

自己組織化マップを用いた システムパラメータの感度解析

SENSITIVITY ANALYSIS FOR NON-LINEAR SYSTEM USING SELF-ORGANIZING MAP

井芹慶彦¹・神野健二²・里村大樹³・BELLIE SIVAKUMAR⁴・西山浩司⁵
Yoshihiko ISERI, Kenji JINNO, Daiki SATOMURA, Bellie SIVAKUMAR
and Koji NISHIYAMA

¹学生会員 工修 九州大学大学院 工学府都市環境システム工学専攻 (〒812-8581 福岡市東区箱崎6丁目10-1)

²正会員 工博 九州大学教授 工学研究院環境都市部門 (同上)

³非会員 国土交通省近畿地方整備局 (〒573-1191 大阪府枚方市新町2丁目2番10号)

⁴非会員 工博 カリフォルニア大学ディビス校 陸空水資源部門 (〒95616 アメリカ合衆国カリフォルニア州ディビス・カリフォルニア大学・陸空水資源部門)

⁵正会員 理修 九州大学助手 工学研究院環境都市部門 (〒812-8581 福岡市東区箱崎6丁目10-1)

Linear or non-linear system equations have been applied for the modeling of hydrological time series. However, the estimation of optimal parameters in the system is one of the most difficult processes in hydrological modeling, because the slight perturbation of the parameters in the system can cause remarkable difference of system solutions. Thereby, in order to assure the reliability of estimated system parameters, it is crucial to investigate the response of the system solutions against perturbation of the system parameters.

In this study, we conducted sensitivity analysis for the perturbation of parameters in non-linear system, using Self-Organizing Map. As the results, the time series generated from semi-optimal parameters were classified depending on their similarities. The parameters which generated approximately actual time series were partially visualized on the 2-Dimensional map.

Key Words : *hydrological modeling, non-linear system, differential equation, sensitivity analysis, Self-Organizing Map*

1. はじめに

水文時系列の特性解析や予測を行うために、様々な方法を用いて水文時系列のモデル化が試みられている。水文時系列をモデル化する方法の一つとして、線形又は非線形のシステム方程式を適用してモデル化する方法がある。特に近年では、カオス力学系理論の水文時系列への適用^{1), 2), 3)}に伴い、カオス力学系微分方程式をシステム方程式として導入する研究例も見られる⁴⁾。

システム方程式とは、例えば降雨量をシステム入力値として、対応する流出量をシステム出力値とするような方程式系のことであり、システム方程式内には一般に複数の未知パラメータが含まれる。従って、システム方程式を用いて実問題をモデル化するには、実データに解が適合するように方程式中に含まれるパラメータを調整する必要がある。しかし、特に非線形項を含むシステム方程式解は一般に、パラメータの変化に対して複雑な応答を示す。また、パラメータの推定法によっては、最適解に対応するパラメータを推定できず、局所解に陥ってしまう可

能性も考えられる。こうしたことから、特に非線形システム方程式における最適パラメータの推定は、水文時系列をモデル化する際に直面する、困難な問題の一つとなっている。

最適パラメータの大域的探索手法として、GA⁵⁾(Genetic Algorithm; 遺伝的アルゴリズム)や、SCE-UA 法⁶⁾ (shuffled complex evolution method developed at The University of Arizona)といった最適化アルゴリズムが提案されており、これら手法は最適化問題において、広範に使われている。しかしこれら手法は、単一の最適パラメータのみを探索するため、最適パラメータ周辺のパラメータに対する解の挙動や分布を知るには、あまり適していない。

そこで本研究では、自己組織化マップを用いて、最適パラメータ周辺のパラメータに対する解の応答を可視化することによる、感度解析アルゴリズムの提案を行う。

本論文の構成は次の通りである。まず、方程式解及びシステムパラメータの分類に用いた、自己組織化マップについて第2章で説明する。第3章では、感度解析アルゴリズムのシミュレーション対象に選定

した、非線形常微分方程式で構成されるシステム方程式（レスラー方程式）の特性について記述する。第4章では、本研究で提案する感度解析アルゴリズムの手順を説明し、第5章でレスラー方程式に感度解析アルゴリズムを適用した結果を示した。

