

用陆地卫星数据描绘渤海湾海河口区 不同污染等级*

平 仲 良

(中国科学院海洋研究所)

Okamoto (1969) 指出, 多变量的主成分具有变差最优性、信息损失最小性、相关最优性和回归最优性等最优性质, 因此多变量的主成分分析被广泛应用于自然科学和社会科学多种领域。由于主成分可以看作所研究的某个复杂事物的一个概括(肯德尔, 1975), 并可根据主成分所提供的数据, 按数值大小排列顺序分级, 因此主成分分析也被遥感界应用于区分不同水质 (Klemas et al, 1981)。

本文以渤海湾西北部海河、永定新河河口区的陆地卫星数据作主成分分析, 以显示河口区海水不同污染等级。同时以河口区各站位实测的多项有机污染数据作主成分变换, 按主成分数值大小, 将各站位分为不同污染等级。计算表明, 遥感数据显示的不同污染等级的地理分布与实测数据是一致的。

一、材料与 方法

记录陆地卫星数据的计算机兼容磁带是由日本遥感科学技术中心提供, 数据记录时间为 1979 年 5 月 19 日。磁带数据由中国科学院空间中心地面部图象处理系统 (SIPS) 打印输出。在卫星图象上, 在天津海河、南排污河河口区以 64, 64 采样, 取 100 个象元 4 个波段的数据。辐射数据已经大气校正, 列于表 1。

主成分变换的方法如下:

1. 将 100 个象元 4 个波段的辐射值看作 4 维空间的变量 x_j 。主成分变换就是要找出一组新变量, 使它们是 x_j 的线性函数, 而且彼此不相关。

$$\xi_i = \sum_{j=1}^4 L_{ij}x_j, \quad i = 1, 2, 3, 4$$

式中 x_1, x_2, x_3, x_4 为第 4, 5, 6, 7 波段的数据。

2. 计算 100 个象元 4 个波段数据的方差和波段间的协方差值 $C_{11}, C_{22}, C_{33}, C_{44}$ 和 $C_{12}, C_{13}, C_{14}, C_{23}, C_{24}, C_{34}$ 。

3. 计算协方差矩阵

* 中国科学院海洋研究所调查研究报告第 1347 号。
收稿日期: 1986 年 7 月 10 日。

表 1 河口区 100 个象元 4 个波段的辐射数据

序号	x_1	x_2	x_3	x_4	序号	x_1	x_2	x_3	x_4	序号	x_1	x_2	x_3	x_4	序号	x_1	x_2	x_3	x_4
1	14	16	8	3	26	11	12	7	2	51	15	12	6	1	76	14	13	5	2
2	13	17	8	3	27	11	13	7	2	52	15	12	5	1	77	14	14	5	3
3	13	14	8	3	28	12	12	7	2	53	16	19	10	1	78	14	13	6	1
4	13	14	7	2	29	11	12	6	0	54	11	13	7	1	79	17	14	7	1
5	13	12	7	3	30	11	9	5	2	55	11	12	6	2	80	14	14	7	3
6	15	13	5	3	31	11	12	6	1	56	10	12	5	3	81	14	13	6	3
7	14	12	6	1	32	11	10	5	0	57	11	13	4	1	82	14	11	6	1
8	14	13	6	2	33	13	11	5	2	58	12	12	5	1	83	14	11	5	1
9	15	14	4	2	34	14	10	6	0	59	11	11	6	1	84	15	24	11	5
10	13	12	6	1	35	15	12	5	3	60	12	10	5	1	85	17	21	10	5
11	7	15	9	4	36	13	11	6	0	61	13	13	6	2	86	14	17	7	3
12	8	14	8	2	37	14	11	7	1	62	13	12	7	1	87	13	12	5	3
13	10	15	8	2	38	15	12	5	1	63	14	11	6	1	88	13	12	4	2
14	10	11	8	3	39	13	13	7	2	64	14	10	5	1	89	13	12	6	1
15	12	15	6	2	40	11	12	7	1	65	14	12	6	1	90	14	11	6	3
16	13	14	8	3	41	11	13	6	3	66	13	12	5	1	91	15	12	5	3
17	13	14	7	2	42	11	11	6	1	67	13	12	6	1	92	14	13	5	2
18	12	14	7	2	43	11	11	5	2	68	15	21	11	3	93	16	14	7	1
19	14	13	5	1	44	11	11	5	1	69	15	16	9	0	94	16	14	5	4
20	13	12	6	3	45	11	10	4	1	70	14	13	8	1	95	15	13	6	2
21	15	13	5	2	46	13	10	6	1	71	10	11	7	4	96	18	15	6	0
22	15	12	4	1	47	14	10	5	0	72	13	12	6	1	97	17	14	6	4
23	15	13	6	2	48	12	11	5	1	73	11	12	5	1	98	16	15	6	2
24	15	13	6	2	49	13	11	5	1	74	12	12	6	3	99	13	12	5	2
25	7	12	9	3	50	15	12	6	2	75	14	12	7	2	100	15	12	5	3

