

時系列データの天底予測のための非線形予測法

鈴木 智也^{†1} 太田 真喜^{†1}

株価などトレンドの強いデータを予測する場合、毎回の変動に注目するよりも、その変動がいつ天底（局所的な極大・極小値）をむかえ、それらがどの程度の値になるのかに注目した方が効率的な場合が多い。しかし、天底を予測するには、どうしても長期予測にならざるをえない。一般的には、非線形予測法を繰り返し適用することで長期予測を実現するが、カオスのような非線形データの場合、予測精度は指数関数的に低下する。そこで本研究では、過去の時系列データの中から客観的に天底を抜き出し、その天底データのみを非線形予測法に利用することで、天底予測を短期予測として取り扱う。これにより、一般的に用いられる反復型の長期予測法と比較して、予測精度などの観点から有用性を期待できる。そこで、数理モデルを用いたシミュレーションにより本手法の有用性を確認し、さらに実データの一例として経済データに対しても本手法を適用した結果、良好な結果を得たので報告する。

Nonlinear Prediction for Top and Bottom Values of Time Series

TOMOYA SUZUKI^{†1} and MASAKI OTA^{†1}

To predict time-series data depending on temporal trends, as stock price and foreign exchange rates, it is often efficient to follow only these top and bottom values more than every time-series datum. Top and bottom values of time series have two important information: The timing and the degree of these values. However, because these top and bottom values do not happen frequently, we have to perform long-term prediction to predict these top and bottom values. In the case of chaotic time series, it is well known that prediction errors grow exponentially by long-term prediction. In the present study, we propose a new prediction method for top and bottom values of time series. We decide top and bottom values numerically, and use only these values for every prediction. By this method, we can consider predictions of top and bottom values as short-term predictions. Moreover, we performed numerical simulations to show the validity of our proposed method with chaotic time series disturbed by observational noise and with foreign exchange rates as real time-series data.

1. はじめに

突然の降水量の急増、大地震の発生、麻疹の患者数の急増など、この実世界にはピークといった突発的な大変動を示す複雑現象が多数存在する。このピークがいつ、どのくらいの規模で発生するのかを事前に予測できれば、対応策を講じる猶予を得ることができる。このようなピークを予測するために最も自然に考えられる方法は、1ステップ予測を反復して行うことで長期予測を実現し、予測された時系列データの中からピークを選ぶ方法である。しかしこの方法は長期予測であるがゆえ、反復における各予測での誤差が拡大していくという欠点がある。特に、予測対象がカオスのような強い非線形性を有する場合、その予測誤差は指数関数的に拡大していくため予測は困難である。この問題点を鑑み、文献 1)、2) では、過去の全データを予測モデルに用いるのではなく、各ピークの状態値とそれらが発生する時間間隔情報のみを予測モデルに用いることで、短期予測として各ピークを予測する方法を提案している。さらに、ピークを検出する方法として、過去のデータを一定の時間幅で区間分けし、各区間の最大値をピークとして扱うというシンプルな方法を採用している。しかしこの方法では、各区間でピークを1つしか選べない（または、選ばなくてはならない）という制約が生じてしまう。

そこで本研究では、ピークのような極大値や極小値を一定の時間幅にとらわれない方法で決定し、さらに強いトレンドを有するデータ（たとえば、株価や商品の購買量などの流行現象）をも効果的に予測できるように、文献 1)、2) の予測モデルにトレンドを除去する操作を導入した新しい予測手法を提案する。トレンドとは、しばらくの間、変動の傾向が偏ることを意味する。しばらくは上昇傾向であるが、やがて下降傾向に転じるなどといった状況である。このようなデータには局所的な極大値、極小値が多数存在するが、突発的な変動ではないのでピークと呼ぶよりも、株式用語にならない、極大値を天井、極小値を底と呼ぶことにする。この天底がいつ、どの程度の規模で発生するのかに着目すれば、時系列データの変動を大まかに効率良く把握することができる。しかし、毎回の極大、極小値を天底値とすると数が多すぎる場合があるので、どの程度の極大値や極小値から天底と見なすのかについて客観的な決定基準が必要となる。そこで本研究では、まず、パラメトリックに天底の決定基準

^{†1} 同志社大学理工学部情報システムデザイン学科

Department of Information System Design, Faculty of Science and Engineering, Doshisha University

を調節できる方法を紹介する(3章)。

天底変動の予測が役立つ対象として、たとえば株価や為替取引価格などの経済データ、降雪量や気温の年次変動などの自然データ、さらには商品の購買量などの流行を反映した社会データがある。経済データに関しては、一般の投資家は日中チャートを眺め続ける時間的余裕はないので、短期的にこまめに投資行動を起こすよりも、長期的な大きなトレンド、つまり天底に着目した方が落ち着いて投資行動を起こせるであろう。さらに、自然データにおける天底値とは、災害を意味する場合が多い。その災害がいつ起こるのかを予測することによって、事前に対策を講じる猶予ができる。また、購買量の天底値を予測できれば、それに合わせて製品の製造量を決定できるので、品切れや売れ残りの削減を期待できる。しかし現状では、天底の予測自体を高精度でできる見込みは少なく、予測手法を改良しつつ予測精度を向上させていく段階にある。

これらの背景に鑑み、本研究では、先に述べた反復型の長期予測による天底予測法に対して、新しい天底予測法を提案する(4章)。さらに、この提案法の有用性を分析するために、予測精度などの観点から反復型長期予測法と比較する(5章)。

2. 非線形予測モデル

本研究では、多くの自然現象の本質は非線形性にあることに鑑み、予測モデルとして非線形予測法を用いる。さらにその中でも比較的拡張がしやすい局所線形近似法を基盤とし、天底予測への応用法を議論する。

