Analysis of Phytoplankton Data from Two Lower Manhattan Sites

Final Report of a Grant from the Hudson River Foundation

Michael Levandowsky, Ph.D.
Haskins Laboratories
Pace University
New York, NY 10038

David Vaccari, Ph.D.
Dept. Civil, Environmental and Ocean Engineering
Stevens Institute of Technology
Hoboken, NJ 07030

March 2004

Acknowledgement: We thank the following students and collaborators, who will be co-authors in the published versions of this work:

Zhau Wang, Stevens Institute of Technology, Hoboken, NJ
Eun Woong Kim and Martin Park, Pace U. and The River Project, N.Y., NY

Preface

We present here two analyses of phytoplankton in the Lower Hudson River estuary. These are based on categorical (presence-absence) data for phytoplankton taxa from weekly samples over an 8-year period, from 1996 to 2003. The first, by the group from Pace University, is a correspondance analysis based on 29 taxa. The second, by the group from Stevens Institute, is a nonlinear multivariate polynomial model of occurrence for 3 of these taxa.

Many studies of phytoplankton populations are concerned with the dynamics of growth and productivity. For this, quantitative data are generally required: cell counts, biomass or pigment concentrations, productivity, utilization of nutrients or other dynamic variables. In the present study, however, we deal with presence-absence data. This raises the question: how meaningful are such data? Can one lump together a sample where a species is extremely abundant, even dominant, with another in which it's represented by only a few cells? Here we explore patterns that can be found with such qualitative data.

Many environmental groups sponsor water monitoring programs by volunteers, who in many cases are not professional scientists. The question naturally arises: how valuable are the data obtained from such projects? Typically, abiotic parameters are measured using inexpensive kits, such as colorimetric pH kits and micro-Winkler kits for dissolved oxygen (DO), and simple equipment such as the Secchi disk. Biological identification is usually based on light microscopy of living material and, in general, qualitative (presence-absence) rather than quantitative biological data are obtained. In the present study we examine several years' worth of such data for the lower Hudson and East Rivers, in New York, and explore its potential information content.

The two analyses are presented separately here, as 2 sections. They have been presented as posters at scientific meetings, and will shortly be submitted for publication.

Section 1. PHYTOPLANKTON PATTERNS IN THE LOWER HUDSON RIVER AND EAST RIVER, N.Y., 1996-2003: THE INFORMATION CONTENT OF CATEGORICAL DATA

Kim, E.-W.[1], Park, M.[2] and Levandowsky, M.. [1,2]

[1] Haskins Labs, Pace U., N.Y., NY 10038
[2] The River Project, N.Y., NY 10013

Abstract

Canonical correspondence analysis (CCA) was used to analyze an 8 year series of weekly phytoplankton samples from surface water at 2 sites in Lower Manhattan during the period 1996-2003. The analysis used presence-absence data of 29 taxa, readily identifiable with the light microscope from living material, and measurements of temperature, salinity, pH, Secchi depth and dissolved oxygen. The 2 sites were: The River Project (Pier 26 on the Hudson River); and the South Street Seaport (Piers 15-16 on the East River). The Hudson sample points formed distinct clusters with respect to salinity, temperature and season, whereas the East River samples, involving the same taxa, did not. The data suggest significant differences in organization of taxonomic contents of phytoplankton at these two sites: lower Hudson River phytoplankton is in some sense much more structured than that of the East River. The analysis also demonstrates that meaningful statistical information can be obtained from such categorical (presence/absence) data.

Introduction.

Approximately weekly measurements of basic abiotic parameters and net phytoplankton samples have been taken in tidal surface waters at two sites in Lower Manhattan starting in1989. Phytoplankton is examined in the light microscope, and identifications made, as feasible, from living material. We see these data as proxy for the sort of data often obtained by volunteer monitoring programs. We show here that correspondence analysis can reveal relationships and interactions of biological interest in such data, which were not obvious from casual observation.

The phytoplankton in the saline waters surrounding New York City were described by Malone and others in the 1970's and 1980's (Malone 1977a,b; Kaneta et al 1985), but conditions have changed significantly since then and relatively little such work has been done in recent years. The present study therefore also provides a baseline description of the present phytoplankton communities, as well as some of the relevant physico-chemical parameters associated with them.

Methods

Sampling protocol and sites.

Weekly plankton samples were taken at 2 sites in Lower Manhattan: The River Project (Pier 26) on the Hudson River (HR); and the South Street Seaport (Piers 15-16) on the East River (ER) (Fig. 1). A fine net (10 µm mesh; 1 m opening diameter) was towed in the upper meter of the water column by hand, from a pier at a slow walking speed. Plankton tows were begun in 1989, but the earlier biological data are considered too fragmentary for use and the present analysis is based on data from 1996 – 2003. Abiotic measurements however are shown in Fig. 2, starting in 1989 (HR) and 1990 (ER).

Identification and choice of taxa for the analysis

Phytoplankton species were identified from live material using an inverted phase contrast microscope, yielding a qualitative (presence-absence) record of the major identifiable fine-net phytoplankton taxa. Sample aliquots were also preserved (Lugol's iodine and buffered formalin). Since living material was observed, using only light microscopy, many diatoms, dinoflagellates and other algal forms were not identifiable to species level, and in a number of cases it was possible only to classify cells at a higher taxonomic level. For purposes of the analysis we used species where available, but higher taxa were also used. Problematic forms, where identification was in doubt were not used. Taxa that occurred only once or a few times in the 7 year period were not used. In addition 2 species, *Skeletonema costatum* and *Paralia sulcata*, were almost always present, and therefore were not used (when they were removed from the analysis the eigenvalues actually increased slightly). After these eliminations, a set of 29 taxa of phytoplankters remained and these were used for the correspondence analysis (Table 1). While biological data are available going back to 1989, earlier data was considered incomplete, and it was decided to use only data from 1996 in the correspondence analysis.

Physico-chemical Measurements

The hydrographic variables measured were salinity, temperature, pH, dissolved oxygen (DO) and turbidity (secchi depth). Temperature was measured with a series of glass mercury and spirit thermometers, calibrated against a certified reference thermometer. DO was measured by azide modification of the Winkler method, using a kit (LaMotte). Salinity was measured with a portable temperature-compensated refractometer (Reichert). Secchi depth was measured using a 20 cm diameter disk with alternating white and black quadrants. pH was measured colorimetrically by Cresol Red, using a kit (LaMotte); this was checked periodically against pH meters in the laboratory and found accurate within the limits of resolution. Before 2000 the pH was recorded to the nearest 0.1 pH unit.

At that point it was decided that extrapolation to the nearest 0.05 pH unit was valid (the change in data resolution is evident by inspection in Fig. 2).

Statistical Analysis

Correspondence analysis (CCA, based on covariances) (Ter Braak 1988) was used. CCA is an ordination technique in which species or other taxa distributions along gradients of environmental variables are assumed to be unimodal, and the extracted species axes are constrained to be linear combinations of environmental (i.e., abiotic, physico-chemical) variables. In the analysis the latter can then be represented as vectors in the multivariable space defined by the taxa. The samples are represented by points in the space, their positions determined by taxa content. The projections of the sample points and environmental vectors in 2 or 3 dimensional subspaces can then be examined for patterns and correlations. We used the CANOCO software of ter Braak (1988) to analyze presence-absence data of 29 taxa for weekly samples taken during 1996-2003. Eigenvalues of the first two axes were 0.434 and 0.337, respectively, accounting for 13.1 percent of the total variance. Cumulative percentage of variance of the taxa-environment correlations for the first 2 axes was 82.2.

Results

Basic Data: Values of surface water temperature, dissolved oxygen (DO), salinity, Secchi depth and pH from 1989 to 2003, in the East River and the Lower Hudson River, are shown in Fig.2. The vertical line indicates the beginning of the period for which suitable phytoplankton data are available for the correspondence analysis. Seasonal variation reflects the effects of spring runoff from rain and melting snow in the watershed. There also appears to have been a trend to higher salinities since 1998. The HR site has recurrent instances of low salinity: episodes of low salinity occurred frequently in winter samples, going below 10 ppt. every year and occasionally reaching undetectable levels. In contrast the ER samples were below 10 ppt. only 6 times in 14 years. This difference no doubt reflects the influence of freshwater from the upriver watershed of the Hudson. There is a tendency to higher turbidity (lower Secchi depth) at the HR site.

Presence-absence data since 1996 for the 29 taxa used in the correspondence analysis are shown in Fig. 3.

Ordinations: Fig. 4 shows the results of 2-dimensional CCA, with outlines around sample points corresponding to 3 temperature ranges. The HR groups are relatively distinct, with very little overlap compared to the ER groups.

Fig. 5. shows seasonal groupings. Here too the HR groups overlap less than the corresponding ER groups.

Fig. 6 shows groups corresponding to 3 salinity ranges for the HR sample points, and to 2 salinity ranges for ER sample points (the ER samples did not have the low salinity range of HR samples).

In Fig. 7 the center of gravities of distributions of the 29 taxa among the samples are shown.

In Fig. 8, we see the results of an analysis in which the 12 sample months are used as nominal environmental variables, and centroids of the sample points corresponding to each month are plotted. Note the counter-clockwise progression of the months in both cases.

Fig. 9 Shows the distributions of individual taxa among the sample points, with density contours calculated by the method of locally weighted regression smoothing (sometimes termed *loess* or *lowess;* Cleveland 1994; Cleveland and Devlin 1888), using the CANOCO program.

Discussion

Differences in Hudson River and East River Patterns

The striking differences seen in Fig. 4 and fig. 6 for the distribution of samples with regard to salinity and temperature are intriguing. Partitioning with regard to salinity in the case of the Hudson samples may be related to the tendency for water masses there to be stratified by density differences. Samples from greater depths usually have substantially higher salinity at the HR site (data not shown), whereas the strong turbulent mixing by tidal currents at the ER site tends to prevent this. Thus water masses are often distinct at the HR site, and our samples would reflect this. In the ER, sample points represent more homogeneous mixtures, and so, even though the same taxa are used, distinct salinity communities are not seen (Fig. 6). Regarding temperature, though, this argument seems less forceful, because we have found that, while salinity can vary greatly with depth in the HR site, temperature is usually rather uniform, differences between surface and bottom water usually being within a degree or two centigrade of each other. Thus, the distinct groups seen at the HR site in Fig. 4 may require further explanation.

