Using chemometrics to identify water quality in Daya Bay, China*

OCEANOLOGIA, 51 (2), 2009.
pp. 217–232.

KEYWORDS
Cluster analysis
Robust principal component analysis
Water quality
Daya Bay (DYB)
South China Sea

Mei-Lin Wu[1]
You-Shao Wang[1,*]
Cui-Ci Sun[1]
Haili Wang[2]
Zhi-Ping Lou[1]
Jun-De Dong[1]

[1] Key Laboratory of Tropical Marine Environmental Dynamics,
South China Sea Institute of Oceanology,
Chinese Academy of Sciences,
Guangzhou 510301, China;

e-mail address: yswang@scsio.ac.cn

*corresponding author

[2] Scripps Institution of Oceanography,
University of California,
San Diego, CA 92093–0218, USA

Received 13 January 2009, revised 18 March 2009, accepted 20 March 2009.

Abstract

In this paper, chemometric approaches based on cluster analysis, classical and robust principal component analysis were employed to identify water quality in

---

* This research was supported by the project of knowledge innovation program of the Chinese Academy of Sciences (No. KZCX2-YW-Q07-02 & No. KSCX2-SW-132), the project of knowledge innovation program of the South China Sea Institute of Oceanology (No. LYQ200701) and the National 908 project (No. 908-02-04-04).

Daya Bay (DYB), China. The results show that these approaches divided water quality in DYB into two groups: stations S3, S8, S10 and S11 belong to cluster A, which lie in Dapeng Cove, Aotou Harbor and the north-eastern part of DYB, where water quality is related mainly to anthropogenic activities. The other stations belong to cluster B, which lie in the southern, central and eastern parts of DYB, where the quality is related mainly to water exchange with the South China Sea. Cluster analysis yields good results as a first exploratory method for evaluating spatial difference, but it fails to demonstrate the relationship between variables and environmental quality on the one hand and the untreated data on the other. However, with the aid of suitable chemometric approaches, the relationship between samples or variables can be investigated. Classical and robust principal component analysis can provide a visual aid for identifying the water environment in DYB, and then extracting specific information about relationships between variables and spatial variation trends in water quality.

1. Introduction

Coastal waters are very vulnerable to pollution caused by wastewater, runoff, effluents, land reclamation, recreation and aquaculture, as well as atmospheric deposition and climate change (Bowen & Depledge 2006, Kuppusamy & Giridhar 2006). To prevent/monitor coastal water pollution, it is imperative to have reliable information on the quality of water for effective management. The measurement of hydrochemistry variables in the marine environment promotes a better understanding of the aquatic environment. These variables produce large sets of data which are often difficult to interpret (Vega et al. 1998, Simeonov et al. 2003, Singh et al. 2004, Shrestha & Kazama 2007). Data interpretation of multidimensional measurements can be approached by the application of chemometric methods such as principal component analysis (PCA) (Simeonov et al. 2003, Stanimirova et al. 2003, Wang et al. 2006, Chau & Muttil 2007, Suikkanen et al. 2007, Wu & Wang 2007, Zhou et al. 2007). PCA is a powerful tool in chemometrics and environmetrics for compressing data and extracting information. It can reduce the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much of the variability present in a data set as possible. This reduction is achieved by transforming the data set into a new set of variables, the principal components (PCs), which are orthogonal (non-correlated) and arranged in decreasing order of importance. However, it is well known that PCA, like any other multivariate statistical method, is sensitive to outliers, missing data, and poor linear correlation between variables due to their poor distribution. As a result, data transformations have a large impact on PCA. In this regard one of the most powerful approaches to improve PCA appears to be robust principal component analysis (RPCA), which diminishes the influence of outliers (Stanimirova et al. 2004, Croux & Ruiz-Gazen 2005, Gong et al. 2005).

In the present study, a robust principal component analysis approach was applied to coastal water quality data in order to acquire a better understanding of the role of water quality as a pollution indicator, and to identify the contribution of natural and anthropogenic factors to water quality variations in temporal and spatial patterns. Additionally, the ecological situation of the coastal regions could be estimated more reliably.

2. Material and methods

2.1. Study area

Daya Bay lies on the southern coast of China (lat. 22.31′12″–22.50′00″N, long. 114.29′42″–114.49′42″E), and is door-shaped. There are two towns (Dapeng town and Nanao) in the western coastal area. The northern and eastern coastal areas belong to Huizhou, Guangdong Province. There are five towns on the bay: Xiachong, Aotou, Nianshan, Xunliao and Gangkou. In recent years, the rapid economic development and anthropogenic activities in Shenzhen and Huizhou have had a great influence on the environment of this bay. For example, two nuclear power plants have come into operation

