



カオス時系列解析とその応用

Chaos Time Series Analysis and Its Applications

(株) あいはら
研究開発チーム チームリーダー

高橋 純 Jun Takahashi

同上 研究開発員

山田泰司 Taiji Yamada

東京大学
生産技術研究所 教授

合原一幸 Kazuyuki Aihara

1 はじめに

時間と共に変動する現象を時系列信号として観測し、その性質を解析することを目的とする「時系列解析」は、様々な分野で必要とされる重要な技術である。その解析技術として、周波数解析などの線形理論に立脚した確率統計的解析手法や確率的モデリング手法などが重要な役割を担ってきた。そして、それらの手法のもとでは、システムの複雑な振舞いは、「確率的なオリジンを有するもの」と解釈される。

複雑な振舞いをする系に対する全く異なる視点として、「決定論的カオス」がある。決定論的カオスとは、システムに確率的な要因が存在しないにも関わらず、その決定論的システム自体が持つ「非線形性」によって複雑で長期予測不可能な振る舞いが生じる現象をいう。決定論的カオスという概念は、複雑な現象が確率的な現象ではなく、決定論に従った現象かもしれないという新たな視点を与えることとなった。そして、以前は単なるノイズと見なされ捨て去られていたような乱雑な振舞いから、システムの決定論的な仕組みが見出される可能性を与えてくれる概念でもある。

決定論的カオスの特徴として重要なものに、アトラクタの幾何学的特徴である自己相似性、動力学的特徴である軌道不安定性、情報論的特徴である短期予測可能性と長期予測不能性などがある¹⁾。軌道不安定性は「初期値に対する鋭敏な依存性」、「バタフライ効果」とも呼ばれ、ごくわずかな初期値の誤差が時間と共に通常指数関数的に拡大され、短期間のうちに対象システムのグローバルスケールにまで拡大されてしまう性質をいう。一方でカオスは、このような軌道不安定性を生じて、状態空間におけるアトラクタの定常的構造は統計的に安定であるという統計的安定性も併せ持つ。

ところで、実世界で何か現象を観測するとき、無限の精度での観測は不可能である。そのため、システムがカオス的であるなら、その特徴である軌道不安定性によって、たとえ対

象のシステムダイナミクスが完全に既知だとしても、長期の予測は不可能ということになる。ただし、カオスは決定論に従って振る舞うので、軌道不安定性の影響が現れるまでの短期であれば、決定論的予測が可能となる。これがカオスの短期予測である。

本稿では、以上のような決定論的カオス理論を基とした「カオス時系列解析²⁾」について解説する。

2 カオス時系列解析

あるシステムを解析するにあたって、そのシステムのすべての状態変数を観測できることはほとんどなく、最悪の場合状態ベクトルから写像(観測器)によって変換された1変数のみしか観測できないという場合も少なくない。このような場合、1変数の観測時系列データから「時間遅れ座標系への変換(埋め込み)」によってアトラクタを再構成することが行われる。再構成されたアトラクタは「埋め込み(Embedding)」が成立していれば、もとのアトラクタと微分同相となり、以下に述べるリアプノフスペクトラムやフラクタル次元などが位相的に保存される。つまり、たとえ1変数しか観測できなくとも、時間遅れ座標系において再構成されたアトラクタで様々な解析を行うことができるということになる。

この埋め込みを実現するための時間遅れ座標系への変換は、具体的には次のように行う。観測された1変数の時系列データを $x(t)$ ($x(t) \in \mathbb{R}$, $t=0, 1, \dots, N-1$) とする。ここで、適当な遅れ時間 τ ごとの d 個の時系列データを取り出し、それから d 次元ベクトル $\mathbf{X}(t) = (x(t), x(t+\tau), \dots, x(t+(d-1)\tau))$ ($\mathbf{X}(t) \in \mathbb{R}^d$, $t=0, 1, \dots, N-1-(d-1)\tau$) を構成する。これが、再構成された d 次元のアトラクタとなる。ここで、 d は再構成次元、 τ は再構成ラグと呼ばれる。

軌道不安定性を表す特徴量は「リアプノフスペクトラム」によって表される。また、アトラクタの幾何学的構造の自己

相似性は「フラクタル次元」によって定量化される。これらの特徴量を求めるアルゴリズムとして、リアプノフスペクトラムの推定には「ヤコビアン行列推定法」、フラクタル次元の推定には「相関積分法」がよく使われる。また、比較的シンプルな決定論的予測手法として「 n 近傍点による局所線形予測」や「ヤコビアン行列推定による決定論的非線形予測」などがあるが、これらのアルゴリズムに共通していることとして、計算の過程で2点間距離が用いられることがあげられる。

