

リカレンスプロット：時系列の視覚化を越えて
Recurrence plots: Beyond visualization of time series

東京大学 生産技術研究所 平田祥人

Yoshito Hirata

Institute of Industrial Science, The University of Tokyo

yoshito@sat.t.u-tokyo.ac.jp

要旨

この原稿では、リカレンスプロットについて紹介する。リカレンスプロットは、非線形時系列解析の手法の 1 つであり、元々時系列データを視覚化するための道具である。近年、リカレンスプロットを用いた時系列解析手法が急速に発達している。また、距離が定義できればリカレンスプロットを求めることができることより、リカレンスプロットは点過程データなど既存の手法があまり用意されていないデータに対しても用いることができる。本原稿では、リカレンスプロットを用いた解析の発展をまとめるとともに、最近の我々の研究を紹介する。

1. はじめに

1987 年に提案されたリカレンスプロット(Eckmann et al. (1987); Marwan et al. (2007)) は、元々時系列データの特徴を視覚化するための 2 次元平面図である。縦軸、横軸とも時間軸であるこの平面図において、リカレンスプロットは、「2 つの対応する時刻の状態間の距離が近ければ対応する場所に点を打ち、そうでなければ点を打たない」とすることで定義できる。数学的には、 M を距離空間、そこで定義される距離を d 、 M 上に値を取る時系列データの i 番目の状態を $x(i) \in M$ とすると、リカレンスプロットは、

$$R(i, j) = \begin{cases} 1, & d(x(i), x(j)) < r(i, j) \\ 0, & \text{otherwise.} \end{cases}$$

と定義できる。ここで $R(i, j) = 1$ のとき (i, j) に点を打ち、 $R(i, j) = 0$ のとき (i, j) に点を打たない。 $r(i, j)$ はしきい値と呼ばれ、Zbilut and Webber (1992) の定義では、 $r(i, j) = r$ のように定数に選ぶ。Eckmann et al. (1987) の定義では、 $r(i, j)$ を i 番目の点とその他の点の距離のうち小さい方から k 番目の距離を用いている。

