IEICE Proceeding Series

Nonlinear time series analysis of marked point process data

Koji Iwayama, Yoshito Hirata, Kazuyuki Aihara

Vol. 2 pp. 189-192
Publication Date: 2014/03/18
Online ISSN: 2188-5079

2013 International Symposium on Nonlinear Theory and its Applications
NOLTA2013, Santa Fe, USA, September 8-11, 2013

Nonlinear time series analysis of marked point process data

Koji Iwayama[†‡], Yoshito Hirata[‡], and Kazuyuki Aihara[‡]

†FIRST, Aihara Innovative Mathematical Modelling Project, Japan Science and Technology Agency
Collaborative Research Center for Innovative Mathematical Modelling,Institute of Industrial Science,
The University of Tokyo 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan
‡Institute of Industrial Science, The University of Tokyo
4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan
Email: koji@sat.t.u-tokyo.ac.jp

Abstract—Marked point process data are time series of discrete events which have some values as marks. A definition of a distance for marked point process data allows us to apply nonlinear time series analysis for marked point process data. However, the existing method uses only approximations of distances due to computational complexity. Here, we calculate exact values of distances and apply some methods of nonlinear time series analysis to a series of local maxima from a Rössler model. Results indicate the effectiveness of distance-based approach to analyze marked point process data.

1. Introduction

A point process produces a time series of discrete events. A typical example of a point process is action potentials of neurons. Like action potentials, point processes, whose events are identical, are called "simple point process." Differently from action potentials, events of earthquake occurrence are not identical, namely, each event is marked by its magnitude and the position of its seismic center. We call such process "marked point process." Due to the discreteness of marked point process data, many methods for time series analysis, which are intended for continuous data, are inapplicable to marked point process data.

However, we can apply many methods to point process data based on its distances. For simple point process data, there exist some distances [1, 2, 3]. Using these distances, many methods have been applied to simple point process data, including clustering [4], fitting to the model [5], multi-dimensional scaling [6], classification [7], and reconstruction of input signals from spike trains [8]. The distance for marked point process data is proposed by Suzuki *et al.* [9]. This distance is the minimum cost required to transform one data into the other one. Based on the distance, they obtained recurrence plots, which visualize recurrences in time series, of exchange tick data of foreign currencies.

Although the definition of distances allows us to apply many methods to marked point process data, it has not been confirmed that results obtained by such analysis correctly reflect properties of the process which generated the observed marked point process data. The results might be spurious and independent of the underlying process. Thus, we investigated the relation between the results obtained by analyses based on the distance and actual properties of the underlying process. We applied some methods of nonlinear time series analysis to a series of local maxima of a Rössler model as an example.

2. Method

2.1. Distance for Marked Point Process Data

The distance for marked point process data [9] is defined as the minimum cost of transformation of one data into another one by shifting, inserting, and removing events. Let $\boldsymbol{x}_i = (x_{i0}, \ldots, x_{im})$ denote the ith event, where x_{i0} is the occurrence time of the event and x_{ij} is the value of its jth mark. When we have two marked point process data $\boldsymbol{X} = \{\boldsymbol{x}_i\}$ and $\boldsymbol{Y} = \{\boldsymbol{y}_i\}$, we consider the set of pairs of events,

$$C = \{(\boldsymbol{x}_k, \boldsymbol{y}_l) \in \boldsymbol{X} \times \boldsymbol{Y} | \boldsymbol{x}_k \neq \boldsymbol{x}_{k'}\ (\forall k \neq k'),\ \boldsymbol{y}_l \neq \boldsymbol{y}_{l'}\ (\forall l \neq l')\}. \quad (1)$$

This set represents the procedure to transform the data $\boldsymbol{X}$ into $\boldsymbol{Y}$. The element $(\boldsymbol{x}_k, \boldsymbol{y}_l)$ in this set means that the kth event of $\boldsymbol{X}$ is shifted to the position of the lth event of $\boldsymbol{Y}$. Events of $\boldsymbol{X}$ not contained in pairs of C are removed and those of $\boldsymbol{Y}$ are inserted. Then, the distance for marked point process data [9] is defined as below,

$$D(\boldsymbol{X}, \boldsymbol{Y}) = \min_C \sum_{(\boldsymbol{x}_k, \boldsymbol{y}_l) \in C} \sum_{n=0}^{m} \lambda_n |x_{kn} - y_{ln}| + |\boldsymbol{X}| + |\boldsymbol{Y}| - 2|C|. \quad (2)$$

Here, λ_n is the cost of shifting the value of the nth mark.