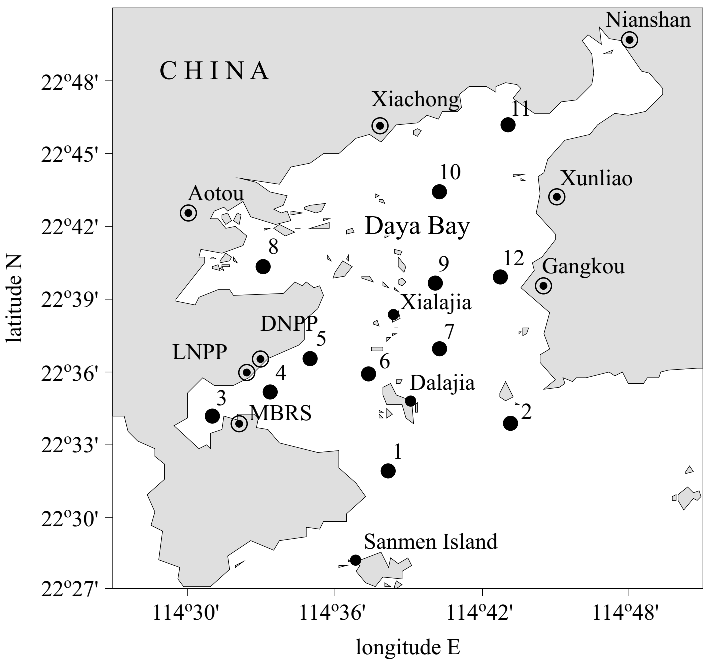

Figure 1. Monitoring stations in Daya Bay (Wang et al. 2006, 2008)

(Wang et al. 2006), and the marine aquaculture industry has increased in importance. The strong north-east monsoon prevails from October to April, while the south-east Asian monsoons blowing in from the south-west predominate from May to September (Wang et al. 2008). In order to evaluate the anthropogenic and natural effects in this bay, the survey stations were located as follows.

A total of 12 monitoring stations are located in DYB (Figure 1) (Wang et al. 2006, 2008). Stations 1, 2, 6, 7, 9, 10, 11 and 12 are in the areas between the mouth and the top of the bay for evaluating the effect of the South China Sea and anthropogenic influence. Stations 3 and 8 are in the aquaculture area for assessing the influence of fish-farming. Stations 4 and 5 are in the outfall and effluent areas of the Lingao Nuclear Power Plant (LNPP) and the Daya Bay Nuclear Power Plant (DNPP) (Wang et al. 2006).

2.2. Analytical methods

Water samples were taken at the surface and from the bottom layers of all stations in both the dry season (January and April) and the wet season (August and November) in 2003. The determinations of pH, water temperature and salinity were performed in situ using the water quality Monitoring System. The other analyses were performed within 48 h after sampling. The analytical parameters were determined in triplicate with reference to current official methods ('The specialties for oceanography survey' GB12763-91, China). The water quality parameters, their units and analytical methods are summarized in Table 1.

2.3. Data treatment

No thermal stratification occurred in DYB, and little difference in hydrographic parameters was observed between surface and bottom waters. Therefore, the averaged data obtained from surface and bottom waters were used in this study.

Standardized skewness and standardized kurtosis were determined to see whether a sample came from a normal distribution. Values of these statistics outside the range of -2 to $+2$ indicate significant departures from normality. Statistical analysis of the data showed that all variables in the original dataset, except pH, are normally distributed. The basic statistics of the data set on water quality are summarized in Table 2.

Since classical PCA is strongly affected by outliers, it is not a robust approach in chemometrics. At the same time, a situation with more variables than observations ($n < p$) is frequently encountered in practical applications: classical PCA might fail if $n < p$. In this study, robust PCA,

Table 1. Physical-chemical and biological parameters determined and analytical methods used (Wang et al. 2006, 2008)

Parameters	Abbreviation	Units	Instrument	Analytical method
temperature	T	°C	A Quanta® Water Quality Monitoring System (Hydrolab Corporation, USA)	
salinity	S	PSU		
pH	pH			
dissolved oxygen	DO	mg dm^{-3}		Winkler titration
chemical oxygen demand	COD	mg dm^{-3}		Potassium dichromate oxidation
5-day biochemical oxygen demand	BOD_5	mg dm^{-3}		5-day incubation, 20
nitrite	NO_2-N	μmol dm^{-3}	A SKALAR auto-analyzer (Skalar Analytical B.V. SanPlus, the Netherlands)	The Griesse-Ilosvay method
nitrate	NO_3-N	μmol dm^{-3}		Cadmium-copper reduction
silicate	SiO_3-Si	μmol dm^{-3}		Silicone-molybdenum blue
ammonia	NH_4-N	μmol dm^{-3}	spectrophotometer	Indophenol blue
phosphate	PO_4-P	μmol dm^{-3}	spectrophotometer	Molybdenum-antimony-ascorbic acid
chlorophyll *a*	Chl *a*	mg dm^{-3}	10-AU Fluorometry (Turner Designs, USA)	Spectrophotometry
total phosphate	TP	μmol dm^{-3}	spectrophotometer	Potassium peroxodisulfate oxidation colorimetry