Qualitative Versus Quantitative Biological Data

What determines whether a species is present, in whatever abundance? On the one hand, a given species might be unable to survive below a certain salinity, or above a certain temperature. But on the other hand, if it's within suitable temperature and salinity ranges and therefore present, its abundance may reflect other types of variables that operate on a trophic level: nutrient concentrations, predator grazing, or light levels perhaps.

Thus there are different sorts of determining variables - some reflected in qualitative data, others in quantitative data. If this is so, then it seems meaningful to analyze qualitative data, because although it doesn't tell you about dynamics, it does say something about evolutionary adaptations of sympatric species to certain aspects of the environment.

Note that, strictly speaking, the term "absence" here is problematic - that is, we can't say that a species is necessarily absent because we did not detect it. It is likely to be present at low concentrations, but, by the laws of probability, not present in our particular sample. So by "absence" we really mean either really absent, or at any rate present at low concentrations. With "presence" though we are on firmer ground - if we saw it was there! So to be quite precise, we should perhaps say that our data indicates which taxa are known to be present. But, it doesn't say in what numbers, so in that sense it's qualitative.

Actually, as one probes a bit, interactions and other complexities appear. For example, one can increase the temperature range for survival and growth in some algal protists by adding trace nutrients - vitamins and trace elements - to the medium (Hutner et al 1957). We should therefore perhaps speak of different *aspects* of determining environmental variables that are revealed by qualitative and quantitative data (Levandowsky 1972).

References

Cleveland, W.S. 1994. The Elements of Graphing Data. Revised Edition. Hobart Press, Summit, NJ

Cleveland, W.S. and S.J. Devlin 1988. Locally-weighted regression: An approach to regression analysis by local fitting. J. Am. Stat. Assoc. 83, 596-610.

Hutner, S.H., H. Baker, A. Aaronson, H. Nathan, E. Rodriguez, S. Lockwood, M. Sanders and R. A. Peterson. 1957. Growing *Ochromonas malhamensis* above 35 C. J. Protozool. 4, 259-269.

Kaneta, P.J., M. Levandowsky and W. Esaias 1985 Multivariate analysis of the phytoplankton community in the New York Bight. Mar. Ecol. Prog Ser. 23, 231-239.

Karentz, D. and T. Smayda 1984. Temperature and seasonal occurrence of 30 dominant phytoplankton species in Narragansett Bay over a 22-year period (1959-1980). Mar. Ecol. Prog. Ser. 18, 277-293.

Levandowsky, M. 1972. Ecological niches of sympatric phytoplankton species. Am. Nat. 106, 71-8

Levandowsky, M. 1972 An ordination of phytoplankton populations in ponds of varying salinity and temperature. Ecology 53, 398-407.

Malone, T.C. 1977a. Environmental regulation of phytoplankton productivity in the lower Hudson estuary. Estuar. Coast. Mar. Sci. 5, 157-171

Malone, T.C. 1977b. Plankton systematics and distributions. MESA New York Bight Atlas Mon. 13, N.Y. State Sea Grant Institute, Albany, NY

Manly, B.F. 1986. Multivariate Statistical Methods. A Primer. Chapman and Hall

Siver, Peter A. 1999. Development of paleolimnological inference models for pH, total nitrogen and specific conductivity based on planktonic diatoms. J. Paleolimnol. 21, 45-59

Ter Braak, C.J.F. 1988 Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. Ecology 67, 1167-1179

Table 1. The 29 Taxa used for correspondance analysis.

Diatoms

Actinoptichus undulatus (= *Actinoptychus senarius*)
Amphirora alata
Asterionella japonica (= *Asterionellopsis glacialis*)
Biddulphia aurita (= *Odontella aurita*)
Chaetoceros debilis
C. socialis
C. subtilis
Corethron criophilum (= *Corethron hystrix*)
Coscinodiscus excentricus
Ditylum brightwelli
Eucampia zodiacus
cf. *Gyrosigma sp.*
Licmophora lyngbyi
Meridion circulare
Nitzchia reversa
cf. *Pseudonitzchia seriata* (= *Nitzchia seriata*)
Rhizosolenia setigera

Chlorophytes

Gomphonema aponina
Pediastrum duplex
Scenedesmus quadricauda

Dinoflagellates

Ceratium fusus
C. tripos
Dinophysis ovum
Exuviaella marina
Gymnodinium splendens (= *Gymnodinium sanguineum*)
Prorocentrum micans

Other

Euglenid
Ebria tripartita
cf. *Phaeocystis*

Note: Some cell types are recognizable from year to year, but their identification is unsure or ambiguous. However, if their potential information value Is judged significant enough they are included in the analysis. Such cases are noted by the prefix “cf.”

Captions

Fig. 1. Location of the two sampling sites.

Fig. 2. Values of physico-chemical parameters , 1989-2003: temperature, dissolved oxygen (DO), salinity, Secchi depth, pH. The vertical line in early 1996 indicates the beginning of phytoplankton data. The correspondance analysis starts then.

Fig. 3. Presence-absence data for 29 taxa. See Table 1 for full names.

Fig. 4. The 2-dimensional CCA. Outlines indicate sample points in certain temperature ranges.

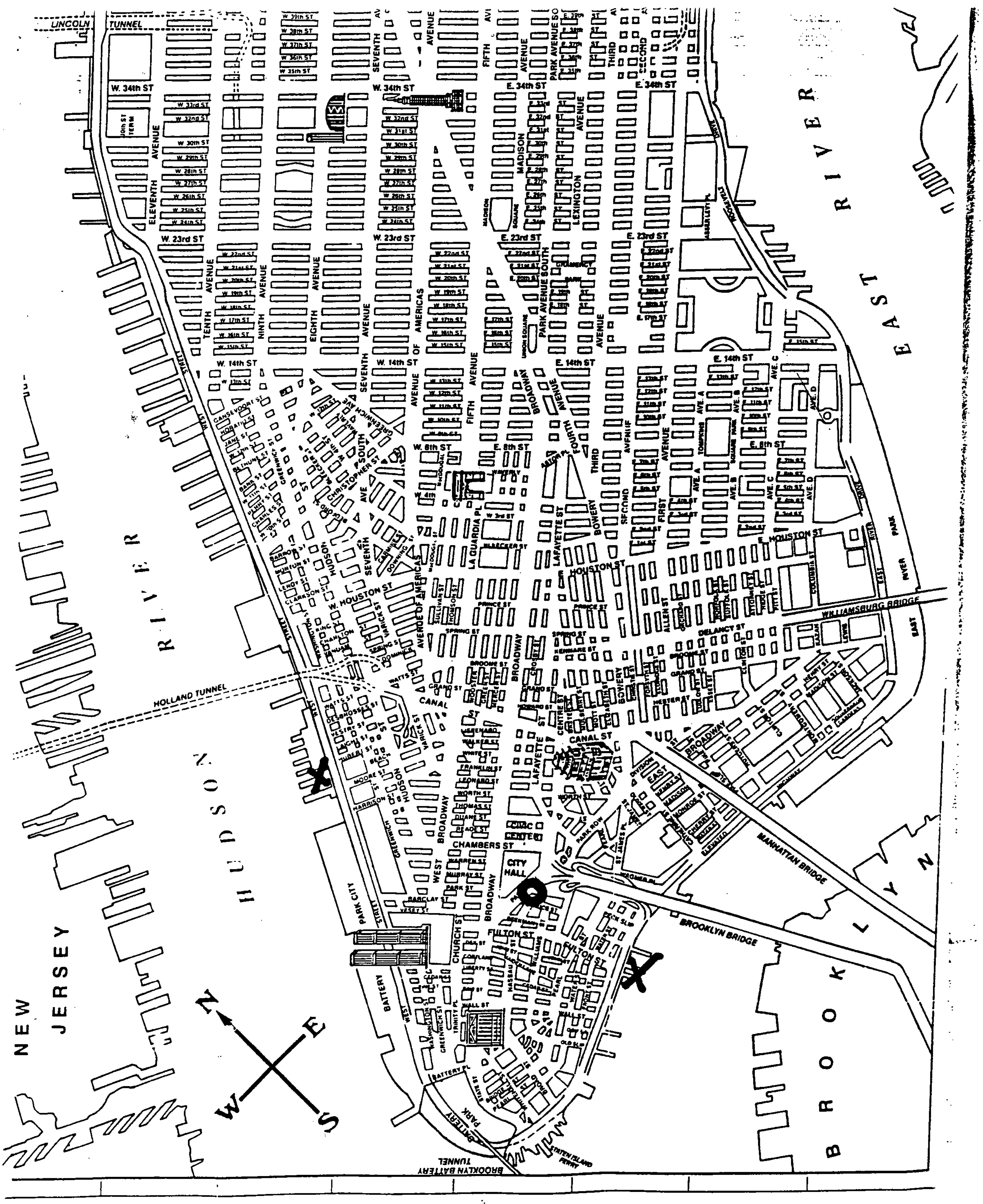
Fig. 5. Same as fig. 4, with seasonal indication. Seasons are here defined as follows: Winter = December through February; Spring = March through May; Summer = June through August; Fall, September through November.

Fig. 6. Same as Fig. 4, partitioned by salinity.

Fig. 7. Centers of gravity for distributions of the 29 taxa among the sample points.

Fig. 8. Results of an analysis in which the month of a sample was treated as a nominal variable.

Fig. 9 (29 parts). Distributions of taxa among sample points. Density contours calculated by the method of *locally weighted regression smoothing* ("loess").