2.1 局所線形近似法

局所線形近似法³⁾は、カオス時系列解析⁴⁾の1手法として広く利用されている。この方法は、対象データの近傍点の将来変動を参照することで、対象データの未来を予測する。その際、あらかじめ観測時系列データ $x(t)$ ($t = 1, 2, \dots, T$) を多次元空間内の軌道 $V(t)$ として表現する必要があるため、次式の埋め込み定理^{5),6)}を用いる。

$$V(t) = \{x(t), x(t-\tau), x(t-2\tau), \dots, x(t-(d-1)\tau)\} \quad (1)$$

ここで、 τ は遅れ時間、 d は埋め込み次元である。埋め込み定理^{5),6)}によって、観測時系列データを多次元ベクトルに再構成し、多変量システムの振舞いを再現するものである。軌道 $V(t)$ の末端 $V(T)$ を予測対象点とし、以下の手順によって将来値を予測する。

STEP 1 近傍点の探索

予測対象点 $V(T)$ とのユークリッド距離が近い近傍点を、過去の軌道より K 個探索す

る。これらの近傍点を $V(T_k)$ ($k = 1, 2, \dots, K$) とする。

STEP 2 近傍点の振舞いをもとに予測値を決定

近傍点の1ステップ後の平均値を、 $V(T)$ の1ステップ予測値 $\tilde{V}(T+1)$ とする。以後、予測値には“ \sim ”をつける。つまり、 $\tilde{V}(T+1)$ は以下のように書ける。

$$\tilde{V}(T+1) = \frac{1}{K} \sum_{k=1}^K V(T_k+1) = \{\tilde{x}(T+1), \dots\} \quad (2)$$

このように、時系列データの1ステップ予測値 $\tilde{x}(T+1)$ を得ることができる。

2.2 最適な遅れ時間 τ 、埋め込み次元 d の決定法

精度の高い予測を行うには、式(1)の軌道 $V(t)$ を適切に構成することが不可欠である。まず、式(1)の各要素は状態空間の次元に対応するので、互いに直交することが望ましい。よって、予測対象の時系列データ $x(t)$ の自己相関関数が最初にゼロクロスする時間(ゼロクロスしない場合には最初に極小値をとる時刻)を最適な遅れ時間 τ として用いるのが一般的である⁴⁾。

次に、最適な埋め込み次元 d の決定法として、FNN法(False Nearest Neighbor法)^{7),8)}が有用である。この手法は、軌道 $V(t)$ の近傍点群 $V(t_k)$ が、埋め込み次元 d を増加させたときに近傍でなくなる場合、それらの $V(t_k)$ を誤り近傍点と見なす。もし、埋め込み次元が最適値以上であれば、この誤り近傍点は存在しないはずなので、最適な埋め込み次元を決める指針となる。

まず、軌道 $V(t)$ とその最近傍点 $V(t_1)$ のユークリッド距離を、 d 次元状態空間内の場合に $D_d(t)$ 、 $d+1$ 次元状態空間内の場合に $D_{d+1}(t)$ とすると、これらの相対距離 $D_L(t)$ を次式で表す。

$$D_L(t) = \sqrt{\frac{D_{d+1}^2(t) - D_d^2(t)}{D_d^2(t)}} \quad (3)$$

文献7)と同様に本研究においても、 $D_L(t)$ が15以上のときに点 $V(t_1)$ を $V(t)$ の誤り最近傍点と見なし、 $t = 1 \sim T$ において誤り最近傍点の個数が0になるように最適な埋め込み次元 d を定めた。

さらに、式(2)の近傍点の個数 K について、近傍点は局所性を保ちつつ予測対象点 $V(T)$ を包囲することが望ましいので、 $K = d + 1$ とした。これにより、近傍点の偏りを防ぎ、

データに含まれるノイズを打ち消す効果を期待できる。

3. 天底の決定法

時系列データにおける天底とは、データの一時的な流れが上昇から下降に転じたときを天井と呼び、逆に下降から上昇に転じたときを底と呼ぶ。しかし、時系列データが正弦波のようになめらかでない場合、すべての極大・極小値を天底とすると、時系列データのほとんどが天底となってしまう。そこである程度の閾値を設けて天底を厳選する必要がある。

まず天底変動の規模を表す指標として、天井値と底値の差をそれらの平均値で標準化することにより定量化する。つまり、天井値を a 、底値を b とおくと、天底変動の規模 λ は次式となる。

$$\lambda = \frac{2(a-b)}{a+b} \quad (4)$$

ただし、 $a \geq b \geq 0$ である。さらに、式 (4) を変形すると、

$$\begin{cases} a = \frac{2+\lambda}{2-\lambda}b \\ b = \frac{2-\lambda}{2+\lambda}a \end{cases} \quad (5)$$

となる。ここで、 i 番目に天井または底となる時刻を t_i とすると、たとえば $x(t_i)$ が底値であれば $b = x(t_i)$ とし、次の天井値を $a = x(t_{i+1})$ とおくことで、 λ の変動幅を持つ天底の関係を定めることができる。しかし、実際には毎回の天底の幅は一定ではないので、次式を天底の決定条件とする。

もし $x(t_i)$ が底値であれば、次の天井値 $x(t_{i+1})$ は、

$$x(t_{i+1}) \geq \frac{2+\lambda}{2-\lambda}x(t_i) \quad (6)$$

を満たし、もし $x(t_i)$ が天井値であれば、次の底値 $x(t_{i+1})$ は、

$$x(t_{i+1}) \leq \frac{2-\lambda}{2+\lambda}x(t_i) \quad (7)$$

を満たすように決定する。ここで、 $t_{i+1} > t_i$ である。

この手法のメリットは客観的に天底を決められるだけでなく、自由に天底の尺度も決定できる点である。 λ の値を操作することで天底の間隔や変動幅を変えられ、 λ を大きくするほど時間間隔と変動幅が大きくなり、 λ を小さくするほどそれぞれも小さくなる。ただし λ の

