PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Biological Studies in the Vicinity of a Shallow-Sea Tidal Mixing Front VI. A General Statistical Study

J. Y. Kassab, C. J. Whitaker and G. D. Floodgate

Phil. Trans. R. Soc. Lond. B 1985 **310**, 521-554
doi: 10.1098/rstb.1985.0130

Phil. Trans. R. Soc. Lond. B **310**, 521–555 (1985) [521]
Printed in Great Britain

BIOLOGICAL STUDIES IN THE VICINITY OF A SHALLOW-SEA TIDAL MIXING FRONT VI. A GENERAL STATISTICAL STUDY

By J. Y. KASSAB[1], C. J. WHITAKER[1] and G. D. FLOODGATE[2]

[1] *Statistics Unit, University College of North Wales, Bangor, Gwynedd LL57 2UW, U.K.*
[2] *Department of Marine Biology, University College of North Wales, Marine Science Laboratories, Menai Bridge, Anglesey, Gwynedd LL59 5EH, U.K.*

(*Communicated by G. E. Fogg, F.R.S.* – *Received* 6 *August* 1984 – *Revised* 10 *April* 1985)

CONTENTS

Studies of the distributional properties, the interrelationships and comparisons in time and space of 21 biological, chemical and physical variables that characterize the activities in a shallow sea tidal mixing front in the Western Irish sea are presented. They represent an attempt at describing and interpreting biologically this complex ecosystem as a whole and particularly to assess and compare the intensity of biological and biochemical activities and differences in distribution of organisms between the upper and lower stratified and mixed water columns on the two sides of the front at

different times of the year. The analyses used were mainly parametric methods, but non-parametric analyses were found to be appropriate in a few cases.

The log-normal distribution tended to fit better than the normal for most variables for each water mass within each cruise. Also different discrete distributions were fitted by the method of maximum likelihood to the bacterioplankton and zooplankton data and the best fits in both cases and for each cruise, where adequately large data was available, turned out to be the negative binomial distribution.

Some of the associations between the variables for separate water masses in each cruise, described by the non-parametric Spearman rank correlations, had meaningful biological interpretations while others did not. Also structural simplification through reducing the dimensionality (15 variables) of the system produced, by using principal component analysis on logarithmically transformed data, a few components that persisted throughout the cruises in the upper stratified water and could be interpreted in ecological terms; notably components showing the effect of physical stratification on biological activity, the depletion of nitrogenous compounds by plankton and the possible effect of protozooplankton grazing.

Comparisons of the levels of biological and biochemical activities determined by parametric, with data logarithmically transformed, and non-parametric one-way analyses of variance (ANOVA) showed significant differences between these levels particularly glucose and urea uptake rates, consistently in the three water masses and in all cruises. The dominant feature was that the upper stratified water was different from the other two water bodies and that the two methods of analysis produced similar results.

The relative importance of the biological variables to differentiate the water masses was assessed by using stepwise discriminant analysis, on data transformed logarithmically, and this confirmed, to a large extent, the results obtained from the analysis of variance comparisons.

Differences between the three lifestages of zooplankton numbers were ascertained by using a randomized block ANOVA on logarithmically transformed data which indicated that these differences were significant in all but one of the eight cruises where data was available. Significantly greater abundance of zooplankton haul numbers were found at the front and the stratified side compared with the mixed side. The diurnal variation of zooplankton numbers modelled by multiple regression analysis with data again transformed logarithmically showed that the numbers depended on depth in the stratified water column but on time in the mixed water column.

The analyses overall showed that the upper stratified water is an area of intense biological activity especially in the vicinity of the front and on the whole has many different characteristics from the rest of the water body and that most variables are closely linked, particularly during stable stratification in the summer.

1. Introduction

An ecosystem can be described in various ways, each of which conveys a greater or lesser amount of information. A simple, verbal account provides the smallest amount of detail. Quantification of each parameter or variable in the system adds further knowledge, while a mathematical analysis generates further insight into the ways in which the variables are distributed, and the relationships between them.

A generalized scheme for the structural analysis of large data sets from marine ecosystems, which deal with temporal and spatial sampling, has been put forward by Bolter *et al.* (1980). Earlier reports (Cassie 1962*a*, *b*, 1963) have described the use of various statistical methods including multivariate techniques to model phyto- and zooplankton distributions in a variety of marine and estuarine conditions. They stress that the data are more often log-normal rather

than normal in distribution, so that models based on the assumption of normality have diminished validity. Nor can the same degree of precision be assumed throughout a sampling programme. Cushing (1962) notes that bad weather can greatly increase the variation between replicates!

Only a few analyses of the distribution of the bacterial component of the plankton have been attempted though the study of bacterial distributions has been established for many years (Kriss 1962). Fisher (1941) first suggested that the bacteria in the water column could best be described by the negative binomial distribution, though only recently has the suggestion been followed up, for example, by El Shaarawi *et al.* (1981) who used it for modelling bacterial density in Lake Erie. Earlier Ashby & Rhodes-Roberts (1976) had found evidence of patchiness in the distribution of bacteria in inshore waters. The negative binomial distribution has previously been found to be the best model for freshwater zooplankton (Colebrook 1960) and has also been fitted to marine zooplankton (Comita & Comita 1957). Rosswall & Kvillner (1978) suggested that principal components and factor analysis should prove to be useful tools for elucidating microbial interactions in both water and soil, while Vaatanen (1980) has applied factor analysis to the microbial communities inhabiting the brackish waters off the coast of southern Finland.

The first five papers in this series (Fogg *et al.* 1985*b*; Egan & Floodgate 1985; Lochte 1985; Turley 1985; Scrope-Howe & Jones 1985) each dealt with a specific biological and biochemical aspect of the ecosystem. This paper deals with the statistical analysis and biological interpretation of the entire data set for the system as a whole. As such the emphasis in this paper is on the general applicability and validity of the statistical techniques to all the variables in the different water masses and for all cruises. At the same time the paper is intended to be read as a constituent part of the whole series. The main objectives of the study are listed below.

(i) To obtain information about the forms of the distributions of the organisms within the ecosystem, for different water columns at different times of the year. Of particular interest are the distributions of bacterioplankton and zooplankton numbers.

(ii) To describe the interrelationships between the biological, biochemical and physical variables within the system as a whole as stratification develops in early spring, stabilizes during the summer months and breaks up in the autumn. Also to determine whether a structural simplification of this complex ecosystem could be achieved by reducing its dimensionality to a few principal components which could be interpreted in meaningful ecological terms.

(iii) To assess the differences in the level of biological activity between the stratified side of the front, stabilized by the formation of a warm surface layer, and the mixed side, which is kept continually mixed by tidal action, and also to identify the relative importance of the variables within the ecosystem.

(iv) To answer various questions about zooplankton numbers such as comparisons between their lifestages, differences in their abundances at the front and the stratified and mixed sides and their diurnal variation over time and depth.

(v) To assess the differences overall between the stratified and mixed water columns.

Both parametric, on logarithmically transformed data, and non-parametric methods were used to answer these questions and the choice of a method was based on its validity and appropriateness to the data at hand.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

2. Statistical methods

The sampling programme has been, essentially, to make a 40 km transect across the front taking samples at various depths at a number of stations at successive stages during the front's development and break up in the years 1980 and 1981. Some preliminary calculations to determine the number of samples, n, needed to detect biologically important differences at the 5% significance level with 80% power for some of the continuous variables indicated that $n = 35$ would be adequate for sampling each of the upper stratified region and the rest of the water mass; giving a total of about 70 samples per cruise. Sampling zooplankton for diurnal variation, with n being much larger than 35, was carried out separately on cruise 3.

The data collected were transferred to the DEC-10 computer at U.C.N.W., Bangor, and were then checked to identify and correct any errors by reference to the original records. The analysis was performed by using various procedures of the statistical package SPSS (Nie *et al.* 1975) and the package MLP (Ross *et al.* 1980).

The distributional properties of the variables observed in this study were of intrinsic interest and these were examined with Kolmogorov–Smirnov tests (Zar 1974) on the raw and transformed data. The aim was also to ensure that the distributional assumptions necessary for the validity of the statistical analyses were met by the observed data.

Spearman rank correlation coefficients, which are measures of the monotonic relationship between pairs of variables, were calculated to show the association between variables. An alternative method, based on the z-transformed product moment correlation, was also used to indicate the important associations.

A principal component analysis, a descriptive multivariate technique, was used on logarithmically transformed data to try to obtain structural simplification by reducing the data from 15 variables to a smaller number of components which may then show underlying biological patterns in the data which could not otherwise be noticed. This procedure was used with varimax rotation of the components to make ecological interpretation of the generated components easier.

A Kruskal–Wallis non-parametric ANOVA was used to test whether there were differences between the three water masses for each of the 12 variables measured in the five cruises in 1980. A non-parametric method of multiple comparison was then used to determine where the differences lay, after an overall significant difference was found. Also parametric ANOVA was used on logarithmically transformed data and the results from the two methods of analysis compared.

Discriminant analysis, a predictive multivariate technique, is used to determine whether there is a difference between groups of observations. The groups here were the three water types defined by using a density measure σ_t, $\bar{v}$ and temperature (Fogg *et al.* 1985*b*). A discriminant analysis determines whether there is a difference between the water types with respect to the other variables measured. The purpose of a stepwise analysis is to give an ordering of importance to the variables in their ability to discriminate between the three water types.

ANOVA for a randomized block design was used to test whether there were differences between the numbers of three life stages of zooplankton on various cruises in 1980 and 1981. The stations were used as the blocking factor in the analysis. A Tukey's simultaneous confidence interval procedure for multiple comparisons, sometimes referred to as 'honest significant difference', was carried out whenever the F-test showed overall significance. This test, suggested

by Fisher and extended by Tukey is more conservative than for example Newman–Keuls or Duncan tests (Winer 1971) since it uses the same critical value (Q, studentized range) based on the number of treatments k for all tests while the other tests use critical values that depend on the number of steps (k or below) between pairs of means when arranged in order of magnitude. Tukey's test, however, has the advantage in that the probability of a type I error is at most equal to the stipulated significance level for all tests simultaneously.

A parametric one-way analysis of variance was applied to logarithmically transformed data for testing differences between the zooplankton haul number in the stratified, front and mixed water masses in 1980. Differences between these numbers for different stations were also tested by using the same procedure.

Multiple regression analysis on data transformed logarithmically was used to model the diurnal variation of zooplankton numbers on depth and time.

Finally, the method of maximum likelihood was used to fit various discrete distributions, including the negative binomial distribution, to the bacterial and zooplankton numbers for the cruises where sample data was adequately large. The calculations, based on a maximum likelihood iterative procedure for estimating for example the exponent K of the negative binomial owing to Bliss & Fisher (1953), which are usually involved and lengthy were made straightforward through the use of a routine in the programme MLP.

3. Results and discussion

(*a*) *Basic information*

Table 1*a* gives basic information about each of the five cruises in 1980. Readings were taken at four depths for each station sampled on every cruise. Because of various difficulties (for example, bad weather conditions and logistic constraints) some stations on the transect were missed but, where possible, transects were repeated or drogue stations were carried out on selected stations (Fogg *et al.* 1985*b*) which made up the sample numbers. After cruise 1 samples were taken at three constant depths while the lowest sample was taken at varying depths 5–10 m above the sea bottom.

The measurements made and the variables derived from them are shown in table 1*b* together with the abbreviations used in this paper.

(*b*) *Definition of the three water types*

The three water types are defined in table 2*a* by using the variables σ_t and $\bar{v}$ for cruises 1–4 and temperature alone for cruise 5 in 1980. Table 2*b* gives the number of observations for each water type together with the number of observations that cannot be classified according to the chosen criteria.

The different water masses are characterized by certain density (σ_t) ranges and they are separated from each other by density discontinuities either at the pycnocline or frontal interface. These density differences are primarily caused by temperature differences in the water masses due to different heat fluxes in the mixed and stratified water columns, which results in heating especially of the surface of the stratified water. Density was chosen in preference to temperature as a defining variable to incorporate any influences salinity may have on the density and to avoid the diurnal temperature fluctuations. Thus, it is possible to distinguish the surface of the stratified water (SSW) from the bottom stratified water (BSW) and the mixed water (MW)

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

TABLE 1. (*a*) TIME PERIOD, STATIONS AND DEPTHS SAMPLED ON THE FIVE CRUISES IN 1980

	cruise 1 $n = 52$	cruise 2 $n = 76$	cruise 3 $n = 84$	cruise 4 $n = 72$	cruise 5 $n = 40$
period...	12–13 March	29, 30 April and 1 May	3–6 June	15–17 July	23, 24 September
station	3–8, 10	1, 3–10	1–10	1–3, 5–10	1–10
depth/m	3.5–4.5	2	2	2	2
	15–20	10	10	10	10
	30–45	20	20	20–21	20
	48–90	46–90	50–75	40–75	45–100

TABLE 1. (*b*) LIST OF VARIABLES MEASURED AND THE DERIVED VARIABLES IN 1980

variable	symbol	unit
salinity	S	parts per thousand
temperature	T	degrees Celsius
nitrate concentration	$[NO_3]$	μmol N l^{-1}
nitrite concentration	$[NO_2]$	μmol N l^{-1}
ammonia concentration	$[NH_4]$	μmol N l^{-1}
urea concentration	U	μmol N l^{-1}
cellular adenosine triphosphate (ATP)	A	μg l^{-1}
chlorophyll *a* concentration	C	μg l^{-1}
phaeopigments concentration	P	μg l^{-1}
bacterial numbers	B_c	$10^8 \times$ cells per litre
bacterial cell volume	B_v	μm^3
zooplankton numbers (pump)	Z_n	numbers per litre
zooplankton biomass (carbon) (pump)	Z_b	μg C l^{-1}
urea uptake rate in the dark	U_d	d^{-1}
urea uptake rate in the light	U_l	d^{-1}
glucose uptake rate	G	d^{-1}
derived variables		
zooplankton haul numbers	H_n	numbers per litre
zooplankton haul biomass	H_b	μg C l^{-1}
$\bar{v}$	$\bar{v}$	J m^{-3}
bacterial biomass carbon	B_b	μg C l^{-1}
density	σ_t	

TABLE 2. (*a*) CRITERIA USED TO DEFINE THE THREE WATER TYPES IN THE FIVE CRUISES IN 1980

	stratified		
cruise	surface (SSW)	bottom (BSW)	mixed (MW)
1	$\sigma_t < 26.812$	$\sigma_t > 26.812$	$\sigma_t > 26.812$
	$\bar{v} < -9$	$\bar{v} < -9$	$\bar{v} > -9$
2	$\sigma_t < 26.701$	$\sigma_t > 26.701$	$\sigma_t > 26.701$
	$\bar{v} < -9$	$\bar{v} < -9$	$\bar{v} > -9$
3	$\sigma_t < 26.369$	$\sigma_t > 26.369$	$\sigma_t > 26.369$
	$\bar{v} < -9$	$\bar{v} < -9$	$\bar{v} > -9$
4	$\sigma_t < 26.001$	$\sigma_t > 26.001$	$\sigma_t > 26.001$
	$\bar{v} < -9$	$\bar{v} < -9$	$\bar{v} > -9$
5	$13.54 < T < 13.78$	$T < 13.54$	$T > 13.78$

TABLE 2. (*b*) NUMBER OF OBSERVATIONS IN EACH WATER TYPE

cruise	SSW	BSW	MW	unclassified
1	15	13	23	1
2	34	18	21	3
3	33	19	32	
4	31	13	21	7
5	13	4	23	

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

by an appropriately chosen critical density value. The latter two water masses are indistinguishable by density ranges. They were further separated by the variable $\bar{v}$, which is a measure of the energy required to bring about total vertical mixing of the water column and is based on the density differences within the water column (Simpson *et al.* 1977). A vertical separation of the water masses between BSW and MW could thus be achieved. During the season with well developed stratification this approach afforded satisfactory definition of the different water types. During the last cruise the stratification was partly broken down by wind mixing and a decrease in solar heating (Fogg *et al.* 1985*b*) and the water masses could no longer be defined by density ranges. It was possible, however, to distinguish them still by their different temperature ranges.

