

## Original Article

# Application of chemometric methods for assessment and modelling of microbiological quality data concerning coastal bathing water in Greece

Agelos Papaioannou,<sup>1</sup> George Rigas,<sup>2</sup> Panagiotis Papastergiou,<sup>3</sup> Christos Hadjichristodoulou<sup>4</sup>

<sup>1</sup>Department of Medical Laboratories and <sup>2</sup>Department of Animal Production, Technological Education Institute of Thessaly, Greece; <sup>3</sup>NHS Trust Microbiology Department, United Lincolnshire Hospitals, Lincoln County Hospital, United Kingdom; <sup>4</sup>Department of Hygiene and Epidemiology, School of Medicine, University of Thessaly, Greece

## Significance for public health

The microbiological protection of coastal bathing water quality is of great interest for the public health authorities as well as for the economy. The present study proves that this protection can be achieved by monitoring only two microbiological parameters, *E. coli* and faecal streptococci/enterococci instead four microbiological parameters (the two mentioned above plus Total coliforms and Faecal coliforms) that are usually monitored today. As a consequence, countries, especially those with large quantities of coastal bathing sites, can perform microbiological monitoring of their bathing waters by checking only the mentioned two parameters, thus ensuring economies of scale. Thus, funds can be used in other actions to preserve the quality of coastal water and human health. This in turn, would aid in the assessment of the quality of coastal bathing waters and provide a more timely indication of bathing water quality, hence contributing to the immediate health protection of bathers.

## Abstract

**Background.** Worldwide, the aim of managing water is to safeguard human health whilst maintaining sustainable aquatic and associated terrestrial, ecosystems. Because human enteric viruses are the most likely pathogens responsible for waterborne diseases from recreational water use, but detection methods are complex and costly for routine monitoring, it is of great interest to determine the quality of coastal bathing water with a minimum cost and maximum safety.

**Design and methods.** This study handles the assessment and modelling of the microbiological quality data of 2149 seawater bathing areas in Greece over 10-year period (1997-2006) by chemometric methods.

**Results.** Cluster analysis results indicated that the studied bathing beaches are classified in accordance with the seasonality in three groups. Factor analysis was applied to investigate possible determining factors in the groups resulted from the cluster analysis, and also two new parameters were created in each group; VF1 includes *E. coli*, faecal coliforms and total coliforms and VF2 includes faecal streptococci/enterococci. By applying the cluster analysis in each seasonal group, three new groups of coasts were generated, group A (ultra-clean), group B (clean) and group C (contaminated).

**Conclusions.** The above analysis is confirmed by the application of discriminant analysis, and proves that chemometric methods are useful tools for assessment and modeling microbiological quality data of coastal bathing water on a large scale, and thus could attribute to effective and economical monitoring of the quality of coastal bathing water in a country with a big number of bathing coasts, like Greece.

## Introduction

In 2010, EU Member States reported 21.063 bathing waters, of which 70% are coastal bathing waters. The majority of coastal bathing waters are located at Mediterranean Sea coasts (about 9900), representing almost two thirds of all reported coastal bathing waters in Europe. A total of 24 countries reported coastal bathing waters. Italy (4896), Greece (2149), France (2012), Spain (1930), Denmark (1054), Croatia (913) and the United Kingdom (596) have the highest number of coastal bathing waters.

The quality of bathing waters is of great importance for several reasons. Contaminated (unclean) water is a major hazard to bathers (causing gastric and skin problems). For the tourist industry, clean and safe water is also a major factor in attracting visitors to an area. Based on risk assessments from the World Health Organization (WHO) and academic research sources, studies suggest that millions of gastrointestinal and severe respiratory diseases are caused by swimming and bathing in wastewater-polluted coastal waters.<sup>1-11</sup>

No indicator has proven perfect for controlling coastal water quality. Indicator bacteria including total coliforms, faecal coliforms, *Escherichia coli* and streptococci/enterococci have been used over time for the assessment of water quality and risk assessment in the prediction of water microbial pollution. Research supports use of *E. coli* and enterococci rather than the broader group of faecal coliforms as indicators of microbiological pollution. Besides their limitations, these indicator bacteria have been used successfully in many countries as a monitoring tool for microbiological contamination of water and prediction of the presence of pathogens.<sup>12-23</sup>

The aim of the present study is to analyse a large number of numerical data that concern measurements of four microbiological quality indicators of seawater (total coliforms, faecal coliforms, *E. coli* and faecal streptococci/enterococci) over a 10-years period using chemometric methods such as cluster analysis, factor analysis and discriminant analysis for the assessment and modelling of these data. The extraction of successful models is of great importance for effective monitoring of coastal bathing water, allowing economies of scale without compromising the health of swimmers.

## Design and methods

### Study areas and sampling programs

Beaches were sampled on a regular basis with an average of 13

samples collected from each beach per year, from predetermined points specified by the competent department of the Hellenic Ministry for the Environment, Physical Planning and Public Works.

### Sample collection and testing

Water samples from regularly monitored beaches were taken from areas of beaches which are most frequently used by bathers. The beaches are mainly visited by bathers during June to September with the highest counts of visitors noted during July and August. During the other months there is minimal to zero visiting of the seawater bathing areas for bathing purposes. Consequently, between the months of October and April only a few water samples were collected from the coastal bathing areas. Water samples were therefore collected from May through to November. The time of sampling was almost the same for each particular beach every time. The majority of samples were taken between 10.30 (a.m.) and 17.30 (p.m.) as this was considered to be the time at which the majority of people engaged in water activities.

A volume of 450 mL of water was collected in sterile bottles of 500 mL capacity. Samples were taken 20-30 cm below the water surface level at locations with a sea depth of 0.8-1.3 meters. Samples were transferred to the laboratory on the same day of collection in a closed Esky cooler, thereby avoiding any disinfecting effect of sunlight and changes to microbial presence. All samples were processed within 24 hours after collection. The microbiological variables of the regular monitored bathing areas can be seen in Table 1.

