气象统计预报

【*Chapter 1 绪论*】

1. **气象统计预报**：利用统计学方法对气象（气候）样本进行分析来估计和推测总体的规律性。
2. **气候统计预测的基本假设**：气候系统的未来状态类似于过去和现在。

【*Chapter 2 气象资料的整理*】

1. **平均值（均值）**是描述某一气候变量样本平均水平的量。它代表样本取值中心趋势的统计量。
2. **中位数**是表征气候变量中心趋势的另一个量。在按大小顺序排列的气候变量中，位置居中的那个数就是中位数。其优点是它不易受异常值的干扰。在样本小的情况下，这一点尤为显著。
3. **距平**：气象上常用的变量，也就是通常说的异常，即对平均值的正常情况的偏差。一组数据中的某一个数 x_i 与均值 $\overline{x}$ 之间的差就是距平，即 $x_{di}=x_i-\overline{x}$。统计诊断中用距平来研究变量的变化幅度。
4. **方差与标准差（均方差）**是描述样本中数据与以均值为中心的平均振动幅度的特征量。在气象中也称标准差为均方差(或均方根误差)。

   方差 $s^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2$　　均方差 $s=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2}$
5. **资料的标准化处理**：在气象要素中，各个要素的单位不一样，平均值及方差也不同。为使它们能在同一水平上进行比较，需要标准化。任何气候变量序列经过标准化处理后，都可以化为平均值为 0、方差为 1 的序列。$(x_i-\overline{x})/s$

6. **协方差**：2 个变量距平向量的内积。n 个样本的资料序列 x_k, x_l，其协方差为

   $s_{kl}=\frac{1}{n}\sum_{i=1}^{n}\left(x_{ki}-\overline{x}_k\right)\left(x_{li}-\overline{x}_l\right)$　协方差是反映两个气象要素异常关系的平均状况，协方差为正则变化一致，协方差为负则变化相反。变量自身的协方差是方差。
7. **相关系数（correlation coefficient）**：协方差是带单位的统计量，不便于比较不同要素的异常关系。利用标准化处理方法，对变量先进行标准化，然后计算协方差（不带单位），这种协方差就是相关系数。相关系数是标准化的协方差。相关系数是描述两个随机变量线性相关的统计量，一般称为相关系数或点相关系数。用 r 表示。设有两个变量序列：

   $x_1, x_2, \mathbf{L}, x_n$ 和 $y_1, y_2, \mathbf{L}, y_n$，相关系数计算公式为：

$$r=\frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\overline{y})^2}}=\frac{\text{cov}(x,y)}{s_x s_y}$$

8. **自协方差**与**自相关系数**是衡量气象要素不同时刻之间的关系密切程度的量。时间序列

$x_i\ (i=1,\cdots,n)$，其时间间隔 j 的自协方差为（n 为时间序列样本容量）

$$s(j)=\frac{1}{n-j}\sum_{i=1}^{n-j}(x_i-\overline{x})(x_{i+j}-\overline{x})$$

自相关系数是描述某一变量不同时刻之间相关的统计量。将滞后长度为 j 的自相关系数记为 r（j）。不同滞后长度的自相关系数可以帮助我们了解前 j 时刻的信息与其后时刻变化间的联系。由此判断由 x_i 预测 x_{j+i} 的可能性。对变量 x，滞后长度为 j 的自相关系数为 $r=\frac{1}{n-j}\sum_{i=1}^{n-j}(\frac{x_i-\overline{x}}{s})(\frac{x_{i+j}-\overline{x}}{s})$ s 是样本序列的标准差，r 一般称为滞后（落后）自相关系数。

9. **落后交叉协方差**与**落后交叉相关系数**：考虑两个变量不同时刻之间的相关密切关系。落后交叉协方差：$s_{xy}(j)=\frac{1}{n-j}\sum_{i=1}^{n-j}(x_i-\overline{x})(y_{i+j}-\overline{y})$

