

気象・海洋データに対する多変量解析*

谷 本 陽 一**

1. 背景

計算機や人工衛星技術の高度化により高頻度高解像度の気象・海洋データセットが数多く提供され、膨大なファイルサイズのデータをときに複数個取り扱う機会が増えてきている。データセットの有用性を知る上で、対象とする領域・時期における変動特性を検証する作業は欠かせない。しかしながら、気象・海洋データは時間変動する3次元場であり、その変動特性は一般に複雑である。このため、平均場や標準偏差といった基礎統計量からだけでは変動特性の全容を知ることができない。多変量解析は、さまざまな変動特性を含むデータセットの中から統計的に主要な成分を効率よく抽出することができるため、全球の客観解析データセットや再解析データセットに対して広く用いられてきた。本解説では、過去の研究で用いられた多変量解析の利点を紹介するとともに、論文の中ではあまり記述されない多変量解析の弱点についても言及する。

2. 主成分解析の概要

多変量解析では最も代表的に用いられる主成分解析(あるいは経験的直交関数解析)の基礎とその利点を概説する。いま、多数の格子点を持つある平面上の変数(例えば、ある圧力面の気温等)の時系列データ x_i^l (l は時刻で $l=1, \dots, N$, i は格子点で $i=1, \dots, M$ とする)を考える。ここで、各格子点での時間平均はゼロとする。2格子点間の相関係数 $r_{ij} = \frac{\sum_{l=1}^N x_i^l x_j^l}{\sigma_i \sigma_j}$ (σ は時間方向の標準偏差)として、すべての2点の組み合わせの r_{ij} を含む $M \times M$ の相関係数行列 V を求める。行列 V を固有値分解して得られる第

k 主成分の固有ベクトルを p_i^k ($k=1, \dots, K$ で、多くの気象・海洋データの場合、有効な K は $K \ll N < M$)とすれば、第 k 主成分の時間スコアは $a_i^k = \sum_{l=1}^M x_i^l p_i^k$ として求められる。元の時系列データは $x_i^l = \sum_{k=1}^K a_i^k p_i^k + \varepsilon_i^l$ (ε_i^l は残差項)のように主成分の重ね合わせとして表現できる。

a_i^k の分散量に相当する固有値が最も大きなものを第1主成分と呼び、その固有ベクトル p_i^1 は M 次元空間で x_i^l の軌跡の拡がりをもっと良く表現する向きの方角ベクトルに相当する。第2主成分以降は、残された主成分の中から最も固有値の大きく、かつより低次の成分と直交するベクトルが順次求められる。

相関係数行列の場合は格子点ごとの分散量のばらつきは無視し、各格子点の変動は規格化されている。格子点ごと分散量のばらつきを考慮するには、相関係数行列の代わりに、分散共分散行列を採用すればよい。

V の行成分を地図上に投影したものは1点相関図と呼ばれる。仮に $M=100$ とし、100枚の1点相関係数図を作成した後、相関係数の空間形態が似ているもの同士をいくつかのグループに整理できたとする。無論、どのグループにも属さない分布図があってもよい。あるグループに50枚以上も属するような場合、その空間形態に伴う変動は“主要な”変動成分と強く関連していると言って良いだろう。実際、第1図に示すように冬季500 hPa高度偏差場に対する第1主成分とその活動中心域を基点とした1点相関係数分布の空間形態はよく似ている¹⁾。Wallace and Gutzler (1981)¹⁾はこれら2つの手法に基づき、500 hPa高度偏差場に見られるテレコネクションパターンを見出した。

主成分解析の代わりとして、1点相関係数の値がなくなるわけではない。1点相関係数は、解析の対象とする領域に依存しない点で優れている。一方、主成分解析は対象とする領域が異なれば、それぞれの領域で主要な成分は異なってくる可能性がある。多数の1

* Applications of modal decomposition analyses to atmosphere and ocean data sets.

** Youichi TANIMOTO, 北海道大学大学院環境科学 院・地球環境科学研究院/海洋研究開発機構地球環境 フロンティア研究センター。

© 2007 日本気象学会

点相関係数分布図から“主要な”成分を取り出すことを考えた場合、これらの2つの手法の相違は、「対象とする領域が異なると、相関係数の空間形態のグループ分けの結果が異なる」こととして解釈できる。