Map of Lower Manhattan, showing the two sampling sites (marked by X's) and Haskins Labs (marked by a circle)

Fig. 1

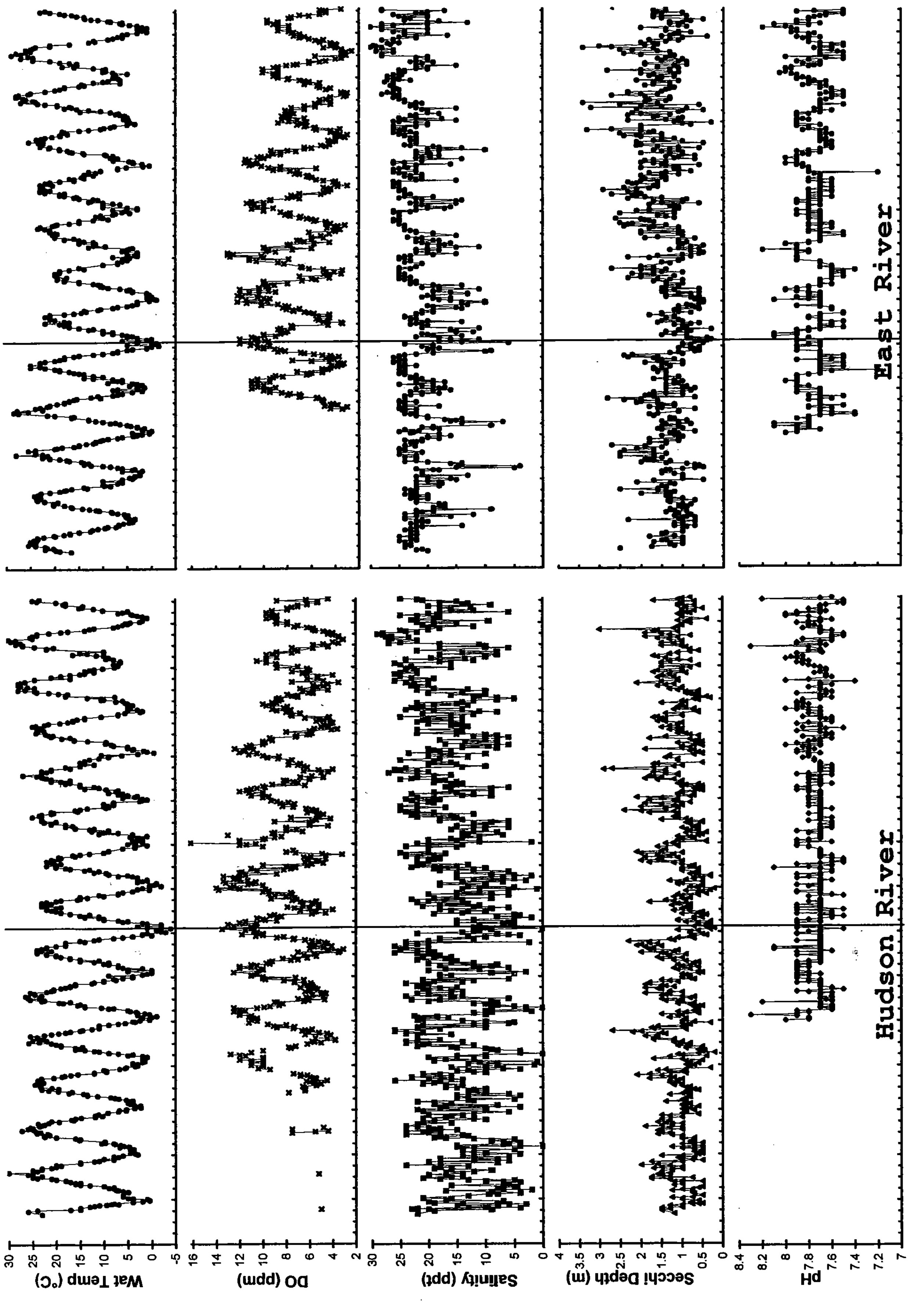

Wat Temp (°C)
30
25
20
15
10
5
0
-5
DO (ppm)
16
14
12
10
8
6
4
2
Salinity (ppt)
30
25
20
15
10
5
0
Secchi Depth (m)
4
3.5
3
2.5
2
1.5
1
0.5
0
pH
8.4
8.2
8
7.8
7.6
7.4
7.2
7
89
90
91
92
93
94
95
96
97
98
99
00
01
02
03
Hudson River
East River

Fig. 2

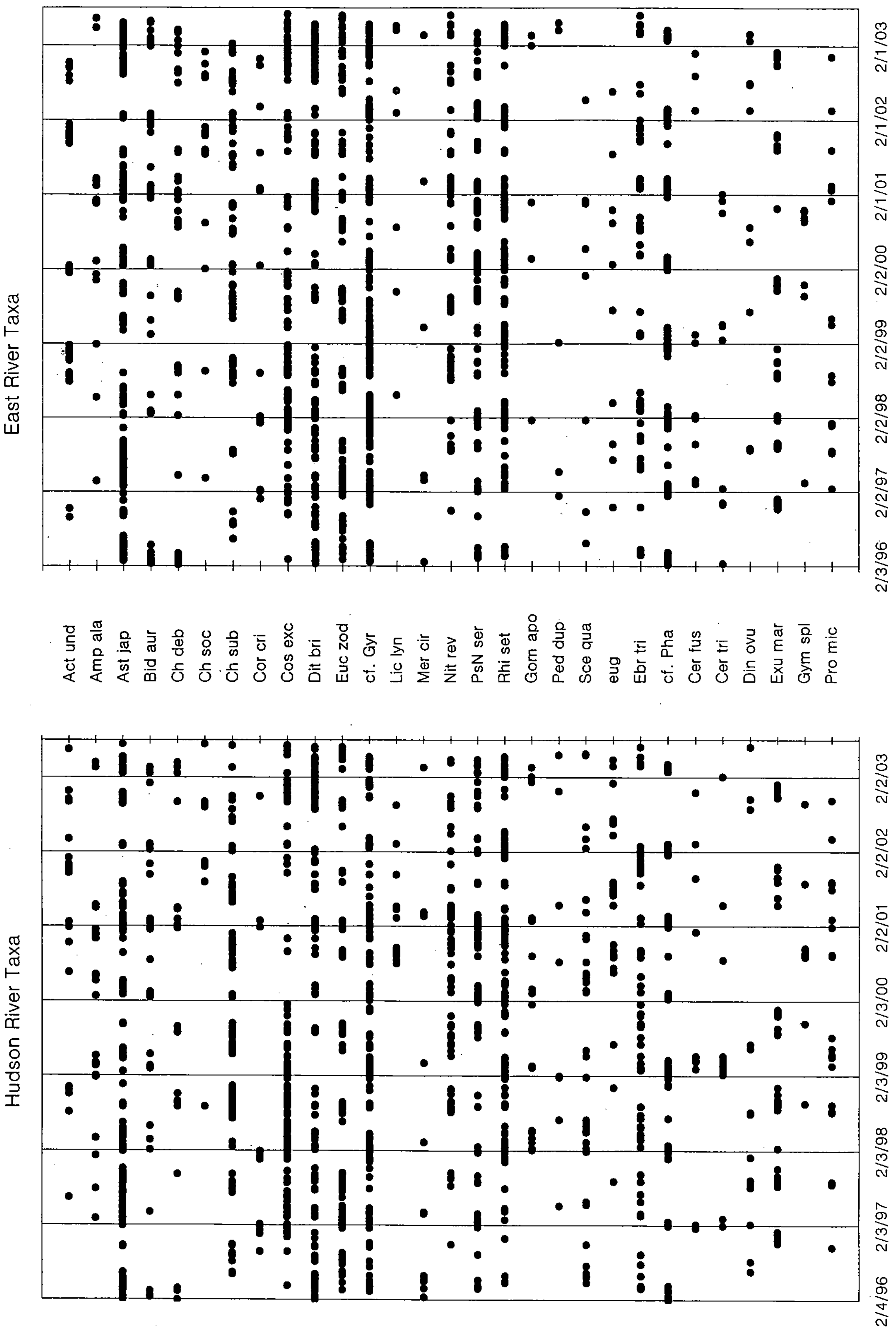

East River Taxa
Hudson River Taxa
Act und
Amp ala
Ast jap
Bid aur
Ch deb
Ch soc
Ch sub
Cor cri
Cos exc
Dit bri
Euc zod
cf. Gyr
Lic lyn
Mer cir
Nit rev
PsN ser
Rhi set
Gom apo
Ped dup
Sce qua
eug
Ebr tri
cf. Pha
Cer fus
Cer tri
Din ovu
Exu mar
Gym spl
Pro mic
2/3/96
2/2/97
2/2/98
2/2/99
2/2/00
2/1/01
2/1/02
2/1/03
2/4/96
2/3/97
2/3/98
2/3/99
2/3/00
2/2/01
2/2/02
2/2/03

Fig. 3

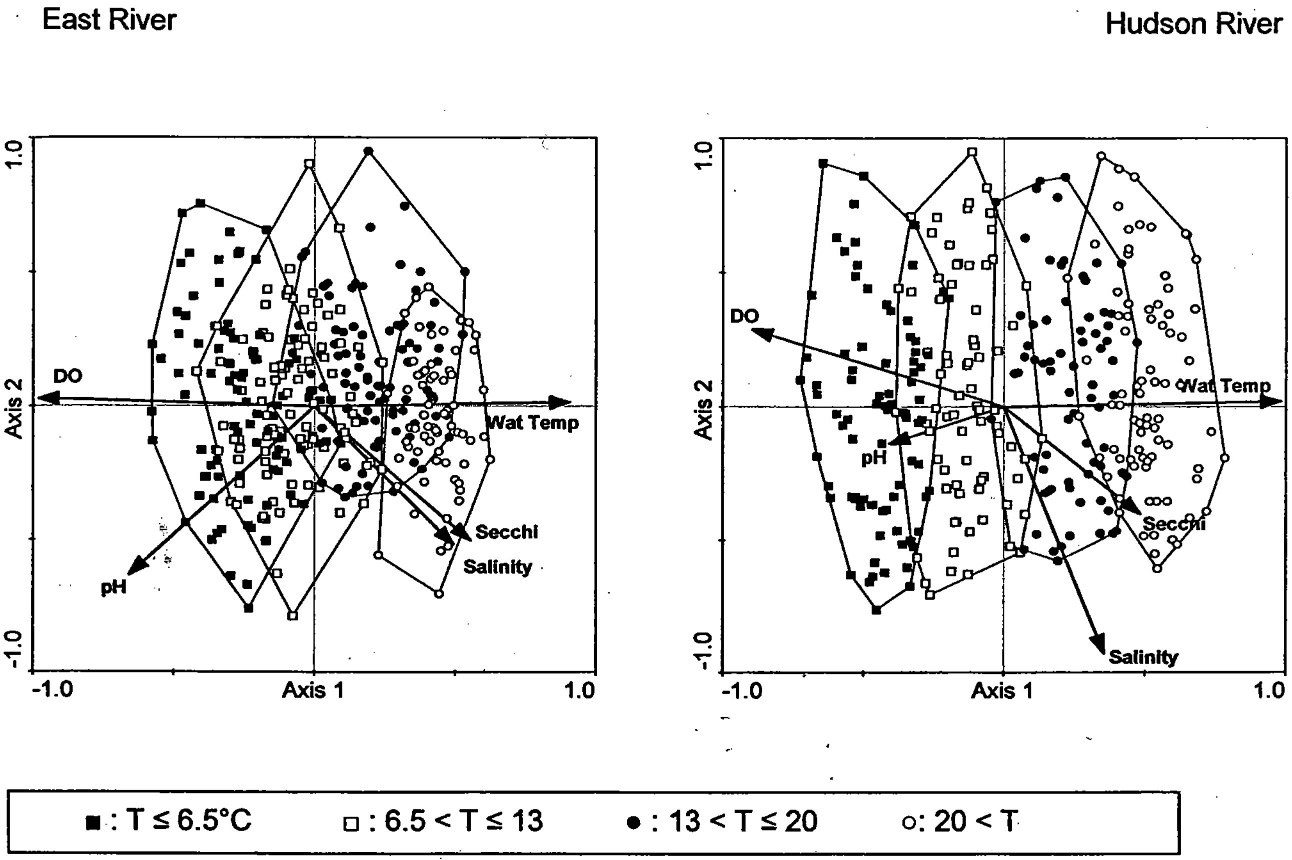

East River
Hudson River
1.0
-1.0
Axis 2
Axis 1
-1.0
1.0
DO
Wat Temp
Secchi
Salinity
pH
■ : T ≤ 6.5°C
□ : 6.5 < T ≤ 13
● : 13 < T ≤ 20
○: 20 < T