とれる値の範囲は時系列データ $x(t)$ の最大値 A 、最小値 B としたとき次式の範囲に限定される。

$$0 < \lambda < \frac{2(A-B)}{A+B} \quad (8)$$

次に、天底の決定基準である式 (6)、(7) を用いて具体的に天底を決めるアルリズムを示す。

STEP 1 λ を設定する

式 (8) の範囲で λ の値を適当に決め、天底の決定基準を設定する。

STEP 2 初期値を仮の天底値とする

天底を決定する時系列データ $x(t)$ の初期値 $x(1)$ を、仮の天底値 $x(t_1)$ とする。ここで t の添字は、天底の順番を表し、 t_1 とは 1 番目の天底が出現した時刻を意味する。 $x(t_1)$ を天井値または底値とするかは、自由に設定してよい。以後、 $x(t_1)$ を仮の天井値とする場合について説明する*1。

STEP 3 $t > t_1$ の範囲から次の底値を決定する

STEP 3-1 底値の候補値を探す

天井値 $x(t_1)$ を式 (7) に代入し、条件を満たす底値 $x(t_2)$ を $x(t_1+1)$ から順に探索する。条件を満たした値は、次の底値の候補 $x(\hat{t}_2)$ として保存しておく*2。もし条件を満たす値がなければ、天底の決定を終了する。

STEP 3-2 候補値から真の底値へ

$t > \hat{t}_2$ の範囲において、条件式 (6) を満たす天井値 $x(t_3)$ を $x(\hat{t}_2+1)$ から順に探索する。

- もし、探索途中で $x(\hat{t}_2)$ 以下の値があれば、それを新しい底値の候補値 $x(\hat{t}_2)$ として更新し、STEP 3-2 の最初に戻る。
- もし、条件式 (6) を満たす値が見つければ、天井値の候補値 $x(\hat{t}_3)$ として保存すると同時に、先の底値の候補値 $x(\hat{t}_2)$ を、真の底値 $x(t_2)$ として決定する。

STEP 4 STEP 3-2 を繰り返す

STEP 3-2 では、 i 番目の天底値が最終決定されると同時に、 $i+1$ 番目の天底値の候補値 $x(\hat{t}_{i+1})$ が選ばれる。そこで再び STEP 3-2 を繰り返すことで、候補値を真の天底値に確定していく。つまり、 $t > \hat{t}_{i+1}$ の範囲において、条件式 (6) または (7) を満たす天

*1 仮の底値とする場合は、STEP 3 中の“天井”と“底”を入れ替えて読んでいただきたい。

*2 天底の候補値には“^”を付けて区別する。

井値 $x(t_{i+2})$ を $x(\hat{t}_{i+1} + 1)$ から順に探索する．STEP 3-2 と同様に，

- もし，探索途中にて $x(\hat{t}_{i+1})$ を更新する値があれば，それを新しい天底の候補値とする．
- もし，天底の条件式を満たす値が見つければ，天底値の候補値 $x(\hat{t}_{i+2})$ として保存すると同時に，候補値 $x(\hat{t}_{i+1})$ を，真の天底値 $x(t_{i+1})$ として決定する．

以上を繰り返すことにより，仮の底値 \rightarrow 天井値 ($i = 2$) \rightarrow 底値 ($i = 3$) \rightarrow 天井値 ($i = 4$) \rightarrow 底値 ($i = 5$) のように交互に天底が決定していき， $t > \hat{t}_{i+1}$ の範囲において上記を満たす値がなければ，天底の決定を終了する．

4. 天底データの予測手法

次の天井値や底値がいつ，どの程度の値になるのかを予測するために，2章で紹介した非線形予測法が利用できる．しかし，本研究で着目する時系列データは強いトレンドを有するので，遅れ時間 τ を決定する際の自己相関関数が短時間ではゼロクロスせず，式 (1) の埋め込み時に τ の値を巨大にせざるをえない場合がよく起こる．この問題を回避するために，データをそのまま埋め込むのではなく，階差をとることでトレンド（低周波）成分を除去しつつ，変動幅として埋め込むことが有効である^{9),*1}．つまり，式 (1) のターケンスの埋め込み定理により，多次元空間内の軌道 $V(t)$ は以下ようになる．

$$V(t) = \{\Delta x(t), \Delta x(t - \tau), \Delta x(t - 2\tau), \dots, \Delta x(t - (d-1)\tau)\} \quad (9)$$

ここで， $\Delta x(t) = x(t) - x(t-1)$ である．さらに，天底は頻繁に発生しないため，一般的に長期予測をせざるをえない．通常，長期予測は1ステップの予測を反復することで実現する⁴⁾．本研究では，これを“反復型長期予測法”と称し，以下の予測アルゴリズムによって天底の将来変動を予測する．

STEP 1 時系列データの前処理

予測対象の時系列データ $x(t)$ ($t = 1, \dots, T$) を階差データ $\Delta x(t)$ に変換することで，データに含まれるトレンド（低周波）成分を除去する．ここで， T は予測開始時刻である．

STEP 2 τ, d の最適値を推定

2.2節で述べた方法を $\Delta x(t)$ に対して適用することにより，式 (9) で用いる τ, d の最適値を推定する．

*1 多くの場合，1階階差をとれば τ の値を十分小さくできる．

STEP 3

STEP 2 で決定した τ, d を用いて， $\Delta x(t)$ から式 (9) の軌道 $V(t)$ に変換する．

STEP 4 1ステップ予測の実行

予測対象点 $V(T)$ を 2.1 節の局所線形近似法のアルゴリズムに適用させることにより1ステップ予測値 $\Delta \tilde{x}(T+1)$ を得る．

STEP 5 予測値に対して1ステップ予測を実行

STEP 4 で得られた予測値を $\Delta x(t)$ ($t = 1, \dots, T$) の最新データとして追加し，新しい時系列データ ($t = 1, \dots, T+1$) に対して再びSTEP3および4を繰り返す．これにより，2ステップ予測値 $\Delta \tilde{x}(T+2)$ が得られる．