2. 自己組織化マップ

自己組織化マップ⁷⁾とは、多次元ベクトルの形で表示されたデータを、その幾何学的特長に応じて分類する手法である。多次元データをその特徴に応じて分類し、低次元のマップ上に視覚化できる所に自己組織化マップの利点がある。2次元の格子に配置されたノードによって、自己組織化マップは通常構成されている。自己組織化マップの適用により、各ノード上に、観測データの支配的なパターンを表すモデル(参照ベクトル)が形成される。また、これらモデルは、互いに類似したものが近接するように、マップ上に配置されていく。マップ上に配置された参照ベクトルを $m_i \in R^n$ とし、 $x \in R^n$ を入力ベクトル(観測データから成るベクトル)とする。このとき、 m_i の更新は以下の更新式に従って行われる。

$$m_i(t+1) = m_i(t) + h_{c(x),i}(x(t) - m_i(t)) \quad (1)$$

上式中、 t は更新ステップであり、 $x(t)$ は更新ステップ t において取り出された観測ベクトルである。また、 $h_{c(x),i}$ は近傍関数であり、添え字 $c = c(x)$ は下記の条件式で定義される。任意の i に対し、

$$\|x(t) - m_c(t)\| \leq \|x(t) - m_i(t)\| \quad (2)$$

上式中、 $m_c(t)$ を勝者ベクトルとよぶ。また、近傍関数として以下のガウス関数がしばしば用いられる。

$$h_{c(x),i} = \alpha(t) \exp\left(-\frac{\|r_i - r_c\|^2}{2\sigma^2}\right) \quad (3)$$

ここで、 $0 < \alpha(t) < 1$ は学習率であり、更新ステップにつれて単調減少させる。 $r_i \in R^2$ および、 $r_c \in R^2$ は、各参照ベクトルの、2次元マップ上での位置をあらわす、位置ベクトルである。 $\sigma(t)$ は近傍関数の幅に対応し、学習率と同じく、更新ステップにつれて単調減少させる。また、本研究では、学習率 $\alpha(t)$ および $\sigma(t)$ には以下の式を用いた。

$$\begin{cases} \alpha(t) = \alpha(0) \frac{T-t}{T}, \\ \sigma(t+1) = 1 + (\sigma(t) - 1) \frac{T-t}{T} \end{cases} \quad (4)$$

なお、式中において T は総学習回数を表す。また本研究では、各ノードの分類精度を測る指標として QE (Quantization Error) を用いた。

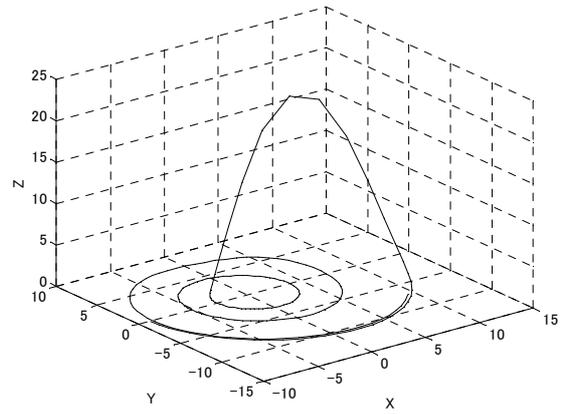


図-1 3次元状態空間に描いた、時間間隔0.1毎のレスラー方程式の解

3. レスラー方程式

レスラー方程式⁸⁾は、化学者O.Rösslerによって提案された数理モデルであり、以下の微分方程式で表される⁹⁾。なお非線形項は第3式中の $z \cdot x$ のみである。

$$\begin{cases} \frac{dx}{dt} = -y - z \\ \frac{dy}{dt} = x + ay \\ \frac{dz}{dt} = b + z(x - c) \end{cases} \quad (5)$$

