

论自然正交函数的稳定性

章基嘉 孙照渤 陈松军

提 要

对亚洲500毫巴候平均环流进行了天气统计学分析^[1]以后,本文进一步利用更新了的资料进行计算分析。指出,原始资料经过中心化处理后,在所对应的时空尺度上,自然正交函数是稳定的。在这一分析基础上,进一步论证了如果所分析的随机过程是平稳的,则其对应的自然正交函数是稳定的。从而为以自然正交分解方法为基础的预报问题提供了物理基础,并可以大大节省计算工作量。

本文还对气象上使用自然正交分解方法的若干问题进行了讨论,为有效地使用自然正交分解方法提出了一些有意义的看法。

一、问题的提出

长期以来,天气预报中使用的统计学方法基本上是以单个测点上的单个要素的观测资料为基础的,近代天气预报中的统计学方法的一个显著特点是以表征大范围天气形势的参数为依据,这是一个从点向面的过渡,有着重要意义。而为取得这样的大形势参数的方法是多种多样的。1959年以来,讨论自然正交分解方法分析气象要素场的变化,以及由此制作天气预报的研究日渐增多^{[2]—[6]}。在我国,文献^[7]较系统地介绍了自然正交分解方法在气象中的应用。由^{[2]—[7]}可知,给定表示气象要素场的量测数据资料,便可用它构成矩阵F。

$$F = \begin{pmatrix} f_{11} & f_{12} & \cdots & f_{1n} \\ f_{21} & f_{22} & \cdots & f_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ f_{m1} & f_{m2} & \cdots & f_{mn} \end{pmatrix} \quad (1)$$

其中 n 为空间点的个数, m 为随时间的测量次数 ($m \gg n$)。利用自然正交分解可以把 F 分解成只依赖时间坐标和只依赖空间坐标的两部分,即

$$\begin{pmatrix} f_{11} & f_{12} & \cdots & f_{1n} \\ f_{21} & f_{22} & \cdots & f_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ f_{m1} & f_{m2} & \cdots & f_{mn} \end{pmatrix} = \begin{pmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ t_{21} & t_{22} & \cdots & t_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ t_{m1} & t_{m2} & \cdots & t_{mn} \end{pmatrix} \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{pmatrix} \quad (2)$$

令

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{pmatrix}$$

是行向量组成的自然正交函数矩阵, 它由给定的气象场的协方差矩阵

$$A = \frac{1}{m} F^T F \quad (3)$$

的特征向量组成。而

$$T = \begin{pmatrix} t_{11} & t_{12} & \cdots & t_{1n} \\ t_{21} & t_{22} & \cdots & t_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ t_{m1} & t_{m2} & \cdots & t_{mn} \end{pmatrix}$$

是时间系数构成的矩阵, 它描写了气象场随时间的变化。只要求出矩阵 T 和 X 以后, 自然正交分解的工作即告完成。目前的一些以自然正交分解方法为基础的预报工作基本上是预告时间系数部分^[1], 然后转化为预告气象要素场未来的变化。很显然, 这样的预告必须有一个条件, 即假设自然正交函数 X 在预报时效内是不变化的。如果自然正交函数本身在预报时效内也随时间变化, 则预报时间系数的途径就失去基础, 变成毫无意义。文献[1]曾指出: 自然正交函数并不是固定不变的, 而是随展开的对象和资料长短而变化的, 因此所谓自然正交函数不依时间变化也仅仅是对所展开的资料序列而言。对时间系数的预报是在假定自然正交函数在以后时刻“稳定”不变的条件下进行的。但是, 这种“稳定性”在所展开的时空尺度上能维持多长时间仍是一个问题, 这个问题与时间系数所允许预报的时效范围有密切关系。

对于 n 个测点给定的 m 次测值, 由 (2) 式可以按自然正交函数展开, 如果随时间延长, 新资料加入原序列以后其自然正交函数没有显著差异, 就称它为稳定的, 否则就称为不稳定的。