このようにカオス時系列解析において基盤となる2点間距離の特性を、2次元平面上で視覚化する解析手法にリカレンスプロットがある³⁻⁵⁾。リカレンスプロットは、一般に再構成ベクトルの時系列 $\mathbf{X}(t)$ ($\mathbf{X}(t) \in \mathbb{R}^d$, $t=0, 1, \dots, n$) について、時刻 i のときの状態 $\mathbf{X}(i)$ と時刻 j のときの状態 $\mathbf{X}(j)$ との間の距離 $|\mathbf{X}(i) - \mathbf{X}(j)|$ によって、2次元平面上の点 (i, j) の状態を決定する。その表現方法は様々なものが提案されているが、よく使われるものの一つに、あるしきい値を設定し、距離 $|\mathbf{X}(i) - \mathbf{X}(j)|$ がそのしきい値以下であったときに、 (i, j) に黒点をプロットするというものがある。

リカレンスプロットは2点間の相関に距離の情報、時間経過に2次元平面を用いて表現するので、プロットした結果得られる空間的なパターンには、周期的な時系列データであれば周期的な空間パターンが現れ、周期的でなければ非周期的な空間パターンが現れる。また、定常な時系列であればリカレンスプロットの空間パターン全体の質感が一様になり、逆に一様でないなら、対象とする時系列データは、非定常である、過渡状態にある、観測時間がシステムの現象を大局的に捉えるには短過ぎる、などであることが示唆される。このように、リカレンスプロットは、その空間パターンから対象時系列データの特徴を定性的に視覚化できる解析法である。

リカレンスプロットは2点間の距離によって解析を行うが、この距離の定義を変更することにより、様々な新しい解析の可能性が出てくる。例えば、同方向性リカレンスプロット⁶⁾は距離の定義を $|\mathbf{X}(i+T) - \mathbf{X}(i) - (\mathbf{X}(j+T) - \mathbf{X}(j))|$ としている。つまり、同じ方向に変化する点同士の距離は小さく、逆方向に変化する点同士の距離は大きくなる。これを、通常のリカレンスプロットと組み合わせ、それぞれの結果の共通集合をとると、2点間の距離が近く、かつ同じ方向に変化しようとする点の集合が得られる。これは同方向的近傍プロット⁶⁾と呼ばれ、決定論的力学系の一つの特性である「近いところにある点は、次の時刻でもやはり近いところに推移している」こと、すなわち決定論性を視覚的に表現していることになる。図1は、決定論的なカオスシステムであるエノン写像のデータと、確率的な正規乱数データについて、リカレンスプロット、同方向性リカレンスプロット、同方向的近

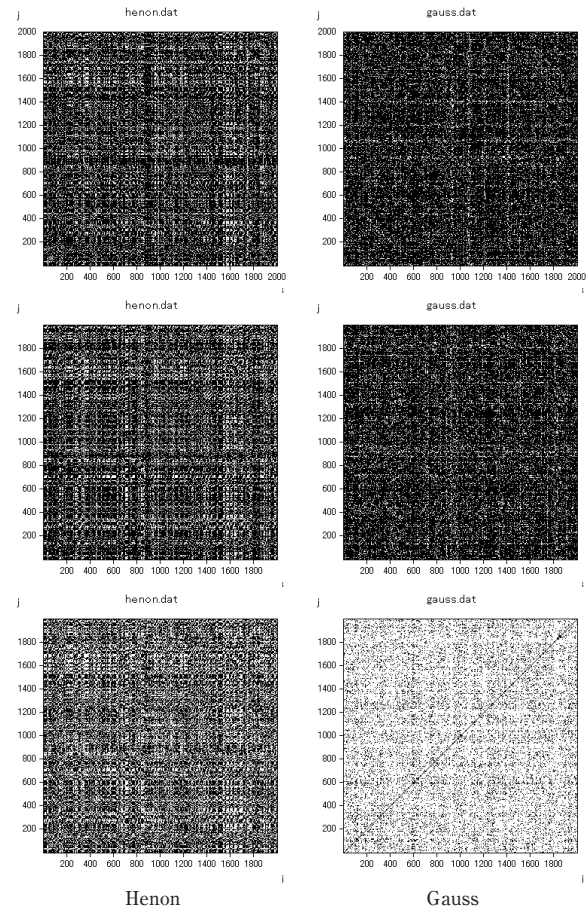


図1 上からリカレンスプロット、同方向性リカレンスプロット、同方向的近傍プロット
左：決定論の時系列データ（エノン写像）、
右：確率的乱数時系列データ

傍プロットを行った結果であり、決定論性が画像の濃度に現れている。

3 ウェーブレット係数列による高炉データの予測

複雑な現象は決して特別なものではなく、世のいたるところに遍在しており、その系から生じる時系列データの解析は、現象の解明さらには解析結果を利用した予測や制御などへの応用を可能とするものである。