STEP 6 長期予測の実行

STEP 5 を P 回繰り返すことで，予測開始時刻 T から P ステップ先までの予測値 $\Delta \tilde{x}(T+p)$ ($p = 1, 2, \dots, P$) を得ることができる．

STEP 7 予測値の後処理

次式より，変動幅の予測値を予測開始点 T 以降の時系列データの振舞いに変更する．

$$\tilde{x}(T+p) = x(T) + \sum_{q=1}^p \Delta \tilde{x}(T+q) \quad (10)$$

STEP 8 長期予測値から天底の予測値を選出

予測に用いた時系列データ $x(t)$ ($t = 1, \dots, T$) に長期予測値 $\tilde{x}(T+p)$ ($p = 1, \dots, P$) を連結した時系列データ ($t = 1, \dots, T+P$) に対して，3章の天底決定アルゴリズムを適用する．予測開始時刻 T の直前の天底値を $x(t_{i'})$ ($t_{i'} = \max\{t_i \leq T\}$) とすると，天底の予測値は $x(t_{i'+1})$ に対応するデータを選出すればよい．以後簡単のため，この予測値を $\tilde{X}_{i'+1}$ と表す．

さらに， $\Delta \tilde{X}_{i'+1} = \tilde{X}_{i'+1} - X_{i'}$ によって天底の変動幅の予測値を求め， $\Delta \tilde{t}_{i'+1} = \tilde{t}_{i'+1} - t_{i'}$ によって天底の時間間隔の予測値を求めることができる．

この手法では，次の天底値 $\tilde{X}_{i'+1}$ が決定されるように予測ステップ数 P を十分大きく設定する必要があるが，予測は長期になるほど困難であるため，多ステップ予測値 $\tilde{x}(T+p)$ の振舞いは乱れていく．たとえば，多ステップの長期予測値の変動が弱まってしまう場合（図1）や発散してしまう場合があり，予測ステップ P を大きく設定しても長期予測値から次の天底の予測値 $\tilde{X}_{i'+1}$ を決定できない場合があるという問題点がある．さらに，多ステップ予測値にかかる計算コストも大きい．

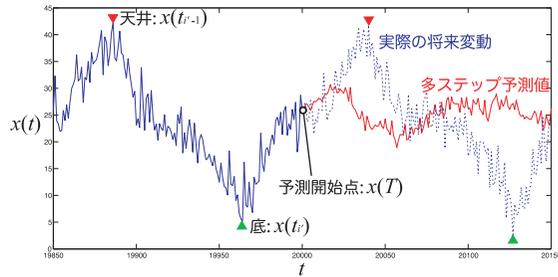


図 1 反復型長期予測法において多ステップの長期予測値が弱まってしまふ様子. この場合, 将来の天底の予測値 $\hat{x}(\hat{t}_{i'+1})$ を決定できない. なお, 5 章で紹介するローレンツ方程式 ($\Xi = 10$ [dB]) の例である
 Fig.1 Figure shows that predicted values of multisteps are weakened by iterating one-step predictions. In this case, we can not predict a next top or bottom value $\hat{x}(\hat{t}_{i'+1})$. This is an example of the Lorenz equation ($\Xi = 10$ [dB]) introduced in Section 5.

次に, 反復型長期予測法のような長期予測を行わずに天底予測を行う手法を提案する. アイデアとして, 時系列データの全要素を用いて予測を行うのではなく, 過去の天底データのみを用いて予測することにより, 次の天底値の予測を 1 ステップ予測として実現する. 以下に予測アルゴリズムを示す.

STEP 1 時系列データの前処理

予測開始時刻を T とすると, 予測対象の時系列データ $x(t)$ ($t = 1, \dots, T$) 中に含まれる天底値 $x(t_i)$ ($i = 1, \dots, i'$) を, 3 章の天底決定アルゴリズムにより決定する. なお, 天底値 $x(t_i)$ を X_i と表す. さらに, 天底データ中に含まれるトレンド成分の除去のために, $\Delta X_i = X_i - X_{i-1}$ によって差分化する.

STEP 2 最適な τ, d を決定

STEP 1 で決定された天底データ ΔX_i に対して, 2.2 節で述べた方法を適用し, 最適な τ, d を決定する.

STEP 3

STEP 2 で決定された τ, d を用いて, ΔX_i に対して, 式 (9) の埋め込み定理を適用する. 予測対象点 $V(i')$ は以下となる.

$$V(i') = \{\Delta X_{i'}, \Delta X_{i'-\tau}, \Delta X_{i'-2\tau}, \dots, \Delta X_{i'-(d-1)\tau}\} \quad (11)$$

STEP 4 1 ステップ予測の実行

予測対象点 $V(i')$ を 2.1 節の局所線形近似法のアルゴリズムに適用することにより, 1 ステップの予測を行うだけで次の天底の変動幅の予測値 $\Delta \tilde{X}_{i'+1}$ を得ることができる.

天底の時間間隔を予測する場合は, STEP 3 において, 予測開始時刻 T より以前に天底が発生した時刻 t_i ($i = 1, 2, \dots, i'$) に対して埋め込み定理を適用すると, 予測対象点 $V(i')$ は以下となる.

$$V(i') = \{\Delta t_{i'}, \Delta t_{i'-\tau}, \Delta t_{i'-2\tau}, \dots, \Delta t_{i'-(d-1)\tau}\} \quad (12)$$

この $V(i')$ に対して STEP 4 を実行することで, 天底の時間間隔の予測値 $\Delta \tilde{t}_{i'+1}$ を得ることができる.

この方法は, 天底値を直接予測することにより天底間をスキップすることができるので, カオスのような軌道不安定性やデータに含まれるノイズにより長期予測が困難なデータに対して有用性を期待できる. また, 計算コストを削減することができる. さらに, 必ず天底の予測値が定まるので, 図 1 に示すような天底の予測値が定まらない問題を解決している. 以上の天底予測手法を“ショートカット予測法”と呼ぶことにする.