关于自然正交函数的稳定性问题, 无论在理论研究中还是在实际应用中, 都是一个急待解决的问题。但是, 这个问题在国内外关于自然正交分解方法的研究中还没有引起应有的重视。本文打算从分析实际资料出发, 对自然正交函数的稳定性问题及其条件进行一些探讨。

二、亚洲 500 毫巴候平均环流自然正交函数的稳定性分析

文献[1]中, 利用 1964—1973 年共 10 年 720 个候的资料, 详细分析了亚洲 500 毫巴候平均环流的自然正交函数, 分析结果表明, 前五项有明确的天气学意义, 在所分析的时

空尺度上它描写了长波系统的状况。为了讨论自然正交函数的稳定性, 我们更新了其中的五年资料, 采用1969年—1978年的十年资料, 共720个候, 重新进行自然正交展开。资料的取法、处理与计算方法均按文献[1]的方法进行。取亚洲地区29个点, 为了消除季节变化, 原始资料经过中心化处理, 即展开的资料为实际高度减去相应的各候多年平均值所得的距平值。

现在把分解得到的前10项特征向量所能表示原始 500 毫巴高度场的方差的比例列在表 1 中。

表 1 特征向量在总方差中所占的比例

序 号	1	2	3	4	5	6	7	8	9	10
λ_1	377.94	231.29	146.11	120.87	80.04	49.98	42.95	36.37	32.07	21.25
$\lambda_1 / \sum_{i=1}^n \lambda_i$	29.8%	18.2%	11.5%	9.5%	6.3%	3.9%	3.4%	2.9%	2.5%	1.7%
$\sum_{i=1}^k \lambda_i / \sum_{i=1}^n \lambda_i$	29.8%	48.0%	59.5%	69.0%	75.3%	79.2%	82.6%	85.5%	88.0%	89.7%