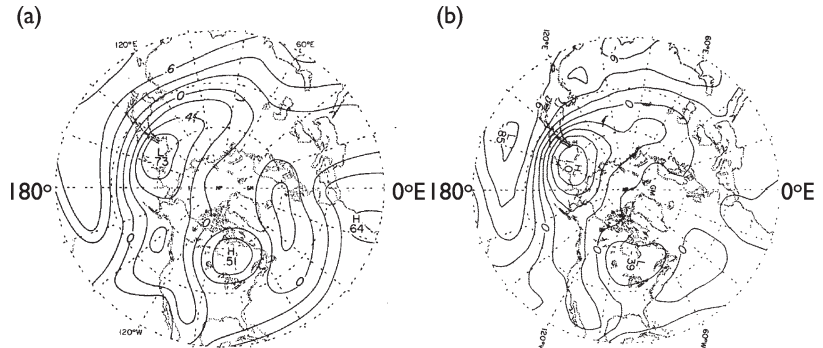
主成分解析の利点は、比較的簡便なプログラミングや統計解析パッケージソフトウェア等を用いることにより、対象とするデータセットの変動特性を空間と時間を分離して表現できることである。観測データや再解析データと数値

モデル出力の時空間変動特性の検証する際などには、それぞれのデータから得られる低次主成分に対する比較から両者間の大まかな相似性を示すことができる。別の利点として、主成分同士は直交なのでそれぞれの主成分を個別に議論できることがあげられる。主成分の独立性の議論については、独立成分分析を気象データに適用した研究成果に詳しい²⁾。

3. 多変量解析の使用上における一般的な注意点

主成分解析の場合、対象とする領域だけでなく、空間分解能、期間、サンプリングが異なると、主要成分の固有ベクトルや時間スコアの結果が異なってくるがあるので、結果の解釈には注意が必要である。特に、ある特徴的な現象を表現するインデックスが主成分解析に基づいて算出された場合（例えば、北極振動/北半球環状モード³⁾、北太平洋海面水温10年規模変動⁴⁾など）は、期間が延長される度に固有ベクトル（つまり、空間形態）が少しずつでも変化するかもしれないし、時間スコアも当初のものとは延長前の期間の値も含めて少しずつ変わってしまう可能性がある。このような事態を避けるためには、当初のインデックスを求めるのに用いた固有ベクトル場を固定すれば、それを延長された期間のデータに投影することで、当初のインデックスは保持される。ただし、その固定された固有ベクトルの場合は、延長された期間でも主要な成分である保証はない。インデックスがある意味一人歩きするような状況になった場合は、多変量解析の時間スコアに固執せず、むしろ簡便に領域平均などを採用した（例えば、東西風指数や南方振動指数など）ほうが期間の延長には対処しやすい。

各主成分の強固さ（robustness）を検証するには、期間を前期と後期に2分し、同じ解析手法を適用し



第1図 (a) 冬季北半球500 hPa 高度偏差場における第1主成分時間スコアに対する相関係数分布図。(b) 同偏差場の60°N, 155°Eを基点とした1点相関係数分布図。Wallace and Gutzler, 1981¹⁾, Fig. 24a と Fig. 27a から引用。どちらも西部北太平洋における2つの活動中心域を示す。

て、それぞれの結果に明確な違いがないことを確かめればよい。期間を2分する代わりに、サンプリングを時間方向あるいは空間方向に1つおきにする方法もある。いずれの検証用データセットはオリジナルデータと全く同じではないため、上述のように各主成分の固有値・固有ベクトルに違いをもたらす得る。しかし、オリジナルデータと検証用データに対する結果において、固有ベクトルの相似性（空間相関が高いことを意味する）や時間スコアの相関が高ければ、オリジナルの結果は統計的に強固だと言える。以上の議論は主成分解析を例に用いているが、多変量解析一般に適用できる。

このように多変量解析で得られた結果は気象学的に必ずしも意味がある保証はなく、統計的な虚像をもたらす可能性を否定できない。例え、上記の検定で統計的に有意とされた場合でも、気象学的有意性は当然ながら別の議論である。北半球環状モード/北極振動に関する有意性の問題については九大・伊藤教授による最近の解説⁵⁾に詳しい。ここでは、2つの統計的に主要な成分が明確に2つの主成分に分離できないときに有意性の問題が生じやすいことについて、赤道を挟む熱帯大西洋の海面水温変動を事例として紹介し、主成分解析が持つ問題点のその回避の方法を解説する。

4. 熱帯大西洋における海面水温変動場に対する適用

エルニーニョ/ラニーニャ現象で知られるように、東部熱帯太平洋では赤道を挟む冷水域上での海面水温変動が大きいため、海面水温変動の極大は赤道上に観測される。これに対し、熱帯大西洋における海面水温変動の極大は赤道にはなく、むしろ赤道から離れた両半球の緯度10-30度の範囲にそれぞれ観測される。熱