Table 2. Descriptive statistics of water quality parameters in Daya Bay

Parameters	Mean	Maximum	Minimum	Stand. dev.	Skewness	Kurtosis
temperature	23.68	25.40	22.81	0.67	1.34	1.88
pH	8.09	8.14	7.89	0.06	−2.66	5.91
salinity	32.81	33.46	31.75	0.43	−1.02	1.44
DO	7.31	8.08	7.01	0.33	1.09	0.39
BOD_5	1.43	1.94	0.99	0.31	0.00	−1.14
COD	0.88	1.30	0.52	0.26	−0.22	−1.16
Chl *a*	3.10	5.86	1.41	1.45	0.60	−0.69
NO_3-N	3.40	4.72	2.78	0.56	0.94	0.54
NO_2-N	0.26	0.43	0.14	0.10	0.24	−1.29
NH_4-N	2.55	3.58	1.95	0.47	0.87	0.24
TP	0.84	1.06	0.63	0.11	0.24	0.35
PO_4-P	0.10	0.15	0.05	0.03	0.37	−1.07
SiO_3-Si	22.93	30.64	19.62	3.17	1.29	0.92

denoted by RPCA, is investigated. RPCA is still effective even if there are a few anomalous observations and $n < p$ (Croux & Ruiz-Gazen 2005, Gong et al. 2005).

In this work, PCA and RPCA are applied to the matrix SA (the 48(12stations × 4seasons) × 13(variable) sample matrix) to obtain the loading matrix V_{pca} and score matrix T_{pca} for PCA, and the loading matrix V_{rpca} and score matrix T_{rpca} for RPCA. In classical PCA and RPCA, we used the annual mean matrix $D_{12\times 13(12\text{stations}\times 13\text{variables})}$ to evaluate the spatial pattern of water quality.

In this present study, cluster analysis, classical PCA and RPCA were employed on our dataset to identify the factors influencing the spatial distribution of water quality.

Data were auto-scaled to avoid misclassification due to wide differences in data dimensionality. The data were normalized with a mean and variance of zero and one, respectively. All the procedures were performed using MATLAB6.5 (Mathworks Inc., USA).

3. Results

3.1. Spatial characteristics for environmental factors

The horizontal distributions of nutrient concentrations in the water column are shown in Figure 2. The spatial distribution of the NO_3-N concentration shows that it increases from the eastern to the western part of DYB (Figure 2a). The spatial distribution of the NH_4-N concentration is similar to that of the NO_3-N concentration (Figure 2b).

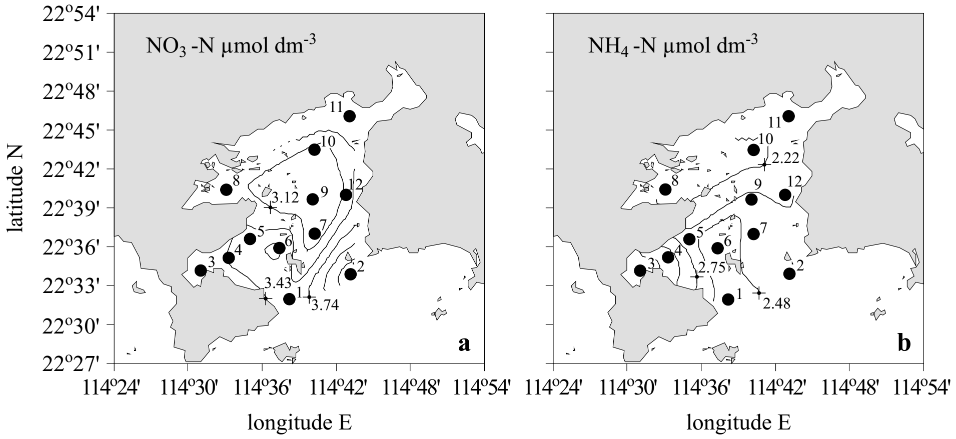

Figure 2. Spatial distributions of NO_3-N and NH_4-N

The PO_4-P concentration shows spatial variations, increasing from the northern region to the mouth of the bay, and again from the eastern to the western part of DYB (Figure 3a). The SiO_3-Si concentration displays spatial variation, decreasing from the northern region to the mouth of the bay (Figure 3b).

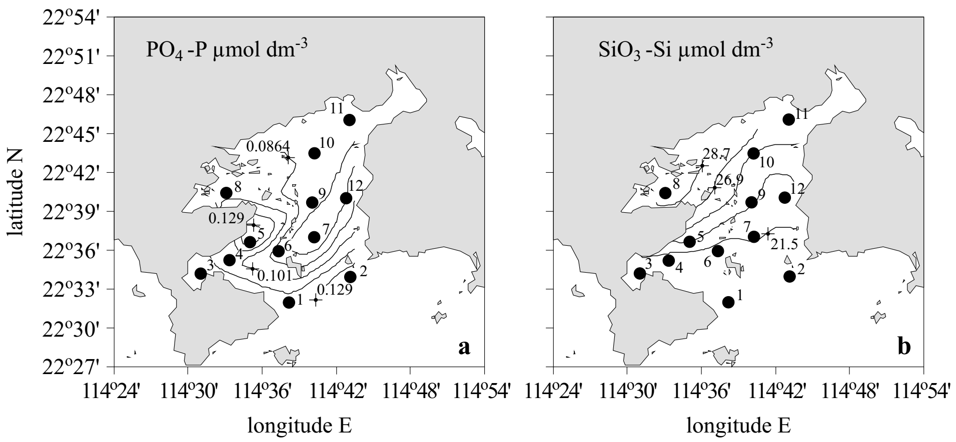

Figure 3. Spatial distributions of PO_4-P and SiO_3-Si

The horizontal distribution of COD concentration in the water column is shown in Figure 4a. This shows that the COD concentration increases from the eastern to the western part of the bay and from the mouth of the bay to its northern region. The horizontal distribution of BOD_5 concentration is similar to that of the COD concentration (Figure 4b).

