Recurrence plots for analysing highly discrete data

Norbert Marwan
Research Deptartment of Complexity Science
Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association
Potsdam, Germany
marwan@pik-potsdam.de
0000-0003-1437-7039

***Abstract*—Discrete event time series pose unique challenges for dynamical systems analysis due to irregular sampling and sparsity. The combination of recurrence plots with edit distance metrics is demonstrated to extend recurrence analysis to highly discretised data, enabling characterisation of event dynamics and detection of phase synchronisation between heterogeneous systems. Applying the method to a coupled Rössler-Lorenz system confirms its validity and highlights its potential for analyzing complex systems. Key challenges in choosing parameters and comparing heterogeneous data are discussed, along with potential areas for future methodological improvements.**

***Index Terms*—recurrence plots, event time series, edit distance, phase synchronization, recurrence quantification analysis, nonlinear time series analysis, dynamical systems, irregular sampling**

I. Introduction

Highly discrete data, such as event time series or extreme events data, play a central role in data science because they encode the temporal organisation of processes generated by underlying dynamical systems. Such data arise across a wide range of disciplines, including finance and medicine (e.g., extreme price movements, ventricular tachyarrhythmias, epileptic seizures), engineering (structural damage events, power grid instabilities), and Earth sciences (geophysical hazards, extreme precipitation, or landslides). This broad applicability makes event time series a fundamental object of analysis and modelling.

Event series may consist of single, discrete or binary events, or events with varying amplitudes, such as extremes, anomalies, or events derived from heavy-tailed distributions. An event series is primarily defined by the set of event time points $\{t_i\}$, which are usually irregularly sampled ($t_{i+1}-t_i \neq$ const.), in contrast to standard time series, which consist of time points and associated amplitude values $\{t_i, x_i\}$. This irregular sampling in event series, together with their often sparse nature (i.e., many zeros between events), poses a significant challenge for conventional time series analysis methods – such as Fourier analysis – which are designed for equidistantly sampled data. Research questions for event data are often similar to those for continuous time series, including data comparison, classification of underlying dynamical behaviour, detection of regime changes, and use in simulations or predictions.

While several approaches address event-data analysis, including probabilistic methods, point process models, and pattern-based prediction algorithms [1]–[5], their application is less straightforward and often limited when the goal is to effectively compare these data with continuous records or reliably characterise their underlying spectral properties and nonlinear dynamics [6]. We address these challenges using recurrence plots adapted with edit distance metrics, which provide a promising alternative [6], [7].

II. Recurrence Plots and the Edit Distance Metric

A *recurrence plot* (RP) is the graphical representation of the recurrence matrix, which visualises pairwise similarity between all states $\vec{x}_i$ and $\vec{x}_j$:

$$R_{i,j} = \Theta\left(\varepsilon - d(\vec{x_i}, \vec{x_j})\right), \tag{1}$$

where $d(\cdot,\cdot)$ is a distance- or cost-based similarity measure and $\Theta(\cdot)$ the Heaviside function, yielding $R_{i,j} = 1$ if the similarity (e.g., defined by a spatial distance measured by the Euclidean norm $d_{L2} = \|\vec{x}_i - \vec{x}_j\|$) falls below a threshold ε [7]. The RP provides a visual fingerprint of a dynamical system's behaviour: its pattern reflects specific dynamics, and the small-scale structures reveal information about predictability and the system's time-scales [7].

For event time series, which consist only of event times $\{t_i\}$ or have strongly non-equidistant sampling, the standard Euclidean norm (or any other metric) is not applicable, because the data is sparse and irregularly spaced. The similarity measure $d(\cdot,\cdot)$ in Eq. (1) must be replaced by a specific metric that measures the coincidence of event sequences.

