雑音を含むカオス時系列データの最小埋込み次元決定方法

伊　藤　憲　一

A Method for Determining the Minimum Embedding Dimension of Chaotic Time Series Corrupted with Measurement Noise

Ken-ichi ITOH

あらまし　カオス時系列データの最小埋込み次元を求める方法として誤り近傍法があるが，この方法は雑音の影響を受けやすいという問題がある。本論文では，この問題解決を図るため，誤り近傍法をもとにした雑音にロバストな方法を提案する。Hénon map および Lorenz model から生成されるカオス時系列データにガウス分布型雑音を加えたデータ，ならびに指尖脈波のデータを用いて評価実験を行い，提案手法の有効性を確認した。
キーワード　カオス，時系列データ，埋込み，誤り近傍法，指尖脈波

1．まえがき

カオス時系列データの解析においては，通常埋込みの手法を用いて相空間上にアトラクタの再構成が行われる[1],[2]。ある１変数の時系列データから最小埋込み次元を求めるために，これまで多くの研究が行われてきた。代表的な方法として，GP法[3]，特異値分解法[4]，誤り近傍法[5]などがある。しかしながら，これらの方法は，最小埋込み次元の決定においてやや客観性に欠けるという問題がある。

この欠点をなくすために，誤り近傍法を基にした改良手法が，Caoにより提案された[6]。この手法は次のような利点を有している。すなわち，（1）埋込みのための時間遅れ以外は，何ら主観的なパラメータを含んでいない，（2）利用できるデータの数にあまり依存しない，（3）高次元のアトラクタから生じた時系列データに対しても適用できるなどである。しかし，実在する時系列データにCaoの手法を適用する場合，次のような問題がある。すなわち，最小埋込み次元の値は，再構成した相空間におけるある点と，その点に最も近いただ一つの点との間の距離を用いて推定されるため，時系列データに含まれる雑音に大きく影響を受ける。実データには通常雑音が含まれているため，そのようなデータの最小埋込み次元を正確に決定することは困難である。

本論文では，Caoの手法を基に，雑音を含むカオス時系列データの最小埋込み次元を決定する方法を提案する。更に，Hénon map[7]およびLorenz model[8]から生成されるカオス時系列データにガウス分布型雑音を加えたデータ，ならびに指尖脈波[9]のデータを用いて評価実験を行い，提案手法の有効性を示す。

2．CAOの手法

ここでは，時系列データから最小埋込み次元を決定するために，Caoにより提案された手法

を説明する。

ある1変数の時系列データ$x_1, x_2, \cdots, x_N$から，一定の時間遅れτを用いて次のd次元ベクトルを作成することにより，アトラクタの再構成を行うことができる[1],[2]。dを埋込み次元と呼ぶ。

$$y_i(d) = (x_i, x_{i+\tau}, \cdots, x_{i+(d-1)\tau}) \quad , \quad i = 1, 2, \cdots, N-(d-1)\tau \tag{1}$$

まず，誤り近傍法[5]の考え方から類推される次式の量$a(i, d)$を計算する。

$$a(i, d) = \frac{\left\| y_i(d+1) - y_{n(i,d)}(d+1) \right\|}{\left\| y_i(d) - y_{n(i,d)}(d) \right\|} \quad , \quad i = 1, 2, \cdots, N - d\tau \tag{2}$$

ここで，$\|\cdot\|$は最大ノルムを表す。すなわち，$\left\| y_k(m) - y_l(m) \right\| = \max_{0 \le j \le m-1} \left| x_{k+j\tau} - x_{l+j\tau} \right|$である。

$y_i(d+1)$は，埋込み次元$d+1$におけるi番目の再構成ベクトルを表す。すなわち，$y_i(d+1) = (x_i, x_{i+\tau}, \cdots, x_{i+d\tau})$である。

$n(i, d)\ (1 \le n(i, d) \le N - d\tau)$はある整数であり，再構成された$d$次元の相空間の中で，最大ノルムで測った時の$y_i(d)$の最近傍点が$y_{n(i,d)}(d)$であることを示す。

次に，$a(i, d)$の平均値$E(d)$を計算する。

$$E(d) = \frac{1}{N - d\tau} \sum_{i=1}^{N-d\tau} a(i, d) \tag{3}$$

さらに，d次元から$d+1$次元への変化を定量化する指標として，次式の$E1(d)$を計算する。

$$E1(d) = E(d+1) / E(d) \tag{4}$$

時系列データがあるアトラクタから生じたものであれば，dがある値d_0より大きくなったときに$E1(d)$が収束し，$d_0 + 1$がこの時系列データの最小埋込み次元となる。

