



気象データ処理法

丸山 健人*

1. はじめに

気象学の諸分野において、観測・実験データに基づく調査・研究に従事している人々にとっては、意識するとしないにかかわらず、“気象データ処理法”の問題に直面している。しかしながら、“気象データ処理法”が特別に論じられることはあまりない。ふつうは、それぞれの分野のデータ解析法の問題として、あるいは気象統計の問題として、あるいはさまざまなモデル実験の検証上の問題として論じられて来た。したがって、実際に自ら扱っている現象についての観測・実験から得られたデータを処理するには、とりあえずはそれまでの人々がやってきた伝統的な方法を踏襲することから始める。それを、単に追試・模倣とのみ見るべきではない。なぜなら、得られた結果は多かれ少なかれデータ処理法に依存するのであって、従前のものと共通性をもつ方法によらないと、結果自体を比較検討することがしばしば困難になるからである。ごく簡単な例で、平均値の求め方ひとつとっても、観測値の各種の加重平均のし方で値も違ってきて、単純な比較ができないことがある。データ処理法が複雑になれば、手法の違いによる結果の違いの吟味も複雑になる。

今日、“気象データ処理法”が特に取り上げられ始めたのには、それなりの事情がある。いうまでもなく、電子計算機=コンピューターの発達と普及である。単に“データ処理法”といえば、コンピューターによるデータ処理法を意味するほどである。自然現象の観測データのなかでも、気象データほど大量のデータが日々得られるものはないであろう。また、大循環モデルなどの数値実験からもぼう大な計算結果が得られる。それゆえ、気象分野においては、つねに最大級のコンピューターが導入されてきた。とはいえ、“気象データ処理法”はわが国ではまだ主要な問題の一つにはなり得ていない。大きな観測実験が企画され実施されても、気象技術者、研究者が

自らも参加して作り上げた大量のデータが、単なる数字の羅列のまま残され、あるいは磁気テープに入ってしまったら、コンピューター・サービス会社に高価な代償を支払わなければ、実際に見ることもできない、というのが現状である。データ処理システムが、単に気象行政部門にとどまらず、気象学の調査・研究部門においても整備されることが求められている。

2. データ処理の各段階

従来、データ処理に直接かかわる問題は、主に統計学の分野から論じられてきた。実際、データ処理にあたっては、多かれ少なかれ統計学の知識が求められ、活用されてきた。しかし、データ処理にあたっては、統計学以前の問題が前面に出てくる。以下、データ処理の段階を追って論じてみたい。

(0) まず、そのデータが利用できる形に再編集されることが必要である。コンピューター時代に入った今日、多くのデータが大量に処理できるようになった反面、前述のようにこの段階でつまづき、処理されず死蔵され、あるいは捨てられていくデータもぼう大なものになっている。気象データ集が、古い時代のものは立派に印刷され製本されて書庫に並んでいるのに、最近のものは印刷も製本も略式になり、さらにはマイクロフィルムになり、あるいは今日では磁気テープになって書架から姿を消しているものもある。印刷されたデータ集なら目で見てデータの概略を知り、必要な部分をコピーすることができる。磁気テープの場合も、ディスプレイ装置などで同様なことが気軽にできるようなシステムがなければならない。同時に、コンピューター・システムは、簡単なプログラムをすることにより、利用者の目的に応じてさまざまな形での再編集を可能にした。図書館、資料室の機能がコンピューター時代にふさわしいように拡張されていくことが求められている。いずれにせよ、利用しやすい形のデータ・セットを用意すること自体、重要な仕事であり、これを公表し、評価する何らかの場を設ける必要があるように思う。

* T. Maruyama, 気象研究所予報研究部

(1) こうして利用できる形になったデータ・セットを用いて、何らかの特徴、なんらかの規則性を見い出す試みがデータ処理の第1段階と言えよう。この段階は、とりわけ気象データ処理の場合、主要な内容を占めている。気象学会での研究発表の大多数はこの段階のものである。気象界に限らず、科学研究から社会生活にいたるまで、この段階のデータ処理が広範に行なわれている。最近では、環境行政上の必要から官庁・自治体等で大気環境を測定し、その解析がデータ処理会社に外注されることが多くなった。内容的にはこの段階のデータ処理であり、官庁・自治体等では副次的な扱いを受けているこの段階のデータ処理がこれらの企業では主要な内容となっており、必要に迫られてさまざまな処理システムや部内資料が作られている。

