大循環ゼミ　用語解説

EOF(Empirical Orthogonal Function：経験的直交関数)解析

2008年2月5日

自然学類4年　大橋 正宏

EOF解析とは

- 統計学でいう主成分分析

- いくつかの地点における時系列データの主要な変動パターンを抽出して変動の特徴を把握するための統計的手法

  例）AO（北半球冬季の海面更正気圧のEOF1）

- 分散共分散行列の固有値問題

アノマリデータ行列の用意

$$\mathbf{X} = \begin{pmatrix} x_1(1) & x_2(1) & \cdots & x_T(1) \\ x_1(2) & x_2(2) & \cdots & x_T(2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(N) & x_2(N) & \cdots & x_T(N) \end{pmatrix}$$

時刻

地点番号

各成分はアノマリデータ

（時間平均は0）

- 横方向：ある地点での時間変化
- 縦方向：ある時刻での地点変化

分散共分散行列Aを作成

$$\mathbf{A} \equiv \frac{1}{T}\mathbf{X}\mathbf{X}^{\mathrm{T}} = \begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T} x_t(1)x_t(1) & \frac{1}{T}\sum_{t=1}^{T} x_t(1)x_t(2) & \cdots & \frac{1}{T}\sum_{t=1}^{T} x_t(1)x_t(N) \\ \frac{1}{T}\sum_{t=1}^{T} x_t(2)x_t(1) & \frac{1}{T}\sum_{t=1}^{T} x_t(2)x_t(2) & \cdots & \frac{1}{T}\sum_{t=1}^{T} x_t(2)x_t(N) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{T}\sum_{t=1}^{T} x_t(N)x_t(1) & \frac{1}{T}\sum_{t=1}^{T} x_t(N)x_t(2) & \cdots & \frac{1}{T}\sum_{t=1}^{T} x_t(N)x_t(N) \end{pmatrix}$$

- 対角成分:ある地点の分散
- 非対角成分:異なる地点の共分散

固有値問題

$$\lambda_1, \lambda_2, \cdots, \lambda_N$$

固有値

$$\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_N$$

固有ベクトル

•分散共分散行列**A**は対称行列

固有値：全て実数

固有ベクトル：異なる固有値に対する固有ベクトルは互いに直交

正規直交展開

- 正規直交展開する ← $|\mathbf{v}_i| = 1$

$$\mathbf{x}_t = s_t(t)\mathbf{v}_1 + s_2(t)\mathbf{v}_2 + \cdots + s_N(t)\mathbf{v}_N$$

$\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_i$ ← モード

$s_i(1), s_i(2), \cdots, s_i(T)$ ← スコア

$$s_i(t) = \mathbf{v}_i \cdot \mathbf{x}_t$$

正規直交展開した意味

$$\mathbf{V}^{-1}\mathbf{A}\mathbf{V} = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_N \end{pmatrix} \qquad \mathbf{V} = \left(\mathbf{v}_1 \mathbf{v}_2 \cdots \mathbf{v}_N\right)^{\mathrm{T}}$$

$$\mathbf{V}^{-1}\mathbf{A}\mathbf{V} = \mathbf{V}^{\mathrm{T}}\mathbf{A}\mathbf{V} = \frac{1}{T}\mathbf{V}^{\mathrm{T}}\mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{V} = \frac{1}{T}\left(\mathbf{V}^{\mathrm{T}}\mathbf{X}\right)\left(\mathbf{V}^{\mathrm{T}}\mathbf{X}\right)^{\mathrm{T}}$$

$$\frac{1}{T}\left(\mathbf{V}^{\mathrm{T}}\mathbf{X}\right)\left(\mathbf{V}^{\mathrm{T}}\mathbf{X}\right)^{\mathrm{T}} = \begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T} s_1(t)s_1(t) & \frac{1}{T}\sum_{t=1}^{T} s_1(t)s_2(t) & \cdots & \frac{1}{T}\sum_{t=1}^{T} s_1(t)s_N(t) \\ \frac{1}{T}\sum_{t=1}^{T} s_2(t)s_1(t) & \frac{1}{T}\sum_{t=1}^{T} s_2(t)s_2(t) & \cdots & \frac{1}{T}\sum_{t=1}^{T} s_2(t)s_N(t) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{T}\sum_{t=1}^{T} s_N(t)s_1(t) & \frac{1}{T}\sum_{t=1}^{T} s_N(t)s_2(t) & \cdots & \frac{1}{T}\sum_{t=1}^{T} s_N(t)s_N(t) \end{pmatrix}$$

つまり

- あるモード i のスコア $s_i(t)$ の分散は固有値 λ_i に等しい
- 異なるモード同士のスコア時系列には相関がない（共分散が0）

$$\mathbf{x}_t = s_t(t)\mathbf{v}_1 + s_2(t)\mathbf{v}_2 + \cdots + s_N(t)\mathbf{v}_N$$

スコアの分散(固有値)が最も大きな項がアノマリ場に最も影響し, この項に含まれる固有ベクトルが最も影響している空間パターン（EOF第一モード）

寄与率

$$\frac{\lambda_i}{\lambda_1+\lambda_2+\cdots+\lambda_i+\cdots+\lambda_N}\times 100 \quad (\%)$$

- 物理場の分散の何割がモードｉ（EOFi）によって説明されるのかがわかる

具体例

- データ:気象研究所の大気海洋結合モデル（MRI-CGCM2.3）における20C3M実験（20世紀）, SRES-A1Bシナリオ実験（21世紀）結果
- 要素:DJF平均の海面更正気圧
- 対象地域:北半球（北緯20度以北）
- 期間:1901年～2100年

