Aquatic Heterotrophic Bacteria: Modeling in the Presence of Spatial Autocorrelation

Pierre Legendre; Marc Troussellier

Limnology and Oceanography, Vol. 33, No. 5 (Sep., 1988), 1055-1067.

Stable URL:
http://links.jstor.org/sici?sici=0024-3590%28198809%2933%3A5%3C1055%3AAHBMIT%3E2.0.CO%3B2-E

Limnology and Oceanography is currently published by American Society of Limnology and Oceanography.

Limnol. Oceanogr., 33(5), 1988, 1055–1067

Aquatic heterotrophic bacteria: Modeling in the presence of spatial autocorrelation

Pierre Legendre
Département de sciences biologiques, Université de Montréal, C.P. 6128, Succursale A, Montréal, Québec H3C 3J7

Marc Troussellier
Laboratoire d'Hydrobiologie marine, U.A. C.N.R.S. 694, Université des Sciences et Techniques du Languedoc, Place E. Bataillon, F-34060 Montpellier Cedex, France

Abstract

Microbial ecologists often obtain data from sampling a piece of geographic space. These are likely to be spatially autocorrelated. Autocorrelation removes degrees of freedom from the usual tests of inferential statistics and can generate spurious correlations among variables, with the consequence that suspected causal relations may not hold. This paper describes methods that can be used to explore the spatial structure of ecological data and to include spatial location as a variable in the study of relationships and models. The relationship between environmental heterotrophic bacteria and phytoplankton, well established in aquatic environments, is re-examined in the Thau brackish lagoon (Mediterranean coast of France). It did not hold for the bacteria growing on bioMérieux nutrient agar (BNA), which are presumably of continental origin; their spatial gradient can only partly be explained by the particulate organic carbon variable (POC) and not at all by phytoplankton biomass (CHL A), despite the existence of a spurious correlation between BNA and CHL A. The spatial gradient of abundance of heterotrophs growing on marine agar (MA), expected to be mostly of marine origin, can be entirely explained by POC and CHL A. Different segments of the bacterial community, both reacting positively to variations of the particulate organic carbon, may follow partly, or not, variations of phytoplankton biomass. The mode of analysis developed here extends to many other spatially distributed processes in ecology and other fields.

The study of aerobic heterotrophs often entails the search for mechanisms explaining variations of abundance observed among samples. Relating these variations to possible causes is usually done by testing implicit or explicit hypotheses that the investigator has in mind. Possible explanatory variables are measured or observed in synchrony with the bacterial variables, and hypotheses are tested, often by computing correlations or simple linear regressions, or in some cases by more complex types of modeling.

Acknowledgments

This work was supported by grants from the France-Québec exchange program and by NSERC grant A7738 to P. Legendre.

Contribution 1 of the ECOTHAU program, Laboratoire d'Hydrobiologie marine, Université des Sciences et Techniques du Languedoc, and contribution 336 of the Groupe d'Écologie des Eaux douces, Université de Montréal.

We thank G. Cauwet and G. Jacques for the use of their organic carbon and chlorophyll data obtained during the ECOTHAU sampling program.

Sampling in the natural environment can be planned in such a way as to grasp variations through space, or through time, or sometimes both. When the data have been obtained and analyzed, and relationships among variables are found, it becomes legitimate to wonder whether we are looking at real relationships or only at spurious correlations induced by a common spatial or temporal structure. To our knowledge, this question has not yet been addressed in the microbial ecology literature. Beyond marine microflora, the mode of analysis developed here extends to many other studies of space-based ecological processes.

Let us look at a few examples from recent microbial literature where this problem is potentially present because the observations are structured through space or through time, which possibly makes the data autocorrelated (*see below*). Common structures that can be identified in space (or in time) for ecological variables are gradients and patches (clumps, aggregates). Patches are

characterized by alternating, significant positive and negative autocorrelation as one goes from short to long distances, while gradients are recognized by significant positive short-distance and negative long-distance autocorrelation (Sokal 1979). One further problem is that by violating the assumption of independence, positive autocorrelation in small distance classes makes such tests as the ANOVA F-test, t-tests, and correlation or regression analyses too liberal; the actual numbers of degrees of freedom are much smaller than the number of samples might suggest. The consequence is that in the presence of positive spatial autocorrelation, differences among group means that in reality are not significant may be found to be so, or correlation and regression coefficients may be declared significantly different from zero when they are not (Bivand 1980; Cliff and Ord 1981). Negative spatial autocorrelation for small distance classes may have the opposite effect. Consequently, data should be tested for spatial independence with spatial correlograms or similar methods before one uses the standard statistical methods that assume the observations to be independent. Examples follow where data should have been tested for spatial independence, but have not. This list is by no means exhaustive.

In the space domain, Mahloch (1974) used multiple regressions to test predictive models of coliform abundances, after collecting data at 20 stations located along the course of the Leaf River, U.S.A. Bölter et al. (1981) computed rank correlation coefficients among 33 variables (chemical, physical, biological) for 12 stations in the Kiel Fjord and the Kiel Bight, F.R.G. Cammen and Walker (1982) computed linear correlations between bacterial counts and suspended particulate matter for 22 stations forming two lines parallel to the long axis of the Bay of Fundy, Canada; they also computed multiple regression models to explain maximal uptake rates with the same data. Linley et al. (1983) analyzed the relation between bacterial abundances and chlorophyll *a,* dissolved organic carbon, and particulate carbon with linear correlations and simple linear regressions on data from a vertical profile (nine depths) in the English Channel. *In the time domain,* Väätänen (1982) computed t-tests and multiple regression models to compare two stations in a coastal archipelago of Finland for 22 variables with an 11-month series of 18 sampling dates; he also computed differences between paired proportions (χ^2-test) for three sites taken two by two with an 18-month series of 55 sampling dates. Wright and Coffin (1983) used a 12-month sampling series to test relations between total bacterial counts, heterotrophic activity, and temperature in the Essex River estuary, U.S.A., with linear correlations and simple linear regressions. Kirchman et al. (1984) used 26 hourly samples (forming two consecutive tidal cycles) to correlate bacterial abundances and chlorinity in a salt marsh of the Cape Cod area, U.S.A. Troussellier et al. (1986) used path analysis, which is based on multiple regressions, to test models of the evolution of bacterial counts in sewage treatment lagoons in France with a 19-month data series made of 41 sampling dates. Finally, workers sometimes analyze data that are possibly autocorrelated *in both the space and time domains.* Miyoshi and Nakamoto (1975) computed a multiple regression model of total bacterial counts involving several explanatory variables from 98 observations drawn from eight sampling stations, five depths, and five sampling dates in the Hiuchi-Nada Sea, Japan. Väätänen (1982) computed multiple regression models to explain the abundances of each of nine subsets of microorganisms, with 24 samples representing six to nine stations visited at three different times.