(2) さて、上記の結果、与えられたデータ・セットについてのさまざまな特徴や規則性が示唆されたならば、それらの統計的有意性を確かめるために計算精度を上げたり、データを増やしたり、あるいは気象学的有意性を確かめるための関連検証を行なう。ここでは統計学や気象学の基礎理論が活用される。気象データの場合、観測・測定のやり直しがきかない、類似条件の再現が必ずしも頻繁でない、などの事情から、統計的有意性の追求が容易でないことが多い。したがって、気象学的有意性の追求に力を注ぐ必要がある。たとえば、静力学の関係が仮定できるとか、ある種の波の伝播方向が理論的に明らかにされているとかの知識がデータ処理のこの段階で積極的に活用される。

(3) データ処理の次の段階は、モデル化である。モデル化への接近は、必ずしもデータ処理の側からだけでなく、理論的研究の側からもなされ、観測・実験に基づく研究と、理論に基づく研究との接点になっている。



研究会などの口頭発表では、上記(1)の段階にとどまる例が多いのに対し、学会誌などの論文発表では、(1)の段階が比較的軽く扱われ、(2)と(3)に重点が置かれる例が多い。そのため、論文集を読むだけでは(1)の段階のことがよくわからない。というよりは、(1)の段階にとどまって(2)(3)へと発展し得なかった数多くの調査・研究がその底辺に秘められていると考えられる。以上、データ処理を段階に分けて論じてみたが、ここで狭い意味で“データ処理法”というなら、(1)の段階のことを指しているように思う。そこで本講では、以下これについて論じてみたい。

3. 探検的データ解析

上記(1)の段階のデータ処理に対応する呼称として、McNeilの入門書¹⁾は“Exploratory data analysis”という語を用いている。とりあえず“探検的データ解析”と訳して本節の標題とした。さて、気象観測データについて、平均値、最高値、最低値、最多値などの統計値を示すことは、伝統的に行なわれてきた。これらはコンピューターの登場するはるか以前から行なわれていたが、これによってデータ・セットに含まれる特徴の概略をつかむことができる。探検的データ解析とはまさにそのようなものであって、コンピューター時代に入ってから、以前なら計算量がぼう大なため広く行なわれなかった相関計算やスペクトル解析なども、探検的データ解析に仲間入りするようになった。そして、これらのデータ処理法はコンピューター・プログラムとして与えられ、計算センターで既成のプログラムとして備えられているものもあれば、解析者自身で作ることもできる。既成のプログラムは、与えられたデータ・セットとはあわせにくいことが多い。その点、自分で作ったプログラムは便利だが、思わぬ誤まりが入りがちで、充分試練をへたものではないと心配なこともある。手作りのプログラムでは、サブ・プログラムを多く使ってわかりやすく作ること、処理結果のプリントは労を惜しまず見やすい形にすることなどが大事であろう。自分の計算を行なうコンピューターや計算センターの機能に合わせて作るのがよい。こうして試練をへた手作りのプログラムをたくさん持つことにより、その解析者はそのセンターを自分のデータ処理センターに変えることができる。

以下、探検的データ解析の標準的な手法について概説する。

(1) 度数分布 気象データにおいては、観測値が暗黙のうちに平均値のまわりに正規分布していることが想定され、それについて議論されることが多い。しかし、実際にどんな分布をしているか確かめてみる。また、コード化されたデータについて各コードの度数を調べてみる。

(2) 総和、平均値、分散、標準偏差 計算上の定義は周知のとおりであるが、分散は平均値からの偏差の二乗平均、標準偏差はその平方根として与えるのがふつうである。実際上の問題は、欠測・異常値の処理である。コンピューター・プログラムには、これらに対する処置を含んでいることが望ましい。多くのコンピューター・システムでは、異常計算(ゼロで割るとか、マイナスの平方根をとるとか)をするとメッセージを出し、一定回数

出すと計算を打ち切ってしまう。とくに異常値については、そのデータについてある程度知っていないとこれをチェック・アウトすることがむずかしい。欠測・異常値の取り扱いについては以下に述べている問題でつきまとう。