Fig. 4

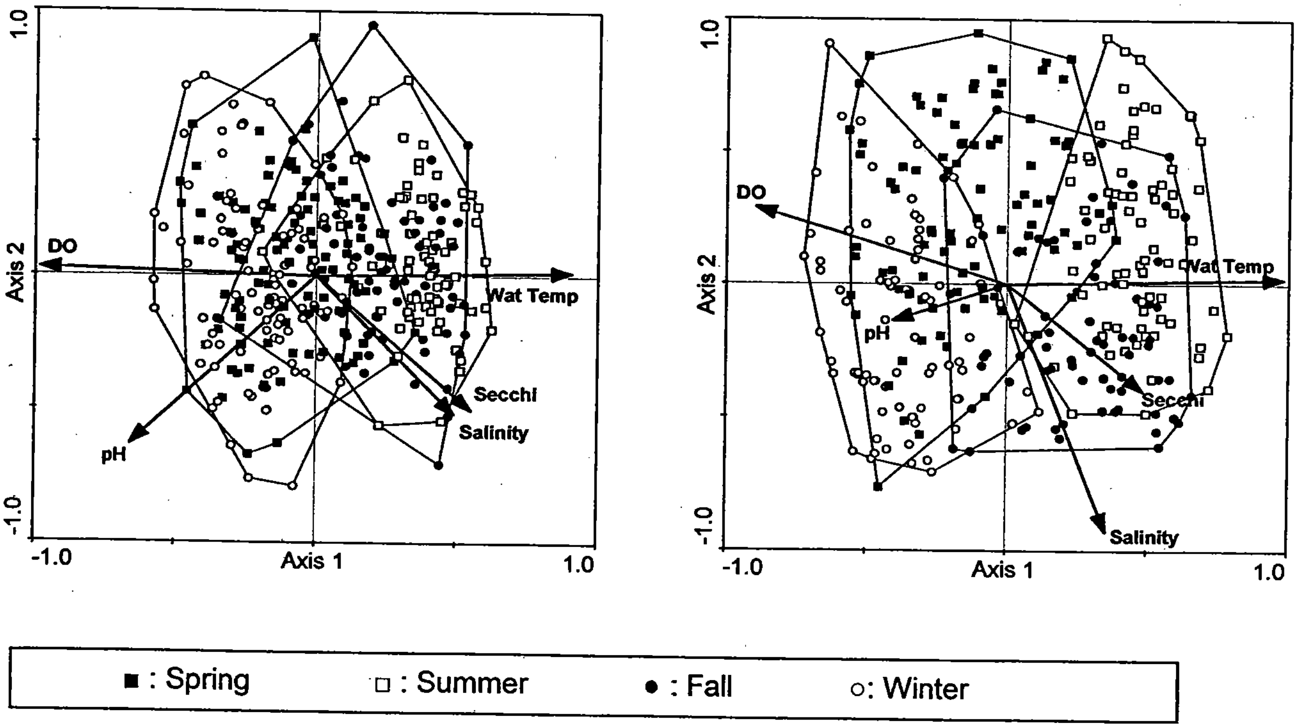

1.0
-1.0
Axis 2
Axis 1
-1.0
1.0
DO
Wat Temp
Secchi
Salinity
pH
■ : Spring
□ : Summer
● : Fall
○: Winter

Fig. 5

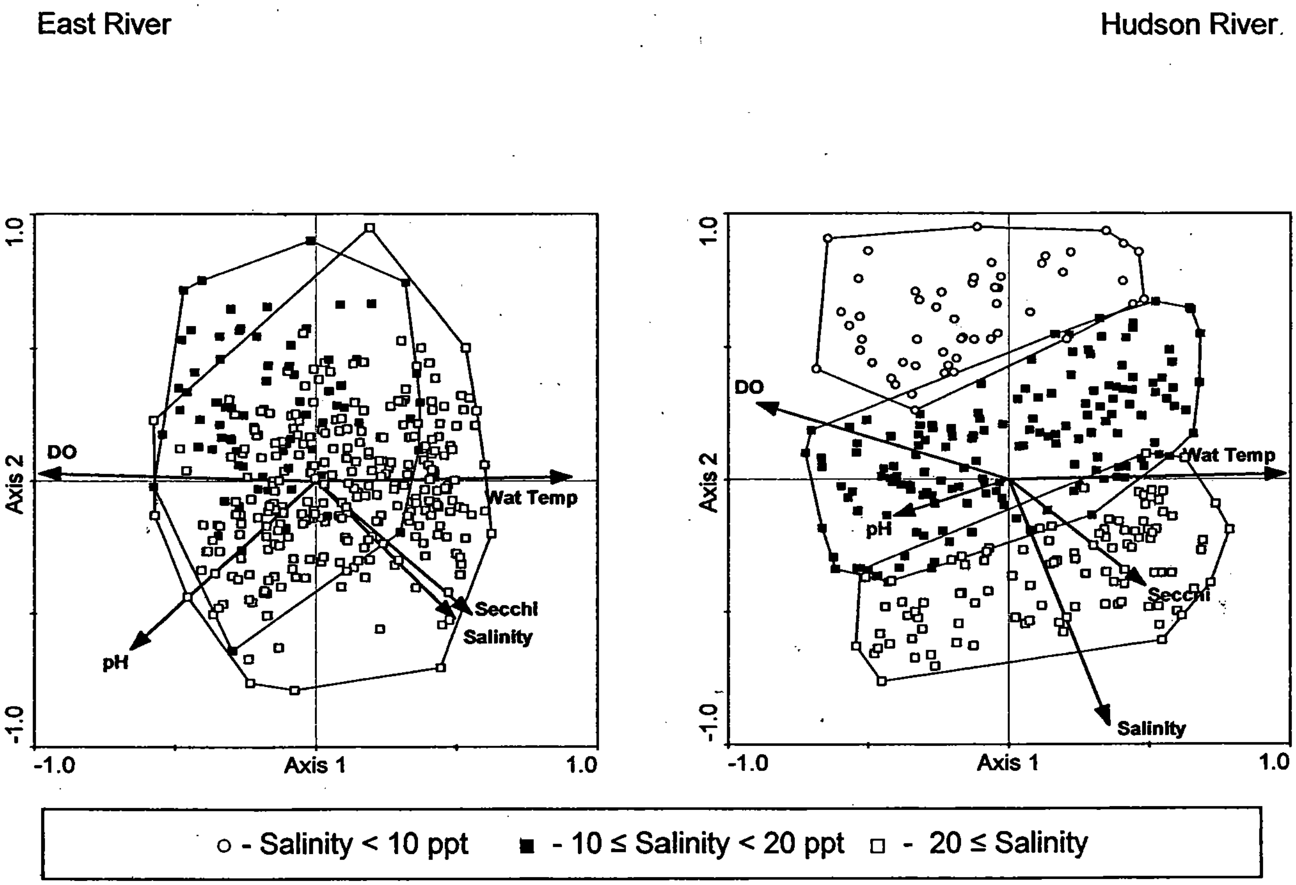

East River
Hudson River
1.0
Axis 2
-1.0
-1.0
Axis 1
1.0
DO
Wat Temp
Secchi
Salinity
pH
○ - Salinity < 10 ppt
■ - 10 ≤ Salinity < 20 ppt
□ - 20 ≤ Salinity