Autocorrelation as a statistical phenomenon has been investigated for a long time by specialists of time series (Box and Jenkins 1970; Fry et al. 1981) and more recently in the case of space by statistical geographers (Cliff and Ord 1973, 1981). Spatial autocorrelation has often been discussed in the context of oceanography (e.g. Jumars et al. 1977; Jumars 1978; Ibanez 1981; Yoder et al. 1987; Legendre and Legendre 1988). The effects of spatial autocorrelation on tests of statistical significance have also been discussed (Bivand 1980; Cliff and Ord 1981; Legendre et al. in prep.); this effect stems from the violation of the assumption of in-

dependence of the observations. Bivand (1980) and Cliff and Ord (1981), in particular, present simulation data and graphs showing the magnitude of the bias generated by various amounts of spatial autocorrelation in the data.

The aims of this paper are two. Ecologically we wish to demonstrate that even after finding significant correlations, some assumed relationships between bacteria and environmental variables can be spurious, implying a common spatial gradient, while others are real. Methodologically we wish to bring a new level of analysis to the problem of correlation and regression, showing how space can be handled in the study of relationships in ecology, and in particular how to handle it as a full-fledged explanatory variable in modeling. What will be said about space can be generalized to time, which is but a simpler (one-dimensional) case.

Materials and methods

On 17 June 1986, the Thau brackish lagoon, located in southern France (43°20′–43°28′N, 3°32′–3°42′E) was subjected to intensive sampling, as part of the ECOTHAU research program. Productivity of the Thau lagoon is a question of economic importance for the Languedoc region, because of large-scale eel fishing and mollusc farming (mussels and oysters) taking place in this 75-km^2 water body. Average depth is 4 m. The lagoon has three communication channels with the Mediterranean Sea (arrows in Fig. 1); it also receives freshwater from the Canal du Midi at its southwestern end and from various streams on its northwestern side. A systematic sampling design was applied and 63 locations, at the nodes of a 1-km regular grid, were visited in <4 h by three teams of investigators working from three boats. All data reported here were collected at 0.5-m depth. The sampling grid covers the whole of this largely enclosed ecosystem.

The bacterial variables studied are the number of colony-forming units of aerobic heterotrophs growing on bioMérieux nutrient agar (low NaCl concentration; they are called *Bna* hereafter) and those growing on marine agar (34‰ salinity; they are designated *Ma*). Water samples were collected in sterile vials, kept at 4°C, and transported to the lab where they were spread-plated with a series of dilutions, within 5 h of sampling. Plates were incubated at lab temperature and counted after 8 d. The nutrient agar medium (bioMérieux) was used to reveal bacteria of continental origin. The marine agar medium (Difco), by contrast, was expected to bring out mostly the aerobic heterotrophs endemic to the marine lagoon, where salinity varies roughly from 31 to 39‰. Difco marine agar was selected because it is a highly standardized product that has been shown to produce, with lagoon marine water, higher counts than either of its competitors, ZoBell medium 2216 and Difco bacto nutrient agar with 30 g liter^{-1} of NaCl added (Troussellier 1987).

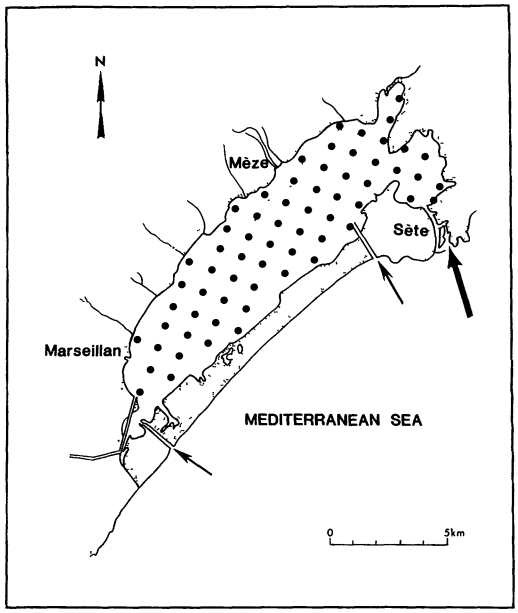

Fig. 1. Map of the 63 sampling stations (black dots) in the Thau lagoon. Dotted line is the 5-m isobath. Arrows represent marine water inputs.

Chlorophyll *a* concentration (variable *Chl a*, in μg liter^{-1}), which is an indirect and approximate measure of phytoplankton biomass, was obtained by acetone extraction and fluorimetric assay (Neveux and Panouse 1987). Particulate organic carbon (variable *POC*, in mg liter^{-1}) was extracted by filtration and assayed with a nondispersive infrared detector (Cauwet 1983); after

removal of *POC,* dissolved organic carbon (variable *DOC,* in mg liter^{-1}) was measured with a flame ionization detector (Cauwet 1984). Finally, the spatial variable (SPACE) is represented by a matrix of geographic distances, computed from the coordinates of the sampling stations on a map.

The Box-Cox method (Sokal and Rohlf 1981) was used to approximate the best normalizing transformation for all variables. Normalization is used here as an empirical way of reducing asymmetry in the frequency distributions of the variables and thus increasing linearity for the product-moment correlation analysis. As confirmed by a Kolmogorov-Smirnov test of normality (Lilliefors 1967), the $y = \log_e(x + 1)$ transformation was found to normalize both bacterial variables as well as Chl *a* concentrations, and it was applied to the data. For the particulate and dissolved organic carbon variables, a log transformation was insufficient to reach normality. The optimal transformations suggested by the Box-Cox method, which indeed normalized the data, were $y = (x^\delta - 1)/\delta$ with $\delta = -0.0918$ for the *POC* variable, and $\delta = -3.78401$ for the *DOC* variable.

