Recurrence analyses of event time-series

Norbert Marwan
Research Deptartment of Complexity Science
Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association
P.O. Box 60 12 03, 14412 Potsdam, Germany
Email: marwan@pik-potsdam.de

The analysis and modelling of event time-series are of broad interest in data science, as they are manifestations of the dynamics of the underlying system. Their relevance spans diverse scientific fields – from financial transactions and customer interactions to cardiac events, system failures, and natural phenomena – making them a ubiquitous and valuable subject of study. Event series may consist of discrete or binary events, or events with varying amplitudes, such as extremes, anomalies, or events derived from heavy-tailed distributions. 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. The critical challenge addressed here is the limitation of conventional time-series analysis methods when applied to sparse, irregular, or discrete event data (although several approaches address event-data analysis [2, 3]), hindering the ability to effectively compare these data with continuous records or reliably characterise their underlying spectral properties and nonlinear dynamics [5]. In this context, recurrence-based methods provide a promising alternative [5, 6].

A recurrence plot (RP) visualises pairwise similarity between all states $\vec{x}_i$ and $\vec{x}_j$: $R_{i,j} = \Theta(\varepsilon - d(\vec{x}_i, \vec{x}_j))$, 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 $d_{L2} = \|\vec{x}_i - \vec{x}_j\|$) falls below a threshold ε [6]. By selecting a similarity measure suitable for discrete data, RPs can also be constructed for event data. A natural choice is the edit distance, an extension of the Levenshtein distance [11]. For two event sequences $S_i = \{t_k^{(i)}\}_{k=1}^{N_i}$ and $S_j = \{t_k^{(j)}\}_{k=1}^{N_j}$ the edit distance d_{ED} is the minimal total cost of transforming S_i into S_j using deletions, insertions, and temporal shifts of events t_k. This metric has been successfully combined with recurrence analysis [1, 10] and allows many advanced data analysis tasks offered by the RP-based framework, thereby extending the powerful tools of nonlinear time-series analysis and phase space characterisation to discrete and event-based processes.

For example, it enables a direct comparison of event with continuous data. One important recurrence-based approach is based on joint recurrence: $JR_{i,j}^{x,y} = d_{ED}(S_i^x, S_j^x) \cdot \Theta(\varepsilon - d_{L2}(\vec{y}_i, \vec{y}_j))$, where x_i denotes an event series and y_i a continuous time-series. It quantifies simultaneous recurrences in both signals and, thus, provides a powerful indicator of generalised synchronisation [9]. This capability is essential for analysing coupled dynamical systems where different components may be measured using heterogeneous data types (discrete events vs.continuous variables). This would also allow future extensions of the JRP approach for detecting coupling directions or even causal links between systems [8]. Nevertheless, the method faces practical challenges such as parameter selection, dimension mismatch, alternative similarity measures, and sensitivity to noise, which will be discussed briefly. Edit-distance-based RPs also enable power-spectrum estimation for spiky signals [4] and facilitate analysis of irregularly sampled data [7], a common challenge, e.g., in astrophysics and palaeoclimate research.

These extensions are demonstrated on climate and palaeoclimate data, where extreme event-like records are analysed together with continuously recorded observations. The results demonstrate that recurrence-based techniques can uncover hidden synchronisation patterns and produce reliable spectral estimates despite the sparsity of the event data.

References

[1] A. Banerjee et al. In: *Nonlin. Proc. Geophys.* 28 (2021), pp. 213–229.

[2] J. Beirlant et al. *Statistics of Extremes: Theory and Applications*. Chichester, UK: Wiley, 2006, p. 512. ISBN: 978-0-470-01237-6.

[3] M. Ciba et al. In: *Neur. Comp.* 32.5 (May 2020), pp. 887–911. ISSN: 0899-7667.

[4] K. H. Kraemer et al. In: *Entr.* 24.11 (2022), p. 1689.

[5] N. Marwan. In: *Front. Appl. Math. Stat.* 9 (2023), p. 1129105.

[6] N. Marwan et al. In: *Phys. Rep.* 438.5–6 (2007), pp. 237–329.

[7] I. Ozken et al. In: *Phys. Rev. E* 91 (2015), p. 062911.

[8] A. M. T. Ramos et al. In: *Phys. Rev. E* 95 (2017), p. 052206.

[9] M. C. Romano et al. In: *Phys. Lett. A* 330.3–4 (2004), pp. 214–223.

[10] S. Suzuki et al. In: *Int. J. Bif. Chaos* 20.11 (2010), pp. 3699–3708.

[11] J. D. Victor and K. P. Purpura. In: *Netw. Comp. Neur. Syst.* 8.2 (Jan. 1997), pp. 127–164. ISSN: 0954-898X.