Fig. 6

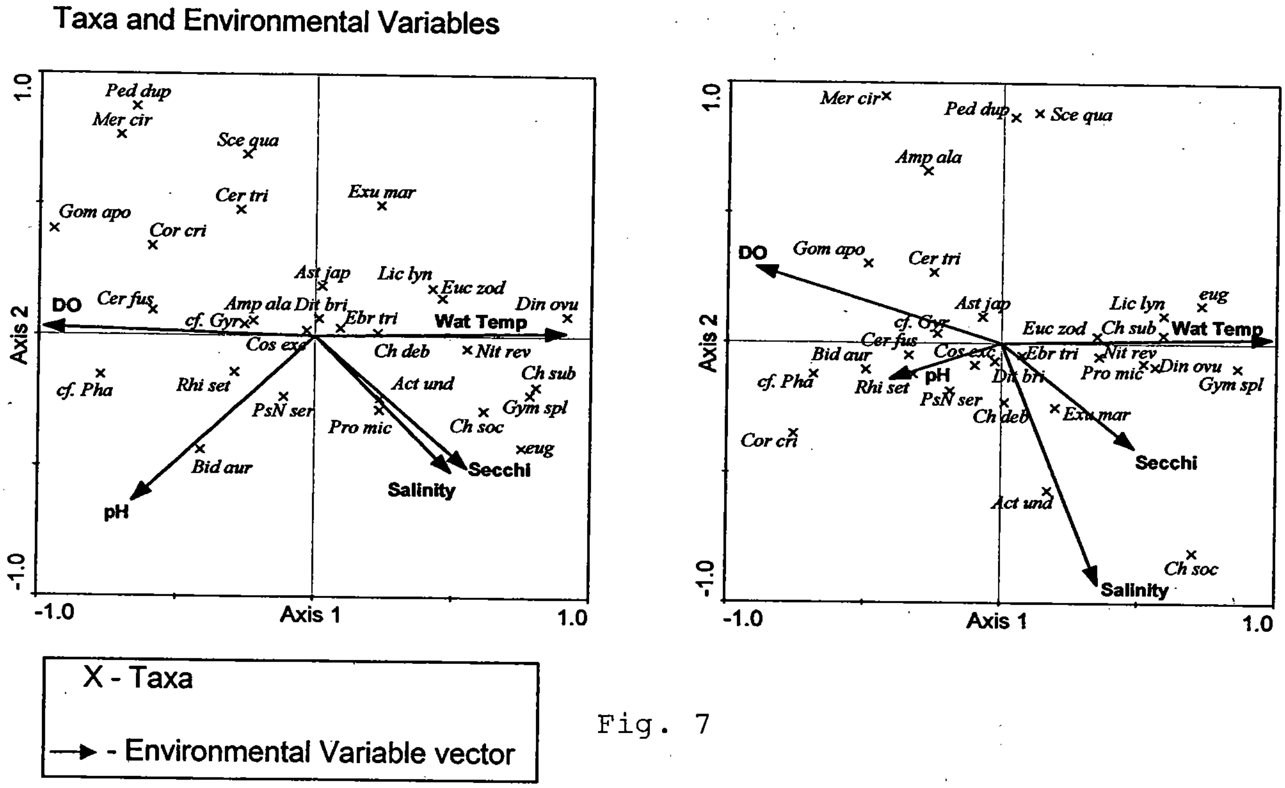

Taxa and Environmental Variables
1.0
Axis 2
-1.0
-1.0
Axis 1
1.0
Ped dup
Mer cir
Sce qua
Cer tri
Exu mar
Gom apo
Cor cri
Ast jap
Lic lyn
Euc zod
DO
Cer fus
Amp ala
Dit bri
Din ovu
cf. Gyr
Ebr tri
Wat Temp
Cos exc
Ch deb
Nit rev
cf. Pha
Rhi set
Act und
Ch sub
PsN ser
Gym spl
Pro mic
Ch soc
eug
Bid aur
Secchi
Salinity
pH
Mer cir
Ped dup
Sce qua
Amp ala
DO
Gom apo
Cer tri
Ast jap
Lic lyn
eug
cf. Gyr
Euc zod
Ch sub
Wat Temp
Cer fus
Bid aur
Cos exc
Ebr tri
Nit rev
cf. Pha
Rhi set
pH
Dit bri
Pro mic
Din ovu
Gym spl
PsN ser
Ch deb
Exu mar
Cor cri
Secchi
Act und
Ch soc
Salinity
X - Taxa
→ - Environmental Variable vector

Fig. 7

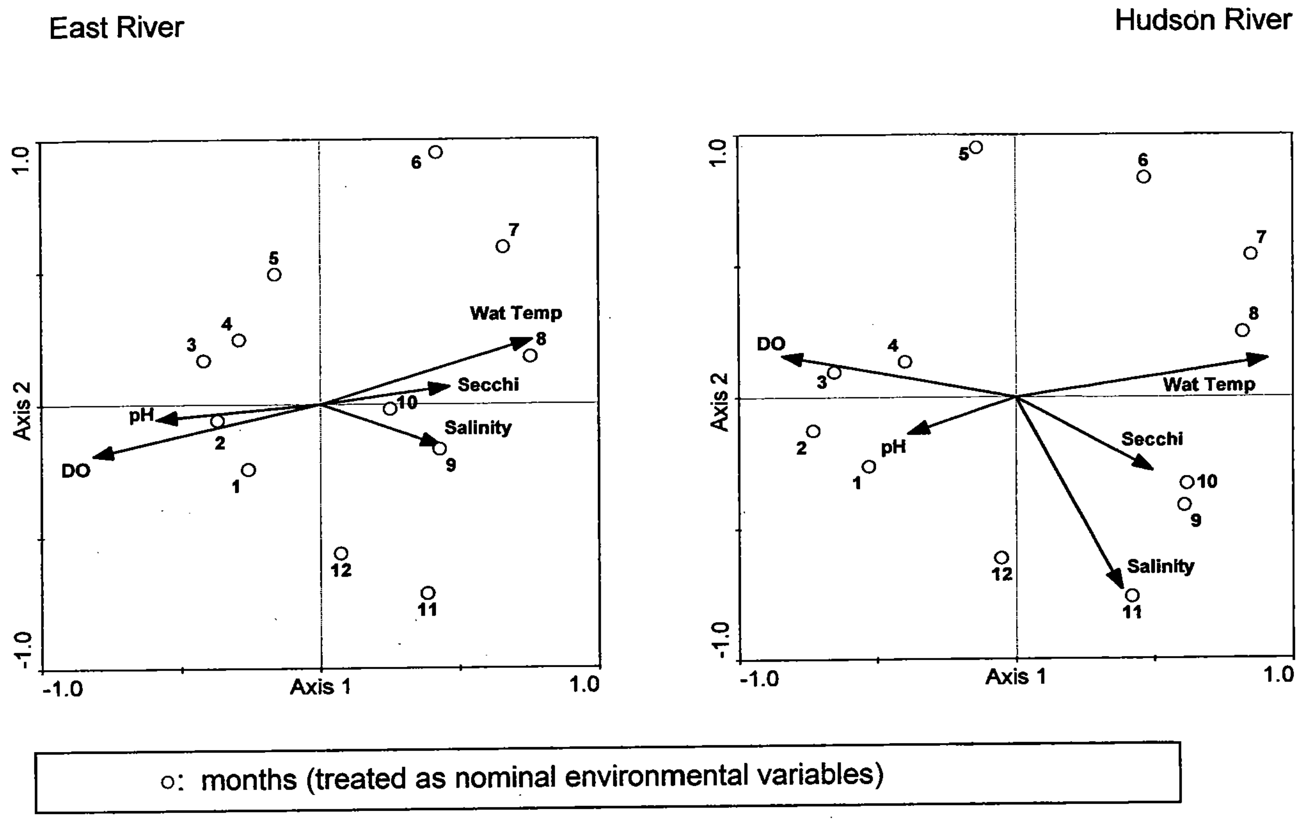
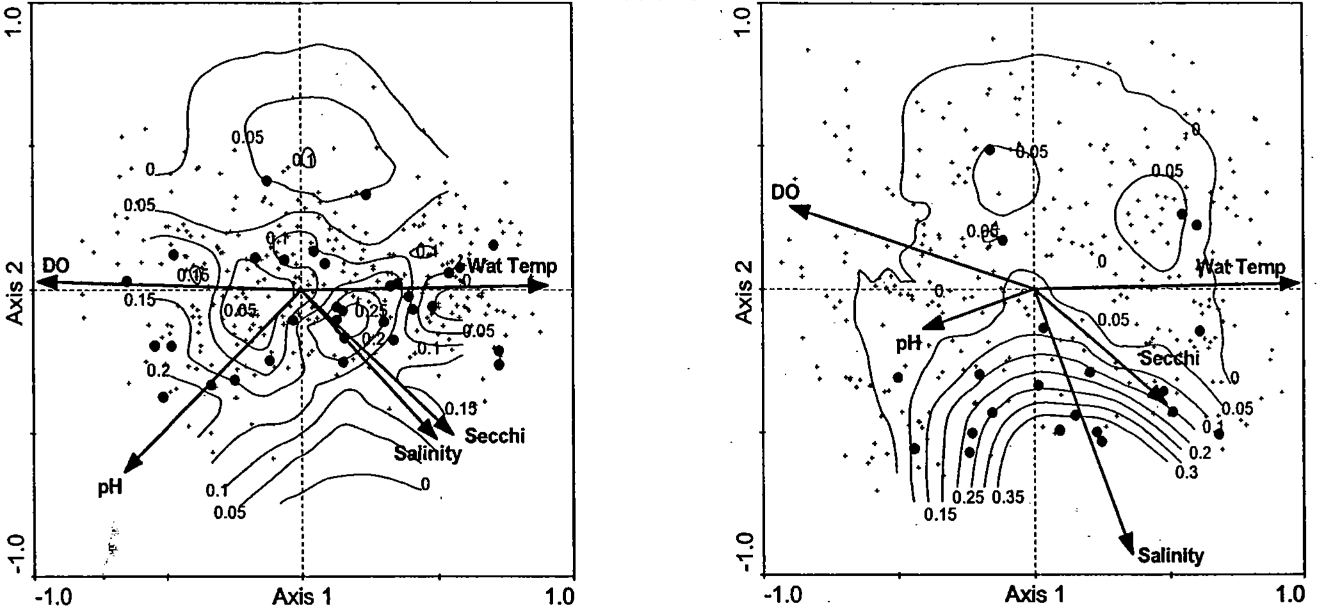

East River
Hudson River
Axis 2
Axis 1
1.0
-1.0
Wat Temp
Secchi
Salinity
pH
DO
1
2
3
4
5
6
7
8
9
10
11
12
o: months (treated as nominal environmental variables)