パラメータ a, b, c を変化させることにより、上式で示されるシステムは様々に変化する。レスラー方程式を $t=0.00 \sim t=20.00$ まで数値積分した解を、3次元状態空間にプロットしたものを図-1に示す。なおパラメータ $(a,b,c) = (0.2, 0.2, 5.7)$ 、初期条件 $(x_0, y_0, z_0) = (-9, 0, 0)$ として計算を行っている¹⁰⁾。また、各変数に対して時間間隔0.1毎に取り出された解が、図-1にはプロットされている。

4. 自己組織化マップを用いた、システムパラメータの感度解析アルゴリズム

本研究で提案する感度解析アルゴリズムは、大別すると二つの目的をもつ。一つ目は、システムパラメータの変動に対してシステム解がとり得る軌道を自己組織化マップを用いて分類し、可視化することである。二つ目は、最適パラメータ周辺に存在し、最適パラメータと推定されるパラメータを、自己組織化マップを用いて分類することである。システム方程式中のパラメータ変化に対する解の感度解析法として、本研究で提案するアルゴリズムの概略は以下の通りである。

(1) パラメータの発生

ここでは、アルゴリズムを示すのが目的であるの

で、何らかの手法により最適パラメータの組合せが既に絞り込めてあると仮定する。つまり、 n 個のパラメータをもつシステムの最適パラメータの組合せ $(P_1, P_2, P_3, \dots, P_n)$ が既知であるとする。また、 $(r_1, r_2, r_3, \dots, r_n)$ を適度に小さな数とする。ここで、 $(P_1 \pm r_1, P_2 \pm r_2, P_3 \pm r_3, \dots, P_n \pm r_n)$ の範囲内でパラメータの組合せを、乱数を用いて十分な個数発生させる。ここで、各パラメータの組合せからなるベクトルをパラメータベクトルと呼ぶことにする。

(2) 時系列の計算

上記手順により発生させたパラメータの組み合わせそれぞれに対して微分方程式を数値積分し、時系列を求める。それら時系列の中から、適当な時間ステップまで解が発散していない時系列を選出し、自己組織化マップによる時系列の分類に用いる。

(3) 自己組織化マップによる時系列の分類

次に、自己組織化マップを用いてそれら収束した時系列を分類する(このマップをマップ1とする)。さらに、このマップをK-means法を用いてクラスター化する。K-means法では、最初にクラスターの中心となるベクトル(中心ベクトル)を適当な数つくる。次に、マップ上の参照ベクトルを用いて中心ベクトルを更新していき、最終的には中心ベクトルを各クラスターの中心としてクラスター化が行われる。

(4) 時系列の類似度に応じた点数の配分

各時系列と真の時系列との類似度に応じて、各時系列を発生させたパラメータに次の要領で点数を与える。真の時系列と各時系列とのRoot Mean Square Error (RMSE)を計算し、その逆数を点数とする。各時系列を発生させたパラメータベクトルの最後の成分として、RMSEの値に応じた点数をパラメータベクトルの末尾に加える。

(5) 自己組織化マップによるパラメータ・点数ベクトルの分類

パラメータベクトルと点数からなるベクトルを自己組織化マップを用いて分類する(マップ2とする)。これにより、「パラメータと時系列に応じた点数」を成分にもつ参照ベクトルを用いて、パラメータと解の関係を表したマップを作成する。

(6) 高得点パラメータを用いた、時系列の再計算

点数に対応する成分が高い値を示している中心ベクトルをマップ2上のクラスターより選出する。選出された中心ベクトルのパラメータの値を取り出し、微分方程式の数値積分を再度行い、時系列を求める。

5. レスラー方程式への感度解析アルゴリズムの適用

本研究では提案するアルゴリズムの性能を評価す

るため、4章で示したアルゴリズムをレスラー方程式に適用した。数値シミュレーションの適用対象にレスラー方程式を用いた理由は次のようである。数値シミュレーションの対象とするシステム方程式は、過度に非線形性の強い方程式でないと同時に、解がある程度豊富な挙動を示すものが望ましいと考えられる。なぜなら、提案するアルゴリズムは方程式解のパターン分類を行うので、非常に非線形性の強い方程式では存在しうる解のパターンが膨大になり、解のパターン分類が困難になる可能性が考えられる。また、解の挙動が非常に単純な方程式系では、本アルゴリズムの性能を評価するには不十分であると考えられる。したがって、非線形項が1項のみであると同時に方程式が豊富な挙動を持つレスラー方程式をシミュレーション対象に選定した。