To apply nonlinear time series analysis methods to marked point process data, we used a sliding window which is shifted by a fixed amount along the time axis. Then we calculated the distance for each pair of windows. When the edge of the window passes through an event, the distance suddenly jumps due to appearances or disappearances of events from the window. We considered events before or after the window to avoid such jumps as the literature [8]. We considered 4 options: include one event before the window; include one event after the window; include events before and after the window; include no events out of the window. We calculated distances for all of 16 possible combinations of these options and adopt the minimum one as the distance between two windows. Finally, we replaced the distance by the shortest route length in the weighted complete graph.

When we calculate the distance for marked point process data, it is required to solve the minimization problem of the cost of transformation. Because this minimization consumes much time, in the original literature [9], authors approximated the distance with a numerical optimization method. To validate analysis based on this distance, we calculated exact values of distances.

2.2. Estimation of the Largest Lyapunov Exponent

Lyapunov exponents characterize the rate of separation of close trajectories of a dynamical system. The sign of the maximal Lyapunov exponent (MLE) indicates the stability of the system. We can estimate the MLE only using distances among states by measuring the rate of separation of the close trajectories directly [10].

We suppose that we have the time series of the state $\boldsymbol{X}(t)$. Let $U(t)$ denote the time indices of neighbors of the state at the time t, i.e., $U(t) = \{t' | D(\boldsymbol{X}(t), \boldsymbol{X}(t')) < \epsilon\}$. Then we compute

$$S(\tau) = \frac{1}{T} \sum_{t=1}^{T} \ln \left(\frac{1}{|U(t)|} \sum_{t' \in U(t)} D(\boldsymbol{X}(t+\tau), \boldsymbol{X}(t'+\tau)) \right). \quad (3)$$

Here, T is the length of the time series. The slope of the plot of $S(\tau)$ vs. τ corresponds to the MLE.

2.3. Recurrence Quantification Analysis

Recurrence plots [11, 12] visualize the recurrence structures of dynamical systems. The recurrence plot of the time series $\boldsymbol{X}(t)$ of length T is the matrix,

$$\boldsymbol{R}_{i,j} = \begin{cases} 1, & D(\boldsymbol{X}(i), \boldsymbol{X}(j)) < \epsilon, \\ 0, & \text{otherwise}, \end{cases} \quad (4)$$

where ϵ is the threshold. We plot a dot at (i, j) if and only if $\boldsymbol{R}_{i,j} = 1$. We can intuitively read out various features of dynamical systems from their recurrence plots. Recurrence quantification analysis (RQA) [12, 13, 14] quantifies structures of recurrence plots.

One of the most important structures is the number of diagonal lines of length l, i.e.,

$$N(l) = \sum_{i=1}^{T} \sum_{j=i+1}^{T} (1 - \boldsymbol{R}_{i-1,j-1})(1 - \boldsymbol{R}_{i+l,j+l}) \prod_{k=0}^{l} \boldsymbol{R}_{i+k,j+k}. \quad (5)$$

A diagonal line of length l indicates that two segments of the trajectory are close during l time steps. Hence, the average diagonal line length corresponds to the mean prediction time.

The τ-recurrence rate, which is defined as

$$RR(\tau) = \frac{1}{T-\tau} \sum_{i=1}^{N-\tau} \boldsymbol{R}_{i,i+\tau}, \quad (6)$$

indicates the probability that the state returns to its neighborhood after τ time steps. The τ-recurrence rate generalizes the auto-correlation function.

2.4. Reconstruction of Original Time Series

Hirata *e*t al. [8] demonstrated that the encoded signals can be reconstructed from neural spike trains on the basis of distances for simple point process data. They converted the spike train into a real-valued time series using the Isomap [15], which is one of the methods for dimensionality reduction.

3. Results

To validate nonlinear time series analysis of marked point process data based on distances, we applied the methods mentioned above to a series of local maxima of the Rössler model [16]. The Rössler model is defined as

$$\begin{aligned} \frac{dx}{dt} &= -y - z, \\ \frac{dy}{dt} &= x + ay, \\ \frac{dz}{dt} &= b + z(x - c). \end{aligned} \quad (7)$$

Here, a, b, and c are parameters. We integrated Eq. (7) with $a = 0.1$, $b = 0.1$, and $c \in \{5, 6, \ldots, 20\}$ for 1000 time units. The bifurcation diagram is illustrated in Fig. 1. We observed various types of behavior including periodic and chaotic dynamics in the range of $c \in [5\ 20]$.

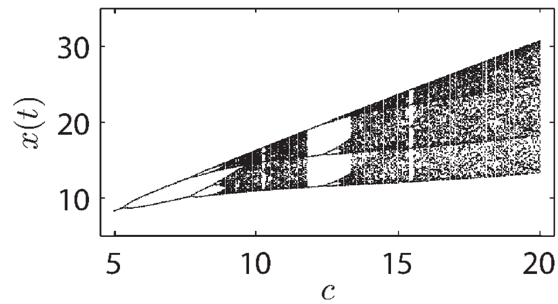

Figure 1: Bifurcation diagram of Rössler model. The vertical axis is the values of local maxima of $x(t)$.