Fig. 8

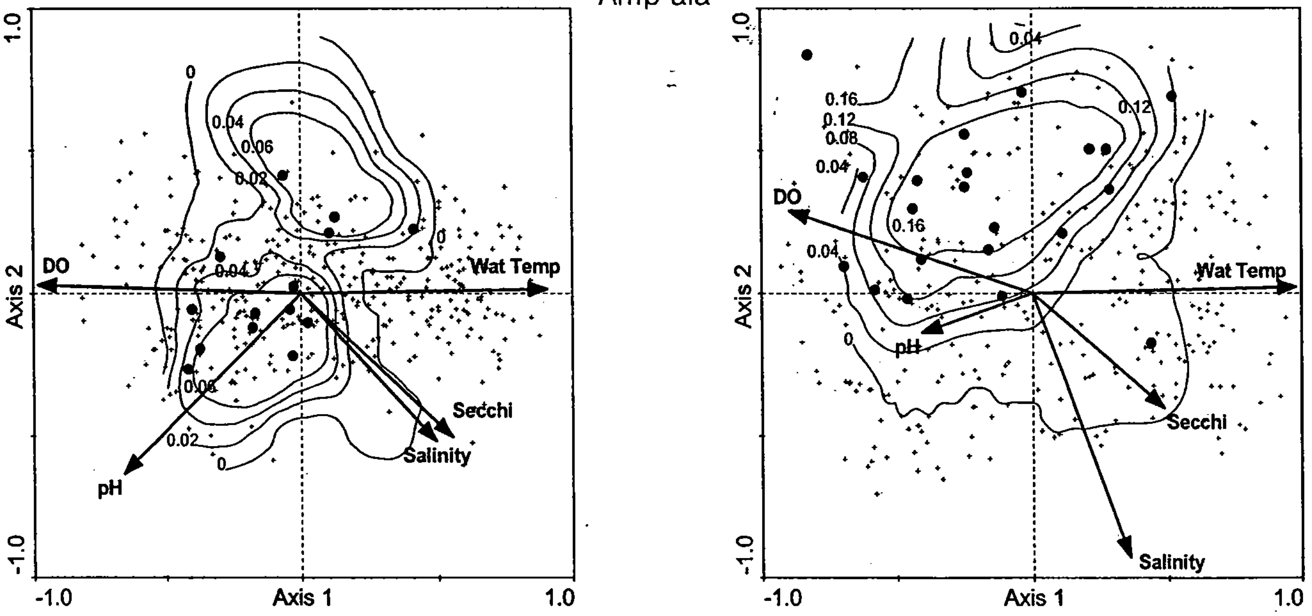
Fig. 9, 29 parts

East River Hudson River

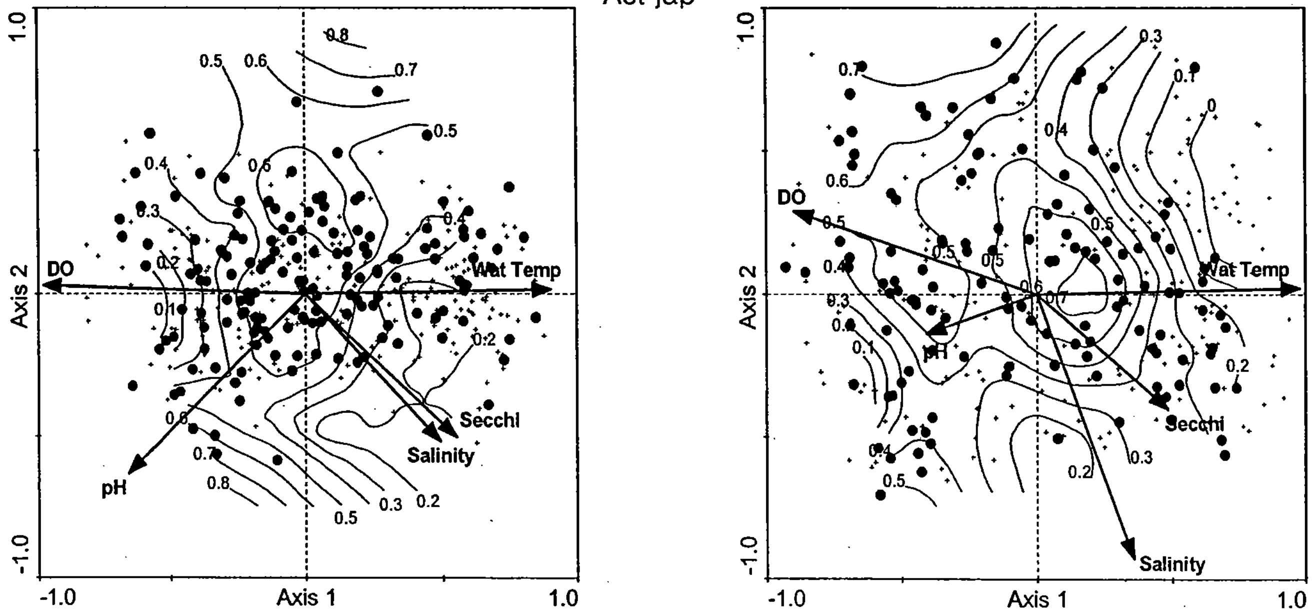
Act und

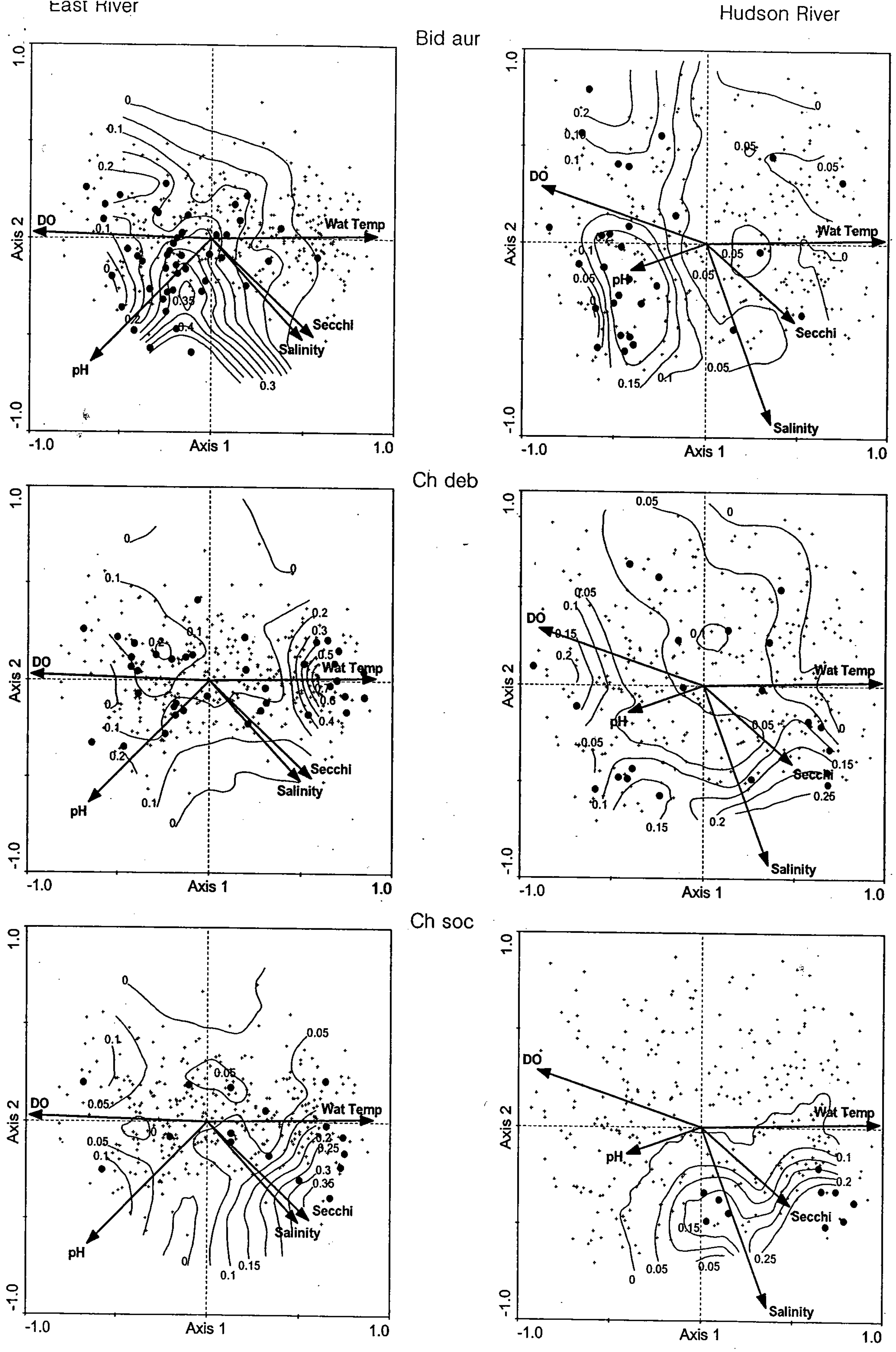


Amp ala



Ast jap



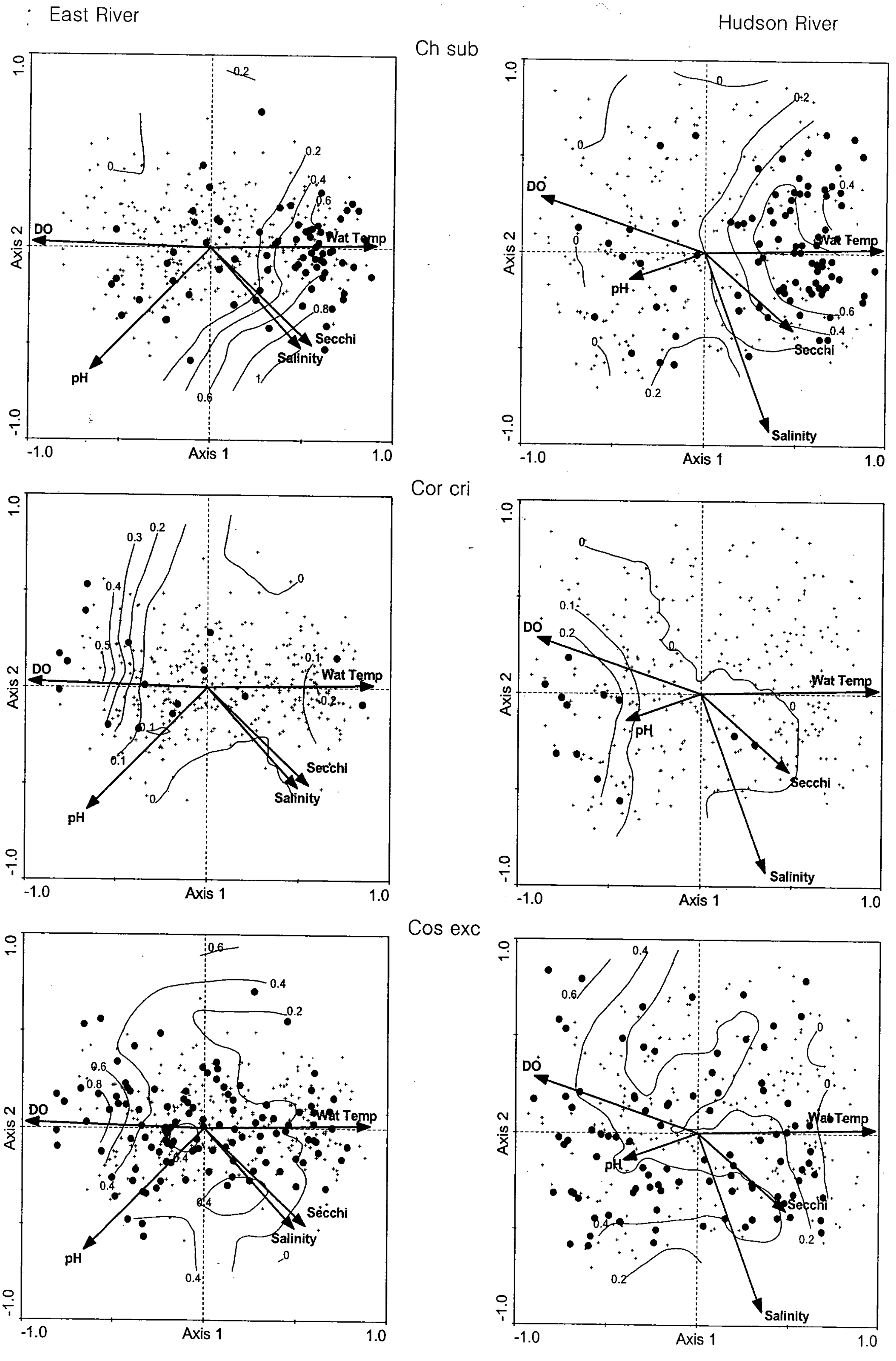

East River

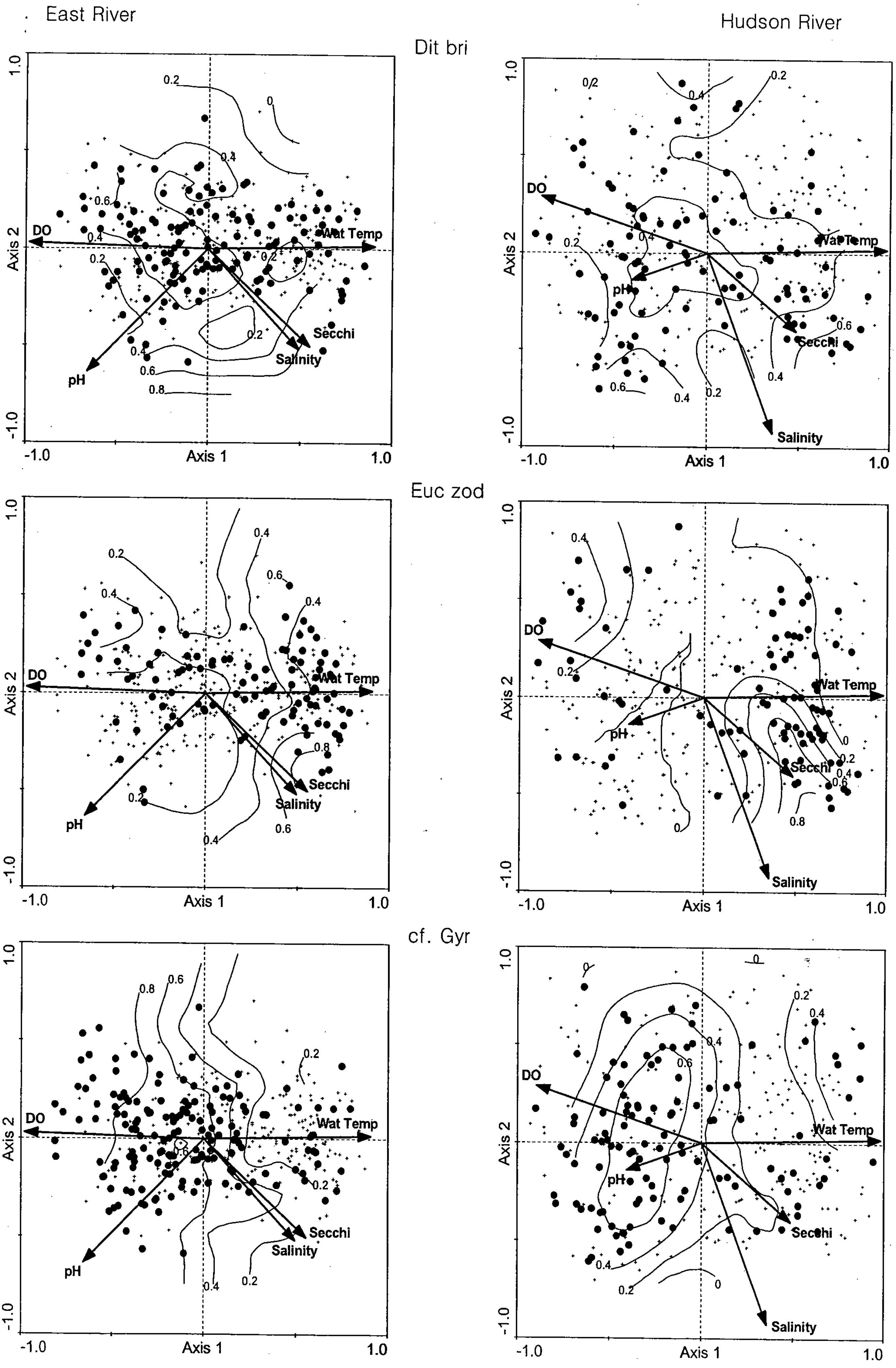
Hudson River

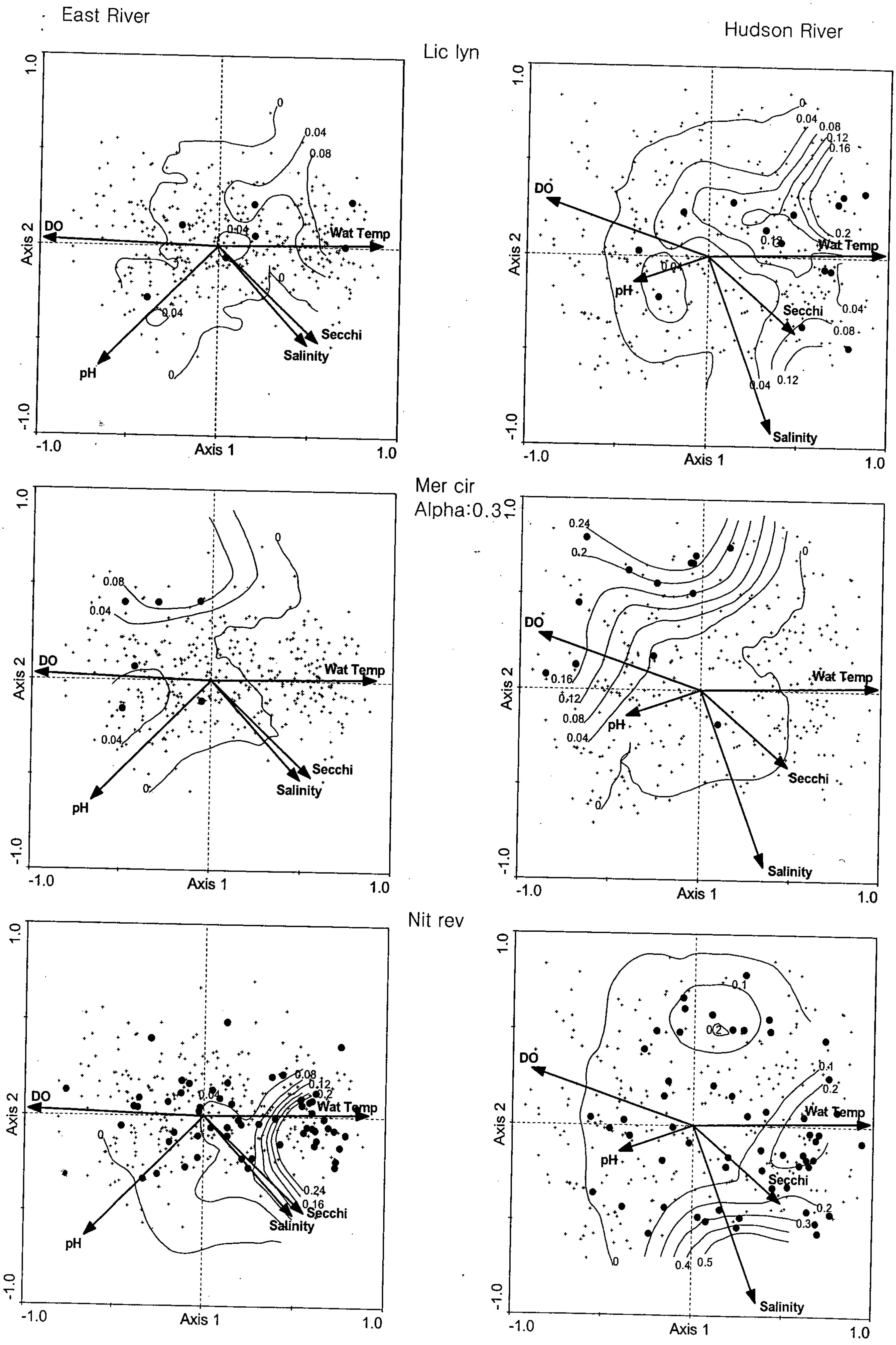
Bid aur

Ch deb

Ch soc


East River
Hudson River
Ch sub
Cor cri
