance.
4. *Validation:* Plot $\xi(\varepsilon)$ versus ε to confirm convergence.

Computational Complexity and Best Practices

- Complexity scales dominated by distance and threshold sweeps.
- Cache distance computations to avoid redundancy across ε .
- Implement core loops in compiled or vectorized form to reduce overhead.

- Ensure lookup tables and counters remain cache-resident for performance.

This algorithmic framework delivers a reproducible and efficient approach to quantifying disorder in time series via recurrence microstates.

Numerical validations using benchmark models

This chapter presents a comprehensive numerical validation framework for the proposed disorder quantifier Ξ , systematically addressing three fundamental questions:

- (i) How robustly does Ξ capture disorder transitions across dynamical paradigms?
- (ii) What are its operational limits under constrained observational conditions?
- (iii) How does it compare to established complexity measures in practical scenarios?

To address these questions, we employ a multi-stage validation strategy across 15 benchmark models spanning discrete maps, continuous flows, stochastic processes, and deterministic-stochastic hybrids. Our methodology adheres to parameter sensitivity, statistical rigor, limited data and cross-paradigm consistency.

The chapter is structured with the Section 5.1 establishing methodological foundations for disorder quantification. Section 5.2 validates statistical tests to sensitivity analysis. Section 5.3 examines quantifier's convergence behavior with respect to data length K .

The core validation spans Sections 5.4-5.7:

- §5.4 evaluates Ξ across deterministic (discrete/continuous) and stochastic models;
- §5.5 compares the disorder measure directly to past dependence of stochastic processes;
- §5.6 tests Ξ on deterministic-stochastic mixtures;
- §5.7 conducts head-to-head comparisons against DET, PE, and SC metrics.

Collectively, these analyses establish the disorder quantifier Ξ as a robust, cross-paradigmatic disorder measure with superior performance in finite-precision observational scenarios. The validation protocol intentionally stresses the quantifier under suboptimal

conditions—short time series, coarse sampling, and mixed dynamics—to demonstrate operational viability for real-world applications.

5.1 Methodological considerations on the disorder quantification

All analyses were performed by computing recurrence microstates with size $N = 4$. This choice provides a balance between capturing sufficient temporal disorder in the time series and maintaining computational efficiency, as larger values of N ($N > 4$) substantially increase computational costs without proportionally enhancing the quantification of disorder.

For $N = 4$, the total number of possible recurrence microstates is 47,416, which are systematically grouped into 147 distinct classes by applying Eq. 4.10. Consequently, the effective number of classes contributing to the entropy computation is $A = 145$. For comparison, when using $N = 3$, the number of distinct classes reduces to $A = 23$, and for $N = 2$, to $A = 4$. The length of the time series analyzed, denoted by K , varies within the range $10 \leq K \leq 10,000$, allowing for a comprehensive assessment of the quantifier's sensitivity across different temporal resolutions.

5.2 Statistical tests for the quantifier evaluation

To evaluate the statistical significance of the differences in the disorder quantifier Ξ computed across various models, we perform rigorous statistical tests based on the Welch's t -test [62]. This choice is motivated by the potential heterogeneity in variances of the quantifier across the different samples analyzed, since different behavior is a common feature in dynamical systems due to the initial conditions, for example.

The Welch's t -test is a robust statistical method for comparing the means of two independent samples when the assumption of equal variances is not tenable. Let $X = x_1, x_2, \dots, x_{n_1}$ and $Y = y_1, y_2, \dots, y_{n_2}$ denote two independent samples, with respective sample means \bar{x} and \bar{y} , sample variances s_X^2 and s_Y^2 , and sizes n_1 and n_2 .

The test statistic is computed as:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_X^2}{n_1} + \frac{s_Y^2}{n_2}}} \quad (5.1)$$

This statistic follows approximately a Student's t -distribution with effective degrees of freedom ν , determined by the Welch-Satterthwaite equation:

$$\nu = \frac{\left(\frac{s_X^2}{n_1} + \frac{s_Y^2}{n_2}\right)^2}{\frac{\left(\frac{s_X^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{s_Y^2}{n_2}\right)^2}{n_2-1}} \quad (5.2)$$

The hypotheses tested are:

- Null hypothesis (H_0): The population means are equal, i.e., $\mu_X = \mu_Y$. In this case, one can not distinguish the mean values of the disorder from the two samples.
- Alternative hypothesis (H_1): The population means are different, i.e., $\mu_X \neq \mu_Y$.

The p -value derived from the t -statistic and ν indicates whether the observed difference is statistically significant. Specifically, we reject H_0 at a significance level $\alpha = 0.05$ if $p < \alpha$, implying sufficient evidence to claim a statistically significant difference between the mean disorder values of the compared groups.

The results are also presented using boxplots overlaid with violin plots. This combination allows the summary statistics, such as the median, interquartile range, outliers, and the full probability distribution of the measurements to be visualized simultaneously. The boxplots highlight central tendencies and variability, while the violin plots illustrate the underlying density of the data.

By confirming the significance of the disorder differences, this statistical procedure enhances the reliability of our interpretations, supporting the identification of model regimes. Furthermore, this approach reinforces the applicability of the disorder quantifier as a robust tool for detecting dynamical shifts in complex data.

5.3 The disorder quantifier and the data size

A key assumption underlying our recurrence-based framework is Hypothesis 4.2.1, which posits that the total data length K is sufficiently large compared to the subsequence length N (i.e., $K \gg N$). To evaluate numerically how well various signals satisfy Hypothesis 4.2.1, we compute the disorder quantifier Ξ for White Gaussian Noise (WGN), Pink Gaussian Noise (PGN), Red Gaussian Noise (RGN), and the logistic map, systematically varying the time series length K (see Fig. 5.1(a)). We compute the quantifier over 30 time series for each model, at a timescale $N = 4$.

White Gaussian Noise, being completely uncorrelated, is the only process for which Ξ approaches its maximum ($\Xi = 1$) as K increases. This behavior confirms that WGN satisfies the ideal disorder condition (Hypothesis 4.1.1); moreover, by the Law of Large Numbers, probability estimates and the resulting entropy-based quantifier converge rapidly to their theoretical values as K grows.

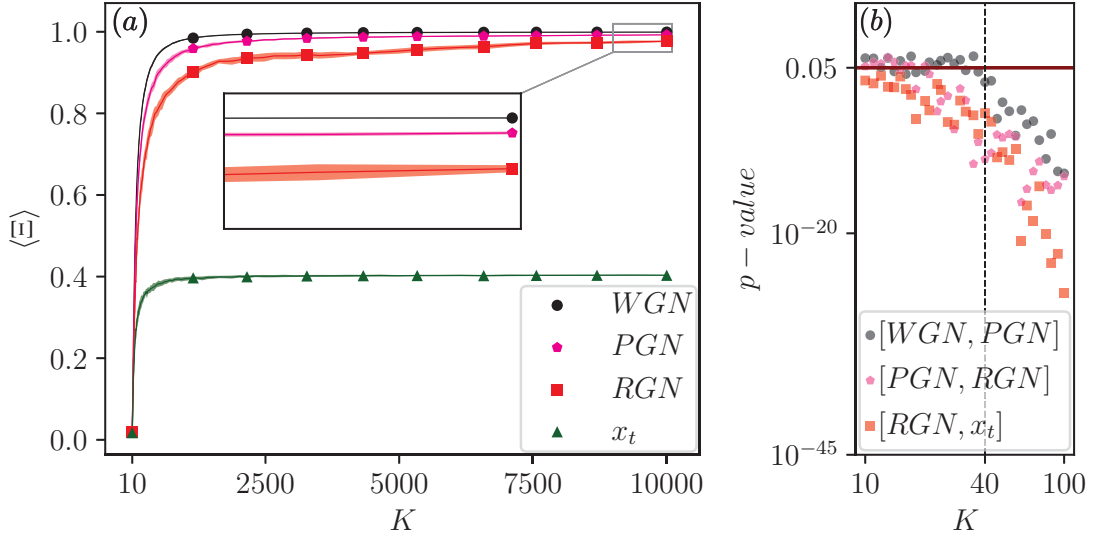


Figure 5.1: (a) The mean of the disorder quantifier $\langle \Xi \rangle$ with its corresponding standard deviation shaded on each curve as a function of the length of the time series K for the logistic map, WGN, PGN and RGN. (b) The p -values of Welch's t -test $[A, B]$ comparing the mean of disorder for the model A and model B as a function of the time series length [63].

By contrast, RGN exhibits stronger temporal correlations than PGN, and accordingly its disorder quantifier remains lower. Since Ξ is designed to increase monotonically with the degree of disorder, PGN—which has weaker correlations than RGN—yields a higher Ξ for the same K . This ordering is consistent with the requirement that signals closer to the ideal disorder condition produce larger disorder values.

Finally, all stochastic signals (WGN, PGN, RGN) yield disorder values that differ markedly from the logistic map, which converges to a value near $\Xi \approx 0.4$ as K increases. The logistic map does not satisfy the disorder condition, and its relatively low Ξ reflects its deterministic (chaotic) dynamics. Thus, the dependence of Ξ on K demonstrates both the convergence properties for stochastic processes and the clear distinction between stochastic and chaotic behavior.

The differences in disorder values among the models are statistically significant, as verified by Welch's t -test $[A, B]$ comparing the mean of disorder for the model A and model B as a function of the time series length. For $K > 40$, all p -values fall below 0.05, indicating that the mean disorder values for WGN, PGN, RGN, and the logistic map can be distinguished with greater than 95% confidence when $K > 40$ (Fig. 5.1(b)).

5.4 The quantifier and the classification of different model regimes

Quantifying dynamical disorder across different classes of systems requires a unified metric capable of capturing both deterministic chaos and stochastic randomness. This section presents a systematic evaluation of the proposed disorder measure Ξ applied to three fundamental paradigms: discrete maps, continuous flows, and stochastic processes. For each paradigm, we examine five canonical systems under controlled parameter variations (p_1 vs p_2) designed to induce graduated disorder levels.

Our experimental design maintains methodological consistency across paradigms: (1) generation of 50 independent realizations per parameter configuration, (2) fixed trajectory length ($K = 3000$ points), and (3) identical statistical visualization (boxplot-violin composites). Crucially, deterministic systems incorporate appropriate transient removal ($T = 1000$) and sampling protocols (time step of 0.5), while stochastic processes emphasize correlation decay properties.

