Journal of the Indian Fisheries Association
21, 1991, 39-43

A MULTIVARIATE STATISTICAL STUDY WITH A FACTOR ANALYSIS OF RECENT PLANKTONIC FORAMINIFERAL DISTRIBUTION IN THE COROMANDEL COAST OF INDIA

K.V. JAYALAKSHMY AND K. KAMESWARA RAO
National Institute of Oceanography Regional Centre, Cochin-682 018, India.

ABSTRACT

A study of planktonic foraminiferal assemblages from 19 stations in the neritic and oceanic regions off the Coromandel coast, Bay of Bengal has been made using a multivariate statistical method termed as factor analysis. On the basis of abundance of 17 foraminiferal species, species were clustered into 5 groups with row normalisation and varimax rotation for Q-mode factor analysis. The 19 stations were also grouped into 5 groups with only 2 groups statistically significant using column normalisation and varimax rotation for R-mode analysis. This assemblage grouping method is suitable because groups of species/stations can explain the maximum amount of variation in them in relation to prevailing environmental conditions in the area of study.

INTRODUCTION

It is well known that recent planktonic foraminifera of the world have been subjected to increasingly detailed study for better interpretation of their distributional and ecological patterns with multivariate statistical techniques such as a principal component analysis (Malmgren and Kennett, 1973; Be' and Hutson , 1977) and a factor analysis (Thiede, 1975). However, such applications on these foraminifers in Indian coastal waters are very little. It is from this viewpoint, the present study which is based on species/stations grouping using factor analysis of the quantitative data of planktonic foraminiferal species in surface plankton tows collected off the east coast of India, has been undertaken to provide some information on this line of work. The purpose of this investigation is to verify whether such faunal grouping or distinct environmental conditions have prevailed in the neretic and oceanic areas separately by using factor analysis which is a method among the various multivariate statistical methods by which one can reduce a great number of species/stations into a few assemblages. These assemblages so obtained can be applied to mapping and the resulting pattern is then compared to any environmental parameter or a group of parameters about which information is available for faunal interpretation. It is also a method to delineate causal nexus. The causal approach to factor interpretation is to impute substantive form to the underlying and unknown causes (Rummel, 1975).

MATERIAL AND METHODS

The data on foraminiferal species are in the form of actual counts of foraminiferal species in 1000 m^3 and consists of 17 species in surface (0-10m) plankton tows from 19 stations of *R.V. Gaveshani* cruise 130 in January 1984, in waters of the Coromandel coast (Fig.1). The data were subjected to row (for Q-mode)