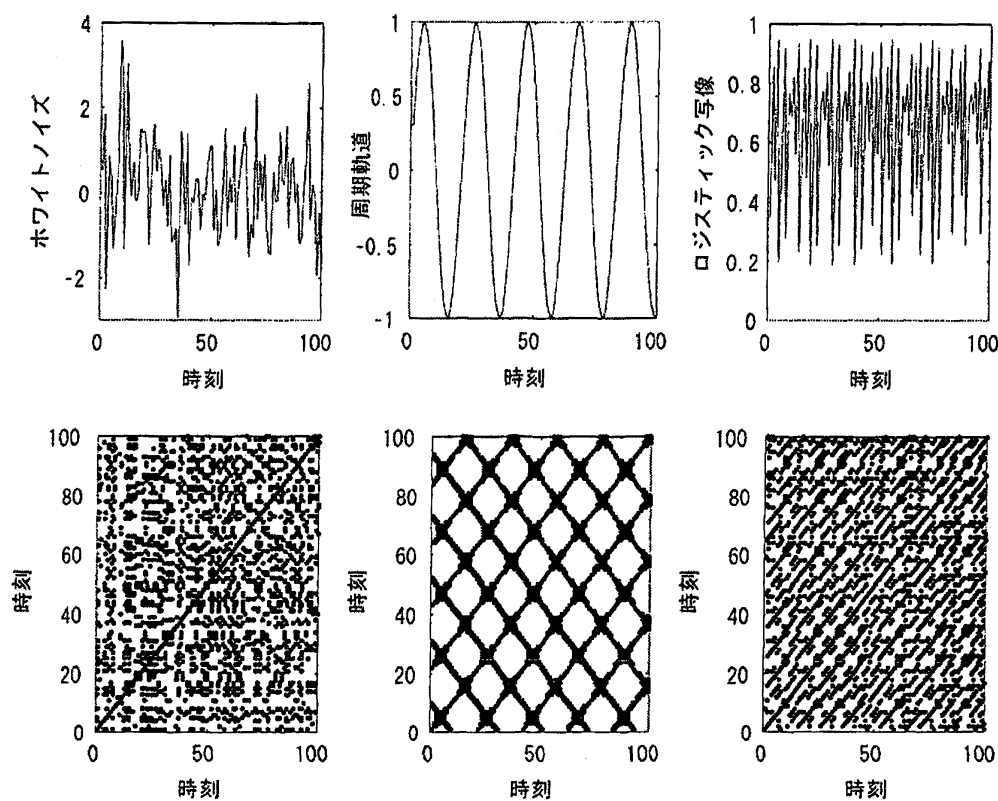


図1：様々な時系列データ(上段)とそのリカレンスプロット(下段)。左からホワイトノイズ、周期的なサイン関数、ロジスティック写像の例。

リカレンスプロットこのようにリカレンスプロットの定義はシンプルであるが、リカレンスプロットからは時系列データの背後にある力学系の様々な情報がわかる。ここで、 M を1次元の実数空間、 d を2つの実数の絶対値、つまり、 $d(x(i), x(j)) = |x(i) - x(j)|$ と選んで例を示す(図1)。ホワイトノイズでは、点が不規則に一樣に広がる。それに対して、周期的なサイン関数では、周期的なパターンの点分布となる。決定論的カオスの時系列データ、例えば、ロジスティック写像の時系列では、斜めの短い線分が多く現れる。このように短い線分が現れるのは、決定論的な時系列データでは2つの異なる点がいったん近傍点になるとしばらく近傍に留まり続けるためである。このように、リカレンスプロットの点の分布を見ることで、時系列データの性質をおおざっぱにつかむことができる。

より、一般的に、 M を m 次元のベクトル空間、 $x_k(i)$ を $x(i)$ の第 k 成分、 d をユークリッド

ド距離、つまり、 $d(x(i), x(j)) = \sqrt{\sum_{k=1}^m (x_k(i) - x_k(j))^2}$ と選んだとしよう。この時、リカレン

ンスプロットは、連続値の時系列情報を 2 値行列情報に変換しているので、情報がかなり落ちてしまっているとい瞬思われるかもしれない。しかし、最近の研究で、リカレンスプロットから元の連続値時系列の概形が復元できることがわかってきた(Hirata et al. (2008); 1次元空間の場合の結果は、Thiel et al. (2004a))。そのため、リカレンスプロットは、時系列データの有効な表現手段となり得る。実際、リカレンスプロットから相関次元や相関エントロピー(Faure and (1998); Thiel et al. (2004b))を推定することができる。

そこで、本原稿の以下の部分では、リカレンスプロットを使った時系列データの解析についてまとめる。2 節では、まず、リカレンスプロットの定量化についてまとめる。3 節では、定量化手法を発展させた、リカレンスプロットを用いた単変量解析についてまとめる。4 節では、リカレンスプロットを用いて外力の再構成をする手法を紹介する。5 節では、リカレンスプロットを用いた多変量解析についてまとめる。6 節では、リカレンスプロットを点過程に対して定義する。7 節で本原稿の結びを述べる。

2. リカレンスプロットの定量化

1992 年以降、リカレンスプロットの点のパターンを定量化する手法が提案されてきている(Webber and Zbilut, 1994; Marwan et al., 2002; Marwan et al., 2009)。これらの大きく分けると、3 種類の量に分けられる。1 つ目は斜めの線の特徴づける量(Webber and Zbilut, 1992)、2 つ目は縦線・横線の特徴づける量(Marwan et al., 2002)、3 つ目はリカレンスプロットをネットワークの隣接行列と見立てそのネットワークの特徴づける量(Marwan et al., 2009)である。ここでは、特に、斜めの線の特徴づける量に関して紹介する。

斜めの線の長さが l の集合は、

$$D(l) = \left\{ (i, j), i=1, 2, \dots, n-1, j=i+1, i+2, \dots, n \mid (1-R(i-1, j-1))(1-R(i+l, j+l)) \prod_{k=0}^{l-1} R(i+k, j+k) = 1 \right\}$$