本章では、4章に示した感度解析アルゴリズムをレスラー方程式に適用した結果を示す。

(1) 計算条件

数値シミュレーションを行う際の計算条件は、以下のように設定した。

a) 自己組織化マップに関する計算条件

マップサイズ: 15×15 (つまり、マップ上のノード数を225とした。また、各ノードはマップを六角形の格子に分割した各格子の中心に配置した)

学習率の式中の初期条件: $\alpha(0)=3.0$

近傍関数の幅の初期条件: $\sigma(0)=3.0$

総学習回数T: $T=500000$

b) レスラー方程式に関する計算条件

初期条件: 初期値は、 $(x_0, y_0, z_0)=(9, 0, 0)$ と与えられていると仮定する

数値積分法: 4次のルンゲ・クッタ法

数値積分時間間隔: $dt=0.01$

(2) 数値シミュレーション

まず、真の時系列の作成を次のように行った。レスラー方程式に対し、真のパラメータとして $(a_{\text{true}}, b_{\text{true}}, c_{\text{true}})=(0.2, 0.2, 5.7)$ を与える。このパラメータを用いて、レスラー方程式を $t=0.00 \sim 20.00$ まで数値積分する。その結果、数値解を表すベクトルとして

$$\begin{cases} \vec{x} = (x_{t=0.00}, x_{t=0.01}, x_{t=0.02}, \dots, x_{t=20.00}) \\ \vec{y} = (y_{t=0.00}, y_{t=0.01}, y_{t=0.02}, \dots, y_{t=20.00}) \\ \vec{z} = (z_{t=0.00}, z_{t=0.01}, z_{t=0.02}, \dots, z_{t=20.00}) \end{cases} \quad (6)$$

が求まる。これら数値解から0.1秒毎の値を取り出して、真の時系列を位相空間上に描いた(図-1を参照)。ここで、0.1秒毎の値を取り出したのは、後に行う自己組織化マップを用いた分類を行う際に用いる、入力データ数を減少させるためである。

a) パラメータの発生

各パラメータの変化範囲 (r_1, r_2, r_3) を、 $(0.5, 0.5, 0.5)$ とする。次に、 $(a_{\text{true}} \pm r_1, b_{\text{true}} \pm r_2, c_{\text{true}} \pm r_3) = (0.2 \pm 0.5, 0.2 \pm 0.5, 5.7 \pm 0.5)$ の範囲内で、少数点以下第3位の精

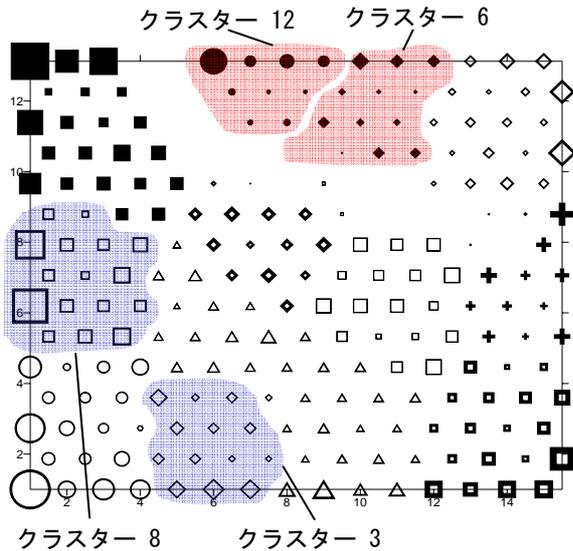


図-2 3436通りの時系列をSOMで分類した後、K-meansを用いてクラスター化したマップ(マップ1)