   落后交叉相关系数：$r_{xy}(j)=\frac{1}{n-j}\sum_{i=1}^{n-j}(\frac{x_i-\overline{x}}{s_x})(\frac{y_{i+j}-\overline{y}}{s_y})=\frac{s_{xy}(j)}{s_x s_y}$

10. **相关系数的检验**：在总体相关系数 $r=0$ 成立的条件下，相关系数 r 的概率密度函数正好是 t 分布的密度函数，因此可以用 t 检验进行显著性检验。步骤如下：

    (1) 计算统计量 $t=\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$

    (2) 它服从自由度为 n-2 的 t 分布，因此可用 t-检验法检验相关系数的显著性程度

    (3) 由给定的显著性水平 a 及自由度为 n-2，查 t 分布表，得到临界值 t_a

    (4) 比较 t 与 t_a，若 $|t|>t_a$，则认为超过给定的显著性水平 a 的信度检验，拒绝原假设，认为相关系数是显著的

    临界相关系数 $r_c=\frac{t_a}{\sqrt{t_a^2+n-2}}$，表示在给定显著性水平 α 和样本数 n 的条件下，相关系数 r 显著区别于 0 的临界值。当 $|r|>r_c$ 时拒绝原假设。

【*Chapter 3 回归分析*】

1. **一元线性回归**：一元回归处理的是两个变量之间的关系(一个预报变量，一个因子)。基本原理：一般来说，对抽取容量为 n 的预报量 y 与预报因子 x 的一组样本，如认为 y 与 x 是一种统计关系，那么预报量的估计量 $\hat{y}$ 与 x 有如下关系：$\hat{y}=b_0+bx$

用离差平方和来刻画全部观测值与回归直线的偏离程度：$Q(b_0,b)=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$

Q 值越小越好，根据极值原理，要求：$\frac{\partial Q}{\partial b_0}=0 \qquad \frac{\partial Q}{\partial b}=0$

得：$b_0=\overline{y}-b\overline{x} \qquad b=\frac{s_{xy}}{s_x^{\ 2}}$

2. **方差分析**：预报量 y 与回归估计值 $\hat{y}$ 以及误差值 e 之间关系：$y=\hat{y}+e$，可得

$s_y^{\ 2}=s_{\hat{y}}^{\ 2}+s_e^{\ 2}$，其中 $s_y^{\ 2}=\frac{1}{n}\sum_{i=1}^{n}(y_i-\overline{y})^2$ 为预报量的方差，

$s_{\hat{y}}^{\ 2}=\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i-\overline{y})^2$ 为回归估计值的方差，$s_e^{\ 2}=\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$ 为误差方差。

$S_{yy}=U+Q$，多元情况下亦成立，其中 $S_{yy}=\sum_{i=1}^{n}(y_i-\overline{y})^2$ 为总离差平方和，

$U=\sum_{i=1}^{n}(\hat{y}_i-\overline{y})^2$ 为回归平方和，$Q=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$ 为残差平方和。

3. **相关系数与线性回归**：用线性组合来反映预报因子与预报量之间的关系，其优劣程度用

回归方差与预报量的方差比来表征：$\frac{s_{\hat{y}}^{\ 2}}{s_y^{\ 2}}=\frac{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}-\overline{y}\right)^2}{\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\overline{y}\right)^2}=\frac{U}{S_{yy}}=r_{xy}^{\ 2}$

$$b=\frac{s_{xy}}{s_x^{\ 2}} \qquad r=\frac{s_{xy}}{s_x s_y} \qquad \frac{\sum_{i=1}^{n}\left(\hat{y}-\overline{y}\right)^2}{\sum_{i=1}^{n}\left(y_i-\overline{y}\right)^2}=\frac{b^2\sum_{i=1}^{n}(x_i-\overline{x})^2}{\sum_{i=1}^{n}\left(y_i-\overline{y}\right)^2}=(\frac{s_{xy}}{s_x^{\ 2}})^2\frac{s_x^{\ 2}}{s_y^{\ 2}}=(\frac{s_{xy}}{s_x s_y})^2$$

$$\frac{U}{r^2}=\frac{S_{yy}}{1}=\frac{Q}{1-r^2} \qquad \frac{s_{\hat{y}}^{\ 2}}{r^2}=\frac{s_y^{\ 2}}{1}=\frac{s_e^{\ 2}}{1-r^2} \qquad b=\frac{s_y}{s_x}r$$ (b 与 r 同号)