The majority of the water samples were collected and analysed by a contracted main private laboratory. Due to the vast number of samples, over 40 public and private authorities were involved in the sampling operation and 11 public and private laboratories including the main contracted laboratory were involved in the testing of the samples. All laboratories processed samples for microbiological analysis in accordance with standard ISO methods for the detection and enumeration of *E. coli* (ECOL), faecal coliforms (FCOL), total coliforms (TCOL) and faecal streptococci/enterococci (STREPT).

### Data collection and validation

Data included in the study were gathered from the archives of the Hellenic Ministry for the Environment, Physical Planning and Public Works and comprised microbiological test results and relevant information recorded during sampling of the regularly monitored coastal bathing areas. All data entries were subjected to data validation and any inaccuracies found in the database, due to data entry errors, were cross checked with result transcripts and corrected.

### Statistical analysis

This study handles the assessment and modelling of the microbiological quality data of 2.149 seawater bathing areas in Greece over 10-year period (1997-2006) by chemometric methods.

Data consisting of the microbiological test results (four microbiological indicators per water sample: TCOL, FCOL, STREPT and *E. coli*) collected from the coastal bathing areas were built into a database.

### Parameters distribution characteristics and data treatment

Most methods such as cluster analysis (CA) and factor analysis (FA) require variables to be at least *column centering* and some of them as discriminant analysis (DA) require variables to conform to a normal distribution.

More specific, for temporal and parameter CA as also for FA, *column standardization* was performed. All parameters were also z-scale standardized (mean=0; variance=1) to minimize the effects of differences in measurement units and variance and render the data dimension-

less. Consequently, each column had a mean of 0 and variance of 1. For spatial CA and DA in each temporal cluster, log transformation and column standardization were performed, thus each column had a mean of 0 and variance of 1. All the calculations and plots in the following sections were done with the SPSS 15.0.

### Cluster analysis

Hierarchical CA, being the most common approach of CA, starts with each case in a separate cluster and joins the clusters together step by step until only one cluster remains.<sup>25-27</sup> In this study, hierarchical CA was performed on the standardized data using Ward's method with squared Euclidean distances as a measure of similarity.<sup>27-34</sup>

### Discriminant analysis

DA constructs a discriminant function for each group of two or more naturally occurring groups as follows:<sup>24,33</sup>

$$f(G_i) = k_i + \sum_{j=1}^n (w_{ij} \times p_{ij})$$

where,  $i$  is the number of groups (G);  $k_i$  is a constant inherent to each group;  $n$  is the number of parameters used to classify a set of data into a given group; and  $w_{ij}$  is the weight coefficient assigned by discriminant analysis to a given parameter ( $p_{ij}$ ).

In this study DA was performed on standardized log-transformed data using the standard, forward stepwise and backward stepwise modes to evaluate both the temporal and spatial variations in water quality. The best discriminant functions for each mode were constructed considering the quality of the classification matrix and the number of parameters. The monitoring sites and periods were the grouping variables and the measured parameters were the independent variables.

### Factor analysis

Although not commonly used in water quality analysis, several studies have used FA to identify primary sources of contamination. FA is also used to find associations between parameters so that the number of measured parameters can be reduced. Known associations are then used to predict unmeasured water quality parameters.<sup>35-38</sup>

In this study, the factor extraction was performed using the method of principal components through varimax rotation. Screeplot criterion was used for determining how many factors to use and how many to ignore, and in our study retains only those factors with eigenvalues more than 0.75. Factor rotation (varimax rotation method) was used to facilitate interpretation by providing simpler factor structure.

## Results and discussion

### Temporal similarity and period grouping (temporal cluster analysis)

Hierarchical CA (single linkage method of linkage, Euclidean distances as similarity measure, standardization of the input data) was used to study the temporal relationships. Temporal CA generated a dendrogram grouping the 6 months into three clusters at (Dlink/Dmax)  $\times 100 < 35$  and the difference between the clusters was significant (Figure 1). As can be seen in Figure 1, the studied period is separated into three clusters as follows: Cluster 1: May; Cluster 2: Jun, Jul and Aug; Cluster 3: Sept and Oct.

The careful consideration of the content of the clusters (Figure 1) offers some interesting conclusions about the data classification. The

first cluster (1<sup>st</sup> period) included May, cluster 2 (2<sup>nd</sup> period) comprised June-August and cluster 3 (3<sup>rd</sup> period) consisted of the two remaining months (September-October).

Therefore, specific patterns of the classified parameters could be offered: 1<sup>st</sup> period or Spring period pattern (May); 2<sup>nd</sup> period or Summer period pattern (June-August); 3<sup>rd</sup> period or Autumn period pattern (September and October).

Hence, the temporal variation in coastal bathing water quality was absolutely determined by local climate (in spring, summer and autumn) or hydrological conditions (dry and wet seasons) because the costal sea water quality was also related to pollution characteristics (such as discharge frequency and type).

### Spatial similarity and site grouping (parameter cluster analysis)

Hierarchical CA on standardized data was also applied to reveal natural groupings (clusters) within the data set of the four microbiological parameters and to examine relationships between them. Parameter CA (Ward's method of linkage, Squared Euclidean distance as similarity measure, standardization of the input data), was conducted for each temporal cluster (first, second and third period) and for the All samples data set, with the same results.

Parameter CA generated a dendrogram for each temporal cluster and for the All samples data set, grouping the four parameters into two clusters at  $(Dlink/Dmax) \times 100 < 35$  and the difference between the clusters was significant. In Figure 2 the hierarchical dendrogram for the clustering of the determined microbiological parameters for all the studied costal sea areas of the 1<sup>st</sup> period is plotted (the other three dendrograms for the temporal clusters are similar).