The analysis of spatial structures can be done in a variety of ways that provide different information. We started by plotting maps of the values of the four variables. These maps were obtained by interpolation, using the SYMAP package (Dougenik and Sheehan 1975). The spatially distributed transformed (above) values of each variable were analyzed to detect significant spatial structures, using spatial correlograms. A spatial correlogram is a graph of autocorrelation values (along the ordinate) as a function of the distance between points (abscissa). Both Moran's (1950) *I* and Geary's *c* coefficients were used in this study, and they evidenced the same type of spatial structure. [Oceanographers will notice that these coefficients differ from the one used by Mackas (1984) in his "dissimilarity correlogram," which is actually a multivariate variogram; as with standard variograms (Matheron 1971), individual values of the statistic in the correlogram of Mackas are not amenable to testing.] In correlogram analysis, one value is computed for all pairs of points located within each given distance class; the values computed for all distance classes are assembled in the correlogram. Each value can also be tested for significance (Cliff and Ord 1973; Sokal and Oden 1978; Legendre and Legendre 1984) after the overall significance of the correlogram has been determined (Oden 1984). Finally, an overall examination of the significant values in a correlogram makes it possible to formulate statements about the underlying spatial structure (Sokal 1979). Correlations among variables were computed with Pearson's product-moment correlation coefficient.

The Mantel (1967) test is primarily useful to look for a spatial trend in data corresponding to some form of diffusive process. Mantel proposed to represent the spatial relationship among sampling localities by a matrix **A** of geographic distances among all pairs of these geographic points; this matrix is referred to as SPACE. Mantel's test looks for a relationship between this matrix of geographic distances and some other distance matrix **B** which is meaningful for the problem at hand. The trend in the data may be linear, in which case geographic distances are used directly for testing, or it may be hypothesized to follow some other relationship, in which case other functions of the geographic distances (D) may be used ($1/D$, or $1/D^2$). In the multivariate case, one would compute one of the many multivariate dissimilarity functions available in the literature as dictated by the nature of the problem; Legendre and Legendre (1983, 1984) and Gower and Legendre (1986) give guidance to the choice of a coefficient. Since this paper deals with the one-variable case, the dissimilarity matrix CHL A for variable *Chl a,* say, is formed by taking the unsigned difference among values of this variable for all possible pairs of stations; this is the common form that most multivariate distance functions boil down to in the univariate case. (Notice that capitals are used throughout to designate dissimilarity or distance matrices, while italics are used for the variables themselves.) The Mantel statistic is simply the sum of the cross-products of the corresponding values in the two matrices **A** and **B** under investigation (for instance, geo-

graphic distance and Chl *a* dissimilarity). The statistic can be normalized to range between −1 and +1 by computing it as a product-moment correlation coefficient between the corresponding values of the two matrices, excluding the main diagonal; the probability associated with this normalized Mantel statistic (*see below*) is exactly the same as for the original Mantel statistic. The Mantel or the normalized Mantel statistic can be tested for significance in one of two possible ways (Mantel 1967): either through a permutation test, or, when the number of data points is large enough as is the case in the present study ($n = 63$), by computing the expected value and variance under the null hypothesis and performing a z-test. The null hypothesis of the test is, in both cases, that the distances in matrices **A** and **B** are not linearly related.

It is well known that two variables may seem related when both are correlated to a third, common cause. Spatial position is a good candidate for causing such spurious correlations: indeed, variables may seem related because they are driven by a common spatial gradient. The statistical problem of computing a partial correlation among variables while controlling for the effect of spatial position has been solved by Smouse et al. (1986). *First,* one computes three matrices **A**, **B** and **C**; **A** contains for instance geographic distances among sampling stations, as above, while **B** and **C** are dissimilarity matrices, computed either from a single variable or from a multivariate data set. *Second,* one computes **B′** which contains the residuals of the regression of the values in **B** on the values in **A**, and **C′** which contains the residuals of the regression of the values in **C** on the values in **A**; finally the standardized Mantel statistic is computed between the values in **B′** and those in **C′**. This procedure is just a way of computing the partial Mantel relation $r_{\mathbf{BC}\cdot\mathbf{A}}$, as it is a standard way of computing a partial correlation coefficient for two variables b and c while controlling the effect of a third variable a. The partial Mantel relation will be denoted $(\mathbf{B}\cdot\mathbf{C})\cdot\mathbf{A}$ for convenience; this symbolism is simpler when full variable names are used (below). *Third,* the probability associated with this partial Mantel statistic is computed in the usual way, either by permutation or using the approximate z-test. Since the test is performed between the two residual matrices **B′** and **C′**, the partial test is then computed in exactly the same way as the simple Mantel test above. The partial Mantel test will be used in the larger context of causal modeling, as described by De Neufville and Stafford (1971) and by Legendre and Legendre (1983, 1984) for partial Pearson correlations. Computations were carried out with the MANTEL-3 program written by A. Vaudor and included in the "R package for multivariate data analysis" distributed by P.L.

Results

Interpolated maps for the four variables *Bna* (aerobic heterotrophs growing on bioMérieux nutrient agar), *Ma* (aerobic heterotrophs growing on marine agar), *Chl a* (chlorophyll *a*) and *POC* (particulate organic carbon) are represented in Fig. 2, based on observations at the 63 sampling stations. These maps differ little from maps obtained by the more sophisticated technique of kriging, which were also produced but are not presented here. They all suggest the presence of a spatial gradient with high values in the northeast part of the lagoon and low values in the southwest. The mean, range, and C.V. for these variables and for *DOC* (dissolved organic carbon) are presented in Table 1.

The correlogram of *Chl a* (Fig. 3) easily passes Oden's (1984) test of overall significance. Further examination clearly shows that observations located near one another over the whole surface of the lagoon have very similar values (significant positive values of the I coefficient at low distance classes). On the contrary, distant observations have dissimilar values of Chl *a,* which translate into significant negative values of I for the large distance classes. This shape for a correlogram, with significant positive values in the low distance classes and significant negative values in the high distance classes, is characteristic of a spatial gradient (Sokal 1979), which statistically confirms the first impressions obtained from the maps. The other three variables, *Bna, Ma,* and *POC,* produced spatial correlograms