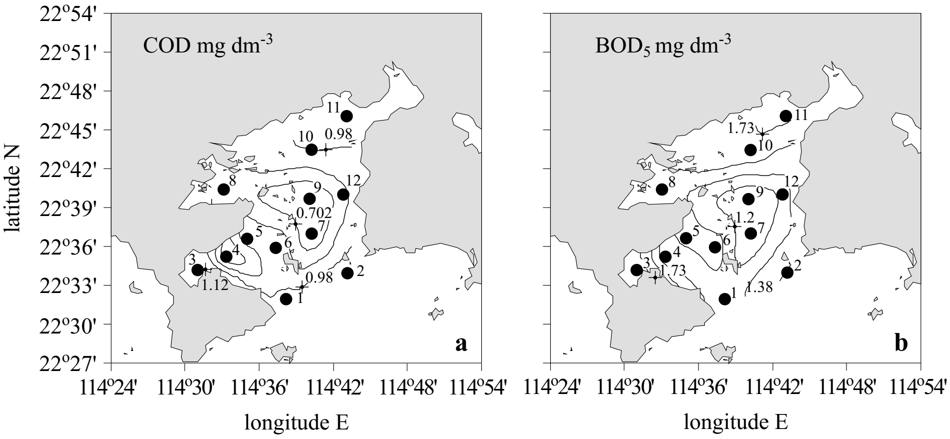

Figure 4. Spatial distributions of COD and BOD_5

3.2. Cluster analysis (CA)

The sampling stations were classified by the use of cluster analysis (the Euclidean distance as a similarity measure and Ward's method of linkage). The cophenetic correlation coefficient is 0.60. Cluster analysis yielded

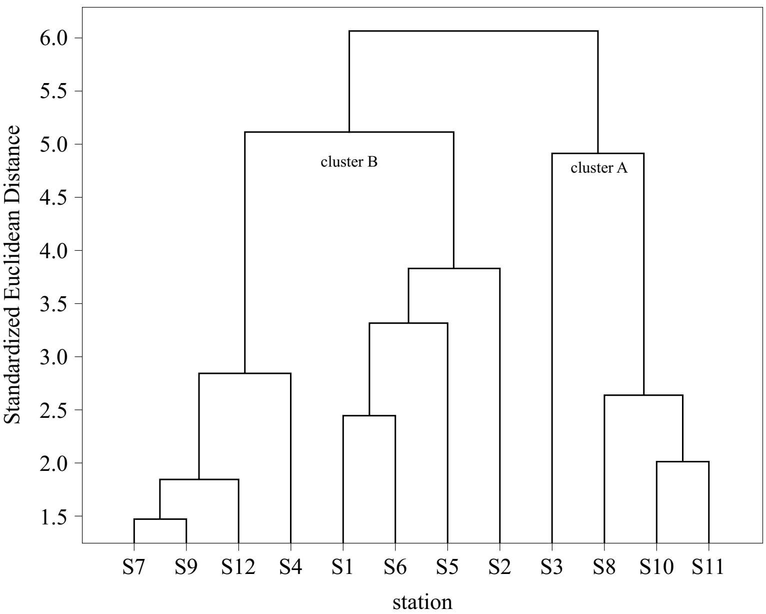

Figure 5. Dendrogram based on Ward's method clustering for 12 samples in Daya Bay in 2003

a dendrogram (Figure 5) in which all 12 sampling stations on the bay were grouped into two significant clusters. The clustering procedure generated two groups of stations in a robust way, as the stations in these groups had similar natural backgrounds and were probably affected by similar sources. Cluster A (S1, S2, S4 ~ S7, S9 and S12) and cluster B (S3, S8, S10 and S11) respectively correspond to regions of relatively low and high pollution. The stations in cluster A are located in the central, eastern and southern parts of DYB. The stations in cluster B are located in the western and northern coastal areas of DYB. S3 and S8 lie in the cage culture areas of Dapeng Cove and the north-western part near Aotou harbor, respectively. S10 is close to Xiachong and the shelf culture area in Xiaojing Wan. S11 was primarily impacted by FanHe (industrial wastewater, agricultural runoff, municipal sewage and aquaculture). A spatial view of the two areas defined in DYB by the cluster analysis, corresponding to clusters A and B, is shown in Figure 6.