本章では, ショートカット予測法の有用性を確認するために, 反復型長期予測法と比較しつつ, 予測精度について検証を行う.

5. 天底予測法の従来法, 提案法の比較実験

5.1 実験方法

予測対象データとして, まずカオスの振舞いを示すローレンツ方程式⁴⁾ をルンゲクッタ法を用いて 40,000 点作成し, 第 1 変数 x のみを用いた. ルンゲクッタ法の時間刻み幅は 0.01 に設定した. なお, ローレンツ方程式は大気変動に関する数理モデルの 1 つであり, 以下のように記述される.

$$\begin{cases} \dot{x} &= -\gamma x + \gamma y \\ \dot{y} &= -xz + rx - y \\ \dot{z} &= xy - bz \end{cases} \quad (13)$$

各パラメータの値は, $\gamma = 10, r = 28, b = 8/3$ とした. さらに, 現実に存在するデータに近づけるために, 次式を用いて正規乱数による観測ノイズを作成し, 元データに付加した.

$$\Xi[\text{dB}] = 10 \log_{10} \frac{\sigma_o^2}{\sigma_\xi^2} \quad (14)$$

ここで、 σ_o^2 は元データの分散、 σ_ξ^2 は観測ノイズの分散である．本実験では、 $\Xi = \infty$ [dB] (観測ノイズなし)、 $\Xi = 10$ [dB]、 $\Xi = 5$ [dB]、 $\Xi = 1$ [dB] のデータを用いた．

さらに実データの 1 例として、円-ドル為替取引価格も予測対象として用いた．このデータは、取引が行われるごとに取引価格 (1 ドルに対する円の価格) が記録されており、1992/10 月中の 40,000 取引データを用いた．

また、天底決定条件の基礎である式 (4) は、正の値どうしの差を評価するものなので、予測対象データを負の値を持たないように変換する必要がある．本研究では、上述の時系列データに対して、最小値を引いたものを予測対象データ $x(t)$ とした．これにより $\min\{x(t)\} = 0$ となるので、式 (8) で示した λ の範囲は、いかなる時系列データにおいても $0 < \lambda < 2$ となる．

各データにおいて、 λ を変動させながら、3 章の方法により天底を決定した様子を図 2 に示す． λ を大きくすると天底の決定基準 (式 (6), (7)) が厳しくなり、天底の個数が少なくなると同時に、天底の変動幅や時間間隔が大きくなるのが分かる．これらの傾向を定量的に示すために、図 3 において、各データの天底の出現率 r 、天底の変動幅の平均値 $\langle |\Delta X_i| \rangle = \langle |X_i - X_{i-1}| \rangle$ 、天底が出現する時間間隔の平均値 $\langle \Delta t_i \rangle = \langle t_i - t_{i-1} \rangle$ の計算結果を示す．天底の出現率 r は、天底の出現回数をデータ長で割ることで算出した．

実際の予測では、予測を行う目的や予測値の特徴を考慮しつつ、ユーザが自由に λ の値を設定すればよい．1 章で述べた例を再掲すると、経済データの予測を行う場合は、どのくらいの時間スケールで取引したいのか (日ごとや月ごと) や、さらには大変動を目論んだ投資にはリスクがともなうので、価格変動幅も考慮しつつ λ の値を設定する必要がある．その際、図 3 のように過去のデータから天底の変動幅や時間間隔の λ 依存性を調べ、予測値の規模を考慮しつつ λ を設定すればよい．自然データを用いて災害などを予測する場合は、大変動を予測することが主であるので λ を大きく設定すればよい．また、商品の購買量を予測しこまめに在庫管理を行いたい場合には、 λ を小さく設定して変動幅や時間間隔の短い予測をする必要がある．このように、各ケースに応じて λ の設定値を可変できる．

次に、図 2 に示されたデータに対して反復型長期予測法およびショートカット予測法を適用し、天底データの予測精度に関する比較を行う．まず、両予測アルゴリズムの STEP 1 で登場する予測開始時刻 T の選び方として、時系列データの後半部分よりランダムに 400 個選出した．後半部分から選出した理由は、予測にはある程度の過去のデータが必要となる