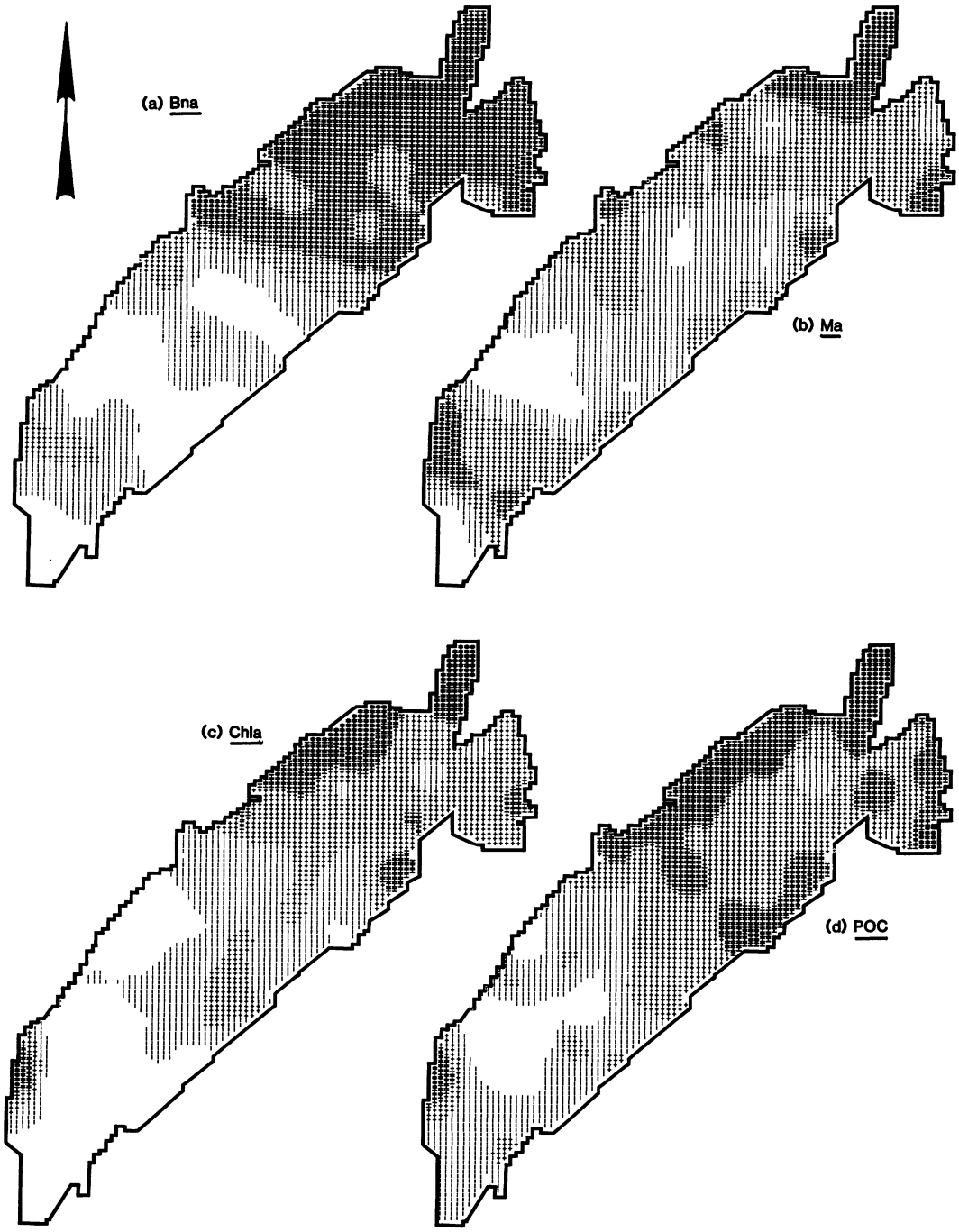

Fig. 2. Interpolated maps of the four variables. Darker areas correspond to higher values of the variables.

that are also characteristic of spatial gradients.

All four variables, *Bna, Ma, Chl a,* and *POC,* are very highly correlated to one another (Table 2). The significance of Pearson's product-moment correlation coefficient is computed in the usual way, without taking into account the fact that all these

Table 1. Basic statistics for the five variables under study (63 sampling stations).

	Min	Mean	Max	C.V. (%)
Bna (N ml^{-1})	1.0	3.2×10^{2}	3.9×10^{3}	207
Ma (N ml^{-1})	4.5×10^{2}	9.5×10^{3}	8.4×10^{4}	170
Chl a (μg liter^{-1})	0.7	2.4	7.5	60
POC (mg liter^{-1})	0.100	0.273	0.690	48
DOC (mg liter^{-1})	2.50	2.98	4.12	11

variables are autocorrelated and form spatial gradients (just as one would have done in most environmental analyses), but the level of significance is corrected to take multiple testing into account (Bonferroni correction: Cooper 1968; Miller 1977). Our purpose is, of course, to show that indications drawn from such analyses can be misleading.

Table 3 shows the results needed to analyze the influences acting on the BNA bacteria, while Table 4 deals with the MA bacteria. In these tables, SPACE is a matrix of geographic distances among sampling stations, while each of the other variables is represented by a matrix of dissimilarities, computed as the unsigned difference among values, as explained in the methods. All analyses were performed with the approximate z-test, after checking that the results were indistinguishable from those obtained with the permutation test.

Discussion

Microbiologists before us have felt the need to include space as an explanatory variable in models. Two cases are Miyoshi and Nakamoto (1975), where "distance from the coast" was included as one of the independent variables in a multiple regression model of total bacterial counts, and Wright and Coffin (1983), where the "distance from the mouth of the estuary" was used in simple linear regressions intended to predict bacterial densities along the course of three different rivers. The various methods of spatial analysis are more versatile, since the analysis is not limited to one-dimensional gradients.

Correlation analysis (Table 2, upper triangle) showed a positive relationship between the heterotrophic bacterial abundance (*Bna* and *Ma*) and Chl *a* concentration (*Chl a*), our indirect measure of phytoplankton biomass. This correlation is well documented in the ecological literature. It has been demonstrated a number of times and in particular for AODC (acridine orange direct counts) by Ferguson and Palumbo (1979) in the marine environment, by Fukami et al. (1983) in a brackish ecosystem, and by Bird and Kalff (1984) in freshwater. The mechanism suggested is that dissolved organic compounds associated with primary production, as well as the particulate matter resulting from decay, are used as substrate by bacterioplankton. Since the model of the control of heterotrophic bacterioplankton by the algal resource is ecologically plausible and is already well supported by the studies mentioned above, ecologists would usually stop their analysis here, satisfied that the correlations in Table 2 corroborate this well-established model.