These three water masses represent the dominant feature of this study. The forms of the distributions of the variables under consideration, their interrelationships and their power to differentiate between the water masses are analysed and discussed in the sections below.

Sample sizes necessary for obtaining the required precision (80% power at 5% significance level) were to a large extent achieved. The exceptions being for cruises 1 and 5 where adverse weather conditions resulted in fewer samples collected.

(*c*) *Descriptive measures and distributional properties*

The basic elementary statistics for nearly all the variables measured in the three water masses were given in the earlier papers. To avoid repetition we give in tables 3*a* and 3*b*, for information only, these statistics for the data overall in each of the 1980 cruises. It should be pointed out that most of the statistical analyses in this study are related to data for separate water masses.

Table 3*a* gives the mean, standard deviation, minimum, maximum and number of missing values for each variable. Table 3*b* gives the same statistics as table 3*a* for the derived variables. It should be noted that zooplankton haul numbers, biomass and $\bar{v}$ are calculated by using the entire water column at a station so the sample sizes are reduced to 13, 19, 21, 18 and 10 for cruises 1, 2, 3, 4 and 5 respectively, in 1980.

In general, the values of the variables measured are, with a few exceptions, unremarkable. Some variables were relatively constant over the five cruises (for example, salinity), some rose consistently (for example, temperature, as expected) while some showed a rise and fall. In addition to the mean varying over the five cruises the standard deviation also varied in the same direction as the mean for most variables which together with the relative sizes of the mean and standard deviation brings into question the normality of many of the variables.

Most of the nutrient values fall within the range that has been observed in the Irish Sea over many years. One exception is the upper end of the nitrate concentration range, which is higher than expected, particularly for cruises 3 and 4. These high values will of course have increased the mean and standard deviation. Similarly the pattern of ammonium concentration is aberrant, and indicates a shift towards the more oxidized species compared with previous years. It should, however, be noted that the figures in the two tables are means over all three water masses. Reference to table 8 of the paper by Fogg *et al.* (1985*b*) shows that the concentration of nutrients can vary by a factor of four between the surface and the bottom water.

The cellular ATP concentrations also have a wide range which can be attributed to the fact that a few large algal or protozoan cells can greatly increase this measurement in any one sample. Since the photometric–luciferase method is capable of detecting very small amounts of living material, the zero values recorded as the minimum for each cruise are probably due to inadequacies in the sampling and extraction procedure.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

TABLE 3. (*a*) ELEMENTARY STATISTICS FOR EACH VARIABLE ON THE FIVE CRUISES IN 1980

	cruise 1 $n = 52$	cruise 2 $n = 76$	cruise 3 $n = 84$	cruise 4 $n = 72$	cruise 5 $n = 40$
S	34.43, 0.10 34.12, 34.59	34.33, 0.14 33.99, 34.64	34.22, 0.12 34.02, 34.35	34.26, 0.06 34.17, 34.37	34.49, 0.08 34.20, 34.67
T	7.98, 0.12 7.67, 8.17	8.45, 0.32 7.68, 9.19	10.29, 0.88 8.44, 12.08	12.20, 0.96 9.55, 13.44	13.88, 0.58 12.11, 14.55
$[NO_3]$	5.10, 1.68 2.00, 8.22	3.32, 1.86 0.09, 7.55 (1)	3.00, 2.21 0.00, 7.79	2.67, 2.29 0.00, 8.43	3.74, 1.24 1.34, 7.06
$[NO_2]$	0.43, 0.20 0.08, 0.72	0.43, 0.11 0.06, 0.71	0.35, 0.10 0.06, 0.50	0.33, 0.15 0.03, 0.62	0.39, 0.17 0.13, 0.80
$[NH_4]$	1.37, 0.72 0.17, 3.56	1.89, 1.04 0.20, 4.28 (1)	1.25, 0.67 0.40, 3.07	1.71, 0.93 0.39, 5.59	1.78, 0.93 0.30, 4.01
U	1.68, 0.63 0.94, 3.77	1.39, 0.60 0.00, 3.20 (2)	1.21, 0.66 0.09, 4.24	1.26, 0.45 0.32, 3.08	1.38, 0.62 0.55, 4.28 (4)
A	0.49, 0.70 0.00, 2.81	4.32, 8.59 0.00, 55.40 (2)	3.17, 4.96 0.00, 34.34 (4)	17.21, 19.10 0.00, 86.17 (4)	†
C	†	1.23, 0.64 0.30, 3.50 (7)	0.90, 0.07 0.10, 2.60 (14)	1.29, 1.66 0.10, 11.20 (27)	0.55, 0.48 0.00, 1.60 (5)
P	†	0.23, 0.29 0.00, 1.30 (9)	0.20, 0.19 0.00, 1.00 (14)	0.26, 0.24 0.00, 1.20 (27)	0.43, 0.47 0.00, 1.80 (5)
B_c	3.65, 2.34 0.00, 12.58	18.78, 21.74 1.20, 93.49 (1)	19.47, 20.19 1.59, 106.1	6.15, 5.37 0.87, 32.17 (1)	4.22, 3.40 0.24, 16.04
B_v	0.18, 0.06 0.00, 0.33 (1)	0.19, 0.04 0.12, 0.31 (3)	0.20, 0.03 0.10, 0.27	0.18, 0.04 0.11, 0.30 (1)	0.17, 0.05 0.10, 0.32
Z_n	0.32, 0.48 0.01, 1.40 (40)	1.41, 1.64 0.02, 8.92 (1)	2.48, 3.55 0.12, 17.48 (4)	1.57, 1.72 0.20, 7.36 (2)	†
Z_b	1.37, 1.13 0.03, 3.35 (40)	39.83, 60.74 0.01, 267.5 (1)	101.1, 154.2 0.42, 820.5 (4)	43.58, 89.18 0.56, 489.5 (2)	†
U_d	0.034, 0.019 0.008, 0.100 (3)	0.058, 0.049 0.004, 0.213 (1)	0.503, 0.683 0.010, 2.439 (2)	0.251, 0.240 0.029, 0.935 (3)	0.140, 0.108 0.019, 0.505 (4)
U_l	0.049, 0.033 0.008, 0.130 (5)	0.110, 0.099 0.007, 0.417 (1)	0.861, 1.096 0.026, 3.448 (5)	0.398, 0.343 0.024, 1.515 (7)	0.138, 0.105 0.021, 0.481
G	0.022, 0.010 0.009, 0.048 (1)	0.320, 0.294 0.012, 1.357 (4)	0.191, 0.138 0.033, 0.529 (2)	0.438, 0.304 0.091, 1.089 (2)	0.068, 0.057 0.013, 0.243

Data are in the form: mean, s.d.
minimum, maximum
(number of missing values).

† No data available.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

TABLE 3. (*b*) ELEMENTARY STATISTICS FOR THE DERIVED VARIABLES ON THE FIVE CRUISES IN 1980

	cruise 1 $n = 52$	cruise 2 $n = 76$	cruise 3 $n = 84$	cruise 4 $n = 72$	cruise 5 $n = 40$
B_b	6.84, 12.05 0.00, 31.30 (1)	37.81, 44.43 2.00, 198.4 (3)	40.92, 44.77 2.30, 211.0	12.14, 12.65 1.50, 60.70 (1)	7.49, 7.19 0.43, 32.97
σ_t	26.85, 0.07 26.72, 26.97	26.70, 0.12 26.44, 26.93	26.31, 0.23 25.87, 26.69	25.99, 0.20 25.70, 26.50	25.84, 0.13 25.63, 26.28
	For the variables listed below the sample sizes are:				
	$n = 13$	$n = 19$	$n = 21$	$n = 18$	$n = 10$
H_n	0.02, 0.02 0.01, 0.03 (11)	1.98, 1.04 0.28, 3.75	1.76, 1.21 0.56, 4.32	0.77, 0.47 0.26, 2.21 (2)	0.40, 0.11 0.24, 0.59
H_b	2.06, 2.88 0.02, 4.10 (11)	226.4, 202.9 13.40, 852.2	217.5, 207.6 3.30, 693.0	60.06, 65.32 1.77, 273.8 (2)	13.35, 26.29 0.82, 76.21
$\bar{v}$	−11.1, 12.1 −35.8, 5.7	−13.2, 32.9 −107.5, 2.6	−81.0, 73.7 −175.6, 1.2	−36.3, 32.4 −76.3, 0.1	−11.2, 14.1 −30.4, 0.7

Data are in the form: mean, s.d.
minimum, maximum
(number of missing values)

Variables that are likely to be contagiously distributed such as bacterial numbers, zooplankton numbers and chlorophyll concentration all have wide ranges as expected. Mean abundances of zooplankton are within previous values obtained in the Irish sea. The large standard deviation is due to differences in abundance in samples taken from the stratified and mixed waters. Both zooplankton numbers and biomass peak in June. The zero number of bacteria shown during cruise 1 was due to one sample with an unusually low count and should be taken to indicate that the true count was less than 10^8 bacteria per litre. The true value would have been found had a larger volume of water been filtered for the count.

The distributional forms of the variables within the water masses and for each cruise were determined by using the Kolmogorov–Smirnov test. The purpose, besides the fact that these distributions are of interest in themselves, was to determine an ad hoc scale on which the distributions are sufficiently near normal for parametric analyses to be valid. It should be added that the method of fitting a suitable model may also be thought of in the generalized linear model context as that of establishing a variance function for the variable, by using, for example, the statistical package GLIM (Baker & Nelder 1978). This proved somewhat difficult to attempt here since the sample sizes for separate water masses are too small. The best-fitted continuous forms for the non-derived variables, based on the *P*-value for the Kolmogorov–Smirnov test, are given in table 3*c* below. Only salinity and temperature are normally distributed throughout, while the rest tend generally to have a log-normal distribution. In the few cases where the square root transformation provided the best fit, this was only marginally better than the logarithmic transformation. Furthermore the logarithmic transformation of the data tended to increase the homogeneity of variance within the water masses for nearly all the variables and was far more effective than the square root transformation.

It should be pointed out that in many cases where either the logarithmic or square root

transformation provided distributions closer to the normal the null hypothesis of a normal distributional form for untransformed data could not be rejected at the 5% significance level. This could well be due to samples being too small for the test to discriminate effectively between the various forms.

It should also be pointed out that when using the logarithmic and square root transformations we added 1 to the values of the variables owing to the presence of a (small) number of zeros in the data. Adding a constant value of 1 to all the variables, though arbitrary, nevertheless allows comparisons of the distributional forms both within and between cruises. However, methods for determining the value of c in, say, $\lg(y+c)$ are available (Atkinson 1983).

TABLE 3. (*c*) DISTRIBUTIONAL PROPERTIES OF VARIABLES BASED ON THE KOLMOGOROV–SMIRNOV TEST

	cruise 1			cruise 2			cruise 3			cruise 4			cruise 5		
variable	SSW	BSW	MW	SSW	BSW	MW	SSW	BSW	MW	SSW	BSW	MW	SSW	BSW	MW
S	N	N	N	N	N	N	N	N	N	N	N	N	N	—	N
T	N	N	N	N	N	N	N	N	N	N	N	N	N	—	N
$[NO_3]$	L	L	L	L	L	N	?	N	N	L	L	N	N	—	N
$[NO_2]$	S	S	N	N	N	L	L	N	N	L	N	L	L	—	N
$[NH_4]$	L	L	N	N	S	L	L	S	L	L	L	L	L	—	L
U	L	L	L	L	L	N	L	L	L	L	N	S	L	—	L
A	?	L	S	L	L	L	L	L	S	S	L	S	—	—	—
C	—	—	—	L	L	L	L	N	L	L	L	L	N	—	N
P	—	—	—	N	L	?	N	L	N	S	L	L	L	—	L
B_c	L	L	L	L	L	L	S	S	S	L	L	L	L	—	L
B_v	S	N	N	L	L	L	N	N	N	L	L	L	L	—	L
Z_n	N	L	L	L	L	L	L	L	L	L	L	L	—	—	—
Z_b	N	N	N	L	S	L	S	S	L	L	S	L	—	—	—
U_d	L	N	L	N	L	N	N	L	L	L	N	N	N	—	L
U_l	N	N	L	N	L	L	N	L	S	N	L	L	N	—	L
G	L	L	L	L	L	S	L	L	?	N	L	?	N	—	L

N, untransformed data are normal.
L, logarithmically transformed data are normal.
S, square-root transformed data are normal
?, None of the above two transformations normalize the data.
—, Data are not available or the sample sizes are too small.