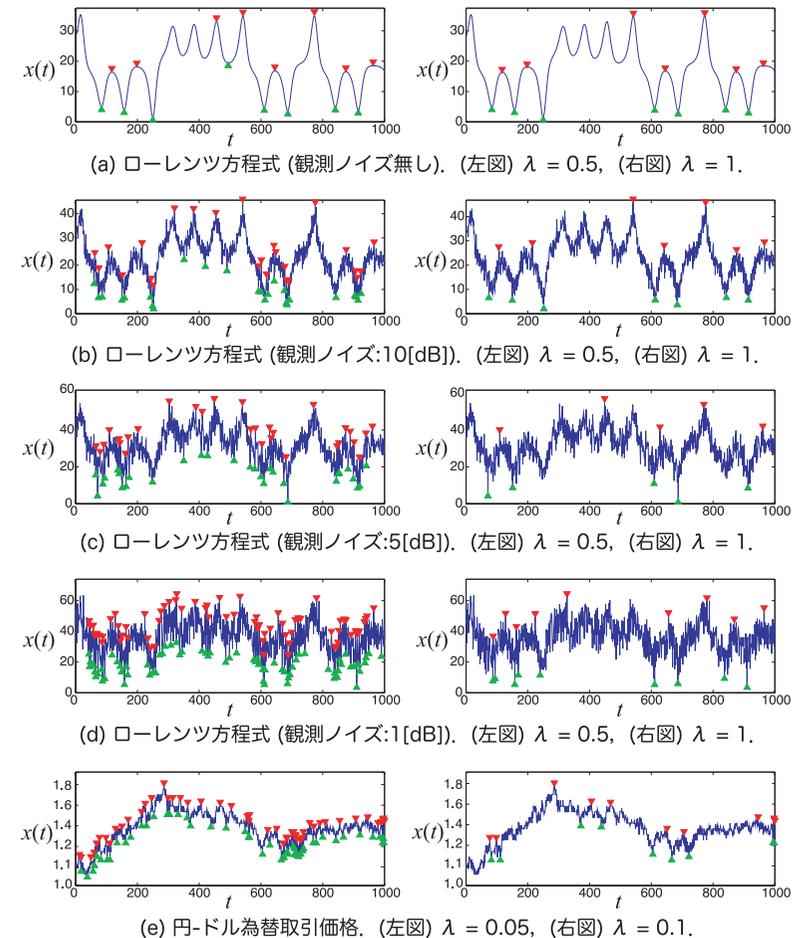


図 2 3 章の方法により決定された天底の様子．下三角印は天井を指し、上三角印は底を指している．ただし、各時系列データ $x(t)$ は前処理として、元データから最小値 $\min\{x(t)\}$ を引いてある．さらに、図 (e) の $x(t)$ の単位は円である

Fig. 2 Top and bottom values decided by the method introduced in Section 3. Each downward triangle means top values, and each upward triangle means bottom values. Here, we subtracted the minimum value $\min\{x(t)\}$ from each time-series data $x(t)$ as a preprocessing. Then, the unit of $x(t)$ in Fig. (e) is Yen.

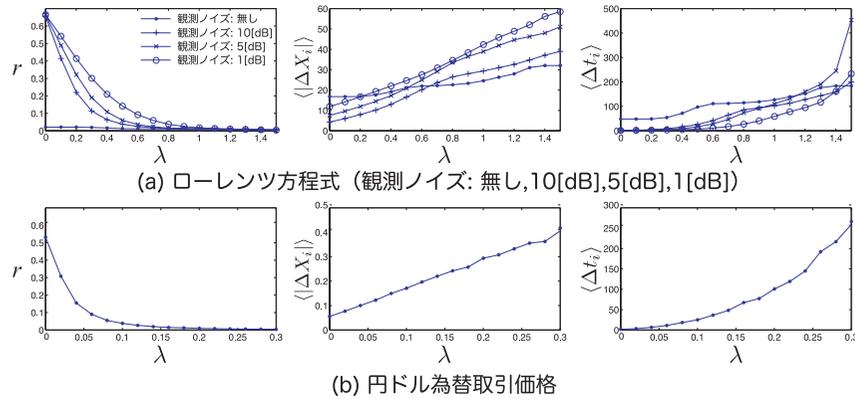


図 3 λ の値に応じた天底の出現率 r , 天底の変動幅の平均値 $\langle |\Delta X_i| \rangle$, 天底が出現する時間間隔の平均値 $\langle \Delta t_i \rangle$
 Fig. 3 The occurrence rate of top and bottom values according to λ . The $\langle |\Delta X_i| \rangle$ means the mean value of movements between top and bottom values, and $\langle \Delta t_i \rangle$ means the mean value of temporal intervals between top and bottom values.

表 1 推定された遅れ時間 τ および埋め込み次元 d の最適値. τ_A および d_A は反復型予測法 (式 (9)) の場合, τ_B および d_B はショートカット予測法 (式 (11), (12)) の場合である. なお, 式 (11) と式 (12) において, 各推定値は同じ値となった

Table 1 Estimated the delay time τ and the embedding dimension d . The τ_A and d_A are optimized for Eq.(9) of the multistep prediction by iterating one-step predictions, and τ_B and d_B are for Eqs. (11) and (12) of the shortcut prediction of our proposed method. Here, these parameters for Eqs. (11) and (12) became the same.

	τ_A	d_A	τ_B	d_B
ローレンツ方程式 ($\Xi = \infty$ [dB])	16	4	1	8
ローレンツ方程式 ($\Xi = 10$ [dB])	1	5	1	8
ローレンツ方程式 ($\Xi = 5$ [dB])	1	5	1	8
ローレンツ方程式 ($\Xi = 1$ [dB])	1	5	1	8
円-ドル為替取引価格	3	10	1	8

からである. 次に, 各予測開始時刻 T に対して, それ以降の全データを用いて反復型長期予測法およびショートカット予測法を適用し, 次の天底の変動幅 ΔX_{i+1} および時間間隔 Δt_{i+1} を予測した.

なお, 各予測法の STEP 2 で推定された遅れ時間 τ や埋め込み次元 d の最適値を表 1 に示す. 各データは STEP 1 において差分化してあるので, τ を小さく設定できる. また, d

に関してもあまり大きな値にはならなかった.

予測精度の評価法として, 以下の相関係数 R と正規化平均二乗誤差 E を用いる.

$$R = \frac{\sum_i (z_i - \bar{z})(\tilde{z}_i - \bar{\tilde{z}})}{\sqrt{\sum_i (z_i - \bar{z})^2} \sqrt{\sum_i (\tilde{z}_i - \bar{\tilde{z}})^2}} \quad (15)$$

$$E = \frac{\sqrt{\sum_i (z_i - \tilde{z}_i)^2}}{\sqrt{\sum_i (z_i - \bar{z})^2}} \quad (16)$$

ここで, z_i は真値, \tilde{z}_i は予測値を表している. さらに, \bar{z} と $\bar{\tilde{z}}$ はそれぞれの平均値である. 相関係数 R は 1 に近いほど予測精度が高く, 振舞いの類似度を重点的に評価する. 正規化平均二乗誤差 E は予測誤差を直接的に評価し, 0 に近いほど予測精度が高い. なお, 正規化平均二乗誤差が 1 以上の場合は, 予測誤差が真値の標準偏差以上に拡大することを意味する.