(3) グラフ、等値線 1次元量ならグラフに描いて、2次元量なら等値線をひいて、そのデータの特徴を知る。データの量が多くなると一枚一枚手で描くのでは探検的データ解析の主旨に反する。プロッターが使えれば便利であるが、ライン・プリンターでも簡便なプログラムでグラフ・等値線を示すことができる。

(4) 直線のあてはめ グラフや相関図に最小二乗法で直線をあてはめることはふつうに行なわれる。気象データの多くは時系列データとして与えられる。単調増加あるいは減少傾向の簡単な指標として直線をあてはめ(リニア・トレンド)、その勾配で示す。

(5) スムーズィング(平滑化) こまかい変動を取り除くために、あるいは逆にトレンドを除いてこまかい変動を残すために、グラフのスムーズィングが行なわれる。コンピューター以前には単純移動平均がよく用いられたが、今では各種の加重移動平均が工夫されている。加重移動平均のプログラムを用いるとき、サイン波を与え振幅がどうかわるかをテストしてみるとよい。

(6) 周期分析、波数分析 気象データの解析において、振動・波動現象の解明は最大の興味の対象の一つであろう。分散あるいは標準偏差がそのデータのばらつき度合を示す伝統的尺度であるとするれば、周期分析・波数分析はそのばらつき方の時間あるいは空間スケールを示すコンピューター時代にふさわしい尺度と言えよう。時系列データのときは周期分析、空間データのときは波数分析、また空間の2次元波数分析や、時間=空間の2次元周期・波数分析などがあり、これらを総称してふつうスペクトル解析と呼んでいる。これについてはさらに次節で述べる。



このほかにも、探検的データ分析の手法としてはいろいろ工夫されてよい。要は、データ・セットがぼう大なとき、まず行なうものである。そうしないと、“木を見て森を見ず”ということになり兼ねない。

4. スペクトル解析

スペクトル解析を実際に行なうてみたいという人には、日野幹雄氏の「スペクトル解析」という本²⁾が役立つであろう。参考書等も多く紹介されている。本節では、気象データ処理法の一つという観点から論じてみたい。任意

の周期函数は一義的に三角函数の級数に展開できる。このことから、与えられたデータ・セット(ふつう時間・空間について等間隔に与えられる)を三角函数の級数に展開すれば一義的にスペクトルが決まるように簡単に考えがちであるが、そうではない。第1に、与えられたデータ・セットは必ず有限の長さ(時間)のものであり、その長さを越えた波長(周期)の現象を記述することができないので、その長さ以下の波数で表わされてしまう。第2に、そのデータ・セットは一定間隔ごとの値であり、その間隔程度、あるいはそれ以下の波長(周期)の現象は記述することができないので、その長さ以上の波数で表わされてしまう(これをエリアジングと呼ぶ)。したがって、スペクトル解析を行なうには、データ・セットの長さを越えるトレンドを取り除き、また、データ間隔に近いスケールの変動はなんらかのスムーズィングを行なってこれを取り除く手続きが必要なことがある。これらの手続きをへてもなお、データ・セットが有限の長さであることに起因して、単一の波が単一の波として分離されない。すなわち、多かれ少なかれひろがりをもったスペクトルとして表わされる。このひろがり具合は以下に述べる手法に依存する。

数学的には、データ・セットの長さを長くするほどひろがり具合が縮まる。しかし、これは同じ波や振動がずっと続いている場合のことであって、実際の気象現象の場合、その波や振動は有限の空間・時間の範囲に限られているのがふつうである。したがって、観測空間や時間を長くしさえすれば現象が明瞭に分離されるというふうにはならない。むしろ、異質の波や振動が加わってきて、結果を複雑にする。たとえば、冬季に顕著な偏西風波、夏季に顕著な偏東風波が現われるような領域でのデータを用いるときは、3カ月ごとの期間についてスペクトル解析を行なう方が1カ年まとめて行なうよりも現象の特徴をよく表わすことができよう。したがって、解析者は手法に依存する特徴とデータ・セットに含まれている現象に依存する特徴とを考慮しながらデータ処理をすすめていくことになる。以下、ふつうに行なわれる手法について概説する。数式やプログラム例は、上記入門書を参照されたい。