↓

EOF解析を行う

計算機でEOF解析をする際の諸注意

- 1地点が表す領域の違い（高緯度になるほど各グリッドあたりの面積が小さくなる）

  →各データに対して $\sqrt{\cos\phi}$ の重みをかける

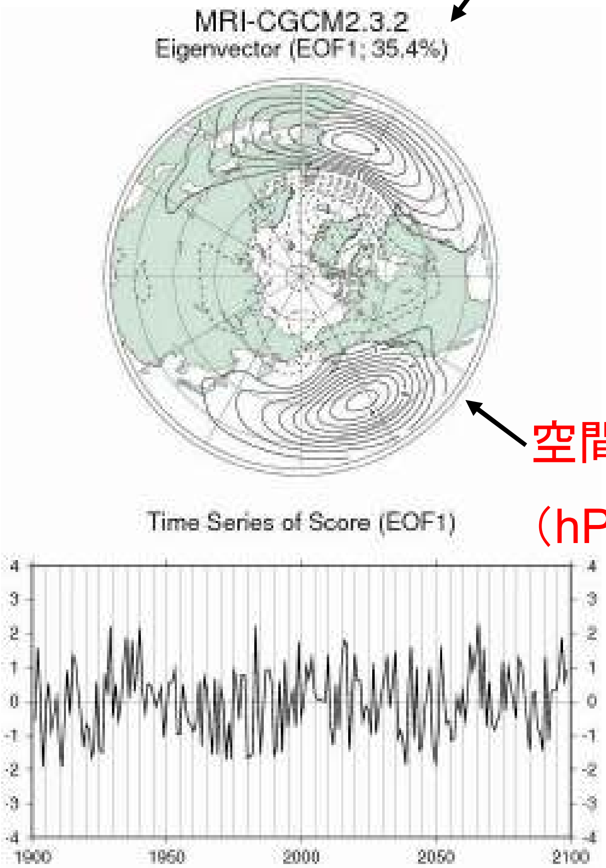
  緯度

- グリッド数が多いと分散共分散行列のサイズが莫大になる（固有値, 固有ベクトルを求められない）

  →地点が数百程度になるように間引く

EOF1

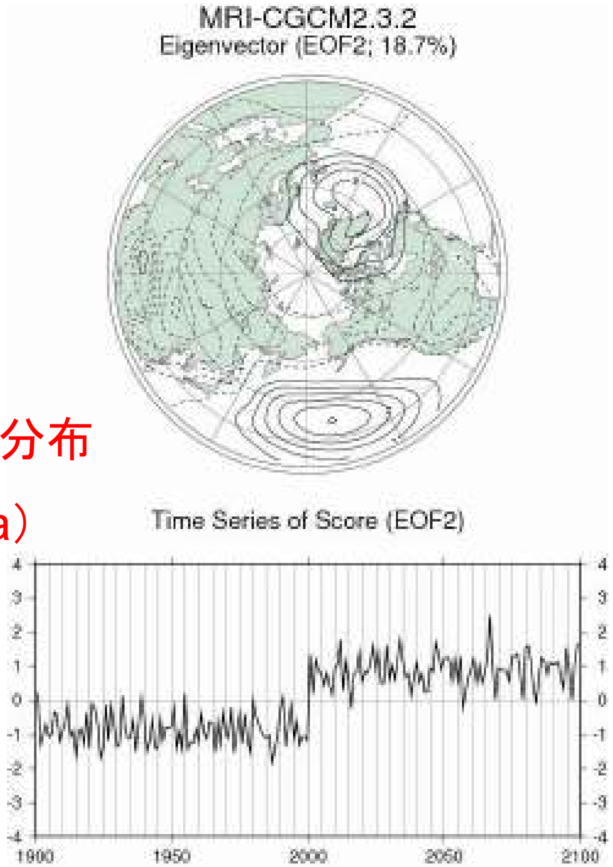
寄与率



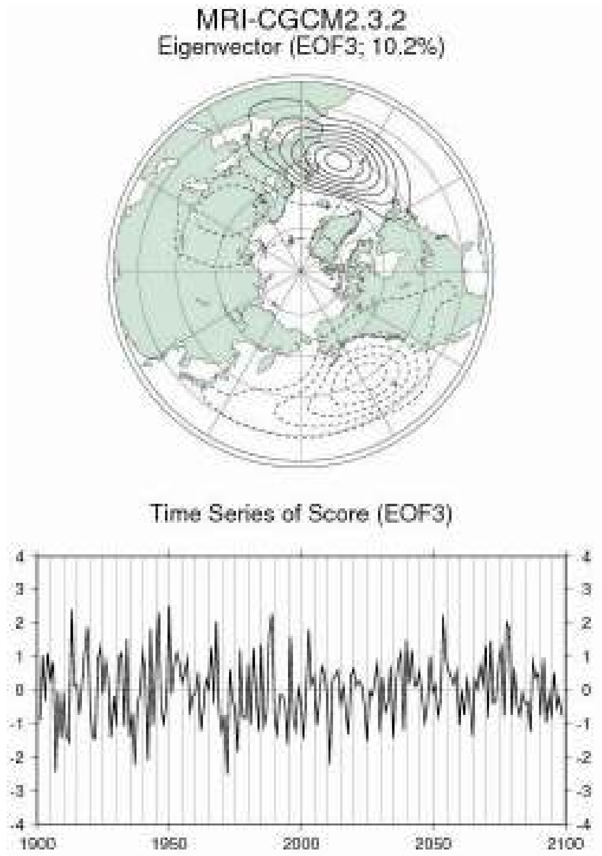
空間分布

（hPa）

EOF2



EOF3



スコア時系列

（無次元）

EOF1：AOパターン

EOF2：PNAパターン？