Introducing SPACE as a variable into the analysis, one finds from the simple Mantel tests (Tables 3 and 4, upper triangles) that the bacterial variables obey the same environmental gradient as phytoplankton. It is only too easy to forget to test the other possible hypothesis—that the correlation of the *Chl a* to bacteria is spurious, both of

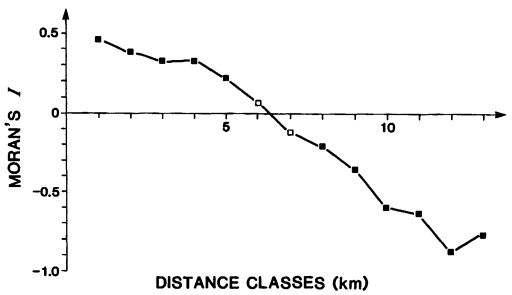

Fig. 3. Spatial correlogram for the *Chl a* variable. Significant autocorrelation coefficients ($P \leq 0.05$)—■; Moran's I coefficient not significantly different from zero—□.

Table 2. Product-moment correlation coefficients and associated probabilities (two-tailed tests) among the five variables *Bna, Na, Chl a, POC,* and *DOC*; $n = 63$.

	Ma	*Chl a*	*POC*	*DOC*
Bna	0.420 $P = 0.00061$*	0.599 $P = 0.00000$*	0.644 $P = 0.00000$*	0.116 $P = 0.36500$ N.S.
Ma	—	0.502 $P = 0.00003$*	0.533 $P = 0.00001$*	0.305 $P = 0.01502$ N.S.
Chl a	—	—	0.729 $P = 0.00000$*	0.117 $P = 0.36148$ N.S.
POC	—	—	—	0.146 $P = 0.25339$ N.S.

* Correlation is significant at the Bonferroni-corrected probability level of (0.05/10 = 0.005) for an overall significance level of 0.05 over 10 simultaneous tests; N.S.—test is not significant.

them being controlled *independently* by environmental factors.

We have chosen the particulate organic carbon variable (*POC*) as the predictive variable for bacterial abundances instead of the dissolved organic carbon (variable *DOC*), because our data show that the *POC* variable is significantly correlated with the *Bna* and *Ma* abundances, while the *DOC* variable is not (Table 2). Since there is positive autocorrelation displayed in the small distance classes of the correlograms of all variables, the effect of positive autocorrelation on tests of statistical significance, such as in linear correlation, is to create a bias in one direction only; positive spatial autocorrelation artifactually could have produced apparently significant correlations, but not the opposite (Bivand 1980), so that we can be confident that *DOC* is really not correlated with our bacterial abundances. On the other hand, Fukami et al. (1981, 1985) have shown, after experimentation, that during phytoplankton degradation, variations in bacterial abundances are more closely related to observed variations of *POC* than to those of *DOC*. This finding does not exclude the possibility that the biodegradable fraction of *DOC* is a more directly accessible organic resource than that of *POC*,

Table 3. Above the diagonal: simple Mantel statistics and associated probabilities. Below the diagonal: partial Mantel statistics and associated probabilities. Tests of significance are one-tailed. The distance matrix held constant in each case is indicated with the dot notation; for instance, the partial Mantel test (BNA·CHL A)·POC is indicated as ·POC in column BNA and line CHL A.

	BNA	CHL A	POC	SPACE
BNA	———	0.258 $P = 0.00000$*	0.315 $P = 0.00000$*	0.521 $P = 0.00000$*
CHL A	·POC = 0.130 $P = 0.00292$† ·SPACE = −0.006 $P = 0.45694$ N.S.	———	0.476 $P = 0.00000$*	0.505 $P = 0.00000$*
POC	·CHL A = 0.226 $P = 0.00000$† ·SPACE = 0.168 $P = 0.00109$†	·BNA = 0.431 $P = 0.00000$† ·SPACE = 0.372 $P = 0.00000$†	———	0.347 $P = 0.00000$*
SPACE	·CHL A = 0.468 $P = 0.00000$† ·POC = 0.462 $P = 0.00000$†	·BNA = 0.449 $P = 0.00000$† ·POC = 0.411 $P = 0.00000$†	·BNA = 0.226 $P = 0.00003$† ·CHL A = 0.141 $P = 0.00128$†	———

* Mantel test is significant at the Bonferroni-corrected probability level of (0.05/6 = 0.00833) for an overall significance level of 0.05 over six simultaneous tests.

† Partial Mantel test is significant at the Bonferroni-corrected probability level of (0.05/12 = 0.00417) for an overall significance level of 0.05 over 12 simultaneous tests; N.S.—test is not significant.

Table 4. As Table 3, but for MA bacteria.

	MA	CHL A	POC	SPACE
MA	——	0.325 $P = 0.00000$*	0.363 $P = 0.00000$*	0.223 $P = 0.00000$*
CHL A	·POC = 0.185 $P = 0.00046$† ·SPACE = 0.252 $P = 0.00001$†	——	0.476 $P = 0.00000$*	0.505 $P = 0.00000$*
POC	·CHL A = 0.250 $P = 0.00001$† ·SPACE = 0.312 $P = 0.00000$†	·MA = 0.407 $P = 0.00000$† ·SPACE = 0.372 $P = 0.00000$†	——	0.347 $P = 0.00000$*
SPACE	·CHL A = 0.073 $P = 0.06015$ N.S. ·POC = 0.111 $P = 0.00591$ N.S.	·MA = 0.469 $P = 0.00000$† ·POC = 0.411 $P = 0.00000$†	·MA = 0.293 $P = 0.00000$† ·CHL A = 0.141 $P = 0.00128$†	——

* As Table 3.
† As Table 3.

but it is likely that the relation between bacteria and *DOC* could only be perceived for finer scales of observation than the one that was used in the present study (1-km sampling grid). On the other hand, the fraction of *DOC* that cannot be degraded by bacteria is possibly greater than that of *POC*; this interpretation would explain why we observe a higher mean concentration of *DOC* than *POC*, while the C.V. of *POC* (48%) is larger than that of *DOC* (11%).

Of course phytoplankton may be one of the fractions, or even the main fraction, of particulate organic carbon. In any case, we learn from Tables 2–4 (upper triangles) that variable *POC* is significantly correlated with *Bna*, *Ma*, and *Chl a*, and that it is also distributed as a spatial gradient. So, at this point, we cannot tell the influence of *Chl a* from that of *POC* on the bacterial variables nor have we tested the alternative to the classical model.