Marked point process data are generated by recording the time and the value of local maxima of $x(t)$ as shown in Fig. 2. A time window of length 20 is shifted by 1 time unit.

We calculated distances between these windows as mentioned above. We set the value of all parameters λ_n in Eq. 2

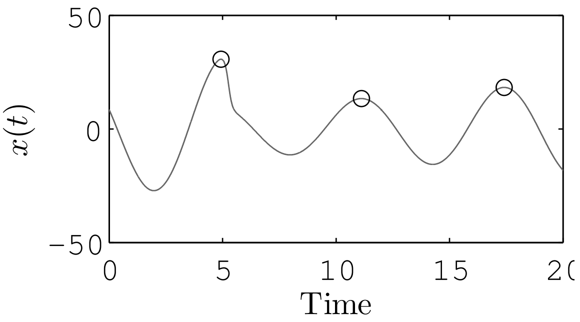

Figure 2: Example of marked point process data. Black circles indicate events, namely local maxima of $x(t)$. The gray curve is the time series of $x(t)$.

to 1. Moreover, we recorded the value of x at the center of a window as the representative of the segment of the trajectory in the window and calculated the Euclidean distances between these points. Hereafter, we refer these representative values of windows and these distances as "original time series" and "original distances," respectively. Distances for the marked point process data and original distances were highly correlated. Correlation coefficients were distributed in the range from 0.83 to 0.96.

Maximal Lyapunov exponents estimated from distances for marked point process data and original time series are plotted in Fig. 3. MLEs estimated from marked point process data almost correspond to those from original time series. Both MLEs are almost 0 when the system is periodic. On the other hand, MLEs take larger value in the chaotic regime.

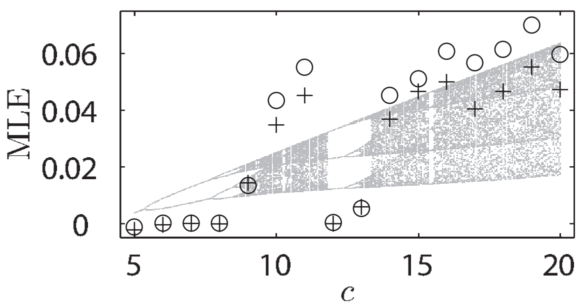

Figure 3: Plot of maximal Lyapunov exponents (MLE) vs. c. MLEs are plotted over the scaled bifurcation diagram (gray dots). Black crosses and circles indicate MLEs estimated from marked point process data and original time series, respectively.

We obtained recurrence plots of marked point process data and original time series. The threshold ϵ in Eq. 4 was set to obtain a probability of 0.01 for plotting a dot. Fig. 4 illustrates examples of recurrence plots of the original time series and the marked point process data. These plot exhibits similar structures.

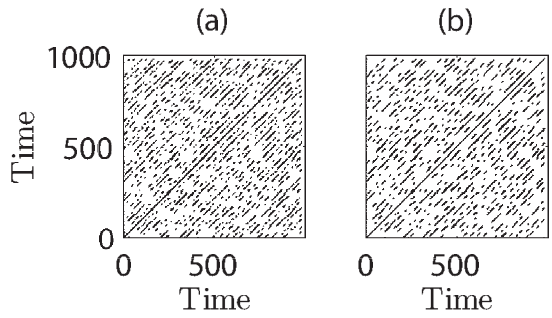

Figure 4: Recurrence plots with $c = 20$. (a) The recurrence plot of the original time series. (b) The recurrence plot of the marked point process data.

Average diagonal line lengths of these recurrence plots are plotted in Fig. 5. Trends of two plots roughly agree. Both curves exhibit smaller values in the chaotic regime.

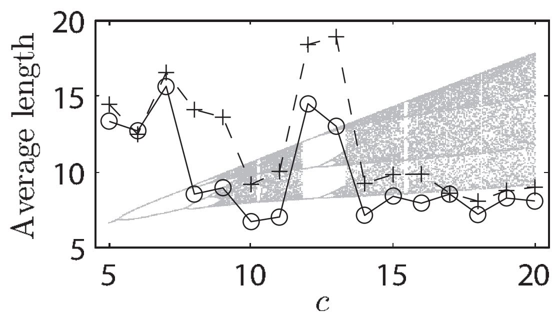

Figure 5: Plot of average diagonal line length vs. c. As the case of MLE, average diagonal line length is plotted over the bifurcation diagram. The solid line and circles indicate average diagonal line lengths of recurrence plots of the original time series. The dashed line and crosses those of marked point process data.