4. **一元线性回归方程的显著性检验**：检验预报因子与预报量是否有线性关系。在总体回归系数为 0 的原假设条件下，统计量 $F=\frac{U/1}{Q/(n-2)}$ 遵从分子自由度为 1，分母自由度为（n－2）的 F 分布。查 F 的分布表，在 $a=0.05$ 下，若 $F>F_a$ 则认为回归方程是显著

的。反之，则不显著。多元线性回归方程的显著性检验类似，其中 $F=\dfrac{U/p}{Q/(n-p-1)}$

5. **复相关系数**：衡量一个预报量与多个变量之间线性关系程度的量。（变量之间的关系归结为多元线性回归方程） 复相关系数是衡量预报量 y 与估计量 $\hat{y}$ 之间线性相关程度的量 $R=\dfrac{\sum_{i=1}^{n}(y_i-\bar{y})(\hat{y}_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}}=\sqrt{\dfrac{U}{S_{yy}}}$ $\quad \dfrac{S_{yy}}{1}=\dfrac{Q}{1-R^2}=\dfrac{U}{R^2}$
6. **逐步回归分析的基本思想**：根据一定的显著性标准，每步只选入一个变量进入回归方程，逐步回归时，由于新变量的引进，可使已进入回归方程的变量变得不显著，从而在下一步给以剔除。因此逐步回归能使最后组成的方程只含有重要的变量。它的自变量都已经经过统计检验，在一定的置信水平下，保证所有的回归系数的总体值均不为零，因此逐步回归分析建立的回归方程也称为最优回归方程。

【*Chapter 4 气候变化趋势分析*】

1. **线性倾向估计**：用 xi 表示样本量为 n 的某一气候变量，用 ti 表示 xi 所对应的时间，建立 xi 与 ti 之间的一元线性回归关系：$\hat{x}_i=a+bt_i\ (i=1,2,\cdots n)$ 其含义是，用一条合理的直线表示 x 与其时间 t 之间的关系。a 是回归常数，b 是回归系数，a 和 b 可用最小二乘法进行估计。对于线性回归计算结果，主要分析回归系数 b 和相关系数 r。
   回归系数 b（倾向值）：b 的符号说明了气候变量 x 的趋势倾向。当 b>0 说明 x 随时间 t 的增加，x 是呈上升趋势；当 b<0，则相反。b 值的大小反映了上升或下降的速率，b 的绝对值越大，表明直线越倾斜。
   相关系数 r：它反映了 x 与 t 间的密切程度。当 r=0，则回归直线平行与 x 轴，说明 x 的变化与时间无关。r>0，说明，x 随时间 t 的增加，x 是呈上升趋势；当 r<0，则相反。R 的绝对值越大，说明 x 与时间的关系越密切。这与 b 所反映的意义是一致的。当然，要判断变化趋势的程度是否显著，就要对 r 进行显著性检验。线性倾向估计的检验（回归方程的检验）。
2. **滑动平均**是趋势拟合技术最基础的方法，它相当于低通滤波器。用确定时间序列的平滑值来显示变化趋势。对样本量为 n 的气候序列 x，其滑动平均序列表示为：
   $$\hat{x}_j=\frac{1}{k}\sum_{i=1}^{k}x_{i+j-1}\text{，}\ (j=1,2,\cdots,n-k+1)$$
   式中 k 为滑动长度，一般取奇数，以使滑动平均对准中间排列。
   经过滑动平均后，序列中短于滑动长度的周期大大削弱，体现出变化趋势来。但经过这种滑动平均后的新序列比原气候时间序列短，原序列两头的信息不能体现。用这种方法求得的是各个时刻的趋势值，而不是具体的数字表达式。
3. **气候变化趋势的显著性检验**：Z 检验