In reporting the elementary statistics in tables 3*a* and 3*b* the standard deviation accompanying the mean is an appropriate measure of scatter of the individual observations only in the cases where the variables are normally distributed, in such cases reporting the standard error as a measure of variability between means would also be appropriate. In the case of a transformed variable it would be more meaningful to report the mean together with the (non-symmetrical) confidence interval (for the mean) calculated from transformed data and then the limits changed back to the original scale (Sokal & Rohlf 1973). In the few remaining cases where the form of the distribution is unknown, it is doubtful if any formalized measure of scatter will have much descriptive meaning except perhaps one based on percentiles, for example, semi-interquartile range. Therefore no single measure of variability is wholly appropriate for the entire data set. Because of this and since we are interested in indicating the 'scale effect', that is, how the standard deviation varies as the mean varies over time for different cruises, partly justifies quoting the standard deviation in tables 3*a* and 3*b*.

(*d*) *Correlation relationships between the variables*

As a result of the findings about the non-normal forms of the distributions of most of the variables in the previous subsection Spearman rank correlation was regarded as a more appropriate measure than the product moment correlation to describe the association between pairs of variables. Correlation matrices were then obtained for each of the three water types on all cruises in 1980, except for the BSW on cruise 5 as there were only four observations. The correlation coefficients, both positive and negative, that were significant at the 5% and 1% levels were extracted from the matrices and these are given in tables 4*a* to 4*e*.

There is in general a great deal of variability between the coefficients both between water types within a cruise and between cruises. In fact a marked increase is noticeable in the number of significant correlations in all three water types as stratification develops, though this is less pronounced in the BSW. Nearly every one of the 17 variables is correlated with at least one other variable in the SSW and MW of cruises 2, 3 and 4. There is no clear reason why this should be so, but it might be expected that, since so many of the variables are interdependent, then as the temperature rises, and stratification develops, so more variables are pulled into step with each other, thus producing a high degree of close linkage.

To interpret the detailed pattern is, however, rather difficult, and although a few correlation coefficients are comprehensible in a biological or biochemical context, other are not. Some examples, where it is possible to explain significant correlations, are as follows.

When chlorophyll concentration and its degraded phaeopigments correlate significantly, they do so negatively as expected and in the SSW and MW only, presumably because in the BSW the light level is too low for algal growth. There is a significant negative correlation between temperature and nitrate in the SSW in cruises 2, 3 and 4 (also in the BSW in cruise 4 and in the mixed water in cruise 5) and the most likely cause of this is the removal of this inorganic nitrogen compound by the phytoplankton during the spring bloom (Fogg *et al.* 1985*b*; Turley 1985). Temperature is significantly positively correlated with urea uptake rate in the dark and in the light and with glucose uptake rate mainly in the stratified water in cruises 2, 3 and 4, and this may be an indication of the rates of metabolic activity of phytoplankton and bacterioplankton being considerably higher in the warmer water mass (Lochte 1985; Turley 1985). Also, as expected, zooplankton numbers and zooplankton biomass are significantly positively correlated, and so are urea uptake rates in the dark and light.

In contrast glucose uptake rate, which might be expected to depend on the concentration of the bacteria to a large extent, only shows a significant correlation in the SSW of cruises 4 and 5. Other correlations are equally puzzling. The density-related variable σ_t correlated negatively with the glucose uptake rate in the SSW of cruise 2. Cruise 3 shows these two variables significantly correlated in the BSW and MW, again negatively. By cruise 4 there is still a negative value in the BSW, but a positive one in the SSW. Another example is that in the SSW temperature is highly positively correlated with chlorophyll in cruise 2, becomes negatively correlated but not significant in cruise 3, then highly significantly negatively so in cruise 4 and returning to positive significance in cruise 5. Yet another unexpected association is the recurring significant correlation between zooplankton numbers and glucose uptake rate, in all three water masses in cruise 2 and in the MW in cruise 4.

Are these correlations spurious, or are they a pointer to mechanisms that are not understood as yet? It would seem that correlation matrices are, in this situation, a not very adequate

TABLE 4. THE SIGNIFICANT SPEARMAN RANK CORRELATION COEFFICIENTS FOR THE THREE WATER TYPES ON EACH CRUISE IN 1980

SSW		BSW		MW	
+ve	−ve	+ve	−ve	+ve	−ve
(a) cruise 1					
σ_t ** S	σ_t * A	σ_t ** S	$[NO_3]$ ** $[NO_2]$	σ_t ** S	S * G
T * U	S * A	S ** T	$[NO_3]$ * $[NH_4]$	S * T	T * $[NH_4]$
B_c * U_l	$[NO_3]$ ** $[NO_2]$	$[NO_2]$ * $[NH_4]$		S * U	T ** G
	$[NH_4]$ * G	U_d ** U_l		$[NH_4]$ ** G	$[NO_3]$ ** $[NO_2]$
		U_d ** G		B_v * U_d	$[NO_3]$ * U
		U_l ** G		U_d ** U_l	$[NO_2]$ * U
					$[NO_2]$ * B_v
					U * G
(b) cruise 2					
σ_t ** S	σ_t ** $[NO_3]$	$[NH_4]$ * U	σ_t ** T	σ_t ** S	S * A
S ** T	σ_t ** Z_n	C ** G	$[NO_3]$ * $[NH_4]$	S * T	T ** U
S * C	σ_t ** G	B_c ** B_v	$[NO_3]$ * C	T * $[NH_4]$	T * A
S ** U_l	S ** $[NO_3]$	Z_n ** U_d	$[NO_3]$ * G	$[NO_3]$ * V_v	$[NO_3]$ ** $[NO_2]$
T ** C	S * $[NO_2]$	Z_n * G	$[NO_2]$ * P	B_c * B_v	$[NO_3]$ * Z_n
T ** U_d	S * U	U_d * U_l	Z_b ** U_d	Z_n ** Z_b	$[NO_3]$ ** G
T ** U_l	S ** G	U_l ** G		Z_n ** G	$[NH_4]$ * U
$[NO_3]$ * $[NH_4]$	T ** $[NO_3]$			Z_b ** G	U ** B_c
$[NO_3]$ * U	T * $[NO_2]$			U_d ** U_l	A * B_v
$[NO_3]$ * G	$[NO_3]$ ** C				C * P
C * U_d	$[NO_3]$ ** U_d				C * U_d
C * U_l	$[NO_3]$ ** U_l				U_d * G
Z_n * Z_b	$[NO_2]$ ** C				
Z_n ** G	$[NO_2]$ * B_c				
U_d ** U_l	B_v * Z_n				
	B_v * Z_b				
(c) cruise 3					
σ_t ** S	σ_t ** T	T ** G	σ_t ** T	σ_t ** S	σ_t ** T
σ_t ** $[NO_3]$	σ_t * U_d	P * G	σ_t ** G	S * U_d	σ_t * $[NO_3]$
S ** C	σ_t ** U_l	B_c * B_v	S * B_v	S * U_l	σ_t * A
S * G	S ** T	B_v * G		T ** G	σ_t ** G
T * $[NH_4]$	S * U_l			$[NH_4]$ ** U	S * B_v
T ** U_d	T ** $[NO_3]$			$[NH_4]$ * G	S * Z_n
T ** U_l	$[NO_3]$ * U_d			A * C	S * Z_b
$[NO_3]$ ** $[NO_2]$	$[NO_3]$ * U_l			P * G	$[NO_2]$ * U_l
U * C	C * B_c			B_c * B_v	C ** P
U * G	B_c * Z_b			B_v ** Z_n	Z_n ** U_d
A * U_d				B_v ** Z_b	Z_b * U_d
C ** G				Z_n ** Z_b	
P * Z_n				U_l * G	
B_c ** B_v					
Z_n ** Z_b					
U_d ** U_l					
(d) cruise 4					
σ_t ** S	σ_t ** T	σ_t ** $[NO_3]$	σ_t ** T	σ_t ** S	σ_t ** T
σ_t ** $[NO_3]$	σ_t ** U_d	σ_t ** U_d	σ_t * $[NO_2]$	T * U	σ_t ** U
σ_t * $[NO_2]$	S ** T	T * $[NO_2]$	σ_t * G	T * P	σ_t ** P
σ_t * A	S ** U_d	T * U_l	T ** $[NO_3]$	$[NO_3]$ ** $[NO_2]$	S ** U
σ_t ** C	S ** U_l	T * G	T ** U_d	$[NO_3]$ ** U_d	S * Z_n
σ_t * B_c	T * $[NO_3]$	$[NO_2]$ ** $[NH_4]$	$[NO_3]$ * $[NO_2]$	U * P	S ** Z_b
σ_t * G	T * $[NO_2]$	$[NO_2]$ * G	$[NO_3]$ ** U_l	U ** Z_b	T * B_v
S ** $[NO_3]$	T ** C	$[NH_4]$ * B_v	$[NO_3]$ * G	P ** Z_n	$[NO_3]$ ** Z_n
S ** A	$[NO_3]$ ** U_d	U ** A	$[NH_4]$ * U_d	P * Z_b	$[NO_3]$ ** G
S ** C	$[NO_3]$ ** U_l	C * Z_n	A ** Z_n	Z_n * Z_b	$[NO_2]$ ** Z_n
S * G	$[NH_4]$ * U	B_c ** B_v	P * Z_b	Z_n ** G	$[NO_2]$ * Z_b

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

TABLE 4 (*cont.*)

SSW				BSW				MW			
+ve			−ve	+ve			−ve	+ve			−ve
(*d*) cruise 4											
T	*	U_d	U ** A	Z_n	*	U_d		Z_b * G			$[NO_2]$ * G
$[NO_3]$	*	A	U * C	U_d	*	G					$[NH_4]$ * U
$[NO_2]$	*	A	U ** B_c								U ** U_l
$[NH_4]$	**	B_c	U ** B_v								A * U_d
A	**	C	C * P								Z_n ** U_d
A	**	B_c									U_d * G
A	*	G									
C	*	B_c									
B_c	*	B_v									
B_c	**	G									
B_v	*	U_l									
Z_n	*	Z_b									
U_d	**	U_l									
(*e*) cruise 5											
σ_t	**	S	$[NO_3]$ * $[NO_2]$	only four measurements were made				σ_t * $[NO_3]$			σ_t ** T
T	*	C	$[NO_2]$ ** P					T ** $[NO_2]$			σ_t * $[NO_2]$
$[NO_3]$	*	P	$[NH_4]$ * P					$[NO_2]$ ** $[NH_4]$			T * $[NO_3]$
$[NO_2]$	**	$[NH_4]$	C * P					$[NO_2]$ ** G			$[NO_3]$ ** $[NO_2]$
$[NO_2]$	*	U_d						U ** U_d			$[NO_3]$ ** $[NH_4]$
C	*	U_d						U_d ** U_l			C ** P
B_c	**	G									B_v * G
U_d	**	U_l									

* $p < 0.05$; ** $p < 0.01$; +ve, positive correlation; −ve, negative correlation.

representation of the relationship between the variables. The situation is so complex and the sampling so infrequent that only very limited insights can be obtained by this method for this frontal system and, possibly, any other ecosystem that has a high turnover rate and may be linked or separated rapidly in time and space. It should also be pointed out that, on theoretical grounds, significance tests on these types of matrices are of doubtful validity since we are simultaneously testing a large number of coefficients (136 in fact) and so about seven coefficients could be expected, by chance, to reach the 5% significance level.

An alternative method to straightforward testing significance that allows for selection of important correlations in a large set was also used. The procedure consists of transforming the product moment correlation coefficients, by using Fisher's z-transformation, to produce approximate normality. The values are then plotted against the expected normal order statistics and the extreme ones in the tails indicate the important coefficients. The results are shown in figures 1*a* to 1*e*.

The number of important coefficients and their patterns are fairly similar to those given in tables 4*a* to 4*e*, where, essentially, some coefficients have meaningful interpretations while others have not in our present state of knowledge of the ecosystem.

