

環境科学的データ解析手法の研究*

田中章司郎

広島大学総合情報処理センター

A Study on the Art of Environmental Data Analysis

Shojiro TANAKA

Information Processing Center, Hiroshima University

Higashi-Hiroshima 724, Japan

要 旨

環境の相互依存性 (interrelations) に関して得られている知識は、環境中で複数の要因が同時に作用している多次元的な本質 (multi-dimensionality) によって、昨今の二酸化炭素排出抑止に関する論議の例を引くまでもなく、驚くほど不足している。

環境の相互依存性と因果関係について実測されたデータを用いて解析して、とりわけ定量的に論議の基盤を提示する必要性が存在することは、明白である。

計量的な立場からの定義として“環境”は、実験計画法の適用が事実上困難であって、相互に依存して確率的に変動する、多変数の同時の観測面をもつ対象である、と意味づけることができる。

このような対象の中から、事後的・帰納的に、定量的な法則性を見出す解析手法を確立することが、本研究の目的である。

そのためには、環境の観測面に関する多様な測定データを入手し、それを解析に適したデータファイルの構成に変換する必要がある。しかし、研究者みずから取れるデータの種類と量には限度があり、データの存在する空間的な範囲も限られている場合が多い。

さいわい、国土数値情報、大気汚染・気象等のモニタデータ、地球観測衛星データ、国勢調査・工業統計・農林業センサスなどの指定統計データが、大量に、かつ急ピッチで集められている。これらの情報には、環境全般 (自然、社会、経済等) にわたっての多種多様な数多くの統計的データが含まれている。また、本研究の解析の目的に適した機械可読型データである場合が多い。それらは、環境のデータ解析に好適である。

それらのデータのうち、標準地域メッシュデータ (格子状方形小区画データ) 及びリモートセンシングデータを用いて解析を行うことは、統計量の空間的な分布の把握を可能とする。このことによって、統計量のみでの解析に対して、新たに常識的・体系的な視点を付加することができる。

本研究のデータは、このため、標準地域メッシュデータ、及びリモートセンシングデータを中心とした。

以下、論文各章毎の内容を要約する。

1. 総合解析データファイルの作成

個別的な細分化された調査研究領域のみではなくて、それらの複合された環境中の関連性を解析する場合、それに適合したデータファイルの構造は、比較的に限定的なものとなる。それは、データを反映する区画同士の整合性を基本とすることであり、行政区画、メッシュ、統計調査区、単位流域、国家単位などの、各々のデータが採取された共通の空間的要素のデータを基盤とすることである。

本研究では、総務庁統計局、建設省国土地理院、国土庁（メッシュ化委員会）、農林統計協会、海上保安庁水路部などが作成したファイルの2,000を超える変数を標準地域メッシュを主体として連結し、他に座標点毎・単位流域毎のデータも、変換により、同時に使用できるように工夫した。

このことによって、第3の変数や潜在変数の影響を考慮する必要が軽減して、おもな環境因子を具体的に単離・発見したり、逆に組み合わせ、変数間の相互の関連性を見出し易くできる。

2. リモートセンシングデータ解析システムの構成

特定の研究室等に限定されず、また大学構成員であれば誰でもが使用できる施設として、本研究は大学計算機センターのホスト計算機及び端末を活用した。最近の主流となったインテリジェント端末（パーソナルコンピュータ）の強力なローカルグラフィック機能と、ホスト（大型計算機）の高速なデータ処理能力の両者を結合させることにより、大量の統計計算と高速な画像表示を両立させることができた。

このシステムは MOS-1、Landsat MSS, TM、NOAA、航空機等のマルチスペクトルスキャナー（MSS）型 デジタルデータ全般について、各種の統計解析・画像表示をする事ができる。

3. リモートセンシングデータ分類手法別精度の検討

建設省国土地理院作成の国土細密数値情報を、画素対画素で一致するように Landsat MSS データと併合した。

このことによって、地上調査データを分類基準データとして、Landsat データのサブシーンを全てカバーできるので、試験領域の全数調査をすることができる。このため分類基準データの量的代表性を考慮しつつ、広義の最尤法に関する8分類手法別・試験領域（母集団）内での分類基準データ（標本）の大きさ別の、計16条件の比較検討をおこなった。

その結果、昭和54年の埼玉県川口市近辺の、土地被覆形態が複雑で多様な 192 平方キロメートルを試験領域とした場合、適用した分類手法・分類条件によって、各土地被覆項目に対する分類精度に大きなばらつきがあり、また、各分類区分の試験領域中における相対度数である事前確率を、その出現頻度に比例してあてはめた場合に、出現頻度を等しいとした場合と比べて、全体として約10～15パーセント精度が向上することが判った。

4. 人口動態と土地利用に関する探索的データ解析

仮説を検証するための確認的データ解析に対して、データから何らかの構造を探索的に探りだすことを目的とした探索的データ解析手法は、その意味からも、非実験的な実際の環境データの解析に、まさしく好都合である。

その応用の対象として、最も身近な人口と土地利用のデータを用いた。昭和50年の広島県のデー

タを探索的に解析した結果、基準地域メッシュあたりの総人口対数値と同メッシュあたりの森林面積比率との間に、右下がりなロジスティック式で近似できるデータ構造が存在することが明らかになった。

5. 人口増加による森林伐採のメカニズムの推定

前章において、単位区画あたりの森林面積比率 (Y) は、同区画あたりの人口 (N) の増加とともに、きれいなシグモイド曲線を描いて変化 (減少) している。そこで、Y を N の関数とする。

増加したヒト一人あたりの森林面積比率の減少量は、移入などによる人口増加の直後の全員のヒト一人ひとりが、増加直前の森林の木々との間で、理論的に出会う機会に比例すると仮定すると、分子衝突説からの類推によって、以下の式を得ることができた:

$$Y = Kr \exp \left[-\frac{r}{2} \{ \exp(X) - 1 \}^2 \right]. \quad \dots\dots (1)$$

ここで

- r: 森林伐採の係数 (deforestation coefficient),
- X: 原点移動後に N を $p=0$ でべき乗変換した変数 $\{ \log_e(N+1) \}$,
- Kr: 初期森林比率値.

ゴンパーツ関数の特殊型である (1) 式を 総合解析データファイルのデータを用いて非線形回帰により評価したところ、メッシュ毎の森林面積比率と同人口密度対数値との関係に影響を及ぼす第3の変数として地形の起伏量が存在するものの、緩斜面の区画では、安定して (1) 式が成立することがわかった。

6. 総合考察

環境データの計量的な取扱いの意義と、そのことによって得られた定量的なモデルを環境計画へ拡張する際の可能性と問題点を検討した。その結果、今まで複雑で因果関係の推定や追跡が困難であった複合環境中の環境因子相互の動態を解明する際に、本研究の手法を適用することが、有効な方法のひとつであることを示した。