度で、一様乱数を5000通り発生させる。この結果、 $(a_1, b_1, c_1), (a_2, b_2, c_2), \dots, (a_{5000}, b_{5000}, c_{5000})$ の5000通りのパラメータが数値計算で用いられる。なお、これら5000通りのパラメータ中の任意のパラメータ (a_n, b_n, c_n) は、 $(-0.30 \leq a_n \leq 0.7, -0.30 \leq b_n \leq 0.7, 5.2 \leq c_n \leq 6.2)$ の範囲にある。

b) 時系列の計算

発生させた5000通りのパラメータそれぞれをレスラー方程式に代入し、 $t=0.00 \sim 20.00$ までの数値解を求める。各パラメータの数値解から、時間間隔0.1秒毎の値を取り出し、時系列を各パラメータの組合せに対して描く。これら時系列のうち、 $t=20.00$ までに、解 (x, y, z) のいずれか一つでも、値が 1.0×10^7 を超えた時系列は発散時系列とした。この発散時系列は、次に行う自己組織化マップを用いた時系列の分類には用いない。その結果、3436個の時系列が収束時系列として求まった。

c) 自己組織化マップによる、時系列の分類

上記3436個の時系列を自己組織化マップにより分類した後、K-meansを用いてクラスター化した。その結果マップ上に14個のクラスターが形成された(図-2)。なお、以降この図をマップ1とよぶ。マップ1において、各ノードのシンボルは、その属するクラスターに応じて変えている。なお、各シンボルの大きさは、そのノードに分類された時系列の数を表している。

まず、マップ1の特性について検討する。マップ1の顕著な特徴として、マップ右上方を斜めに横断する空白部分が見られた。このことから、この空白部分を境にして、全く異質の時系列が存在することが示唆される。

次に、マップ1のクラスター化の性能を調べるため、各クラスター内に分類された参照ベクトルと、その参照ベクトルが属するクラスターの中心ベクトルとのユークリッド距離の平均値を計算した。この値が小さいほど、そのクラスター内の分類精度は高

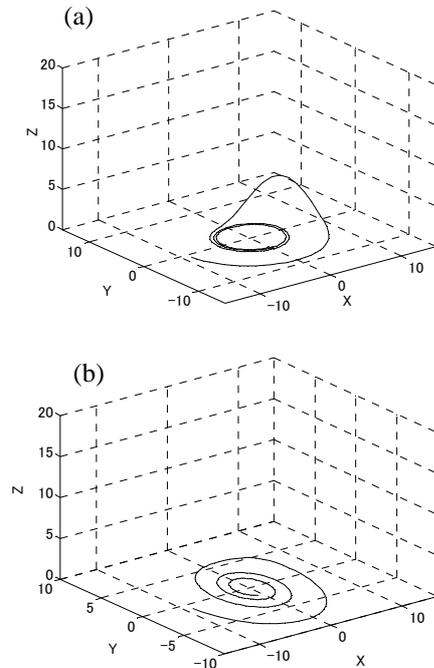


図-3 3次元状態空間に描いた(a)クラスター3の中心ベクトル、(b)クラスター8の中心ベクトル

いと考えられる。クラスター3, 8が上位1, 2位の小さなユークリッド距離を示し、クラスター6, 12が下位1, 2位の大きなユークリッド距離を示した。これらクラスター3, 8, およびクラスター6, 12の中心ベクトルを三次元位相空間上にプロットしたものを、図-3(a), (b)および図-4(a), (b)にそれぞれ示す。図-3(a)と図-3(b)及び平均ユークリッド距離の値から、クラスター3と8にはそれぞれ図-3(a), (b)で示される明確な軌道を持つ参照ベクトルが精度よく集まっていることが分かる。一方、クラスター6と12の中心ベクトルは図-4(a), (b)に見られるように明確な軌道を持っておらず、クラスター化した際の分類性能も悪い。また、クラスター6と12の中心ベクトルの軌道は発散に向かっていくようであった。さらに、クラスター内の各ノードのQEが高い値であったことも考慮すると、クラスター6, 12内には発散に向かう様々な軌道を持った時系列が分類されていると考えられる。