(*e*) *Reduction of the variables to a few ecological components*

The varimax rotated component matrices of a principal component analysis on all five cruises are shown in tables 5*a* to 5*e*. The tables give only the relatively numerically large (0.35 or more or −0.35 or less, where 0.35 is the critical value for a correlation coefficient with 30 degrees of freedom) coefficients, or loadings, of linear functions defining the components. The numerical size of a loading corresponding to a variable reflects the relative importance of that variable

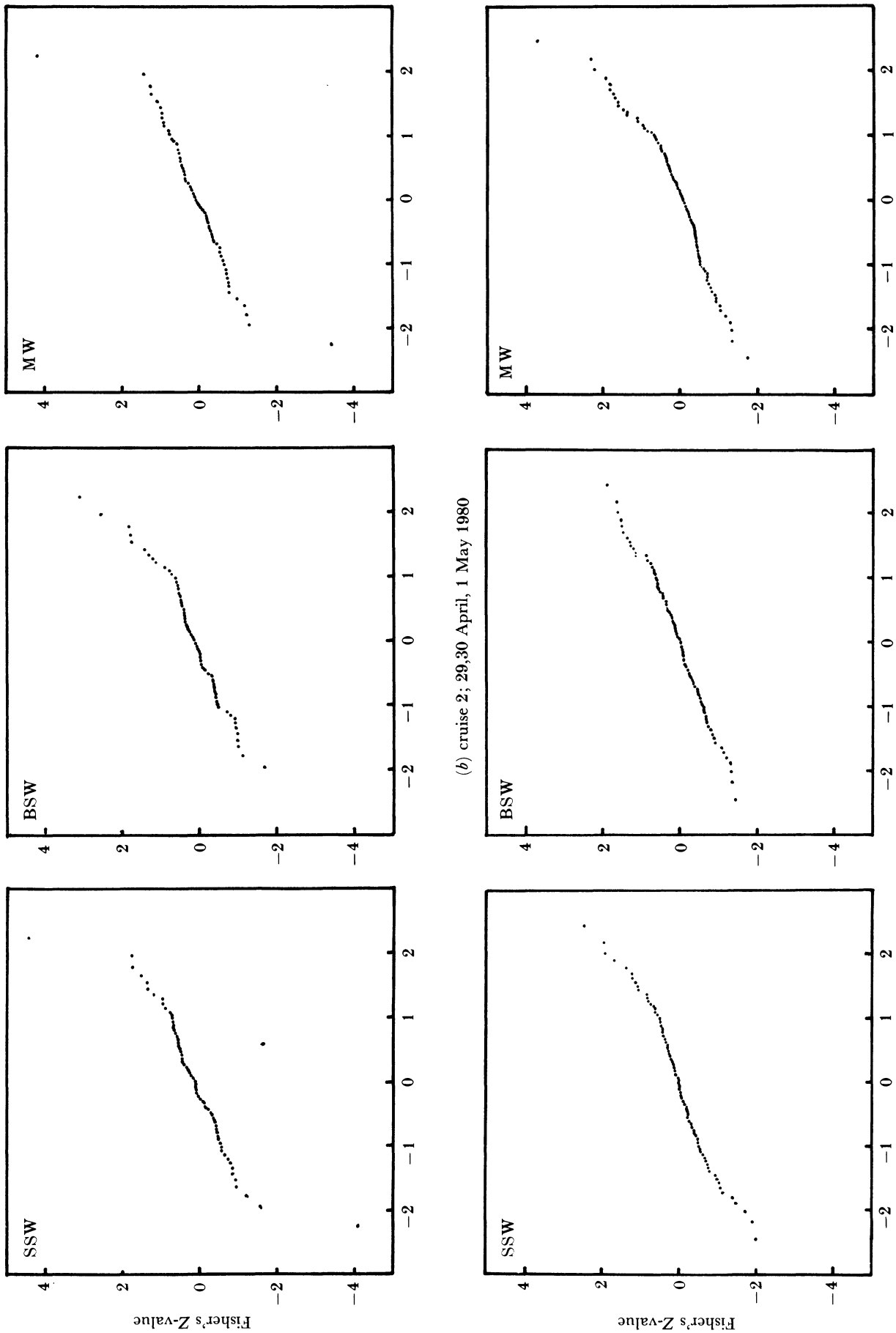

FIGURE 1 (*a*) and (*b*). For legend see page 536.

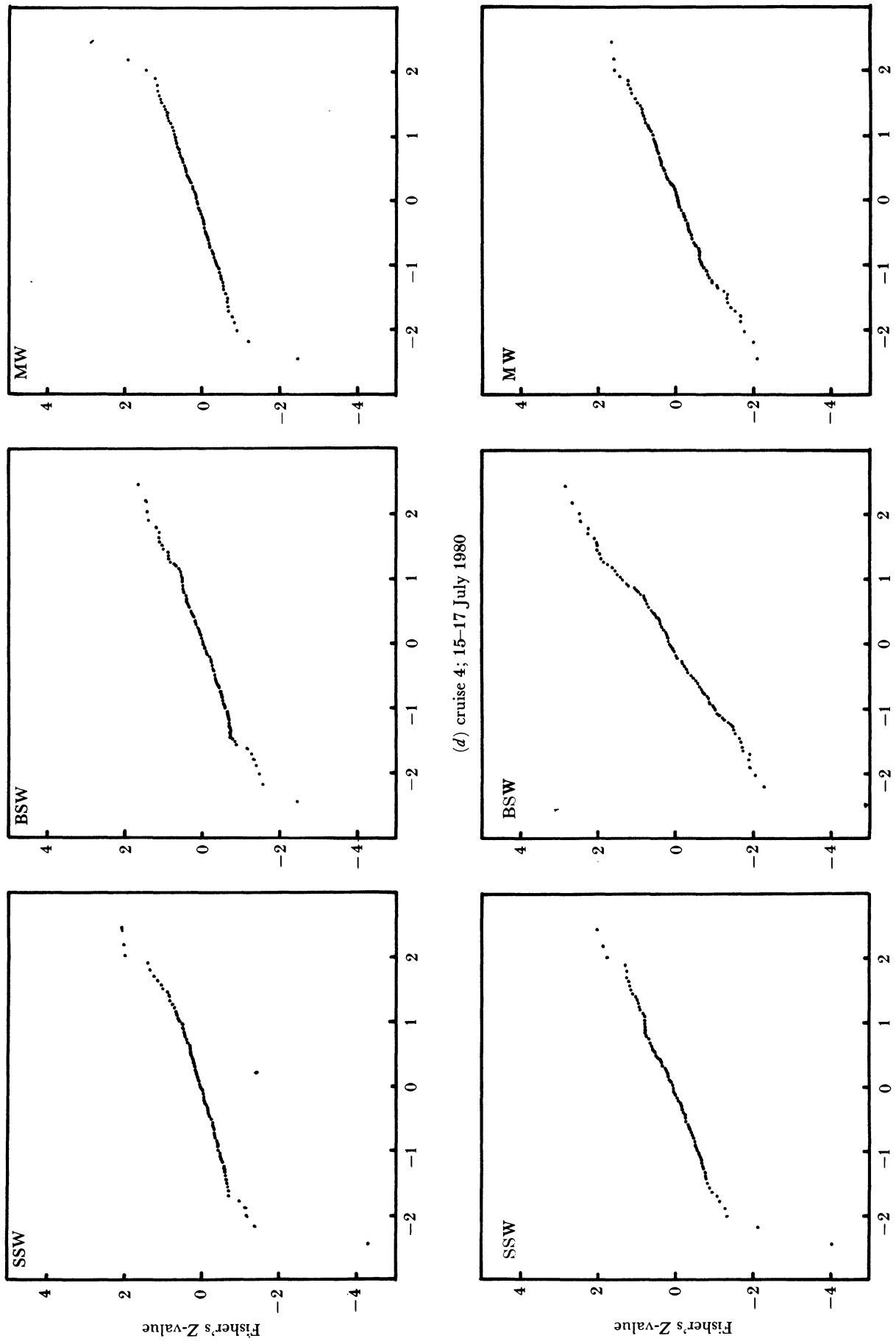

FIGURE 1 (c) and (d). For legend see page 536.

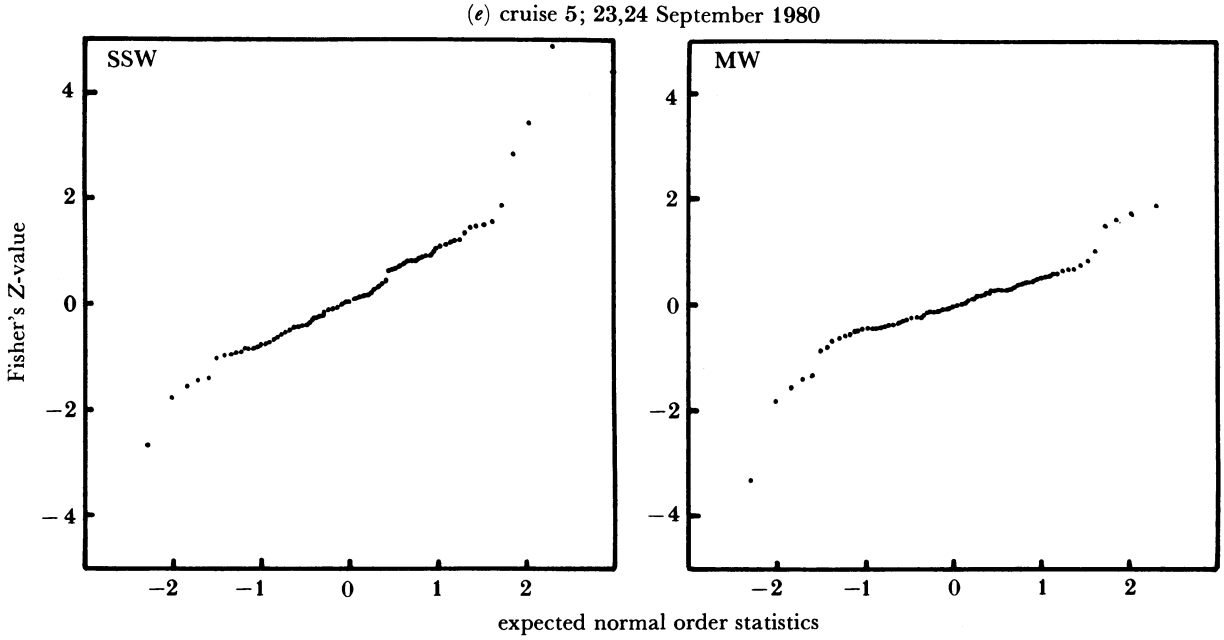

FIGURE 1. (*a*)–(*e*). Probability plots of Fisher's *z*-transformation of the product moment correlation coefficient for the water types during the five cruises in 1980 in the Western Irish Sea.

as a constituent of a component. The tables also include the eigenvalue corresponding to each component, the percentage variability explained by each component (100 × (eigenvalue)/number of eigenvalues)), cumulative percentage variability and the communality; the latter gives the proportion of variance of a variable accounted for by the components and is calculated as the sum of squares of component loadings for that variable. In each analysis only the first four (most important) components were retained as these had eigenvalues greater than the average value of one.

The number of observations were not sufficiently large in comparison with the number of variables to allow analysis on each water type (see table 2*b*). As a result, the bottom stratified and mixed waters were combined for cruises 2, 3 and 4 and all three water masses were aggregated for cruises 1 and 5 for 1980. It should be pointed out that the grouping together of the data for the BSW and the MW has some justification, as indicated by the first conclusion in the next section.

Because of the previous findings about the distributional properties, a logarithmic transformation to base 10 was used on the variables indicated to normalize the data, with 1 added to the raw data because of the presence of zero values. The density measure (untransformed) was included in preference to temperature and salinity for the reasons given in §2*b*, and the derived variables were left out for obvious reasons.

The results of the analysis show consistency in that the first four components presented in the tables explain between 65 % and 69 % of the total variability in the data for the SSW, while a slightly smaller variability (57–70 %) was accounted for by the first four components in the BSW and MW combined for the three middle cruises. Some of the possible ecological interpretations of the components, in physical, chemical and biological terms and based on the relative sizes and signs of the loadings are as follows.

(i) In cruise 1 (entire water mass) the first component explains 33.1 % of the total variability

in the data. This component is a contrast between σ_t on the one hand and urea uptake in the dark, in the light, and glucose uptake rate on the other. It represents the effect of stratification on the biological activity and may be labelled as 'biological stratification'. An explanation is that in early spring stratification of the water column was weak and especially in waters with high density; that is, in the deeper waters, biological activity, as measured by the uptake of urea and glucose, was low (Turley 1985; Lochte 1985).

The second component, which explains an additional 15.5% of the total variation, is a 'chemical contrast' which compares nitrate with nitrite and ammonia concentrations.

The third component accounts for a further 15.1% of the variability and is an index of urea, bacterial numbers and bacterial cell volumes; a mixture of chemical and biological variables, indicating some association.

The fourth component is a 'direct measure' of cellular ATP which is a constituent of all living matter but which rapidly decays when cells die.

The four components account for 72.9% of the total variability in the data for the 11 variables involved.

(ii) In the SSW of cruise 2, the 'biological stratification' component indicated in cruise 1 appears as the second component, which this time includes a high positive loading for zooplankton numbers.

The first component in this cruise accounts for 22.4% of the total variability in the data for the 15 variables included in the analysis. It is a contrast of the chemical measurements nitrate and, to a lesser extent, ammonia and urea concentrations with chlorophyll and the uptake rates of urea. It may be interpreted as the depletion of the nitrogenous compounds by phytoplankton and may be labelled as 'chemical stratification'.

Component 3 is a biological–chemical contrast comparing mainly nitrite with chlorophyll and bacterial numbers, which is not very easily understood since the concentrations of nitrite are too low to contribute very much to bacterial growth.

Component 4 is a contrast of bacterial cell volume and zooplankton biomass and, to a lesser extent, zooplankton numbers. This component also appears in later cruises but is often weak. However, as it occurs regularly it may be an indication of the removal of the larger bacterial cells by zooplankton grazing and may be labelled as 'biological contrast'. The grazing is not necessarily due to zooplankton itself but there may be an intermediate link; that is, protozooplankton, between bacteria and zooplankton (Lochte 1985).

The components in the MW and BSW combined are more difficult to interpret as the combination of the relatively important variables within each component varies from cruise to cruise. However, the first component here (cruise 2) is basically similar to component 1; that is, 'chemical stratification', found in the SSW for this cruise, but weaker.

(iii) In the SSW of cruise 3, component 1 may again be labelled as 'biological stratification'. It expresses the effect of the uptake of urea by phytoplankton. This is due to the lack of other nitrogen sources in the less dense surface waters (there is a negative loading of nitrate and nitrite in accordance with σ_t).

Component 2 shows the grazing effect referred to in cruise 2.

Component 3 is an index of a mixture of biological and chemical variables.

Component 4 is an index of chemical variables.

(iv) In cruise 4, component 1 in the SSW is essentially a biological–nutrient contrast, similar to the 'chemical stratification' component referred to in (ii) above. It suggests both bacterioplankton and algae growing, directly or indirectly, at the expanse of urea.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

TABLE 5. PRINCIPAL COMPONENT ANALYSIS FOR EACH CRUISE IN 1980 SHOWING THE VARIMAX ROTATED COMPONENT LOADINGS NUMERICALLY GREATER THAN OR EQUAL TO 0.35, THE COMMUNALITIES AND THE VARIABILITY EXPLAINED BY THE FIRST FOUR COMPONENTS

(L (variable) indicates that the variable was transformed by adding 1 and taking the logarithm to base 10.)

