II－6

旬降水量時系列のパターン解析とシミュレーション

金沢大学工学部　学生員　上谷　昌史
同　　上　正　員　高瀬　信忠
同　　上　正　員　宇治橋康行

1．まえがき

水文時系列中に存在する変動パターンに着目し、時系列をベクトル系列と見なし、パターン認識の概念を用いたデータの模擬発生手法がUnnyらにより提案された。著者らは、複雑な変動パターンを示すわが国の水文データに適用するため、Unnyらの手法の修正法を提案し、わが国の月降水量データに対し適用し良好な結果を得た。しかし、パターン認識概念に基づく旬単位データに適用する場合、旬単位データの示す強い非正規性のために、通常の変数変換を用いてもパターンクラスのパターン内構造を多変数正規分布で表せないことがほとんどであり、直接旬単位データの解析は問題が残されていた。本研究では、パターンの正規化と特徴抽出を行うことによりパターン内構造を多変数正規分布とし、水文時系列のパターン解析およびデータの模擬発生を行う方法について述べる。

2．特徴抽出とデータシミュレーション

始めに、特徴抽出とパターンの正規化について述べる。特徴抽出にはさまざまの方法があるが、ここではKarhunen-Loeve展開（以下、K－L展開という）を用いた。n個の属性（x_1，・・・，x_n）を持つn次元ベクトル空間R^n上のパターンベクトル　X＝（x_1，・・・，x_n）は、一般に、R^nの一つの正規直交基底を用いて一意に展開される。この展開の正規直交基底として、Xの分散共分散行列の固有値に対応する正規化された固有ベクトルを基底に用いる展開である。すなわち、K－L展開は、

$$X = Z_1\mu_1 + \cdots + Z_n\mu_n \qquad (1)$$
$$Z_i = \mu_i{}^t X$$

と表される。ここに、μ_iはXの分散共分散行列の正規化された固有ベクトルであり，tはベクトルの転置を表す。K－L展開による特徴抽出とは、式（1）のμ_iの係数Z_iの最初のm個を抽出しこれを特徴とするものである。

K－L展開は、パターンの各要素の測定単位により展開係数が異なるという問題がある。この欠点をなくすための1つの手段としてパターンの正規化が行われる。正規化は、

$$X^{*} = W\cdot(X - EX) \qquad (2)$$

$$w_{kk} = 1/\{\sigma_k{}^2\sum_{k=1}^{n}(1/\sigma_k{}^2)\} \qquad w_{kk}\in W \quad k=1,\cdots,n$$

$$w_{ij} = 0.0 \qquad i\neq j \qquad w_{ij}\in W \quad i,j=1,\cdots,n$$

の様に表される。ここで，Xはパターンベクトル、X^{*}は正規化されたパターンベクトル、Wはw_{ij}を要素に持つ変換行列、$\sigma_k{}^2$はパターンベクトルの第k要素の分散、nはパターンベクトルの次数である。この変換は、パターン空間内のパターンを互いにもっとも近くなるようにする変換であり、Clustering-Transformation とも言われる。

データシミュレーションについては、m個の特徴は月単位の場合と同様に、パターンクラスの多変数正規性に基づき、多変数正規乱数を用いて発生させ、残りのn－m個の要素については、各々の要素が独立に正規分布に従っていると仮定し、正規乱数を用いて発生させた。この仮定は厳密には正しくないが、n－m個の要素は認識、分類には重要ではない要素なので許される仮定である。

3．解析結果と考察

解析に用いたデータは、金沢の１８８６年から１９８６年迄の１００年間の旬降水量データである。データのシーズン分割は、月単位データでの解析結果を参考に、１８８６年１２月上旬より９旬づつを１シーズ