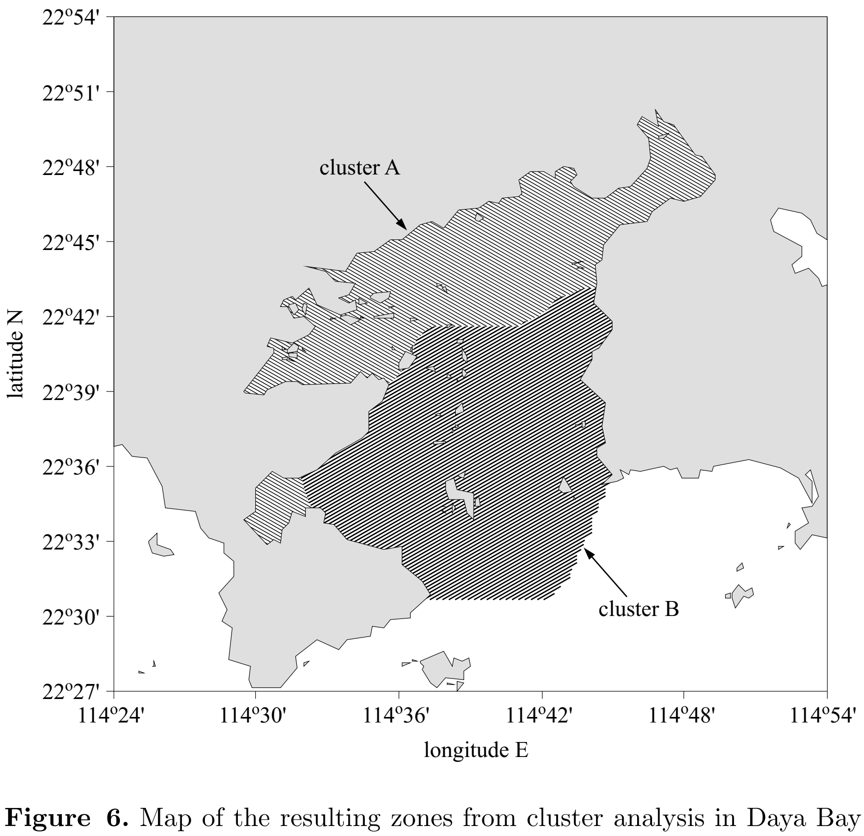

Figure 6. Map of the resulting zones from cluster analysis in Daya Bay in 2003 (Wang et al. 2006, 2008)

3.3. Classical PCA and RPCA

Bartlett's sphericity test carried out on the correlation matrix of variables show the calculated value $\chi^2 = 625.01$, which is greater than the critical value $\chi^2 = 124.34$ ($\alpha = 0.05$ and 78 degrees of freedom), thus indicating that the variables are not orthogonal but correlated; this therefore allows the data variability to be explained with a smaller number of variables than by principal component analysis. The loadings of the four retained PCs with classical PCA are shown in Table 3. PC1 (25.53% of the variance) is due mainly to T, S and NO_3-N. PC2 (21.64% of the variance) is characterized by DO and SiO_3-Si. PC3 explains 15.91% of the variance, due mainly to NO_2-N and NH_4-N, and is named the 'N-containing factor'. PC4 explains 10.50% of the variance; it is due mainly to parameters of organic pollution (DO, BOD_5 and COD) and reflects contributions from urban and industrial wastewater drainage.

Table 3. Loadings of 13 physical-chemical parameters on the the first four PCs in classical PCA

Parameters	PC1	PC2	PC3	PC4
temperature	0.37	−0.34	0.18	−0.21
pH	0.26	0.15	−0.20	0.29
salinity	−0.42	−0.16	−0.26	0.11
DO	−0.12	0.47	−0.01	0.37
BOD_5	0.31	−0.30	−0.18	0.47
COD	0.20	−0.34	−0.08	0.54
Chl *a*	0.34	0.31	0.23	0.11
NO_3-N	0.46	−0.01	0.16	−0.27
NO_2-N	0.20	−0.12	−0.54	−0.26
NH_4-N	0.14	0.14	−0.47	−0.19
TP	0.20	0.29	0.27	0.10
PO_4-P	0.15	0.26	−0.21	−0.11
SiO_3-Si	0.14	0.35	−0.34	0.02
eigenvalue	3.3190	2.8133	2.0686	1.3649
variance (%)	25.53	21.64	15.91	10.50
cumulative (%)	25.53	47.17	63.08	73.58

Figure 7 shows the PC1–PC2 scores and loading projection plots with classical PCA. The loadings of variables were zoomed in 1.8 times; it is intuitively clear that the loading of some variable contributes more to a station's score. The first two scores of the stations show that the marine aquaculture stations (S3 and S8) and the non-aquaculture stations (the other stations) are widely separated. S3 and S8 are located in Dapeng

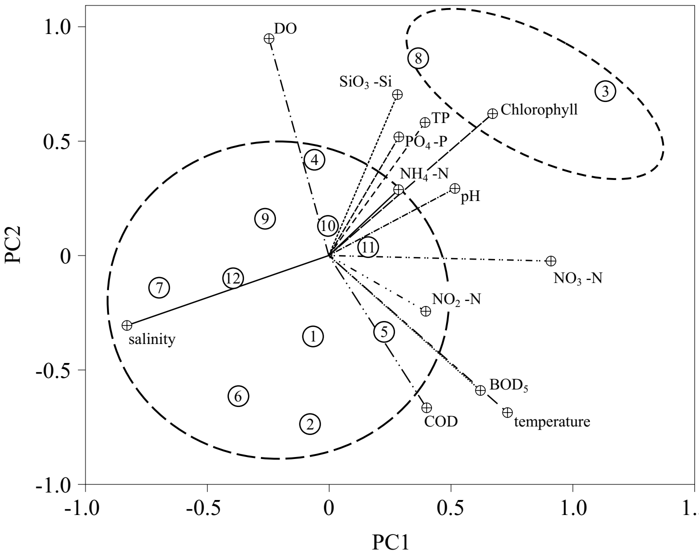

Figure 7. Loadings of variables and scores of twelve stations for the first two PCs, respectively