【*Chapter 5 气候突变检测*】

1. **滑动 t-检验**：考察两组样本平均值差异是否显著来检验突变。基本思想：把气候序列中

两段子序列均值有无显著差异看作来自两个总体均值有无显著差异的问题来检验。对于具有 n 个样本量的时间序列 x，设置一个基准点，基准点前后两段子序列 x1,x2 的样本分布为 n1,n2，则 $t=\dfrac{\overline{x}_1-\overline{x}_2}{s\cdot\sqrt{\dfrac{1}{n1}+\dfrac{1}{n2}}}$ 遵从自由度 n1+n2-2 的 t 分布，其中 $s=\sqrt{\dfrac{n_1s_1^2+n_2s_2^2}{n_1+n_2-2}}$

2. **Mann-Kendall 法**：非参数统计检验方法（无分布检验）。(1) 计算顺序时间序列的秩序列 Sk (2) 计算逆序时间序列的秩序列 Sk (3) 给定显著性水平，将计算的 UFk 和 UBk 及显著水平临界线（2 条直线）画到一张图上。若 UFk 和 UBk 出现交点，且交点位于临界线之间，那么交点对应的时刻即为突变开始的时间。

【*Chapter 6 气候序列周期分析（谱分析）*】

1. **功率谱分析**：以傅立叶变换为基础的频域分析方法，将时间序列的总能量分解到不同频率上的分量，根据不同频率的波的方差贡献诊断出序列的主要周期。
2. **离散功率谱估计**：

   (1) 根据公式计算傅里叶系数 a_k, b_k

   (2) 计算相应的功率谱值 $\hat{s}_k=\dfrac{1}{2}(a_k{}^2+b_k{}^2)$

   (3) 确定周期，周期与波数的关系为 $T_k=n/k$

   (4) 作周期与功率谱图
   (5) 显著性检验（F 检验），确定显著周期
3. **连续功率谱估计**：

   (1) 计算最大滞后时间长度为 m 的落后自相关系数 $r(j)=\dfrac{1}{n-j}\sum_{t=1}^{n-j}\left(\dfrac{x_t-\overline{x}}{s}\right)\left(\dfrac{x_{t+j}-\overline{x}}{s}\right)$

   (2) 计算粗谱估计值
   (3) 平滑粗谱估计值

   (4) 确定周期，周期与波数的关系为 $T_k=2m/k$

   (5) 作周期与功率谱图
   (6) 显著性检验，确定显著周期
4. **功率谱检验**：离散功率谱用 F 检验。连续功率谱检验：
   如果序列的滞后相关系数 r(1)为较大正值，表明序列具有持续性，用红噪音标准谱检验；如果 r(1)接近 0 或为负值，表明序列无持续性，用白噪音标准谱检验。红噪音过程是一阶马尔科夫过程或一阶自回归模型，而白噪音过程是一种非周期性过程。
5. **交叉谱**分析是揭露两个时间序列在不同频率上互关系的一种分析方法。
6. **奇异谱**(Singular Spectrum Analysis, SSA)是从时间序列的动力重构出发，并与经验正交函数(Empirical Orthogonal Function, EOF)相联系的一种统计方法。优点：滤波器不需要预先给定，而是根据资料自身最优确定适合确定和寻找噪声系统中的弱信号；不需要作时间序列由不同频率正弦波叠加而成的假定；对嵌套空间维数 m 的限定，可以使得对振荡的转换进行时间定位。SSA 是一种特别适合于识别隐含在气候序列中的弱信号，是一种研究周期振荡现象的新统计技术。先作扩展矩阵，再做 EOF 分解，找周期。

7. **小波分析**：小波分析(Wavelet Analysis)亦称多分辨率分析。分析时间序列在局部时段的频率特征，能表示强度和初位相随时间变化，称时－频局部分析。
优点：一般的功率谱方法对时间域上分辨不清的信号，通过频域分析便可以清晰地描述信号的频率特征，但是几乎不能获取信号在任一时刻的频率特征。而小波分析则可以分析时间序列在局部时段的频率特征。