例えば、鉄鋼業におけるシンボリック的存在である高炉を例にしよう。高炉のステーブ冷却水の給排水温度差変動の一例を図2に示す⁷⁾。このような変動は、高炉内の複雑な反応の様子を反映していると考えられる。図3は、ステーブ冷却水給排水温度差変動を次元3、ラグ5で再構成したアトラクタである。また、図4はそのリカレンスプロットの結果である。これを見ると、その空間パターンは全体の質感が一様ではなく、これはデータの非定常性を反映している。文献7)においても、この時系列全体では定常ではないことが示されて

いる。

さて、この高炉時系列データをそのまま単純な決定論的非線形予測手法に適用しても、再構成状態空間内で明確なアトラクタが見出せていないので、適切な予測は望めそうにない。そこで本稿では、「ウェーブレット係数列による予測法⁸⁾」の適用を試みる。この手法は低周波域と高周波域が異なる法則に従うシステムの予測などに有効な手法である。

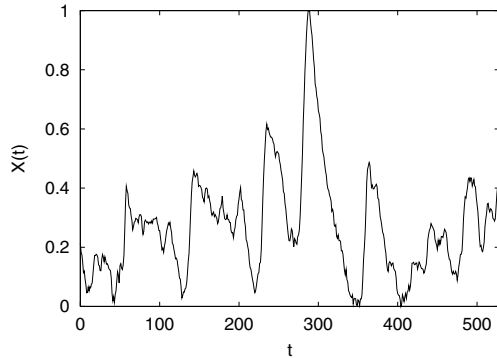


図2 ステープ冷却水の給排水温度差変動の例

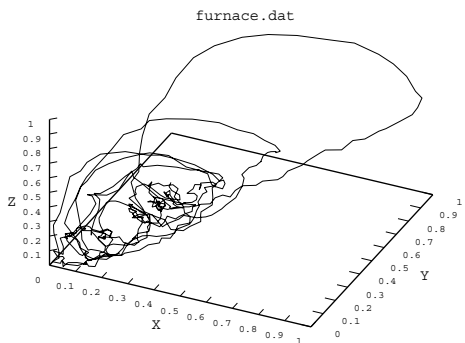


図3 ステープ冷却水の給排水温度差変動を再構成次元3、ラグ5で再構成したアトラクタ

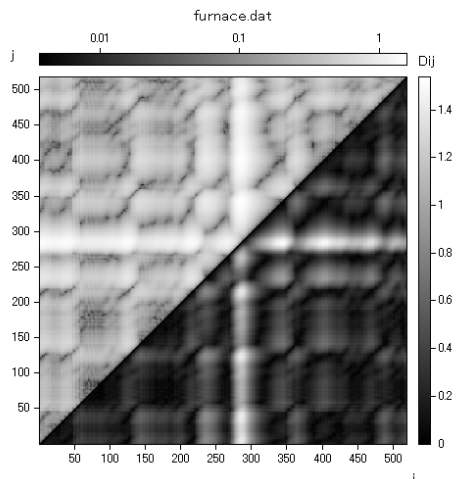


図4 ステープ冷却水の給排水温度差変動のリカレンスプロット

まず、時系列データ $x(t)$, $t=0, 1, \dots, N-1$ を (2進) ウェーブレット変換で時間域で局在した周波数成分の和の形に分解し、各周波数成分の係数列 $d_j(k)$, $c_R(k)$ をそれぞれ時間遅れ座標系へ変換したもののから予測値 $\hat{d}_j(k)$, $\hat{c}_R(k)$ を算出する。これらを (2進) ウェーブレット逆変換して元の時系列に対する予測値の時系列データ $\hat{x}(t)$ を構成する⁸⁾。ここで、スケール j は対数周波数、 $d_j(k)$ はスケール j の周波数成分時系列、 $c_R(k)$ は $j < R$ の $d_j(k)$ をまとめた低周波数成分時系列である。採用する予測手法は各周波数成分ごとの特性に応じて確率的予測法か決定論的予測法かを検討すべきであるが、本稿では単純な n 近傍点による局所線形予測を用いる。また、高炉時系列データの始めの256点を予測データベースとして用いる。