Cos exc
Axis 1
Axis 2
-1.0
1.0
DO
Wat Temp
pH
Secchi
Salinity
0
0.1
0.2
0.3
0.4
0.5
0.6
0.8
1


East River
Hudson River
Dit bri
Euc zod
cf. Gyr
Axis 1
Axis 2
-1.0
1.0
DO
Wat Temp
pH
Secchi
Salinity
0
0.2
0.4
0.6
0.8

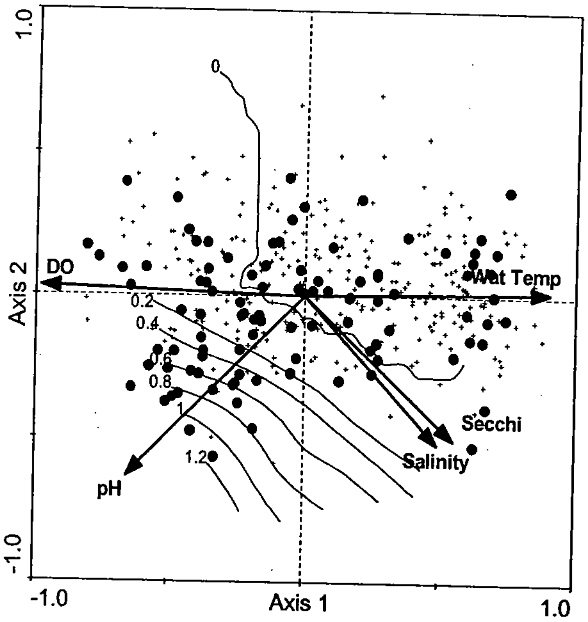

East River
Hudson River
Lic lyn
Mer cir
Alpha:0.3
Nit rev
Axis 1
Axis 2
-1.0
1.0
DO
Wat Temp
pH
Secchi
Salinity
0
0.04
0.08
0.12
0.16
0.2
0.24
0.1
0.3
0.4
0.5