(*a*) cruise 1

entire water mass

variable	components 1	2	3	4	communality
σ_t	−0.78	—	—	—	0.62
L([NO_3])	—	−0.89	—	—	0.91
L([NO_2])	—	0.93	—	—	0.93
L([NH_4])	—	0.71	0.45	—	0.72
L(U)	—	—	0.68	—	0.54
L(A)	—	—	—	0.98	0.97
L(B_c)	0.36	—	0.63	—	0.52
L(B_v)	—	—	0.70	—	0.57
L(U_d)	0.79	—	—	—	0.69
L(U_l)	0.89	—	—	—	0.83
L(G)	0.79	—	—	—	0.70
eigenvalue	3.64	1.71	1.66	1.01	
percentage variance	33.1	15.5	15.1	9.2	
cumulative percentage variance	33.1	48.6	63.7	72.9	

(*b*) cruise 2

variable	surface stratified water components 1	2	3	4	communality	mixed and bottom stratified water components 1	2	3	4	communality
σ_t	—	−0.83	—	—	0.84	—	—	—	0.64	0.47
L([NO_3])	−0.76	—	0.41	—	0.85	−0.71	—	—	—	0.61
L([NO_2])	—	—	0.79	—	0.65	—	—	—	−0.79	0.68
L([NH_4])	−0.40	—	—	—	0.31	—	—	0.44	—	0.29
L(U)	−0.36	—	0.46	0.40	0.53	—	−0.38	—	−0.48	0.49
L(A)	—	0.45	−0.37	0.35	0.49	—	−0.47	—	—	0.32
L(C)	0.52	—	−0.64	—	0.78	—	0.72	—	—	0.70
L(P)	—	—	0.42	—	0.24	—	−0.59	—	0.44	0.68
L(B_c)	—	—	−0.56	—	0.32	—	0.66	—	0.45	0.68
L(B_v)	—	—	—	0.84	0.70	—	0.55	—	—	0.37
L(Z_n)	—	0.77	—	−0.40	0.87	0.79	—	—	—	0.67
L(Z_b)	—	—	—	−0.76	0.61	—	—	0.83	—	0.74
L(U_d)	0.85	—	—	—	0.78	—	—	−0.80	—	0.77
L(U_l)	0.90	—	—	—	0.81	0.75	—	—	—	0.74
L(G)	—	0.94	—	—	0.92	0.88	—	—	—	0.80
eigenvalue	3.35	2.71	2.00	1.64		3.19	2.20	1.99	1.64	
percentage variance	22.4	18.0	13.3	10.9		21.3	14.6	13.2	10.9	
cumulative percentage variance	22.4	40.4	53.7	64.6		21.3	35.9	49.1	60.0	

(*c*) cruise 3

variable	surface stratified water components 1	2	3	4	communality	mixed and bottom stratified water components 1	2	3	4	communality
σ_t	−0.70	0.39	—	—	0.69	—	—	—	0.81	0.77
L([NO_3])	−0.63	—	—	0.55	0.79	—	—	0.62	—	0.40
L([NO_2])	−0.43	—	—	0.74	0.75	−0.36	—	0.69	—	0.73
L([NH_4])	—	—	—	0.79	0.72	0.57	—	—	—	0.53
L(U)	—	0.71	—	—	0.55	—	—	—	—	0.20
L(A)	0.36	—	—	0.46	0.43	—	0.60	—	—	0.41
L(C)	—	0.81	—	—	0.74	—	0.51	−0.53	—	0.71
L(P)	—	—	0.65	—	0.50	0.63	—	0.38	—	0.58

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

TABLE 5. (*cont.*)

(*c*) cruise 3

variable	surface stratified water components 1	2	3	4	communality	mixed and bottom stratified water components 1	2	3	4	communality
L(B_c)	—	—	−0.60	—	0.50	—	0.72	—	—	0.53
L(B_v)	—	−0.48	—	—	0.40	0.38	0.62	—	—	0.62
L(Z_n)	—	—	0.82	—	0.69	0.65	—	—	—	0.48
L(Z_b)	—	—	0.78	—	0.66	0.67	—	—	0.38	0.61
L(U_d)	0.90	—	—	—	0.89	—	—	0.65	—	0.53
L(U_l)	0.86	—	—	—	0.79	—	—	—	−0.78	0.70
L(G)	—	0.77	—	—	0.67	0.87	—	—	—	0.81
eigenvalue	3.17	2.78	2.20	1.59		3.23	2.25	1.71	1.40	
percentage variance	21.1	18.6	14.7	10.6		21.6	15.0	11.4	9.3	
cumulative percentage variance	21.1	39.7	54.4	65.0		21.6	36.6	48.0	57.3	

(*d*) cruise 4

variable	surface stratified water components 1	2	3	4	communality	mixed and bottom stratified water components 1	2	3	4	communality
σ_t	0.46	−0.51	—	0.48	0.79	−0.88	—	—	—	0.84
L([NO_3])	—	−0.84	—	—	0.75	−0.56	−0.71	—	—	0.82
L([NO_2])	0.36	−0.48	—	—	0.52	0.75	—	—	—	0.73
L([NH_4])	0.46	—	—	—	0.40	—	—	0.76	—	0.73
L(U)	−0.78	—	—	—	0.61	—	0.40	—	0.66	0.66
L(A)	0.78	—	—	—	0.70	0.40	—	—	0.49	0.53
L(C)	0.59	—	—	0.65	0.82	0.83	—	—	—	0.82
L(P)	—	—	—	−0.89	0.85	—	—	—	0.74	0.60
L(B_c)	0.78	—	—	—	0.77	—	0.45	0.62	—	0.69
L(B_v)	0.60	—	0.44	—	0.55	—	—	0.80	—	0.64
L(Z_n)	—	—	−0.77	—	0.65	—	0.78	—	—	0.74
L(Z_b)	—	—	−0.83	0.43	0.87	−0.40	0.64	—	—	0.63
L(U_d)	—	0.84	—	—	0.76	−0.86	—	—	—	0.80
L(U_l)	—	0.86	0.35	—	0.90	0.53	—	—	—	0.42
L(G)	0.39	—	0.40	—	0.37	—	0.87	—	—	0.84
eigenvalue	3.98	3.22	1.69	1.44		4.45	2.75	1.77	1.54	
percentage variance	26.5	21.5	11.3	9.6		29.6	18.3	11.8	10.2	
cumulative percentage variance	26.5	48.0	59.3	68.9		29.6	47.9	59.7	69.9	

(*e*) cruise 5

variable	entire water mass components 1	2	3	4	communality
σ_t	—	—	0.78	—	0.73
L([NO_3])	—	−0.86	—	—	0.83
L([NO_2])	—	0.87	—	—	0.82
L([NH_4])	—	0.72	—	—	0.71
L(U)	—	—	—	0.80	0.68
L(C)	0.47	0.60	0.36	—	0.72
L(P)	—	—	−0.80	—	0.73
L(B_c)	0.62	—	—	0.45	0.69
L(B_v)	—	—	—	0.70	0.55
L(U_d)	0.92	—	—	—	0.88
L(U_l)	0.89	—	—	—	0.85
L(G)	0.70	0.35	—	—	0.63
eigenvalue	4.72	1.79	1.23	1.08	
percentage variance	39.4	14.9	10.3	9.0	
cumulative percentage variance	39.4	54.3	64.6	73.6	

Component 2 is 'biological stratification', indicating the effect of stratification on biological activity.

Component 3 is a 'biological contrast', showing the effect of zooplankton grazing on the bacterial population.

Component 4 is essentially a contrast between chlorophyll and its degraded form phaeopigments. It may be explained by the fact that in July the bloom was declining and degradation products were increasing.

(v) During the last cruise, wind mixing started to break down the physical stratification and for the first time there is no contrast between density-related σ_t with glucose and urea uptake rates. But some form of biological stratification is represented by component 3 which contrasts σ_t with phaeopigments.

Component 1 is an index of biological variables, component 2 is mainly a chemical contrast, while component 4 is an index of a mixture of biological and chemical variables.

(*f*) *Comparisons between the water masses*

In this subsection we shall compare the levels of biological and biochemical activities in the three water columns by using both non-parametric and parametric ANOVA, the latter on logarithmically transformed data. The validity of the assumptions, advantages and disadvantages of the two methods are discussed and the results compared. Also the use of the more recent methods of generalized linear models are referred to briefly.

(i) *Non-parametric ANOVA*

In general, non-parametric methods are less powerful than parametric ones, but are valid under less strict assumptions. The two basic assumptions of a one-factor parametric ANOVA are normality and homogeneity of variance within factor levels, here the three water masses. When these assumptions are not satisfied then, usually, the data are either transformed to meet these assumptions or non-parametric ANOVA is used. Table 3*c* shows that the form of the distribution of every variable changes within a cruise or from cruise to cruise and in few cases the variables cannot be transformed to the normal by taking logarithms or square roots. Furthermore Bartlett's test (Snedecor & Cochran 1980) has shown that heterogeneity of variance cannot be corrected by either of these transformations in the cases of bacterial cell volume, urea uptake in the light and glucose uptake rate in cruise 1; bacterial cell volume, glucose and urea uptake rates in cruise 2; nitrate, cellular ATP, chlorophyll, zooplankton numbers, glucose and urea uptake rates in cruise 3; chlorophyll, zooplankton numbers and urea uptake rates in cruise 4, and urea, bacterial cell volume, glucose and urea uptake rates in cruise 5. Therefore these violations of the assumptions, together with the fact that sample sizes are dissimilar (the latter reduces the robustness of the *F*-test further), would to some extent make the use of parametric ANOVA of questionable validity, at least for some of the variables. So to give one consistent test for the differences between the three water types on all variables (except for $\bar{v}$, σ_t, salinity and temperature, the variables that define the water types) a non-parametric analysis of variance was considered best. Even when such a test is used there is still the requirement that the distribution should have the same shape, but their medians can be different. This was only questionable for chlorophyll on cruise 4, where the distribution in the SSW was greatly influenced by one very high value of 11.2. The effect it had can be seen when the means are 1.41, 0.69 and 1.25 in the SSW, BSW and MW, respectively, while the corresponding medians are 0.717, 0.525 and 1.18 (table 6*d*).

For these reasons a Kruskal–Wallis non-parametric analysis of variance was used to test for differences between the three water types and the statistic derived from the test is compared with the χ^2 distribution with two degrees of freedom. When the three water types were significantly different overall at the 5% level, a non-parametric multiple comparison procedure, based on confidence intervals for ranks, was used (Miller 1966) to determine which water types differ from each other. The results of the multiple comparisons, together with the observed significance levels and the median values, are given in tables 6*a* to 6*e*. In these, for convenience, a shortened notation for the three water types was used with the letters S, B and M denoting the SSW, BSW and MW, respectively.

In the last column of the tables, where multiple comparisons are given, for example, in table 6*a*, S compared with M for the response variables zooplankton numbers, glucose and urea uptake rates means that the SSW is significantly different from the MW for these variables. Multiple comparisons were not performed when the Kruskal–Wallis test result overall was not significant at the 5% level.

The main conclusions to be drawn from these comparisons and their interpretation in biological and chemical terms over the five cruises may be summarized as follows.

(i) The most notable feature throughout the five cruises is that when there are differences it is usually the SSW that is different from the other two. Furthermore, during cruises 2, 3 and 4 there are more variables which show this difference and, in most cases, with higher significance (but larger sample sizes) than cruises 1 and 5. This indicates that the SSW becomes increasingly distinct in its chemical and biological properties from the adjacent water masses during the summer; cruise 1 showing only the onset of stratification, cruise 5 being affected by the erosion of the stratification and only intermediate cruises reflecting the situation during stable stratification. As discussed in Fogg *et al.* (1985*a*), Turley (1985) and Lochte (1985), an increasing degree of interdependent relationships between different chemical and biological components is likely to occur during the period of stable stratification.

(ii) The uptake of glucose by bacteria is the most sensitive measurement for differentiating water masses, with very high overall significance for all five cruises in 1980. When stratification was at its maximum (cruise 3), all three water types could be separated by this variable. The SSW and MW were separated on every cruise; SSW and BSW on two occasions and the BSW and MW on one. In all cases the glucose median values were highest in the SSW and lowest in the MW, with intermediate values occurring in the BSW.

(iii) The urea uptake rates, whether in the light or dark, were also very sensitive, but did not distinguish any of the water masses during the final cruise and, as for the glucose uptake rate, the SSW had the highest median values while the MW had the lowest. In contrast, the urea concentration did not separate the water types on any cruise.

(iv) Table 6*a* shows a significant difference between the SSW and the MW for zooplankton numbers, supporting the conclusions of Scrope–Howe & Jones (1985). As zooplankton populations increase in abundance significant separation of the SSW and BSW and the SSW and the MW is possible (table 6*b*). This is due to zooplankton increases in the SSW occurring before population increases in the MW. The later cruises 3 and 4 (there were no data for cruise 5) show that both zooplankton numbers and zooplankton biomass can be used to separate the water masses very effectively, with the former in the SSW being significantly higher than those in the BSW and MW, while the latter follows a different pattern where the significant differences are between the SSW and MW and between the BSW and MW.

(v) Since increased primary productivity is often cited as the prime cause of the different

TABLE 6. KRUSKAL–WALLIS ONE-WAY ANALYSIS OF VARIANCE BETWEEN THE THREE WATER TYPES IN EACH CRUISE IN 1980

(The figures in parentheses are the number of missing values. S, surface stratified water; B, bottom stratified water; M, mixed water. S *v.* M, median of a response variable in S is significantly different from the median in M, etc.)