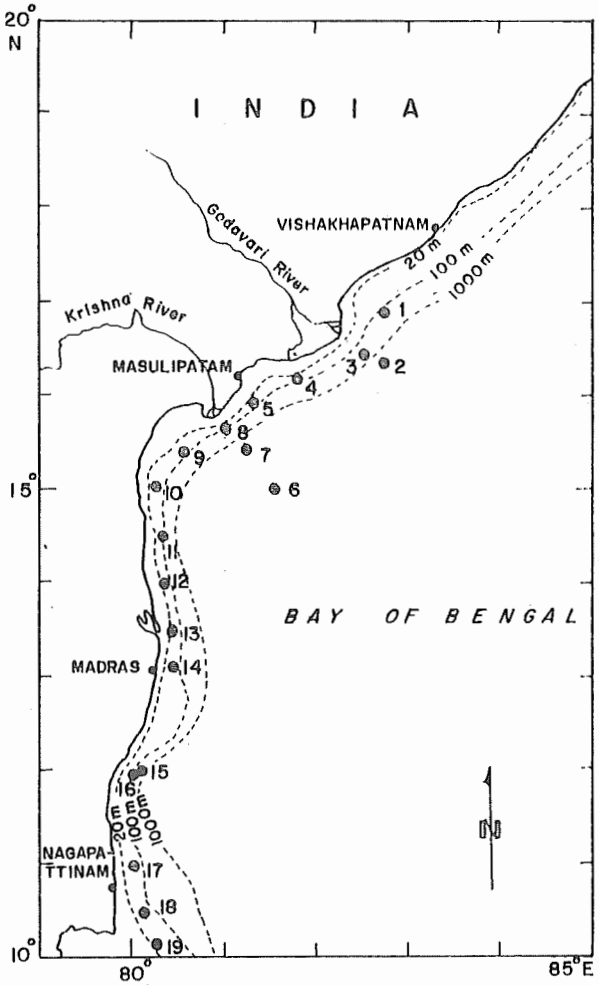

Fig. 1 : *Map showing sampling sites.*

and column (for R-mode) normalisation as transformation of data (Fernandes and Mahadevan, 1984), data being overdispersed (variance > mean) (Cassie, 1959 and Williamson, 1961). The transformed data had been represented as 17x19 matrix for Q-mode factor analysis and as 19x17 matrix for R-mode for grouping of species and stations respectively. To get unique groups for species and stations the criteria of varimax rotation to simple structure had been applied (Kaiser, 1958). The factor analysis model used was:

$$Y_{ij} = \sum_{q=1}^{q=m} a_{jq} x f_{qi} + b_j x s_j + c_j x e_j$$

where i=1,2,---s (no. of species), j=1,2,---n (no. of stations) for R-mode and i=1,2,---n and j=1,2,---s for Q-mode analysis, Yij is the transformed observation for ith species at the jth station for R-mode and it is the transformed observation for the jth species at the ith station for Q - mode. The range for i and j are as given above in each case (Angel and Fasham, 1973). To determine factor analysis , total variance was split into common variance and the specific plus error variance. Common variance called communality is defined (Harman, 1967) as :

$$h_j = \sum_{q=1}^{q=m} a^2_{jq}$$

here, ajq is the factor loading for jth species (j=1,2,---s) in Q -mode and for jth station (j=1,2,---n) in R-mode. The number of factors "m" extracted was determined by the number of eigen values $\lambda<1$ which are $\lambda>1$ because $\lambda < 1$ will not be statistically significant (Guttman, 1954). 5 factors calculated for Q -mode were statistically significant. But in R-mode only 2 factors were found to be significant since $\lambda<1$ for the other 3 factors. Factor loadings of factors 1 to 5 for stations were plotted with respect to latitude (N) and depth (m) (Fig.2)

Factor score distribution of species and stations were studied and verified for normality based on skewness (Snedecor and Cochran, 1967)

RESULTS AND DISCUSSION

Analysis of quantitative data of foraminiferal species in surface tows shows 5 major groups which can explain 83.93% of information about them based on maximum factor loadings and high communality (Table I).

In groups 1,2 and 3 only species with high negative factor loadings, while in groups 4 and 5 only species with high positive factor loadings are present. The five groups represent 5 distinct clusters which are coexisting, since they have the same sign for factor loadings in each group and are almost similar to the clusters of species obtained (Rao *et al.*, 1989) using multilinkage cluster analysis (Sanders, 1978) after log transformation of data (Clifford and Stephenson, 1975). It is observed that factor groups 1 and 4 show similarity with clusters 2 and 1 respectively and they include highly coexisting species which are almost similar in abundance. Cluster 4 (*H. pelagica* and *G. tumida*) is not grouped in factor analysis because eigen values corresponding to these species are not significant. Factor group 2 is not clustered in multilinkage analysis since *G. ungulata* shows more coexistence with *P. obliquiloculata* than with *B. digitata*. Factor groups 3 and 5 contain single species *P. obliquiloculata* and *O. universa* correspondingly which are rare. *G glutinata*, which has been clustered with *O. universa*, is grouped under factor 1 due to its greater abundance. *P. obliquiloculata* is not clustered with any species because of its low value of correlation coeffecient with rest of the species. Range percentage of occurrence is given for species within parentheses in each group.