The following subsections validate Ξ 's discriminative capacity through three critical lenses: sensitivity to parameter-induced disorder transitions (Sections 5.4.1-5.4.2), differentiation between deterministic and stochastic regimes (Section 5.4.3), and robustness under constrained observational conditions ($K < 10^4$, coarse sampling). Collectively, these analyses establish Ξ as a cross-paradigm disorder metric with significant implications for characterizing complex dynamics.

5.4.1 Disorder in discrete deterministic systems

We evaluate five paradigmatic discrete maps, including those with more intricate dynamics such as the hyperchaotic Towel map and the Standard map, the latter exhibiting a highly “cloud-like” phase-space structure due to its strong diffusive behavior. The formal definitions and parameter settings for each map are provided in Section 3.1.

In each case, two parameter values, p_1 and p_2 , are chosen so that p_2 is known to produce a higher degree of dynamical disorder compared to p_1 . For each map and each parameter value, we generated 50 independent time series of length $K = 3000$ to compute the disorder quantifier Ξ .

Figure 5.2 presents the results as combined boxplots and violin plots, illustrating the distribution of Ξ across those realizations. The blue boxplots and violin plots correspond to the parameter value p_1 for each model, while the red plots refer to p_2 , the inducing parameter value of higher disorder. Boxplots denote interquartile ranges while violin contours show kernel density estimates.

In all cases, increasing the control parameter from p_1 to p_2 yields a statistically significant rise in Ξ , confirming that the quantifier effectively captures the expected

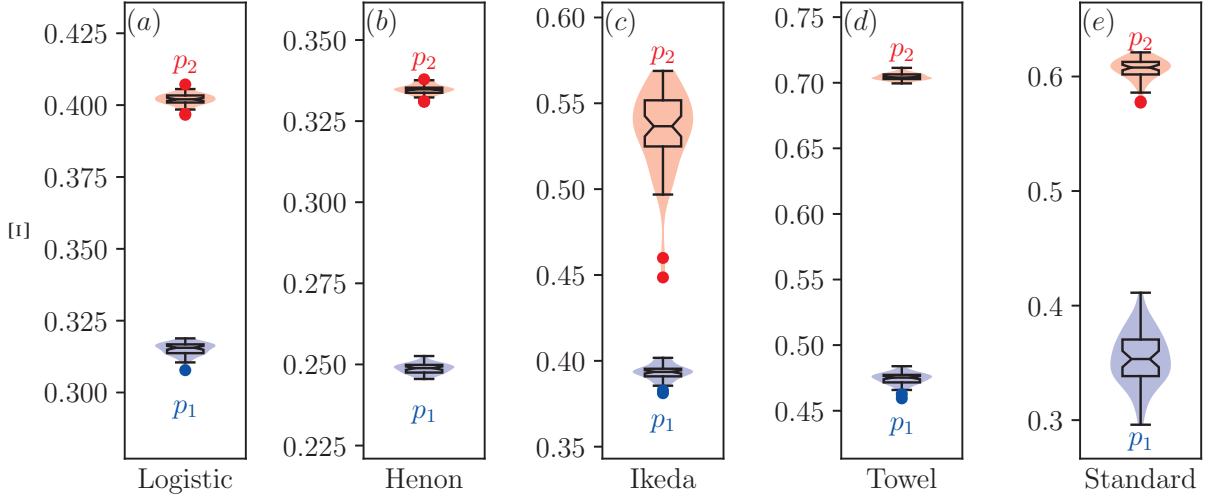


Figure 5.2: Boxplots with underlaid violin plots showing the distribution of the disorder quantifier Ξ for each discrete dynamical system under parameter values p_1 and p_2 [63].

increase in disorder. Notably, the highest disorder values are observed for the hyperchaotic Towel map, reflecting its highly irregular dynamics.

5.4.2 Disorder in continuous deterministic systems

We examine five continuous dynamical systems, each evaluated at two distinct control-parameter values, p_1 and p_2 , where p_2 is known to produce a higher level of dynamical disorder than p_1 . Formal definitions and specific parameter settings for these systems are provided in Section 3.2.

For each model and each parameter value, we generated 50 independent trajectories, discarding a transient of 1,000 time units. Initial conditions were chosen at random, except in the Double-Gyre flow, where they were selected within the chaotic sea. Each trajectory was sampled at fixed time intervals of 0.5 and truncated to a total length of $K = 3000$ points, ensuring that our analysis applies to relatively short, low-temporal-resolution datasets ($K < 10^4$).

Figure 5.3 presents boxplots with overlaid violin plots of the disorder quantifier Ξ for each system under both p_1 and p_2 .

In every case, the disorder measure Ξ is significantly larger at p_2 than at p_1 , demonstrating that the proposed quantifier reliably detects the expected increase in disorder even for short, coarsely sampled time series.

5.4.3 Disorder in stochastic processes

We extend our analysis to stochastic processes, evaluating two parameter values (p_1 and p_2) where p_2 enhances disorder through reduced temporal correlations or weakened

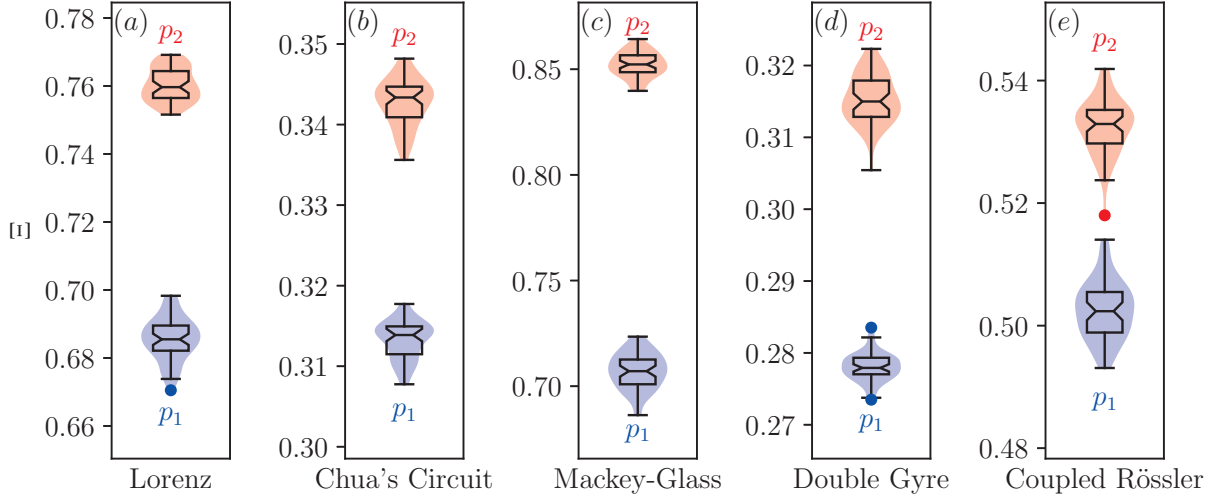


Figure 5.3: Boxplots with violin overlays showing the distribution of the disorder quantifier Ξ for each continuous dynamical system under parameter values p_1 (blue) and p_2 (red) [63].

dependence on past states. In this case, the conceptualization of disorder fundamentally differs from deterministic interpretations, being intrinsically linked to decay properties of autocorrelation functions.

Figure 5.4 presents results from 50 realizations per parameter-process pair, each of length $K = 3000$:

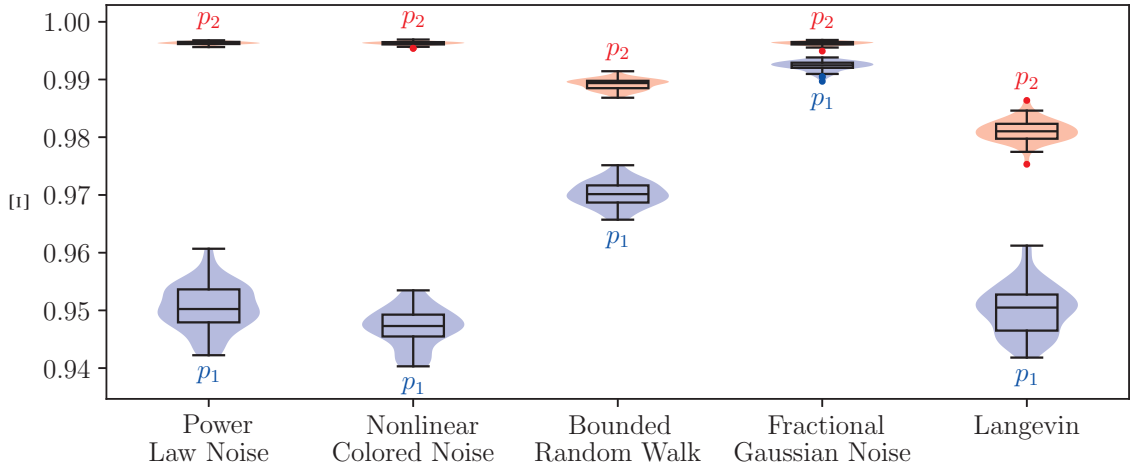


Figure 5.4: Boxplots with violin plots showing the values of the proposed quantifier Ξ for each stochastic process and its parameter values p_1 and p_2 [63].

The disorder measure Ξ exhibits statistically significant increases when transitioning to higher-disorder parameter p_2 across all processes. Notably, Ξ values for stochastic systems exceed 0.94 – substantially higher than those observed in deterministic systems ($\Xi < 0.865$) despite identical sampling constraints ($K = 3000$, $\Delta t = 0.5$).

Even the hyperchaotic Towel map (maximum deterministic $\Xi = 0.712$) remains well

below minimally disordered stochastic cases. This pronounced separation confirms Ξ 's discriminative capacity between stochastic and deterministic disorder paradigms under finite-resolution observations.

5.5 Disorder and past dependence

The autoregressive (AR) model, defined in Section 3.3.2, is a simple stochastic process that allows direct control over the degree of past dependence through the parameter ϕ . The disorder condition of Hypothesis 4.1.1 corresponds to the limiting case of no memory, i.e., $\phi = 0$.

Since the disorder quantifier Ξ is constructed to increase monotonically as this memory-less condition is approached, the AR model serves as an appropriate testbed for numerically validating our framework. To this end, we computed Ξ for 30 realizations of AR model time series, each of length 3,000, using a timescale of $N = 4$.