其中 $\sum_{i=1}^n \lambda_i = \sum_{i=1}^{29} \lambda_i = 1267.60$

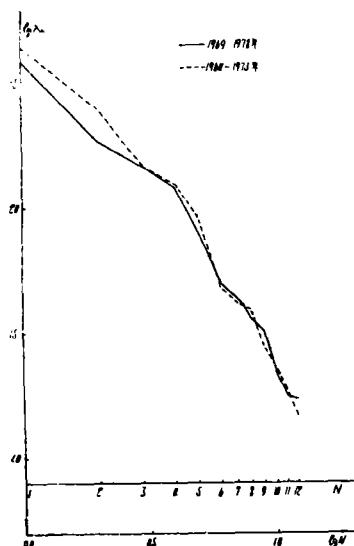


图1 对数座标中不同序号的 λ 比较

为了比较两次分解的情况和所占比例的差异程度, 把使用不同资料展开的两次结果中的特征值 λ_N 和序号N的关系表示在对数座标中(图1)。

由图1可以看出, 两条曲线是十分接近的。这说明它们表示的逼近过程是相近的。取前五个自然正交函数, 就表示出了候平均环流的基本特征。

图2—图6分别给出两次展开的前五个特征向量形势图(图6, 略)。

由图中可以看出, 它们表示的环流形势是十分一致的。为了从数量上给以说明, 我们使用了t检验法, 检验成对数据之间的差异显著性^[9]。

设第一次展开的自然正交函数的某一特征向量的分量为

$$x_1 \quad x_2 \quad \cdots \quad x_n;$$

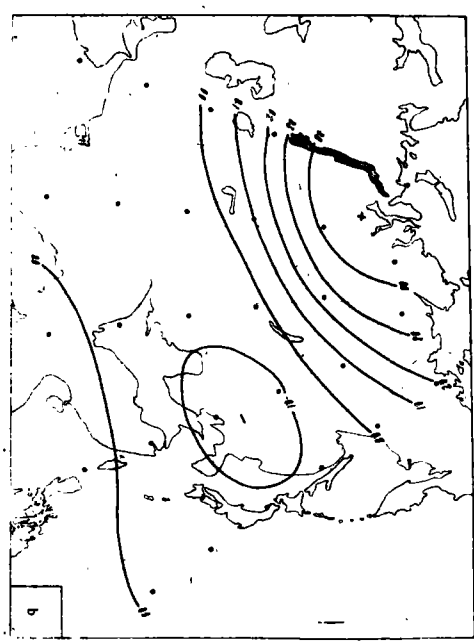
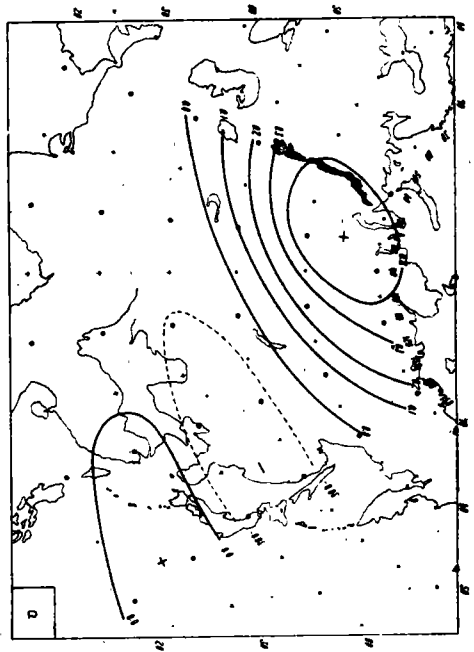


图2 第一特征向量形势图比较
a) 1964—1973年
b) 1969—1978年

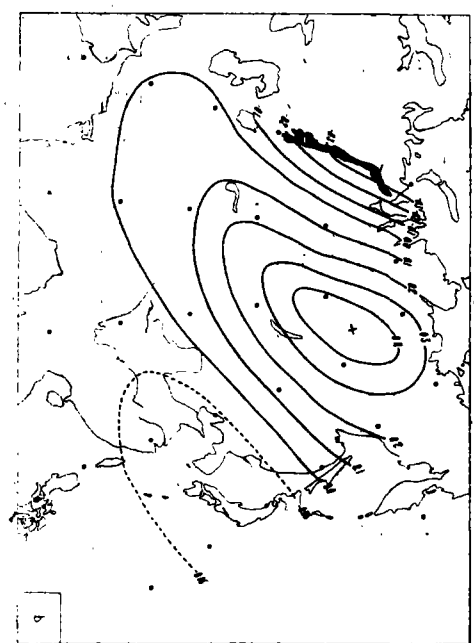
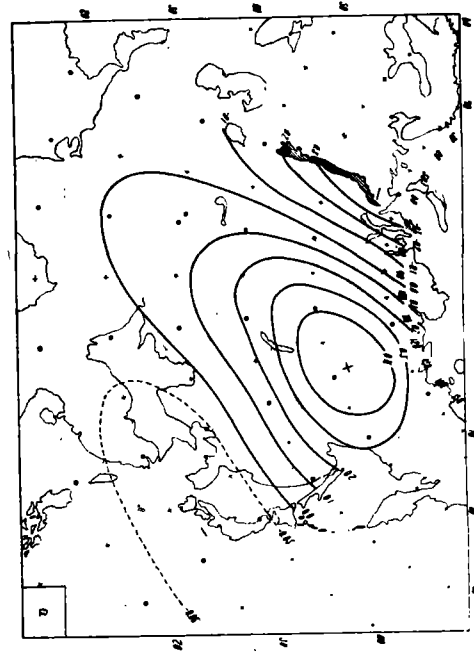