(1) 単純フーリエ分析 与えられたデータ・セットに対し、フーリエ級数の定義どおり展開し、その振幅と位相を示す方法である。コンピューター時代以前から行なわれてきた手法であり、与えられたデータ・セットの中に1ないし整数個の波がちょうど入っているという場合

に便利である。たとえば、4カ年分の時系列データがあれば、4年周期、2年周期、4/3年周期、1年周期……というふうにスペクトル値が定義される。4/3年周期を別にすれば、4年周期と2年周期と1年周期の振幅を比較したり、1年周期の位相角を要素間あるいは観測地点で比較したりすることができる。単純フーリエ分析は、データ・セットの長さの半分の項数の波まで、たとえば上述の例でデータ間隔が0.5月とすれば1カ月周期の項まで展開できるが、気象学上有意なスペクトル値が期待されるのは、展開できるもののはじめから半分、上述の例では2カ月周期の項あたりまでと考えてよい。項数がすすむと周期の差が縮まってきて、たとえば第23項目は2.09月、第24項目は2.00月、第25項目は1.92月というふうになり、気象学的に意味をもつ現象としては区別する意味が失われてくる。そこで、項数のすすんだところでは、個々のスペクトル値よりもその周辺いくつかのスペクトル値の振幅が高まっているかどうか注目する。すなわち、その周辺のスペクトル値は、ある気象現象の独立した測定値と考えてみる。あるいは、データ・セットのなかの一部分に集中した波(波束)を表わすものと見ることでもできる。

(2) パワー・スペクトル、クロス・スペクトル 上述の考え方をわかりやすくするため、単純フーリエ分析の結果をさらに若干の計算処理をして表示する。まず、振幅の2乗の半分をパワー・スペクトルと定義する。パワー・スペクトル値の全部を積算すると分散(variance)になるので、分散に対するいろいろな周期帯の寄与の度合を示すとみることができ、パワー・スペクトル値の相並ぶ数個をスムージングして示すのがふつうである。このスムージングの幅は、スペクトル値の統計的信頼度を与える尺度になる。幅を大きくすると統計的信頼度は増すが、スペクトルの分解度が落ち、気象学的に有意な現象を分離することができなくなる。気象学的に有意な現象を分離するには、2量についてのデータ・セットを用い、フーリエ係数を組み合わせ、スムージング幅を与えることによって求まるクロス・スペクトル値(振幅・位相差、コヒアレンス)の評価が有用である。コヒアレンスは2量間の関係の度合を示す。これらの評価する方法がコンピューター時代に入って開発された。

(3) ラグ相関分析 与えられたデータ・セットのラグ(ずらし)相関を求めフーリエ変換する方法。Blackman・Tukeyが1950年代の後期に考案したので、Blackman and Tukey法とも呼ばれる。自己ラグ相関にラグ・ウ

ィンドーという加重(スペクトルのひろがりをおさえるもの)を行なった上でフーリエ変換することにより、パワー・スペクトルを与える。2量間の相互ラグ相関を求め、一定の手続をへてフーリエ変換するとクロス・スペクトルを与える。分解の度合を増すにはラグ相関をとる長さを長くすればよいが、スペクトルの統計的信頼度が落ちるので、見たい現象のスケールとデータの長さを検討してラグの長さをきめる。経験的には、データの長さの1/10以下が望ましいと言われるが、大規模大気じょう乱を扱う場合は、1/4程度まで行なわれている。

(4) FFT (Fast Fourier Transform) 上記(2)(3)の方法は計算効率が悪いので、フーリエ変換を大量に行なうとき効率的に処理する方法として、1965年 Cooley・Tukeyにより考案されたという。データの長さを2のべき(2^P)にそろえることにより表われる同じ数値の計算処理が一括してできるというのがその原理であるが、既成のプログラムはデータの長さが必ずしも2のべきになっていなくても一定の条件付で計算できるようになっている。

(5) MEM (Maximum Entropy Method) 上記の方法とは発想が異なり、情報エントロピーという量を定義し、エントロピーが増大しないように(最大値をもつように)共分散関数を定義してパワー・スペクトルを求める方法。1967年、地震波の解析に関してJ.P. Burgにより提案されたという。データの長さに関わりなく分解の度合がきわめてよいのが特徴である。