$$C_x = \begin{pmatrix} C_{11} & C_{12} & C_{13} & C_{14} \\ C_{21} & C_{22} & C_{23} & C_{24} \\ C_{31} & C_{32} & C_{33} & C_{34} \\ C_{41} & C_{42} & C_{43} & C_{44} \end{pmatrix}$$

的 4 个特征值。4 个特征值由方程

$$\begin{vmatrix} C_{11} - \lambda & C_{12} & C_{13} & C_{14} \\ C_{21} & C_{22} - \lambda & C_{23} & C_{24} \\ C_{31} & C_{32} & C_{33} - \lambda & C_{34} \\ C_{41} & C_{42} & C_{43} & C_{44} - \lambda \end{vmatrix} = 0$$

求得。4 个特征值 $\lambda_1 = 7.351$, $\lambda_2 = 3.788$, $\lambda_3 = 0.902$, $\lambda_4 = 0.593$ 。

4. 由 4 个特征值分别计算 4 个主成分。由方程组

$$\begin{aligned} (C_{11} - \lambda)L_{11} + C_{12}L_{12} + C_{13}L_{13} + C_{14}L_{14} &= 0 \\ C_{21}L_{11} + (C_{22} - \lambda)L_{12} + C_{23}L_{13} + C_{24}L_{14} &= 0 \\ C_{31}L_{11} + C_{32}L_{12} + (C_{33} - \lambda)L_{13} + C_{34}L_{14} &= 0 \\ C_{41}L_{11} + C_{42}L_{12} + C_{43}L_{13} + (C_{44} - \lambda)L_{14} &= 0 \end{aligned}$$

$$\text{及 } C_{11}^2 + C_{12}^2 + L_{13}^2 + L_{14}^2 = 1$$

求得 $L_{11} = 0.364$, $L_{12} = 0.831$, $L_{13} = 0.373$, $L_{14} = 0.194$ 。则第一主成分

$$\xi_1 = 0.364x_1 + 0.831x_2 + 0.373x_3 + 0.194x_4$$

用同样的方法可以求得第二、三、四主成分。

$$\xi_2 = -0.895x_1 + 0.186x_2 + 0.367x_3 + 0.171x_4$$

$$\xi_3 = 0.0413x_1 - 0.114x_2 - 0.298x_3 + 0.947x_4$$

$$\xi_4 = 0.238x_1 - 0.498x_2 + 0.542x_3 + 0.634x_4$$

对几个主成分的物理含义作如下探讨：第一主成分与 4 个波段的辐射值呈正的线性关系，且主要与第 5 波段的数据有关，它主要反映海水中悬浮泥沙含量的信息。第二主成分与第 4 波段数据呈负相关关系，而与第 5, 6, 7 波段数据呈正相关关系，它反映枯水期污水的光学性质，因而主要反映海水不同程度的污染及其影响。第三主成分主要由第 7 波段辐射数据决定，它主要反映海水表层和表面性质。第四主成分主要受噪声控制。因本文主要研究河口区海水污染问题，因此除第二主成分外，对其他主成分不作进一步讨论。

5. 将河口区每一象元 4 个波段数据代入第二主成分，得到如图 1 所示的河口区第二主成分图象。



图 1 渤海湾西北部河口区 1979 年 5 月 19 日陆地卫星数据第二主成分图象

6. 将河口区每一象元的第二主成分按其数值大小分级, 以三种不同的标记表示污染区、排污影响区和未污染区三个等级, 得到如图 2 所示的河口区不同污染等级的分布图。