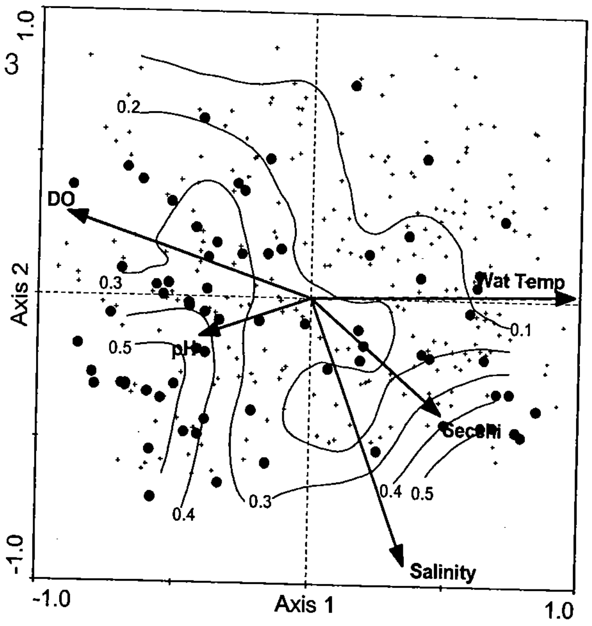
East River

Hudson River

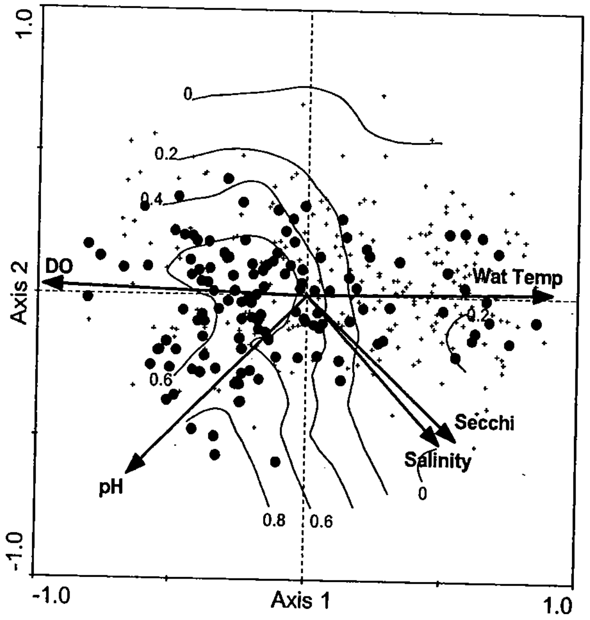
PsN ser

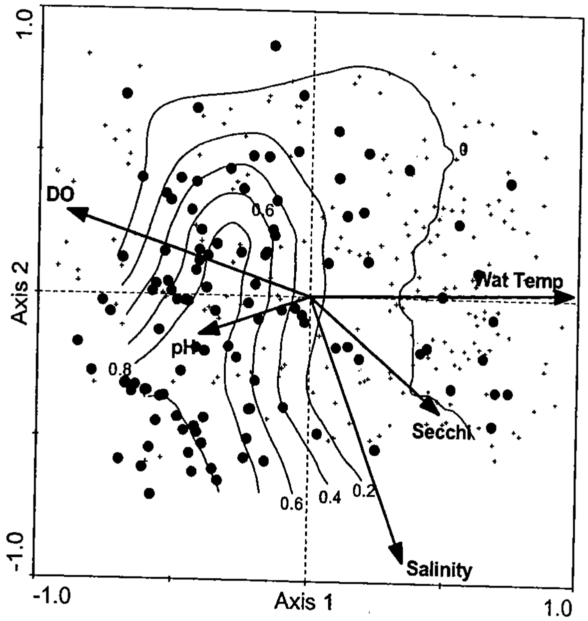
Alpha=0.3





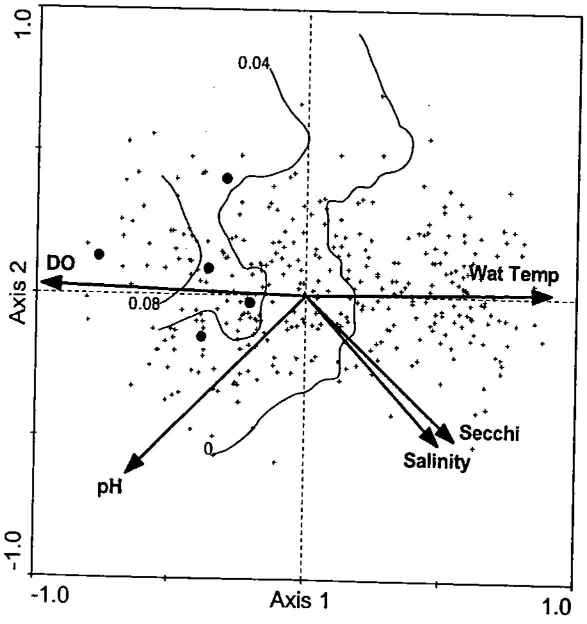
Rhi set

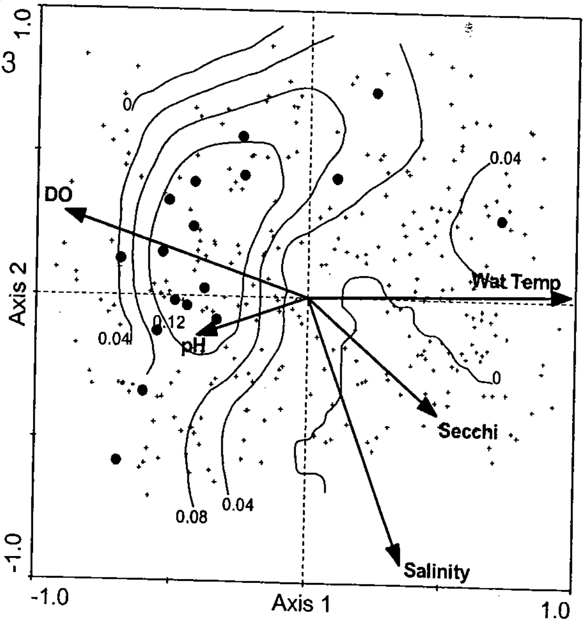




Gom apo

Alpha=0.3

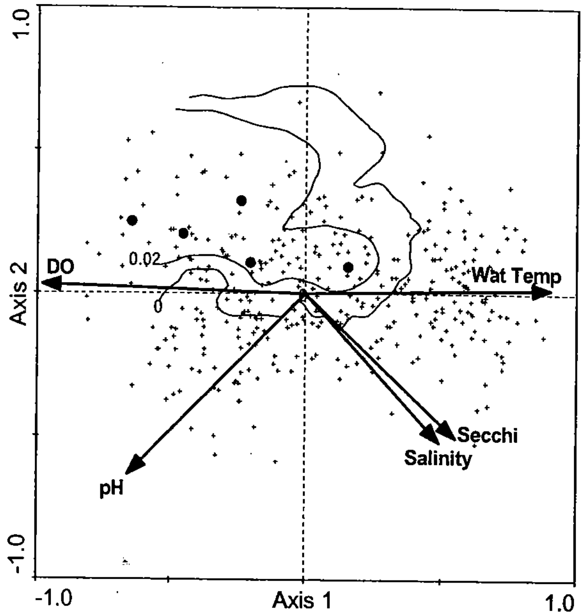


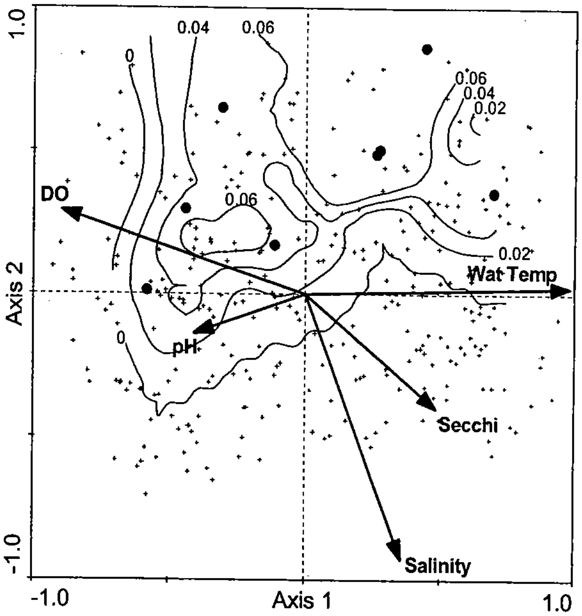


East River

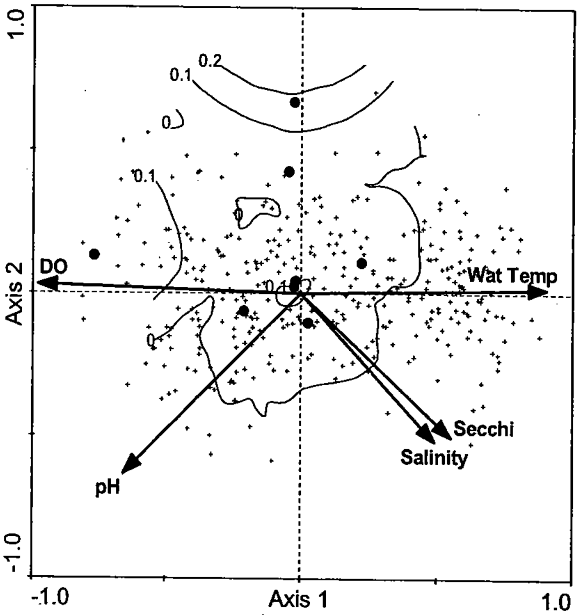
Hudson River

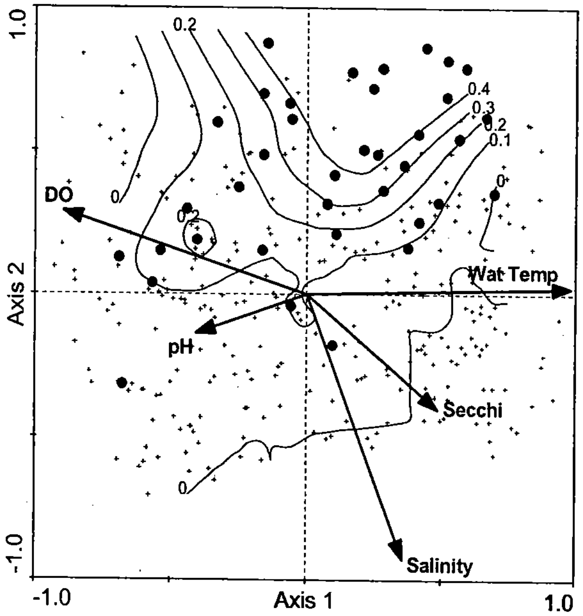
Ped dup





Sce qua





eug