(*a*) cruise 1

	medians for each water type				
variable	S $n = 15$	B $n = 13$	M $n = 23$	significance level	significant comparisons
$[NO_3]$	5.94	4.85	4.18	0.1	—
$[NO_2]$	0.330	0.430	0.532	n.s.	—
$[NH_4]$	1.14	1.20	1.60	n.s.	—
U	1.88	1.65	1.28	0.1	—
A	0.003	0.430	0.317	0.1	—
B_c	3.19	4.16	2.22	0.1	—
B_v	0.161	0.186	0.195 (1)	0.1	—
Z_n	0.600 (12)	0.065 (8)	0.006 (19)	0.05	S *v.* M
Z_b	2.01 (12)	1.320 (8)	0.580 (19)	n.s.	—
U_d	0.049	0.032 (1)	0.023 (2)	0.001	S *v.* M
U_l	0.085 (3)	0.053	0.027 (2)	0.001	S *v.* M
G	0.026 (1)	0.019	0.014	0.001	S *v.* M

(*b*) cruise 2

	medians for each water type				
variable	S $n = 34$	B $n = 18$	M $n = 21$	significance level	significant comparisons
$[NO_3]$	2.12	3.40 (1)	4.09	0.001	S *v.* M
$[NO_2]$	0.395	0.495	0.408	0.05	S *v.* B, B *v.* M
$[NH_4]$	2.02	1.95 (1)	1.73	n.s.	—
U	1.34 (1)	1.50 (1)	1.35	n.s.	—
A	1.32 (1)	0.435	0.395 (1)	0.05	S *v.* M
C	1.05 (4)	0.875 (3)	1.30	0.01	B *v.* M
P	0.175 (4)	0.200 (5)	0.038	n.s.	—
B_c	8.80 (1)	9.82	17.49	0.01	S *v.* M
B_v	0.179 (3)	0.195	0.182	n.s.	—
Z_n	1.60 (1)	0.540	0.199	0.001	S *v.* B, S *v.* M
Z_b	8.29 (1)	39.03	2.34	0.1	—
U_d	0.080	0.010 (1)	0.034	0.001	S *v.* B, S *v.* M
U_l	0.181 (1)	0.030	0.041	0.001	S *v.* B, S *v.* M
G	0.291 (3)	0.192 (1)	0.203	0.01	S *v.* M

(*c*) cruise 3

	medians for each water type				
variable	S $n = 33$	B $n = 19$	M $n = 32$	significance level	significant comparisons
$[NO_3]$	0.160	4.48	4.63	0.001	S *v.* B, S *v.* M
$[NO_2]$	0.273	0.430	0.400	0.001	S *v.* M, B *v.* M
$[NH_4]$	1.17	1.48	0.665	0.001	S *v.* B, B *v.* M
U	1.09	1.12	0.975	n.s.	—
A	2.68	1.65	1.24 (4)	n.s.	—
C	1.15 (3)	0.492 (6)	0.679 (5)	0.001	S *v.* B, S *v.* M
P	0.125 (3)	0.288 (6)	0.116 (5)	0.1	—
B_c	13.91	17.74	11.54	n.s.	—
B_v	0.201	0.203	0.202	n.s.	—
Z_n	3.69	0.880	0.420 (4)	0.001	S *v.* B. S *v.* M
Z_b	113.80	46.87	10.10 (4)	0.001	S *v.* M, B *v.* M
U_d	1.14	0.064	0.071 (2)	0.001	S *v.* B, S *v.* M
U_l	2.05 (1)	0.060 (2)	0.180 (2)	0.001	S *v.* B, S *v.* M, B *v.* M
G	0.318	0.158	0.055 (2)	0.001	S *v.* B, S *v.* M, B *v.* M

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

TABLE 6. (*cont.*)

(*d*) cruise 4

variable	medians for each water type S $n = 31$	B $n = 13$	M $n = 21$	significance level	significant comparisons
$[NO_3]$	0.240	4.71	3.69	0.001	S *v.* B, S *v.* M
$[NO_2]$	0.229	0.250	0.460	0.001	S *v.* M, B *v.* M
$[NH_4]$	1.34	1.04	1.59	n.s.	—
U	1.28	1.35	1.27	n.s.	—
A	19.06 (2)	1.73 (2)	13.65	0.01	S *v.* B, B *v.* M
C	0.717 (9)	0.525 (6)	1.18 (9)	0.1	—
P	0.217 (9)	0.175 (6)	0.250 (9)	n.s.	—
B_c	6.31 (1)	4.90	3.08	0.01	S *v.* M
B_v	0.179 (1)	0.164	0.168	n.s.	—
Z_n	2.19 (2)	0.430	0.590	0.001	S *v.* B, S *v.* M
Z_b	28.24 (2)	19.16	2.99	0.01	S *v.* M, B *v.* M
U_d	0.375	0.156	0.073 (2)	0.001	S *v.* B, S *v.* M
U_l	0.644 (5)	0.101 (1)	0.232 (1)	0.001	S *v.* B, S *v.* M
G	0.707 (1)	0.201	0.146 (1)	0.001	S *v.* B, S *v.* M

(*e*) cruise 5

variable	medians for each water type S $n = 13$	B $n = 4$	M $n = 23$	significance level	significant comparisons
$[NO_3]$	2.63	5.81	4.25	0.001	S *v.* B, S *v.* M
$[NO_2]$	0.497	0.215	0.348	0.05	S *v.* B
$[NH_4]$	2.39	2.08	1.12	0.001	S *v.* B
U	1.35 (3)	1.05 (1)	1.28	n.s.	—
C	1.05 (1)	0.300	0.367 (4)	0.01	S *v.* M
P	0.200 (1)	0.017	0.333 (4)	0.05	—
B_c	6.22	0.815	2.63	0.01	S *v.* B, S *v.* M
B_v	0.166	0.145	0.168	n.s.	—
U_d	0.247	0.102	0.089 (4)	n.s.	—
U_l	0.190	0.103	0.089	0.1	—
G	0.126	0.065	0.031	0.001	S *v.* M

biological structures of the stratified and mixed water bodies, chlorophyll might be expected to show marked differences within the water bodies. In fact, the ANOVA indicated that this variable had approximately the same sensitivity as bacterial numbers whose means show a high degree of homeostasis (Egan & Floodgate 1985). Bacterial cell volume showed no significant difference between the water masses in any of the five cruises.

(vi) Among the chemical variables measured, nitrate and nitrite concentrations showed a high consistent ability to differentiate the water types but cellular ATP and ammonia less so.

The above results suggest that the Kruskal–Wallis ANOVA technique is a useful method for analysing both the chemistry and the biology of the frontal system.

However, these results merely indicate which water masses are significantly different from each other and do not give a measure of the accuracy of these differences. This measure (that is, by how much the population medians differ) may be given by a confidence interval by using a Mann–Whitney U-test for equality of population medians for the variables in the three water masses (Sachs 1982). Briefly, the procedure is to set up, say, H_0: $M_S - M_M = c$ against H_1: $M_S - M_M \neq c$ (M_S and M_M are the population medians for a variable in the SSW and MW respectively) at the 5% significance level, add c to every observation for the data in the MW and carry out a Mann–Whitney U-test. Then the set of all values of c for which H_0 is not rejected

would give the 95% confidence interval for the difference $M_S - M_M$. These confidence intervals are straightforward but rather tedious to calculate and so are best obtained by using computers.

(ii) *Parametric ANOVA*

The application of the non-parametric ANOVA, in preference to the parametric one, above, was essentially due to the complexity of the ecosystem which necessitated making as few assumptions as possible. However, *F*-tests for the equality of factor level means are fairly robust to departures from the assumptions of normality and homogeneity of variance especially if sample sizes are equal (Box 1953) (the latter is not the case in our data). Furthermore, table 3*c* indicates that if any distributional form for our error terms is to be assumed then it would be the log-normal. Therefore, the results from a one-way parametric ANOVA on logarithmically transformed data were also considered to be of interest, at least for comparison purposes with those from the non-parametric ANOVA. Scheffe's multiple comparison procedure (Neter & Wasserman 1974) based on 95% simultaneous confidence intervals was used to identify where differences lay whenever the overall *F*-test was significant at the 5% level. This test is somewhat conservative but has the advantage that the probability of a type I error for all the simultaneous tests here does not exceed 5%. The results from the two different methods turned out to be quite similar, particularly in cruises 3 and 4 during stable stratification. In fact, when using parametric ANOVA the only differences are that ammonia is no longer significantly different in the three water masses in cruise 3, while in cruise 4 the BSW is not significantly different from the MW for cellular ATP and zooplankton biomass, which makes the parametric ANOVA somewhat less powerful. This interesting feature was also noticeable in the other cruises especially in the cases where heterogeneity of variance was not significant. On the other hand, parametric ANOVA was more powerful in the cases where heterogeneity was significant, particularly for glucose and urea uptake rates, which could be because violation of the homogeneity of variance assumption increases the probability of a type I error. These features provide further evidence for the usefulness of non-parametric ANOVA, at least for interpreting some biological aspects of this complex ecosystem.

It should be pointed out that the remark made in §3c about reporting the results when data are transformed applies equally here; that is, the means of the raw data for the three water masses should be reported with their confidence intervals calculated from logarithmically transformed data and then the limits changed back to the original scale.

(iii) *Generalized linear models*

When data are transformed to meet the assumptions of a parametric ANOVA, error distributions are sometimes compromised and inferences would apply only approximately to the original untransformed data. However, recently developed techniques called generalized linear models allow the construction of appropriate ANOVA with non-normal error distributions, for example, Poisson, gamma and negative binomial distributions. Furthermore, computational facilities for dealing with these techniques are now readily available through procedures in the statistical packages GLIM and GENSTAT (Alvey *et al.* 1977). We should add, however, that these methods are not wholly appropriate to our data since, as stated earlier, error distributions change within and between cruises. But preliminary investigations on a few appropriately selected variables have indicated that these methods give fairly similar results to those obtained by the standard two we have used above.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

(*g*) *Variables that discriminate between the water masses*

Tables 7*a* to 7*c* give the results of a stepwise discriminant analysis using Wilks' λ criterion on the three water types for cruises 2, 3 and 4 in 1980. Cruises 1 and 5 were not used as they had missing variables. Salinity, temperature, $\bar{v}$ and σ_t were not included in the analysis as these variables are used to define the three water masses (table 2*a*). Since there are only three groups (water masses), the maximum number of discriminant functions that can be obtained is two (the number of groups minus one). The values given in the tables are the weighting coefficients of the standardized variables defining linearly the discriminant functions; they indicate the relative contribution of each variable to the function. The stepwise ordering of the variables in the tables shows the importance with which the analysis selected them for their discriminant power, the first variable being the most important, the second variable the next most important, etc. All variables selected by the analyses were significant at the 0.1 % level for all three cruises. The centroids summarize the location of the three water masses in the space (graph) defined by the two funtions and are the mean discriminant scores for the observations within each water mass. The analyses for the three cruises are fairly consistent in that the first discriminant function separates out the SSW while the second separates out the BSW. All 14 variables were significant at least once in the three analyses but only cellular ATP, chlorophyll and zooplankton biomass were significant on all three analyses.

Since the ordering of the variables shows their importance in discriminatory power, it is interesting to compare them with the result of the Kruskal–Wallis analyses. Broadly the same characteristics are found in both tests. Glucose and urea uptake rates, zooplankton numbers and their biomass are the most discriminatory of the biological variables, as are nitrate and nitrite among the chemical ones. Chlorophyll and cellular ATP appear as more discriminatory in this case than in the ANOVA, while bacterial numbers less so.

(*h*) *Comparisons of the zooplankton lifestages*

At each station in a cruise the haul number of zooplankton in each of the lifestages nauplii, copepodite and adult were counted for the purpose of comparing them. The distributions of these numbers tended to be log-normal and the assumption of a randomized block ANOVA using the stations as the blocking factor were satisfied when the data were transformed to a logarithmic scale. The mean number of zooplankton for the stations in each lifestage for the entire water mass is given in table 8 together with the significance level for the overall differences between lifestages and multiple comparisons. Data were also available for cruises 7–10 made in 1981 (Fogg *et al.* 1985*b*) and so these were included in the analysis, while cruise 1 was left out because of too many missing values.

Table 8 shows that there are overall significant differences between the three life-cycle stages of copepods for all but one of eight cruises. A common feature of the results of the multiple comparison procedure based on Tukey's 95 % simultaneous confidence intervals given in the last column of table 8 shows that when there are differences then it is the intermediate lifestage copepodite (C) which is different from the other two in all cases except for cruise 9 where the number of nauplii (N) is significantly higher than the number of adults (A). It should be added that these confidence intervals indicate not only where the differences between the population means lie but also provide a measure of the accuracy of the differences with 95 % confidence.

Comparisons between stations sampled during the same cruise have been used elsewhere

TABLE 7. STEPWISE DISCRIMINANT ANALYSIS FOR CRUISES 2, 3 AND 4 IN 1980

(L (variable) indicates that the variables were transformed by adding 1 and taking the logarithm to base 10.)

(*a*) cruise 2

variable ordering	standardized discriminant functions 1	2
L(U_1)	0.61	0.57
L(B_c)	−0.41	0.63
L(Z_n)	0.57	−0.20
L(B_v)	0.21	−0.71
L(C)	−0.64	0.52
L(NO_2)	−0.36	−0.07
L(Z_b)	0.09	−0.42
L(NO_3)	−0.25	0.41
L(A)	0.28	0.15

centroids for each water type

	function 1	2
SSW ($n = 23$)	1.85	0.18
BSW ($n = 10$)	−0.76	−1.53
MW ($n = 20$)	−1.74	0.55

84.8 % of observations were classified in the correct water type.

(*b*) cruise 3

variable ordering	standardized discriminant functions 1	2
L(G)	1.25	0.57
L([NH_4])	0.56	0.56
L(Z_b)	0.24	0.71
L(P)	−0.42	0.04
L(Z_n)	0.70	−0.51
L(C)	−0.57	−0.43
$L(U_1)$	−0.67	−0.88
L(U_d)	1.18	0.15
L(A)	−0.48	0.15

centroids for each water type

	function 1	2
SSW ($n = 29$)	2.80	−0.45
BSW ($n = 11$)	−0.61	2.43
MW ($n = 19$)	−3.92	−0.71

98.3 % of observations were classified in the correct water type.

(*c*) cruise 4

variable ordering	standardized discriminant functions 1	2
L([NO_3])	1.13	−0.40
L(Z_b)	−0.87	−0.31
L([NO_2])	0.34	0.71
L(U)	0.74	0.14
L(A)	−0.63	0.21
L(C)	0.29	0.58
L(B_c)	0.22	−0.54

centroids for each water type

	function 1	2
SSW ($n = 13$)	−2.73	−0.06
BSW ($n = 5$)	3.00	−1.94
MW ($n = 8$)	2.57	1.31

100 % of observations were classified in the correct water type.