More specific, from the dendrogram of Figure 2 it could be concluded that the parameters at each temporal cluster and the All samples group are separated into two similar clusters as follows: Cluster 1 (three parameters are included): ECOL, FCOL, TCOL; Cluster 2 (one parameter is included): STREPT.

Therefore, specific patterns of the classified microbiological parameters could be offered: *Coliforms* pattern (including ECOL, FCOL and TCOL); *Streptococci* pattern (including STREPT).

### Factor analysis

Usually, the typical classification approach of clustering is accompanied by FA, which is a typical projection and modelling approach. FA was applied to standardized datasets (4 microbiological parameters) to examine differences between the three studied periods and the All samples group and moreover, to identify the latent factors.

Before conducting the FA, the Kaiser-Meyer-Olkin (KMO) and Bartlett's sphericity tests were performed on the parameter correlation matrix to examine the validity of the FA. The KMO results for the three temporal clusters and All samples were 0.694; 0.703; 0.719 and 0.710 respectively, and those for Bartlett's sphericity were 9.036; 1.2263; 8.175 and 27.765 ( $P < 0.05$ ), respectively, indicating that FA may be useful in providing significant reductions in dimensionality.

Based on the screeplot criterion, only the VFs with eigenvalues over 0.75 were considered significant.

FA yielded two VariFactors (VFs) for each of the above data sets (Spring, Summer, and Autumn period and All samples group), explaining 92.14%; 97.40%; 94.50% and 94.90% of the total variance, respectively.

Table 1 summarized the FA results comprising the loadings, eigenvalues, percentage of total variance (the loadings which absolute value was more than 0.7 of the total variance was significant).

In general, FA confirms the results obtained by CA. From Table 1 it is seen, that the pollution structure of the three periods was similar to that of the All samples group and almost the same in the degree of pol-

lution.

The linkage between the microbiological parameters in the four groups is shaped as follows (Table 1).

*First period:* VF1, which explained 72.73% of the total variance (TV), shows how high  $[0.8 \leq \text{Loading (L)}]$  ECOL, FCOL and TCOL coincide with costal bathing areas having low STREPT ( $L=0.21$ ) (this VF could be called *Coliforms* factor). Additionally, VF2 (19.41% of the TV) shows that high STREPT ( $L=0.98$ ) is met in costal sea having low ECOL ( $L=0.15$ ), FCOL ( $L=0.18$ ) and TCOL ( $L=0.28$ ) (*Streptococci* factor).

*Second period:* VF1, which explained 75.85% of the TV, shows how high ECOL, FCOL and TCOL coincide with costal seas having low STREPT ( $L=0.16$ ) (*Coliforms* factor) and VF2 (21.55% of the TV) shows that high STREPT ( $L=0.99$ ) is met in costal bathing areas having low ECOL ( $L=0.10$ ), FCOL ( $L=0.13$ ) and TCOL ( $L=0.29$ ) (*Streptococci* factor).

*Third period:* VF1, which explained 74.17% of the TV, shows how high ECOL, FCOL and TCOL coincide with costal seas having low STREPT ( $L=0.19$ ) (*Coliforms* factor) and VF2 (20.33% of the TV) shows that high STREPT ( $L=0.98$ ) is met in costal bathing areas having low ECOL ( $L=0.12$ ), FCOL ( $L=0.15$ ) and TCOL ( $L=0.34$ ) (*Streptococci* factor).

*All samples:* VF1, which explained 74.15% of the TV, shows how high ECOL, FCOL and TCOL coincide with costal seas having low STREPT ( $L=0.18$ ) (*Coliforms* factor) and VF2 (20.75% of the TV) shows that high STREPT ( $L=0.98$ ) is met in costal bathing areas having low ECOL ( $L=0.12$ ), FCOL ( $L=0.15$ ) and TCOL ( $L=0.39$ ) (*Streptococci* factor).

It is easily seen that the major groups of microbiological parameters interpreted by parameter CA for the studied groups (Figure 2) are also involved in the VFs loadings presented in Table 1. Thus, the classification scheme obtained by parameter CA is confirmed by FA. This confirmation is an important hint that the microbiological parameters tested are indeed related and form groups of similar indicative properties.

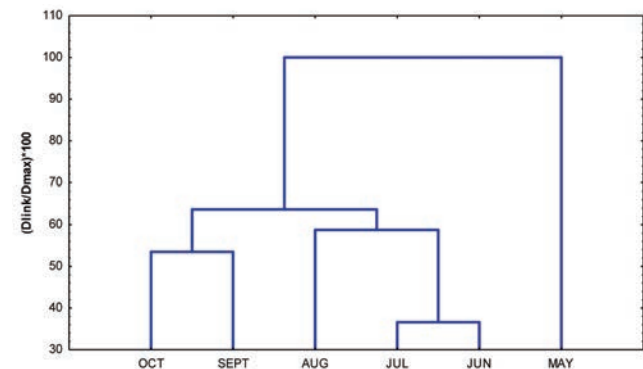


Figure 1. Dendrogram showing temporal clustering of monitoring periods (JUN = June, JUL = July, AUG = August, SEPT = September and OCT = October).

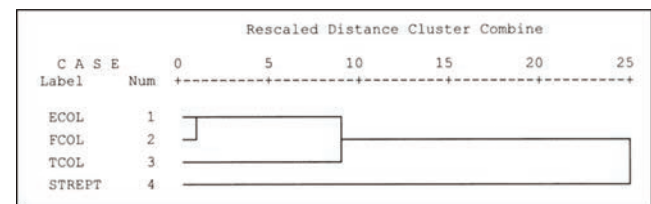


Figure 2. Dendrogram for the microbiological parameters in all the studied costal bathing areas of the 1st period.