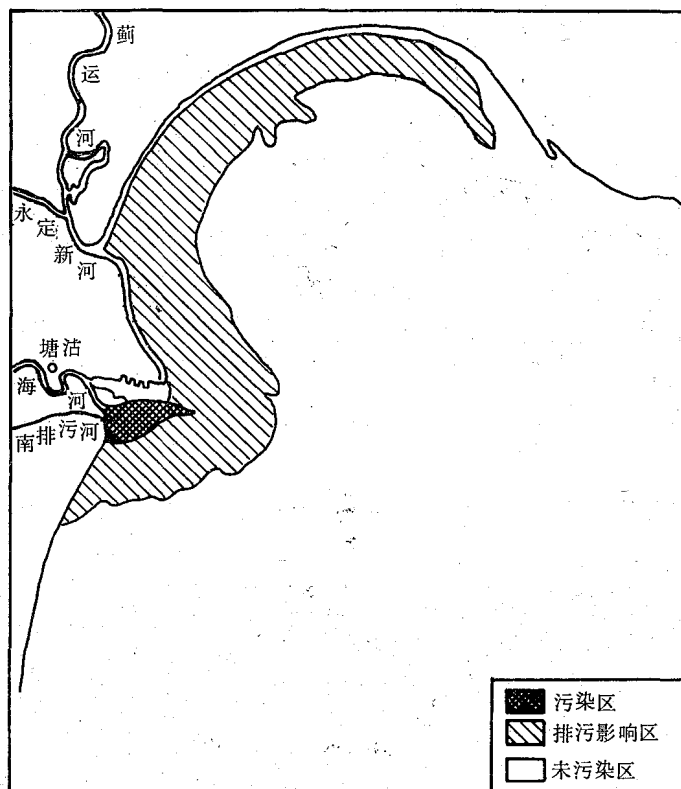


图 2 渤海湾西北部河口区 1979 年 5 月 19 日陆地卫星数据第二主成分所显示的不同污染等级

二、海洋实测数据的主成分分析

表 2 所列有机污染数据是由中国科学院海洋研究所“渤海湾河口区有机污染及其生态效应”课题组 1984 年 5 月在河道段 5 个站位和河口区 23 个站位实测所得。选择 COD (化学耗氧量)、BOD (生物耗氧量)、TN (总氮)、TP (总磷)、CB (大肠菌群数的对数)、DS (浮游藻类的多样性指数) 6 个参数作主成分分析是因为 COD, BOD, TN, TP 公认是水质污染的主要指标, CB 和 DS 公认是海洋受陆源性有机污染的指标 (钱振儒等, 1985)。另一个原因是这些参数彼此之间的相关系数较高, 说明它们的变化规律很一致, 数据是可靠的, 所以它们的第一主成分能全面、确切地反映有机污染状况。

实测数据的主成分变换方法与卫星遥感的多光谱数据的变换类同, 不再赘述。但有两个重要差别须要指出。一是遥感数据的尺度是相同的, 因此在进行多光谱数据的主成分变换时使用了协方差矩阵; 而海洋实测的化学、生物数据的尺度是不同的, 采用协方差矩阵就使变换结果过分地突出了那些数值大的量的作用。为了弥补这种方法的不足, 须要将初始变量标准化, 使其方差为 1, 即使用相关矩阵计算主成分。二是污水流入海洋,

表 2 1984 年 5 月河口区各站位的有机污染数据

站位	COD (mg/L)	BOD (mg/L)	TN (mg/L)	TP (mg/L)	CB (lg 个数/100 mL)	DS
3	1.38	0.44	41.20	1.71	1.602	1.1
5	2.88	0.93	56.40	1.07	2.362	1.7
6	2.12	1.00	37.60	0.57	1.602	1.9
7	2.08	0.85	39.60	0.78	2.176	2.1
8	2.04	0.72	28.40	1.34	1.602	1.5
10	1.04	1.28	46.80	0.98	1.602	1.1
11a	10.08	5.92	111.20	3.45	3.380	0.4
12	2.28	1.34	59.20	1.82	1.845	1.9
13	1.18	0.77	40.00	1.32	1.845	1.9
14	1.56	0.60	51.60	1.68	1.602	1.9
15	1.36	1.05	33.60	1.10	1.602	2.0
16	1.18	0.98	34.80	0.78	1.602	1.6
18	1.68	0.53	44.00	1.34	1.602	1.7
19	1.60	0.77	46.40	1.49	1.602	2.0
20	1.54	0.48	39.60	0.59	1.602	3.0
21	1.86	0.74	29.60	0.80	1.602	2.3
22	1.6	0.31	21.60	0.40	1.602	1.3
C	13.52	2.06	142.00	23.36	5.041	1.2