As the parameter ϕ increases, the process exhibits stronger dependence on its past values, leading to a systematic decrease in the disorder quantifier. This inverse relationship is clearly observed in Fig. 5.5, where the mean disorder $\langle \Xi \rangle$ declines monotonically with increasing ϕ .

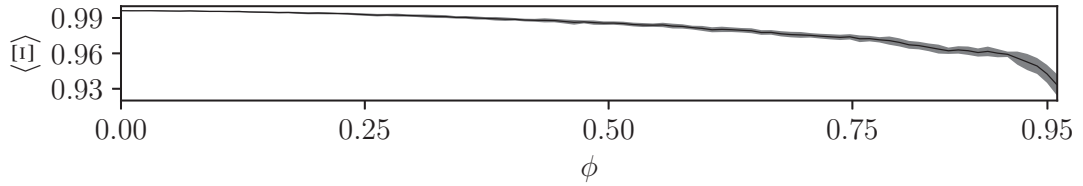


Figure 5.5: The mean of the disorder quantifier $\langle \Xi \rangle$ for the autoregressive model. The disorder decreases as the parameter ϕ increases [63].

5.6 Disorder in data from deterministic-stochastic mixtures

Because the disorder quantifier Ξ is sensitive both to transitions between deterministic chaos and purely random processes, and to different forms of temporal correlation in noise, we now investigate its behavior on time series that combine deterministic chaotic dynamics with stochastic perturbations. Specifically, we assess the robustness of Ξ when a chaotic signal is corrupted by noise with varying degrees of temporal dependence.

To this end, we employ the noisy logistic model (Eq. 3.3.3), in which a chaotic logistic map trajectory is perturbed by additive noise $\eta_t(\phi)$ generated by an autoregressive (AR) process for correlation parameters $\phi \in \{0.00, 0.25, 0.50, 0.75, 0.90\}$.

We systematically vary the variance ratio $\frac{\sigma_\eta^2}{\sigma_x^2}$, where σ_η^2 is the variance of $\eta_t(\phi)$ and σ_x^2 is the variance of the purely chaotic logistic map x_t . As this ratio increases from 0 (purely deterministic) to 1 (noise variance equal to signal variance), the combined process transitions from deterministic chaos to a predominantly stochastic regime.

For each combination of ϕ and variance ratio, we generate 30 independent realizations of length $K = 3000$ and compute the disorder quantifier Ξ using a recurrence timescale of $N = 4$. Figure 5.6(a) displays the mean disorder $\langle \Xi \rangle$ as a function of σ_η^2/σ_x^2 for each value of ϕ .

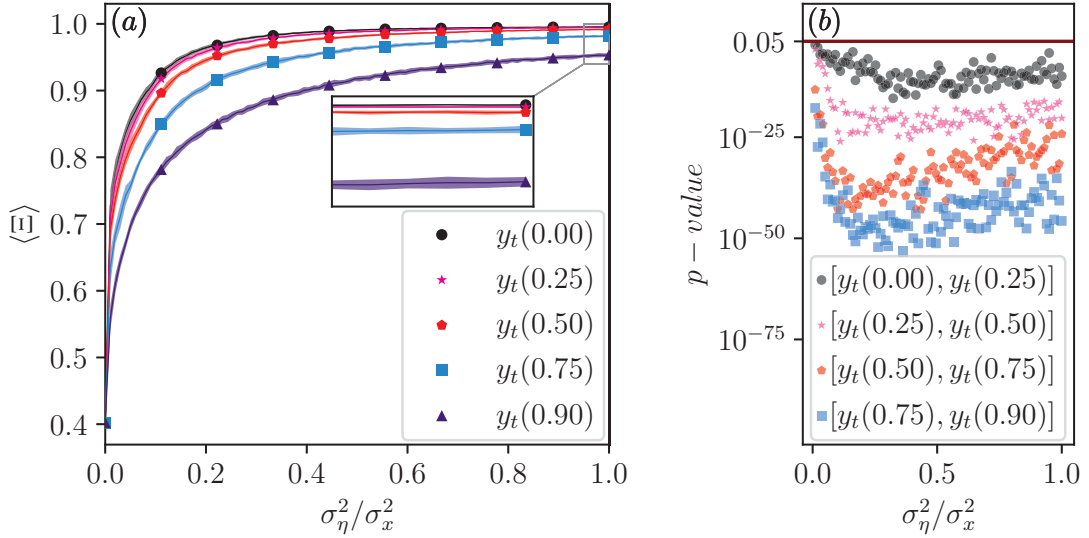


Figure 5.6: (a) Mean disorder quantifier $\langle \Xi \rangle$ for the noisy logistic model $y_t(\phi) = x_t + \eta_t(\phi)$, plotted against the variance ratio σ_η^2/σ_x^2 . Each curve corresponds to a different AR correlation parameter ϕ . (b) Corresponding p -values from Welch's t -test comparing the mean disorder between pairs of models as a function of σ_η^2/σ_x^2 [63].

The disorder increases monotonically with the variance ratio σ_η^2/σ_x^2 , but the slope of this increase depends on the temporal correlation ϕ : signals with higher ϕ (stronger memory) exhibit a slower rise in Ξ compared to those with weaker correlation.

Figure 5.6(b) reports the p -values of Welch's t -test for pairwise comparisons of mean disorder between different values of ϕ at each variance ratio. For sufficiently large variance ratios, the differences in $\langle \Xi \rangle$ between models become statistically significant ($p < 0.05$), confirming that the quantifier reliably distinguishes processes with varying degrees of correlation even when deterministic and stochastic components are mixed.

5.7 Comparisons with prior measures

We conduct analyses on the same datasets generated in the previous responses, comparing with four well-known methods defined in the Chapter 2: the permutation entropy

(PE) [26], the multivariate permutation entropy (MvPE) [27], the statistical complexity (SC) [16], and the determinism by recurrence quantification (DET) [25]. All methods were implemented using the Julia language packages `ComplexityMeasures.jl` [29] and `RecurrenceAnalysis.jl` [30].

To facilitate comparison, we summarize in Fig. 5.7 the disorder values obtained for the models that most challenge alternative methods. It consistently increases as the control parameter shifts from p_1 to p_2 (Fig. 5.7), reflecting the quantifier’s sensitivity to changes in system behavior. Notably, all disorder values calculated for the paradigmatic stochastic processes exceed $\Xi = 0.94$, demonstrating the expected high disorder in these cases. In contrast, the dynamical systems analyzed yield disorder values below $\Xi = 0.8$, even when the datasets are undersampled.

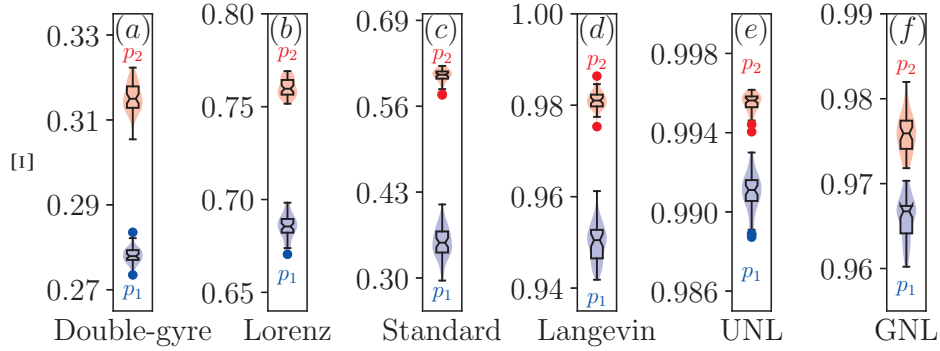


Figure 5.7: Boxplots with violin plots showing the values of the proposed disorder quantifier for critical cases [63].

These results confirm a clear and significant distinction between the disorder levels of typical stochastic processes and deterministic dynamics. They also highlight the robustness of Ξ in quantifying order–disorder phenomena across diverse scenarios, supporting its utility as a reliable measure to differentiate between fundamentally different classes of processes.

5.7.1 Testing determinism (DET)

We assess the determinism (DET) metric for critical test cases using the established boxplot-violin visualization framework. Figure 5.8 presents distributions of DET values across selected systems under parameter configurations p_1 and p_2 .

The Standard map exhibits characteristic deterministic behavior with elevated DET values ($\text{DET} > 0.5$), showing decreased determinism under disorder-inducing parameter p_2 . Conversely, the delay-differential Mackey-Glass system displays anomalous behavior: its p_2 configuration yields significantly lower DET values than p_1 , contradicting both the proposed disorder quantifier Ξ (which increases at p_2) and the structurality and entropy measures [28].

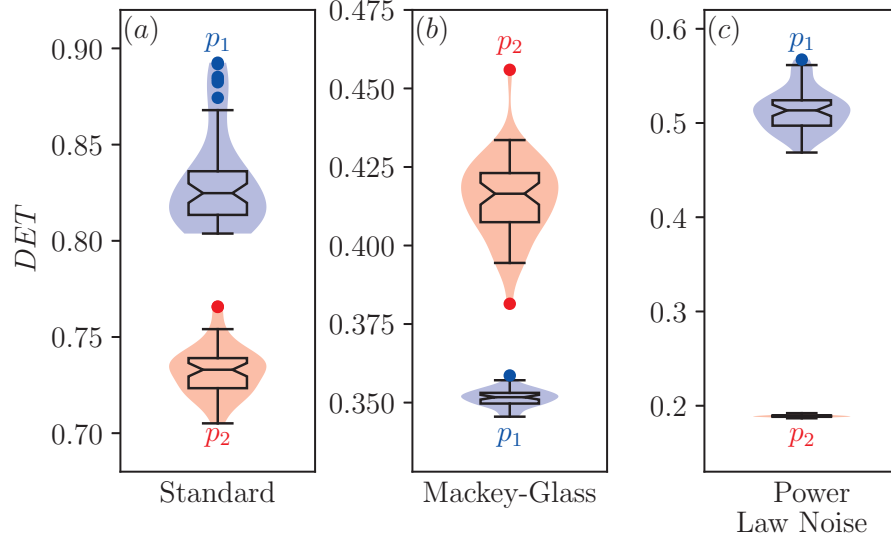


Figure 5.8: Boxplots with violin plots showing the values of the Determinism (DET) for critical cases [63].

Notably, Mackey-Glass determinism under p_2 falls below that of correlated power-law noise, suggesting potential limitations of DET as a universal determinism indicator in systems with complex temporal organization.