To make clear the relation between a causal model and the partial Mantel tests, let us examine the series of possible models among three variables, BNA, CHL A, and SPACE, and the predictions each one makes about the direct and partial relationships that can be computed among them. In Fig. 4, all the possible three-variable models are listed, except those where SPACE would be a dependent variable, which do not make sense. Indeed, it is not enough to demonstrate that the data support the model that we hypothesized for each case; we also have to show that all other possible alternative models are not supported. We are left with seven different models. For each one, causal analysis makes predictions involving the values of the simple correlations (De Neufville and Stafford 1971) and of the partial correlations (Legendre and Legendre 1983, 1984). We feel authorized to use these predictions here because the standardized simple Mantel statistic is computed as a product-moment correlation coefficient while the partial Mantel statistic is computed as a partial correlation coefficient.

The first six models in Fig. 4 are not supported by the data, because one or several of their predictions are not realized; the results of actual computations are listed for each model, indicating which predictions are realized and which ones are not. When we state that a relation is equal to zero, we mean that it is not significantly different from zero in Table 3. With model 7, all eight predictions that can be made about the values of the simple Mantel and partial Mantel statistics are realized in the results of Table 3, so that this model—and this model only—cannot be rejected. The first conclusion that we reach is the following: contrary to generally accepted ideas about AODC, the subclass of the AODC bacteria that grows on bioMérieux nutrient agar and is presum-

Model	Predictions of model if model is true	Model	Computed results	Model	Computed results
1 2 → 3	1•3 ≠ 0	1) SPACE BNA → CHL A	yes	2) SPACE CHL A → BNA	yes
	2•3 ≠ 0		yes		yes
	1•2 = 0		no		no
	(1•2)•3 ≠ 0		yes		yes
	(2•3)•1 ≠ 0		no		no
	(1•3)•2 ≠ 0		yes		yes
	(1•3)•2 ≥ 1•3		no		no
	(2•3)•1 ≥ 2•3		no		no
1 → 2 → 3	1•2 ≠ 0	3) SPACE → BNA → CHL A	yes	4) SPACE → CHL A → BNA	yes
	2•3 ≠ 0		yes		yes
	1•2 ≥ 1•3		yes		no
	2•3 ≥ 1•3		no		no
	(1•3)•2 = 0		no		no
	(1•2)•3 ≠ 0		yes		yes
	(2•3)•1 ≠ 0		no		no
	(1•2)•3 ≤ 1•2		yes		yes
	(2•3)•1 ≤ 2•3		yes		yes
	(1•2) x (2•3) ≈ (1•3)		0.135 ≠ 0.505		0.130 ≠ 0.521
1 → 2 ↔ 3	1•2 ≠ 0	5) SPACE → CHL A → BNA	yes	6) SPACE → BNA → CHL A	yes
	1•3 ≠ 0		yes		yes
	2•3 ≠ 0		yes		yes
	(1•2)•3 ≠ 0		yes		yes
	(1•3)•2 ≠ 0		yes		yes
	(2•3)•1 ≠ 0		no		no

Model	Predictions of the model if the model is true	Computed results
7) SPACE → CHL A, BNA	SPACE•CHL A ≠ 0	yes
	SPACE•BNA ≠ 0	yes
	(SPACE•CHL A)•BNA ≠ 0	yes
	(SPACE•BNA)•CHL A ≠ 0	yes
	(CHL A•BNA)•SPACE = 0	yes
	(SPACE•BNA)•CHL A ≤ SPACE•BNA	yes
	(SPACE•CHL A)•BNA ≤ SPACE•CHL A	yes
	(SPACE•CHL A) x (SPACE•BNA) ≈ (CHL A•BNA)	0.263 ≈ 0.258

Fig. 4. Among the seven possible models of causal relationships for the variables SPACE, CHL A, and BNA, the first six are not supported by the data since some of the computed results do not agree with the predictions of the models.

ably predominantly of continental origin rather than endemic to the marine lagoon is distributed along a spatial gradient which is independent of that of the phytoplankton biomass. This absence of a causal link between CHL A and BNA could mean that the BNA bacteria are unable to use phytoplankton products to support their own growth in the marine environment. Either they are impaired by salinity, or the organic matter content produced by phytoplankton in the marine lagoon is too dilute compared to what is found in the eutrophied freshwaters where these bacteria are usually found growing in large numbers.

Since the partial Mantel test is a new technique to most ecologists, we tried to see if the same results could be obtained with a more standard technique. Instead of using the dissimilarity matrices (represented by capital roman letters), we removed the effect of *Chl a* from the *Bna* variable by linear regression and computed the residual *Bna* variability. From the residual *Bna*, we computed a dissimilarity matrix as explained in the methods and compared that to the geographic distance matrix with a simple Mantel test. The Mantel statistic was 0.134 with a significant probability value of 0.00190, which shows again a linear effect of spatial location in *Bna* remains after controlling for the influence of *Chl a*.

In order to explain the spatial gradient of the BNA bacteria, we added the particulate organic carbon variable (POC) into the analysis. Since it is not yet known how to compute partial Mantel tests while controlling for several matrix variables, or the equivalent of multiple regression involving dissimilarity matrices (for path analysis), we resorted to computing three-variable models for all combinations of the four variables BNA, CHL A, POC, and SPACE. The results of these analyses, based on the Mantel

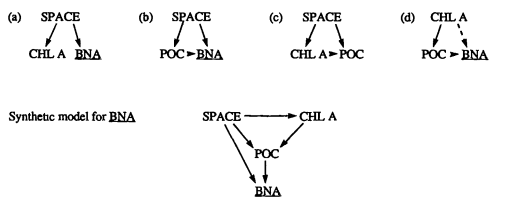

Fig. 5. The four three-variable models (a–d) are assembled into a synthetic model for the bacteria growing on bioMérieux nutrient agar (target distance matrix BNA, underlined). Significant causal links are indicated by arrows (dashed arrow weakly significant).