注：C 站为永定新河河道站。

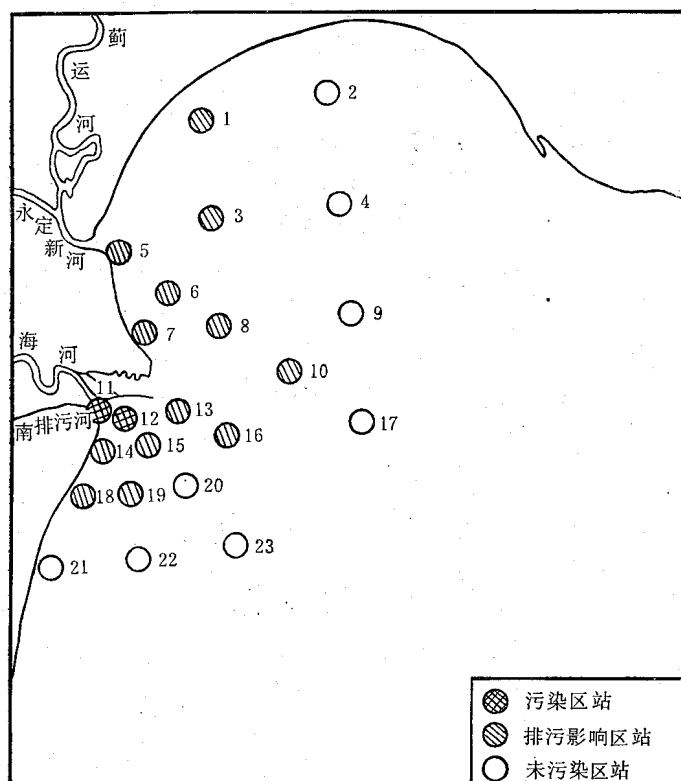


图 3 渤海湾西北部河口区各站位不同主成分数值范围,即不同污染等级

使得比较浑浊的沿岸区海水在光谱绿区的反射率降低,红区和近红外区的反射率提高,污水成分愈高,这种效应就愈显著。而遥感数据的第二主成分正好反映这种情况,因此用第二主成分分布图显示河口区海水受城市污水污染的状况。而海洋实测数据的第一主成分则反映各污染参数对海洋污染的综合“贡献”,所以用第一主成分反映各站位海水污染状况。还有一点须要特别指出,求高于 5×5 阶矩阵的特征值是件非常麻烦的事,本文采用高尔腊伊(1973)介绍的强特征值的计算方法比较容易地求得特征值的解。由此求得的多变量第一主成分

$$Z = 0.4431\text{COD} + 0.4477\text{BOD} + 0.4383\text{TN} + 0.3926\text{TP} \\ + 0.4247\text{CB} - 0.2758\text{DS}$$

式中变量 COD, BOD 等为已经过标准化的数据。由此计算得到各站位第一主成分值,然后将主成分值划分为 3 个区间,用不同标记表示渤海湾西北部河口区各站位的不同污染等级,如图 3 所示。

三、几点结论

1. 由图 2 可见,在海河、南排污河河口,塘沽新港港池以南,有一近直角三角形区(网格标记),显示遥感数据第二主成分数值最高。由图 3 可见,海河、南排污河河口的 11, 12 站位实测数据的第一主成分数值最高,为城市污水直接污染区。该区特点一是海水光学性质直接受污水影响,绿波段辐射值较低,红波段及近红外波段辐射值较高。二是泥沙含量低, COD, BOD, TN, TP, CB 数值大, DS 数值小。例如 11a 站,其指示有机体含量的 COD, BOD 值为未受城市污水影响的 20, 22 站的 4—9 倍,代表有机营养成分含量的 TN, TP 值为 5—6 倍,大肠菌群数的对数为 3.25 倍,而浮游植物的多样性指数和叶绿素 a 浓度反而较低。三是污水不易扩散。这个三角形区北面有塘沽新港的南阻浪堤,西面海岸向陆地缩进,是渤海湾沿岸流不能到达的死角。