A powerful choice for adapting RPs to discrete and event data is the *edit distance*, an extension of the Levenshtein distance [8], [9]. The event time series must first be segmented into sequences S_i within a time window of length T_w. The edit distance d_{ED} between two event sequences $S_i = \{t_k(i)\}_{k=1}^{N_i}$ and $S_j = \{t_k(j)\}_{k=1}^{N_j}$ is defined as the minimal total cost required to transform S_i into S_j using deletions, insertions, and temporal shifts of events (Fig. 1):

$$d_{ED}(S_i, S_j) = \min\Big\{\lambda_s(N_i + N_j - 2N_{(i,j)}) + \sum_{a,b\in\mathcal{C}} \lambda_0 \|t_a^{(i)} - t_b^{(j)}\|\Big\}, \tag{2}$$

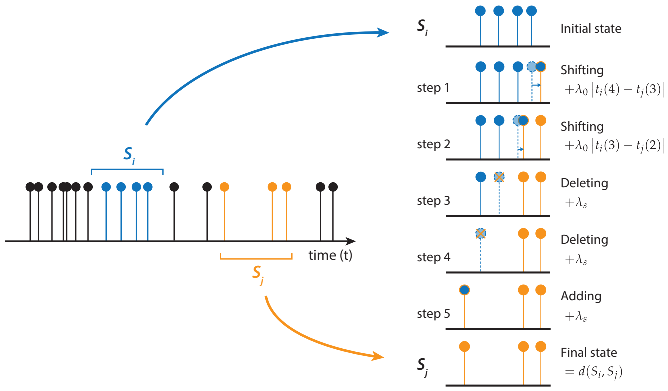

Fig. 1. Edit distance as cost-based similarity between event sequences S_i and S_j from an event series (left). Events can be shifted, added or deleted, and their amplitude adjusted. All these operations have costs. The minimum cost is used as the distance (right).

where λ_s is the cost of deletion/insertion, and λ_0 is the cost for shifting events in time. The cost λ_s can be set to 1, reducing the number of free parameters to just one. For the remaining parameter λ_0, a reasonable choice is the number of events normalised by the total duration of the considered event series (i.e., $t(\text{end}) - t(1)$) [10]. N_i and N_j are the number of events in segments S_i and S_j, respectively, and $N(i,j)$ the number of events in S_i and S_j to be shifted, which form the set $\mathcal{C}$, with a and b the indices of the events in these sequences S_i and S_j. The sequences S_i and S_j are extracted from the series of events by using overlapping intervals of length T_w which are subsequently moved by a step of s_w. While formal guidelines for choosing T_w are not yet available, it should be sufficiently large to capture multiple events, typically $5\dots10$ times the mean inter-event interval. In the specific application of edit-distance-based RPs, it can be found that smaller step sizes, e.g., $s_w = 2\dots3$, often yield more robust results. While larger steps reduce overlap and increase independence between windows, a very small s_w can help capture subtle temporal variations in the event sequences.

In essence, d_{ED} measures the minimum cost to align two event sequences starting at time points i and j through the aforementioned operations. Considering all pairwise combinations of i and j, this specific metric d_{ED} can be naturally integrated into recurrence analysis [9]–[13], allowing RPs to be computed directly from event data.

An extension of the edit distance can also take amplitude variations into account [9], by adding an additional term

$$\lambda_k \left\| \vec{x}_a^{(i)} - \vec{x}_b^{(j)} \right\| \quad \text{for} \quad (a,b) \in \mathcal{C}$$

to the cost function, Eq. (2), with $\vec{x}_a^{(i)}$ the values of events at time a in sequence i (and analogously for $\vec{x}_b^{(j)}$), which can also be multi-dimensional. However, this introduces a further complexity, as it mixes two different aspects of the data: the temporal pattern of event sequences and amplitude differences. The optimal choice of the corresponding parameters λ_0 and λ_k might be less clear then, but have to be used to balance between these aspects.

III. Applications and Potential of Recurrence Analysis

The integration of the edit distance metric into the RP definition extends the powerful tools of nonlinear time series analysis to discrete and event-based processes.

A. Unveiling Synchronization Patterns

One important application is the direct comparison and measurement of interrelationships between heterogeneous time series – specifically, event data and continuous time series. Even when the underlying dynamics remain chaotic and their amplitudes uncorrelated, these systems can adjust their timing in a coordinated manner, an important phenomenon known as phase synchronisation [14].