d) 時系列の類似度に応じた点数の配分

各パラメータが発生させた時系列と、真の時系列とのRMSEを計算する。各時系列を発生させたパラメータベクトルの第4成分に、このRMSEの逆数を点数として与える。なお、ステップb)にて除外した、発散時系列を生成するパラメータの時系列に与える点数は全て0点とした。次に行うパラメータ・点数ベクトルの分類では、発散時系列を生成したパラメータもSOMへの入力ベクトルとして用いた。

e) 自己組織化マップによるパラメータ・点数ベクトルの分類

パラメータと点数からなる5000通りのベクトルをSOMを用いて分類した後、K-meansを用いてクラスター化した。その結果マップ上に18個のクラスター

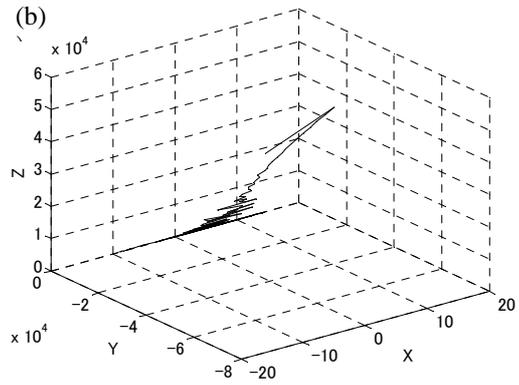
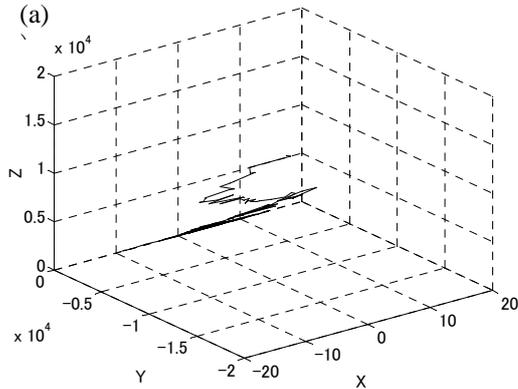


図-4 3次元状態空間に描いた、(a)クラスター6の中心ベクトル、(b)クラスター12の中心ベクトル

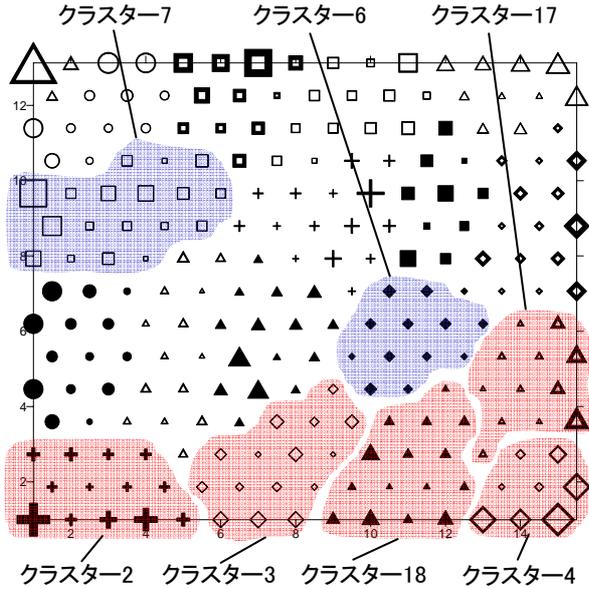


図-5 パラメータと点数から成る5000個のベクトルをSOMで分類した後、K-meansを用いてクラスター化したマップ(マップ2)

が形成された(図-5). なお、このマップをマップ2とする. 図中の各シンボルの意味はマップ1と同様である. マップ2を作成する目的は、真の時系列に近い時系列を生成させるパラメータの組合せをマップ上に分類することにある. また、発散時系列を生成するパラメータもSOMへの入力に含めることで、発散パラメータの組合せを同時に分類することも目的とする. 図-5より、マップ2上ではマップ1に比べある程度均等にデータが分類されており、マップ1に現れた明確な境界はないことが分かる.