The degree of pollution in the monitoring sites for 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> period, according to the sp-cluster A, B and C defined by the 1<sup>st</sup> period, was as follows:

**First period:** The degree of pollution (as the average value of VFs) in the monitoring sites for 1<sup>st</sup> period differed significantly among the three sp-clusters, and sites received more pollution from VF1 (ECOL, TCOL and FCOL) and VF2 (STREPT) in the sp-cluster C<sub>1</sub> than in the sp-cluster A<sub>1</sub> and B<sub>1</sub>, and more in the sp-cluster B<sub>1</sub> than in the A<sub>1</sub>. Cases distributed in the region of larger values of VF1 were almost all collected from sp-cluster B<sub>1</sub> and C<sub>1</sub>.

**Second period:** The degree of pollution (as the average value of VFs) in the monitoring sites for 2<sup>nd</sup> period differed significantly among the three sp-clusters, and sites received more pollution from VF1 in the sp-cluster C<sub>1</sub> than in the sp-cluster A<sub>1</sub> and B<sub>1</sub>, and more in the sp-cluster B<sub>1</sub> than in the A<sub>1</sub>. Moreover, sites received more pollution from VF2 in the sp-cluster C<sub>1</sub> than in the sp-cluster A<sub>1</sub> and B<sub>1</sub>, and more in the sp-cluster A<sub>1</sub> than in the B<sub>1</sub>. Cases distributed in the region of larger values of VF1 were almost all collected from sp-cluster B<sub>1</sub> and C<sub>1</sub>.

**Third period:** The degree of pollution (as the average value of VFs) in the monitoring sites for 3<sup>rd</sup> period differed significantly among the three sp-clusters, and sites received more pollution from VF1 and VF2 in the sp-cluster C than in the sp-cluster A<sub>1</sub> and B<sub>1</sub>, and more in the sp-cluster B<sub>1</sub> than in the A<sub>1</sub>. Cases distributed in the region of larger values of VF1 were almost all collected from sp-cluster B<sub>1</sub> and C<sub>1</sub>.

The degree of pollution in the monitoring sites for sp-clusters A, B and C defined by the 1<sup>st</sup> period, according to the three periods (1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup>), was as follows:

**Sp-cluster A defined by the 1<sup>st</sup> period [or Group A (A<sub>1</sub>):** The degree of pollution (as the average value of VFs) in the monitoring sites for group A differed significantly among the three periods, and sites received more pollution from VF1 (ECOL, FCOL and TCOL) and VF2 (STREPT) during the second (June-August) and the third (September-October) periods than in the first period (May). The factor scores for the three periods were not significantly regular or distinct. Cases distributed in the region of larger values of VF1 were almost all collected from second and third period.

**Sp-cluster B defined by the 1<sup>st</sup> period [or Group B (B<sub>1</sub>):** The degree of pollution (as the average value of VFs) in the monitoring sites for group B differed significantly among the three periods, and sites received more pollution from VF1 during the second and the third periods than in the first period. Moreover, sites received more pollution from VF2 during the first period than in the second and in the third periods, and more in the third period than in the second. The factor scores for the three periods were not significantly regular or distinct. Cases distributed in the region of larger values of VF1 were almost all collected from second and third period.

**Sp-cluster C defined by the 1<sup>st</sup> period [or Group C (C<sub>1</sub>):** The degree of pollution (as the mean average value of VFs) in the monitoring sites

for group C differed significantly among the three periods, and sites received more pollution from VF1 during the first period than in the second and the third periods. Moreover, sites received more pollution from VF2 during the first period than in the second period and in the third period received not significantly different pollution as the other two. Cases distributed in the region of larger values of VF1 were collected from all three periods but mostly from 1<sup>st</sup> and 2<sup>nd</sup> period.

### Spatial cluster analysis (on temporal clusters)

Spatial CA was conducted on standardized log-transformed data for each temporal cluster (1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> period). Before carrying out Spatial CA on the data sets of the three temporal clusters, the following data subsets were created: the costal sea sites, of which the value of at least one of the four microbiological parameters was greater than the permitted limit (according the EU Directive 2006/7/EC) was subtracted from each temporal cluster.

Spatial CA produced a dendrogram with two spatial clusters for each temporal cluster (or two spatial clusters were created for each temporal cluster). The above subtracted costal sea sites of each temporal cluster created the third cluster in each period.

Table 2 shows the statistical descriptive of the four microbiological parameters per studied temporal cluster (period) and per spatial cluster of each period.

Next, Table 3 shows the statistical descriptive of the four microbiological parameters per studied temporal cluster (period) and per spatial cluster of the 1<sup>st</sup> period.

Consequently, spatial CA identified similar monitoring sites considering the effects of temporal differences in spatial CA.

### Identification of the pollution pattern in costal bathing areas in the three different periods

As we see above, temporal CA generated a dendrogram grouping the six months into three clusters at (Dlink/Dmax) × 100 < 35, and the difference between the clusters was significant (Figure 2). Cluster 1 (the first period) comprised May, cluster 2 (the second period) included June-August and cluster 3 (the third period) consisted of the two remaining months (September-October). Spatial CA identified similar monitoring sites considering the effects of temporal differences in spatial CA.

Spatial similarity analysis was conducted for each temporal cluster (first, second, and third period) with different results. Therefore, in continuance the three spatial clusters of the first time period were used for the clustering of the second and third period. In this way the All samples group was achieved to be clustered according the clusters of the 1<sup>st</sup> period (A, B and C). Group A comprised 276, group B contained 1546 and group C included 96 monitoring sites (costal bathing areas).