5.7.2 Testing permutation entropy (PE)

The Figure 5.9 presents comparative distributions of Permutation Entropy (PE) and Multivariate Permutation Entropy (MvPE) for critical test cases under parameter configurations p_1 and p_2 .

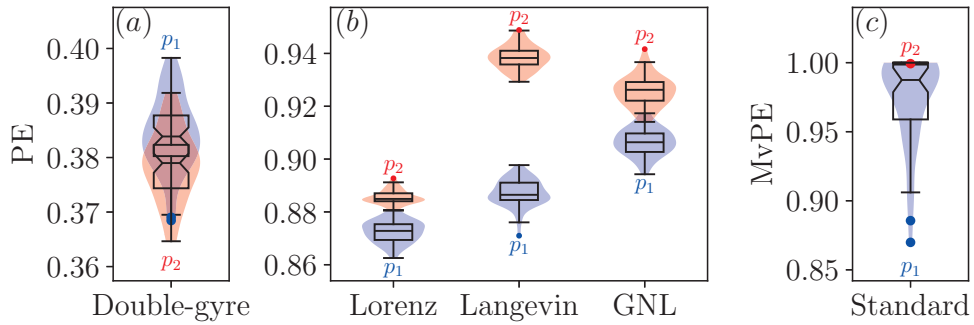


Figure 5.9: Boxplots with violin plots showing the values of the Permutation Entropy (PE) and Multivariate Permutation Entropy (MvPE) for critical cases [63].

The analysis reveals three significant limitations of entropy-based measures:

1. Low discriminative power: The Double-Gyre system shows substantial overlap in PE distributions between p_1 and p_2 , failing to detect the expected disorder increase [28];
2. Paradigm ambiguity: Lorenz system PE values are statistically indistinguishable from the stochastic Langevin equation, confounding deterministic chaos with stochasticity;

3. Anomalous saturation: The Standard Map attains maximum theoretical MvPE despite being deterministic, while the GNL shows parameter overlap.

These observations suggest fundamental limitations in entropy-based measures for distinguishing dynamical regimes and parameter-induced disorder transitions, particularly in systems with complex temporal structures.

5.7.3 Testing statistical complexity (SC)

The Figure 5.10 presents distributions of Statistical Complexity (SC) and its multivariate counterpart (MvSC) computed in conjunction with the permutation entropy metrics analyzed in Section 5.7.4.

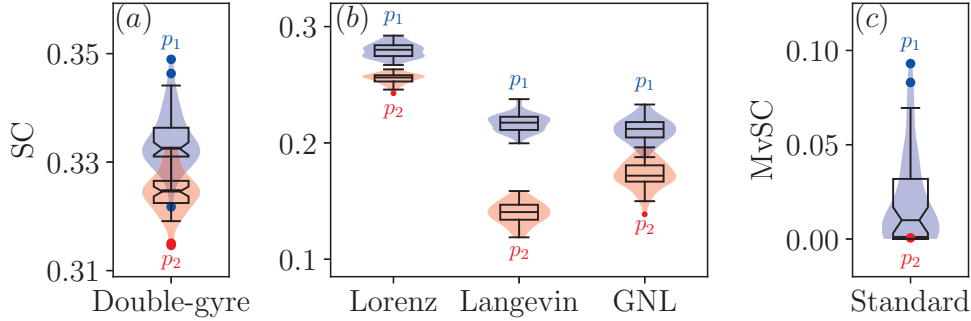


Figure 5.10: Boxplots with violin plots showing the values of the Statistical Complexity (SC) for critical cases [63].

The results are similar to those obtained by the PE. While the joint PE-SC plane approach [16] theoretically enhances characterization, it introduces significant practical limitations:

- Indirect interpretation of disorder;
- Obfuscation of temporal disorder progression;
- Increase dimensionality complicates real-world analysis.

These findings suggest that complexity measures, whether univariate or multivariate, provide limited direct insight into disorder dynamics when used in isolation, while their combined implementation sacrifices interpretative clarity for marginal diagnostic gains.

5.7.4 Testing structurality (Δ) and dynamical complexity (DC)

Although the structurality measure Δ performs satisfactorily in several scenarios, such as those illustrated in Fig. 5.11(a) and (b), where its results are in close agreement with the proposed disorder quantifier (cf. Fig. 5.7(a) and (c)), it exhibits notable limitations in

discriminating between certain types of signals. Specifically, when applied to the uniform-noisy logistic (UNL) and Gaussian-noisy logistic (GNL) models, Δ produces overlapping values across these cases, as shown in Fig. 5.11(c) and (d).

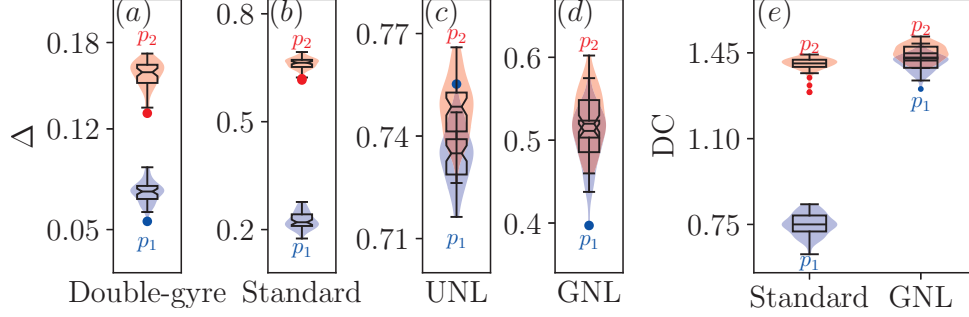


Figure 5.11: Boxplots with violin plots showing the values of the structrality (Δ) and dynamical complexity (DC) for critical cases [63].

A similar pattern emerges when considering the dynamical complexity (DC). While DC performs well with the standard map, successfully capturing the expected increase in disorder, it also yields overlapping values in the GNL scenario (see Fig. 5.11(e)). This observation indicates that both Δ and DC tend to produce results for GNL that fall within the same range as those obtained for the standard map. Consequently, deterministic and stochastic signals with differing underlying structures can be assigned comparable complexity scores. These findings suggest that, despite their utility in certain applications, Δ and DC have limited discriminative power in contexts where fluctuations and randomness dominate the dynamics.

5.7.5 Summarizing the comparative tests

Table 5.1 synthesizes the comparative performance evaluation across all tested metrics, highlighting distinctive advantages of the proposed disorder measure Ξ when applied to short ($K = 3000$) and undersampled (time step 0.5) time series across deterministic and stochastic scenarios.

This systematic evaluation demonstrates that Ξ uniquely satisfies all five criteria, outperforming in quantifying disorder transitions across deterministic and stochastic systems. Crucially, it achieves robust performance despite stringent observational constraints, establishing it as a general disorder metric for multivariate time series analysis under limited-data conditions while maintaining consistent performance, statistical significance, and interpretability.

Table 5.1: Summary of the comparison between the proposed quantifier Ξ and similar traditional methods.

Performance criteria	Ξ	DET	PE	MvPE	SC	Δ	DC
Correct and statistically significant distinction of disorder-inducing parameter values (p_1 and p_2) for all tested models	Yes	No	No	No	No	No	No
Correct distinction of deterministic and stochastic typical cases	Yes	No	No	No	Yes*	No	No
Independence of dimensionality	Yes	Yes	No	Yes	No	Yes	Yes
Optimal and inherent selection of its parameters	Yes	No	No	Yes	Yes	Yes	Yes
Direct interpretation via a single measure	Yes	Yes	Yes	Yes	No*	Yes	No

*If used together with PE.

Real-world application to understand Earth's dynamics

The Cenozoic Era, spanning approximately the last 66 million years, represents a critical period in Earth's paleoclimate history, marked by significant transitions such as the Paleocene-Eocene Thermal Maximum, the onset of Antarctic glaciation, and the Quaternary glacial-interglacial cycles. Understanding these climatic shifts is essential for reconstructing past environmental conditions, identifying natural climate drivers, and contextualizing current anthropogenic changes. However, the analysis of paleoclimate data faces several challenges: the datasets are typically irregularly sampled, contain substantial noise from sedimentary and diagenetic processes, and are finite and incomplete due to the inherent limitations of geological archives. Moreover, distinguishing deterministic patterns, such as astronomical forcing, from stochastic variability remains a central problem, requiring robust quantitative methods capable of handling complex, noisy, and often non-stationary time series.

6.1 The Cenozoic dataset

6.1.1 Data acquisition and structure

The Cenozoic Global Reference benthic foraminifer carbon and oxygen Isotope Dataset (CENOGRID) [21] is derived from a synthesis of high-resolution stable isotope records ($\delta^{18}\text{O}$ and $\delta^{13}\text{C}$) obtained from 14 deep-sea sediment cores collected during the International Ocean Discovery Program (IODP) and its predecessors. These records ensure precise orbital tuning with age uncertainties ranging from ± 100 kiloyear (kyr) to ± 10 kyr. The dataset corrects for interspecies isotopic offsets by prioritizing analyses of *Cibicidoides* and *Nuttallides* species, while gaps in critical intervals (e.g., late Miocene) are filled with new benthic isotope measurements. This approach resolves orbital-scale climate variability (10^3 – 10^6 years) and addresses longstanding challenges of fragmented records and coarse

temporal resolution in earlier compilations.

6.1.2 Physical interpretation and geological stages

The CENOGRID physically represents global temperature trends (via $\delta^{18}\text{O}$, adjusted for ice-volume effects) and carbon cycle dynamics (via $\delta^{13}\text{C}$). It identifies four distinct climate states: (1) *Hothouse* (56–47 Ma), characterized by extreme warmth ($> 10^\circ\text{C}$ above modern) and frequent hyperthermal events; (2) *Warmhouse* (66–56 Ma and 47–34 Ma), marked by sustained warmth ($> 5^\circ\text{C}$ above modern) and stable carbon-temperature coupling; (3) *Coolhouse* (34–3.3 Ma), defined by Antarctic ice sheet expansion and eccentricity-to-obliquity transitions; and (4) *Icehouse* (post-3.3 Ma), dominated by Northern Hemisphere glaciation and high-latitude feedbacks.