Species belonging to group 1 such as *G. aequilateralis* (9.62-66.68%), *G. sacculifer*

Table I : *Factor loadings for foraminiferal species after varimax rotation (Q-mode)*

Factor	Species	Max. factor loading	Communality (%)	Eigen value	Closeness ratio (%)
1	*G. bulloides*	-0.8684	85		
	G. quinqueloba	-0.5914	58		
	G. aequilateralis	-0.9459	96		
				8.192	48.2
	G. ruber	-0.9240	92		
	G. sacculifer	-0.9516	96		
	G. glutinata	-0.9047	91		
2.	*B. digitata*	-0.9326	98		
				2.020	60.1
	G. ungulata	-0.6065	83		
3.	*P. obliquiloculata*	-0.8914	88	1.621	69.61
4.	*G. falconensis*	0.7327	97		
	G. conglobatus	0.9476	95	1.401	77.84
5.	*O. universa*	0.9818	99	1.034	83.93

(7.91-50.00%), *G. glutinata* (6.25-34.46%), *G. ruber* (2.94-17.64%), *G bulloides* (0.78-8.69%) and *G. quinqueloba* (0.53-3.45%) in the order of dominance are the most abundant species. Presence of *G. bulloides* at stations 1, 6 and 8 where it is highly abundant indicates that these stations are in the upwelling area.

Species of group 2 viz. *B. digitata* (0-2.38%) and *G. ungulata* (0.39-2.38%) showing low dominance are found only at station 4.

Group 3 contains only one species - *P. obliquiloculata* (2.22-2.26%) is a low dominant one, but homogeneously distributed species at stations 1 and 3.

Group 4 includes *G. falconensis* (O. 39-2.38%) and *G. conglobatus* (0.39-5.27%) and is confined to stations 7 and 8. This is a cluster which is more significantly coexisting in the oceanic than in nearshore waters. The only species occurring in group 5 — *O. universa* (0-1.04%) is a low dominant one.

Similarly 5 major groups of stations explaining about 92.70% of information regarding the same were obtained, based on maximum factor loadings from R - mode factor analysis (Table -2). Here only first 2 groups are statistically significant. The five groups are identified by the common species occurring in the stations of each group and range percentage of abundance given within brackets for the species is for the stations in each group.

Table II : *Factor loading for different sites after varimax rotation (R-mode)*

Factor	Sampling sites	Maximum factor loading	Communality(%)	Eigen value	Closeness ratio(%)
1	2	-0.6431	83		
				14.99	78.91
	6	-0.6458	83		
2	13	-0.8808	95		
				1.01	84.24
	18	-0.6911	86		
3	10	-0.6412	84		
	11	-0.6143	76	0.75	88.20
	14	-0.7542	90		
4	4	-0.8033	92	0.47	90.65
5	1	0.4874	53		
	5	0.7733	90		
	9	0.6560	78		
	12	0.7694	90	0.39	92.70
	16	0.7836	90		
	17	0.7018	77		
	19	0.7943	89		

Group 1 includes deep stations 2 and 6. The abundant species occurring in these stations are *G. aequilateralis* (15.94-32.35%), *G. ruber* (8.94-17.64%), *G. sacculifer* (7.91-14.70%), *N. dutertrei* (0.00-4.40%) and *G. glutinata* (11.76-30.22%) while the remaining species are sparse in occurrence. Considering all the sites, *G. ruber* and *N. dutertrei* show maximum abundance at the station 2 and 6 respectively . *N. dutertrei* being a productivity -cum-upwelling indicator species is found at station 6 which is in the upwelling region.

Stations 13 and 18 which constitute group 2 show dominance of *G. aequilateralis* (27.29-66.68%) and *G. sacculifer* (33.31-45.46%) and the fauna consists of these species only.