窗口傅立叶变换：$F(w,b)=\frac{1}{\sqrt{2p}}\int_R f(t)\bar{y}\left(t-b\right)e^{-iwt}dt$

$\bar{y}(t)$ 是固定的，称为窗函数，起到时限的作用。随时间 b 的变化，确定的时间窗口在 t 轴上移动，逐步对 f(t)进行变化。窗口的大小及形状均固定。缺点：对所有频率共用一个相同宽度或相同衰减速度的窗口，窗口的宽度决定了加窗傅立叶变化能够辨认的频率范围。小波变换是为了克服加窗傅立叶变换的缺点发展起来的，它使得窗口的宽度随频率增高而缩小。

若函数 $\bar{y}(t)$ 满足下列条件的任意函数 $\int_R y(t)dt=0$ $\quad\int_R\frac{\left|\hat{y}(w)\right|^2}{w}dw<\infty$

$\hat{y}(w)$ 是 $y(t)$ 的频谱，$y_{a,b}(t)=\left|a\right|^{-\frac{1}{2}}y\left(\frac{t-b}{a}\right)$ 为连续小波，y 叫基小波或母小波。

连续小波是双窗函数，一个是时间窗，一个是频率谱，其振荡随 1/a 增大而增大。a 是频率参数，b 是时间参数。

函数 f(t)的小波变化的连续形式为：$w_f(a,b)=\left|a\right|^{-\frac{1}{2}}\int_R f(t)\bar{y}(\frac{t-b}{a})dt$

小波变化函数是通过对母小波的伸缩和平移得到的。

小波变换的离散形式为：$w_f(a,b)=\left|a\right|^{-\frac{1}{2}}\Delta t\sum_{i=1}^{n}f(i\Delta t)y(\frac{i\Delta t-b}{a})$

小波能量谱 $E(a,b)=\frac{1}{a^2}\left|W_f(a,b)\right|^2$

计算步骤（离散小波变换）：

(1) 根据研究问题的时间尺度确定出频率参数 a 的初值和 a 增长的时间间隔；

(2) 选定并计算母小波函数；

(3) 将确定的频率 a，研究对象序列 f(t)及母小波函数代入离散小波变换公式，计算小波变换 Wf(a,b)

(4) 将计算结果按时间(b)，和频率 a (周期)画图分析时间序列的周期变换

【*Chapter 7 气象变量场时空结构分离*】

1. **主分量（主成分）分析(Principal Component Analysis,PCA)**：将随时间变化的气象场分解为空间函数和时间函数部分。空间函数部分不随时间变化；时间函数部分由空间点(变量)的线性组合所构成，称为主分量 (主成分)。研究主分量随时间变化的规律可以代替对场的随时间变化的研究。
设我们研究的对象是某一气象要素场，场中有 p 个空间点，抽取样本容量为 n 的样本，记 p 个空间点上的要素为 $x_1,x_2,\mathbf{L},x_p$， 由这 p 个变量线性组合成一新变量

$y = v_1x_1 + v_2x_2 + \cdots + v_px_p$，如果 y 满足方差极大的要求，则称 y 为原 p 个变量的主分量。在条件 $v'v = 1$ 下，求 y 方差的极大值，可得 p 个变量的协方差阵 S 的 p 个特征向量 $v_1, v_2, \cdots, v_p$，于是得到 p 个主分量 $y_k = v_k'x$ (k=1,2,⋯,p) $y = V'x$

$$V = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1p} \\ v_{21} & v_{22} & \cdots & v_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ v_{p1} & v_{p2} & \cdots & v_{pp} \end{bmatrix}, \quad \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_p \end{bmatrix} = \begin{bmatrix} v_{11} & v_{21} & \cdots & v_{p1} \\ v_{12} & v_{22} & \cdots & v_{p2} \\ \cdots & \cdots & \cdots & \cdots \\ v_{1p} & v_{2p} & \cdots & v_{pp} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_p \end{bmatrix}$$