パワー・スペクトルを分解度を上げて求めたいというときは、MEM法がますます使われるようになってきている。クロス・スペクトルを求めるときには(2)~(4)の方法を用いる。既成のプログラムが公表されているので、たいいていの場合に間に合う。実際の計算をする場合には、サイン波の組み合わせや、既知のデータ・セットをいくつか試験的に通してみたほうがよい。結果がどのように出るかを知っておくことは、本番の計算結果の理解を助けるし、まずプログラム・カードの誤まり(パンチ・ミス等)の発見に役立つ。もっとも、MEM法の場合はサイン波のみを与えると分解が良すぎてスペクトル値が定義されないこともある。

6. むすび

以上、“気象データ処理法”の標題のもとに書きすんできた。別の執筆者が書けば別の書き方があると思う。スペクトル解析としても、サイン級数に展開する方法の

ほかに、任意の直交関数系に展開する経験的直交関数系展開（因子分析，主成分分析）の方法があり，天気図のパターン分類等に広く活用されつつある。また，気象庁のルーチン・データの客観解析についても独特なデータ処理法がある。これらについては，別の執筆者が予定されている。最後に，筆者自身の“気象データ処理法”との出会いを述べてむすびとしたい。

(1) 物理実験学 1960年，平田森三先生のこの講義は他の学問分野の講義とは異質の感じを持った。図書室の利用法や参考文献の整理のしかた，数の教え方のようなお話もあった。ディメンジョン解析，測定精度とそのばらつき等が主要な内容であったが，この講義が筆者にとっては“データ処理”の問題への最初の接近だったように思う。

(2) 実験整約法 1961年，坪井忠二先生のこの講義は，まさに今日でいう“データ処理法”であった。まだコンピューター時代に入っていなかった当時のことで，手まわし計算器の使い方のようなお話が多かったが，内容的には，度数分布，平均，分散，標準偏差，相関係数，直線や曲線のあてはめ方（フーリエ解析もサイン波のあてはめとして説明された）などであった。フーリエ分析は孔のあいた紙を何種類か作り，その孔に出てくる数を加算して係数を求めた（FFTの発想である）。

(3) FORTRAN 入門 1962年，松野太郎先生のこの講義（演習）は，コンピューター時代への展望を与えてくれるものであった。実際に筆者にとってこれが威力を発揮したのは，1966年以後の熱帯高層気象データの処理であった。その後，筆者の作成したデータ・カードはいつも場所をたくさんとして迷惑がられている。

(4) データ・ゼミ 1968年，筆者が東京学芸大学の学生有志を集めて一緒に行なったゼミ。当時同大学にはコンピューターが入っていなかったが，コンピューター時代を意識して行ない，同大学の学生有志が行なっている気象観測のデータが処理された。今日では，これらのデータはコンピューターで処理されるようになっていく。

(5) 情報科学入門 1970年，島貫陸氏が行なった講義（演習）。気象学を情報科学の見地から論じたものであった。この内容は同氏の著書「数理情報科学」³に盛り込まれている。量としての数字とコードとしての数字をたくみに使って気象データを表わすことが論じられた。



コンピューター時代以前からコンピューター全盛時代まで，振り返ってみると“データ処理法”のなかに一貫

した精神が貫かれていることに気づく。それは，一見乱雑にみえる数字の羅列から何らかの特徴，規則性を見出すことであって，そこが単なる“コンピューター利用法”とは区別されるところであろう。本講が，コンピューターを自由に利用できる環境にいる人々のみための入門講座にならないことを望むものである。

謝 辞

本講をまとめるにあたり，気象研究所予報研究部長内田英治氏はじめ，同研究部の方々に御意見，御助言をいただきました。また，日本科学技術研修所の板宮憲一氏には有意義な御討論をいただきました。

文 献

他の講座と違って，本講では入門書等を列記することはしない。以下に掲げるものは筆者が比較的愛用，愛読してきたもので，上記講義等とともに筆者にさまざまなアイデアを提供してくれたものである。

1 McNeil, D. R., 1977: Interactive Data Analysis, Wiley-Interscience, 186 pp.

“Exploratory Data Analysis”という規定がおもしろく的を得ている。Princeton Universityにおける講義ノートをまとめたものという。ただし内容は，演習例とコンピューター・プログラムが中心である。

2 日野幹雄, 1977: スペクトル解析, 朝倉書店, 300 pp.