ンとし、１年を４シーズンに分割した。得られた４００個のパターンベクトルに対して式（２）の正規化とＫ－Ｌ展開を行った。図－１に特徴Ｚの分散の相対分布を示す。図より、第５要素から分散の相対的な大きさが、分散の大きさが一様な場合の値０．１１より小さくなっていることが分かる。この結果から、式（１）の展開の係数Ｚｉのはじめの４項を特徴として抽出し、４次元の特徴ベクトルを構成した。得られた特徴ベクトルに対し、ＩＳＯＤＡＴＡを用いて分類を行った結果、４００個の特徴ベクトルは１９のパターンクラスに分類された。Kolmogorov-Smilnov検定を用いて各パターンクラスの多変数正規性を検定した結果、各パターンクラスとも多変数正規分布であると見なされた。

パターン解析の結果に基づき、２で述べた手順に従ってデータシミュレーションを行なった。１００年間のシミュレーションを５０回行なった結果を表－２および図－２、図－３に示す。

表より、時系列レベルにおいては、平均、標準偏差、ラグ１の自己相関係数、およびハースト係数の再現性は良好であるが、歪度、尖度等の高次のモーメントの再現性が悪いことが分かる。旬レベルでは、図－２より、平均はすべて観測値がシミュレーションデータの９５％信頼限界内にあり、標準偏差についても、図３よりシミュレーションデータの９５％信頼限界から外れるものが数個あるものの、再現性は良好であると言える。

４．結　語

クラスタリングトランスフォメーションとＫ－Ｌ展開による特徴抽出を行うことにより、非正規性の強い旬単位データに対してもパターン認識の概念に基づくデータシミュレーション法が適用できることを示した。

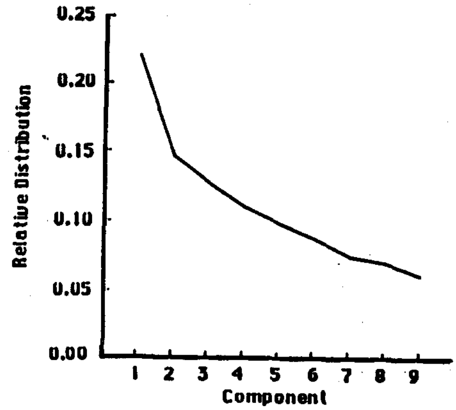

図-1 分散の相対分布

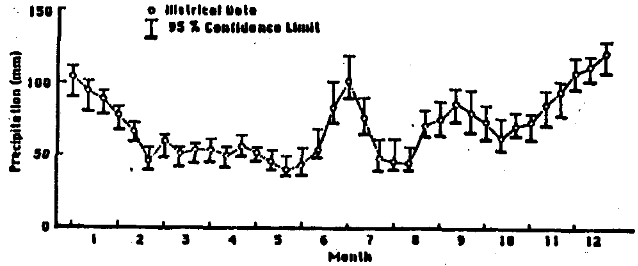

図-2 旬レベルでの平均値の比較

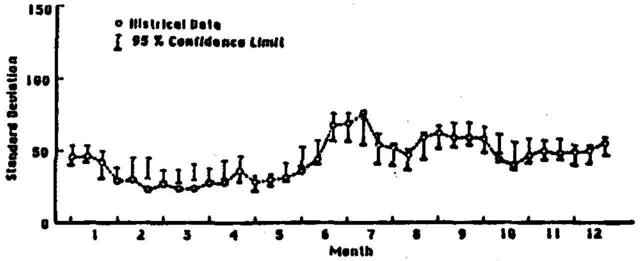

図-3 旬レベルでの標準偏差の比較

表-1 時系列レベルでの統計量の比較

Statistics	Mean	S.D.	Skew.	Kurt.	A.C.	Hurst C.
Generated data	71.5	52.5	0.4935	3.5703	0.2100	0.5403
S.D. of G.D.	0.9	0.8	0.0568	0.1555	0.0171	0.0034
Histrical data	71.3	51.8	1.4499	6.5670	0.1917	0.5646