The Cenozoic era is also divided into stages linked to the driving triggers of transitions, such as astronomical forcing, tectonic reorganizations, and biogeochemical feedbacks. A detailed description of each stage based on the book “Geologic Time Scale 2020: Volume 2” [64] is provided below:

- **Dan (Danian; 66–61.6 Ma):** Following the Cretaceous–Paleogene (K/Pg) mass extinction, the Danian marks a period of ecological recovery and gradual warming. $\delta^{13}\text{C}$ records indicate a stabilization of the global carbon cycle after the abrupt perturbations associated with the extinction event. Concurrently, $\delta^{18}\text{O}$ data show a progressive return to warmer ocean temperatures from the relatively cool conditions that characterized the latest Cretaceous. This stage is also defined paleontologically by the rapid diversification of planktonic foraminifera and calcareous nannoplankton, which serve as key biostratigraphic markers.
- **Sel (Selandian; 61.6–59.2 Ma) and Tha (Thanetian; 59.2–56 Ma):** The late Paleocene is characterized by sustained warming, culminating in a series of prelude events to the Paleocene-Eocene Thermal Maximum (PETM). The increasing trend in $\delta^{13}\text{C}$ excursions suggests intensified carbon cycle variability, possibly linked to volcanic outgassing or methane hydrate destabilization. These stages are distinguished paleontologically by turnover events in benthic foraminifera and diversification of mammals, particularly archaic ungulates, which reflect significant terrestrial ecosystem restructuring.
- **Ypr (Ypresian; 56–47.8 Ma):** The early Eocene represents the peak of the Cenozoic “Hothouse” conditions, exemplified by the PETM, a rapid warming event associated with a sharp negative $\delta^{13}\text{C}$ excursion (up to -4‰) and elevated $\delta^{18}\text{O}$ values indicating deep-sea temperatures exceeding 12°C . Paleontologically, this stage is notable for the first appearance of modern mammalian orders such as primates,

perissodactyls, and artiodactyls, as well as significant marine plankton turnover linked to ocean acidification and temperature changes.

- **Lut (Lutetian; 47.8–41.2 Ma) and Bar (Bartonian; 41.2–37.8 Ma):** The middle Eocene initiates a long-term cooling trend. $\delta^{18}\text{O}$ values increase by approximately 1‰, signaling a gradual decline in global temperatures and the initial buildup of ephemeral Antarctic ice sheets. These stages are delineated by key paleontological markers, including the diversification of early cetaceans and sirenians in marine environments, as well as floral shifts from tropical to more temperate assemblages in mid-latitudes.
- **Pri (Priabonian; 37.8–33.9 Ma):** The late Eocene culminates in the Eocene-Oligocene Transition (EOT), a major climatic reorganization characterized by a sharp positive shift in $\delta^{18}\text{O}$ (about +1.5‰). This isotopic shift reflects both significant cooling and the first permanent formation of a large-scale Antarctic ice sheet. Biostratigraphically, the Priabonian is marked by a significant extinction event among deep-sea benthic foraminifera and nannoplankton, as well as turnover in terrestrial mammal faunas in response to climatic deterioration.
- **Rup (Rupelian; 33.9–27.8 Ma) and Cha (Chattian; 27.8–23.0 Ma):** The Oligocene is typified by the establishment of the “Coolhouse” climate state, marked by relatively stable but cooler conditions compared to the Eocene. The coupling between $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$ records indicates the development of complex feedbacks between the carbon cycle and expanding polar ice sheets. Paleontological data define these stages through significant faunal exchanges between continents (e.g., the Grande Coupure in Europe), as well as changes in marine microfossil assemblages reflecting cooler, more seasonally variable climates.
- **Aqu (Aquitanian; 23.0–20.4 Ma) and Bur (Burdigalian; 20.4–14.9 Ma):** The early Miocene witnesses a transient return to warmer conditions during the Miocene Climatic Optimum (MCO), with $\delta^{18}\text{O}$ values decreasing to less than –0.5‰, implying reduced ice volume and warmer oceans. These stages are paleontologically distinguished by the radiation of modern cetaceans and pinnipeds, as well as terrestrial mammal diversification, notably of ruminants and early hominoids. Marine biostratigraphy is also refined by planktonic foraminiferal zonations.
- **Lan (Langhian; 14.9–13.8 Ma) and Ser (Serravallian; 13.8–11.6 Ma):** The middle Miocene is marked by the Middle Miocene Climate Transition (mMCT) around 13.9 Ma, with $\delta^{18}\text{O}$ increasing by approximately +0.8‰. This event represents the expansion of Antarctic ice sheets to near-modern extents and a shift towards obliquity-paced glacial-interglacial cycles. Paleontologically, these stages are defined by notable

turnovers in marine microfauna, especially the extinction and radiation of certain foraminiferal and nannoplankton groups, alongside the continued diversification of terrestrial mammals.

- **Tor (Tortonian; 11.6–7.2 Ma) and Zan (Zanclean; 7.2–3.6 Ma):** The late Miocene to early Pliocene period is characterized by further depletion of $\delta^{13}\text{C}$ (around -1‰), likely associated with decreased organic carbon burial, as well as a gradual trend in $\delta^{18}\text{O}$ indicating the progressive development of Northern Hemisphere glaciation. Paleontological markers include the Messinian Salinity Crisis near the Tortonian-Zanclean boundary, major faunal exchanges such as the Hipparion dispersal, and continued diversification of hominids and other primates, indicating increasingly complex terrestrial ecosystems.

6.1.3 Scientific relevance and prior insights

The CENOGRID dataset has significantly advanced our understanding of Cenozoic climate sensitivity and the role of orbital forcing. Previous studies [65] lacked the resolution necessary to directly link atmospheric CO_2 fluctuations with ice sheet feedbacks or to quantify the nonlinear responses of the climate system to astronomical forcing. In contrast, CENOGRID provides clear evidence that, during warm climate periods, low-latitude processes such as monsoon systems were the primary drivers of eccentricity-paced carbon cycling, whereas obliquity became increasingly influential as polar ice sheets expanded. The dataset also improves the chronological precision of major climate events, including the Middle Miocene Climate Transition (13.9 million years ago) and the Plio-Pleistocene intensification of glaciation.

Despite some spatial limitations—being mainly composed of Atlantic and Pacific sediment records—CENOGRID serves as a critical benchmark for calibrating and validating Earth system models, thereby enhancing predictions of climate thresholds under future elevated CO_2 scenarios. Its integration of astronomically tuned stratigraphy highlights the essential role of high-resolution data in paleoclimatic reconstructions.

Notably, a recurrence quantifier known as Determinism (DET) has already been applied to the CENOGRID dataset (Fig.6.1). DET increases when the system’s behavior becomes more deterministic or predictable. The key finding from this analysis [21] is that polar ice volume exerts a major influence on the predictability of Earth’s climatic response to astronomical forcing. This is reflected by increased fluctuations in the DET measure between approximately 50 and 35 million years ago, followed by an overall decline in predictability over subsequent stages. However, identifying and quantifying the major climate transitions embedded in this dataset remains an open and challenging problem in paleoclimatic research.

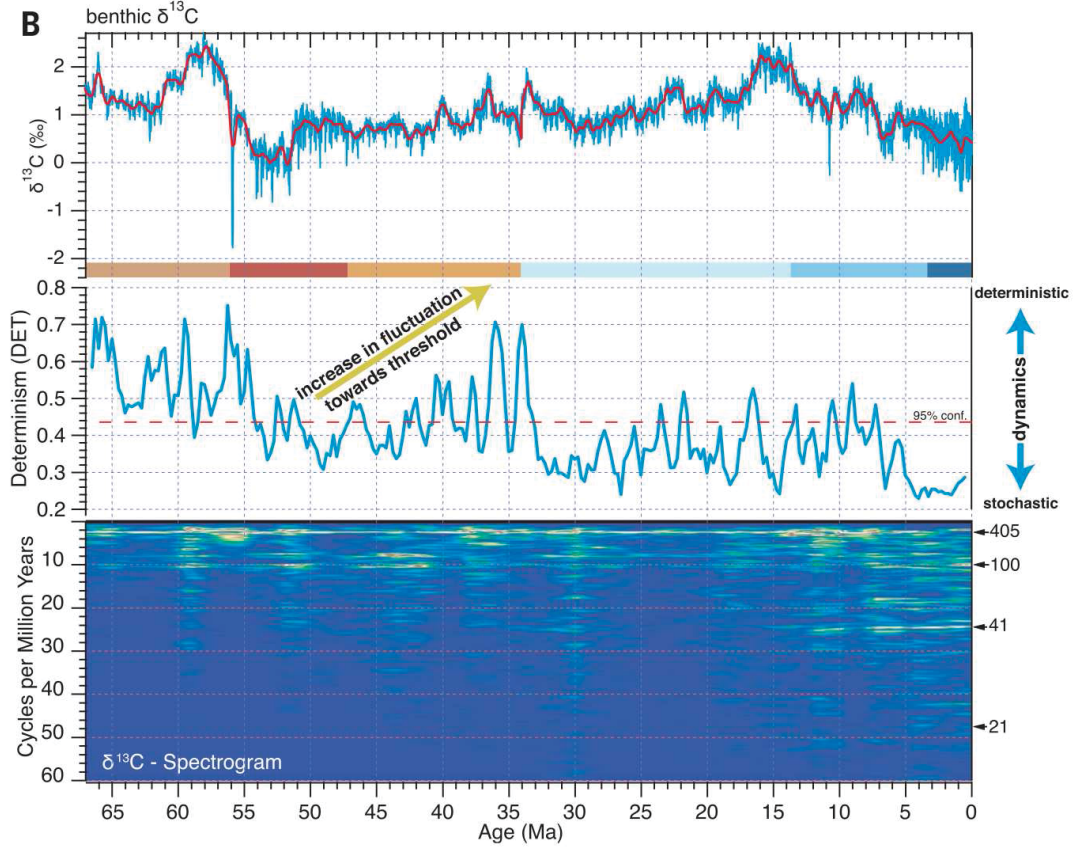


Figure 6.1: Adapted figure [21] showing the carbon data (CENOGRID), the DET measure, and the spectrogram.

6.2 Disorder in the Cenozoic dataset

The application of the disorder quantifier to the CENOGRID dataset yields critical insights into the interplay between stochastic and deterministic drivers throughout Cenozoic climate evolution. As illustrated in Fig. 6.2, the disorder quantifier exhibits temporal variations that reflect shifts in the relative contributions of deterministic forcings (e.g., orbital cycles, CO_2 trends) versus stochastic processes (e.g., regional feedbacks, unforced variability). The transition from disorder levels comparable to Red Gaussian Noise (RGN; characterized by long-term autocorrelation) in the early Cenozoic to those approaching White Gaussian Noise (WGN; indicative of uncorrelated variability) during the Neogene (Fig. 6.2) highlights a fundamental shift in the climate system's behavior. This evolution suggests a progressive decoupling of high-latitude ice sheet dynamics from low-latitude orbital forcing, as the growing influence of cryospheric feedbacks introduced new modes of variability [66].