図5は2進ウェーブレット変換により得られた各周波数成分の係数列である。これにより元の時系列データが低周波域と高周波域に分解されている様子がわかるが、まずはじめに各係数列のリカレンスプロットの様相を観測してみよう。図6の各リカレンスプロットをみると、高周波域のリカレンスプロットの質感が比較的一様になっている一方で、低周波域のリカレンスプロットには周期的な様相が現れている。

図7は高炉時系列データ256点から32期先までを、ウェーブレット変換を用いる方法と2進ウェーブレット変換を用いる方法により予測した結果である。これをみると、どちらの方法でも20期先までは良好な予測であるが、30期先にみられる(元データで最大ピーク値をとる)変動を予測するには至らなかった。これは、その変動がそれに至るまでの時間には見られない現象であることを示唆するが、少なくとも短期予測可能性があるという結果は文献7)と同様である。

高炉データのウェーブレット係数列をみても、各周波数成分間の相関が見受けられる。その情報を抽出し予測アルゴリズムに導入できれば、より有効な時系列予測とシステム同定を実現できる可能性があり、今後の課題である。

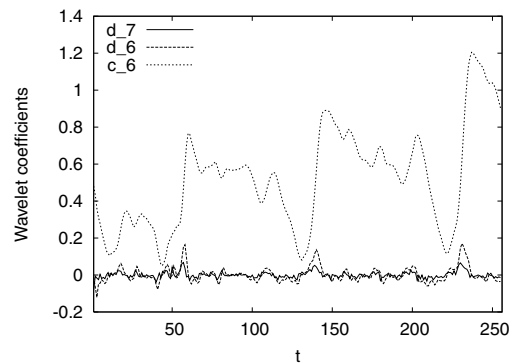


図5 高炉データを2進ウェーブレット変換 ($R=6$) により分解した係数列 ($N=256$)
実線: $j=7$ 周波数成分、破線: $j=6$ 周波数成分、点線: 低周波数成分。解析手法の詳細は文献8)を参照。

4 リカレンスプロットの ゲノムシーケンス解析への応用

リカレンスプロットはカオス時系列解析において基本となる手法であるが、その手法自体はシンプルでわかりやすいものであるため、他の解析への応用も可能である。

例えば、ゲノムシーケンスは4種類の塩基 (A, T, G, C) の並びで構成されている。これを、一種の記号時系列と捉え、それに対してリカレンスプロットを適用する。まず、塩基配列から配列ベクトルを構成する。長さ N の塩基配列を $X_1, X_2, \dots, X_N$ とする。 X_i は一塩基を表し、A, T, G, C のいずれかである。ここで、次元を d 、ラグを τ とし、配

列ベクトル $\mathbf{x}_i = (X_i, X_{i+\tau}, \dots, X_{i+(d-1)\tau})$ を構成する。

リカレンスプロットは2点間の距離によって点をプロットするが、対象が通常の時系列データの場合、ユークリッド距離がよく使われる。この距離の定義を変えることによって、リカレンスプロットの結果から得たい情報を変えることができる。例えば、記号列間に定義される距離の一つにハミング距離がある。これは、同一長の記号列 $\mathbf{u} = u_1 u_2 \dots u_n$ と $\mathbf{v} = v_1 v_2 \dots v_n$ の対応する位置にある n 個の記号対 u_j と v_j ($j=1, 2, \dots, n$) が、 $u_j \neq v_j$ となる数で定義される。配列ベクトル $\mathbf{x}_i$ は次元 d 分の長さがあるので、そのハミング距離は最小0から、最大 d までの値をとる。

ハミング距離の定義から自明であるが、これを距離とした場合、記号が一致している数が多いほど距離は小さくなる。これを利用したアプリケーションの一つとして、ゲノムシーケンスにおけるタンデム・リピート⁹⁾の検出が考えられる。タンデム・リピートとは、ある長さの塩基配列が繰り返し連続して現れる状態をいう。タンデム・リピートは、リカレンスプロットでは、対角線に平行に連続してプロットされた線が等間隔に繰り返し並ぶようなパターンとして現れることが予想される。図8は出芽酵母ゲノム¹⁰⁾の一部を解析した結果で、上記のようなパターンが現れている部分を拡大して表示したものである。実際に、その部分の塩基配列を取り出してみると1周期が72塩基で、繰り返し回数が6回のタンデム・リピートになっていることが確認された。