帯大西洋の双極子と呼ばれる海面水温変動は比較的早くから知られており、大西洋兩岸のアマゾンやサハルの降水変動との関連が指摘されていた。

エルニーニョに伴う東部赤道太平洋の海面水温正偏差は赤道を挟んで両半球側に観測される。赤道の南北に位置するこれらの偏差はエルニーニョという同じ現象の一部であることに殆ど異論はない。一方、熱帯大西洋の場合、赤道の極側に観測される両半球の海面水温変動が、熱帯大西洋の双極子に伴う同一の現象の一部とみなすか^{6,7,8)}、それぞれ独自の別の現象であるとみなすか^{9,10,11)}について意見が分かれていた。

第2図aは大西洋上で帯状平均した北半球側と南半球側の海面水温偏差の2次元プロットである。もし、双極子形態が唯一の卓越する現象ならば、最適な直線は $y=-x$ で表される。しかし、第2図aのプロットはほぼ全ての象限に散布し、無相関に近い同心円上の分布となっている。

2次元空間に対する主成分解析は、プロットの分布をよく表現する2つの直交する直線を見つけることを意味する。2次元空間でほぼ同心円状に分布するプロットでは、直交する直線の組み合わせは無数に選ぶことができ、その代表的な組み合わせとして2例上げることができる(第2図b)。

①横軸に対し±45度の角度をもった直線の組み合わせ。つまり、北半球と南半球の変動が同符号(①-M)と逆符号(①-D)の成分がある。赤道を挟んで同符号(逆符号)となる例はエルニーニョ(双極子形態)である。

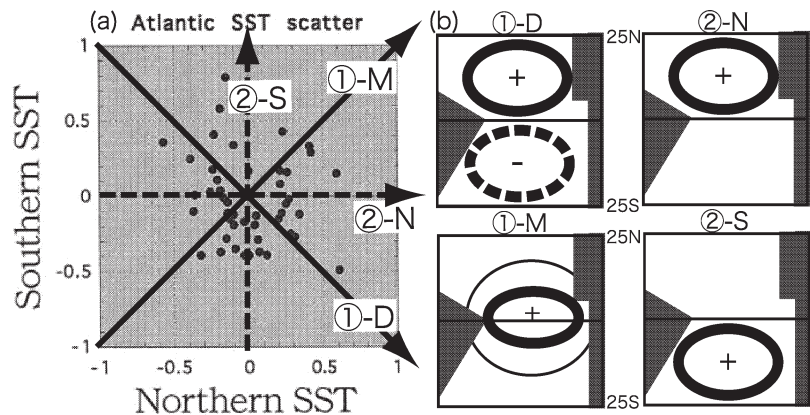
②縦軸と横軸に平行な直線の組み合わせ。つまり、北半球(②-N)と南半球(②-S)がそれぞれ独自に変動する成分である。

実際、熱帯大西洋の多格子点における海面水温変動場に対して主成分解析を施した場合でも、本質的には2次元プロットの分布を表す2つの直交直線を見出すことと同じため、データセットの期間や対象とする範囲、さらには使用する海面水温変動場のデータセットにより、①を支持する結果^{6,7,8)}と②を支持する結果^{9,10,11)}に分かれてしまう。つまり、多変量解析を熱帯大西洋海面水温変動場に単純に適用している限り、どち

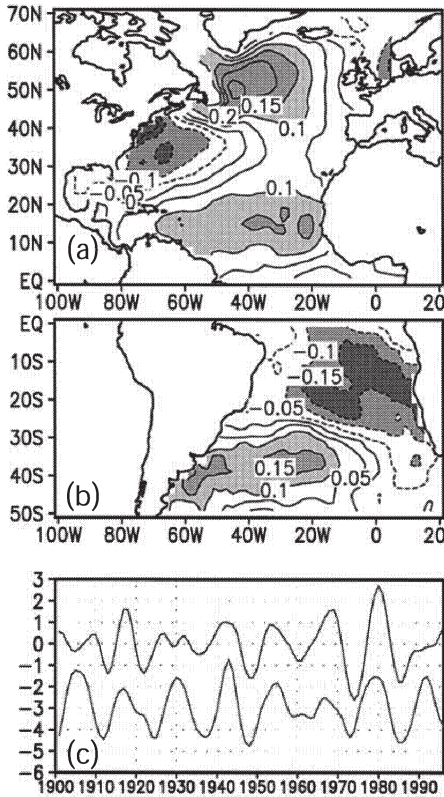
らがより正しいかを明らかにする結果は得られない。

このような問題は、変動場を表すデータセットに何か工夫をすることで回避できる場合がある。熱帯大西洋における海面水温変動の双極子形態を示唆する解析結果では、この変動の時間スケールはおおよそ10年であった。一方、赤道を挟み同符号の海面水温偏差を持ちいわば大西洋版エルニーニョとも言える形態を示唆する解析結果はエルニーニョ/ラニーニャの典型的な時間スケールよりさらに短く2-3年程度の時間スケールを示していた。このように水温偏差の空間形態によって時間スケールが異なることを利用すると、主成分解析を用いても2つの変動をうまく分離することができる。具体的には、各格子点における時系列に対して時間フィルターを施し、時間スケールが分離された低周波と高周波の変動場に対してそれぞれ主成分解析を行う。前者の第1主成分は双極子の形態であり、後者の第1主成分は赤道を挟み同符号であった¹²⁾。