実際にスペクトル解析をやってみようという人にはとりつきやすい本である。豊富な実例と参考文献リスト，Blackman and Tukey法，FFT法MEM法，Munk法のコンピューター・プログラム例がついていて，すぐ役立つ。

3 島貫陸, 1971: 数理情報科学, オーム社, 188 pp.

副題に“データの理解からコンピューターへ”とあるように，データ（数字の配列，量としての数値とコードとしての数値）に限りない愛着をよせ，論じられている。与えられたデータを処理する立場からだけでなく，自らデータ・セットを作る場合にも参考になる。

4 渡辺次雄, 1958: 近代気象調査法, 技報堂全書, 302 pp.

5 荒井隆夫, 渡辺次雄, 1960: 天気学, 技報堂全書, 357 pp.

上記2冊は，気象データを具体的に処理しようというとき役立つ示唆に富んだ本。ただし，コンピューター時代以前のものなので部分的，断片的にとどまっている例も多い。

6 日本気象学会編, 1972: 気象力学に用いられる数値計算法, 気象研究ノート, 110, 158 pp.

気象データ処理を直接の目的としてまとめられたものではないが，境界値問題，初期値問題，数

値予報モデルの時間積分、気象じょう乱のスペクトル解析など、気象データ処理にかかわる事項が数値計算法を示して取り扱われている。

- 7 日本気象学会 編, 1973: 大気境界層内の理論と乱れの測定, 気象研究ノート, 114, 110 pp.

スペクトル解析はとりわけ乱流観測データ解析の主要な手法である。解析上の諸問題がのべられており、乱流以外の対象を扱う時にも参考になる指摘が多い。

- 8 日本気象学会 編, 1977: スペクトル解析, 気象研究ノート, 131, 74 pp.

この中で扱われている林良一氏の空=時間2次元スペクトル解析法は、大気大循環モデル実験結果の処理に適用されたものであるが、多方面での活用が期待される。

- 9 丸山健人, 1975: 大規模大気擾乱のスペクトル解析, 天気, 22, 267-280.

筆者が大規模波動擾乱の解析に用いた方法を解説したものであるが、解析法そのものは Munk *et al.* (1958)* が海洋波浪の解析に適用したものである。サイン波や乱数を与えるとどういう結果になるか例示してある。基本的には Blackman and Tukey 法であり、これの FFT 版や MEM 版が出されることを期待したい。

* Munk, W.H., F.E. Snodgrass, and M.J. Tucker, 1958: Spectra of low-frequency ocean waves, Bull. Scripps Inst. Oceanogr., 7, 283-362.

- 10 Hayashi, Y., 1977: Space-time power spectral analysis using the maximum entropy method, J. Met. Soc. Japan, 55, 415-420.

空=時間2次元パワー・スペクトルをMEM法で求める方法。プログラム例付。



続 気象学入門講座

これからの予定

(太字は既に掲載されたもの、カッコ内は掲載された巻号)

- 気象学へのガイダンス (25. 4)
 [基礎コース]
 気象解析の手引き (25. 5)
 気象力学・気象熱力学 (25. 6)
 気象放射学
 高層大気物理学入門 (25. 5)
 雲物理学・降水物理学 (25. 8)
 大気電気学・大気化学 (25. 12)
 気象観測と気象器械
 気象統計について (25. 7)
 気候学
 生活と気象 (25. 6)
 [アドヴァンスト・コース]
 気象予測論 (25. 7)

- 回転流体力学を学ぶために(25. 6)
 対流論 (25. 6)
 中小規模現象の気象学 (25. 11)
 大気大循環論 (26. 2)
 エーロゾルの気象学
 気候変動論
 熱帯気象学 (25. 8)
 高層大気力学の諸問題 (25. 9)
 高層大気物性 (26. 3)
 大気境界層の物理
 衛星気象学
 レーダ気象学
 惑星気象学 (25. 7)
 自動気象観測(隔測)・通報システム

- 応用気象学
 大気汚染の気象学
 実験気象学 (25. 10)
 天候・気候変化の気象学
 海洋気象学 (25. 9)
 極気象学
 気象災害論 (25. 9)
 気象教育論
 気象データ処理法 (26. 4)
 [研究のすすめ方]
 最近の気象資料
 論文の書き方
 気象学教科書・参考書のリスト