決定論的なデータと確率的なデータを区別するために，もう1つの指標$E^*(d)$が用いられる。

$$E^*(d) = \frac{1}{N - d\tau} \sum_{i=1}^{N-d\tau} \left| x_{i+d\tau} - x_{n(i,d)+d\tau} \right| \tag{5}$$

さらに，次式の指標$E2(d)$が計算される。

$$E2(d) = E^*(d+1) / E^*(d) \tag{6}$$

ランダムデータの場合，過去の値と将来の値とが独立であるから，いかなるdに対しても$E2(d)$は1になる。一方，決定論的なデータの場合，$E2(d)$はdに関係があるため，$E2(d) \neq 1$となるdが存在する。

ある時系列データの最小埋込み次元を決定し，同時に決定論的データとランダムデータを区別するためには，$E1(d)$と$E2(d)$の両方を計算する必要がある。

3．提案手法

Caoの手法は，まえがきで述べたように従来の手法に比べていくつかの利点があり，人工的につくられた時系列データの最小埋込み次元を決定するには大変有効である。しかし，実在する時系列データにCaoの手法を適用するのは困難と考えられる。すなわち，Caoの手法では再構成した相空間において，ある点$y_i(d)$に最も近いただ一つの点が選択され，式(2)の$a(i, d)$の値が計算される。このため，$a(i, d)$の値は実在する時系列データに通常含まれている雑音に大きく影響を受け，$E1(d)$の値を正しく計算するのは困難と考えられる。

ここでは、Caoの手法を基に，時系列データに含まれる雑音にロバストな最小埋込み次元決定方法を提案する。d次元の相空間において，ある点$y_i(d)$の近傍点を近い順にk個探索し，これを$y_{n_j(i,d)}(d)\quad (j = 1, 2, \cdots, k)$とする。近傍点の探索には，ユークリッド距離を用いる。$y_i(d)$と$y_{n_j(i,d)}(d)$との間のユークリッド距離の平均値を，次式により計算する。

$$z_i = \frac{1}{k} \sum_{j=1}^{k} \left\| y_i(d) - y_{n_j(i,d)}(d) \right\| \tag{7}$$

次に，短時間s を経過した後の$y_{i+s}(d)$ と$y_{n_j(i,d)+s}(d)$ との間のユークリッド距離の平均値を，次式により計算する。

$$z_{i+s}=\frac{1}{k}\sum_{j=1}^{k}\left\| y_{i+s}(d)-y_{n_j(i,d)+s}(d)\right\| \tag{8}$$

式(3)の代わりに次式を用いて，$E(d)$ を計算する。

$$E(d)=\frac{1}{N-(d-1)\tau-s}\sum_{i=1}^{N-(d-1)\tau-s}\frac{z_{i+s}}{z_i} \tag{9}$$

また，式(5)の代わりに次式を用いて，$E^*(d)$ を計算する。

$$E^*(d)=\frac{1}{k(N-(d-1)\tau-s)}\cdot\sum_{i=1}^{N-(d-1)\tau-s}\sum_{j=1}^{k}\left| x_{i+(d-1)\tau+s}-x_{n_j(i,d)+(d-1)\tau+s}\right| \tag{10}$$

$E1(d)$ と$E2(d)$ の計算には，式(4)と式(6)をそのまま用いる。

以上のように，提案手法では，再構成した相空間において複数個の近傍点を探索し，これを時系列データの最小埋込み次元の決定に用いるため，時系列データに含まれる雑音にロバストであると想定される。提案手法では，近傍点の数k をいくつに設定するかについての検討が必要である。k が小さすぎると$E1(d)$ の値は雑音に大きく影響を受ける。一方，k が大きすぎると，$y_i(d)$ から離れた点が$y_i(d)$ の近傍点として選ばれるため，$E2(d)$ の値はd の値に関わらずほぼ１になり，決定論的データとランダムデータとを区別することが不可能となる。k の最適値は，調査対象となる時系列データにある程度依存すると考えられるため，試行実験によりk の値を決定することが重要となる。

４．評価実験

提案手法の有効性を確認するために，Hénon map[7]およびLorenz model[8]から生成される時系列データにガウス分布型雑音を加えたデータ，ならびに指尖脈波[9]のデータを用いて評価実験を行った。