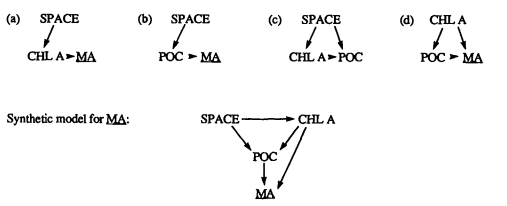

Fig. 6. As Fig. 5, but for the bacteria growing on marine agar (target distance matrix MA, underlined).

test results reported in Table 3, are presented in Fig. 5. One model has been selected in each case as being supported by the data, after a reasoning process similar to that illustrated in Fig. 4. Model a has already been discussed above. In models b, c, and d, the mathematics did not dictate the direction of the arrows that are drawn between variables; it dictated only their presence or absence. Drawing causal arrows from CHL A and from POC to BNA, and not the other way around, comes from the fact that BNA is the target variable of the model: in other words BNA is the variable whose variations we want to explain. It is not meant to negate the possible positive influence that the BNA bacteria may have on either CHL A or POC, be it by liberating metabolites that can help phytoplankton growth or by creating particulate organic matter (which would increase *POC*) through their metabolism or decay. The arrows from CHL A to POC on the other hand describe the fact that phytoplankton is most likely a part of the amount of particulate organic carbon that was measured, so that the variations of CHL A cause at least in part the variations observed in POC and not the opposite. If it had three full arrows, model d, which does not take spatial structure into account, would be one of the classical models of aquatic bacterial ecologists; the dashed arrow in model d from CHL A to BNA means that the partial Mantel test (CHL A·BNA)·POC in Table 3 is only weakly significant, being very close to the Bonferroni-corrected level of significance.

The four three-variable models lead to the synthetic model shown in Fig. 5. This model contains all the causal arrows of the smaller models, except for the weaker (dashed) link from CHL A to BNA because that link disappears when one controls for the effect of SPACE (model a). The CHL A variable has no direct influence on the BNA bacteria (as shown and discussed above), its influence being included in and measured through the influence of the particulate organic carbon (POC variable). Part of the POC gradient may reflect continental run-offs, explaining why BNA is related to POC. And, despite the significant partial influence of POC on BNA (partial Mantel statistic = 0.168, when controlling for SPACE), the variability in the spatial gradient of the BNA that remains unexplained, as measured by the partial Mantel statistic (SPACE·BNA)·POC = 0.462, is much more important.

So much for the *Bna* bacteria. Let us examine now the factors influencing the abundance variability of the bacteria growing on marine agar (variable *Ma*), which are presumably mostly of marine origin. We do not mean to say, of course, that all bacteria growing on marine agar are necessarily of marine origin, since we have shown that some bacteria from an urban wastewater treatment center can grow on marine agar (Troussellier 1987). Mean values for the *Bna* and the *Ma* bacteria (Table 1), however, show that the *Bna* bacteria can at best contribute little to the *Ma* counts.

Data analysis was carried out in the same way as for the *Bna* bacteria. The chief feature of these models (Fig. 6) is that the spatial gradient of the MA distance matrix, displayed in Fig. 2b and demonstrated in the spatial correlogram as well as in the simple Mantel test (Table 4, upper triangle), is to-

tally "explained away" by CHL A and POC. This situation is opposite to that observed for the BNA bacteria, where most of the spatial gradient remained unexplained after controlling for CHL A or for POC. Again, if the POC spatial gradient is the result of the phytoplankton gradient plus a hydrological gradient reflecting continental runoffs, it may explain why the MA bacteria that can use organic particulate substrate are still significantly related to POC, even when controlling for the effect of CHL A.

Conclusions

We have clearly shown here that even highly significant correlations among environmental variables can be spurious (*Bna·Chl a*) while others can be real (*Ma·Chl a*), so that it is important to pursue the analysis further, examining other hypotheses. When data represent samples taken from a piece of geographic space, as is so often the case in environmental studies, the analysis of the spatial structure itself is a good way of detecting spurious correlations, spatial position being a good candidate for generating such false relations because of the autocorrelated nature of most environmental variables. We have also shown how spatial position can be introduced as a variable in the study of relationships in microbial ecology and in particular in the framework of modeling.

Even though the relationship between total counts of bacteria (AODC) and phytoplankton has been demonstrated a number of times in aquatic environments, we have failed to find it in the study of bacteria responding to bioMérieux nutrient agar (BNA distance matrix). These bacteria, which are presumably of continental origin, have been shown to display a spatial gradient that can only partly be explained by the particulate organic carbon variable (POC) and not at all by phytoplankton biomass (variable CHL A), despite the existence of a spurious correlation between BNA and CHL A. This spatial gradient is probably linked to the diffusion of continental water through the lagoon; the partial explanation of the variations of the BNA bacteria by the POC variable would then be that POC is partly a tracer of continental water. On the other hand, the spatial gradient of the environmental heterotrophs growing on marine agar (MA distance matrix), which are expected to be mostly of marine origin, can be entirely explained by POC and CHL A. This study has then shown that different segments of the bacterial community that both react positively to variations of particulate organic carbon may vary in their dependence on phytoplankton biomass.

References

BIRD, D. F., AND J. KALFF. 1984. Empirical relationships between bacterial abundance and chlorophyll concentration in fresh and marine waters. Can. J. Fish. Aquat. Sci. **41:** 1015–1023.

BIVAND, R. 1980. A Monte Carlo study of correlation coefficient estimation with spatially autocorrelated observations. Quaest. Geogr. **6:** 5–10.

BÖLTER, M., AND OTHERS. 1981. Structure analysis of shallow water ecosystems: Interaction of microbiological, chemical and physical characteristics measured in the overlying waters of sandy beach sediments. Estuarine Coastal Shelf Sci. **13:** 579–589.

BOX, G. E. P., AND G. M. JENKINS. 1970. Time-series analysis, forecasting and control. Holden-Day.

CAMMEN, L. M., AND J. A. WALKER. 1982. Distribution and activity of attached and free-living suspended bacteria in the Bay of Fundy. Can. J. Fish. Aquat. Sci. **39:** 1655–1663.

CAUWET, G. 1983. Distribution de COD et COP en Méditerranée occidentale. Rapp. P.-V. Cons. Int. Explor. Mer 28, p. 101–105.

———. 1984. Automatic determination of dissolved organic carbon in sea water in the sub-parts per million range. Mar. Chem. **14:** 297–306.

CLIFF, A. D., AND J. K. ORD. 1973. Spatial autocorrelation. Pion.

———, AND ———. 1981. Spatial processes: Models and applications. Pion.

COOPER, D. W. 1968. The significance level in multiple tests made simultaneously. Heredity **23:** 614–617.

DE NEUFVILLE, R., AND J. H. STAFFORD. 1971. Systems analysis for engineers and managers. McGraw-Hill.