実用の観点から予測精度を評価する場合, 予測対象に応じて妥当な評価基準が異なると考えられる. たとえば株価予測の場合, 株価変動の絶対的な値よりも上下変動といった変動パターンを予測することが重要である. 将来値の上昇または下降を予測できれば, 変動幅を過小・過大評価したとしても損はしない. つまり株価予測の場合には, 相関係数 R が予測精度を分析するうえで適している. 一方, 商品の購買量などの予測精度を評価する場合, 正規化平均二乗誤差 E が適している. 将来購買量の過小評価は, 在庫切れを引き起こし販売チャンスを失う. 将来購買量の過大評価は, 在庫を無駄に多く製造してしまい, 製造コストの分だけ損をする. つまり予測値と正解値の絶対的な差が損失となる.

本研究では予測対象を特定せず, 一般のデータに対して反復型長期予測法およびショートカット予測法の有用性を評価したいので, 相関係数 R と正規化平均二乗誤差 E の両観点から分析する. なお, 統計的な誤差評価法の域を超え, 具体的な利益や損失をもって予測法の有用性を議論するには, 予測値の活用方法を設定する必要がある. これは強く予測対象に依存し, また予測法とは異なる方法論として議論されるべきなので, 本研究では取り扱わない.

さらに反復型長期予測法では, 図 1 に示すように天底の予測自体ができない場合があるという問題点があるため, 予測ができなかった回数を予測試行回数 (= 400) で割ることににより, 予測失敗率 m を算出する. なお, ショートカット予測法では次の天底値を直接予測することにより, 誤差を含むものの, 必ず予測値が算出される. よって, $m = 0$ である. また, 反復型長期予測法では予測ができなかった場合を取り除いて, 予測精度である R と E を算出した.

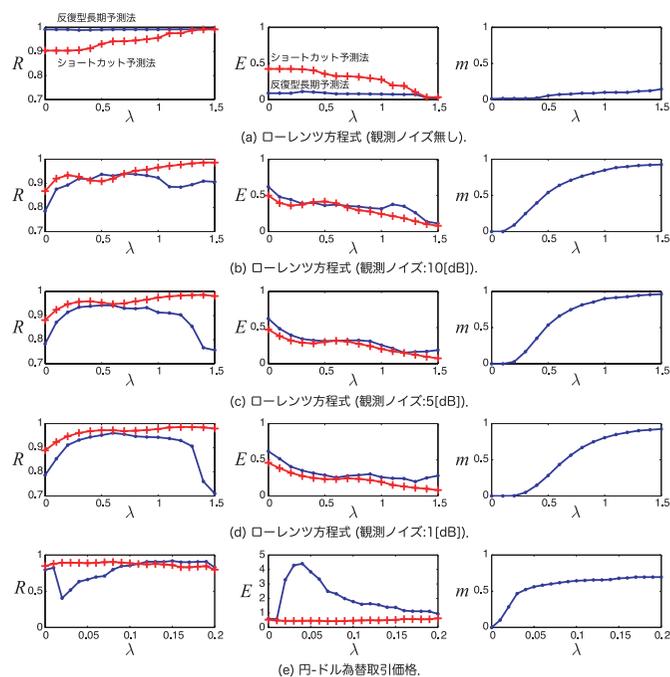


図 4 天底の変動幅 ΔX_i の予測結果と反復型長期予測法における予測不能率 m . 点線は反復型長期予測法を表し, 十字線はショートカット予測法を表している

Fig. 4 The R and E mean prediction accuracies of movements between top and bottom values ΔX_i . Each dotted line shows accuracies of the multistep prediction by iterating one-step predictions, and each cross line shows those of the shortcut prediction of our proposed method. Moreover, m means the rate of unsuccessful multistep predictions by iterating one-step predictions.

5.2 実験結果

天底の変動幅 ΔX_i に関する予測結果を図 4 に, 時間間隔 Δt_i に関する予測結果を図 5 に示す. それぞれの結果より得られる知見は以下のとおりである.

反復型長期予測法では, λ が大きくなるほど予測失敗率 m が増加していく. λ が大きくなるほど, より遠い未来の大変動を予測することになるので, 多ステップ予測が必要になる. この多ステップ予測においてカオスの有する軌道不安定性が影響するだけでなく, 各ステップの予測における近傍点の選択の際, 観測ノイズにより最適な近傍点が選べず, 予測誤

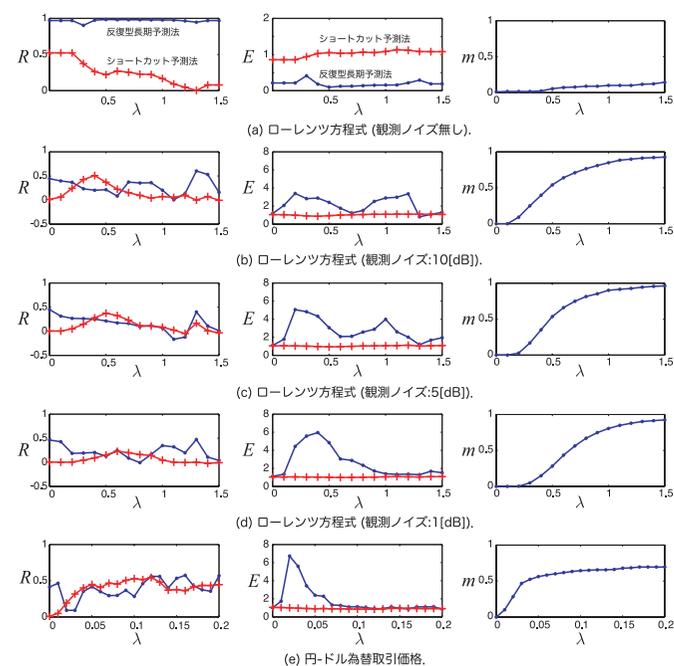


図 5 天底の時間間隔 Δt_i の予測結果と反復型長期予測法における予測不能率 m . 点線は反復型長期予測法を表し, 十字線はショートカット予測法を表している

Fig. 5 The same as Fig. 4, but the case of predicting temporal intervals between top and bottom values Δt_i .

差が拡大するためと考えられる.