2. **主成分的性质：**(1) 各主分量的方差分别为原 p 个变量的协方差矩阵的特征值，不同的主成分量彼此是无关的；p 个主分量的协方差 $V'SV = \Lambda = \begin{bmatrix} l_1 & & \\ & \ddots & \\ & & l_p \end{bmatrix}$

(2) 各主分量的方差贡献大小按矩阵 S 特征值大小顺序排列；第 k 个主分量的方差贡献大小为 $R_k = l_k / \sum_{i=1}^{p} l_i$，由于 $l_1 \ge l_2 \ge \cdots \ge l_p$，得前 m 个主分量的累积方差贡献百分率为 $G(m) = \dfrac{\sum_{i=1}^{m} l_i}{\sum_{i=1}^{p} l_i}$

(3) p 个主分量的总方差与原 p 个变量的总方差相等。$Tr(S) = Tr(\Lambda)$

3. **经验正交分解(Empirical Orthogonal Function ,EOF)：**是针对气象要素场进行的，其基本原理是把包含 p 个空间点(变量)的场随时间变化进行分解。利用经验正交函数展开，就是把 X 分解成正交的空间函数 F 与正交的时间函数 T 的乘积。即 X=FT

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} = \begin{bmatrix} f_{11} & f_{12} & \cdots & f_{1m} \\ f_{21} & f_{22} & \cdots & f_{2m} \\ \vdots & \vdots & & \vdots \\ f_{m1} & f_{m2} & \cdots & f_{mm} \end{bmatrix} \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ t_{21} & t_{22} & \cdots & t_{2n} \\ \vdots & \vdots & & \vdots \\ t_{m1} & t_{m2} & \cdots & t_{mn} \end{bmatrix}$$

根据正交性，F 和 T 应满足下面条件：

当 k=l 时，$\sum_{i=1}^{m} f_{ik} f_{il} = 1$；当 k≠l 时，$\sum_{j=1}^{n} t_{kj} t_{lj} = 0$

X=VT，其中 V 为空间函数，T 为时间函数。

计算步骤：

(1) 对原始资料矩阵 X 作距平或标准化处理。然后计算其协方差矩阵 $S = XX^T$，S 是 $m \times m$ 的实矩阵；

(2) 用求实对称矩阵的特征值和特征向量的方法求出 S 阵的特征值 Λ 和特征向量 V；

(3) 矩阵Λ为对角阵，对角元素即为$S = XX^T$的特征值$l = (l_1, l_2, \mathbf{L}, l_m)$。将特征值按大小排列为$l_1 \ge l_2 \ge \mathbf{L} \ge l_m \ge 0$

(4) 根据特征向量的性质$V'V = VV' = I$，可得$T = V'X$；

(5) 计算每个特征向量的方差贡献和前 p 个特征向量的累积方差贡献。

(6) 凭借气候学知识对前几项有意义的模态（mode）作分析。

每个特征向量所对应的时间向量（称为时间系数）就反映了 X 的空间区域中由此特征向量所表示的空间型的时间变化特征。系数绝对值越大表明对应时刻（某月、某年）这种空间分布型（模态）越明显（典型）。

EOF 优点：(1) 分解不需要固定的函数；(2) 能在有限的区域内对不规则分布的站点进行分解；(3) 由于 EOF 的收敛速度很快，一般特征值较大的前几个模态就能反映出气候场 X 的主要特征了；(4) 分离出的空间模态具有一定的物理意义。

4. **扩展经验正交分解(EEOF)**：EOF 得到气候变量场空间上的分布结构，是固定时间形式的空间分别结构。EEOF 充分利用变量场时间上的自相关及交叉相关，可以得到变量场的移动性分布结构。
5. **旋转经验正交分解(REOF)**：EOF 展开的前几个特征向量，可以最大限度第表征气候变量场整个区域的变率结构，但分离出的空间分布结构不能清晰表示不同地理地区的特征。利用旋转 EOF(REOF)分析，可以克服 EOF 分析中由于区域和时间取样不一致而造成的误差。REOF 的空间分布结构清晰，可以较好地反映不同地域的变化，及不同地域的相关分布状况。如果 EOF 分析截取了前 K 个空间型，累积解释场得总方差已达一定要求(比如达 80％)，可否将这 K 个空间型再作调整，使得调整后得 K 个空间型累积解释原场的总方差的百分率保持不变，而单个空间型尽量反映场的局部相关结果。即使 V 分量方差极大，而主成分中是 T 分量方差极大。
6. **主振荡型分析 (POP)**: 气候系统的主要过程是由 1 阶马尔科夫过程所描述的线性动力过程，其他次要过程被认为是随机噪音过程。将多维时间序列看成 1 阶马尔科夫过程，建立一阶自回归方程，并以其自回归系数阵的特征向量为基础进行正交展开，提取随时间变化的主要模态。POP = EOF+自回归滑动平均