DOUGENIK, J. A., AND D. E. SHEEHAN. 1975. SYMAP user's reference manual. Lab. Computer Graphics Spatial Anal. Grad. School Design, Harvard Univ.

FERGUSON, R. L., AND A. V. PALUMBO. 1979. Distribution of suspended bacteria in neritic waters south of Long Island during stratified conditions. Limnol. Oceanogr. **24:** 697–705.

FRY, J. C., N. C. B. HUMPHREY, AND T. C. ILES. 1981. Time-series analysis for identifying cyclic components in microbial data. J. Appl. Bact. **50:** 189–224.

FUKAMI, K., U. SIMIDU, AND N. TAGA. 1981. Fluctuation of the communities of heterotrophic bacteria during the decomposition process of phytoplankton. J. Exp. Mar. Biol. Ecol. **55:** 171–184.

———, ———, AND ———. 1983. Change in a bacterial population during the process of degradation of a phytoplankton bloom in a brackish lake. Mar. Biol. (Berl.) **76:** 253–255.

———, ———, AND ———. 1985. Microbial decomposition of phyto- and zooplankton in seawater. 2. Changes in the bacterial community. Mar. Ecol. Prog. Ser. **21:** 7–13.

GOWER, J. C., AND P. LEGENDRE. 1986. Metric and Euclidean properties of dissimilarity coefficients. J. Classif. **3:** 5–48.

IBANEZ, F. 1981. Immediate detection of heterogeneities in continuous multivariate, oceanographic recordings. Application to time series analysis of changes in the bay of Villefranche sur Mer. Limnol. Oceanogr. **26:** 336–349.

JUMARS, P. A. 1978. Spatial autocorrelation with RUM (Remote Underwater Manipulator): Vertical and horizontal structure of a bathyal community. Deep-Sea Res. **25:** 589–604.

———, D. THISTLE, AND M. L. JONES. 1977. Detecting two-dimensional spatial structure in biological data. Oecologia **28:** 109–123.

KIRCHMAN, D., B. PETERSON, AND D. JUERS. 1984. Bacterial growth and tidal variation in bacterial abundance in the Great Sippewissett Salt Marsh. Mar. Ecol. Prog. Ser. **19:** 247–259.

LEGENDRE, L., AND P. LEGENDRE. 1983. Numerical ecology. Develop. Environ. Model. V. 3. Elsevier.

———, AND ———. 1984. Ecologie numérique, 2e éd. Tome 2: La structure des données écologiques. Collect. Écol. 13. Masson.

LEGENDRE, P., AND L. LEGENDRE. 1988. Structure spatio-temporelle des variables en océanographie: Problèmes d'analyse numérique et méthodes d'analyse spatiale, in press. *In* S. Frontier [ed.], Biométrie et océanographie. Actes Colloq. Biom. Oceanogr. Soc. Fr. Biom.

LILLIEFORS, H. W. 1967. The Kolmogorov-Smirnov test for normality with mean and variance unknown. J. Am. Stat. Assoc. **62:** 399–402.

LINLEY, E. A. S., R. C. NEWELL, AND M. I. LUCAS. 1983. Quantitative relationships between phytoplankton, bacteria and heterotrophic microflagellates in shelf waters. Mar. Ecol. Prog. Ser. **12:** 77–89.

MACKAS, D. L. 1984. Spatial autocorrelation of plankton community composition in a continental shelf ecosystem. Limnol. Oceanogr. **29:** 451–471.

MAHLOCH, J. L. 1974. Comparative analysis of modeling techniques for coliform organisms in streams. Appl. Microbiol. **27:** 340–345.

MANTEL, N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Res. **27:** 209–220.

MATHERON, G. 1971. The theory of regionalized variables and its applications. Cah. Centre Morphol. Math. Fasc. 5. ENSMP, Paris.

MILLER, R. G., JR. 1977. Developments in multiple comparisons. J. Am. Stat. Assoc. **72:** 779–788.

MIYOSHI, H., AND K. NAKAMOTO. 1975. Factors influencing bacterial distribution in the sea of Hiuchi-Nada area. Bull. Jpn. Soc. Sci. Fish. **41:** 645–652.

MORAN, P. A. P. 1950. Notes on continuous stochastic phenomena. Biometrika **37:** 17–23.

NEVEUX, J., AND M. PANOUSE. 1987. Spectrofluorometric determination of chlorophylls and pheophytins. Arch. Hydrobiol. **109:** 567–581.

ODEN, N. L. 1984. Assessing the significance of a spatial correlogram. Geogr. Anal. **16:** 1–16.

SMOUSE, P. E., J. C. LONG, AND R. R. SOKAL. 1986. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. Syst. Zool. **35:** 627–632.

SOKAL, R. R. 1979. Ecological parameters inferred from spatial correlograms, p. 167–196. *In* G. P. Patil and M. L. Rosenzweig [eds.], Contemporary quantitative ecology and related ecometrics. Stat. Ecol. Ser. V. 12. Int. Co-op. Publ.

———, AND N. L. ODEN. 1978. Spatial autocorrelation in biology. 1. Methodology. Biol. J. Linn. Soc. **10:** 199–228.

———, AND F. J. ROHLF. 1981. Biometry, 2nd ed. Freeman.

TROUSSELLIER, M. 1987. Dynamique des populations et des communautés bactériennes sous les effets des facteurs environnementaux le long d'un gradient aquatique continent-étang saumâtre. Thesis, Univ. Sci. Technol. Languedoc, Montpellier.

———, P. LEGENDRE, AND B. BALEUX. 1986. Modeling of the evolution of bacterial densities in an eutrophic ecosystem (sewage lagoons). Microb. Ecol. **12:** 355–379.

VÄÄTÄNEN, P. 1982. Effects of freshwater outflows on microbial populations in the Tvärminne archipelago, southern Finland. Holarct. Ecol. **5:** 61–66.

WRIGHT, R. T., AND R. B. COFFIN. 1983. Planktonic bacteria in estuaries and coastal waters of northern Massachusetts: Spatial and temporal distribution. Mar. Ecol. Prog. Ser. **11:** 205–216.

YODER, J. A., C. R. MCCLAIN, J. O. BLANTON, AND L.-Y. OEY. 1987. Spatial scales in CZCS-chlorophyll imagery of the southeastern U.S. continental shelf. Limnol. Oceanogr. **32:** 929–941.

Submitted: 19 October 1987
Accepted: 4 March 1988
Revised: 22 June 1988