しかし, 図 4, 5 の (a) に示すように, ノイズのない純粋なカオスの場合, 反復型長期予測法を用いても予測失敗率 m は小さく, さらにショートカット予測法より予測精度は高い. これは, 軌道不安定性が反復型長期予測法の多ステップ予測に与える影響より, ショートカット予測法において天底変動のみを予測に用いるといった, いわば情報の削減の方が予測精度に悪影響を及ぼさずと考えられる.

一方, 図 4, 5 の (b) ~ (e) において, 観測ノイズや実データに含まれる非カオスの要素を有するデータの予測結果を示す. おおむねショートカット予測法の方が優れていることが分かる. これは, 観測ノイズなどが含まれると, 予測における最適な近傍点を選択するこ

とが困難となる．反復型長期予測法では多ステップ予測の回数が多い(つまり λ が大きい)ほど, 近傍点選択の誤りを繰り返してしまい, 予測誤差を拡大させてしまう．ショートカット予測法では, 予測に用いる情報を減らす代わりに, 近傍点の選択は 1 回でよいという利点がある．そのため, 反復型長期予測法と比較して, ショートカット予測法の予測精度は λ の値にあまり依存しない．

しかし, 天底の時間間隔の予測においては, 正規化平均二乗誤差 E の結果よりショートカット予測法の有用性は明らかであるが, 相関係数 R の結果では両予測法に差異は見られない．これは反復型長期予測法において, $|\Delta \tilde{X}_i|$ および $\Delta \tilde{t}_i$ が過剰に大きくなりがちであることが原因である．つまり, 上昇や下降といった変動の傾向に関する精度については, 両予測法に大差はないが, 正規化平均二乗誤差 E により誤差を直接評価すると, 反復型長期予測法において予測値が大きくなりすぎた分が誤差として算出されている．その結果, 天底変動の傾向を予測できていたとしても変動幅や時間間隔を過大評価しているため, 実際の将来の天底を, 天底として認識できずに見過ごしてしまう危険性がある．反復型長期予測法の予測値が過剰に大きくなる問題は, 図 1 のような天底の予測値を決定できないという問題と同様に, 多ステップ予測値が不安定になることが原因だと考えられる．さらに 5.1 節で述べたように, 予測精度の評価基準として R よりも E の方が適しているデータを予測対象とした場合, このような問題が起こらないという点でショートカット予測法は有用である．

予測精度以外の利点として, ショートカット予測法は天底予測を毎回行うことができ, さらに反復型長期予測法より予測回数を大幅に削減できるので, 計算コストの面でも有利である．

6. ま と め

本研究では, 数理モデルであるローレンツ方程式と実データである円-ドル為替取引価格を解析対象とし, 局所線形近似法を用いて天底予測を行った．その際, 1 ステップ予測を繰り返すことで長期的に次の天底を予測する反復型長期予測法に対し, 過去の天底情報のみを予測に用い, 1 ステップ予測で次の天底を予測するショートカット予測法を提案した．

反復型長期予測法による天底予測では, 基本的に長期予測となるので予測難易度が向上し, 長期予測値から天底の予測値を決定できないという問題があるが, ショートカット予測法を用いることによってこの問題を解決することができた．また, 実データのようにノイズなど, 非カオス的な要素が含まれている場合, 提案法では局所線形近似法において近傍選択の誤りを繰り返さないため, 反復型長期予測法より安定した予測を行うことができる．さら

に, 反復型長期予測法より予測回数を大幅に削減できるので, 計算コストの面でも有利である．

なお, 本研究の一部は, 文科省科研費若手研究(スタートアップ)(No.18800061)のご支援により行われました．

参 考 文 献

- 1) Kanno, Y., Ikeguchi, T. and Uchida, A.: Application of Nonlinear Modeling Methods for Event Sizes and Event Timings, *Proc. 2002 International Symposium on Nonlinear Theory and its Applications (NOLTA '02)*, pp.243–246 (2002).
- 2) Kanno, Y. and Ikeguchi, T.: Nonlinear Prediction on Maximum Timing of Complex Time Series, *Proc. 9th International Conference on Neural Information Processing (ICONIP '02)* (2002).
- 3) Lorenz, E.N.: Atmospheric predictability as revealed by naturally occurring analogues, *Journal of the Atmospheric Sciences*, Vol.26, pp.636–646 (1969).
- 4) 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, 産業図書 (2000).
- 5) Sauer, T., Yorke, J.A. and Casdagli, M.: Embedology, *Journal of Statistical Physics*, Vol.65, No.3/4, pp.579–616 (1991).
- 6) Takens, F.: Detecting strange attractors in turbulence, *Dynamical Systems of Turbulence, Lecture Notes in Mathematics*, Vol.898, pp.366–381, Springer-Verlag (1981).
- 7) Kennel, M.B., Brown, R. and Abarbanel, H.D.I.: Determining embedding dimension for phase-space reconstruction using a geometrical construction, *Physical Review A*, Vol.45, No.6, pp.3403–3411 (1992).
- 8) Abarbanel, H.D.I. and Kennel, M.B.: Local false nearest neighbors and dynamical dimensions from observed chaotic data, *Physical Review E*, Vol.47, No.5, pp.3075–3068 (1993).
- 9) 松葉育雄: 非線形時系列解析, 朝倉書店 (2000).

(平成 20 年 4 月 17 日受付)

(平成 20 年 5 月 27 日再受付)

(平成 20 年 6 月 25 日採録)



鈴木 智也 (正会員)

昭和 51 年生。平成 17 年東京理科大学大学院理学研究科物理学専攻博士課程修了。理学博士。同年東京電機大学工学部電子工学科助手，平成 18 年より同志社大学理工学部情報システムデザイン学科専任講師，現在に至る。複雑系，時系列解析，非線形現象等に関する研究に従事。電子情報通信学会，日本物理学会，人工知能学会各会員。



太田 真喜

昭和 60 年生。平成 20 年同志社大学工学部情報システムデザイン学科卒業。同年 4 月よりアルプス電気株式会社勤務。