と書ける。斜めの線の長さを用いて定義される量は、主に 4 つある。1 つ目は、斜めの線をなす点の割合 (DET) である。これは、

$$DET = \frac{\sum_{l \geq 2} l |D(l)|}{\sum_{l \geq 1} l |D(l)|}$$

と書ける。ここで、 $|A|$ は、集合 A の要素の数を表す。時系列が決定論的であるとき、DET は大きな値を示す傾向にある。

2 つ目は、斜めの線の長さの平均 L で、

$$L = \frac{\sum_{|I|} |D(I)|}{\sum_{|I|} 1}$$

と書ける。時系列が決定論的であるとき、 L は大きめの値を取る。

3 つ目の量は、斜めの線の最大長 $L_{\max}$ で、記号では、

$$L_{\max} = \max\{|I| \mid D(I) \neq \emptyset\}$$

と書ける。ここで、 $\emptyset$ は空集合を示す。この量は、リアプノフ指数と反比例すると考えられている。

4 つ目の量は、斜めの線の長さに関するエントロピーである。この量は、

$$p(I) = |D(I)| / \sum_{|I|} |D(I)|$$

とする時、

$$ENTR = -\sum_{|I|} p(I) \log p(I)$$

と定義される。この量も、リアプノフ指数と反比例すると考えられている。

様々な量が提案されて来ているが、一貫して言えるのは、打たれている点のパターンを表面的にとらえるものであり、提案されている定量化と背後に存在する力学系の特徴がうまく関係づけられていない。こういう問題の動機づけにより、背後に存在する力学系の特徴とうまく関係づけられるような解析手法を開発するに至った。

3. 単変量時系列データのためのリカレンスプロット

この章では、1 つの時系列が与えられた時に分かる背後の力学系の性質に関して、我々が開発した手法をまとめる。

3.1 系列相関

系列相関は、斜めの線が連続しやすい傾向にあるかどうかを調べることによって検定できる (Hirata and Aihara, 2011)。リカレンスプロット上で点が打たれる確率を p とする。この時、斜めに 2 つ点が並ぶ確率は、 p^2 と書くことができる。実際にリカレンスプロットの

左上半分に斜めに2つ点が並んだ数を数え、 n_d とする。リカレンスプロットの左上半分の上で2つ点が並ぶ可能性のある場所の数を $m_d = \frac{1}{2}(n-1)(n-2)$ とすると、時系列の長さが十分長い時、系列相関がないという帰無仮説の下では、 n_d は確率 p^2 、大きさ m_d の2項分布に従う。 m_d が十分大きな時、この2項分布は、平均 $m_d p^2$ 、分散 $m_d p^2(1-p^2)$ の正規分布で近似できる。よって、

$$z_d = \frac{n_d - m_d p^2}{\sqrt{m_d p^2(1-p^2)}}$$

と変数変換すると、 z_d は、平均0、分散1の正規分布となる。このように、 z_d という量を扱うことによって、サロゲートデータなしに系列相関を検定することができる。

図1の例では、ホワイトノイズのp値が0.58になるのに対し、周期関数とロジスティック写像ではp値は0.001以下である。

3.2 カオス性

リカレンスプロットの点のパターンから、元の力学系がDevaneyのカオス(Devaney, 1989)の定義と無矛盾であるかどうか調べることができる(Hirata and Aihara, 2010a)。Devaneyのカオスの定義は、位相推移性、周期点の稠密性、初期値鋭敏性からなる。Hirata and Aihara (2010b)では、これらの概念を有限長の時系列において定義できるように定義を緩めた。その概念をそれぞれr位相推移性、周期点のr稠密性、r初期値鋭敏性と書く。これらの概念は、リカレンスプロットの点のパターンを見ることにより解釈できる。リカレンスプロットの言葉で解釈すると、r位相推移性とは、リカレンスプロットの最大インデックスの最小値が最小インデックスの最大値よりも大きいことである。周期点のr稠密性とは、斜めに点が連続する傾向にある所を周期解だと思いそのインデックス集合の和集合を求めると、時系列のインデックス集合全体と一致することである。r初期値鋭敏性とは、どの斜めの線も必ず途中で途切れていることであると解釈できる。

図1のロジスティック写像のリカレンスプロット例では、r位相推移性、周期点のr稠密性、r初期値鋭敏性がそれぞれ満たされているのに対し、ホワイトノイズの例ではr位相推移性と周期解のr稠密性が満たされず、周期関数の例ではr初期値鋭敏性が満たされない。

このように、リカレンスプロットを用いることで、比較的簡単に決定論的カオスを特徴づけることができる。

4. リカレンスプロットを用いた外力の再構成

リカレンスプロットの強みは、空間スケール以外の時系列の情報のほとんどがリカレンスプロットに含まれていることである。よって、リカレンスプロットから元の時系列が復元できる。この性質を利用することで、観測されていない時間的にゆっくり変化する外力を、変化の激しい外力によって駆動される力学系の観測から再構成することができる。ここで活躍するのが、外力に関する埋め込み定理である (Casdagli, 1997; Stark, 1999; Hegger et al., 2000)。