The proposed method Ξ does not only supports the overall decrease in predictability [21], but also reveals major climate transitions, as minima in disorder often align with well-established stage divisions. This relationship between disorder/determinism and paleoclimatic stages could not be assessed using previous methods like DET and opens

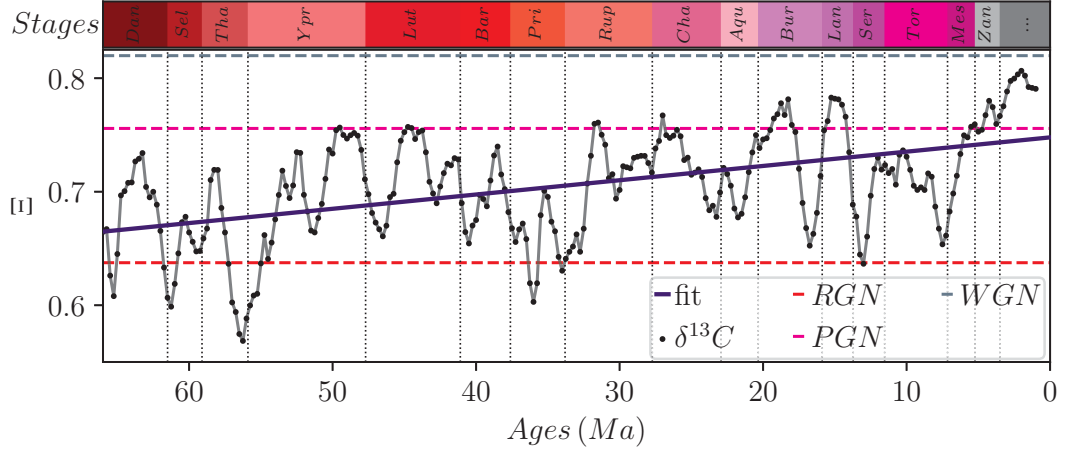


Figure 6.2: The disorder quantifier Ξ applied to benthic foraminiferal carbon isotope ($\delta^{13}\text{C}$) data as a function of geological age (in million years ago, Ma). The quantifier was computed using a sliding window of length $K = 200$ (2 Ma), with a timescale $N = 4$ (0.04 Ma). The background depicts the mean disorder levels of White Gaussian Noise (WGN), Pink Gaussian Noise (PGN), and Red Gaussian Noise (RGN), with $K = 200$. A linear fit highlights the overall trend. Vertical dotted lines denote stage boundaries.

the door to exploring additional phenomena via the CENOGRID dataset.

6.2.1 Dominant triggers and disorder minima

Stage transitions identified within the CENOGRID (Fig. 6.2) correspond to intervals where a single forcing mechanism exerted a predominant influence, suppressing stochastic variability and reducing overall disorder. For example, the Eocene-Oligocene Transition (EOT) at ~ 34 Ma—characterized by a pronounced $\delta^{13}\text{C}$ minimum and a sharp $\delta^{18}\text{O}$ increase—coincides with a marked disorder minimum (Fig. 6.2). This reflects the dominance of CO_2 -driven Antarctic glaciation [67], which effectively overwhelmed regional climatic variability. Similar behavior is observed in the Noised Logistic Map model (Fig. 5.6), where increasing the variance of a primary driver (σ_x^2) relative to background noise (σ_η^2) induces a transition from irregular (noisy) to more regular (deterministic chaos) regimes, manifesting as reduced disorder. In climatic terms, this parallels the emergence of CO_2 thresholds or tectonic reorganizations (e.g., Himalayan uplift [68]) capable of restructuring global climate dynamics.

Representative case studies include:

- Paleocene-Eocene Thermal Maximum (PETM; ~ 56 Ma): A transient disorder minimum accompanies the PETM's abrupt $\delta^{13}\text{C}$ excursion, reflecting the dominance of methane hydrate destabilization as the primary carbon source [69]. The subsequent recovery phase is marked by rising disorder, indicating the re-emergence of stochastic ocean-atmosphere interactions.

- Middle Miocene Climate Transition (mMCT; ~ 14 Ma): The mMCT is characterized by a $\delta^{18}\text{O}$ increase associated with Antarctic ice sheet expansion. Disorder decreases as obliquity-paced ice volume changes supersede eccentricity-driven carbon cycle variability [70].
- Plio-Pleistocene Glacial Intensification (~ 3.3 Ma): The expansion of Northern Hemisphere ice sheets is concomitant with a shift toward WGN-like disorder, indicating an enhanced role for stochastic ice-albedo feedbacks and a concomitant reduction in system predictability [71].

6.2.2 Theoretical and practical implications

The disorder quantifier framework offers a conceptual and methodological bridge between empirical paleoclimate data, dynamical systems, and stochastic processes. It reveals that Cenozoic climate states result from a dynamic interplay between deterministic forcings and stochastic resilience. Periods of low disorder (e.g., EOT, PETM) represent intervals wherein external perturbations (e.g., CO_2 drawdown, orbital reconfigurations) exceeded the climate system's capacity to absorb variability through internal feedbacks. Conversely, phases of elevated disorder (e.g., late Miocene) correspond to transitional regimes where multiple drivers interact nonlinearly, enhancing system complexity.

These findings underscore the importance of quantifying disorder in paleoclimate archives to better identify tipping points and refine projections of anthropogenic climate trajectories [72]. Therefore, the proposed quantifier of disorder provides a consistent way to relate data fluctuations in the Paleoclimate to transitions and behavioral changes, which was previously not possible with traditional approaches.

Conclusions and perspectives

In this work, we have embraced the ideal disorder condition—interpreted as the complete statistical independence of all past states—as the foundational principle for defining a quantifier using recurrences. By formalizing this condition, we have shown that the group of all permutations acts naturally on the space of recurrence matrices, yielding equivalence classes of equiprobable microstates. From these classes, we proposed the disorder quantifier Ξ defined as the average information entropy across all equivalence classes. By construction, Ξ increases monotonically as the data’s behavior approaches the ideal of full disorder. Crucially, we demonstrated that maximizing this entropy-based measure is not merely convenient but necessary for its optimal estimation. The resulting quantifier is thus parameter-free (apart from a chosen timescale) and provides a direct, interpretable measure of disorder in any numerical dataset.

Through extensive numerical experiments on both deterministic and stochastic benchmark models, we have validated that Ξ reliably distinguishes chaotic, temporally correlated, and uncorrelated signals even when time series are relatively short ($K < 10^4$). For autoregressive noise, variations in the memory parameter ϕ produced predictable, monotonic decreases in Ξ . When a chaotic logistic map was corrupted by noise with differing correlation strengths, Ξ correctly tracked the gradual shift from deterministic chaos toward stochastic scenarios. Comparisons with traditional measures—Permutation Entropy (PE) [26], Multivariate Permutation Entropy (MvPE) [27], Statistical Complexity (SC) [16], and Recurrence Quantification’s Determinism (DET) [25]—demonstrated that Ξ outperforms them in sensitivity, interpretability, and robustness, while requiring no user-tuned parameters.

We further applied Ξ to paleoclimate data from the Cenozoic era (the CENOGRID dataset [21]). Our analysis captured known major climate transitions—such as the Paleocene-Eocene Thermal Maximum, the Eocene-Oligocene Transition, and the Middle Miocene Climate Transition—through local minima in disorder. These results confirm that Ξ not only discriminates between deterministic and stochastic regimes in synthetic data but also meaningfully reflects Earth’s system dynamics as recorded in geological archives.

Taken together, these findings establish Ξ as a versatile tool for quantifying disorder across a wide range of complex systems. Its parameter-free nature reduces the risk of bias introduced by arbitrary choices, while its basis in maximum entropy ensures a clear, physically grounded interpretation. Because Ξ can be computed for data of any dimension with minimal preprocessing, it holds promise for disciplines where fast, reliable characterization of temporal complexity is essential.

Looking ahead, the framework introduced here can be further expanded into a unified methodology for the quantitative analysis of complex systems. Ongoing work will integrate concepts from nonlinear dynamics, statistical physics, information theory, and recurrence analysis to develop novel algorithms. These algorithms aim to characterize not only disorder but other temporal symmetries in both experimental and simulated data, thereby overcoming limitations of current approaches that rely on arbitrary parameters and often lack clear physical interpretability.

Validation of this broader methodology will proceed on three fronts. First, rigorous mathematical verification will confirm that the maximum-entropy formulations yield consistent, well-posed quantifiers across diverse system classes. Second, computational benchmarks on standard chaotic and stochastic models will test the method's accuracy, efficiency, and robustness under controlled conditions. Third, applications to real-world datasets in climatology, neuroscience, and plasma physics will demonstrate practical utility. In climatology, Ξ could refine our understanding of tipping points and improve forecasts of abrupt climate shifts. In neuroscience, it could help distinguish healthy from pathological brain dynamics. In plasma physics, it could quantify the onset of turbulence and other complex behaviors.

By unifying these elements into a coherent theoretical and computational framework, this line of research promises to advance interpretable data science within complex systems. Ultimately, it can provide a principled basis for comparing dynamic properties across domains, inform the design of experiments, and guide practical applications such as improved climate forecasting and neural signal analysis. Through these contributions, we hope to push the frontier of how disorder and complexity are quantified, interpreted, and leveraged in scientific discovery.