Stations 10,11 and 14 which form group 3 are dominated by the species, *G. aequilateralis* (21.87-23.53%), *G. ruber* (2.94-16.67%), *G. sacculifer* (36.84-50.00%), *G. tumida* (2.08-2.94%) and *G. glutinata* (6.24-17.64%) with higher abundances of the above mentioned species at station 14.

In station 4 which forms group 4 includes all dominant species such as *G. aequilateralis* (28.57%) *G. sacculifer* (22.22%), *G. glutinata* (16.67%), *G. sacculifer* (13.34%), *G. quinqueloba* (4.76%), *G. ruber* (2.38%) and *G. ungulata* (2.38%). As a consequence of its location in the upwelling area, group 4 has high diversity and high evenness.

Group 5 contains stations 1,5,9,12,16, 17 and 19 which have low factor loadings with a range of 0.4874 — 0.7943 and have all the abundant species with almost the same abundance at all the stations except at station 1 where species *G. bulloides, G. aequilateralis, G. ruber* and *G. glutinata* occur nearly 8 to 9 times more than that occuring in other stations in this group and *G. aequilatea* ralis and *G. glutinata* constitute, more than double in concentration at station 9 compared to that occurring at other stations of this group.

Fig. 2 depicts the contours corresponding to factor loadings for 5 factors of different stations with reference to latitude (N) and depth (m). From this mapping it is obvious that there is a marked difference between the nereitic and oceanic regions in that it shows all the inshore stations have the same range for the factor loadings which are higher than that of oceanic regions.

As stated earlier the groups of species obtained from factor analysis are similar to groups of species obtained by cluster analysis. Likewise stations with high factor loadings in various groups derived from factor analysis (Table II) have great affinity as observed from a trellis

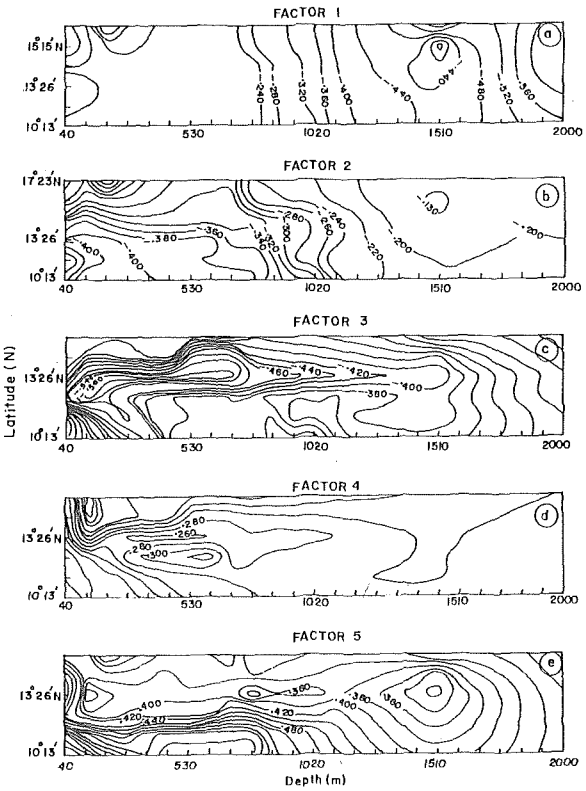

Fig. 2 : *Mapping of factor loadings of stations in relation to depth (m) and latitude (N) for factors 1 to 5 (Figs. a to e) from R - mode factor analysis.*

diagram (Rao *et al.*, 1989). It further implies that the species belonging to a group require the same environmental characteristics (Castro *et al.*, 1991). Similarly stations coming under each group have almost the same environmental conditions favourable to species dominating that region.

It is found that factor score distribution of factors 2, 3, 4 and 5 for stations is highly abnormal, being negatively skewed for factors 2 and 3 and positively skewed for factors 4 and 5 while that of factor 1 is almost symmetrical. Therefore at open sea sites, species population is a single homogenous one with less variation in the species distribution pattern. But in nearshore waters it is a continuously varying population due to species competition and mixing.