2. 自海河、南排污河河口的网格区向外为斜线区,由河口沿岸向北延伸,直到湾顶,为污水继续稀释、扩散和影响区,区内氮、磷营养盐含量较高,浮游植物生长茂盛,叶绿素 a 浓度较高,氨氮营养盐的含量与叶绿素 a 浓度呈正相关关系。

在污水扩散和影响区,污水与海水的量相比,只占很小的百分比,特别是在枯水期,污水不至于直接影响到海水的颜色。但它通过丰富的营养盐影响海水中叶绿素浓度,从而改变海水的光学性质,与背景海水相区别,所以污水影响区亦能从遥感图象上显示出来。污水扩散影响区的图象亦显示沿岸流的流向(见图 2)。

3. 从斜线区再向外为未污染区(未作标记),区内海水的光学性质,生物、化学参量基本未受城市污水影响。由图 3 可见,20, 22 等站虽与海河、南排污河河口接近,但这些站位实测的污染数据的第一主成分数值很小,它们处于图 2 的未污染区。其原因是由南排污河排出的城市污水在向东南方向扩散的同时,在渤海湾沿岸流作用下向北运移,由河口排出的污水实际上并未到达这些站位(平仲良, 1983; 1984)。

4. 研究表明,不是任何一幅图象在任何情况下作一下主成分变换即能显示污水扩散形式和等级(或分类类别)。它需要两个条件:一是卫星记录了(或卫星数据能反映)污水

排放或稀释扩散的景象，二是卫星数据的某一主成分反映污水的光谱特性；二者缺一不可，否则不会得到满意的结果。

参 考 文 献

- 平仲良, 1983. 从 ERTS 图象上观测渤海湾表层流. 海洋与湖沼 14(3): 297—304.
 平仲良, 1984. 渤海湾航空遥感的图象分析——渤海湾近岸叶绿素 *a* 分布图. 海洋科学集刊 23: 51—56.
 肯德尔, M., 1975. 多元分析. 中国科学院计算中心概率统计组译, 1983. 科学出版社, 16—37 页.
 高尔腊伊, A. R., G. A. 瓦特桑, 1973. 矩阵特征问题的计算方法. 唐焕文等译, 1983. 上海科学技术出版社, 45—55 页.
 钱振儒、张景镛, 1985. 城市生活污水细菌在海水中死亡机制的研究. 海洋环境科学 4(2): 28—33.
 Klemas, V. *et al*, 1981. Drift and dispersion studies of ocean-dumped waste using Landsat imagery and current drogues. *Photogrammetric Engineering and Remote Sensing* 47(4): 533—542.
 Okamoto, M., 1969. *Multivariate Analysis II* (Ed. by P. R. Krishnaiah), Academic Press, New York. 转引自科学通报 29(8): 449—451.

DESCRIBING DIFFERENT POLLUTED GRADATION OF THE SEA WATER IN THE ESTUARY AREA OF THE HAIHE RIVER IN THE BOHAI BAY WITH LANDSAT DATA*

Ping Zhongliang

(Institute of Oceanology, Academia Sinica)

ABSTRACT

Principal component analysis of Landsat data obtained over the estuary area of the Haihe River-Yongdingxin River in the northwest of the Bohai Bay shows the different pollution gradations of the sea water. Principal component transformation of the organic pollution data measured *in situ* was also performed. The values of the first principal component were divided into three regions to show the different pollution gradations at these experimental sites.

The results of the investigation show that the geographical distribution of the different pollution gradations obtained from remotely sensed data agrees with that from sea-truth measured data.

The investigation comes to these conclusions: 1. In the estuary of the Haihe River and the South Sewerage River, the optical property of the sea water is directly affected by urban sewage, the values of COD, BOD, TN, TP, CB are high, and the values of DS are low. The water quality is the worst. This is a polluted area by urban sewage. 2. The urban sewage affected area is out of the polluted area, it originates from the estuary, and extends southeastwards, then spreads northwards to the apex of the Bohai Bay along the coast. In this area, the contents of nutrient substance of N, P and the chlorophyll-*a* concentration are higher. 3. The rest is a non-polluted area, where the optical property of sea water is not affect by sewerage, the biological and chemical data of water are near the background levels, the water quality is good.

* Contribution No. 1347 from the Institute of Oceanology, Academia Sinica.