Ao and Aotou Bay, respectively. They are associated mainly with pH, chlorophyll and nutrients. The first chlorophyll loading contributed more to the score of S3 than the other factors did. The annual mean chlorophyll (5.86 μg dm^{-3}) is the highest at S3. The second silicate loading contributed more to the score of S8 than the other factors did. Annual mean silicate (30.64 μmol dm^{-3}) is higher at S8 than at the other stations. S5 is located in the outfall of high temperature effluent from DNPS; the annual mean temperature (25.40°) is higher at this station than at the others. The temperature loading is strongly positive and negative in the first and second PCs, respectively. The fifth station is predominantly related to temperature.

RPCA of the entire data set (Table 4) evolved three PCs with eigenvalues greater than one explaining about 76.78% of the total variance in the water-quality data set. The first PC accounting for 24.43% of the total variance was correlated (loading $>$ 0.40) with NH_4-N and SiO_3-Si. The second PC accounting for 19.06% of total variance was correlated with temperature, DO and SiO_3-Si. The third PC accounting for 13.65% of total variance was correlated with BOD_5, NO_2-N and NH_4-N. The fourth PC accounting for 11.23% of total variance was correlated with salinity and TP. The fifth PC accounting for 8.41% of total variance was correlated with COD and PO_4-P.

The matrix of scores provides information about the distribution of patterns or sources of contamination among samples. The matrix of loadings defines the contribution of the original variables to each one of these

Table 4. Loadings of 13 physical-chemical parameters on the first four PCs in RPCA

Parameters	PC1	PC2	PC3	PC4	PC5
temperature	−0.11	−0.47	−0.25	0.14	−0.36
pH	−0.16	0.12	−0.08	0.27	−0.11
salinity	0.22	0.15	0.03	0.46	0.10
DO	0.03	0.48	0.18	−0.11	0.07
BOD_5	−0.28	−0.10	−0.46	0.31	0.28
COD	−0.13	−0.13	−0.28	0.01	0.51
Chl *a*	−0.27	0.23	−0.32	−0.01	−0.34
NO_3-N	−0.30	−0.31	0.03	−0.25	−0.03
NO_2-N	−0.37	−0.29	0.45	0.30	0.16
NH_4-N	−0.40	−0.01	0.53	0.04	0.02
TP	−0.37	0.20	−0.16	−0.49	0.36
PO_4-P	−0.19	−0.03	−0.01	−0.30	−0.45
SiO_3-Si	−0.43	0.46	−0.06	0.32	−0.18
eigenvalue	3.0515	2.3807	1.7057	1.4028	1.0507
variance (%)	24.43	19.06	13.65	11.23	8.41
cumulative (%)	24.43	43.48	57.14	68.37	76.78

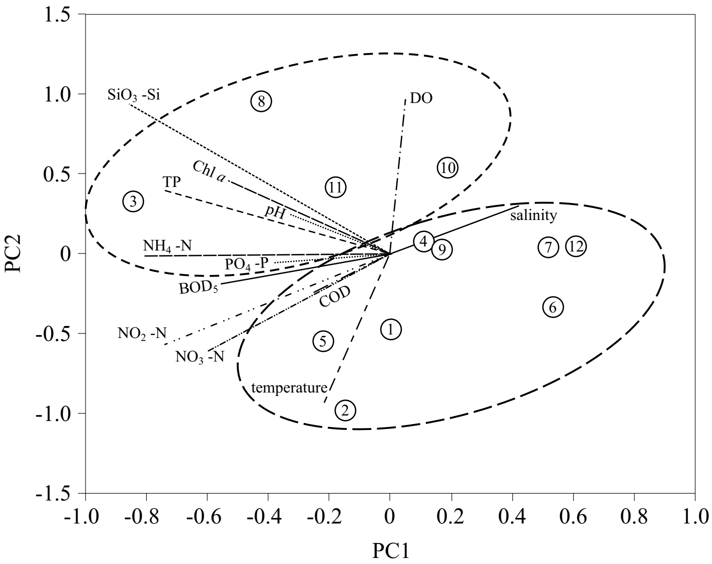

Figure 8. Loadings of variables zoomed in 5.0 times and scores of twelve stations for the first two PCs, respectively

contamination patterns or sources. The loadings of variables were zoomed in 5.0 times, which shows clearly that the loading of some variables contribute more to a station's score (Figure 8). The data are distributed in a limited region of space spanned by the two well-defined PC axes. The scores of stations S3, S8, S10 and S11 have negative and positive values in PC1 and PC2, respectively. These stations are located in the west and north of DYB. The scores of the other stations (S1, S2, S4-S7, S9 and S12) have negative values in PC2; these stations are located in the south and east of DYB.

4. Discussion

The results of CA, classical PCA and RPCA show that two clusters can describe marine aquaculture and other human activities taking place in this area (S3, S8, S10 and S11) and the non-aquaculture area (Figure 5). Therefore, it is clear that the scores of the stations are in relation to the importance of the variables.