FA (PCA method) was carried out for the source identification in the

**Table 1. Loadings, eigenvalues, and percentage of total variance (TV) of the measured parameters on significant VFs of 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> period, and all samples.**

Parameter	First period		Second period		Third period		All samples	
	VF1	VF2	VF1	VF2	VF1	VF2	VF1	VF2
ECOL	0.962	0.147	0.985	0.099	0.968	0.124	0.971	0.120
FCOL	0.964	0.180	0.984	0.134	0.972	0.152	0.976	0.146
TCOL	0.837	0.283	0.922	0.289	0.867	0.340	0.886	0.289
STREPT	0.205	0.976	0.163	0.985	0.186	0.979	0.178	0.982
Eigenvalue	2.909	0.776	3.034	0.862	2.967	0.813	2.966	0.830
% TV	72.728	19.412	75.851	21.550	74.172	20.329	74.152	20.753

VF, varifactors; ECOL, E. coli; FCOL, faecal coliforms; TCOL, total coliforms; STREPT, faecal streptococci/enterococci.

monitoring sites. Before conducting the FA, the Kaiser-Meyer-Olkin (KMO) and Bartlett's sphericity tests were performed on the parameter correlation matrix to examine the validity of the PCA. The KMO results for groups A, B, C and All samples were 0.598; 0.700; 0.664 and 0.710 respectively, and those for Bartlett's sphericity were 2.561,70; 20.183,60; 1.612,22 and 27.764,82 ( $P < 0.05$ ), indicating that PCA may be useful in providing significant reductions in dimensionality.

FA was applied to standardized data sets (4 parameters) to examine differences between groups A, B and C and identify the latent factors. Based on the screeplot for the FA and the eigenvalues-0.75 criterion, only the VFs with eigenvalues over 0.75 were considered essential.

Table 4 summarizes the FA results comprising the loadings, eigenvalues and percentages of total variance and the loadings of which the absolute value is more than 0.7 was highlighted.

FA of the four data sets yielded two VFs for the groups A, B, C and All samples, explaining 87.02%, 94.10, 96.03% and 94.90% of the total variance in the respective costal bathing areas water quality data sets.

According to the descriptive of the microbiological parameters and FA results (pollution sources) the 276 monitoring sites of group A corresponded to cleaner costal bathing areas, the 1546 sites of group B corresponded to relatively cleaner costal bathing areas and the 96 sites in group C corresponded to relatively polluted costal bathing areas.

Considering the last results of FA, the degree of pollution in the monitoring sites for groups A, B and C was as follows:

**Group A:** The degree of pollution (as the mean average value of VFs) in the monitoring sites for group A differed significantly among the three periods, and sites received more pollution from VF1 (ECOL and FCOL) and VF2 (TCOL and STREPT) during the second (June–August) and the third (September–October) periods than in the first period (May).

**Group B:** The degree of pollution (as the mean average value of VFs) in the monitoring sites for group B differed significantly among the three periods, and sites received more pollution from VF1 (ECOL, TCOL and FCOL) during the second and the third periods than in the first period. Moreover, sites received more pollution from VF2 (STREPT) during the first period than in the second and in the third periods, and more in the third period than in the second.

**Group C:** The degree of pollution (as the mean average value of VFs) in the monitoring sites for group C differed significantly among the three periods, and sites received more pollution from VF1 (ECOL, FCOL and TCOL) during the first period than in the third period and during the second period received not significant different pollution than each of the other two separately. Moreover, sites received more pollution from VF2 (STREPT) during the first period than in the second and third periods.

### Discriminant analysis: spatial variations in costal seawater quality

based on spatial CA. The objectives of the DA were to test the significance of discriminant functions and determine the most significant parameters associated with the differences among clusters.

The spatial DA was performed using the standardized log-transformed data of the four parameters after classification into the three major clusters (first, second and third period) obtained from the spatial CA. Clusters formed the dependent categorical, and the measured parameters were the independent variables. Wilks' lambda and the Chi-square for the discriminant functions (DFs), obtained from the standard and stepwise modes of DA for the three periods (first, second and third period), ranged from 0.337 to 0.972, and from 55.108 to 2078.428, the two methods, respectively, at  $P < 0.0001$ , suggesting that the spatial DA was credible and effective.

Using the stepwise mode, the same results were received, which means that all four microbiological parameters are significant for the determination of the differences among the clusters. The standard and

**Table 2. Statistical descriptive [mean, standard error (SE), standard deviation (SD) and CI-95% for mean] of the four microbiological parameters per studied temporal and spatial cluster.**