Hénon mapは，次の２次元写像である。

$$x_{n+1}=y_n+1-Ax_n^2\ ,\quad y_{n+1}=Bx_n \tag{11}$$

式(11)において，パラメータ値$A=1.4$ ，$B=0.3$ ，初期値$x_0=0.3$ ，$y_0=0.3$ としたときのx を時系列データとした。

Lorenz modelは，次の３変数微分方程式である。

$$\dot{x}=-\sigma(x-y)\ ,\quad \dot{y}=-y-xz+rx\ ,\quad \dot{z}=xy-bz \tag{12}$$

式(12)において，パラメータ値$\sigma=10$ ，$r=28$ ，$b=8/3$ ，初期値$x=0.1$ ，$y=0$ ，$z=0$ ，時間刻み$\delta t=0.01$ として４次のルンゲ・クッタ法でx の時間変化を求め，これを時系列データとした。

雑音データを生成するために，上記のHénon mapとLorenz modelの時系列データに，１％と３％のガウス分布型雑音を加えた。時系列データのデータ長N は，1,000と10,000に設定した。時間遅れτ は，Hénon mapでは1に，Lorenz modelでは10に設定した。経過時間s は，τ と同一とした。

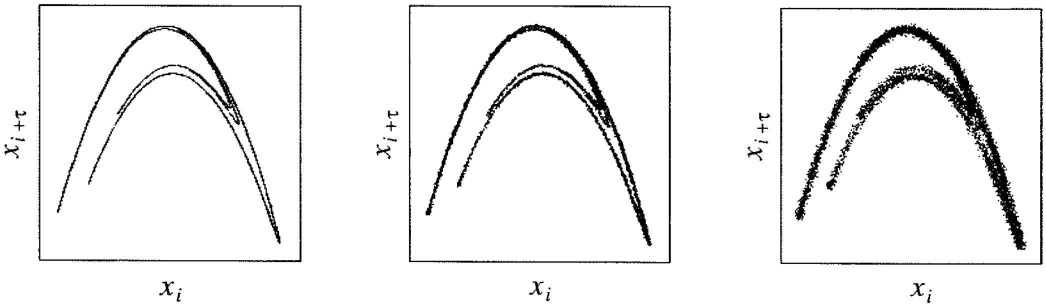
Hénon mapおよびこれにガウス分布型雑音を加えた時系列データの相図を図１に示す。Lorenz modelおよびこれにガウス分布型雑音を加えた時系列データの相図を図２に示す。

指尖脈波のデータについては，これまでの研究[9]によりカオス的特徴を有することが指摘されており，最小埋込み次元は４であると推定されている。

本実験では，３人の健常者の指尖脈波について，各々安静状態で測定したデータを用いた。データのサンプリング時間は5msである。パラメータN ，τ ，s の値は，Lorenz modelの場合と同一とした。