TABLE 8. COMPARISON OF LIFESTAGES OF ZOOPLANKTON HAUL NUMBERS ($\lg(H_n+1)$) BY USING A RANDOMIZED BLOCK ANOVA, WITH STATIONS AS BLOCKS, FOR THE ENTIRE WATER MASS FOR EIGHT CRUISES IN 1980 AND 1981

(N, nauplii; C, copepodite; A, adult. C *v.* A, population mean of c is significantly different from that of A, etc.)

cruise	number of stations	means for the lifestages N	C	A	significance level	significant comparisons
2	10	163.8	26.7	431.3	0.001	C *v.* A
3	10	1032.1	25.3	249.4	0.001	C *v.* N, C *v.* A
4	8	117.1	23.1	134.9	0.01	C *v.* A
5	10	51.0	13.3	20.7	0.01	C *v.* N
7	10	82.2	186.5	83.8	0.01	C *v.* N, C *v.* A
8	10	44.6	93.6	50.7	0.1	—
9	10	112.7	79.2	44.5	0.05	N *v.* A
10	10	29.0	61.6	11.5	0.01	C *v.* A

(Scrope–Howe & Jones 1985) to suggest that zooplankton peaks occurring on the fronts are due to population growth rather than mechanical aggregation.

It may be of interest to point out that the data were also analysed by using the analogous non-parametric ANOVA, namely Friedman's two-way ANOVA, and the results turned out to be fairly similar but the test was less sensitive and produced fewer significant comparisons as one would expect.

(*i*) *Comparison of zooplankton haul numbers in the stratified, front and mixed water types*

The number of zooplankton, as measured by haul numbers, in the stratified, front and mixed water types were counted for all 1980 cruises combined. The mean numbers are shown in table 9, as is a parametric analysis of variance on the logarithmically transformed data (the transformation was needed to normalize the data and equalize the variances) which shows significant differences at the 0.1 % level. Scheffe's multiple pairwise comparison procedure was used to show that the mean value in the mixed water was significantly lower than the stratified and the front water types, but there was no significant difference between the latter two. This was further confirmed by a similar procedure with the stations as the factor levels for cruises 2, 3 and 4.

This supports the conclusions reached by Scrope–Howe & Jones (1985) that zooplankton abundance and population peaks differ in the mixed and stratified waters, but that those of the front reflect changes in the stratified water closely for most of the year.

TABLE 9. PARAMETRIC ANALYSIS OF VARIANCE FOR THE HAUL NUMBERS OF ZOOPLANKTON ($\lg(H_n+1)$) IN THE STRATIFIED, FRONT AND MIXED WATER TYPES IN 1980

	stratified	front	mixed
means	1309.4	1659.4	581.0
n	43	14	43

$F = 17.5$, significance of $F = p < 0.001$.

The mean in the mixed water is significantly lower than that in the front and stratified waters. The difference between the latter two is not significant.

(*j*) *Diurnal variation of zooplankton numbers over time and depth*

The results of a multiple regression analysis to determine the relation between the number of zooplankton at different depths over a 36 hour period are shown in table 10. The two time series data sets were taken on cruise 3 and were at stations 2 and 9 in the stratified and mixed waters respectively. Station 2 was sampled 19 times at two-hourly intervals at depths of 2, 15, 20, 30, 60 and 80 m, while depths at station 9 were 2, 10, 20, 30, 40 and 50 m and were mostly sampled again at intervals of two hours. To normalize the data the logarithmic transformation of the number of zooplankton plus 1 was used and a stepwise regression analysis performed on the time, depth, square of time, square of depth and the product of time and depth. The analysis showed that for the stratified water column the relation depended significantly on the square of depth only, but for the mixed water column on time and the square of time. The relation for the stratified water had a higher coefficient of determination ($R^2 = 0.55$) compared to that for the mixed water ($R^2 = 0.17$), indicating that the data is a better fit to the model in the stratified water. Also in both water columns there was some indication that zooplankton numbers had a cyclical variation of approximately 24 h.

The usual tests for the validity of the assumptions of the normal error model, including graphical analysis of residuals, were performed. These assumptions were found to be reasonably satisfied except that in the mixed water column the Durbin–Watson test (Neter & Wasserman 1974) for autocorrelation indicated that the error terms were positively correlated over time and so the fitted model in this case should be treated with reserve.

Table 10. Stepwise regression analysis for zooplankton ($\lg(Z_n+1)$) diurnal variation on cruise 3 in 1980

(Independent variables: time, depth, time^2, depth^2, time × depth.)

water type	n	significant effects	F value	significance	R^2
stratified (station 2)	114	depth^2	137.9	0.001	0.55
mixed (station 9)	106	time^2	6.1	0.05	0.06
		time	10.2	0.001	0.17

	equation
stratified	$\lg(Z_n+1) = 2.53-0.00015 \times \text{depth}^2$
mixed	$\lg(Z_n+1) = 2.13-0.076 \times \text{time}+0.004 \times \text{time}^2$

An explanation of the above results is that vertical migration was occurring on the stratified side of the front when chlorophyll was low, so depth and changes in depth with time were the important features relating to zooplankton. Whereas on the mixed side of the front vertical migration did not occur because of tidal mixing so that zooplankton densities were related to time rather than depth. A further factor complicating the fluctuations in zooplankton numbers in both types of water is the patchy nature of phytoplankton distribution.

(*k*) *Distributions of bacterial and zooplankton numbers and their best fits*

In §3*c* we used the logarithmic and square-root transformations to obtain some information about the non-normal distributional forms of the variables. These transformations, particularly the logarithmic one, provided a good approximation to the distribution of the variables

bacterial and zooplankton numbers. In this subsection we shall fit by the method of maximum likelihood a number of possible discrete distributions to the data of these planktonic organisms and identify the ones that fit best.

The Poisson (purely random) distribution for bacterial numbers was tested first by using Fisher's index of dispersion $((n-1) \times \text{variance/mean})$ as the test statistic which has an approximate χ^2 distribution with $n-1$ degrees of freedom. This was found to be very highly significant, especially for cruises 2–5, indicating that the Poisson assumption is not tenable. The distributions were then fitted by the method of maximum likelihood to a number of contagious distributions such as the negative binomial, Neyman type A and Polya–Appeli. For all five cruises the negative binomial turned out to be the best fit to the observed distributions. It should be pointed out that some degrees of freedom (d.f.) were lost as a result of pooling adjacent classes in the right tail of the distribution to satisfy the theoretical requirement of a minimum expected frequency of five per cell. Also due to grouping the data for these analyses the mean values given in the table differ slightly from those given in table 3*a*. The details of the analysis, including the inverse of the estimated exponent K and the observed P-value for testing the negative binomial hypothesis, are given in table 11.

TABLE 11. FITTING THE NEGATIVE BINOMIAL DISTRIBUTION TO BACTERIAL CONCENTRATION BY THE METHOD OF MAXIMUM LIKELIHOOD FOR THE ENTIRE WATER MASS IN 1980

	cruise				
	1	2	3	4	5
n	52	75	84	71	40
mean	3.21	18.27	18.96	5.44	3.60
variance	5.10	357.94	340.30	19.93	9.68
Fisher's index	81.03	1449.78	1489.71	256.45	104.87
$1/K$	0.18	1.39	1.14	0.49	0.47
χ^2	5.35	5.56	8.44	5.68	1.18
d.f.	3	5	6	7	3
P-value	0.15	0.35	0.21	0.58	0.75

Figures 2*a* to 2*e* illustrate graphically the differences between the Poisson and the negative binomial distributions when fitted to the observed frequency distributions and show the inadequacy of the Poisson fit. In March (cruise 1) the bacterial numbers exhibit a unimodal pattern with a significant skewness to the right. But by the beginning of the summer when stratification had developed (cruises 2 and 3) the pattern changes markedly, becoming a reversed J-shaped distribution and with over a fivefold increase in the mean values (see also table 5*a*). By midsummer and early autumn (cruises 4 and 5) the numbers return gradually to their earlier (March) values but with greater patchiness.

Similar analyses were performed on zooplankton numbers. The Poisson assumption was again rejected at a very high significance level and the negative binomial provided the most satisfactory fit in each case (the near-perfect fit for cruise 3 is partly due to pooling of expected frequencies to meet the theoretical requirements). The data were again grouped and so the mean values given are slightly different from those in table 3*a*. Cruises 2, 3 and 4 only were considered here because there were no observations in cruise 5 and too few in cruise 1. The details of the analyses are given in table 12.

Figures 3*a* to 3*c* indicate that during stable stratification (cruises 2–4) the zooplankton

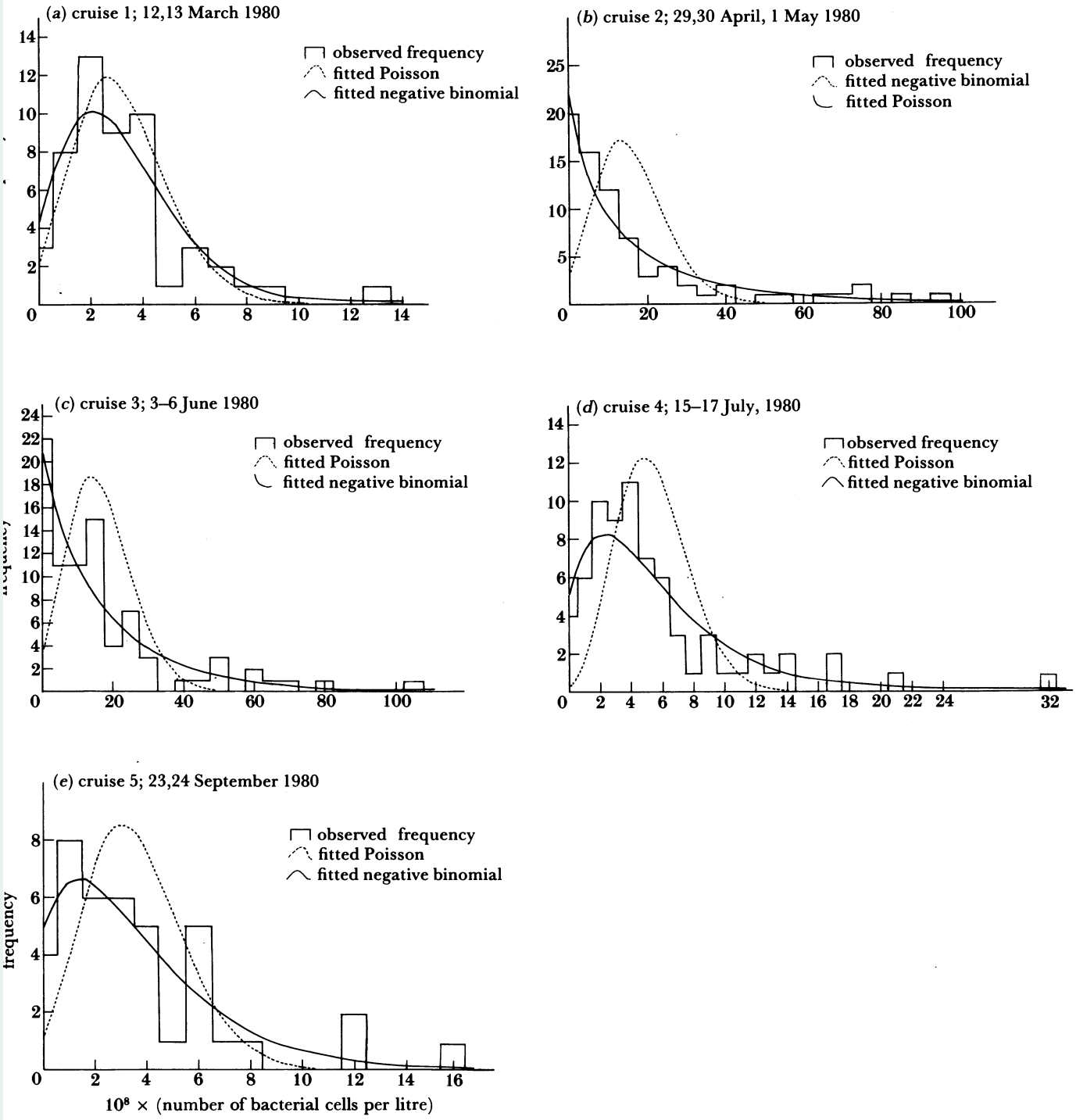

FIGURE 2. (*a*)–(*e*). Comparisons between Poisson and negative binomial distributions fitted by the method of maximum likelihood to the observed frequency distributions of bacterioplankton numbers at five different times of the year in 1980 in the Western Irish Sea.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

TABLE 12. FITTING THE NEGATIVE BINOMIAL DISTRIBUTION TO ZOOPLANKTON NUMBERS BY THE METHOD OF MAXIMUM LIKELIHOOD FOR THE ENTIRE WATER MASS IN 1980

	cruise 2	3	4
n	75	80	70
mean	0.96	2.00	1.06
variance	2.07	14.08	3.66
Fisher's index	159.56	556.16	238.25
$1/K$	1.20	3.03	2.33
χ^2	1.25	0.07	2.55
d.f.	2	3	1
P-value	0.54	0.99	0.11

No observations were made on cruise 5 and only 12 on cruise 1.