In classical PCA, Chl *a* had higher positive loadings in PC2 than PC1, so stations loaded with a higher concentration of Chl *a* are distributed in the right upper quadrant. The loading of Chl *a* makes an important contribution to the scores of S3, S8 in PC1 and PC2. The scores of S10 and S11 are due mainly to nutrients and Chl *a* (see Figure 7). Similarly, in RPCA, Chl *a* had higher negative and positive loadings in PC1 and PC2, respectively, so the stations loaded with a higher concentration of Chl *a* are distributed in the left upper quadrant. The loading of Chl *a* makes an important contribution to the scores of S3, S8, S10 and S11 in PC1 and PC2, respectively (Figure 8).

Nutrient loadings make an important contribution to the scores of S3, S8, S10 and S11 in PC1. The scores of S10 and S11 are due principally to nutrients and Chl *a* (see Figures 7 and 8). The clockwise Euler Residual Current in spring, summer and autumn in DYB (Xu 1989) carries river-borne nutrients entering the west and north of the bay through this area; meanwhile, tidal currents carry nutrients from the South China Sea through the bay as well, which may be the reason for the higher concentration of Chl *a* there (Qiu et al. 2005). The PO_4-P concentration is higher in cluster A than in cluster B (Figure 3a). Domestic wastewaters, particularly those containing detergents, as well as industrial effluents and fertilizer run-off contribute to elevated levels of phosphates in the water column. Phosphate concentrations can indicate the presence of predominantly anthropogenic pollutants (Iscen et al. 2008). The NH_4-N and NO_3-N concentrations have similar spatial distributions that increase from the south and east to the west and north of DYB (Figure 2): this is due mainly to industrial effluents and sewage. The amount of NH_4-N from industrial wastewaters is 34 000 000 kg year^{-1}, and the total nitrogen and NH_4-N contents in domestic pollution are 807 600 and 646 800 kg year^{-1} (Zheng et al. 1998) in DYB, respectively. In DYB, expanding aquaculture activities are aggravating eutrophication in some parts of the Bay, and the phytoplankton biomass remains high in Dapeng Ao and Aotou Bay (Song et al. 2004). The silicate loading has greatly contributed to S8, terrestrial sources being among the main origins of silicate (Wang et al. 2004). Nutrients play an important role in the scores of the marine aquaculture area (S3 and S8). The scores of S10 and S11 are also due mainly to nutrients and Chl *a* (see Figures 7 and 8).

The BOD_5 and COD loadings make a strong contribution to the scores of S3, S8, S10 and S11 in classical PCA and RPCA. COD and BOD_5 are the traditional methods of obtaining information on bulk organic matter in water (Kotti et al. 2005), and both variables are considered important indicators of water quality. Both COD and BOD_5 concentrations have similar spatial distributions in DYB. The most polluted sites are S3, S8, S10 and S11 (higher COD and BOD_5 values). This was to be expected, since marine aquaculture and other activities take place in these areas. COD and BOD_5 loadings in PC1 in both classical PCA and RPCA are closely related to the scores of stations S3, S8, S10 and S11. This result suggests that the water quality at these stations is dominated by anthropogenic variables (COD, BOD_5). The concentrations of parameters related to anthropogenic pollution like BOD_5 and COD are higher in the west and north of DYB than in its southern and eastern parts (Figure 4). This result indicates that the contamination is derived mainly from municipal wastewater and fish-farming.

BOD_5 may also measure oxygen consumption by reduced forms of nitrogen (e.g., nitrite) (Kotti et al. 2005). Accordingly, in the nitrification process, NO_2-N is an unstable intermediary which tends to convert to NO_3-N. Nitrates and nitrites are considered jointly due to their conversion from one form to the other in the environment (Iscen et al. 2008). As high NO_3-N, NH_4-N and NO_2-N levels are associated with human activities, the water quality in the western and northern parts of DYB is strongly related to them.

The result of CA offers a reliable classification of water quality in the whole region and will make it possible to design a future spatial sampling strategy in an optimal manner. This reduces the number of sampling sites in the monitoring network and the cost of the risk assessment procedure (Simeonov et al. 2003, Singh et al. 2004). However, classical PCA and RPCA can extract specific information about relationships between variables and spatial variation trends in water quality. In comparison with classical PCA, the clustering seen from the PC1–PC2 score projection graphs with RPCA is better than that with classic PCA.

5. Conclusions

The result of RPCA is in good agreement with the CA results. CA yields good results as a first exploratory method to evaluate spatial differences, but it fails to show up the details of these differences (Singh et al. 2004). Classical PCA and RPCA can evaluate the incidence of each group in the overall change in water quality and support specific information on spatial differences in water quality. Water quality in DYB was divided into two groups by robust principal component analysis: S3, S8, S10 and S11 belong to cluster A, which lie in Dapeng Cove, Aotou Harbor and the north-eastern parts of DYB, where water quality is related mainly to anthropogenic activities. The other stations belong to cluster B, which lie in the southern, central and eastern parts of DYB, where the quality is related largely to water exchange from the South China Sea. Robust principal component analysis as an important tool for information extraction presents a novel approach for understanding a complex data matrix, even though the situation with more variables than observations ($n < p$) is frequently encountered in practical applications. Classical PCA and RPCA can extract specific information about relationships between variables and spatial variation trends in water quality.