次にマップ2における18個の各クラスターの中心ベクトルの得点成分を調べた所、クラスター2, 3, 4, 17, 18が下位1~5位の特に低い点数を示していた. このことから、マップ2の下部には発散時系列を含む、真の時系列とは全く異なる時系列を生成するパラメータが分類されていると考えられる. これらパラメータの一例として、クラスター4内に属する参照ベクトル中のパラメータ成分を、図-6にレーダーチャートで示す. なお、図-6において各軸はパラメータの各成分を表しており、軸は各成分の最

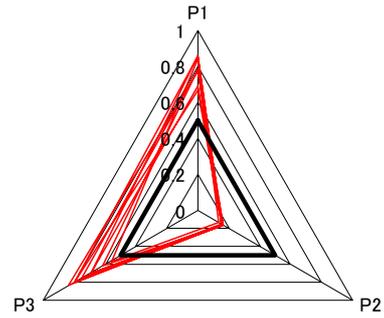


図-6 クラスター4に属する参照ベクトルのレーダーチャート

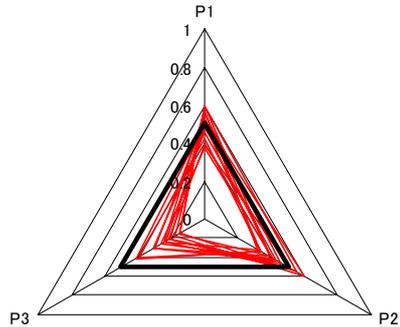


図-7 クラスター6に属する参照ベクトルのレーダーチャート

大値が1、最小値が0へと線形スケールしている. また、図中に黒の太線で示しているのは真のパラメータである. 図-6よりクラスター4内では、パラメータの第2成分は分散が小さく、第1成分と第3成分はやや分散が大きくなっている. このことから、クラスター4内の参照ベクトルからは、パラメータの第2成分が発散時系列の生成に大きく関与していることが示唆される. 次に、18通りのクラスター中の中心ベクトルの得点成分が高いものを調べた所、クラスター6, 7がそれぞれ0.312, 0.412と他のクラスターに比べ高い点数を示していた. ここで、クラスター6, 7に属する参照ベクトルのパラメータ成分をそれぞれ図-7, 図-8に示す. 図-7より、クラスター6に属する参照ベクトルの第3パラメータ成分は全て真の値よりかなり小さい. しかし、第1, 第2パラメータ成分が真の値 ± 0.1 前後であれば、真の時系列にある程度近い時系列を生成できるようなのである.

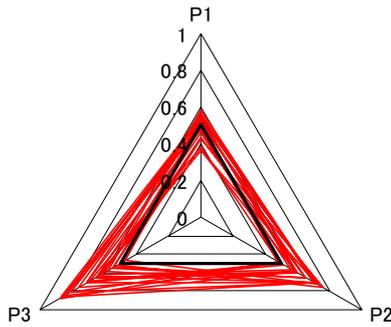


図-8 クラスタ7に属する参照ベクトルのレーダーチャート

図-8より、第1パラメータ成分が真の値±0.1程度で、第2、第3パラメータはともに真の値より大きい場合も真の時系列に近い時系列を生成できるようである。

f) 高得点パラメータを用いた時系列の再計算

18個のクラスタの各中心ベクトルの内、得点成分が高い値を示していた、クラスタ6、7のパラメータ成分を用いて時系列を再計算した(図-9)。なお、クラスタ6、7の中心ベクトルのパラメータ成分はそれぞれ、Cluster6=(0.194, 0.154, 5.448), Cluster7=(0.189, 0.327, 5.884)である。図-9より、これらクラスタ6、7の中心ベクトル中のパラメータを用いて、真の時系列に非常に良く似た時系列を発生させることが分かる。

(3) 結果の考察

本研究では、レスラー方程式から生成された時系列に対し、提案するアルゴリズムを適用した。その結果、レスラー方程式が生成する時系列を分類できた。また、真の時系列と類似(または相違)した時系列を生成するパラメータの組合せを分類できた。