Factor score distribution of species being not significantly skewed justifies that distinct environmental conditions prevail in oceanic and nereitic areas off the coast but they do not change continuously with latitude.

Thus, the five groups of species obtained from factor analysis are five species assemblages dominating specific areas with varying environmental conditions and the two significant groups of stations in this study give a clear-cut demarcation between oceanic and neritic species assemblages. Factor analysis results are more useful because factor scores provide objective average distribution patterns for various species associations which may be useful for modelling and simulation studies. Further, unlike cluster analysis it does not impose a hierarchical structure on the data for which there is no prior evidence and it gives an easier interpretation of data.

ACKNOWLEDGEMENTS

The authors are grateful to Dr. B.N. Desai, Director and Dr. V.N. Sankaranarayanan, Scientist-in-Charge, National Institute of Oceanography, Regional Centre, Cochin, for encouragement and facilities.

REFERENCES

Angle, M.V. and **Fasham, M.J.R.** 1973. Sond Cruise 1965: Factor and cluster analyses of the plankton results; a general summary. *J. mar. Biol. Ass. U.K.* **53**:185-231.

Be, A.W.H. and **Hutson, W.H.**, 1977. Ecology of planktonic Foraminifera and biogeographic patterns of life and fossil assemblages in the Indian Ocean. *Micropaleont.* **23**: 369-414.

Cassie, R.M. 1959. An experimental study of factors inducing aggregation in marine plankton. *N.Z.J. Sci.* **2**: 339-365.

Castro, L.A., Bernal, P.A. and **Gonzalez, H.E.** 1991. Vertical distribution of copepods and the utilisation of chlorophyll *a* rich layer within Concepcion Bay, Chile. *Estuar. Coast. Shelf. Sci.* **32**:243-256.

Clifford, H.T. and **Stephenson, W.** 1975. *An Introduction to Numerical Classification,* Academic Press, New York., 202 pp.

Fernandes, A.A. and **Mahadevan, R.** 1984. *Program Package for Factor Analysis.* Tech. Rep., (National Institute of Oceanography, Dona Paula, Goa, India). 1984, 54 pp.

Guttman, L. 1954. Some necessary conditions for common factor analysis. *Psychometrika,* **19**:149-161.

Harman, H.H. 1967. *Modern Factor Analysis.* University of Chicago, Chicago. 474 pp.

Kaiser, H.F. 1958. The varimax criterion for analytical rotation in factor analysis. *Psychometrika,* **23**:181-200.

Malmgren, B.A. and **Kennett, J.P.** 1973. Recent planktonic foraminiferal distribution in high latitudes of the south Pacific: A multivariate statistical study. *Paleogeogr. Paleoclimatol. Palaecol.* **14**:127-136.

Rao, K.K., Jayalakshmy, K.V. Kumaran, S., Balasubramanian, T. and Kutty, M.K. 1989. Planktonic Foraminifera in waters off the Coromandel coast, Bay of Bengal. *Indian J. Mar. Sci.* **18**: 1-7.

Rummel, R.J. 1975. *Applied Factor Analysis,* John Wiley & Sons, New York, 792 pp.

Sanders, H.L. 1978. Florida oil spill impact on the Buzzards Bay benthic fauna, west Falmouth. *J. Fish. Res. Board Can.* **35**:717-730.

Snedecor, G.W. and Cochran, W.G. 1967. *Statistical Methods.* Oxford and I.B.H. Publishing Co., Calcutta, 593 pp.

Thiede, J. 1975. Distribution of Foraminifera in surface waters of a coastal upwelling area. *Nature,* Lond. **253**:712-714.

Williamson, M.H. 1961. A method for studying the relation of plankton variations to hydrography. *Bull. mar. Ecol.* **5**:224-229.