图3 第二特征向量形势图比较
a) 1964—1973年
b) 1969—1978年

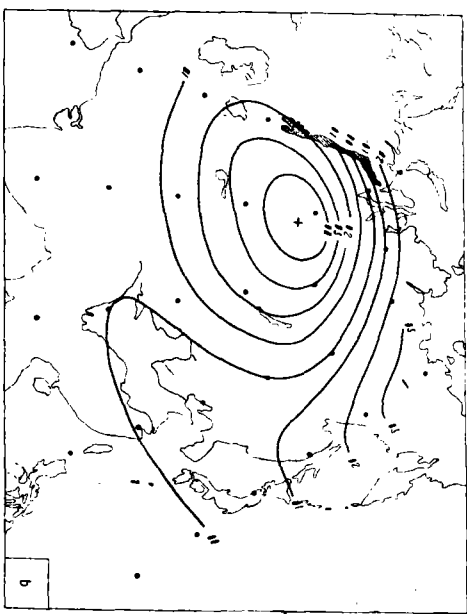
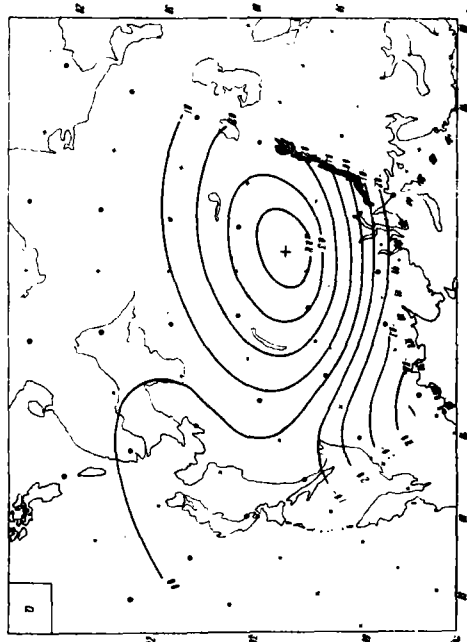


图 4 第三特征向量形势图比较
a) 1964—1973年
b) 1969—1976年

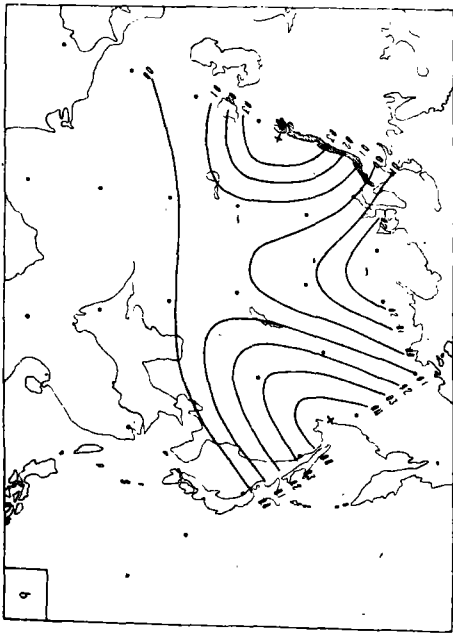
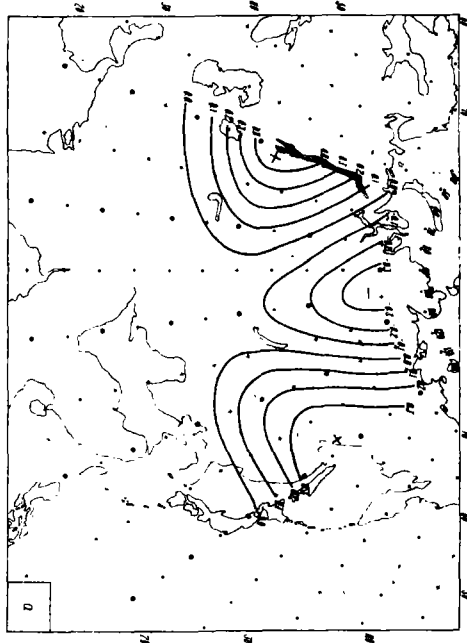