記号時系列データ解析のための距離として様々な定義が考えられるが、例えばハミング距離において比較する記号対を逆方向からとするようなことも考えられる。この場合、 $u_j \neq v_{n+1-j}$ となるものの数という定義になる。また、対象がゲノムシーケンスであることから、単純な比較ではなく、記

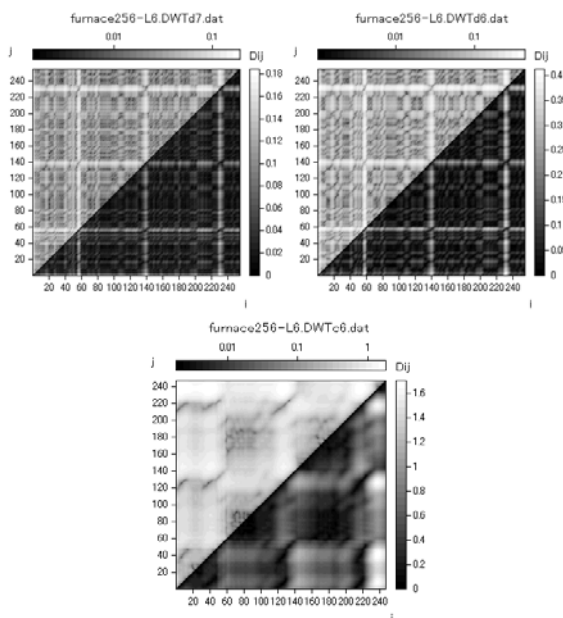


図6 高炉データを2進ウェーブレット変換 ($R=6$) により分解した系列それぞれを再構成次元3で時間遅れ座標系変換したベクトルのリカレンスプロット ($N=256$)
順に $j=7$ 周波数成分 (ラグ1)、 $j=6$ 周波数成分 (ラグ1)、低周波数成分 (ラグ5)

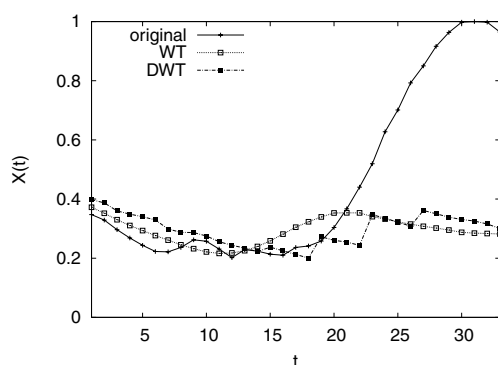


図7 高炉データをウェーブレット変換により分解した系列それぞれの非線形予測系列をウェーブレット逆変換して得られた予測値 ($N=256$, $M=32$, $n=32$)
+: オリジナル, $\square$: ウェーブレット変換 ($R=8$)、 $\blacksquare$: 2進ウェーブレット変換 ($R=6$)

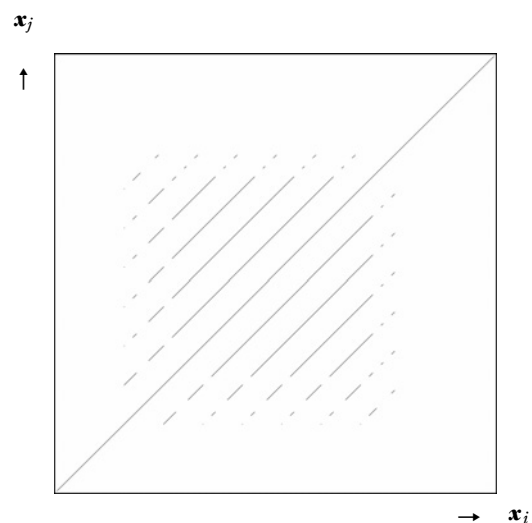


図8 出芽酵母ゲノムのタンデム・リピートの検出 (10番染色体713700~714500)