The RP framework allows us to directly compare the phases of dynamical systems, even when their dynamics are not coherent or, as for highly discrete data, not easy to estimate. This can be done by comparing the diagonal-line-based recurrence rate (τ-*recurrence rate*), which represents the probability that states recur after some time τ [15], [16],

$$RR_\tau = \frac{1}{N-\tau} \sum_{i=1}^{N-\tau} R_{i,i+\tau}. \tag{3}$$

Here, τ is the distance of a diagonal from the main diagonal in the RP. Comparing $RR_\tau^{(x)}$ of system x with $RR_\tau^{(y)}$ of system

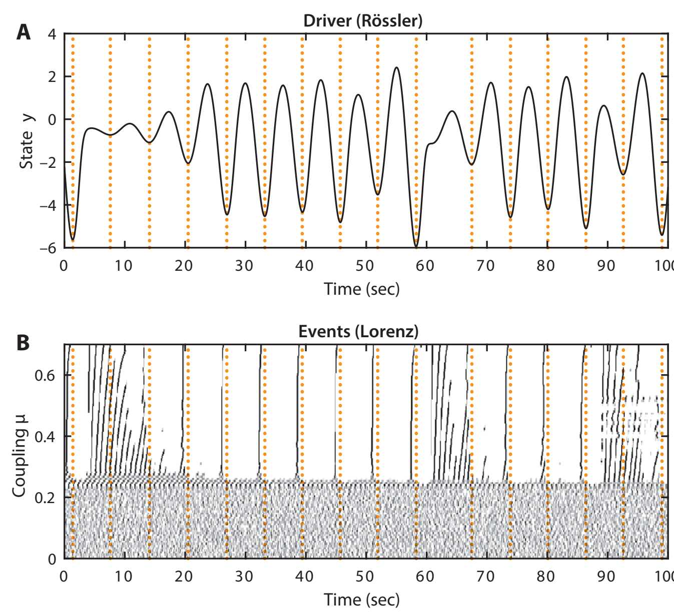

Fig. 2. Event data from a coupled Rössler-Lorenz model, where the Rössler system (A) drives the Lorenz system. The maxima of the first component of the Lorenz system serve as events (B, a black point marks an event). As the coupling parameter μ increases, phase synchronisation begins to emerge between the two systems, indicated by phase-locking of the events – visible by the accumulation of dots in lines for larger values of μ_0 and their constant distance to the dotted lines (sampling time $\Delta t = 0.1$).

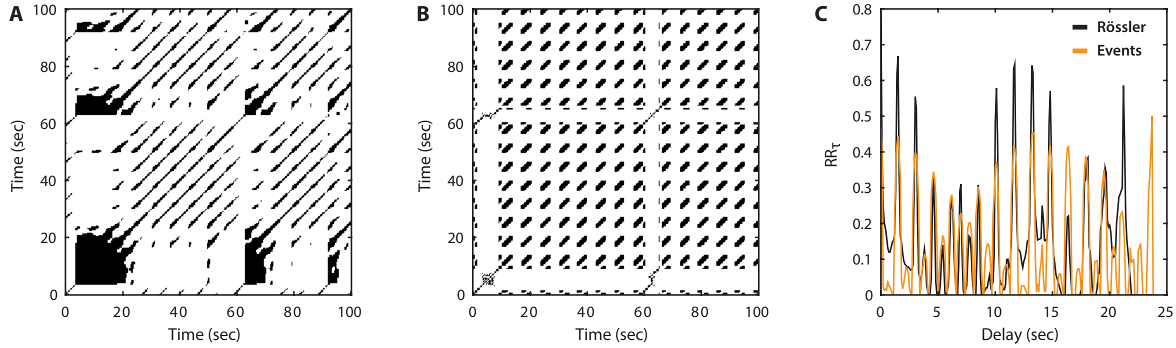

Fig. 3. Recurrence plots and synchronisation measure CPR for the coupled Rössler-Lorenz system. (A) RP of the (continuous) first component of the Rössler system, (B) edit-distance-based RP of the event series extracted from the Lorenz system for a coupling $\mu = 0.5$, using a sequence length of $T_w = 3.5$ sec and a step size of $s_w = 0.4$ sec. The RPs show some similarity in the vertical distance between the diagonal line structures and in the disruptions (vertical white bands) around 5 sec and 60 sec. (C) τ-recurrence rate RR_τ obtained from the RPs of the continuous Rössler time series and the event series (as shown in panels A and B).

y, phase synchronisation can be detected. The standard way to do this is to use Pearson correlation [15], [17], but other measures are also suitable, such as Spearman rank correlation, Kullback–Leibler or Hellinger distance [16]. This results in the *correlation of probability of recurrences* (CPR),

$$CPR = \varrho\left(RR^{(x)}_{\tau>\tau_{\min}}, RR^{(y)}_{\tau>\tau_{\min}}\right), \tag{4}$$

with $\varrho(\cdot,\cdot)$ a suitable correlation measure between the RR_τ series, excluding the initial part up to $\tau_{\min}$ to avoid trivial correlations [16], [17].

The potential of this approach is demonstrated here by an unidirectionally coupled Rössler-Lorenz system,

$$\begin{aligned} \dot{x}_1 &= x_2 + a x_1, & \dot{y}_1 &= -\sigma(y_1 - y_2), \\ \dot{x}_2 &= -x_3 - x_1, & \dot{y}_2 &= u(\rho - y_3) - y_2, \\ \dot{x}_3 &= b + x_3(x_2 - c), & \dot{y}_3 &= u y_2 - \beta y_3 \end{aligned}$$

with x the Rössler and y the Lorenz system, the coupling

$$u = \mu(x_1 + x_2 + x_3) + (1 - \mu) y_1,$$

μ the coupling coefficient, and the parameters $a = 0.45$, $b = 2.0$, $c = 4.0$, $\sigma = 10.0$, $\rho = 28.0$, $\beta = 2.0$. The Rössler system x drives the Lorenz system y. To obtain event data, the local maxima of the Lorenz system are considered to be events (Fig. 2). For weak coupling, phase synchronisation can be found, leading to a phase locking between the Rössler system and the events extracted from the Lorenz system.

The RPs calculated from the continuous Rössler system and the event series exhibit some similarities (Fig. 3). Both feature diagonal line structures, which indicate similar state evolutions at different times. The vertical distances between these line structures represent return times [7] and appear to be similar in both RPs. Furthermore, vertical white bands interrupt the more uniform structures in the RPs. These disruptions generally align in both RPs around 5 sec and 60 sec.

The τ-recurrence rate curves offer a quantitative way to compare return times, reflecting the average phase distribution. The curves for the Rössler and the event series align closely, indicating phase synchronisation (Fig. 3C).

The CPR values for increasing coupling show an abrupt jump shortly after the onset of phase synchronisation (Fig. 4B), as indicated by a change of the third-largest Lyapunov exponent from zero to negative values (Fig. 4A). To assess the significance of CPR values, it is recommended to design a surrogate-based hypothesis test. A simple surrogate data ensemble can be constructed by shuffling the events (i.e., assigning random time points to the available events), thereby ensuring a null-model of no synchronisation between the event series and the driver. The results show a clearly significant change in the CPR values (Fig. 4B), confirming the method's effectiveness in detecting phase synchronisation even in heterogeneous data.

An alternative approach using RPs is to test for simultaneous recurrences in both signals. This provides a powerful indicator of generalised synchronisation [18]. By combining the event-based edit distance metric for system x with the standard distance metric for continuous system y, the JRP is

$$JR^{x,y}_{i,j} = \Theta\left(\varepsilon - d_{L2}(\vec{x}_i, \vec{x}_j)\right) \cdot d_{ED}(S^y_i, S^y_j), \tag{5}$$

where $\vec{x}_i$ denotes the continuous time series (or phase space vectors) and y_i the event series. This definition requires both matrices, $\Theta\left(\varepsilon - d_{L2}(\vec{x}_i, \vec{x}_j)\right)$ and $d_{ED}(S^y_i, S^y_j)$, to have matching sizes, which poses a challenge in optimally selecting the parameters T_w and s_w that needs to be addressed systematically in future studies.

These capabilities are essential for analysing coupled dynamical systems where, for example, climate indices (continuous data, x) are compared with the occurrence of extreme events (discrete events, y), such as atmospheric rivers or flush floods. This approach is promising for future extensions, including detecting coupling directions or even causal links between systems [19], [20].

B. Quantitative Characterisation of Event Dynamics

The structural properties of RPs can be quantified using recurrence quantification analysis (RQA) measures. These

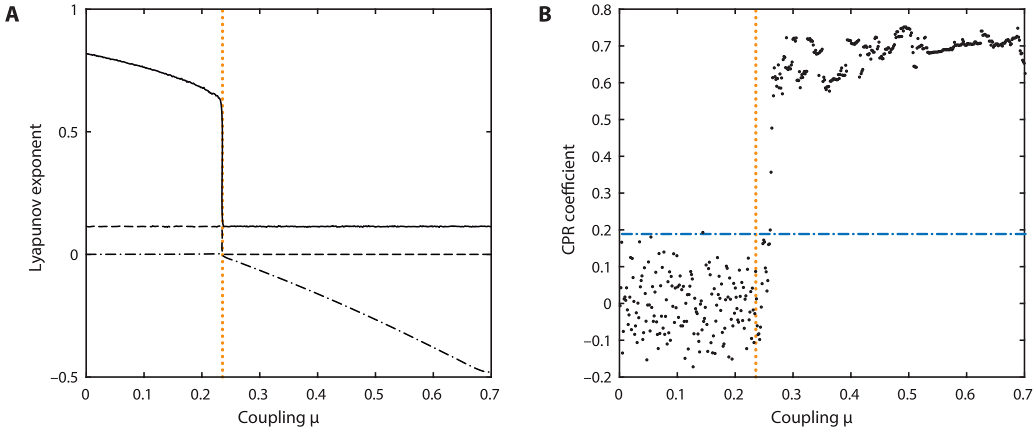

Fig. 4. (A) First three largest Lyapunov exponents of the Rössler-Lorenz system, indicating the onset of phase synchronisation at a critical value of the coupling coefficient $\mu_c = 0.236$ (dotted line). (B) CPR measure calculated using the RPs of the continuous Rössler system and the event series (CPR based on Pearson correlation and $\tau_{\min=1}$ sec). The dotted line indicates the onset of the phase synchronisation as found by the Lyapunov exponents, and the dash-dotted line represents the 95% significance level (based on 50 surrogate realisations of a null-model, ensuring no synchronisation).

measures are key to classifying data and identifying regime transitions.

a) Serial Dependence (Determinism): Diagonal lines in an RP represent the tendency for system states to remain close neighbours in phase space over time, indicating serial dependence or deterministic evolution. The RQA measure *determinism* quantifies the fraction of recurrence points forming these diagonal lines, allowing the study of temporal dependencies within event series or irregularly sampled time series [10], a common challenge, e.g., in astrophysics and palaeoclimate research. High determinism indicates that similar event patterns repeat reliably, suggesting predictable underlying dynamics. For sparse event data where traditional autocorrelation fails, this offers a natural alternative [13].

b) Spectral Estimation: RPs can also be used to estimate the power spectral density of event series [21]. The τ-recurrence rate, which is the density of recurrence points along the diagonals, is related to the auto-correlation function. By calculating this probability of recurrence using the edit distance, the power spectrum can be estimated via Fourier transform or other decomposition methods, thus, providing a spectral view of spiky signals where traditional methods fail [21], [22].

IV. Challenges and Future Perspectives

Despite the advances, several challenges remain when applying recurrence analysis to event time series.

A. Parameter Selection and Sensitivity

An important challenge for future work is to systematically assess the influence not only of the edit-distance parameters themselves, but also of the sequence length T_w and the degree of overlap between sequences, controlled by the step size s_w. Both parameters can substantially affect the structure and interpretation of edit-distance-based RPs. Moreover, when extending edit distance to incorporate amplitude information, a key open issue is the development of robust strategies for parameter selection that appropriately balance the editing of event occurrences against the measurement of amplitude differences.

B. Data Sparsity and Non-Stationarity

Event series often suffer from missing or sparse data. Non-stationary processes, where the statistical properties (e.g., event distribution) change over time, can lead to periods without events. This can result in non-defined costs $d(S_i, S_j)$ if the chosen time window T_w is too small. The optimal selection of the window length T_w and the development of correction and gap-filling schemes are crucial future research directions [6].

C. Handling Heterogeneous Data

Comparing RPs derived from sparse event data with RPs from continuous data remains challenging for the application of the JRP approach, due to the potentially large difference in sequence length. Approaches are needed to match the sizes of event-based RPs with those of continuous data, potentially involving coarse-graining, interpolation, or specific window selection schemes.

D. Methodological Extensions

Future methodical developments should focus on:

a) Alternative Event Metrics: Testing and comparing other similarity measures like event synchronisation, event coincidence analysis, or Needleman-Wunsch distance, which might be better suited for specific event characteristics [2]. Additionally, exploring connections to point process frameworks – such as linking recurrence-based measures to likelihood estimators – could provide a statistically rigorous foundation for event series analysis and open further bridges between dynamical systems and probabilistic modelling approaches.

b) Timing Jitter: Real-world event series often have uncertainty in timing (jitter). Future concepts should integrate this, perhaps through Monte Carlo or Bayesian approaches, leading to recurrence matrices that represent the probabilities of recurrences rather than binary information [23].

c) Machine Learning Integration: RPs of event series can be converted into images and used as input for machine learning workflows (e.g., classification tasks), leveraging the geometrical structures unveiled by the recurrence method [24].

E. New Perspectives in Studying Nonlinear Systems

In chaotic dynamical systems, phase locking can be interrupted by phase slips – brief desynchronisations where the locked phase relationship temporarily breaks down before re-establishing. While such phenomena have been extensively studied in continuous-time systems [14], their manifestation and significance in discrete event time series remain unexplored. The irregular spacing and sparsity of event data may produce unique phase slip signatures distinct from continuous-system counterparts. Recurrence-based methods are ideally positioned to detect and characterise these transient desynchronisation events, offering new perspectives on the stability and robustness of coupled event-generating processes.

V. Conclusion

In conclusion, recurrence analysis is a relatively novel yet rapidly developing framework, with numerous powerful extensions emerging over the past two decades [25]. Particularly when using tailored metrics such as the edit distance, it provides a robust and versatile approach for characterising the complex dynamics of highly discretised event time series, especially in regimes where alternative methods reach their limits. Addressing the remaining challenges related to parameter selection, data sparsity, and the comparison of heterogeneous data will further unlock its potential across a wide range of scientific disciplines.

Acknowledgment

I thank Deniz Eroglu, Abhirup Banerjee, K. Hauke Krämer, Tobias Braun, Lütje Lange, and Franziska Niederstedt for fruitful discussions. The research was partly supported by the DAAD grant 57705568: "Recurrence quantifiers as features for machine-learning-decision-making processes".

References

[1] J. Beirlant, Y. Goegebeur, J. Segers, J. L. Teugels, D. De Waal, and C. Ferro, *Statistics of Extremes: Theory and Applications*. Chichester, UK: Wiley, 2006.

[2] M. Ciba, R. Bestel, C. Nick, G. F. de Arruda, T. Peron, C. C. Henrique, L. d. F. Costa, F. A. Rodrigues, and C. Thielemann, "Comparison of Different Spike Train Synchrony Measures Regarding Their Robustness to Erroneous Data From Bicuculline-Induced Epileptiform Activity," *Neur. Comp.*, vol. 32, no. 5, pp. 887–911, may 2020.

[3] M. Ghil, P. Yiou, S. Hallegatte, B. D. Malamud, P. Naveau, a. Soloviev, P. Friederichs, V. Keilis-Borok, D. Kondrashov, V. Kossobokov, O. Mestre, C. Nicolis, H. W. Rust, P. Shebalin, M. Vrac, a. Witt, and I. Zaliapin, "Extreme events: dynamics, statistics and prediction," *Nonlin. Proc. Geophys.*, vol. 18, no. 3, pp. 295–350, 2011.

[4] R. G. Andrzejak and T. Kreuz, "Characterizing unidirectional couplings between point processes and flows," *Europhysics Letters*, vol. 96, no. 5, p. 50012, 2011.

[5] R. Quian Quiroga, T. Kreuz, and P. Grassberger, "Event synchronization: A simple and fast method to measure synchronicity and time delay patterns," *Physical Review E*, vol. 66, no. 4, p. 041904, 2002.

[6] N. Marwan, "Challenges and perspectives in recurrence analyses of event time series," *Front. Appl. Math. Stat.*, vol. 9, p. 1129105, 2023.

[7] N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, "Recurrence Plots for the Analysis of Complex Systems," *Phys. Rep.*, vol. 438, no. 5–6, pp. 237–329, 2007.

[8] J. D. Victor and K. P. Purpura, "Metric-space analysis of spike trains: theory, algorithms and application," *Netw. Comp. Neur. Syst.*, vol. 8, no. 2, pp. 127–164, jan 1997.