Parameter, spatial cluster	N.	Mean	SD	SE	95% CI for mean	
<b>First period</b>						
<b>ECOL</b>						
A1	276	8.53	3.60	0.22	8.11	8.96
B1	1546	33.73	12.36	0.31	33.12	34.35
C1	96	175.32	184.83	18.86	137.87	212.77
All samples	1918	37.19	53.86	1.23	34.78	39.61
<b>FCOL</b>						
A1	276	9.82	3.98	0.24	9.35	10.29
B1	1546	36.63	14.31	0.36	35.92	37.35
C1	96	190.36	188.38	19.23	152.19	228.53
All samples	1918	40.47	56.55	1.29	37.94	43.00
<b>TCOL</b>						
A1	276	27.98	19.84	1.19	25.63	30.33
B1	1546	50.24	44.89	1.14	48.00	52.48
C1	96	272.37	237.65	24.26	224.22	320.52
All samples	1918	58.16	83.42	1.90	54.42	61.89
<b>STREPT</b>						
A1	276	6.85	3.36	0.20	6.45	7.24
B1	1546	13.82	5.23	0.13	13.56	14.08
C1	96	22.81	9.61	0.98	20.87	24.76
All samples	1918	13.27	6.25	0.14	12.99	13.55
<b>Second period</b>						
<b>ECOL</b>						
A2	1598	29.26	9.95	0.25	28.78	29.75
B2	237	56.32	13.15	0.85	54.64	58.00
C2	83	187.99	275.40	30.23	127.86	248.13
All samples	1918	39.48	66.52	1.52	36.50	42.46
<b>FCOL</b>						
A2	1598	31.46	11.00	0.28	30.92	32.00
B2	237	71.00	14.23	0.92	69.18	72.82
C2	83	226.79	306.48	33.64	159.86	293.71
All samples	1918	44.80	76.23	1.74	41.39	48.21
<b>TCOL</b>						
A2	1598	46.80	24.81	0.62	45.58	48.02
B2	237	103.76	42.36	2.75	98.34	109.18
C2	83	320.76	387.69	42.55	236.11	405.42
All samples	1918	65.70	102.26	2.33	61.12	70.27
<b>STREPT</b>						
A2	1598	10.07	3.75	0.09	9.88	10.25
B2	237	16.91	4.98	0.32	16.27	17.55
C2	83	21.78	14.20	1.56	18.67	24.88
All samples	1918	11.42	5.77	0.13	11.16	11.68
<b>Third period</b>						
<b>ECOL</b>						
A3	1121	26.04	9.00	0.27	25.51	26.57
B3	699	46.23	16.20	0.61	45.02	47.43
C3	98	138.29	197.54	19.95	98.69	177.89
All samples	1918	39.13	52.33	1.19	36.79	41.47
<b>FCOL</b>						
A3	1121	28.80	9.41	0.28	28.25	29.36
B3	699	56.68	18.99	0.72	55.27	58.09
C3	98	185.57	237.41	23.98	137.97	233.17
All samples	1918	46.97	65.15	1.49	44.06	49.89
<b>TCOL</b>						
A3	1121	44.94	19.53	0.58	43.79	46.08
B3	699	91.70	45.97	1.74	88.28	95.11
C3	98	283.34	310.61	31.38	221.07	345.61
All samples	1918	74.16	93.40	2.13	69.98	78.34
<b>STREPT</b>						
A3	1121	9.53	2.36	0.07	9.40	9.67
B3	699	15.67	4.62	0.17	15.33	16.01
C3	98	20.08	8.25	0.83	18.43	21.74
All samples	1918	12.31	5.11	0.12	12.08	12.54

stepwise modes of DA constructed DFs containing all parameters. The discriminant functions (DFs), using the four discriminant variables (microbiological parameters) yielded classification matrices (CMs) correctly assigning 93.60%, 90.80% and 88.20% of all the cases for the 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> period respectively.

## Conclusions

This research concerns the study of the seawater quality at bathing coasts, based on four microbiological parameters. Multivariate statistical analysis methods were applied to group the bathing beaches in order to accomplish the assessment and modelling of the microbiological quality data of them.

CA was applied to investigate the effect of seasonality on water quality of the bathing beaches surveyed. CA results indicated that the studied bathing beaches are classified in accordance with the seasonality in three groups. The first group consists of May, the second comprises June to August and the third group September and October (Figure 2). Therefore, it appears that the water quality of coastal bathing is completely dependent on seasonality and varies from spring to summer and then autumn.

By implementing CA in each seasonal group of bathing coasts and all coasts together, a new grouping arises clearly clustering the determined parameters in two new groups. The first group includes parameters ECOL, FCOL and TCOL, while the second includes the parameter STREPT (Figure 2). Furthermore, FA was applied to investigate possible determining factors in each of above groups that resulted from the CA. As shown in Table 1, in each group two new parameters were created (VF1 and VF2), where VF1 includes ECOL, FCOL and TCOL and VF2 includes STREPT.

By applying the CA in each seasonal group, three new groups of coasts were generated, group A (ultraclean), group B (clean) and group C (infected). The above analysis is confirmed by the application of DA.

CA and DA, as well as the FA give identical results, grouping the studied parameters in two groups or two new parameters, respectively. Based on the results it can be concluded that the study of water of the bathing beaches as to its microbiological quality does not require the identification of all four parameters as it is sufficient to identify only two: ECOL and STREPT. Taking into account all the studied beaches, ECOL interprets approximately 74% and STREPT about 21% of the total variance, while both together interpret nearly 95% of the total variance in the quality of the coastline (Table 1).

Considering the above, it is understood that the results of this research can contribute to economies of scale in determining the water quality of coastal bathing waters.

Specifically, in Greece the bathing coasts are audited at least 13 times a year for the above four microbiological parameters in accordance with the law (the Greek legislation is in line with the new Directive 2006/7/EU regarding the management of bathing water quality and repealing, Directive 76/160/EEC). Given that today in Greece four microbiological parameters are identified and the results of this research show that only two are needed, cost saving that can be achieved is approximately 50%. Therefore, from an approximate initial cost of 1.700.000 Euro (based on 2.149 swimming locations tested 13 times a year (28.037 samples, ~60 euro per sample) about 850.000 Euro can be economized without risking the health of bathers, since the determination of the parameters ECOL and STREPT interprets nearly 95% of the variation in quality throughout the studied coasts. Moreover, the parameter ECOL by itself seems to be a reliable quality marker, since it corresponds to 74% of the total variance, according FA results. In this case the cost reduction would be about 75% or around 1.250.000

**Table 3. Statistical descriptive [mean, standard error (SE), standard deviation (SD) and CI-95% for mean] of the four microbiological parameters per studied temporal cluster (period) and per spatial cluster of the 1<sup>st</sup> period.**