熱帯大西洋におけるこの問題は北半球側と南半球側の双方を同時に主成分解析の対象海域に含めるために生じるので、予め赤道の北側と南側に分けた後、それぞれの領域で主成分解析を施してもよい。2つの領域の空間形態や時間スコアはいわば独立に求められるので、それぞれの結果において何らかの関連が予め定められているわけではない。しかしながら、このように領域を分離した場合の主成分解析の結果では、それぞれの第1主成分の時間スコアの相関は有意であり、かつ10年スケールの変動場における空間形態は赤道の北と南で逆の符号を示した¹³⁾(第3図)。以上のように



第2図 (a) 熱帯大西洋海面水温偏差(°C)の散布図。横軸と縦軸はそれぞれ北半球側(北緯8-20度)、南半球側(南緯の8-20度)経度平均の偏差を表す。矢印は、(b)に示されるそれぞれの空間形態が2次元プロットにおいて表される方向を示す。Tanimoto and Xie, 2002¹⁴⁾, Fig. 1aから引用。(b) 熱帯大西洋海面水温変動の概念図。①-D(①-M)は赤道を挟む逆符号(同符号)の変動形態を表す。②-Nと②-Sは北半球と南半球で独立の変動形態を表す。



第3図 大西洋海面水温偏差場に対する主成分解析の(c)時間スコアと(a, b) それに対する相関係数分布図。(a) 0-70°N領域に対する第1主成分。(b) 0-50°S領域に対する第1主成分。(c) 上側(下側)の曲線は0-70°N領域(0-50°S領域)に対する時間スコア。

時間スケールや領域を分離した変動場を対象にした場合にはどちらも①の説を支持する解析結果となった。第2図aの一見無相関なプロットは、赤道を挟み逆符号となる10年スケール変動(①-D)と同符号となる数年の変動(①-M)の重ね合わせとして解釈できる。

5. おわりに

どの見方が科学として正しいかは、その背後にあるメカニズムを議論することが不可欠である。熱帯大西

洋の海面水温変動の場合、赤道を横切る南北風が生じれば、大気海洋結合系における風-蒸発-海面水温フィードバックを通して、海面水温偏差の双極子形態が維持される。赤道上の大気に南北風を遮る壁があるわけではないので、①の説をより支持していると考えられる。熱帯大西洋における海面水温偏差の形成には、下層雲や大気海洋間の湿度偏差が強く関わっているという指摘もあり、問題の決着は必ずしもついでない。

これらのメカニズムの検証にもまた観測データの解析が必要とされる。現実には、ある特定のメカニズムだけが機能している場合は希で変動場は一般に複雑であり、観測データからの検証の際にも多変量解析がよく用いられる。ただし、上記の熱帯大西洋海面水温場の例だけでなく、単純な多変量解析の適用は変動場の姿を歪めてしまう場合もあり得る。得られた多変量解析の結果に対する統計的な検定とともに、背後にあるメカニズムに対する検討も含めた解析や議論を相補的に組み合わせることにより、より確かな成果としていくことが重要であろう。

参考文献

- Wallace, J. M. and D. S. Gutzler, 1981: *Mon. Wea. Rev.*, **109**, 784-812.
- Mori, A. *et al.*, 2006: *SOLA*, **2**, 5-8.
- Thompson, D. W. J. and J. M. Wallace, 1997: *Geophys. Res. Lett.*, **25**, 1297-1300.
- Mantua, N. J. *et al.*, 1997: *Bull. Amer. Meteor. Soc.*, **78**, 1069-1079.
- 伊藤久徳, 2007: *天気*, **54**, 35-51.
- Servain, J., 1991: *J. Geophys. Res.*, **96**, 15137-15146.
- Nobre, P. and J. Shukla, 1996: *J. Climate*, **9**, 2464-2479.
- Chang, P., L. Ji and H. Li, 1997: *Nature*, **385**, 516-518.
- Houghton, R. W. and Y. M. Tourre, 1992: *J. Climate*, **5**, 765-771.
- Mehta, V. M., 1998: *J. Climate*, **11**, 2351-2375.
- Enfield, D. B. *et al.*, 1999: *J. Geophys. Res.*, **104**, 7841-7848.
- Tanimoto, Y. and S.-P. Xie, 1999: *J. Meteor. Soc. Japan*, **77**, 31-46.
- Tanimoto, Y. and S.-P. Xie, 2002: *J. Meteor. Soc. Japan*, **80**, 1199-1219.