【*Chapter 8 两气候变量场相关模态的分离*】

1. **奇异值分解式定理**：EOF 分析中变量场的协方差矩阵为实对称矩阵，可以采用 Jacobi 方法求解。而涉及两个场分析的协方差矩阵往往是实非对称的，对于这类矩阵需要利用奇异值分解式求解其特征值和特征向量。

   设 2 变量场的交叉协方差阵 Vpxq 为任意非零矩阵，有$V_{p\times q} = L_{p\times q} \wedge_{p\times q} R'_{p\times q}$为 V 的奇异值分解。L 和 Q 分别是 p 和 q 阶正交方阵，Λ是 pxq 阶的对角阵$\wedge = diag(l_1, l_2 ...., l_p)$，称$\Lambda$为 V 的奇异值。

   由于$V'V = R\wedge' L'L\wedge R' = Rdiag(l_1^2, l_2^2, ..., l_p^2)R'$，$l_1^2, l_2^2 ...., l_p^2$是$V'V$的特征值，且$l_1^2 \ge l_2^2 \ge .... \ge l_p^2$，R 是对应的特征向量。

2. **典型相关分析(CCA, Canonical Correlation Analysis)**：将原来较多的变量转化为少数几个典型变量，通过研究典型变量之间的相关系数，分析两组变量间的相关关系。CCA 是

一种具有坚实数学基础，推理严谨，能够有效提取两组变量或变量场相关信号的有用工具。基本思想是：对两组变量分布作线性组合构成新的一对变量 u1,v1，使得它们之间有最大相关系数。在分别作与 u1,v1 正交的线性组合 u2,v2,使它们之间有其次大的相关系数。如此下去，直到认为合适为止,ui,vi, i=1,2⋯.就称为典型变量。

$u_1 = c_{11}x_1 + c_{21}x_2 + \cdots + c_{p1}x_p = c_1'X$ ，$v_1 = d_{11}y_1 + d_{21}y_2 + \cdots + d_{p1}y_p = d_1'Y$，

具有性质$\frac{1}{n}u_1u_1' = 1$，$\frac{1}{n}v_1v_1' = 1$，$r_1 = \frac{1}{n}u_1v_1' = \frac{1}{n}(c_1'X)(d_1'Y)' = c_1'S_{12}d_1 \to$ 最大

的新变量u_1, v_1称为典型因子，称它们之间的相关系数为典型相关系数。

典型因子的特点：(1) 各典型因子是标准化变量；(2) 各对典型因子之间的相关系数反映两组变量线性组合相互关系的密切程度；(3) 不在同一对中各典型因子相互是无关的。

3. **奇异值分解（Singular Value Decomposition)**：设有两个变量场 X,Y，

$X = (x_1, x_2, \cdots, x_p)$ p 个空间点，$Y = (y_1, y_2, \cdots, y_p)$ q 个空间点，样本容量都是 n，X,Y 都是标准化变量，SVD 的方法是基于对上述两个变量的交叉协方差的奇异值分解基础上的：$S = L\begin{bmatrix} \wedge_m & 0 \\ 0 & 0 \end{bmatrix}R'$，分量形式：$S = \sum_{k=1}^{m} l_k l_k r_k' \quad m \le \min(p,q)$ 向量l_k 和 r_k 都有 m 个，都相互正交($l_i, l_j \quad r_i, r_j$)，分别称为<u>左奇异向量</u>和<u>右奇异向量</u>。SVD 的目的是寻找两变量场的线性组合，即由左、右两场分布构造两矩阵$U = L'X$ ，$V = R'Y$