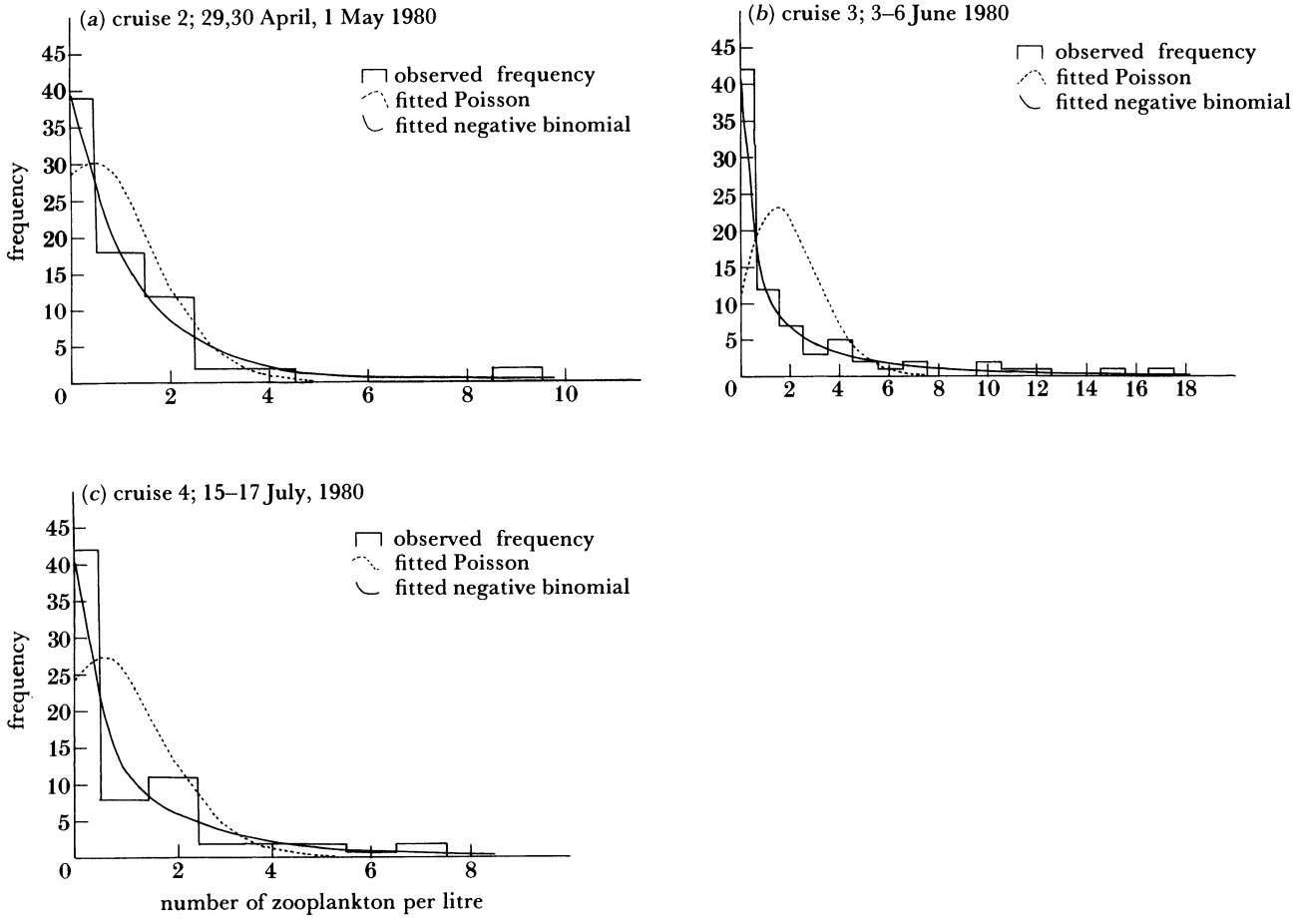

FIGURE 3. (*a*)–(*c*) Comparisons between Poisson and negative binomial distributions fitted by the method of maximum likelihood to the observed frequency distributions of zooplankton numbers at three times during the year during stable stratification in 1980 in the Western Irish Sea.

numbers had a reversed J-shaped distribution with the data in the earlier cruises (2 and 3) showing greater patchiness and that the negative binomial distribution again provides a superior fit to the Poisson.

The analyses for both types of planktonic organisms were performed on the entire water mass for all cruises since sample sizes were not sufficiently large to allow tests for separate water masses.

4. Conclusion

It is suggested that an insight could be developed into this complex frontal ecosystem through statistical analyses including some non-parametric methods. The conclusions from these analyses regarding the various objectives set out in the introduction may be summarized as follows.

(i) The distributional properties observed over time and space indicate that most of the variables are skewed to the right and tend to have a log-normal rather than the normal distribution. But the complexity of the marine biology of the ecosystem is such that several years data at least is required before adequate information is obtained on the forms of the distributions. However, further analyses through fitting the more appropriate discrete distributions by the method of maximum likelihood provide strong and consistent evidence that the bacterioplankton and zooplankton numbers have a negative binomial distribution at different times of the year. It may be of interest to note that the best fitted negative binomial distributions for bacterial counts were obtained in the July and September cruises while for zooplankton number these occurred in the earlier cruises in May and June. This perhaps reflects the greater stability of the water column in the later part of the summer favouring patchiness in bacterial distribution, whereas aggregation of zooplankton depends more on behavioural characteristics. The addition of the bacterioplankton to the planktonic organisms known to be contagious in distribution is a significant advance.

(ii) Associations, described by correlation coefficients, between variables within water masses and for different cruises produced too many significant coefficients which could not be explained in our present state of knowledge of the ecosystem. However, certain general features were discernible. The most important of these were the occurrence of more significant correlations in the surface stratified than the other two water types and that considerably more significant correlations occurred during stratification in cruises 2, 3 and 4 than in cruises 1 and 5.

The complex structure of the ecosystem was simplified by reducing the 15 variables under study to a few meaningful principal components that accounted for a substantial part of the variability in the data for the surface stratified water. A number of components were persistent over some of the five cruises. These were a 'biological stratification' component which occurred regularly and showed the effect of physical stratification on the biological activity, especially uptake rates and plankton numbers; a 'chemical stratification' component which showed the depletion of the nitrogenous compounds by plankton; a 'chemical contrast' which compared various chemical variables; and a 'biological contrast' which indicated the removal of the larger bacterial cells by, possibly, protozooplankton grazing. The lack of a measure of protozoan biomass and activity is a notable omission and should be rectified in any future investigation. In contrast, hardly any components in the combined bottom stratified and mixed waters could be interpreted in any meaningful way. This could, in part, be due to the small number of observations relative to the number of variables involved. The question of what should be the

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

optimum ratio of these numbers is rather difficult to answer. Kendall (1975) suggested a rough rule of thumb of about 10 times as many observations as there are variables. So any future programme should be planned with this in mind if principal component analysis is to be used to its fullest advantage.

(iii) The levels of the biological and biochemical activities in the surface stratified water were often significantly different from those in the other water columns especially during stable stratification in cruises 2, 3 and 4. Bacterial and zooplankton numbers were more consistent than bacterial cell volume and zooplankton biomass, indicating that the number of organisms rather than their biomass is relevant. The potential for discriminating water masses shown by the microbial activity measurements glucose and urea uptake rates, by using both ANOVA and discriminant analysis, is intriguing and merits further research. The fact that the most persistent effect of stratification on biological activity was observed on uptake rates rather than on biomass measurements is worth noting. This indicates that the metabolic activity of the planktonic organisms gives a clearer picture of the response of the biology to its physical and chemical environment than biomass or standing crop estimates. The biomass may be subject to increase by growth or decrease by grazing, whereas metabolic rates are a function of the physiology of the organism itself alone.

(iv) Significant differences between the three lifestages of zooplankton numbers were found in all but one of the eight cruises in the two successive years for which data were available. Zooplankton haul numbers in the stratified, front and mixed water masses were highly significantly different, with the frontal zone showing the greatest abundance followed closely by the stratified side. On the basis of this result, together with other evidence obtained from detailed examination of the level of activity on the stratified side near the front, it would in any future research be worth investigating the front as a separate water mass, sampling it at least as extensively as the other water masses.

Modelling the diurnal response of the zooplankton numbers to depth and time in the stratified and mixed water columns during the early June cruise indicated that the fitted models were both quadratic; the numbers in the stratified water depended on depth while those in the mixed water depended on time. A more complete picture of the relationship between zooplankton numbers and time and depth may, in any future investigation, be achieved by a time series analysis of data collected over a longer period of, say, five days.

(v) The overall evidence is that the surface stratified water is a zone of intense biological activity, especially near the front, and the level of this activity is quite different from those in the bottom stratified and mixed water columns. Furthermore, there is in this zone a close coupling between the physical, chemical and biological variables during stable stratification, a matter that will be discussed in more detail by Fogg *et al.* 1985*a*).

We are indebted to our colleagues Professor G. E. Fogg, Dr D. A. Jones, Dr C. M. Turley, Dr K. Lochte, Mr B. Egan and Mr S. Scrope–Howe for their valuable comments on the biological and practical interpretation of some of the statistical results.

We are also grateful to the referee for his useful and constructive comments which have led to an improvement in the presentation of this paper.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

References

Alvey, N. G. *et al.* 1977 *GENSTAT. A general statistical program*, Harpenden: Rothamsted Experimental Station.

Ashby, R. E. & Rhodes-Roberts, M. E. 1976 The use of analysis of variance to examine the variations between samples of marine bacterial populations. *J. appl. Bact.* **41**, 439–451.

Atkinson, A. C. 1983 Diagnostic regression analysis and shifted power transformations. *Technometrics* **25**, 23–33.

Baker, R. J. & Nelder, J. A. 1978 *The GLIM system, release* 3, *generalized linear interactive modelling*. Oxford: Numerical Algorithms Group.

Bliss, C. I. & Fisher, R. A. 1953 Fitting the negative binomial distribution to biological data. *Biometrics* **9**, 177–200.

Bolter, M., Meyer, M. & Probst, B. 1980 A statistical scheme for structural analysis in marine ecosystems. *Ecol. Modelling* **9**, 143–151.

Box, G. E. P. 1953 Non-normality and tests on variance. *Biometrika* **40**, 318–335.

Cassie, R. M. 1962*a* Frequency distribution models in the ecology of plankton and other organisms. *J. Anim. Ecol.* **31**, 65–92.

Cassie, R. M. 1962*b* Microdistribution and other error components of C^{14} primary production estimates. *Limnol. Oceanogr.* **7**, 121–130.

Cassie, R. M. 1963 Microdistribution of plankton. *Oceanogr. mar. Biol. ann. Rev.* **1**, 223–252.

Colebrook, J. M. 1960 Plankton and water movements in Windermere. *J. Anim. Ecol.* **29**, 217–240.

Comita, G. W. & Comita, J. J. 1957 The internal distribution of a calanoid copepod population, and a description of a modified Clarke–Bumpus plankton sampler. *Limnol. Oceanogr.* **2**, 321–332.

Cushing, D. H. 1962 Patchiness. *Rapp. P.-v. Réun. Cons. perm. int. Explor. Mer* **153**, 152–164.

Egan, B. & Floodgate, G. D. 1985 Biological studies in the vicinity of a shallow sea tidal mixing front. II. The distribution of bacteria. *Phil. Trans. R. Soc. Lond.* B. **310**, 435–444.

El-Shaarawi, A. H., Esterby, S. R. & Dutka, B. J. 1981 Bacterial density in water determined by Poisson or negative binomial distributions. *Appl. environ. Microbiol.* **41**, 107–116.

Fisher, R. A. 1941 The negative binomial distribution. *Ann. Eugen.* **11**, 182–187.

Fogg, G. E., Egan, B., Floodgate, G. D., Jones, D. A., Kassab, J. Y., Lochte, K., Rees, E. I. S., Scrope-Howe, S. & Turley, C. M. 1985*a* Biological studies in the vicinity of a shallow-sea tidal mixing front. VII. The frontal ecosystems. *Phil. Trans. R. Soc. Lond.* B **310**, 555–571.

Fogg, G. E., Egan, B., Hoy, S., Lochte, K., Scrope-Howe, S. & Turley, C. M. 1985*b* Biological studies in the vicinity of a shallow-sea tidal mixing front. I. Physical and chemical background. *Phil. Trans. R. Soc. Lond.* B **310**, 407–433.

Kendall, M. G. 1975 *Multivariate analysis*. London: Charles Griffin.

Kriss, A. E. 1962 *Deep sea microbiology*. London: Oliver and Boyd.

Lochte, K. 1985 Biological studies in the vicinity of a shallow-sea tidal mixing front. III. Seasonal and spatial distribution of heterotrophic uptake of glucose. *Phil. Trans. R. Soc. Lond.* B, **310**, 445–469

Miller, R. G. 1966 *Simultaneous statistical inference*. New York: McGraw-Hill.

Neter, J. & Wasserman, W. 1974 *Applied linear statistical models*. Illinois: Irwin.

Nie, N. H. *et al.* 1975 *SPSS, statistical package for the social sciences*. New York: McGraw-Hill.

Ross, G. J. S. *et al.* 1980 *MLP, maximum likelihood program*. Harpenden: Rothamsted Experimental Station.

Rosswall, T. & Kvillner, E. 1978 Principal-components and factor analysis for the description of microbial populations. *Adv. microb. Ecol.* **2**, 1–48.

Sach. L. 1982 *Applied statistics: a handbook of techniques*. Heidelberg: Springer-Verlag.

Scrope-Howe, S. & Jones, D. A. 1985 Biological studies in the vicinity of a shallow-sea tidal mixing front. V. Composition, abundance and distribution of zooplankton in the western Irish sea, April 1980 to November 1981. *Phil. Trans. R. Soc. Lond.* B **310**, 501–519.

Simpson, J. H., Hughes, D. G. & Morris, N. C. G. 1977 The relation of seasonal stratification to tidal mixing on the continental shelf. In *A voyage of discovery* (ed. M. Angel); suppl. to *Deep Sea Res.* pp. 327–340.

Snedecor, G. W. & Cochran, W. G. 1980 *Statistical methods*. Ames: Iowa State University Press.

Sokal, R. R. & Rohlf, F. J. 1973 *Introduction to biostatistics*. San Francisco: Freeman.

Turley, C. M. 1985 Biological studies in the vicinity of a shallow-sea tidal mixing front. IV. Seasonal and spatial distribution of urea and its uptake by phytoplankton, *Phil. Trans. R. Soc. Lond.* B **310**, 471–500.

Vaatanen, P. 1980 Factor analysis of the impact of the environment on microbial communities in the Tvarminne area, southern coast of Finland. *Appl. environ. Microbiol.* **40**, 55–61.

Winer, B. J. 1971 *Statistical principles in experimental design*. New York: McGraw-Hill.

Zar, J. H. 1974 *Biostatistical analysis*. Eaglewood Cliffs, N.J.: Prentice-Hall.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES