

东海表层多年平均水温的奇异 矢量分析和水系划分*

甘子钧 毛汉礼
(中国科学院海洋研究所)

在海洋学和气象学的研究中,海洋表层水温是最重要而又易于观测的基本参数之一,对其观测数据进行收集、整理和分析研究,向为海洋和气象业务部门所重视。就东海来说,日本气象厅长崎海洋气象台自 1953 年以来即已系统收集该海区的表层水温观测资料,逐旬发布《东海表层水温旬报》,并先后出版了三卷详尽的资料汇编:《东海逐旬表层

表 1 文献 [8, 9, 10, 17] 提出的东海表层水系划分方案

作者及水系划分的主要依据	水系名称	代表站 (°N, °E)
I. 菊池繁雄 (1959) ^[8] ; 1953—1959 年表层平均水温特征: 1. 年平均分布, 2. 月平均分布, 3. 年极值及年较差分布, 4. 周年、半年及 1/4 年谐波的振幅 和初位相分布	黄海冷水	(31°, 125°)
	对马暖流	(33°, 128°)
	黑潮暖流	(26°, 126°)
	流 隔	(32°, 127°)
II. 小泉政美 (1964) ^[9] ; 1953—1962 年表层平均水温特征: 1. 年平均分布, 2. 各月标准偏差分布	黑 潮 型	(26°30', 124°00'; 29°30', 129°30')
	黄 海 冷 水 型	(33°00', 123°00'; 31°00', 125°00')
	黑潮-大陆沿岸水流隔	(29°00', 126°00')
	对马暖流源地	(31°00', 128°00'; 33°30', 128°30')
	黄海暖流	(33°45', 124°15')
III. 佐原勉 (1974, 1975) ^[17, 10] ; 1953—1972 年表层平均水温特征: 1. 年平均分布, 2. 月平均分布, 3. 各月标准偏差分布, 4. 各月水温与前月较差的分布	大陆沿岸水系	(32°, 128°)
	对马暖流水系	(32°, 128°)
	黑潮水系:	
	黑潮流入水	(25°, 124°)
	黑潮陆坡水	(28°, 127°)
	黑潮流出水	(30°, 130°)
	黑潮-大陆沿岸混合体	(29°, 124°)

* 中国科学院海洋研究所调查研究报告第 655 号。本文曾在中国海洋湖沼学会 1979 年学术年会上宣读。

本所张启龙、朱兰部同志参加了资料整理和上机计算工作,杜湄山同志清绘插图,范顺庭、翁学传等同志曾同作者进行了有益的讨论,谨此志谢。

收稿日期: 1980 年 6 月 18 日。

水温五年报(1953—1957)》^[5]，《东海及黄海逐旬表面水温十年报(1953—1962)》^[6]，《东海表层水温二十年报(1953—1972)》^[7]。

现有关于东海表层水系配置以及与之有关的海流系统和海况变动等方面的知识,大多是通过这些资料的分析研究而得到的。其中,特别应该提到的是日本学者菊池繁雄(1959)^[8]、小泉政美(1964)^[9]和佐原勉(1974, 1975)^[10,17]的工作。他们分别依据上列三项资料,对东海表层水温旬报中的一百多个测站的水温季节变化作了常规统计分析,并在

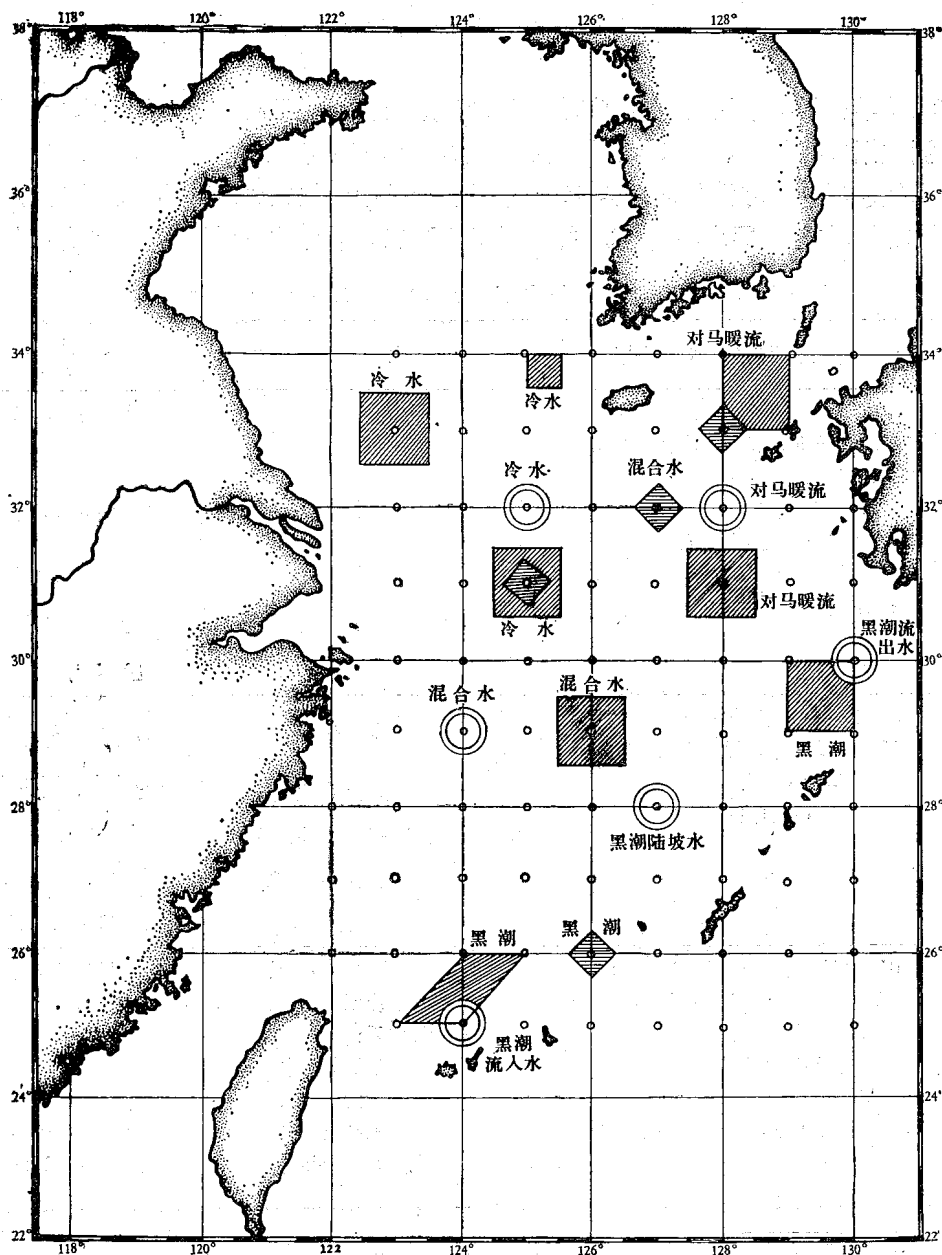


图1 文献[8, 9, 10, 17]提出的各水系代表站

■ 按[8]; ◆ 按[9]; ◎ [按[17,10]

这些分析基础上定性地勾划出东海表层水系配置和海流系统的大致轮廓。这些研究成果迄今仍是讨论东海水文特征的重要依据。

然而,上述三位作者提出的东海表层水系划分方案,尽管大体相同,但在测站类别的划分或合并以及代表站的选择等具体问题上却有颇大的出入,如表 1 及图 1 所示。在我们看来,这三种基于表层水温季节变化统计特征而提出的水系划分方案之所以有如此显著的差别,其原因主要不在于资料的年限不一,而是在于缺乏客观的分类判据。迄今为止,在海洋水文特征的分析研究中,水系(或水团)划分判据的选择是一件相当主观的事情。一般都是根据分析者的经验和直观判断来作出选择。表 1 列举的三位作者所依据的统计特征,除菊池繁雄(1959)^[8]采用的水温年变化调和常数以外,大多并不提供独立的信息,因而未能充分揭示各测站表层水温季节变化之间错综复杂的关系。况且,这三位作者在具体分析各测站的各种统计特征的相似性或差异性时,并未给出客观的判据,而只是先验地从不同的角度对这些特征加以删简或强调,这就给测站的聚类 and 相应代表站的选择带来了不同程度的片面性,以致得到不同的聚类结果。

为了避免上述关于东海表层水系划分方案所蕴含的不确定性,我们试图采用数据矩阵的奇异值分解技术,对东海 82 个 $1^\circ \times 1^\circ$ (经、纬度)网格点的多年(1953—1972) 旬平均表层水温季节距平数据矩阵进行正交分解,从而确定表征该海区表层水温时-空变异的前几对最具特征意义的奇异矢量,以之作为东海表层水系划分的一项客观依据。

一、分析方法

本文分析采用的基本数学工具是矩阵论中的奇异值分解定理(例如,可参见文献 [3, 4, 14, 15] 等)。这一定理可表述如下:

若 A 为 $m \times n$ 阶矩阵,其秩为 r , 且 $m \geq n$, 则必存在一个 $m \times m$ 阶正交矩阵 U 和一个 $n \times n$ 阶正交矩阵 V , 使

$$U'AV = S \quad (1)$$

$$A = USV'; \quad (2)$$

这里,“ $'$ ”表示矩阵转置; S 为 $m \times n$ 阶对角矩阵,它的主对角元素恰有 r 个非零,其值为正,即:

$$S = \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix}, \quad (3)$$

$$\Sigma = \text{diag}[\sigma_1, \sigma_2, \dots, \sigma_r], \quad (4)$$

其中, $\sigma_i = \sqrt{\lambda_i}$, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > 0$ 是 $A'A$ 的非零特征值全体,也是 AA' 的非零特征值全体,因为

$$\begin{aligned} A'A &= VSU'USV' = VS^2V' \\ &= V \begin{bmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{bmatrix} V' = V \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix} V', \end{aligned} \quad (5)$$

$$AA' = USV'VSU' = US^2U'$$

$$= U \begin{bmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{bmatrix} U' = U \begin{bmatrix} A & 0 \\ 0 & 0 \end{bmatrix} U', \quad (6)$$

$$A = \Sigma^2 = \text{diag}[\lambda_1, \lambda_2, \dots, \lambda_r]. \quad (7)$$

式(5)和式(6)表明, U 和 V 各列相应为 AA' 和 $A'A$ 的特征矢量。因此,通常称式(1)和式(2)中的对角矩阵 S 诸元素 $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > \sigma_{r+1} = \dots = \sigma_n = 0$ 为矩阵 A 的奇异值 (singular values); U 和 V 各列相应称为 A 的左奇异矢量和右奇异矢量 (left and right singular vectors); 而恒等式(2)则称为 A 的奇异值分解 (singular value decomposition)。显然,由式(2)容易得知

$$U = AVS^{-1}, \quad (8)$$

$$V = A'US^{-1}. \quad (9)$$

此外,由于 A 的秩 $r \leq n$, 故当 $r = n$ 时, A 称为满秩矩阵;而当 $r < n$ 时则称为亏秩矩阵。

$m \times n$ 阶矩阵的奇异值分解,由于奇异值具有很好的数值稳定性^[18],对于线性代数计算,特别是对于多元观测数据的分析处理来说,是一种值得重视的工具。

如所周知,通常由观测或数值计算结果组成的数据矩阵,即使它本来是亏秩的,但由于其元素不可避免地含有观测或计算舍入的随机误差,结果必然导致它在数值上是满秩的。换言之,它的所有奇异值都大于0。在这种情况下,对于 $m \times n$ 阶数据矩阵 $A_{m \times n}$,一般地有

$$A_{m \times n} = U_{m \times m} S_{m \times n} V'_{n \times n} = \sum_{i=1}^n \sigma_i u_i v'_i, \quad (10)$$

其中,

$$U_{m \times m} = [u_1, u_2, \dots, u_m], \quad (11)$$

$$V_{n \times n} = [v_1, v_2, \dots, v_n], \quad (12)$$

$$S_{m \times n} = \begin{bmatrix} \Sigma_{n \times n} \\ 0_{(m-n) \times n} \end{bmatrix}, \quad (13)$$

$$\Sigma_{n \times n} = \text{diag}[\sigma_1, \sigma_2, \dots, \sigma_n], \quad (14)$$

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n > 0, \quad (15)$$

这里,为明确起见,矩阵的阶数由其下标表出。于是,也就产生一个如何确定数据矩阵 A 的“有效秩”(effective rank)问题。

从原则上说,确定任一矩阵的秩有许多方法。而利用奇异值分解来确定矩阵的秩,其最大好处在于只需对某些较小的奇异值是否可以忽略不计作出判断,并且由此可以得到数据矩阵“有效秩”问题的最小二乘解,而无需按矩阵秩的一般定义对其列(行)矢量或代数余子式进行相应的运算。

事实上,奇异值分解式(10)表明,数据矩阵 $A_{m \times n}$ 可表示为 n 对奇异矢量外积项的线性组合。其中,每一外积项 $u_i v'_i$ 的秩均等于1,其欧氏范数亦等于1,而线性组合的系数(权重)则是相应的奇异值 σ_i 。由于 σ_i 是按递降次序排列的,若取式(10)的前 k 项,

并记为

$$\mathbf{A}_{m \times n}^{(k)} = \mathbf{U}_{m \times k} \mathbf{S}_{k \times k} \mathbf{V}_{k \times n}' = \sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}_i' \quad k < n \quad (16)$$

则可证明

$$\|\mathbf{E}_{m \times n}^{(k)}\|^2 = \|\mathbf{A}_{m \times n} - \mathbf{A}_{m \times n}^{(k)}\|^2 = \sum_{i=k+1}^n \sigma_i^2 = \text{minimum}. \quad (17)$$

这里，

$$\|\mathbf{A}_{m \times n}\|^2 = \text{tr}[\mathbf{A}'\mathbf{A}] = \text{tr}[\mathbf{A}\mathbf{A}'] = \sum_{i=1}^n \sigma_i^2 \quad (18)$$

$$\|\mathbf{A}_{m \times n}^{(k)}\|^2 = \text{tr}[\mathbf{A}^{(k)'}\mathbf{A}^{(k)}] = \text{tr}[\mathbf{A}^{(k)}\mathbf{A}^{(k)'}] = \sum_{i=1}^k \sigma_i^2, \quad (19)$$

而 $\|\cdot\|^2$ 即矩阵 $[\cdot]$ 的欧氏范数， $\text{tr}[\cdot]$ 表示矩阵的迹； $\|\mathbf{E}_{m \times n}^{(k)}\|^2$ 即表示由于保留式 (10) 中前 k 项产生的截尾误差的平方。显然，比值

$$e_k = \|\mathbf{A}_{m \times n}^{(k)}\|^2 / \|\mathbf{A}_{m \times n}\|^2 = \sum_{i=1}^k \sigma_i^2 / \sum_{i=1}^n \sigma_i^2 \quad (20)$$

则给出式 (10) 的前 k 项在 n 项总方差中所占的比重。由于式 (17) 表明以 $\mathbf{A}_{m \times n}^{(k)}$ 逼近 $\mathbf{A}_{m \times n}$ 具有最小二乘性质，且 $k < n$ ，因此，可称 $\mathbf{A}_{m \times n}^{(k)}$ 为 $\mathbf{A}_{m \times n}$ 的“最小二乘降秩逼近”或“渐近奇异值分解”。

在通常情况下，确定截尾展式 (16) 的适当项数 k 不会有多大的困难。对于一个数据矩阵来说，尽管它的奇异值全都大于 0，但我们总可以按式 (17) 或式 (20) 依次取 $k = 1, 2, \dots$ 算出 $\|\mathbf{E}_{m \times n}^{(k)}\|^2$ 或 e_k ，并考察它们随着 k 的增加而递降的规律，从而根据 $\|\mathbf{E}_{m \times n}^{(k)}\|^2$ 或 e_k 是否小于事先给定的某一误差的要求来决定一个合适的 k 值，作为该数据矩阵的“有效秩”。

确定了数据矩阵的“有效秩”可以给我们带来很大的好处。从数据处理的角度来说，原始数据矩阵 $\mathbf{A}_{m \times n}$ 在计算机中要占用 $m \times n$ 个存储单元，而 $\mathbf{A}_{m \times n}^{(k)}$ 仅需占用 $k(m+1)$ 个存储单元；当 m 和 n 都相当大而 $k \ll \min(m, n)$ 时，用 $\mathbf{A}_{m \times n}^{(k)}$ 近似地代替 $\mathbf{A}_{m \times n}$ 可以大大节省机器的存储量和计算时间。另一方面，从多元观测数据的表示来说， $\mathbf{A}_{m \times n}^{(k)}$ 则给出了在最小二乘意义下，将原始数据矩阵的行矢量从原来的 n 维空间向 $k < n$ 维子空间的投影，从而按最小可能畸变的方式简化并精练了原始观测数据。在这一点上，渐近奇异值分解的效果相当于多元统计分析^[1]中的主成分分析 (principal component analysis) 或因子分析 (factor analysis)。事实上，在式 (16) 中如记

$$[\mathbf{L}(Q)]_{m \times k} = \mathbf{U}_{m \times k} \mathbf{S}_{k \times k} = \mathbf{A}_{m \times n} \mathbf{V}_{n \times k}, \quad (21)$$

$$[\mathbf{F}(Q)]_{n \times k} = \mathbf{V}_{n \times k}, \quad (22)$$

则有

$$\mathbf{A}_{m \times n}^{(k)} = [\mathbf{L}(Q)]_{m \times k} [\mathbf{F}(Q)]_{k \times n}' ; \quad (23)$$

或者，类似地记

$$[\mathbf{L}(R)]_{m \times k} = \mathbf{U}_{m \times k}, \quad (21')$$

$$[\mathbf{F}(R)]_{n \times k} = \mathbf{V}_{n \times k} \mathbf{S}_{k \times k} = \mathbf{A}_{n \times m}' \mathbf{U}_{m \times k}, \quad (22')$$

则有

$$\mathbf{A}_{m \times n}^{(k)} = [\mathbf{L}(R)]_{m \times k} [\mathbf{F}(R)]_{k \times n}' \quad (23')$$

显然, 这里式 (23) 和 (23') 也就是多元统计分析中的主成分分析, 或即气象学文献中熟知的经验(自然)正交函数分解^[1,16,19]; 而按因子分析^[12]的术语来说, $\mathbf{L}(Q)$ 和 $\mathbf{L}(R)$ 相当于因子载荷矩阵 (matrix of factor loadings), $\mathbf{F}(Q)$ 和 $\mathbf{F}(R)$ 相当于因子得分矩阵 (matrix of factor scores), k 则相当于因子数。如在数据矩阵中, m 行表示研究对象的 m 个个体, 而 n 列表示对这些个体进行测定的 n 个变量的记录, 则式 (23) 和 (23') 分别相应于“ Q 型”和“ R 型”因子分析的主成分分解。由于奇异值具有很好的数值稳定性, 而且奇异值分解可以“一举两得”地进行主成分分析以及“ Q 型”和“ R 型”因子分析, 因此, 在计算上要比主成分分析和因子分析更灵活而精确。

二、东海表层多年旬平均水温季节变化的奇异值分解

本文的主要目的, 在于运用上述数据矩阵的奇异值分解技术来处理东海表层多年旬平均水温数据, 确定其最具特征意义的时-空变异格局, 从而为东海表层水系的地理区划提供一项客观依据。

1. 资料来源

为了得到比较有代表性的分析结果, 我们选用了图 2 所示东海区 82 个测站 (经纬度网格点) 的多年旬平均水温资料。各测站的具体数据摘自文献 [7], 结果得到一个 82×36 阶数据矩阵

$$\mathbf{T}_{m \times n} = \begin{bmatrix} t_{1,1} & t_{1,2} & \cdots & t_{1,n} \\ t_{2,1} & t_{2,2} & \cdots & t_{2,n} \\ \cdots & \cdots & \cdots & \cdots \\ t_{m,1} & t_{m,2} & \cdots & t_{m,n} \end{bmatrix} = [t_{p,q}] \quad (24)$$

$$p = 1, 2, \cdots, m; q = 1, 2, \cdots, n; m = 82 (\text{站}); n = 36 (\text{旬}).$$

这里, 矩阵元素 $t_{p,q}$ 即第 p 站的第 q 旬多年 (1953—1972) 平均表层水温, 其中, $t_{6,15}$ 和 $t_{6,28}$ 缺值, 我们按行内插值补齐; $\mathbf{T}_{m \times n}$ 的行矢量

$$\mathbf{t}_p = [t_{p,1} \ t_{p,2} \ \cdots \ t_{p,36}] \quad (p = 1, 2, \cdots, 82 \text{ 站})$$

给出第 p 站多年旬平均表层水温的季节变化; 而列矢量

$$\mathbf{t}_q = [t_{1,q} \ t_{2,q} \ \cdots \ t_{82,q}]' \quad (q = 1, 2, \cdots, 36 \text{ 旬})$$

则表示第 q 旬多年平均表层水温的地理分布。

2. 数据分析

为了便于分析计算, 我们首先对原始数据矩阵 $\mathbf{T}_{m \times n}$ 作如下简单变换:

$$\hat{\mathbf{T}}_{m \times n} = \mathbf{T}_{m \times n} - \bar{\mathbf{T}}_{m \times n} \quad (25)$$

$$\hat{\mathbf{Q}}_{m \times n} = \mathbf{D}_{m \times n}^{-1/2} \hat{\mathbf{T}}_{m \times n} \quad (26)$$

这里, $\bar{\mathbf{T}}_{m \times n}$ 表示 $m = 82$ 个测站的各旬平均数据, 它的所有行矢量均等于

$$\bar{\mathbf{t}}_p = \mathbf{1}' \mathbf{T}_{m \times n} / \mathbf{1}' \mathbf{1} = [\bar{t}_{p,1} \ \bar{t}_{p,2} \ \cdots \ \bar{t}_{p,36}] \quad (27)$$

其中, $\mathbf{1}' = [1, 1, \cdots, 1]$ 是一个 $m = 82$ 维的单位行矢量; 而 $\mathbf{D}_{m \times m}$ 则是一个对角矩

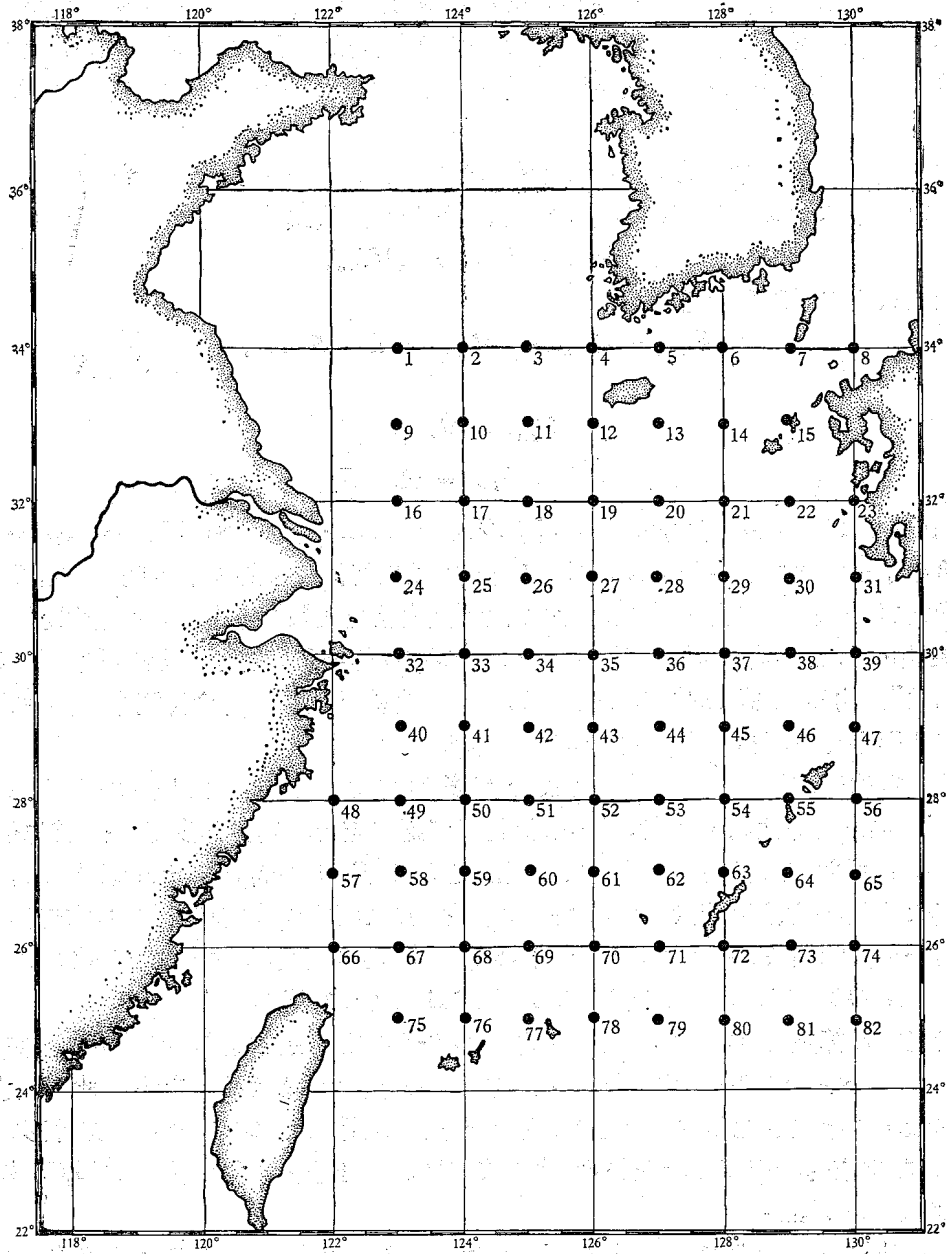


图2 测站位置图

阵

$$D_{m \times m} = \text{diag} [\hat{T}_{m \times n} \hat{T}'_{n \times m}] \quad (28)$$

由此可见,矩阵 $\Theta_{m \times n}$ 是对原始矩阵 $T_{m \times n}$ 按列求距平后再按行进行规格化的结果,即东海表层多年平均水温季节距平的规格化数据矩阵。容易得出,

$$\begin{aligned} C_{m \times m} &= \Theta_{m \times n} \Theta'_{n \times m} = D_{m \times m}^{-1/2} \hat{T}_{m \times n} \hat{T}'_{n \times m} D_{m \times m}^{-1/2} \\ &= [\cos \alpha_{p_1, p_2}], \quad (p_1, p_2 = 1, 2, \dots; m = 82) \end{aligned} \quad (29)$$

给出任意两测站 (p_1, p_2) 多年旬平均表层水温季节距平 i_{p_1} 和 i_{p_2} 的相似系数矩阵, 而其元素 $c_{p_1, p_2} = \cos \alpha_{p_1, p_2}$ 则等于在 $n = 36$ (旬) 维坐标中任意两测站 (p_1, p_2) 表层多年旬平均水温季节距平矢量的夹角余弦。此外, 由于 $\Theta_{m \times n}$ 是按行规格化的, 故 $C_{m \times m}$ 的迹

$$\text{Tr}[C_{m \times m}] = \text{Tr}[\Theta_{m \times n} \Theta'_{n \times m}] = m = 82. \quad (30)$$

在算出矩阵 $\Theta_{m \times n}$ 后, 我们对其进行奇异值分解。按式 (10) 有

$$\Theta_{m \times n} = U_{m \times m} S_{m \times n} V'_{n \times n}, \quad (31)$$

而相应的最小二乘降秩逼近则为

$$\Theta_{m \times n}^{(k)} = U_{m \times k} S_{k \times k} V'_{k \times n} = L_{m \times k} F'_{k \times n}. \quad (32)$$

其中,

$$F_{n \times k} = V_{n \times k}, \quad (33)$$

$$L_{m \times k} = U_{m \times k} S_{k \times k} = \Theta_{m \times n} F_{n \times k}. \quad (34)$$

显然, 在式 (31) 中, 左、右奇异矢量矩阵 $U_{m \times k} = [u_1 u_2 \cdots u_k]$ 和 $V_{n \times k} = [v_1 v_2 \cdots v_k]$, 分别给出了东海表层水温季节距平时-空变异中最具特征意义的 k 个空间变异格局和时间变异格局。因此, 我们可以相应地称之为“空间格局函数”和“时间格局函数”。而式 (34) 给出的 $L_{m \times k}$ 则是 $U_{m \times k}$ 的各列按奇异值加权的结果, 故可称其各列为“加权空间格局函数”。此外, 由式 (34) 还可以看出, $L_{m \times k}$ 也就是我们所讨论海区中各测站的表层水温季节距平与时间格局函数的相似系数矩阵。

关于数据矩阵的奇异值分解, 已设计有多种算法。其中, 尤以 Golub 和 Reinsch (1970)^[13] 提出的算法精度较高, 近年来得到了广泛的应用。其计算步骤是: 首先借助于 Householder 反射变换直接将矩阵 $\Theta_{m \times n}$ 化为上双对角型 (upper bidiagonal form), 然后采用 QR 型迭代法算出此双对角矩阵的奇异值, 并从而回代算出相应的左、右奇异矢量。有关这一算法的细节, 在文献 [15] 的第 18 章中有详细的讨论, 并附有 FORTRAN 语言程序, 而 ALGOL 语言程序则见于 Golub 和 Reinsch (1970)^[13]。

我们的计算是在 DJS-6 机上进行的。DJS-6 机的算法语言与 ALGOL 60 基本相同^[2]。上述 Golub 和 Reinsch (1970) 给出的程序只要稍加修改即可应用。此外, 考虑到一般的最小二乘法逼近和主成分分析(或因因子分析)总是引出 $\Theta'_{n \times m} \Theta_{m \times n}$ 这种形式的矩阵, 而且要求它的特征值和特征矢量, 为比较起见, 我们还采用了如下步骤计算了式 (32) 中的 $L_{m \times k}$ 和 $F_{n \times k}$:

(1) 计算 $n \times n$ 阶实对称矩阵 $\Theta'_{n \times m} \Theta_{m \times n}$, 按常用的 Jacobi 算法计算其全部特征值 λ_i 和相应的特征矢量 $f_i (i = 1, 2, \cdots, n)$ 。

(2) 按公式

$$e_k = 1 - \sum_{i=1}^k \lambda_i / \sum_{i=1}^n \lambda_i \quad k = 1, 2, \cdots, \leq n$$

算出规格化逼近误差平方 e_k ; 然后根据 e_k 是否小于某一给定的误差要求决定一个合适的 k 值。

(3) 取前 k 个特征矢量 $f_i (i = 1, 2, \cdots, k \leq n)$ 构成矩阵 $F_{n \times k} = [f_1, f_2 \cdots f_k]$,

并按下式算出 $L_{m \times k}$ ：

$$L_{m \times k} = \Theta_{m \times n} F_{n \times k}.$$

我们对矩阵 $\Theta_{m \times n}$ 和 $\Theta'_{n \times m} \Theta_{m \times n}$ 相应算得的奇异值 σ_i ，特征值 λ_i 以及规格化逼近误差平方 e_i ，如表 2 所示。为节省篇幅起见，表中仅列出 $i = 1, 2, \dots, 10$ 的计算结果，而表列值则取至小数点后第五位为止。

表 2 $\sigma_i, \sigma_i^2 / \sum_{i=1}^{36} \sigma_i^2, \sum_{i=1}^k \sigma_i^2 / \sum_{i=1}^{36} \sigma_i^2$ 和 e_k 的数值表

i, k	$\sigma_i^2 = \lambda_i$	$(\sigma_i^2 / \sum_{i=1}^{36} \sigma_i^2) \%$	$(\sum_{i=1}^k \sigma_i^2 / \sum_{i=1}^{36} \sigma_i^2) \%$	$e_k \%$
1	71.47822	87.16856	87.16856	12.83144
2	5.38416	6.56605	93.73461	0.26539
3	3.30279	4.02780	97.76241	2.23759
4	0.83890	1.02305	98.78546	1.21454
5	0.22708	0.27692	99.06238	0.93762
6	0.16259	0.19827	99.26065	0.73935
7	0.10994	0.13407	99.39472	0.60528
8	0.08774	0.10696	99.50168	0.49832
9	0.06744	0.08225	99.58393	0.41607
10	0.05335	0.06506	99.64899	0.35101

注： $\sum_{i=1}^{36} \sigma_i^2 = \sum_{i=1}^{36} \lambda_i = \text{tr}[\Theta' \Theta] = \text{tr}[\Theta \Theta'] = 82$

由表 2 可以看出， $\Theta_{m \times n}$ 的奇异值分解展式 (31) 的收敛速度很高，如取其前两项 ($k = 2$)，相对误差平方约 6.27%；而前三项 ($k = 3$) 的相对误差平方仅 2.24%。换言之，对于原来 $n = 36$ 维的东海表层多年平均水温季节距平规格化数据矩阵，只要采用 $k = 3$ 的“最小二乘降秩逼近”，即可保留原来信息的 97.76% 以上。这样，我们便可在不高于 3 维的欧氏空间中，绘出相应于东海表层多年平均水温季节距平时-空变异的 82 个点子(或矢量)，而不会丢掉它们在原来 36 维欧氏空间中分布格局的基本特征。

三、结果的讨论

1. 平均状况

在讨论上述关于 $\Theta_{m \times n}$ 的奇异值分解结果之前，我们先看一下对于原始数据矩阵 $T_{m \times n}$ 及其列(季节)距平矩阵 $\hat{T}_{m \times n}$ 的一些常规统计特征，即

(1) 多年平均表层水温 $t_{p,q}$ 的列(季节)平均 $M(t_q)$ 及其标准偏差 $S(t_q)$ ：

$$M(t_q) = \frac{1}{m} \sum_{p=1}^m t_{pq}, \quad (35)$$

$$S(t_q) = \sqrt{\frac{1}{m-1} \sum_{p=1}^m [t_{pq} - M(t_q)]^2}, \quad (36)$$

(2) 多年旬平均表层水温季节距平 $\bar{i}_{pq} = (t_{pq} - \bar{t}_q)$ 的行(站)平均 $M(\bar{i}_p)$ 及其标准偏差 $S(\bar{i}_p)$:

$$M(\bar{i}_p) = \frac{1}{n} \sum_{q=1}^n \bar{i}_{pq}, \quad (37)$$

$$S(\bar{i}_p) = \sqrt{\frac{1}{n-1} \sum_{q=1}^n [\bar{i}_{pq} - M(\bar{i}_p)]^2} \quad (38)$$

在上列各式中, $p = 1, 2, \dots, m$; $q = 1, 2, \dots, n$; $m = 82$ (站); $n = 36$ (旬)。

图 3 给出了东海 82 个测站表层多年旬平均水温 $M(t_q)$ 及其标准偏差 $S(t_q)$ 的季节

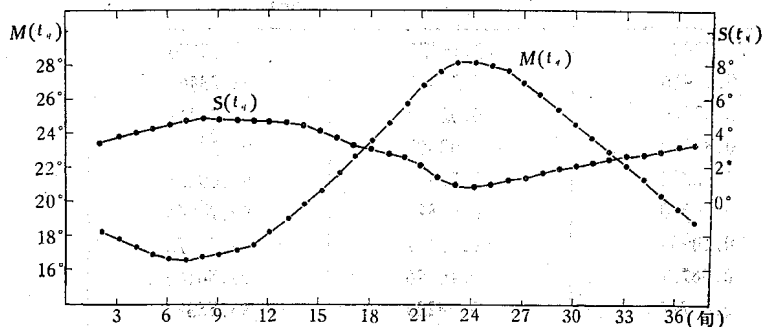


图 3 东海表层 82 个测站多年旬平均水温的均值 $M(t_q)$ 和标准偏差 $S(t_q)$ 的季节变化

变化。由此可以看出,它们具有明显的季节性,并且两者的变化趋势是反相的: $M(t_q)$ 的最高值出现于 8 月中旬 ($q = 23$), 最低值出现于 2 月下旬 ($q = 6$); 而 $S(t_q)$ 则与之相反,最高值出现于 3 月中旬 ($q = 8$), 最低值出现于 8 月中旬。

图 4 和 5 分别给出了东海 82 个测站表层多年旬平均水温季节距平的行平均 $M(\bar{i}_p)$ 和标准偏差的地理分布。应该指出,这里 $M(\bar{i}_p)$ 的分布特征与文献 [10, 17] 给出的东海表

层多年旬平均水温年平均 $\bar{i}_p = \frac{1}{n} \sum_{q=1}^n t_{pq}$ 的分布特征是相似的。但由于 $M(\bar{i}_p)$ 是各

测站季节距平的年平均值,因而能更清晰地显示出以 $M(\bar{i}_p) = 0$ 为界的两个正、负距平区,即通常所谓“外海(黑潮)暖水系”和“陆缘(黄海)冷水系”;而与之相应的暖脊和冷槽的走向则反映出这两大水系的配置格局。值得注意的是,图 3 给出的 $S(\bar{i}_p)$ 分布还进一步刻划了这两大水系的主体所在及其过渡区的界限。例如,若取 $S(\bar{i}_p) = 1^\circ\text{C}$ 等值线所限区域作为冷、暖水系的过渡区,则在这一区域中还可进一步划分出由 $S(\bar{i}_p) = 0.5^\circ\text{C}$ 等值线包围的两个独立子区,即“对马暖流和黄海暖流源地”以及由 $(28^\circ\text{N}, 125^\circ\text{E})$ 和 $(29^\circ\text{N}, 126^\circ\text{E})$ 所代表的“流隔”区。

2. 季节距平的时-空变异特征

图 6—8 分别给出了对东海表层 82 个测站多年旬平均水温季节距平规格化数据矩阵 $\Theta_{m \times n}$ 进行奇异值分解得到的前三对奇异矢量 ($\mathbf{l}_i, \mathbf{f}_i$) 的图象。其中, $\mathbf{l}_i$ 表示规格化左奇异矢量 $\mathbf{u}_i$ 与奇异值 σ_i 的乘积,即按奇异值加权的空间格局函数

$$\mathbf{l}_i = \sigma_i \mathbf{u}_i = [l_1, l_2, \dots, l_m, i]' \quad m = 82, \quad i = 1, 2, 3;$$

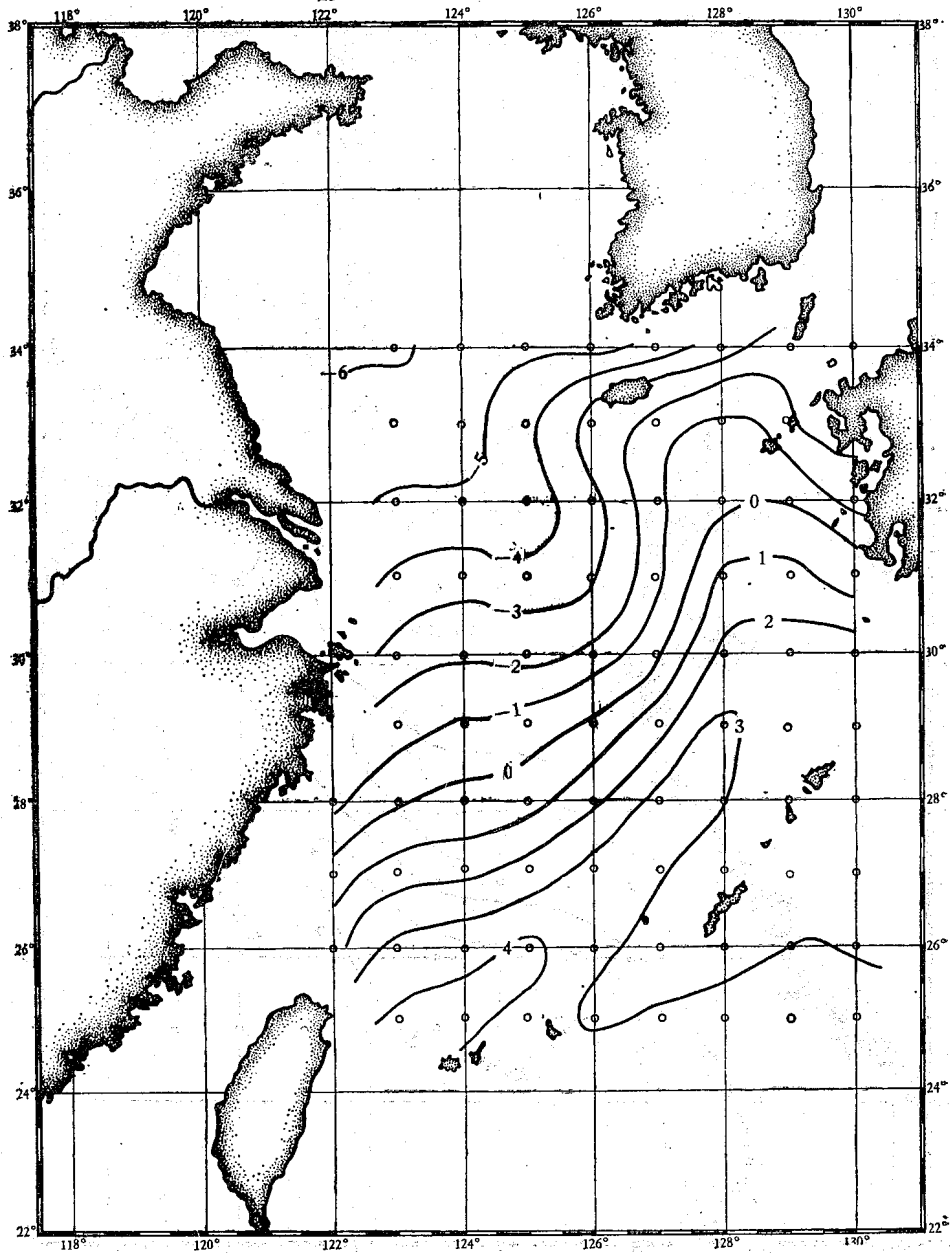


图4 东海表层82个测站多年旬平均水温季节距年平均 $M(i_p)$ 的地理分布

而 f_i 为规格化右奇异矢量 v_i ，即时间格局函数

$$f_i = v_i = [f_{1,i} f_{2,i} \cdots f_{n,i}]' \quad n = 36, i = 1, 2, 3.$$

前面已经指出， L_i 表征了第 i 个时间格局函数 f_i 对各测站表层规格化水温季节距平的贡献，或即第 p 测站的规格化水温季节距平

$$\theta_p = [\theta_{p,1} \theta_{p,2} \cdots \theta_{p,n}], \quad (p = 1, 2, \cdots, 82, n = 36)$$

与 f_i 的相似系数。因此，在 L_i 的图象中， $l_{p,i}$ 取正值的区域表明该区表层水温季节距平的

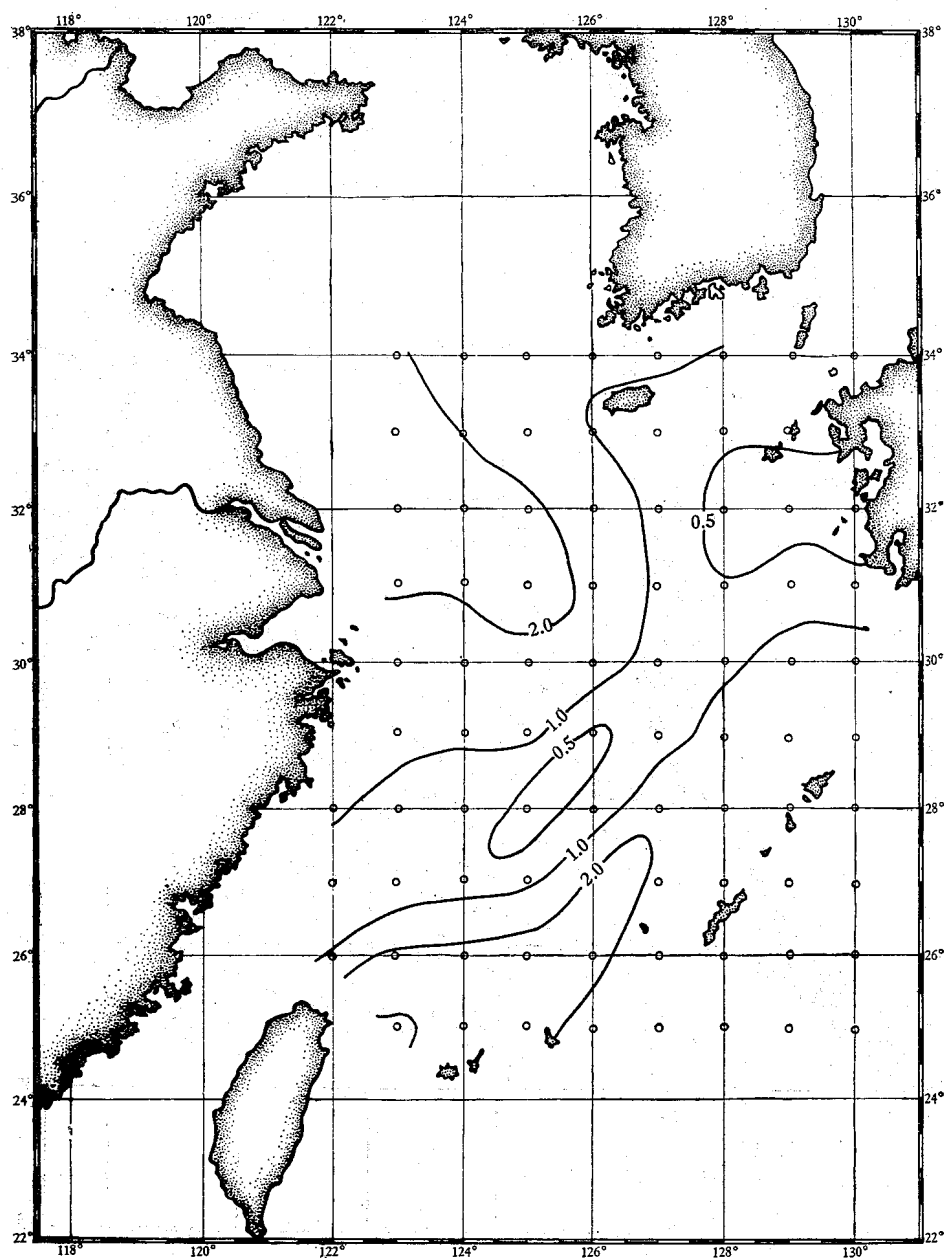


图5 东海表层82个测站多年旬平均水温季节距平标准偏差 $S(l_p)$ 的地理分布

变化过程和 f_i 同相, 而 $l_{p,i}$ 取负值的区域则反相; 并且 $l_{p,i}$ 的绝对值越接近于1, 则该测站的表层水温季节距平的变化过程与 f_i 越相似。

(1) 第1时-空变异格局函数(图6)对东海表层多年旬平均规格化水温季节距平总方差的贡献高达87.17%(参见表2)。这里, 值得注意的是, 第1时间格局函数 f_i 具有明显的季节变化的特征(图6b), 并且它的变化趋势和整个东海82个测站多年旬平均表层

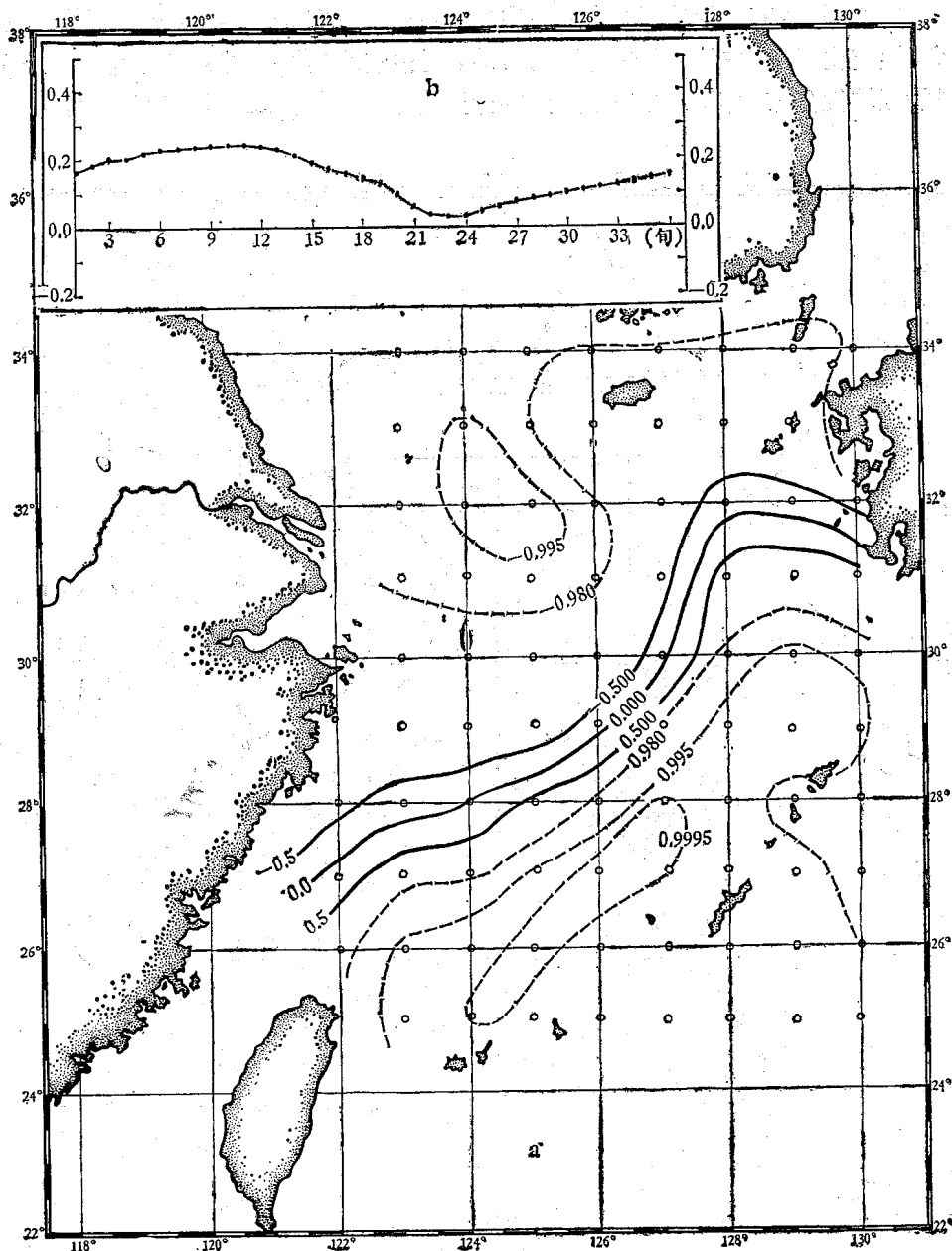


图 6 a. 第 1 加权空间格局函数 l_1 ;
b. 第 1 时间格局函数 f_1

水温季节变化的标准偏差 $S(t_q)$ 几乎完全一致(参见图 3)。同样,第 1 空间格局函数 l_1 也和图 4 所示东海表层多年平均水温季节距平的年平均分布颇相一致,并且具有明显的区域性分布规律。由图 6a 可以看出, $l_{p,1} = 0$ 等值线横亘东海中部,呈西南—东北走向,将整个东海表层划分为两个水温季节距平变异相位相反的区域。显然,这里 $l_{p,1} > 0$ 的区域相当于外海(黑潮)暖水,其中心位置如 $l_{p,1} = 0.9995$ 这一封闭等值线所示;反之, $l_{p,1} < 0$ 的区域相当于陆缘(黄海)冷水,其中心位置大体在 (32°N, 125°E) 附近。

(2) 第 2 时-空变异格局函数(图 7)的方差贡献约 6.57%。其中, 第 2 时间格局函数 f_2 的季节性变化规律也是显而易见的(图 7b), 而相应的空间格局函数 l_2 也同 l_1 一样, 具有明显的区域性分布规律。图 7a 清晰地表明, $l_{p,2} = 0$ 等值线横亘东海中部, 呈西北—

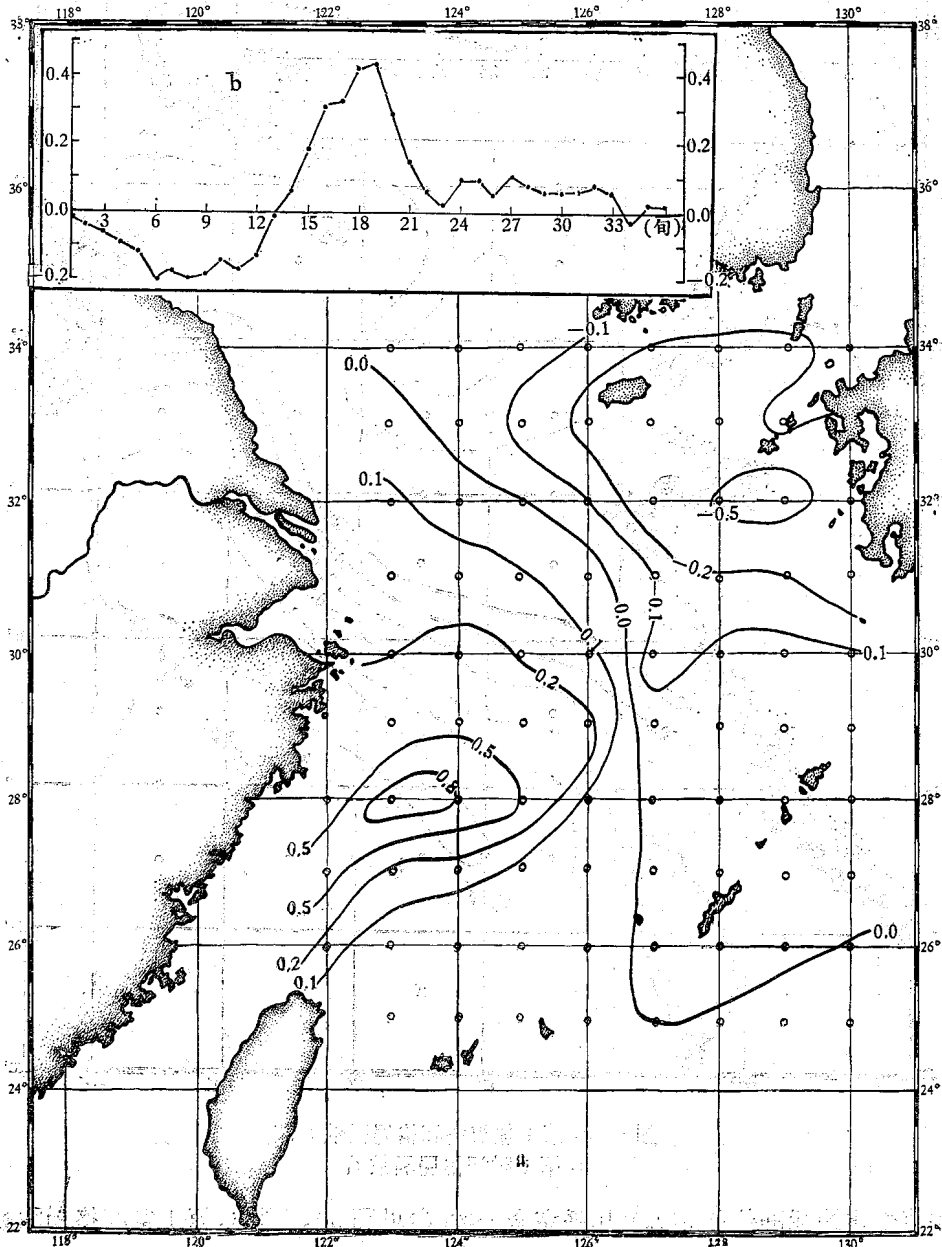


图 7 a. 第 2 加权空间格局函数 f_2 ;
b. 第 2 时间格局函数 l_2

东南走向, 将东海表层由 f_2 定义的水温季节距平变异划分为两个相位相反的区域, 即位于中国大陆一侧的西南陆缘区 ($l_{p,2} > 0$) 和日本九州一侧的东北陆缘区 ($l_{p,2} < 0$)。在这

两个区域中,各有一个封闭性的极值中心,它们分别对应于我国浙、闽外海的台湾暖水区和日本九州以西五岛列岛南部的对马暖流源地。

3. 第3时-空变异格局函数(图8a, 8b)的方差贡献较小(仅约4.03%),但仍具有较明显的规律性。特别是 $l_{p,3}$ 的分布,在东海中部呈现出一个取西南—东北方向延伸的椭圆

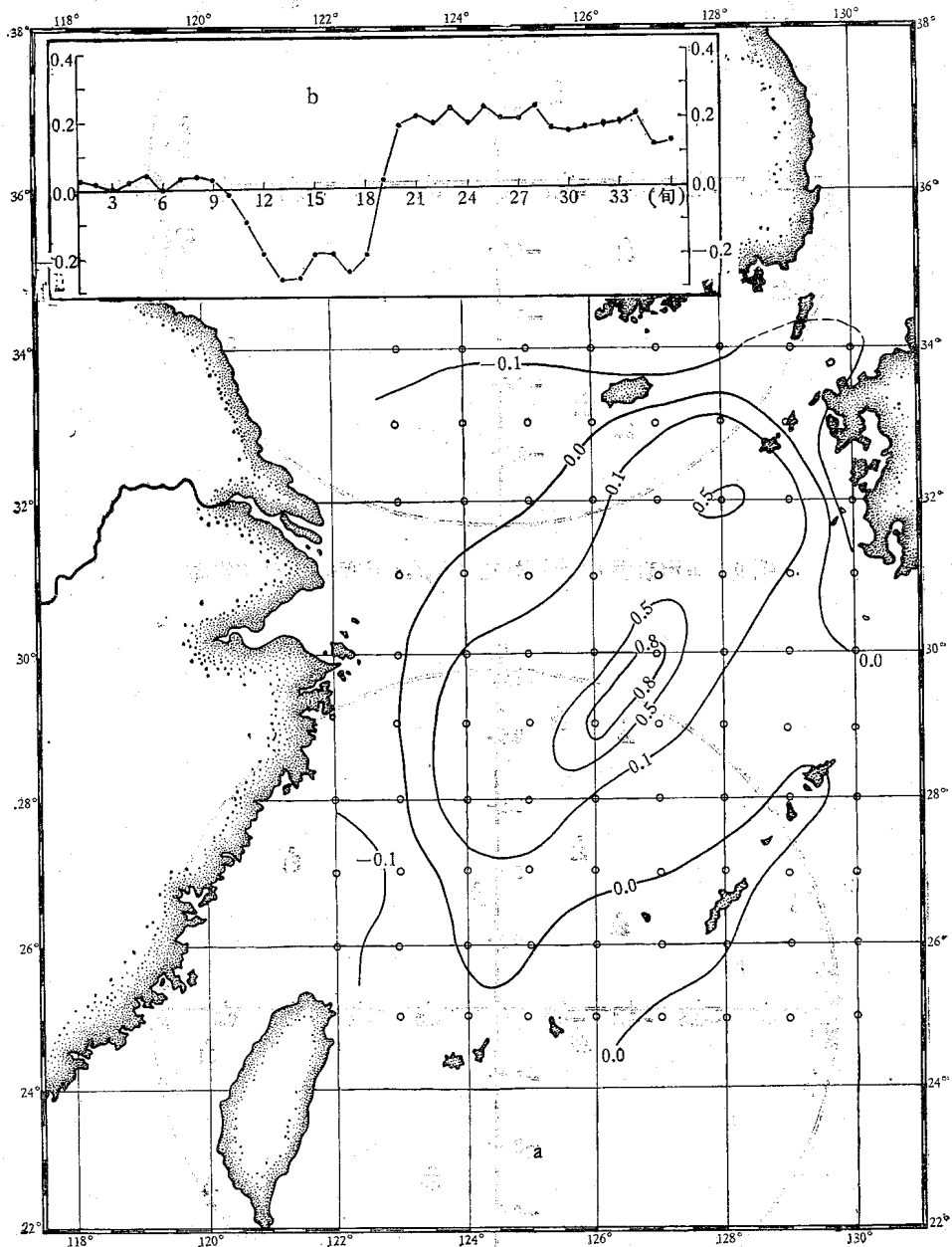


图8 a. 第3加权空间格局函数 l_3 ;
b. 第3时间格局函数 f_3

圆形高值中心,其极值位置在(29°N, 126°E)和(30°N, 127°E)两站之间。此外,在

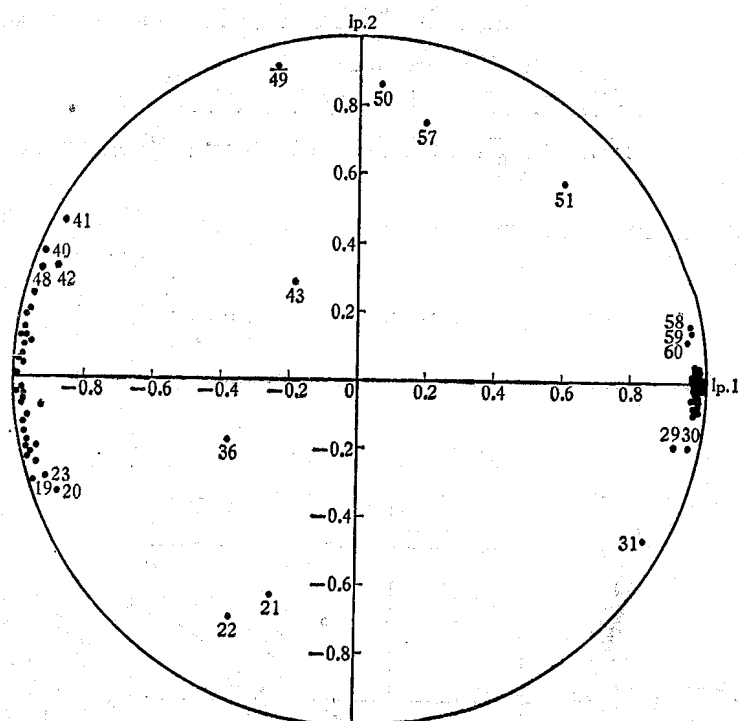


图 9 东海表层 82 个测站在 (f_1, f_2) 空间中的分布格局

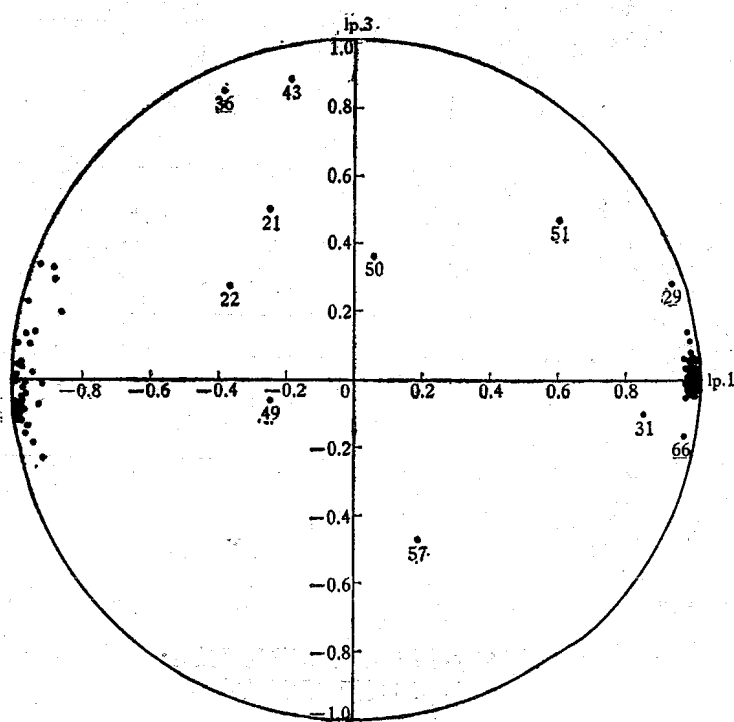


图 10 东海表层 82 个测站在 (f_1, f_2) 空间中的分布格局

(32°N, 128°E) 附近还有一个较弱的封闭性中心。对比图 6 可以看出, 这两个高值中心恰好位于外海(黑潮)暖水和陆缘(黄海)冷水的分界线 ($l_{p,1} = 0$) 附近。

上述第 1, 2, 3 对时-空变异格局函数对东海表层多年平均水温规格化季节距平总方差的累积贡献等于 97.76%; 其余高阶时-空变异格局函数 (l_i, f_i) ($i = 4, 5, \dots, 36$) 的方差贡献累积百分比仅约 2.24%, 可略去不计。

3. 测站的相似聚类

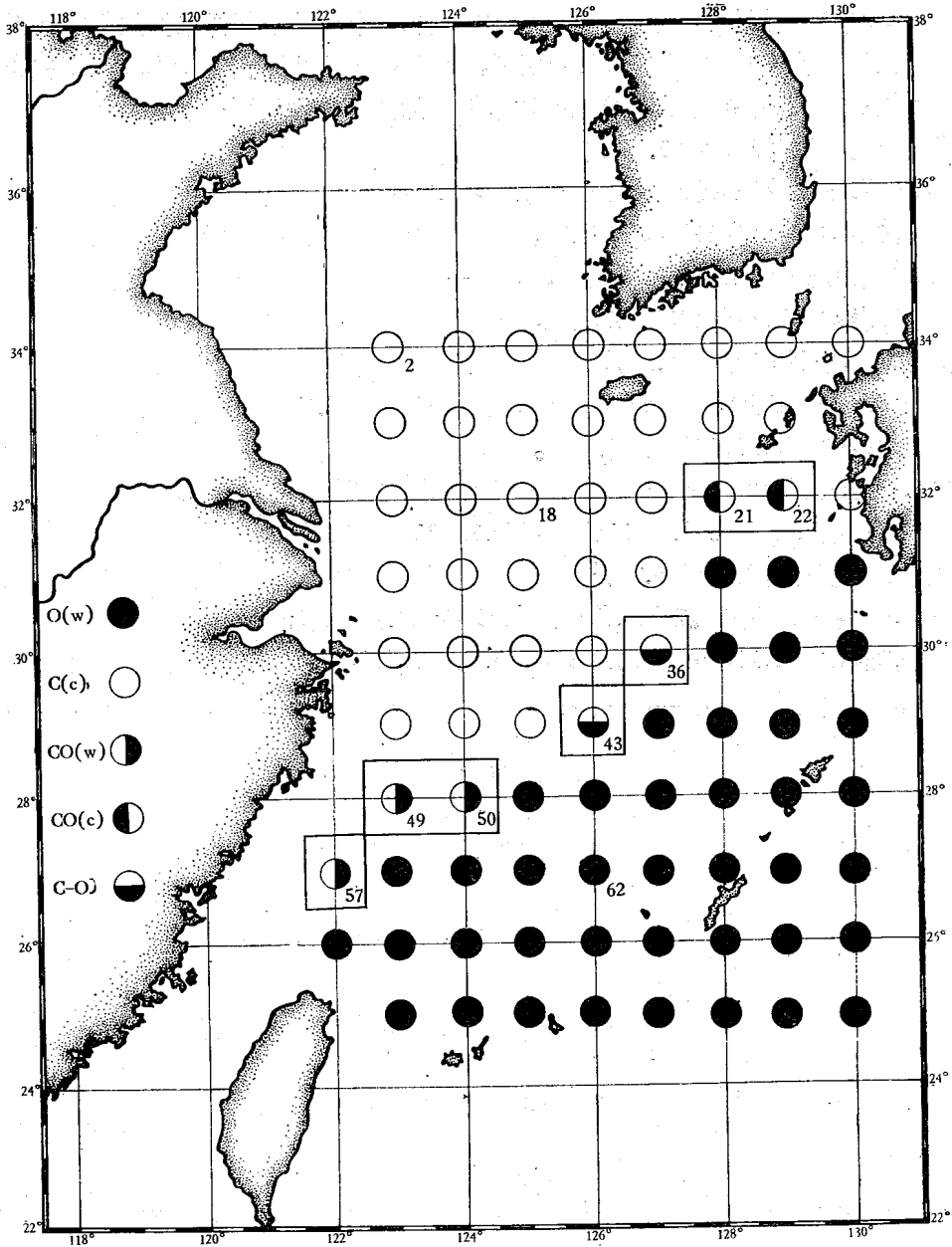


图 11 按加权空间格局函数 l_1, l_2, l_3 划分的东海表层水系配置图

如前所述,我们在概念上可以将东海表层 82 个测站多年的 36 旬平均水温季节距平规格化数据矩阵 $\Theta_{m \times n}$ 表示为 36 维欧氏空间坐标中的 82 个矢量(或点)的散布图形;而通过对这一数据矩阵的奇异值分解的结果表明,我们可将这一高维空间矢量(或点)的散布图形投影到以正交单位矢量 f_1, f_2, f_3 为基底的三维空间中而去而仍能保留它们在原来高维空间分布格局的基本特征。为直观起见,图 9 和图 10 分别绘出了在 (f_1, f_2) 和 (f_1, f_3) 这两个二维欧氏空间中 82 个测站的分布格局。图中,每一点的坐标相应为 $(l_{p,1}; l_{p,2})$ 和 $(l_{p,1}; l_{p,3})$, $p = 1, 2, \dots, 82$ 。显然,在这两幅图中,我们无需采用任何聚类分析(cluster analysis)的计算程序即可直观地看出这 82 个测站是怎样聚类的,并且不难选出每一聚类相应的代表站。这样,我们便可将东海表层多年平均水温季节距平的时-空变异按其相对于 f_1, f_2 和 f_3 的亲疏程度逐级划分为如下几种类型:

I. 第一级划分(等级水平 87.17%)

I-1. 外海暖水(黑潮暖水) O(w); $l_{p,1} > 0.5$;

代表站 No.61(27°N, 126°E)。

I-2. 陆缘冷水(黄海冷水) C(c); $l_{p,1} < -0.5$;

代表站 No.18(32°N, 125°E)。

II. 第二级划分(等级水平 6.57%)

II-1. 陆缘-外海西部混合暖水(台湾暖流水) CO(w); $l_{p,2} > 0.5$;

代表站 No.49(28°N, 123°E)。

II-2. 陆缘-外海东部混合冷水(对马暖流源地) CO(c); $l_{p,2} < 0.5$;

代表站 No.21(32°N, 128°E), No.22(32°N, 129°E)。

III. 第三级划分(等级水平 4.03%)

III-0. 陆缘-外海中部混合水(“流隔”) C-O; $l_{p,3} > 0.5$;

代表站 No.36(30°N, 127°E), No.43(29°N, 126°E)。

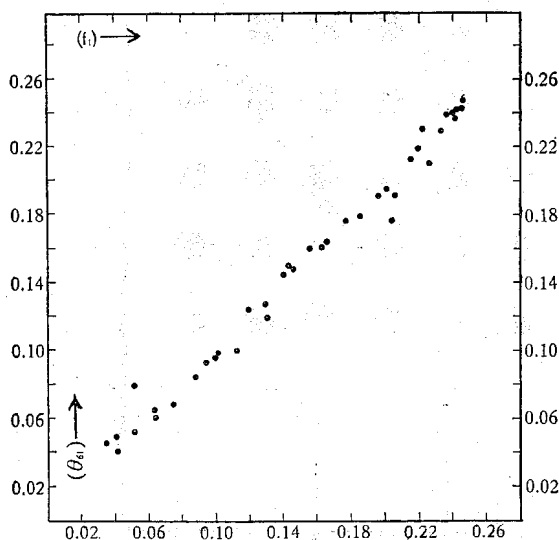


图 12 O(w)代表站 No.61 的规格化表层水温季节距平年变化与 f_1 的对比

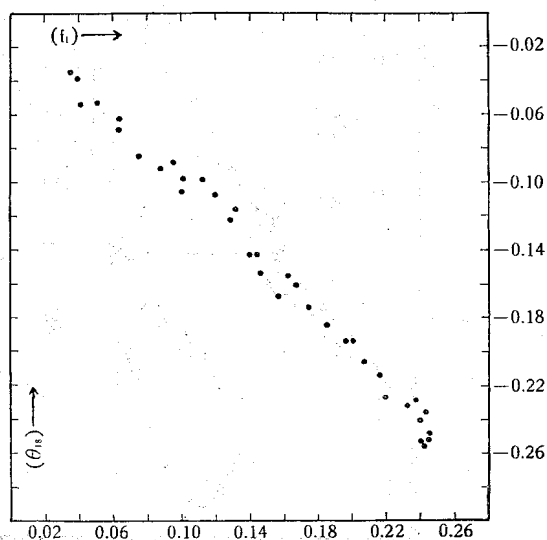


图 13 C(c)代表站 No.18 的规格化表层水温季节距平年变化与 f_1 的对比

准此,可将上述测站聚类结果绘成东海表层水系配置图,如图 11 所示,而各水系代表站的规格化水温季节距平年变化相应与第 1, 2, 3 时间变异格局函数的对比则见于图 12—15。

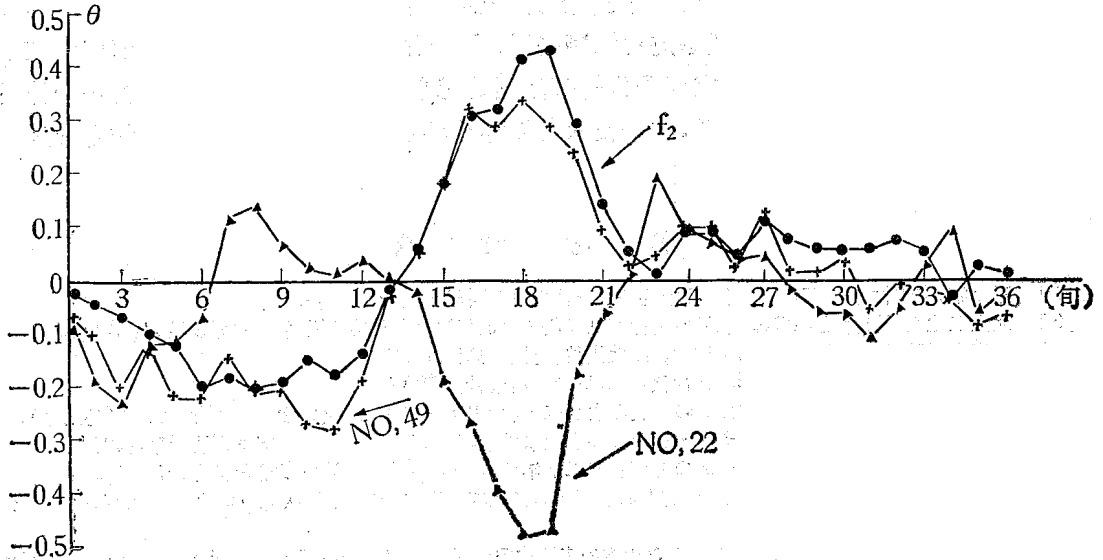


图 14 CO(w)代表站 No.49 和 CO(c)代表站 No.22 的规格化表层水温季节距平年变化与 f_2 的对比

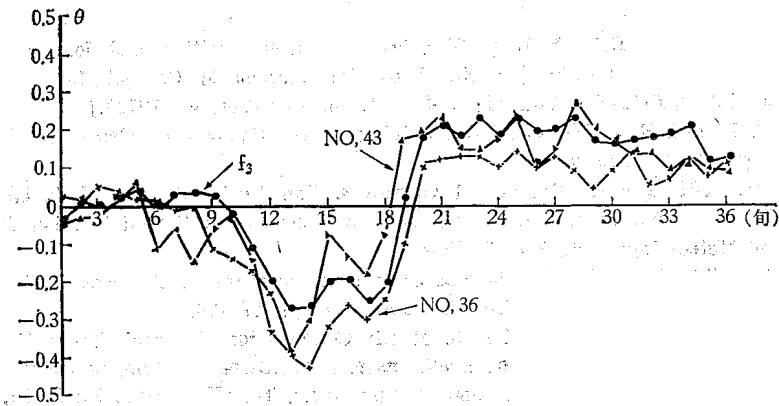


图 15 C-O 代表站 No.36, No.43 的规格化表层水温季节距平年变化与 f_3 的对比

四、结 语

本文应用数据矩阵的奇异值分解技术分析了东海表层多年(1953—1972)表层旬平均水温的时-空变异特征。结果表明,该海区表层多年旬平均水温规格化季节距平数据矩阵可以足够近似地由其前三对奇异矢量(即空间格局函数和时间格局函数)的外积项的线

性组合表出。准此,我们可以在仅仅“损失”原信息 2.24% 的情况下,将东海表层 82 个测站逐级划分为五种类型,从而为东海表层的划分提供了一项客观依据。和以往基于该海区表层水温季节变化而提出的一些划分结果^[8,9,10,17]相比较,本文得到的结果似更合理。

数据矩阵的奇异值分解(或奇异矢量分析)技术,具有良好的数值稳定性,对于多元观测数据的统计分析和图象处理等应用问题来说,是一种强有力的工具。和目前广泛应用的特征矢量分析、主成分分析或因子分析等数据处理方法相比较,奇异矢量分析是一种更一般、更灵活的方法。可以预料,这一技术在海洋观测数据处理和水文要素场的客观分析等方面将有广泛的应用。

参 考 文 献

- [1] 王宗皓、李麦村等, 1974. 天气预报中的概率统计方法. 科学出版社, 206 页。
- [2] 北京工业大学计算站, 1976. 怎样使用计算机 (DJS-6 机算法语言及使用说明). 科学出版社, 240 页。
- [3] 南京大学数学系计算数学专业, 1978. 线性代数. 科学出版社, 201 页。
- [4] 蒋尔雄、高坤敏、吴景琨, 1978. 线性代数, 人民教育出版社, 479 页。
- [5] 长崎海洋气象台, 1958. 東支那海の旬表面水温五年報 (1953—1957). 気象庁海洋報告, 第 7 号(特別号)。
- [6] 长崎海洋气象台, 1963. 東シナ海・黄海の旬表面水温十年報 (1953—1962). 気象庁技術報告, 第 31 号。
- [7] 长崎海洋气象台, 1975. 東シナ海の海面水温二十年報 (1953—1972). 気象庁技術報告, 第 89 号。
- [8] 菊池繁雄 (Kikuchi, S.), 1959. 東支那海表面水温に関する研究 (I)—平均状態の特性. 研究時報 11(3): 169—183。
- [9] 小泉政美 (Koizumi, M.), 1964. 東シナ海の表面水温の標準偏差について. Studies on Oceanography. Hidaka Jubille Committi, Tokyo, pp.140—144.
- [10] 佐原勉, (Sawara, T.), 1975. 同[7]。
- [11] Anderson, T. W., 1958. An Introduction to Multivariate Statistical Analysis. Wiley, 374pp.
- [12] Cattell, R. B., 1965. Factor analysis: An introduction to essentials. Biometries 21: 190—215, 405—435.
- [13] Golub, G. H. and C. Reinsch 1970. Singular value decomposition and least squares solutions. Numer. Math. 14: 403—420. [in Handbook for Automatic Computation. (Editors: J. H. Wilkinson and C. Reinsch) Vol. II: 134—151, Springer-Verlag (1971).]
- [14] Lawson, C. L. and R. J. Hanson, 1974. Solving Least Squares Problems. Prentice-Hall, Englewood Cliffs. N. J., 340pp.
- [15] Lanczos, C., 1961. Linear Differential Operators. Van Nostrand, New York, 564pp.
- [16] Lorenz, E. N., 1956. Empirical Orthogonal Functions and statistical Weather Predictor. M. I. T. Dep. of Meteorology, Sci. Rep. 1, 49pp.
- [17] Sawara, T. (佐原勉), 1974. On the sea surface temperature in the East China Sea for the year 1953—1972. Oceanographical Magazine 26(1): 17—36.
- [18] Stewart, G. W., 1973. Introduction to Matrix Computation. Academic Press, New York, 564pp.
- [19] Багров Н. А., 1959. Аналитическое представление последовательности метеорологических полей посредством естественных ортогональных составляющих. Тр. ЦИП, Вып. 74: 3—24.

OBJECTIVE ANALYSIS AND CLASSIFICATION OF THE MEAN SEA SURFACE TEMPERATURE IN THE EAST CHINA SEA BY APPLYING THE SINGULAR-VALUE DECOMPOSITION TECHNIQUE*

Gan Zijun and Mao Hanli

(*Institute of Oceanology, Academia Sinica*)

ABSTRACT

The seasonal changes of the mean 10-day sea surface temperature in the East China Sea for the years 1953—1972 were studied by applying the singular-value decomposition technique. Main results obtained are as follows:

In the singular-vector analysis of the mean 10-day sea surface temperature in the East China Sea, the first three singular-vectors (or seasonal march vectors) account for 97.76% of the variance of the data.

The geographical distribution in the weightings (or spatial pattern vectors) on these vectors indicates a markedly regional organization.

An objective classification of the surface water mass in the East China Sea was provided by its seasonal temperature change pattern.

It may be safely said that the singular-value decomposition technique may be used much more extensively in oceanographic investigations.

*Contribution No. 655 from the Institute of Oceanology, Academia Sinica.