外力が加わった系に対して十分大きな埋め込み次元を用いた遅れ座標を構成すると、外力と観測した力学系の状態を組み合わせた状態が再構成される。外力の変化が十分ゆっくりなとき、このような遅れ座標を用いて求めたリカレンスプロットは、外力のリカレンスプロットと良く似たものになることが知られている (Casdagli, 1997)。これは、遅れ座標で近傍に来る点が外力と観測した力学系の状態が同時に近傍に来る点であるため、遅れ座標で求めたリカレンスプロットが外力のリカレンスプロットの部分集合になることによって、このため、外力を受ける系の観測を用いて求めたリカレンスプロットを粗視化し、これを Hirata et al. (2008) や Tanio et al. (2009) の手法を使って時系列に変換すれば、外力を復元することができる。

ここで、外力の復元の例を示す。ここでは、Henon 写像に Lorenz' 63 モデルを用いてゆっくりとした外力を加えた。図 2 の灰色の線に外力として用いた Lorenz' 63 モデルの時系列を示す。この Henon 写像から得られた時系列を 10 次元の遅れ座標を使って埋め込み、リカレンスプロットを求めた。これを粗視化すると、図 3 のようになる。このリカレンスプロットは、Lorenz' 63 とよく似たリカレンスプロットになっていることがわかる。この粗視化したリカレンスプロットを Hirata et al. (2008) を用いて時系列に変換すると、図 2 の黒線のようにになる。Lorenz' 63 モデルの外力がよく再構成されていることがわかる。