Further, we calculated correlation coefficients between the τ-recurrence rates of the original time series and that of the marked point process data. The minimum value was 0.96. Thus, two τ-recurrence rates were highly correlated in all cases.

Finally, we attempted to reconstruct the original time series of $x(t)$ from the marked point process data. The first 100 time units of the original time series and the reconstructed time series of the case of $c = 20$ are plotted in Fig. 6. The reconstructed time series agreed well with the original time series. Moreover, correlation coefficients between the original and reconstructed time series are larger than 0.95 in all cases.

4. Conclusion

By defining a distance, we can apply many methods of nonlinear time series analysis to marked point process

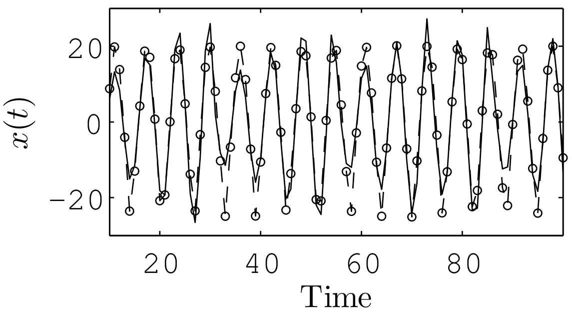

Figure 6: Reconstruction of the original time series from the marked point process data. The first 100 time units of the original time series of $x(t)$ with the parameter $c = 20$ is plotted by the black solid curve. The dashed line and black circles indicate the reconstructed time series from marked point process data.

data. However, it was unknown whether the results obtained by such analysis actually reflect the underlying dynamics which generated the observed marked point process data or not.

We applied some methods of nonlinear time series analysis based on distances for data to marked point process data generated by Rössler model. Results indicated that characteristics estimated from marked point process data agreed well with those estimated from the original time series. Moreover, those characteristics could distinguish between periodic and chaotic behavior of the original dynamics.

Acknowledgments

This research is supported by the Aihara Innovative Mathematical Modelling Project, the Japan Society for the Promotion of Science (JSPS) through the "Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program)," initiated by the Council for Science and Technology Policy (CSTP).

References

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

[2] M.C.W. Van Rossum, "A novel spike distance," *Neural. Comput.*, vol.13, no.4, pp.751–764, 2001.

[3] J.D. Victor, "Spike train metrics," *Curr. Opin. Neurol.*, vol.15, no.5, pp.585–592, 2005.

[4] K. MacLeod, A. Bäcker, and G. Laurent, "Who reads temporal information contained across synchronized and oscillatory spike trains?," *Nature*, vol.395, no.6703, pp.693–698, 1998.

[5] J. Keat, P. Reinagel, R.C. Reid, and M. Meister, "Predicting every spike: a model for the responses of visual neurons," *Neuron*, vol.30, no.3, pp.803–817, 2001.

[6] D. Aronov, D.S. Reich, F. Mechler, and J.D. Victor, "Neural coding of spatial phase in V1 of the macaque monkey," *J. Neurophysiol.*, vol.89, no.6, pp.3304–3327, 2003.

[7] R. Narayan, G. Graña, and K. Sen, "Distinct time scales in cortical discrimination of natural sounds in songbirds," *J. Neurophysiol.*, vol.96, no.1, pp.252–258, 2006.

[8] Y. Hirata and K. Aihara, "Representing spike trains using constant sampling intervals," *J. Neurosci. Meth.*, vol.183, no.2, pp.277–286, 2009.

[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. Bifurcat. Chaos.*, vol.20, no.11, pp.3699–3708, 2010.

[10] H. Kantz, "A robust method to estimate the maximal Lyapunov exponent of a time series," *Phys. Lett. A*, vol.185, no.1, pp.77–87, 1994.

[11] J.-P. Eckmann, S.O. Kamphorst, and D. Ruelle, "Recurrence plots of dynamical systems," *Europhys. Lett.*, vol.4, no.9, pp.973–977, 1987.

[12] 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.

[13] J.P. Zbilut and C.L. Webber Jr., "Embeddings and delays as derived from quantification of recurrence plots," *Phys. Lett. A*, vol.171, no.3, pp.199–203, 1992.

[14] C.L. Webber Jr. and J.P. Zbilut, "Dynamical assessment of physiological systems and states using recurrence plot strategies," *J. Appl. Physiol.*, vol.76, no.2, pp.965–973, 1994.

[15] J.B. Tenenbaum, V. de Silva, and J.C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol.290, no.5500, pp.2319–2323, 2000.

[16] O.E. Rössler, "An equation for continuous chaos," *Phys. Lett. A*, vol.57, no.5, pp.397–398, 1976.