号対が相補対で無いものの数とする定義が考えられる。相補対とは塩基においてAとT (U)、GとCの組み合わせの対をいう。これらを組み合わせて、 u_j と v_{n+1-j} が相補対でないものの数を「逆方向相補対距離」として定義する。これにより、ゲノムシーケンスにおけるステム・ループ構造の検出が可能となることが期待される。ステム・ループ構造とは、RNAの塩基配列が、部分的に相補対逆繰り返し状態となっている場合で、相補対の配列による分子内二重らせんのステム構造と、それ以外の一本鎖によるループ構造で構成される¹¹⁾。タンデム・リピートは同じ方向に向かって一致している塩基が配列されているが、ステム・ループ構造では相補対が互いに逆方向に配列されている。このことから、ステム・ループ構造は、リカレンスプロットでは、対角線に直角に連続してプロットされた線として現れると考えられる。図9は出芽酵母ゲノム¹⁰⁾の一部を解析した結果で、上記のようなパターンが現れている部分を表示したものである。実際に、その部分の塩基配列を取り出してみるとステム・ループ配列構造になっていることが確認された。

5 おわりに

カオス時系列解析について、その概要をリカレンスプロットを中心に述べてきた。本稿では、リカレンスプロットの定性的な解析について述べてきたが、これを定量的な解析とする研究も進められている^{6,12)}。

また、新しい解析技術として、ウェーブレット係数列を用いたカオス時系列の予測法と、リカレンスプロットのゲノムシーケンス解析への応用について紹介した。これらを含めて、カオス時系列解析は、新しい理論や解析技術が次々と提

案されており、その応用も含めて今後ますます発展していくことが期待されている。

謝辞

ウェーブレット係数列を用いたカオス時系列の予測手法について御助言いただいた増田直紀博士、ゲノム解析について御助言いただいた市瀬夏洋博士に感謝致します。

参考文献

- 1) 合原一幸：カオス学入門，放送大学教育振興会，(2001)
- 2) 池口徹，山田泰司，小室元政：カオス時系列解析の基礎と応用，合原一幸編，産業図書，(2000)
- 3) J. P. Eckmann, S. O. Kamphorst and D. Ruelle : Recurrence plots of dynamical systems, Europhysics Letters, 4 (1987) 9, 973.
- 4) M. Koebbe and G. Mayer-Kress : Use of recurrence-plots in analysis of time-series data, Nonlinear modeling and forecasting, ed. by M. Casdagli & S. Eubank, Addison-Wesley, (1992), 361.
- 5) M. C. Casdagli : Recurrence plots revisited, Physica D, 108 (1997), 12.
- 6) 寶来俊介，山田泰司，合原一幸：同方向性リカレンスプロットによる決定論性解析，電気学会論文誌C, 122-C (2002) 1, 141.
- 7) T. Miyano, S. Kimoto, H. Shibuta, K. Nakashima, Y. Ikenaga and K. Aihara : Time series analysis and prediction on complex dynamical behavior observed in a blast furnace, Physica D, 135 (2000) 3-4, 305.
- 8) 増田直紀，合原一幸：ウェーブレット係数列を用いたカオス時系列の予測，電子情報通信学会誌A, J82-A (1999) 11, 1710.
- 9) T. Yoshida, N. Obata and K. Oosawa : Color-coding Reveals Tandem Repeats in the *Escherichia coli* Genome, J. Mol. Biol., (2000), 298, 343.
- 10) Dennis A. Benson, Ilene Karsch-Mizrachi, David J. Lipman, James Ostell and David L. Wheeler : GenBank, Nucleic Acids Research, 31 (2003) 1, 23.
<http://www.ncbi.nlm.nih.gov/Genbank/>
- 11) 中村春木，中井謙太：バイオテクノロジーのためのコンピュータ入門，コロナ社，(1995)
- 12) 山田泰司，寶来俊介，高橋純，合原一幸：工学から見たカオスと複雑系，知能と情報，日本知能情報ファジィ学会誌, 15 (2003) 6, 607.

(2004年2月26日受付)

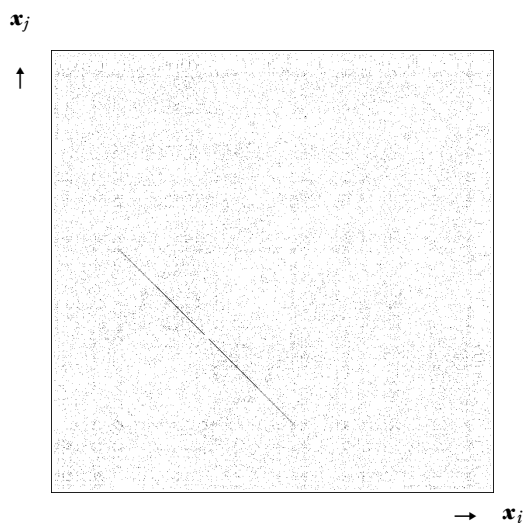


図9 出芽酵母ゲノムのステム・ループ構造の検出
(16番染色体840000～870000)