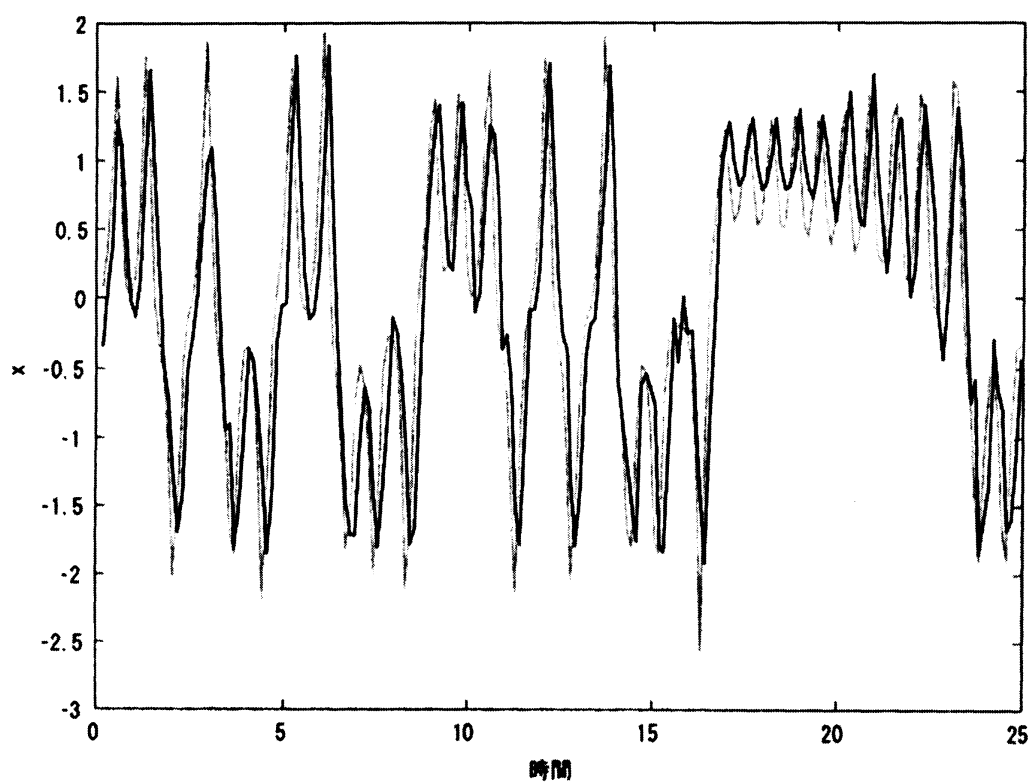


図 2 : 外力として用いた Lorenz' 63 (灰色) と再構成された外力 (黒線)。

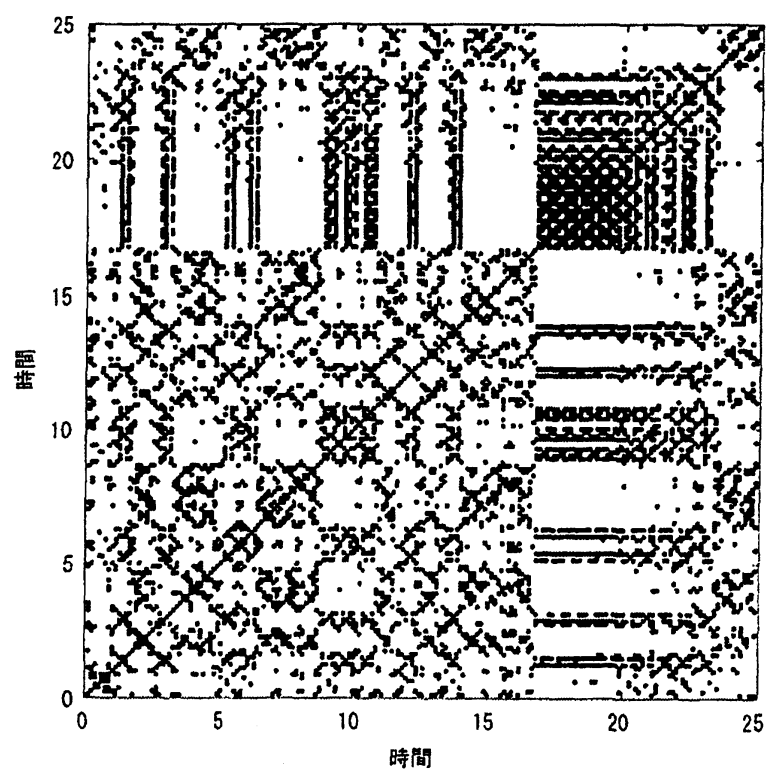


図 3 : Lorenz' 63 モデルによって駆動された Henon 写像から求めたリカレンスプロット。

5. 多変量時系列データのためのリカレンスプロット

リカレンスプロットは多変量時系列データ解析にも拡張されている。2つの拡張がある。1つ目は、クロスリカレンスプロットである (Zbilut et al., 1998; Marwan and Kurths, 2002)。クロスリカレンスプロットでは、縦軸・横軸に異なる時系列データを取る。距離空間 M 上に値を取る2つの時系列データの i 番目の状態を $x(i), y(i) \in M$ とする。この時、クロスリカレンスプロットは、

$$C(i, j) = \begin{cases} 1, & d(x(i), y(j)) < r(i, j) \\ 0, & \text{otherwise.} \end{cases}$$

となる。2つの時系列が同様の変化を示した時に、斜めの線が現れる。

もう1つの解析手法は、ジョイントリカレンスプロットである (Romano et al., 2004)。ジョイントリカレンスプロットは、2つのリカレンスプロットの直積として定義される。つまり、2つのリカレンスプロット両方で点が打たれている場合にのみ、ジョイントリカレンスプロットで点を残す。数学的には、次のように定義される。2つの時系列データ $x_1(i) \in M_1$, $x_2(i) \in M_2$ から求めたリカレンスプロットをそれぞれ $R_1(i, j)$, $R_2(i, j)$ とする。この時、ジョイントリカレンスプロットは、 $J(i, j) = R_1(i, j)R_2(i, j)$ と定義できる。2つの時系列で同じ時刻に再帰が起こる傾向にある時、ジョイントリカレンスプロットに残る点の割合が多くなる。

昨年、Hirata and Aihara (2010b)において、我々は、ジョイントリカレンスプロットを用いて、方向性の結合を検定する手法を提案した。この手法の特徴は、観測されていない共通の第3のダイナミクスがあり、それが観測された2つのダイナミクスに影響を与えている場合にも、第3のダイナミクスの存在を推定することができる点にある。

6. 点過程データのためのリカレンスプロット

この章では、不規則な間隔で観測が得られるような現象のことを点過程と呼ぶことにする。点過程の現象は、例えば、神経の発火、地震、為替取引等様々なものがある。しかし、時系列データ解析は今まで主に観測が一定時間間隔おきに得られる現象のみを中心に開発されてきたため、点過程データを解析する良い手法があまりない。そこで、我々は、点過程データに対しても距離が定義できる (Victor and Purpura, 1997; Hirata and Aihara, 2009; Suzuki et al., 2010) ことに着目し、点過程データに対してもリカレンスプロットを定義することを提案した (Suzuki et al., 2010)。この章では、点過程間の距離を紹介するとと