以上から、本アルゴリズムを実データに適用した場合の利点を、例えば神野らの研究⁴⁾に即して考えると次の点が挙げられる。神野らは、線形化したレスラー方程式を黒点時系列の予測方程式に用い、その際システムパラメータの推定には拡張カルマンフィルターを用いた。この研究で本アルゴリズムを併用することで次の二点が期待できる。一つは、用いた方程式が生成する時系列の傾向を知ること、用いた方程式(線形化したレスラー方程式)が黒点時系列(観測時系列)の予測方程式として妥当かを検討できる。もう一つは、推定されたパラメータ周辺のパラメータが生成する時系列をマップ上で確認することで、推定されたパラメータの信頼性の評価に役立つことができるという点である。

5. 結論

本研究では、自己組織化マップを用いて、非線形微分方程式で表されたシステム方程式のパラメータ変化に対する解の感度解析を行った。

まず、乱数を用いて発生させた5000通りのパラメータをレスラー方程式に代入して得られた数値解

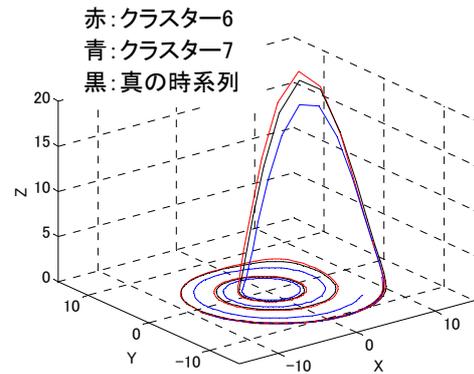


図-9 三次元状態空間上に描いた、クラスタ6と7の中心ベクトルのパラメータを用いて計算した時系列

からなる時系列を自己組織化マップを用いて分類した。その結果、マップ上に形成された境界により、発散及び収束時系列は明瞭に分類されていた。さらに、時系列が特定の軌道を持って収束する場合をマップ上で分類することができた。

また、微分方程式のパラメータに真の時系列への時系列の類似度から定めた点数(RMSE)を加えて自己組織化マップで分類したところ、真の時系列に非常に良く似た時系列を発生させるパラメータの組合せや、発散時系列を発生させるパラメータの組合せをマップ上に分類できた。

参考文献

- 1) Jayawardena, A. W., Lai, F.: Analysis and prediction of chaos in rainfall and stream flow time series, *Journal of Hydrology*, Vol.153, pp.23-52, 1994.
- 2) Kawamura, A., McKerchar, A.I., Spigel, R.H. and Jinno, K.: Chaotic characteristics of the Southern Oscillation Index time series, *Journal of Hydrology*, Vol.204, pp.168-181, 1997
- 3) Sivakumar, B., Liong, S.-Y., Liaw, C.-Y.: Evidence of chaotic behavior in Singapore rainfall, *J. Am. Water Resour. Assoc.*, Vol.34(2), pp.301-310, 1998.
- 4) Jinno, K., Xu, S., Berndtsson, R., Kawamura, A. and Matsumoto, M.: Prediction of sunspots using reconstructed chaotic system equations, *Journal of Geophysical Research - Space Physics*, Vol.100, No.8, pp.14,773-14,781, 1995
- 5) Tsutui, S. and Goldberg, D. E.: Search space boundary extension method in real-coded genetic algorithms, *Information Sciences*, Vol.133, pp.229-247, 2001.
- 6) Duan, O., Sorrosian, S. and Gupta, V. K.: Optimal use of the SCE-UA global optimization method for calibrating watershed models, *Journal of Hydrology*, Vol.158, pp.265-284, 1994.
- 7) Kohonen, T.: The self-organizing map, *Neurcomputing*, Vol.21, pp.1-6, 1998.
- 8) Rössler, O. E.: An equation for continuous chaos, *Physic letters*, Vol.57A, pp.397-398, 1976.
- 9) 合原和幸編: カオス時系列の基礎と応用, 産業図書, 2000.
- 10) Sprott, J. C.: Chaos and time-series analysis, Oxford University Press, Oxford, pp.93-94 and pp.431, 2003.

(2006. 9. 30受付)