II-107　旬単位水文量時系列の変動パターンの解析とデータの模擬発生

建設省　正員　上谷 昌史
金沢大学工学部　正員　高瀬 信忠
同上　正員　宇治橋康行

１．はじめに

筆者らはこれまでにパターン認識概念を用いた水文時系列の変動パターンの解析とデータの模擬発生手法を提案し、我が国の月単位水文量に適用し良好な結果を得てきた。この手法を渇水あるいは渇水にかかわる少雨の解析に応用していく場合月単位だけでなく旬あるいは半旬単位の解析が必要となる。この場合、旬あるいは半旬単位の水文量は指数分布に近く強い非正規性を示すため、パターン内構造として当てはめるべき多変数の確立分布とその母数の推定に問題がある。ここでは、パターンの正規化と特徴抽出を行うことにより旬単位水文量の変動パターンの解析とデータの模擬発生を行う手法について述べる。

２．K－L変換による特徴抽出

n個の属性（$x_1, \cdots, x_n$）を持つn次元ベクトル空間R^n上のパターンベクトル $X=(x_1, \cdots, x_n)$ 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をK－L展開（Karhunen-Loeve展開）と呼ぶ。すなわち、

$$X = Z_1\mu_1 + \cdots + Z_n\mu_n \quad (1), \qquad Z_i=\mu^t X \quad (2)$$

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、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 １）Xとその第m項近似との２乗平均誤差を最小とする、２）式（３）で示されるエントロピーを最小にするという２つの意味で最良の展開基底である。

$$H(\mu)=-\Sigma\rho_i \log\rho_i \quad , \quad \rho_i=\sigma_i^2/\Sigma\sigma_i^2 \quad (3)$$

ここに、σ_i^2は式（１）のZ_iの分散である。Z_iの分散σ_i^2は特徴の相対的重要度を示すものであり、より少ない特徴でパターンを近似するという立場からは分散が少数のσ_iに集中していることが望ましい。

K－L展開を行う前にClustering-Transfoamationと言われる次式の正規化を行なった。

$$X^*=W(X-EX) \quad (4)$$
$$w_{ij}=1/\{\sigma_i^2\Sigma(1/\sigma_i^2)\}, \quad i=j$$
$$w_{ij}=0 \quad , \quad i\neq j$$

ここに、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、σ_i^2はパターンベクトルの第i要素の分散である。

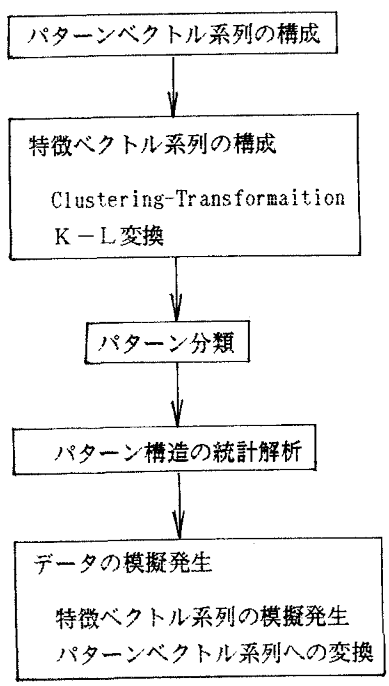

図-1　解析のフローチャート

３．パターン解析とデータの模擬発生手順

パターン解析の手順は、パターンベクトルの代わりに２で得られた特徴ベクトルを対象とする点を除けば月単位の場合と同様で

ある。分類手法としてはISODATAアルゴリズムを用いた。データの模擬発生は以下のようにして行なった。

１）ｍ次元の特徴ベクトル系列を月単位の場合と同様、パターン解析によって得られたパターン内およびパターン間の統計構造に従って模擬発生させる。

２）残された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して模擬発生させる。

３）得られたｎ次元の特徴ベクトル系列をパターンベクトルに変換する。

２）の過程は厳密には正しくないが、特徴抽出の後残されたｎ－ｍ個の要素は認識・分類には重要ではない要素であるので許される仮定であろう。解析のフローチャートを図-１に示す。

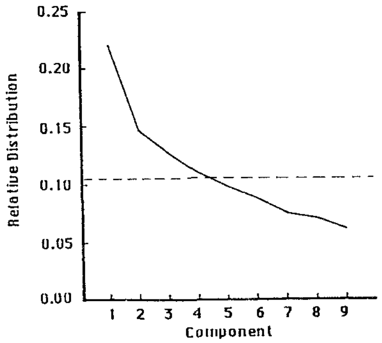

図-２　分散の相対分布

４．実データへの適用結果と考察

解析に用いたデータは、金沢の１８８６年から１９８６年までの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し１２月上旬から９旬３ヶ月を１シーズンとして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正規化とＫ－Ｌ変換を行った。図-２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大きさが各要素の分散が一様な場合の値０．１１より小さくなっていることが分かる。このことから第５要素以下の要素は認識において重要ではないと考えられる。この結果から特徴ベクトルの次元をｍ＝４と定めた。この４次元の特徴ベクトルを用いてパターン解析を行った結果、４００個のベクトルは１９のパターンクラスに分類され、各パターンクラスのパターン内構造は多変数正規分布に従うと見なされた。データの模擬発生は１００年間のデータを５０回発生させた。観測データと模擬発生データの統計量の時系列レベルでの比較を表-１に、旬レベルでの比較を図-３および図-４に示す。結果を見ると、時系列レベルでは高次のモーメントの再現性は悪いが２次までのモーメントは十分再現されている。旬レベルでは平均値の再現性は十分であり、分散の再現性の良くない旬がいくつか見られるが全体として再現性は良好であるといえる。

５．おわりに　パターンベクトルの正規化とＫ－Ｌ変換による特徴抽出を行うことにより、非正規性の高い旬単位水文データに対してもパターン認識概念に基づく時系列の変動パターンの解析とデータの模擬発生手法の有効性が示された。

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

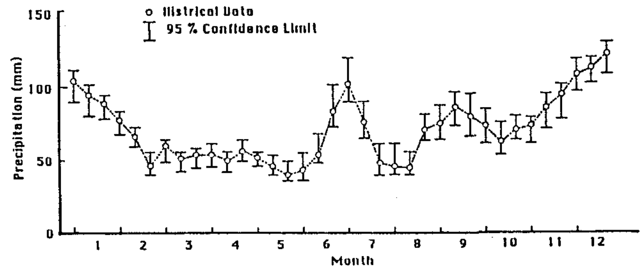

図-３　旬レベルでの平均値の比較

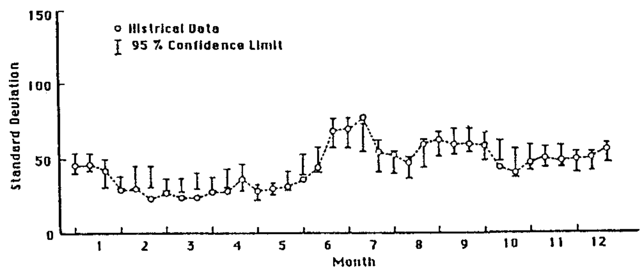

図-４　旬レベルでの分散の比較