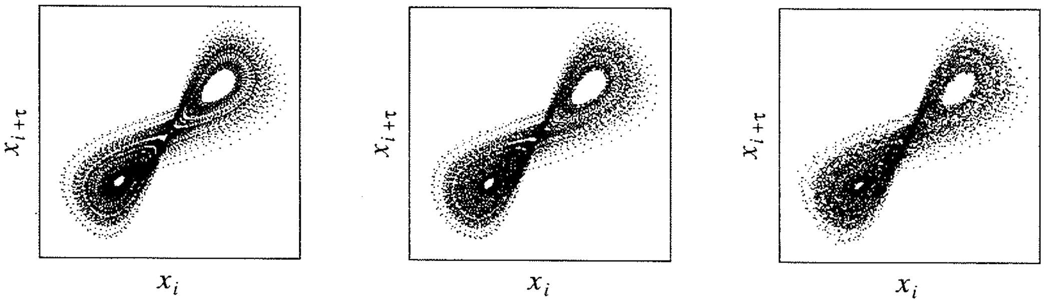
指尖脈波データの相図を図３に示す。

本実験では，まず近傍点の数k の値を決定するために，Hénon mapおよびLorenz modelの



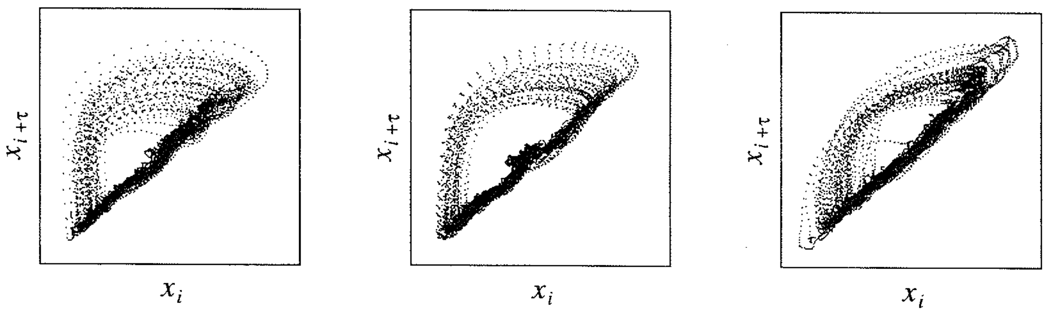
Original data　　+ 1% Gaussian white noise　　+ 3% Gaussian white noise

図1　Hénon mapの相図
Fig. 1　Phase plots of the Hénon map.



Original data　　+ 1% Gaussian white noise　　+ 3% Gaussian white noise

図2　Lorenz modelの相図
Fig. 2　Phase plots of the Lorenz model.



Subject A　　Subject B　　Subject C

図3　指尖脈波データの相図
Fig. 3　Phase plots of the pulsation of human finger capillary vessels.

データ，指尖脈波データ，ランダムデータを用いて，$E2(1)$ とk の関係を調べた。なお，指尖脈波データは，図３のSubject Bのデータを使用した。結果を図４に示す。ランダムデータの場合，近傍点の数k の値に関わらず，$E2(1)$ の値はほぼ１となる。一方，Hénon mapおよびLorenz modelのデータ，指尖脈波データの場合，k がデータ長N の約10%以下では，$E2(1)$ の値は１より小さくなる。k がN の5%になると，$E2(1)$ の値は0.5から0.6の間となる。本実験では，雑音にロバストであり，かつ決定論的データとランダムデータとを区別できるようにするため，k の値をN の5%に設定した。

Hénon mapのデータについて，$E1(d)$ と埋込み次元d との関係を求めた結果を図５に示す。図５において，従来手法とはCaoの手法を意味する。これは，以降の図でも同様である。Lorenz modelのデータについて，$E1(d)$ とd との関係を求めた結果を図６に示す。図５，図６から，雑音を加える前のオリジナルデータに対しては，従来手法および提案手法ともに最小埋込み次元を正しく決定できることがわかる。すなわち，Hénon mapの場合は２であり，Lorenz modelの場合は３である。しかし，ガウス分布型雑音を加えたデータに対して従来手法を用いると，$E1(d)$ が収束するd の値は正しい値よりも大きくなる。従って，従来手法は雑音を含むデータに対して有用ではない。一方，提案手法は，ガウス分布型雑音を加えたデータに対しても最小埋込み次元を正しく決定できる。

指尖脈波データについて，$E1(d)$ と埋込み次元d との関係を求めた結果を図７に示す。提案手法では，これまでの研究[9]の結果と同様に最小埋込み次元はほぼ４となる。一方，従来手