图 5 第四特征向量形势图比较
a) 1964—1973年
b) 1969—1976年

第二次展开的自然正交函数同一序号的特征向量的分量为

$$y_1 \quad y_2 \quad \cdots \quad y_n;$$

令 $Z_i = x_i - y_i$ $i=1, 2, \cdots, n;$

由概率论可知, 可以假定随机变量 Z 服从正态分佈 $N(m, \sigma^2)$ [9]。

为了检验两个自然正交函数的同一序号的向量的分量之间的差异性, 原假设为两向量没有差异。即 $MZ=0$ 。

由量测数据 Z 的数学期望和方差的估计值分别为

$$\bar{Z} = \frac{1}{n} \sum_{i=1}^n Z_i$$

$$s^2 = \frac{1}{n} \sum_{i=1}^n (Z_i - \bar{Z})^2$$

则统计量

$$t = \frac{\bar{Z}}{s} \sqrt{n-1}$$

服从自由度为 $n-1$ 的 t 分佈, 由此出发, 进行统计假设检验。

前五个自然正交函数的检验结果列于表2中。

表2 自然正交函数差异性检验计算表

项 目	函数	第一向量	第二向量	第三向量	第四向量	第五向量
$\bar{Z}$		0.0047	-0.0004	0.0028	-0.0166	-0.0097
s		0.0216	0.0407	0.0386	0.0439	0.0431
t		1.153	-0.052	0.110	-2.000	-1.191

取 $\alpha=0.05$ 时, $t_\alpha=2.048$

由表2可以看出, $\alpha=0.05$ 时, $|t| < t_\alpha$, 即全部自然正交函数的分量之间没有显著差异。

由前面的分析可以看出, 按文献[1]的方法分析的500毫巴候平均环流自然正交函数在一定的信度上可以认为是稳定的。

在我们的计算方案中, 对展开的资料是采用更新的办法, 换句话说, 在两次不同的展开中, 使用的资料仅有原来资料数的1/2是相同的。若计算的方案是采用增加原来资料数的1/2的话, 增加资料后展开的自然正交函数与原来资料的自然正交函数也不会有显著差异, 即这种自然正交函数是稳定的。

三、自然正交函数稳定的条件

在用实际资料分析亚洲500毫巴候平均环流自然正交函数的基础上, 我们进一步研

究自然正交函数稳定性的条件。

由(3)式可以看出,自然正交函数 X 是协方差矩阵

$$A = \frac{1}{m} F^T F$$

的特征向量。其中矩阵 A 的元素

$$a_{ij} = \frac{1}{m} \sum_{t=1}^m (f_{ti} - \bar{f}_i) (f_{tj} - \bar{f}_j) \quad (4)$$

设 λ_p 为 A 的第 p 个特征值,属于 λ_p 的特征向量为

$$X_p = (x_{p1} \ x_{p2} \ \cdots \ x_{pn})$$

$$\text{则} \quad A X_p^T = \lambda_p X_p^T \quad (5)$$

对于 A 的所有特征值的特征向量有

$$A X^T = \Lambda X^T \quad (6)$$

由此可见,对于不同的 A 就有不同的特征值和特征向量。为了由协方差矩阵 A 研究特征向量,(1)式所给定的气象要素场量测数据

$$\begin{pmatrix} f_{11} & f_{12} & \cdots & f_{1n} \\ f_{21} & f_{22} & \cdots & f_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ f_{m1} & f_{m2} & \cdots & f_{mn} \end{pmatrix}$$