References

Bowen R. E., Depledge M. H., 2006, *Rapid assessment of marine pollution (RAMP)*, Mar. Pollut. Bull., 53 (10–12), 631–639.

Chau K. W., Muttil N., 2007, *Data mining and multivariate statistical analysis for ecological system in coastal waters*, J. Hydroinform., 9 (4), 305–317.

Croux C., Ruiz-Gazen A., 2005, *High breakdown estimators for principal components: the projection-pursuit approach revisited*, J. Multivar. Anal., 95 (1), 206–226.

Gong F., Wang B. T., Fung Y. S., Chau F. T., 2005, *Chemometric characterization of the quality of the atmospheric environment in Hong Kong*, Atmos. Environ., 39 (34), 6388–6397.

Iscen C. F., Emiroglu Ö., Ilhan S., Arslan N., Yilmaz V., Ahiska S., 2008, *Application of multivariate statistical techniques in the assessment of surface water quality in Uluabat Lake, Turkey*, Environ. Monit. Assess., 144 (1–3), 269–276.

Kotti M. E., Vlessidis A. G., Thanasoulias N. C., Evmiridis N. P., 2005, *Assessment of river water quality in Northwestern Greece*, Water Resour. Manag., 19 (1), 77–94.

Kuppusamy M. R., Giridhar V. V., 2006, *Factor analysis of water quality characteristics including trace metal speciation in the coastal environmental system of Chennai Ennore*, Environ. Int., 32 (2), 174–179.

Qiu Y. W., Wang Z. D., Zhu L. S., 2005, *Variation trend of nutrient and chlorophyll contents and their effects on ecological environment in Daya Bay*, J. Oceanogr.-Taiwan Strait, 24 (2), 131–139, (in Chinese).

Shrestha S., Kazama F., 2007, *Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan*, Environ. Modell. Softw., 22 (4), 464–475.

Simeonov V., Stratis J. A., Samara C., Zachariadis G., Voutsa D., Anthemidis A., Sofoniou M., Kouimtzis Th., 2003, *Assessment of the surface water quality in Northern Greece*, Water Res., 37 (17), 4119–4124.

Simeonova P., Simeonov V., Andreev G., 2003, *Water quality study of the Struma river basin, Bulgaria (1989–1998)*, Centr. Eur. J. Chem., 1 (2), 121–136.

Singh K. P., Malik A., Mohan D., Sinha S., 2004, *Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India) – a case study*, Water Res., 38 (18), 3980–3992.

Song X. Y., Huang L. M., Zhang J. L., Huang X. P., Zhang J. B., Yin J. Q., 2004, *Variation of phytoplankton biomass and primary production in Daya Bay during spring and summer*, Mar. Pollut. Bull., 49 (11–12), 1036–1044.

Stanimirova I., Walczak B., Massart D. L., Simeonov V., 2004, *A comparison between two robust PCA algorithms*, Chemometr. Intell. Lab., 71 (1), 83–95.

Suikkanen S., Laamanen M., Huttunen M., 2007, *Long-term changes in summer phytoplankton communities of the open northern Baltic Sea*, Estuar. Coast. Shelf Sci., 71 (3–4), 580–592.

Vega M., Pardo R., Barrado E., Deban L., 1998, *Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis*, Water Res., 32 (12), 3581–3592.

Wang Y. S., Lou Z. P., Sun C. C., Sun S., 2008, *Ecological environment changes in Daya Bay, China, from 1982 to 2004*, Mar. Pollut. Bull., 56 (11), 1871–1879.

Wang Y. S., Lou Z. P., Sun C. C., Wu M. L., Han S. H., 2006, *Multivariate statistical analysis of water quality and phytoplankton characteristics in Daya Bay, China, from 1999 to 2002*, Oceanologia, 48 (2), 193–211.

Wang C. H., Qi Y. Z., Li J. T., Xu N., Chen J. F., 2004, *Analysis and evaluation of trophic status in aquaculture areas of Daya Bay*, Mar. Environ. Sci., 23 (2), 25–28, (in Chinese).

Wu M. L., Wang Y. S., 2007, *Using chemometrics to evaluate anthropogenic effects in Daya Bay, China*, Estuar. Coast. Shelf Sci., 72 (4), 732–742.

Xu G. Z., 1989, *Environments and resources of Daya Bay*, Anhui Sci. Publ., HeFei, China, 1–28, (in Chinese).

Zheng Q. H., He Y. Q., Zhang G. X., 1998, *Impact on the changes of chemical composition of seawater from waste water discharged in Daya Bay*, [in:] *Annual research reports: Marine Biology research station at Daya Bay (II)*, J. Pen & Z. Wang (eds.), Sci. Publ., Beijing, China, 102–112, (in Chinese).

Zhou F., Guo H. C., Hao Z. J., 2007, *Spatial distribution of heavy metals in Hong Kong's marine sediments and their human impacts: A GIS-based chemometric approach*, Mar. Pollut. Bull., 54 (9), 1372–1384.