もに、点過程データの解析例を紹介する。

点過程間の距離の基本は、2つの点過程の窓を比較し、一方の窓の中にあるイベントに編集を加えて行って、もう一方の点過程の窓を作っていく時の編集のコストの最小値を距離として定義するものである(図4参照)。修正の仕方としては3通りあり、イベントの挿入・削除にはコスト1、移動には移動量に比例したコストを割り当てる。神経の発火などの値を伴わない(マークなし)点過程に関して距離を定義したのが、Victor and Purpura (1997)の仕事である。これをイベントに値を伴うマーク付き点過程に拡張したのが我々(Suzuki et al., 2010)の仕事である。

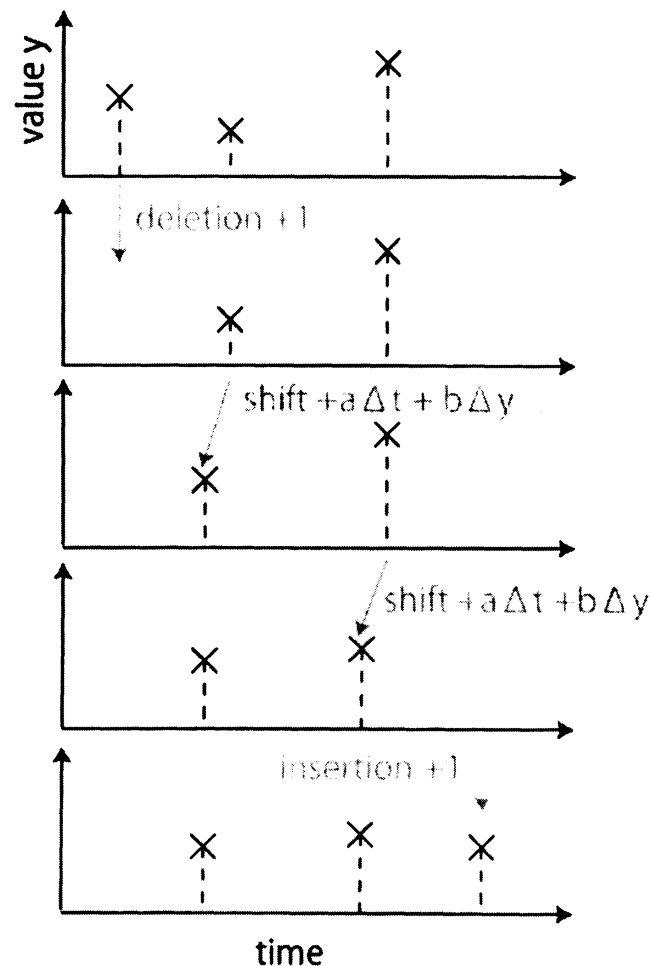


図4：点過程の距離の計算例

窓を一定間隔ずつ動かして行って、2つの窓の場所間で上の距離を求めれば、点過程の時系列データからリカレンスプロットを描くことができる。但し、点過程に基づく距離を用いてリカレンスプロットを描く時には、少しコツがいる。各時刻で距離の分布が大きく異なる

るのが常であることから、各時刻で一定のしきい値を用いてしまうと、点の四角いかたまりをたくさん含むようにリカレンスプロットが得られてしまうことがある。そこで、各時刻において、 k 個の近傍点をプロットするようにしてリカレンスプロットを求めることにすることが考えられる。このようにすれば、点過程の時系列データから細かい構造を読み取ることができる。

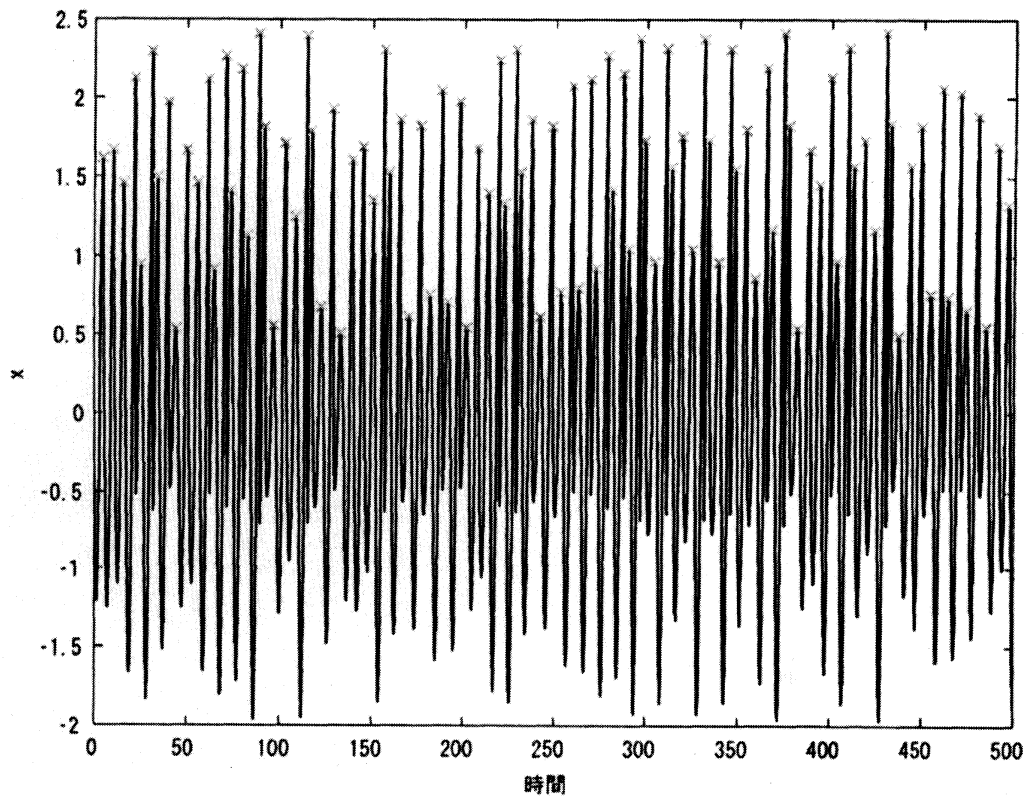


図5 : Rossler モデルから生成した時系列データ(黒線)とその極大値の点過程(灰色の×)。点過程データに関しても、例を示す。ここでは、図5で示すような Rossler モデルの時系列データから極大値を取る時刻とその値を点過程として取り出す。元のサンプリング間隔が一定の時系列データから求めたリカレンスプロットと極大値の点過程から求めたリカレンスプロットを図6で比較する。この2つのリカレンスプロットはとても似た性質を持っていることがわかる。どちらのリカレンスプロットも 3.2 節で紹介した手法を使うと、Devaney のカオスと consistent であることが示せる。このように、マーク付き点過程の距離を用いたリカレンスプロットは、点過程時系列データから元のダイナミクスの性質を調べる上で強力な武器となり得る。

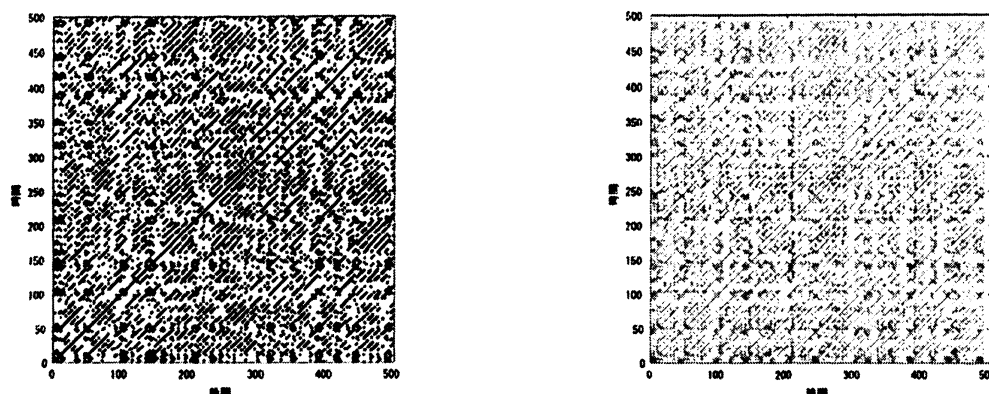


図6：Rossler モデルの連続値の時系列データから求めたリカレンスプロット(左)と極大値の点過程から求めたリカレンスプロット(右)。ここでは、どちらのリカレンスプロットに対しても、それぞれインデックスの値に依存しない一定のしきい値を使用した。

7. 結び

リカレンスプロットは、時系列を視覚化する以上に、背後の力学系の性質を特徴づける上で大変有益な道具である。特に、点過程データなど既存の時系列解析手法では取り扱うことが難しい対象に対して、広く利用できる方法論を与える可能性がある。この解説記事が今後のリカレンスプロットを用いた時系列解析手法の発展のきっかけになれば幸いである。