可以看成是 n 维随机过程 $(f_1(t), f_2(t), \cdots, f_n(t))$ 的量测数据,其协方差函数为

$$K_{ij}(t_1, t_2) = M [(f_i(t_1) - \bar{f}_i) (f_j(t_2) - \bar{f}_j)] \quad (7)$$

$$i=1, 2, \cdots, n.$$

$$j=1, 2, \cdots, n.$$

若随机过程是平稳的,则其协方差函数仅仅是时间间隔的函数^[10]。

设 $t_2 - t_1 = \tau$ 则

$$K_{ij}(\tau) = M [(f_i(t_1) - \bar{f}_i) (f_j(t_1 + \tau) - \bar{f}_j)] \quad (8)$$

对于 $\tau=0$,即 $t_1 = t_2 = t$,则

$$K_{ij} = M [(f_i(t) - \bar{f}_i) (f_j(t) - \bar{f}_j)]$$

若满足各态历经性质,则 K_{ij} 的估计值为

$$K_{ij} = \frac{1}{m} \sum_{t=1}^m (f_{ti} - \bar{f}_i) (f_{tj} - \bar{f}_j) \quad (9)$$

比较(4)式和(9)式,可以看出

$$a_{ij} = K_{ij}$$

即平稳过程的协方差函数不随时间变化,与时间起点无关。

当量测的气象要素场数据增加1次以后,其协方差函数为

$$K'_{ij} = \frac{1}{m+1} \sum_{t=1}^{m+1} (f_{ti} - \bar{f}_i) (f_{tj} - \bar{f}_j) \quad (10)$$

由平稳过程理论可知

$$K'_{ij} = K_{ij} \quad (11)$$

也就是说,增加1次量测数据以后,协方差矩阵A并不变化,因此,其特征向量也是不变的,所以这种自然正交函数是稳定的。

由此可见,自然正交函数稳定的条件是所分析的气象要素场序列为一平稳过程。

四、几个问题的讨论

1.前面的分析指出,自然正交函数稳定的条件是所分析的气象要素场序列为一平稳过程。但是,实践表明,大气过程往往是非平稳的,正是大气运动过程的非平稳性,给气象预报带来了原则性的困难,无论使用统计学的,天气学的,动力学的方法都会遇到这样的困难。这种非平稳性,当然也使得自然正交函数是不稳定的。但是,我们可以使用一些变换的方法,例如我们在分析亚洲500毫巴候平均环流的自然正交函数中使用的中心化方法,把非平稳的过程转化为渐近平稳的过程。这样,当分析的资料数N足够大的话,在外推时效不太长的情况下,自然正交函数可以保持较好的稳定性。前面分析的500毫巴候平均环流的自然正交函数的稳定性就证实了这一点。

2.前面已经指出,自然正交函数的稳定性研究对于以自然正交分解方法为基础的预报问题提供了物理基础,无疑,这对预报问题是有意义的。另一方面还应指出,自然正交函数的稳定性研究可以大大简化计算。下面作一简单说明。

由(2)式,设增加了一次观测数据,为

$$(f_{m+1\ 1} \ f_{m+1\ 2} \ \cdots \ f_{m+1\ n})$$

若自然正交函数是稳定的,则新增加的系数为

$$(t_{m+1\ 1} \ t_{m+1\ 2} \ \cdots \ t_{m+1\ n})$$

$$t_{m+1\ i} = \sum_{j=1}^n f_{m+1\ j} \cdot x_{ij}$$

$$i=1, 2, \cdots, k.$$

其中k为所选取的比較有意义的自然正交函数的个数。在我们的工作中, $k=5$, 由此可见,计算时间系数只要做 $K \cdot n$ 次乘法和 $k(n-1)$ 次加法,全部时间系数就求出来了。

由此可以看出,在自然正交函数稳定的条件下,增加了资料以后,不需要重新计算协方差矩阵和重新计算特征值和特征向量,而这些计算量在没有电子计算机的情况下是很难设想的。通过稳定性研究以后,在已分解的基础上,若增加了新的资料以后,只要