Parameter, spatial cluster	N.	Mean	SD	SE	95% CI for mean	
<b>First period</b>						
<b>ECOL</b>						
A1	276	8.53	3.60	0.22	8.11	276
B1	1546	33.73	12.36	0.31	33.12	1546
C1	96	175.32	184.83	18.86	137.87	96
All samples	1918	37.19	53.86	1.23	34.78	1918
<b>FCOL</b>						
A1	276	9.82	3.98	0.24	9.35	276
B1	1546	36.63	14.31	0.36	35.92	1546
C1	96	190.36	188.38	19.23	152.19	96
All samples	1918	40.47	56.55	1.29	37.94	1918
<b>TCOL</b>						
A1	276	27.98	19.84	1.19	25.63	276
B1	1546	50.24	44.89	1.14	48.00	1546
C1	96	272.37	237.65	24.26	224.22	96
All samples	1918	58.16	83.42	1.90	54.42	1918
<b>STREPT</b>						
A1	276	6.85	3.36	0.20	6.45	276
B1	1546	13.82	5.23	0.13	13.56	1546
C1	96	22.81	9.61	0.98	20.87	96
All samples	1918	13.27	6.25	0.14	12.99	1918
<b>Second period</b>						
<b>ECOL</b>						
A2	276	16.77	9.96	0.60	15.59	17.95
B2	1546	40.04	53.79	1.37	37.35	42.72
C2	96	95.74	193.02	19.70	56.64	134.85
All samples	1918	39.48	66.52	1.52	36.50	42.46
<b>FCOL</b>						
A2	276	18.62	11.63	0.70	17.25	20.00
B2	1546	45.27	62.77	1.60	42.13	48.40
C2	96	112.53	214.82	21.93	69.00	156.06
All samples	1918	44.80	76.23	1.74	41.39	48.21
<b>TCOL</b>						
A2	276	60.41	56.98	3.43	53.66	67.16
B2	1546	59.98	72.29	1.84	56.38	63.59
C2	96	172.87	323.11	32.98	107.40	238.34
All samples	1918	65.70	102.26	2.33	61.12	70.27
<b>STREPT</b>						
A2	276	11.41	7.98	0.48	10.47	12.36
B2	1546	11.05	4.90	0.12	10.81	11.30
C2	96	17.31	7.71	0.79	15.75	18.87
All samples	1918	11.42	5.77	0.13	11.16	11.68
<b>Third period</b>						
<b>ECOL</b>						
A3	276	16.35	10.85	0.65	15.07	17.64
B3	1546	40.99	51.76	1.32	38.41	43.58
C3	96	74.64	92.25	9.42	55.95	93.33
All samples	1918	39.13	52.33	1.19	36.79	41.47
<b>FCOL</b>						
A3	276	19.25	12.72	0.77	17.75	20.76
B3	1546	49.06	65.05	1.65	45.81	52.30
C3	96	93.12	109.07	11.13	71.02	115.22
All samples	1918	46.97	65.15	1.49	44.06	49.89
<b>TCOL</b>						
A3	276	55.87	39.82	2.40	51.16	60.59
B3	1546	71.73	85.07	2.16	67.49	75.98
C3	96	165.78	210.21	21.45	123.19	208.37
All samples	1918	74.16	93.40	2.13	69.98	78.34
<b>STREPT</b>						
A3	276	10.79	3.78	0.23	10.34	11.24
B3	1546	12.22	4.78	0.12	11.98	12.45
C3	96	18.18	8.54	0.87	16.45	19.91

Euro. These funds can be used in other actions to preserve the quality of coastal water and human health. This in turn, would aid in the assessment of the quality of coastal bathing waters and provide a more timely indication of bathing water quality, hence contributing to the immediate health protection of bathers.

The results of this study show that observing the quality of coastal bathing water can be accomplished by specifying and monitoring the parameters ECOL and STREPT. As a consequence, countries, especially those with large quantities of coastal bathing sites, can perform microbiological monitoring of their bathing waters by checking only the mentioned two parameters, thus ensuring economies of scale.

Correspondence: Agelos Papaioannou, Clinical Chemistry Section, Department of Medical Laboratories, Education and Technological Institute of Thessaly, 41110 Larissa, Greece.

Tel.: +30.241.068.4448 - Fax: +30.241.068.4650.

E-mail: papaioannou@teilar.gr

Key words: public health, chemometric methods, coastal bathing quality, bacterial indicators, Mediterranean.

Contributions: the authors contributed equally.

Conflict of interests: the authors declare no potential conflict of interests.

Received for publication: 23 September 2014.

Accepted for publication: 1 December 2014.

©Copyright A. Papaioannou et al., 2014

Licensee PAGEPress, Italy

Journal of Public Health Research 2014; 3:357

doi:10.4081/jphr.2014.357

This work is licensed under a Creative Commons Attribution NonCommercial 3.0 License (CC BY-NC 3.0).