Appendix: Published articles

A.1 Quantifying disorder in data (*Physical Review Letters*)

This study was published in *Physical Review Letters* [63]. The abstract is as follows:

PHYSICAL REVIEW LETTERS **135**, 097401 (2025)

Quantifying Disorder in Data

João Vitor Vieira Flauzino^{1,2,3,4}  Thiago Lima Prado^{1,4}  Norbert Marwan,^{3,5,6}
Jürgen Kurths,^{2,3,*} and Sergio Roberto Lopes^{1,4,†,‡} 

¹*Department of Physics, Federal University of Paraná, 815 31-980 Curitiba, Brazil*

²*Department of Physics, Humboldt University Berlin, 12489 Berlin, Germany*

³*Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, 14473 Potsdam, Germany*

⁴*Interdisciplinary Center for Science, Technology and Innovation CICTI, Federal University of Paraná, 815 31-980 Curitiba, Brazil*

⁵*Institute for Physics and Astronomy, University of Potsdam, Potsdam 14476, Germany*

⁶*Institute for Geosciences, University of Potsdam, Potsdam 14476, Germany*

 (Received 12 November 2024; revised 11 March 2025; accepted 29 July 2025; published 26 August 2025)


The quantification of disorder in data remains a fundamental challenge in science, as many phenomena yield short length datasets with order-disorder behavior, significant (un)correlated fluctuations, and indistinguishable characteristics even when arising from distinct natures, such as chaotic or stochastic processes. In this Letter, we propose a novel method to directly quantify disorder in data through recurrence microstate analysis, showing that maximizing this measure is essential for its optimal estimation. Our approach reveals that the disorder condition corresponds to the action of the symmetric group on recurrence space, producing classes of equiprobable recurrence microstates. By leveraging information entropy, we define a robust quantifier that reliably differentiates between chaotic, correlated, and uncorrelated stochastic signals even using just small time series. Additionally, it uncovers the characteristics of corrupting noise in dynamical systems. As an application, we show that disorder minima over time often align with well-known stage transitions of the Cenozoic era, indicating periods of dominant drivers in paleoclimatic data.

DOI: 10.1103/1y98-x33s

A.2 Analytical results in calculating the entropy of recurrence microstates (*Physica A*)

This paper [73] was published in *Physica A: Statistical Mechanics and its Applications*. The abstract and some authorship details are provided below.


Physica A 638 (2024) 129629


ELSEVIER

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Physica A


journal homepage: www.elsevier.com/locate/physa



Analytical results in calculating the entropy of recurrence microstates

Felipe Eduardo Lopes da Cruz^{*}, João Vitor Vieira Flauzino, Sergio Roberto Lopes, Thiago de Lima Prado

Departamento de Física, Universidade Federal do Paraná, Curitiba, PR, 81531-980, Brazil



ARTICLE INFO

Keywords:
Recurrence analysis
Recurrence entropy
Approximation method

ABSTRACT

Since the development of recurrence plots (RP) and recurrence quantification analysis (RQA), there has been a growing interest in many areas in studying physical systems using recursion techniques. In particular, as part of the RQAs, we observed the development of the concept of recurrence microstates, defined as small blocks obtained from a recurrence graph. It can be shown that some other RQAs can be calculated as a function of recurrence microstates, and the probabilities of occurrences of these microstates can define an information entropy, the so-called entropy of recurrence microstates. It was also observed that recurrence microstates and recurrence entropy can distinguish between correlated and uncorrelated stochastic and deterministic states, due to their symmetry properties. In this paper we propose analytical expressions for calculating the entropy of recurrence microstates, avoiding the need to sample a large set of recurrence microstates. The results can be particularly important in quantifying small amounts of data, where significant sampling of microstates may not be possible. In this paper we propose analytical expressions to compute the entropy of recurrence microstates avoiding the need to sample a large set of recurrence microstates. We show that our results are accurate for cases where the probability distribution function is known. For other situations, the results can be calculated approximately. Another important fact is that the approximate results can be generalized to any size of microstate, making them a powerful tool for calculating the entropy of recurrence. Our approximate methods allow us to know what these properties are and how to exploit this quantifier in the best possible way, with minimal memory usage. Finally, we show that our analytical results are in remarkable agreement with numerical simulations.

A.3 Recurrence microstates for machine learning classification (*Chaos*)

This paper [74] was published in *Chaos: An Interdisciplinary Journal of Nonlinear Science*. The abstract and some authorship details are provided below.

Chaos

ARTICLE

pubs.aip.org/aip/cha

Recurrence microstates for machine learning classification

Cite as: Chaos 34, 073140 (2024); doi: 10.1063/5.0203801

Submitted: 16 February 2024 · Accepted: 6 July 2024 ·

Published Online: 19 July 2024



C. S. Spezzatto,¹ J. V. V. Flauzino,¹ G. Corso,² B. R. R. Boaretto,³ E. E. N. Macau,³ T. L. Prado,^{1,4,5,6} and S. R. Lopes^{1,5,6}

AFFILIATIONS

¹Department of Physics, Federal University of Paraná, 81531-980 Curitiba, Brazil²Biophysics and Pharmacology Department, Federal University of Rio Grande do Norte, 59078-900 Natal, Rio Grande do Norte, Brazil³Institute of Science and Technology, Federal University of São Paulo, 12231-280 São José dos Campos, São Paulo, Brazil⁴Department of Physics, University Rey Juan Carlos, Móstoles, 28933 Madrid, Spain⁵Interdisciplinary Center for Science, Technology and Innovation CICTI, Federal University of Paraná, 81531-980 Curitiba, Brazil⁶Potsdam Institute for Climate Impact Research—Telegraphenberg, 14473 Potsdam, Germany^aElectronic mail: thiago@fisica.ufpr.br^bAuthor to whom correspondence should be addressed: lopes@fisica.ufpr.br

ABSTRACT

Recurrence microstates are obtained from the cross recurrence of two sequences of values embedded in a time series, being the generalization of the concept of recurrence of a given state in phase space. The probability of occurrence of each microstate constitutes a recurrence quantifier. The set of probabilities of all microstates are capable of detecting even small changes in the data pattern. This creates an ideal tool for generating features in machine learning algorithms. Thanks to the sensitivity of the set of probabilities of occurrence of microstates, it can be used to feed a deep neural network, namely, a microstate multi-layer perceptron (MMLP) to classify parameters of chaotic systems. Additionally, we show that with more microstates, the accuracy of the MMLP increases, showing that the increasing size and number of microstates insert new and independent information into the analysis. We also explore potential applications of the proposed method when adapted to different contexts.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0203801>

Bibliography

- [1] T. Shinbrot and F. J. Muzzio, “Noise to order,” *Nature*, vol. 410, pp. 251–258, Mar 2001.
- [2] D. Sornette, *Critical phenomena in natural sciences: chaos, fractals, selforganization and disorder: concepts and tools*. Springer Science & Business Media, 2006.
- [3] S. Kos and P. Littlewood, “Hear the noise,” *Nature*, vol. 431, pp. 29–29, Sep 2004.
- [4] S. A. Crooker, D. G. Rickel, A. V. Balatsky, and D. L. Smith, “Spectroscopy of spontaneous spin noise as a probe of spin dynamics and magnetic resonance,” *Nature*, vol. 431, pp. 49–52, Sep 2004.
- [5] E. C. Barrows, “Stochasticity in nature, and its consequences,” in *Maximum Entropy and Bayesian Methods: Laramie, Wyoming, 1990*, pp. 191–197, Springer, 1991.
- [6] N. Marwan and J. Kurths, “Comment on “stochastic analysis of recurrence plots with applications to the detection of deterministic signals” by Rohde et al. [Physica D 237 (2008) 619–629],” *Physica D*, vol. 238, no. 16, pp. 1711–1715, 2009.
- [7] C. Letellier, *Chaos in Nature*. World Scientific, 2013.
- [8] T. N. Palmer, “Stochastic weather and climate models,” *Nature Reviews Physics*, vol. 1, pp. 463–471, Jul 2019.
- [9] C. E. Shannon, “A mathematical theory of communication,” *Bell System Technical Journal*, vol. 27, no. 623, p. 218, 1948.
- [10] C. Bandt and B. Pompe, “Permutation entropy: a natural complexity measure for time series,” *Physical Review Letters*, vol. 88, no. 17, p. 174102, 2002.
- [11] G. Corso, T. L. Prado, G. Z. d. S. Lima, J. Kurths, and S. R. Lopes, “Quantifying entropy using recurrence matrix microstates,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 28, no. 8, p. 083108, 2018.

- [12] X. Xu, J. Zhang, and M. Small, “Superfamily phenomena and motifs of networks induced from time series,” *Proceedings of the National Academy of Sciences*, vol. 105, no. 50, pp. 19601–19605, 2008.
- [13] Y. Hirata, “Recurrence plots for characterizing random dynamical systems,” *Communications in Nonlinear Science and Numerical Simulation*, vol. 94, p. 105552, 2021.
- [14] T. L. Prado, B. R. R. Boaretto, G. Corso, G. Z. dos Santos Lima, J. Kurths, and S. R. Lopes, “A direct method to detect deterministic and stochastic properties of data,” *New Journal of Physics*, vol. 24, no. 3, p. 033027, 2022.
- [15] S. M. Pincus, “Approximate entropy as a measure of system complexity,” *Proceedings of the National Academy of Sciences*, vol. 88, no. 6, pp. 2297–2301, 1991.
- [16] O. A. Rosso, H. Larrondo, M. T. Martin, A. Plastino, and M. A. Fuentes, “Distinguishing noise from chaos,” *Physical Review Letters*, vol. 99, no. 15, p. 154102, 2007.
- [17] M. B. Kennel and S. Isabelle, “Method to distinguish possible chaos from colored noise and to determine embedding parameters,” *Physical Review A*, vol. 46, no. 6, p. 3111, 1992.
- [18] D. T. Kaplan and L. Glass, “Direct test for determinism in a time series,” *Physical Review Letters*, vol. 68, pp. 427–430, Jan 1992.
- [19] D. Toker, F. T. Sommer, and M. D’Esposito, “A simple method for detecting chaos in nature,” *Communications Biology*, vol. 3, p. 11, Jan 2020.
- [20] S.-F. Liang, H.-C. Wang, and W.-L. Chang, “Combination of eeg complexity and spectral analysis for epilepsy diagnosis and seizure detection,” *EURASIP Journal on Advances in Signal Processing*, vol. 2010, pp. 1–15, 2010.
- [21] T. Westerhold, N. Marwan, A. J. Drury, D. Liebrand, C. Agnini, E. Anagnostou, J. S. Barnet, S. M. Bohaty, D. De Vleeschouwer, F. Florindo, *et al.*, “An astronomically dated record of earth’s climate and its predictability over the last 66 million years,” *Science*, vol. 369, no. 6509, pp. 1383–1387, 2020.
- [22] P. Grassberger and H. Kantz, “Generating partitions for the dissipative h  non map,” *Physics Letters A*, vol. 113, no. 5, pp. 235–238, 1985.
- [23] F. Christiansen and A. Politi, “Guidelines for the construction of a generating partition in the standard map,” *Physica D: Nonlinear Phenomena*, vol. 109, no. 1, pp. 32–41, 1997. Proceedings of the Workshop on Physics and Dynamics between Chaos, Order, and Noise.