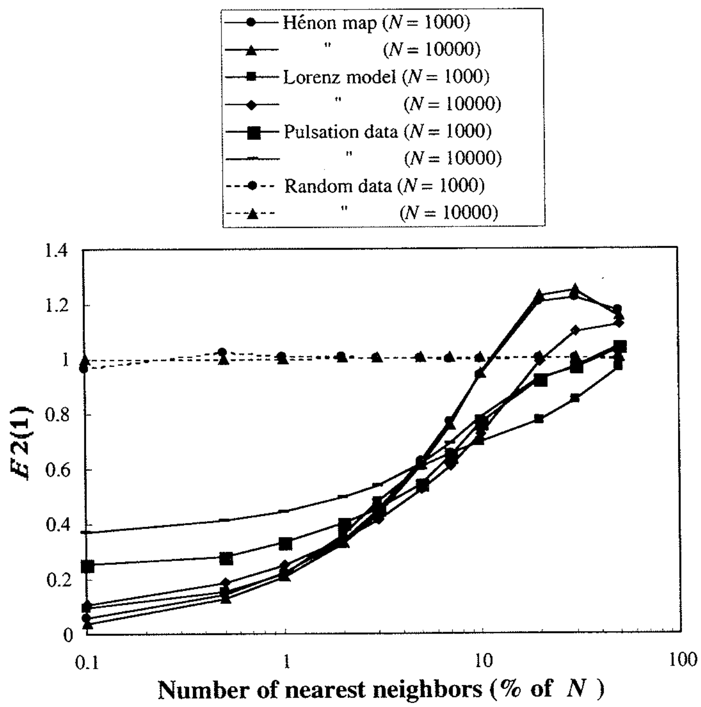
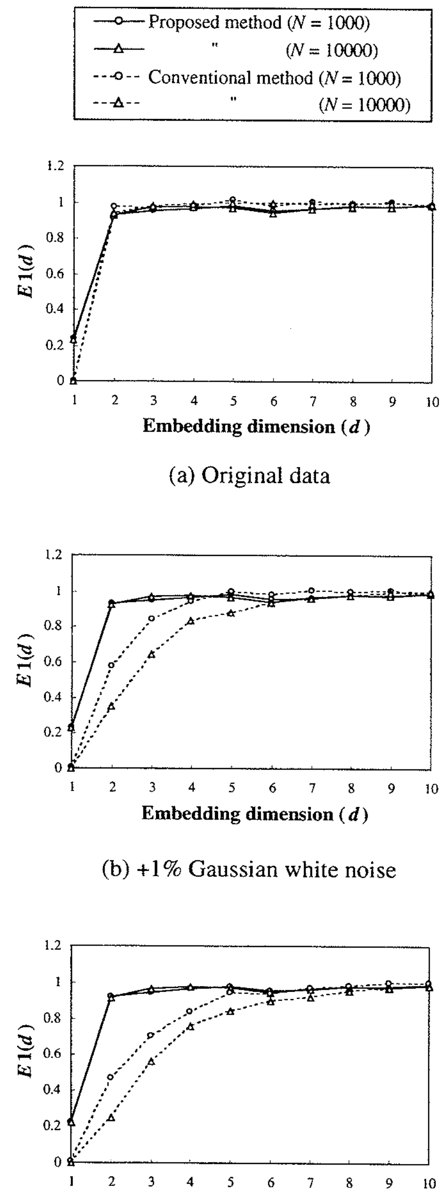


図４　近傍点の数と$E2(1)$の関係

Fig. 4　$E2(1)$ vs. number of nearest neighbors.



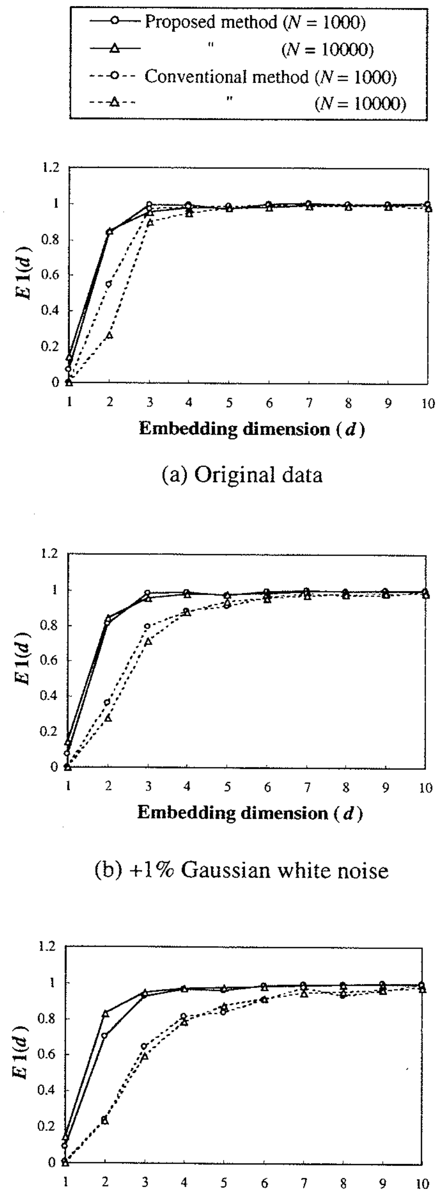
(a) Original data

(b) +1% Gaussian white noise

(c) +3% Gaussian white noise

図5　Hénon mapにおける埋込み次元と$E1$(d)の関係

Fig. 5　$E1$(d) vs. embedding dimension for the Hénon map.



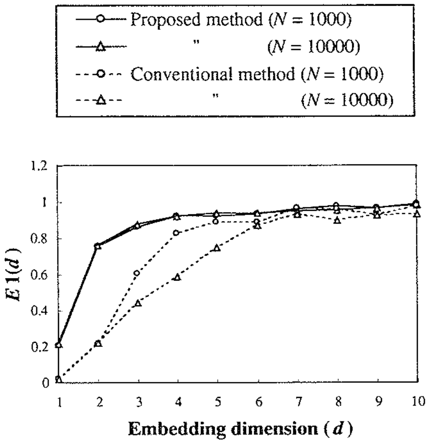
(a) Original data

(b) +1% Gaussian white noise

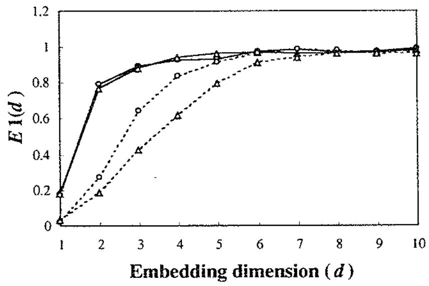
(c) +3% Gaussian white noise

図6　Lorenz modelにおける埋込み次元と$E1$(d)の関係

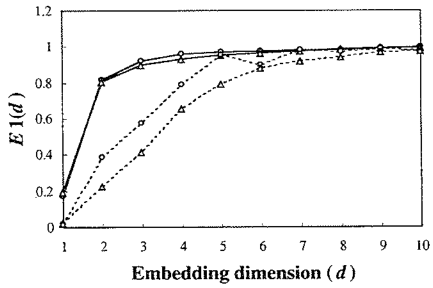
Fig. 6　$E1$(d) vs. embedding dimension for the Lorenz model.



(a) Subject A



(b) Subject B



(c) Subject C

図7　指尖脈波データにおける埋込み次元と$E1$(d)の関係

Fig. 7　$E1$(d) vs. embedding dimension for the pulsation of human finger capillary vessels.

法を用いると，$E1(d)$ が収束するd の値は 4 よりも大きくなり，誤った結果となる。この誤りは，指尖脈波データに含まれている雑音に起因するものと考えられる。

以上の評価実験より，提案手法は時系列データに雑音が含まれる場合でも最小埋込み次元を正確に決定できることを確認した。

5．むすび

本論文では，Caoの手法を基に，雑音を含むカオス時系列データの最小埋込み次元を決定する方法を提案した。更に，Hénon mapおよびLorenz modelから生成されるカオス時系列データにガウス分布型雑音を加えたデータ，ならびに指尖脈波のデータを用いて評価実験を行い，提案手法の有効性を示した。

今後の課題は，提案手法を他の実データにも適用し，評価を行うことである。

文献

[1] F. Takens, “Detecting strange attractors in turbu-lence,” in Dynamical Systems and Turbulence, eds. D. A. Rand and L. S. Young, *Lecture Notes in Mathematics*, vol.898, pp.366-381, Springer, Berlin, 1981.

[2] T. Sauer, J. A. Yorke, and M. Casdagli, “Embedology,” *J. Stat. Phys.*, vol.65, no.3,4, pp.579-616, 1991.

[3] P. Grassberger and I. Procaccia, “Measuring strange-ness of strange attractors,” *Physica*, vol.9D, pp.189-208, 1983.

[4] D. S. Broomhead and G. P. King, “Extracting qualita-tive dynamics from experimental data,” *Physica*, vol.20D, pp.217-236, 1986.

[5] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, “Determining embedding dimension for phase-space reconstruction using a geometrical construction,” *Phys. Rev. A*, vol.45, no.6, pp.3403-3411, 1992.

[6] L. Cao, “Practical method for determining the mini-mum embedding dimension of a scalar time series,” *Physica D*, vol.110, pp.43-50, 1997.

[7] M. Hénon, "A two-dimensional mapping with a strange attractor," *Commun. Math. Phys.*, vol.50, pp.69-77, 1976.

[8] E. N. Lorenz, "Deterministic non-periodic flow," *J. Atmos. Sci.*, vol.20, pp.130-141, 1963.

[9] I. Tsuda, T. Tahara, and H. Iwanaga, "Chaotic pulsa-tion in human capillary vessels and its dependence on mental and physical conditions," *Int. J. Bifurca-tion and Chaos*, vol.2, no.2, pp.313-324, 1992.