为了唯一分解 S 矩阵，令 L,R 为正交化向量的条件，即 $LL' = I \quad RR' = I$

同时使矩阵 U,V 之间有极大化协方差$\mathrm{cov}(U,V) = L'SR \to \max$ 转化为求条件极值。

根据条件极值的求解，有$\begin{cases} S^T L = l_k R \\ SR = l_k L \end{cases} \quad k = 1, 2, \cdots, m$

若写成矩阵形式，即$S = L\begin{bmatrix} \wedge_m & 0 \\ 0 & 0 \end{bmatrix}R'$

l 为奇异值分解中的奇异值，$l_1 \ge l_2 \ge \cdots \ge l_m > 0$

实对称矩阵$S^T S$ 的特征值和特征向量为 l_k 和 R，SS^T 的特征值和特征向量为 l_k 和 L，

两个矩阵的特征值 l_k 是相同的，即$\begin{cases} (SS^T - l_k I)L = 0 \\ (S^T S - l_k I)R = 0 \end{cases}$，

左变量场展开式 $X = LU$ ，式中U 为左场的时间系数矩阵，记为向量形式：

$$U(t)=\begin{bmatrix} u_1(t) \\ u_2(t) \\ \vdots \\ u_m(t) \end{bmatrix}$$

同理，右变量场展开式$Y=RV$，式中V为右场的时间系数矩阵，记为向量形式：

$$V(t)=\begin{bmatrix} v_1(t) \\ v_2(t) \\ \vdots \\ v_m(t) \end{bmatrix}$$

SVD 相当于将左、右变量场分解为左、右奇异向量的线性组合。每对奇异向量和相对应的时间系数确定了一对 SVD 模态。

每对奇异向量方差贡献为：$SCF_k = l_k^2 / \sum_{i=1}^{m} l_i^2$

前 k 对奇异向量的累积方差贡献：$CSCF_k = \sum_{i=1}^{k} l_i^2 / \sum_{i=1}^{m} l_i^2$

由 SVD 得到的时间系数矩阵 U,V，就可以定义每对奇异向量的时间系数 U 和 V 之间的相关系数：$r_k(U,V)=\dfrac{E\left[u_k(t)v_k(t)\right]}{E\left[u_k(t)^2\right]^{1/2} E\left[v_k(t)^2\right]^{1/2}}$

rk(U,V)表示每对奇异向量间线性组合相关关系的密切程度，与 CCA 中的典型相关系数类似，反映典型变量场总体相关状况。

异性相关系数：$r_k(X,V)$，$r_k(Y,U)$，代表两变量场相互关系的分布结构，显著相关区则是两变量场相互作用的关键区域。

同性相关系数：$r_k(X,U)$，$r_k(Y,V)$，每对奇异向量的时间系数与该场之间的相关分布，就是该对向量的空间分布型，它们在一定程度上代表了两变量场的遥相关型。

SVD 计算步骤：

(1) 由 X,Y 计算交叉协方差矩阵 S；
(2) 对 S 进行奇异值分解，得到奇异值和 L,R；
(3) 根据 X,L 和 Y,R 计算时间系数 U,V；
(4) 计算 U,V 相关系数，同性和异性相关；
(5) 计算方差贡献和累积方差贡献；
(6) 用 Monte Carlo 技术进行显著性检验；
(7) 结果分析

SVD 结果分析：

(1) 从奇异向量的方差贡献及累积方差贡献了解某一对显著 SVD 模态及前几对显著模态所占的方差比例；
(2) 由奇异向量时间系数之间的相关系数 rk 了解两变量场的显著空间分布型总体的相关程度；

(3) 分析异性相关系数场，寻找一个场对另一个场的相互影响关键区；
(4) 利用同性相关系数寻找两个变量的遥相关特征

【图解案例】
见 PPT 讲义