- [24] N. Marwan and K. H. Kraemer, “Trends in recurrence analysis of dynamical systems,” *The European Physical Journal Special Topics*, vol. 232, pp. 5–27, Feb 2023.
- [25] N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, “Recurrence plots for the analysis of complex systems,” *Physics Reports*, vol. 438, no. 5-6, pp. 237–329, 2007.
- [26] C. Bandt and B. Pompe, “Permutation entropy: a natural complexity measure for time series,” *Physical Review Letters*, vol. 88, no. 17, p. 174102, 2002.
- [27] S. He, K. Sun, and H. Wang, “Multivariate permutation entropy and its application for complexity analysis of chaotic systems,” *Physica A: Statistical Mechanics and its Applications*, vol. 461, pp. 812–823, 2016.
- [28] C. Letellier, I. Leyva, and I. Sendiña-Nadal, “Dynamical complexity measure to distinguish organized from disorganized dynamics,” *Physical Review E*, vol. 101, no. 2, p. 022204, 2020.
- [29] G. Datseris and K. A. Haaga, “Complexitymeasures.jl: scalable software to unify and accelerate entropy and complexity timeseries analysis,” *arXiv preprint arXiv:2406.05011*, 2024.
- [30] G. Datseris, “Dynamicalsystems.jl: A Julia software library for chaos and nonlinear dynamics,” *Journal of Open Source Software*, vol. 3, no. 23, p. 598, 2018.
- [31] S. Schinkel, O. Dimigen, and N. Marwan, “Selection of recurrence threshold for signal detection,” *The European Physical Journal Special Topics*, vol. 164, no. 1, pp. 45–53, 2008.
- [32] I. Andreadis, A. D. Fragkou, and T. E. Karakasidis, “On a topological criterion to select a recurrence threshold,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 30, no. 1, 2020.
- [33] R. Delage and T. Nakata, “A variable threshold for recurrence based on local attractor density,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 32, no. 9, 2022.
- [34] M. Thiel, M. C. Romano, J. Kurths, R. Meucci, E. Allaria, and F. T. Arecchi, “Influence of observational noise on the recurrence quantification analysis,” *Physica D: Nonlinear Phenomena*, vol. 171, no. 3, pp. 138–152, 2002.
- [35] T. Prado, G. Corso, G. dos Santos Lima, R. Budzinski, B. Boaretto, F. Ferrari, E. Macau, and S. Lopes, “Maximum entropy principle in recurrence plot analysis on stochastic and chaotic systems,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 30, no. 4, 2020.

- [36] R. M. May, “Simple mathematical models with very complicated dynamics,” *Nature*, vol. 261, no. 5560, pp. 459–467, 1976.
- [37] K. T. Alligood, T. D. Sauer, J. A. Yorke, and D. Chillingworth, “Chaos: an introduction to dynamical systems,” *SIAM Review*, vol. 40, no. 3, pp. 732–732, 1998.
- [38] C. S. Hsu, *Cell-to-cell mapping: a method of global analysis for nonlinear systems*, vol. 64. Springer Science & Business Media, 2013.
- [39] K. Ikeda, “Multiple-valued stationary state and its instability of the transmitted light by a ring cavity system,” *Optics communications*, vol. 30, no. 2, pp. 257–261, 1979.
- [40] O. E. Rossler, “An equation for hyperchaos,” *Physics Letters A*, vol. 71, no. 2-3, pp. 155–157, 1979.
- [41] E. Ott, *Chaos in dynamical systems*. Cambridge University Press, 2002.
- [42] B. V. Chirikov, “A universal instability of many-dimensional oscillator systems,” *Physics reports*, vol. 52, no. 5, pp. 263–379, 1979.
- [43] C. Coulliette and S. Wiggins, “Intergyre transport in a wind-driven, quasigeostrophic double gyre: An application of lobe dynamics,” *Nonlinear Processes in Geophysics*, vol. 7, no. 1/2, pp. 59–85, 2000.
- [44] O. E. Rössler, “An equation for continuous chaos,” *Physics Letters A*, vol. 57, no. 5, pp. 397–398, 1976.
- [45] M. C. Mackey and L. Glass, “Oscillation and chaos in physiological control systems,” *Science*, vol. 197, no. 4300, pp. 287–289, 1977.
- [46] E. N. Lorenz, “Deterministic nonperiodic flow 1,” in *Universality in Chaos, 2nd edition*, pp. 367–378, Routledge, 2017.
- [47] T. Matsumoto, “A chaotic attractor from Chua’s circuit,” *IEEE transactions on circuits and systems*, vol. 31, no. 12, pp. 1055–1058, 2003.
- [48] M. Chitre, “Signalanalysis.jl,” 2024. GitHub repository. Available at: <https://github.com/org-ar1/SignalAnalysis.jl> [Accessed: 2023-09-16].
- [49] J. Sprott, “Chaos and time-series analysis,” Oxford University Press, 2003.
- [50] G. E. P. Box and G. M. Jenkins, “Time series analysis: Forecasting and control,” Prentice Hall PTR, 1994.
- [51] F. Hooge, T. Kleinpenning, and L. K. Vandamme, “Experimental studies on 1/f noise,” *Reports on Progress in Physics*, vol. 44, no. 5, p. 479, 1981.

- [52] U. S. Freitas, C. Letellier, and L. A. Aguirre, “Failure in distinguishing colored noise from chaos using the “noise titration” technique,” *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, vol. 79, no. 3, p. 035201, 2009.
- [53] D. Toker, F. T. Sommer, and M. D’Esposito, “A simple method for detecting chaos in nature,” *Communications Biology*, vol. 3, no. 1, p. 11, 2020.
- [54] C. Schölzel, “Nolds documentation,” 2020. Available at: https://nolds.readthedocs.io/_/downloads/en/0.5.2/pdf/.
- [55] F. Molz, H. Liu, and J. Szulga, “Fractional Brownian motion and fractional Gaussian noise in subsurface hydrology: A review, presentation of fundamental properties, and extensions,” *Water Resources Research*, vol. 33, no. 10, pp. 2273–2286, 1997.
- [56] R. Zwanzig, “Brownian motion and Langevin equations,” *Nonequilibrium Statistical Mechanics*, vol. 3, 2001.
- [57] B. B. Mandelbrot and J. W. Van Ness, “Fractional Brownian motions, fractional noises and applications,” *SIAM review*, vol. 10, no. 4, pp. 422–437, 1968.
- [58] B. E. Sagan, *The symmetric group: representations, combinatorial algorithms, and symmetric functions*, vol. 203. Springer Science & Business Media, 2013.
- [59] A. Politi, “Quantifying the dynamical complexity of chaotic time series,” *Physical Review Letters*, vol. 118, no. 14, p. 144101, 2017.
- [60] T. L. Prado, G. Z. dos Santos Lima, B. Lobão Soares, G. C. do Nascimento, G. Corso, J. Fontenele-Araujo, J. Kurths, and S. R. Lopes, “Optimizing the detection of nonstationary signals by using recurrence analysis,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 28, no. 8, p. 085703, 2018.
- [61] N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, “Recurrence plots for the analysis of complex systems,” *Physics Reports*, vol. 438, no. 5, pp. 237–329, 2007.
- [62] J. H. Zar, *Biostatistical Analysis*. Prentice-Hall, Inc., 2007.
- [63] J. V. V. Flauzino, T. L. Prado, N. Marwan, J. Kurths, and S. R. Lopes, “Quantifying disorder in data,” *Physical Review Letters*, vol. 135, p. 097401, Aug 2025.
- [64] F. M. Gradstein, J. G. Ogg, M. B. Schmitz, and G. M. Ogg, *Geologic Time Scale 2020: Volume 2*. Elsevier, 2020.
- [65] J. Zachos, M. Pagani, L. Sloan, E. Thomas, and K. Billups, “Trends, rhythms, and aberrations in global climate 65 ma to present,” *Science*, vol. 292, no. 5517, pp. 686–693, 2001.

- [66] M. E. Raymo, W. Ruddiman, N. Shackleton, and D. Oppo, “Evolution of atlantic-pacific $\delta^{13}\text{C}$ gradients over the last 2.5 my,” *Earth and Planetary Science Letters*, vol. 97, no. 3-4, pp. 353–368, 1990.
- [67] P. N. Pearson, G. L. Foster, and B. S. Wade, “Atmospheric carbon dioxide through the eocene–oligocene climate transition,” *Nature*, vol. 461, no. 7267, pp. 1110–1113, 2009.
- [68] R. Fortey, *Life: A Natural History of the First Four Billion Years of Life on Earth*. Vintage, 1999.
- [69] G. R. Dickens, J. R. O’Neil, D. K. Rea, and R. M. Owen, “Dissociation of oceanic methane hydrate as a cause of the carbon isotope excursion at the end of the paleocene,” *Paleoceanography*, vol. 10, no. 6, pp. 965–971, 1995.
- [70] A. Holbourn, W. Kuhnt, M. Schulz, and H. Erlenkeuser, “Impacts of orbital forcing and atmospheric carbon dioxide on miocene ice-sheet expansion,” *Nature*, vol. 438, no. 7067, pp. 483–487, 2005.
- [71] G. Roe, “In defense of milankovitch,” *Geophysical Research Letters*, vol. 33, no. 24, 2006.
- [72] T. M. Lenton, H. Held, E. Kriegler, J. W. Hall, W. Lucht, S. Rahmstorf, and H. J. Schellnhuber, “Tipping elements in the earth’s climate system,” *Proceedings of the National Academy of Sciences*, vol. 105, no. 6, pp. 1786–1793, 2008.
- [73] F. E. L. da Cruz, J. V. V. Flauzino, S. R. Lopes, and T. de Lima Prado, “Analytical results in calculating the entropy of recurrence microstates,” *Physica A: Statistical Mechanics and its Applications*, vol. 638, p. 129629, 2024.
- [74] G. S. Spezzatto, J. V. V. Flauzino, G. Corso, B. R. R. Boaretto, E. E. N. Macau, T. L. Prado, and S. R. Lopes, “Recurrence microstates for machine learning classification,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 34, no. 7, 2024.