[9] S. Suzuki, Y. Hirata, and K. Aihara, "Definition of distance for marked point process data and its application to recurrence plot-based analysis of exchange tick data of foreign currencies," *Int. J. Bif. Chaos*, vol. 20, no. 11, pp. 3699–3708, 2010.

[10] I. Ozken, D. Eroglu, T. Stemler, N. Marwan, G. B. Bagci, and J. Kurths, "Transformation-cost time-series method for analyzing irregularly sampled data," *Phys. Rev. E*, vol. 91, p. 062911, 2015.

[11] A. Banerjee, B. Goswami, Y. Hirata, D. Eroglu, B. Merz, J. Kurths, and N. Marwan, "Recurrence analysis of extreme event-like data," *Nonlin. Proc. Geophys.*, vol. 28, pp. 213–229, 2021.

[12] D. Eroglu, F. H. McRobie, I. Ozken, T. Stemler, K.-H. Wyrwoll, S. F. M. Breitenbach, N. Marwan, and J. Kurths, "See-saw relationship of the Holocene East Asian-Australian summer monsoon," *Nature Communications*, vol. 7, p. 12929, 2016.

[13] J. Zhang, K. Liu, M. Wang, K. Li, F., J. Ludescher, J. Kurths, and N. Marwan, "High predictability potential of highly synchronized widespread floods in monsoon regions," *Journal of Hydrology*, vol. 668, p. 135006, 2026.

[14] M. G. Rosenblum, A. S. Pikovsky, and J. Kurths, "Phase Synchronization of Chaotic Oscillators," *Physical Review Letters*, vol. 76, p. 1804, 1996.

[15] M. C. Romano, M. Thiel, J. Kurths, I. Z. Kiss, and J. L. Hudson, "Detection of synchronization for non-phase-coherent and non-stationary data," *Europhysics Letters*, vol. 71, no. 3, pp. 466–472, 2005.

[16] A. M. Nkomidio, E. J. Ngamga, B. R. N. Nbendjo, J. Kurths, and N. Marwan, "Recurrence-Based Synchronization Analysis of Weakly Coupled Bursting Neurons Under External ELF Fields," *Entropy*, vol. 24, no. 2, p. 235, 2022.

[17] B. Goswami, G. Ambika, N. Marwan, and J. Kurths, "On interrelations of recurrences and connectivity trends between stock indices," *Physica A*, vol. 391, pp. 4364–4376, 2012.

[18] M. C. Romano, M. Thiel, J. Kurths, and W. von Bloh, "Multivariate Recurrence Plots," *Phys. Lett. A*, vol. 330, no. 3–4, pp. 214–223, 2004.

[19] Y. Zou, M. C. Romano, M. Thiel, N. Marwan, and J. Kurths, "Inferring Indirect Coupling by Means of Recurrences," *International Journal of Bifurcation and Chaos*, vol. 21, no. 4, pp. 1099–1111, 2011.

[20] A. M. T. Ramos, A. Builes-Jaramillo, G. Poveda, B. Goswami, E. E. N. Macau, J. Kurths, and N. Marwan, "Recurrence measure of conditional dependence and applications," *Phys. Rev. E*, vol. 95, p. 052206, 2017.

[21] K. H. Kraemer, F. Hellmann, M. Anvari, J. Kurths, and N. Marwan, "Spike spectra for recurrences," *Entr.*, vol. 24, no. 11, p. 1689, 2022.

[22] J. P. Zbilut and N. Marwan, "The Wiener-Khinchin theorem and recurrence quantification," *Physics Letters A*, vol. 372, no. 44, pp. 6622–6626, 2008.

[23] B. Goswami, N. Boers, A. Rheinwalt, N. Marwan, J. Heitzig, S. F. M. Breitenbach, and J. Kurths, "Abrupt transitions in time series with uncertainties," *Nature Communications*, vol. 9, p. 48, 2018.

[24] N. Hatami, Y. Gavet, and J. Debayle, "Classification of time-series images using deep convolutional neural networks," *Proceedings of SPIE*, vol. 10696, p. 106960Y, 2018.

[25] N. Marwan and K. H. Kraemer, "Trends in recurrence analysis of dynamical systems," *European Physical Journal – Special Topics*, vol. 232, pp. 5–27, 2023.