做不多的乘法和加法, 分解工作就能完成了。这在广大基层台站也能使用这个方法提供了根据。

3. 自然正交分解理论指出, 使用协方差矩阵

$$A = \frac{1}{m} F^T F$$

进行分解计算。由于使用的原始数据不同, 结果当然也不同, 有的使用原始数据分解^[4]

有的使用距平分解^[5], 还有的对协方差矩阵进一步求相关矩阵进行分解。文献〔5〕指出, 用相关矩阵分解是不恰当的。我们在实践中体会到, 使用相关矩阵分解, 反映的是气象要素场相关程度, 因为相关矩阵中的元素是相关系数, 其对应的自然正交函数的物理意义不直观。如果用原始场求出协方差矩阵, 其优点是物理意义直观, 第一项表示了平均状况, 它的权重也大, 但是, 由于气象要素场一般都存在季节变化, 非平稳性十分明显, 使得自然正交函数是不稳定的, 因此, 这种办法是不可取的。我们认为从消去季节变化的中心化数据出发, 求其协方差矩阵进行分解, 既能使得天气学意义比较直观, 又能保持自然正交函数在一定的时效内是稳定的, 给分析工作带来了方便。

4. 尺度对应原理在天气预报中起着重要作用, 在自然正交分解中也是如此。它主要反映在格点的取法上, 一般地说, 点子应取在系统经过的路径上或系统比较稳定的位置上, 点子不应取得过密, 只要能反映出所要分析的系统中心附近的等值线的情况就可以了。如若点子取得过密, 会使得协方差矩阵 A 趋向于退化, 收敛变慢, 以至于取了许多项自然正交函数以后, 所表示总方差的比例仍不能令人满意。这个现象对于变化比较缓慢, 不同点上的要素值之间相关性较密切的要素来说, 如海表温度资料, 就更是如此。

另外, 从理论上来说, 自然正交分解的范围是不受任何限制, 但是, 这仅是对计算工作而言的, 为了能充分表达天气学意义, 所取的范围应该是受相同系统作用的区域, 这大体上相当于一个自然天气区域的范围, 为了消除边缘的影响, 所取区域可比自然天气区域略小一点。

5. 在分析工作中还发现, 采用消去季节变化的数据进行分解, 确能使得自然正交函数成为稳定的, 但是另一方面这又抑制了要素的季节变化。因此, 必要时还可以按季节作分解, 关于这方面的工作将在以后另文讨论。

参 考 文 献

- [1] 章基嘉、孙照渤、兰国明, 亚洲 500 毫巴候平均环流的天气统计学分析, 南京气象学院学报, 1979, 1。
- [2] Н. А. Багров Аналитическое Представление Последовательности Метеорологических Последствий Посредством естественных ортогональных составляющих Труды ЦИП вып. 74, 1959.
- [3] М. И. Юдин, А. Б. Мещерская Некоторые Оценки естественных составляющих

- щих как предикторов и предиктантов. Труды ГГО вып. 273. 1972.
- [4] Н.А.Багров редсказание месячного количества осадков Метеор.и.Гидрол. No7. 1966.
 - [5] J.M.Craddock and C.R.Flood Eigenvector for representing the 500mb geopotential surface over the Northern Hemisphere Quart,J.R.Met.soc. (1969) 95.
 - [6] J.E.Kutzback Large—scale Features of Monthly Mean Northern Hemisphere Anomaly maps of sea—level pressure,Monthly Weather Rev. 98 (1970).
 - [7] 章基嘉, 苏联和其它国家若干长期预报方法简介, 华东区域气象中心编印《长期预报方法选编》, 1973, 11.
 - [8] 王宗皓、李麦村等编著, 天气预报中的概率统计方法, 科学出版社, 1974.
 - [9] M.费史著(王福保译), 概率论及数理统计, 上海科学技术出版社, 1962.
 - [10] “回归分析”与“时间序列分析”, 南京气象学院统计预报短训班讲义(油印本), 1973, 4.