## References

- Bouvy M, Briand E, Boup MM, et al. Effects of sewage discharges on microbial components in tropical coastal waters (Senegal, West Africa). *Mar Freshwater Res* 2008;59:614-26.
- Beversdorf LJ, Bornstein-Forst SM, McLellan SL. The potential for beach sand to serve as a reservoir for *Escherichia coli* and the physical influences on cell die-off. *J Appl Microbiol* 2007;102:1372-81.
- Cabelli V. Health effects criteria for marine recreational waters. U.S. Environmental Protection Agency; 1983. EPA-600/1-80-031.
- Kay D, Fleischer JM, Salomon RL, et al. Predicting likelihood of gastroenteritis from sea bathing: results from randomised exposure. *Lancet* 1994;344:905-9.
- Lipp EK, Farrah SA, Rose JB. Assessment and impact of microbial fecal pollution and human enteric pathogens in a coastal community. *Mar Pollut Bull* 2001;42:286-93.
- Lipp EK, Kurz R, Vincent R, et al. The effects of seasonal variability and weather on microbial fecal pollution and enteric pathogens in a subtropical estuary. *Estuaries* 2001;24:266-76.
- Schiff KC, Weisberg SB, Dorsey JH. Microbiological monitoring of marine recreational waters in southern California. *Environ Manage* 2001;27:149-57.
- Shuval H. Estimating the global burden of thalassogenic diseases: human infectious diseases caused by wastewater pollution of the marine environment. *J. Water Health* 2003;1:53-64.
- Smith LM, Macauley JM, Harwell LC, Chancy CA. Water quality in the near coastal waters of the Gulf of Mexico affected by Hurricane Katrina: before and after the storm. *Environ Manage* 2009;44:149-62.
- Wade TJ, Pai N, Eisenberg JN, Colford JM Jr. Do US Environmental Protection Agency water quality guidelines for recreational waters prevent gastrointestinal illness? (A systematic review and meta-analysis). *Environ Health Perspect* 2003;111:1102-9.
- Ostoich M, Aimo E, Fassina D, et al. Biologic impact on the coastal belt of the province of Venice (Italy, Northern Adriatic Sea): preliminary analysis for the characterization of the bathing water profile. *Environ Sci Pollut Res* 2011;18:247-59.
- Basset A. Aquatic science and the water framework directive: a still open challenge towards ecogovernance of aquatic ecosystems. *Aquatic Conserv: Mar Freshw Ecosyst* 2010;20:245-9.
- Chandran A, Hatha AAM, Varghese S. Increased prevalence of indicator and pathogenic bacteria in Vembanadu Lake: a function of salt water regulator, along south west coast of India. *J Water Health* 2008;6:539-46.
- Cinotto PJ. Occurrence of fecal-indicator bacteria and protocols for identification of fecal-contamination sources in selected reaches of the West Branch Brandywine Creek, Chester County, Pennsylvania: US Geological Survey Scientific Investigations Report 2005-5039. p 91.
- EU Directive 2006/7/EC of the European Parliament and of the Council. Directive of 15 February concerning the management of bathing water quality and repealing Directive 76/160/EEC. *Official J European Communities* 2006;L64:37-51.
- U.S. Environmental Protection Agency (USEPA). Ambient water quality criteria for bacteria. Washington, DC; 1986. EPA 440/5-84-002.
- U.S. Environmental Protection Agency (USEPA). Quality criteria for water Washington, DC; 1986. EPA 440/5-86-001.
- WHO. Monitoring bathing waters: a practical guide to the design and implementation of assessments and monitoring programmes. Geneva: WHO; 2000.
- WHO. Water quality: guidelines, standards and health. London: IWA; 2001.
- WHO. Guidelines for safe recreational-water environments: Vol. 1. Coastal and fresh-waters. Geneva: WHO; 2003.
- WHO. Addendum to the WHO guidelines for safe recreational water environments. Vol. 1. Coastal and fresh waters. WHO, Geneva; 2009.
- Papastergiou P, Mouchtouri V, Karanika M, et al. Analysis of seawater microbiological quality data in Greece from 1997 to 2006: association of risk factors with bacterial indicators. *J Water Health* 2009;7:514-26.
- Almeida C, Gonzalez SO, Mallea M, Gonzalez P. A recreational water quality index using chemical, physical and microbiological parameters. *Environ Sci Pollut Res* 2012;19:3400-11.
- Lattin J, Carroll D, Green P. Analyzing multivariate data. New York: Duxbury; 2003.
- McKenna J. An enhanced cluster analysis program with bootstrap significance testing for ecological community analysis. *Environ Modell Softw* 2003;18:205-20.
- Otto M. Multivariate methods. In: Kellner R, Mermet JM, Otto M, Widmer HM, Eds. *Analytical Chemistry*. Weinheim: Wiley-VCH; 1998.
- Astel A, Tsakovski S, Simeonov V, et al. Multivariate classification and modeling in surface water pollution estimation. *Anal Bioanal Chem* 2008;390:1283-92.
- Kowalkowski T, Zbytniewski R, Szpejna J, Buszewski B. Application of chemometrics in river water classification. *Water Res* 2006;40:744-52.
- Shrestha S, Kazama F. Assessment of surface water quality using multivariate statistical techniques: a case study of the Fuji River

- Basin, Japan. *Environ Modell Softw* 2007;22:464-75.
30. Simeonov V, Stratis JA, Samara C, et al. Assessment of the surface water quality in Northern Greece. *Water Res* 2003;37:4119-24.
  31. Singh KP, Malik A, Mohan D, Sinha S. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India): a case study. *Water Res* 2004;38:3980-92.
  32. Venugopal T, Giridharan L, Jayaprakash M. Application of chemometric analysis for identifying pollution sources: a case study on the River Adyar, India. *Marine Freshwater Res* 2009;60:1254-64.
  33. Wunderlin DA, Diaz MP, Ame MV, et al. Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquia river basin (Cordoba-Argentina). *Water Res* 2001;35:2881-94.
  34. Zhou F, Guo HC, Liu Y, Jiang YM. Chemometrics data analysis of marine water quality and source identification in Southern Hong Kong. *Mar Pollut Bull* 2007;54:745-56.
  35. Gupta I, Dhage S, Kumar R. Study of variations in water quality of Mumbai Coast through multivariate analysis techniques. *Indian J Mar Sci* 2009;38:170-7.
  36. Jagadeesan L, Manju M, Perumal P, Anantharaman P. Temporal variations of water quality characteristics and their principal sources in tropical Vellar Estuary, south east coast of India. *Res J Environ Sci* 2011;5:703-13.
  37. Papaioannou A, Dovriki E, Rigas N, et al. Assessment and modeling of groundwater quality data by environmental methods in the context of public health. *Water Resour Manag* 2010;24:3257-78.
  38. Papaioannou A, Mavridou A, Hadjichristodoulou C, et al. Application of multivariate statistical methods for groundwater physicochemical and biological quality assessment in the context of public health. *Environ Monit Assess* 2010;170:87-97.

Non commercial use only