なお、3 節の全部、また、1 節、4 節、5 節と 6 節の一部が我々の貢献である。

謝辞

本研究は、科研費若手研究(B) 21700249 の助成を受けたものである。

参考文献

- M. C. Casdagli: Recurrence plots revisited, *Physica D*, 108, 12–44 (1997).
- R. L. Devaney: *An Introduction to Chaotic Dynamical Systems*, Addison-Wesley, Reading, Massachusetts, 1989.
- J.-P. Eckmann, S. Oliffson Kamphorst, D. Ruelle: Recurrence plots of dynamical systems, *Europhysics Letters*, 5, 973–977 (1987).
- P. Faure, H. Korn: A new method to estimate the Kolmogorov entropy from recurrence plots: its application to neuronal signals, *Physica D*, 122, 265–279 (1998).
- R. Hegger, H. Kantz, L. Matassini, T. Schreiber: Coping with nonstationarity by

overembedding, *Physical Review Letters*, **84**, 4092–4095 (2000).

- Y. Hirata, S. Horai, K. Aihara: Reproduction of distance matrices from recurrence plots and its applications, *European Physical Journal – Special Topics*, **164**, 13-22 (2008).
- Y. Hirata, K. Aihara: Representing spike trains using constant sampling intervals, *Journal of Neuroscience Methods*, **183**, 277-286 (2009).
- Y. Hirata, K. Aihara: Devaney's chaos on recurrence plots, *Physical Review E*, **82**, 036209 (2010a).
- Y. Hirata, K. Aihara: Identifying hidden common causes from bivariate time series: A method using recurrence plots, *Physical Review E*, **81**, 016203 (2010b).
- Y. Hirata, K. Aihara: Statistical tests for serial dependence and laminarity on recurrence plots, *International Journal of Bifurcation and Chaos*, in press, (2011).
- N. Marwan, M. Thiel, N. R. Nowaczyk: Cross recurrence plot based synchronization of time series, *Nonlinear Processes in Geophysics*, **9**, 325–331 (2002).
- N. Marwan, N. Wessel, U. Meyerfeldt, A. Schirdewan, J. Kurths: Recurrence plot based measures of complexity and its application to heart rate variability data, *Physical Review E*, **66**, 026702 (2002).
- N. Marwan, M. C. Romano, M. Thiel, J. Kurths: Recurrence plots for the analysis of complex systems, *Physics Reports*, **438**, 237–329 (2007).
- N. Marwan, J. F. Donges, Y. Zou, R. V. Donner, J. Kurths: Complex network approach for recurrence analysis of time series, *Physics Letters A*, **373**, 4246–4254 (2009).
- M. C. Romano, M. Thiel, J. Kurths, W. von Bloh: Multivariate recurrence plots, *Physics Letters A*, **330**, 214–223 (2004).
- J. Stark: Delay embeddings for forced systems. I. Deterministic forcing, *Journal of Nonlinear Science*, **9**, 255-332 (1999).
- S. Suzuki, Y. Hirata, 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, *International Journal of Bifurcation and Chaos*, **20**, 3699-3708 (2010).
- M. Tanio, Y. Hirata, H. Suzuki: Reconstruction of driving forces through recurrence plots, *Physics Letters A*, **373**, 2031–2040 (2009).
- M. Thiel, M. C. Romano, J. Kurths: How much information is contained in a recurrence plot?, *Physics Letters A*, **330**, 343-349 (2004a).
- M. Thiel, M. C. Romano, P. L. Read, J. Kurths: Estimation of dynamical invariants without embedding by recurrence plots, *Chaos*, **14**, 234–243 (2004b).
- J. Victor, K. Purpura: Metric-space analysis of spike trains: theory, algorithms and

application, *Network* 8, 127-164 (1997).

- C. L. Webber Jr., J. P. Zbilut: Dynamical assessment of physiological systems and states using recurrence plot strategies, *Journal of Applied Physiology*, 76, 965-973 (1994).
- J. P. Zbilut, C. L. Webber Jr.: Embeddings and delays as derived from quantification of recurrence plots, *Physics Letters A*, 171, 199-203 (1992).
- J. P. Zbilut, A. Giuliani, C. L. Webber Jr.: Detecting deterministic signals in exceptionally noisy environments using cross-recurrence quantification, *Physics Letters A*, 246, 122-128 (1998).