Is species level identification essential for environmental impact studies?

S. Ajmal Khan

Identification of organisms up to the species level is essential in biological studies and it is generally believed that omission of species level identification will lead to loss of information. However when there is a need to make a quick assessment of the impact of a natural calamity like tsunami on biological organisms, species level identification takes time. What is the way out then? This article explains with examples the advantages of species removal, methods for species removal, aggregation of data at higher levels of taxonomy like genus, family and phylum and suitability of aggregation data for univariate, graphical and multivariate methods of analysis. Such an approach is called as Taxonomic Sufficiency (TS). This article discusses the advantages and disadvantages of TS. The TS approach can be used in India not only with respect to tsunami but also El Nino, oil spill and earthquake.

Keywords: Environmental impact assessment, species level identification, taxonomic sufficiency.

SPECIES name is a functional label using which various pieces of information concerning that organism can be retrieved and assembled. Identification of species is essential in descriptive, taxonomical, ecological and biodiversity studies. It is also of help in finding out the species, which characterize different habitats/environs. Endemism (endemic species) is better explained and understood at species level. During analysis of data, it is important to include all the species present in each site as the omission of species level identity will affect the outcome of univariate and distributional methods[1]. Maurer[2] opined that exclusion of species level information during the monitoring studies leads to unacceptable loss of ecological information. Resh and Unzicker[3] noted that freshwater macroinvertebrate congeners often show a wide range of pollution tolerances and concluded that species-level identification is essential for water quality monitoring. It has also been reported that grouping species into higher levels tends to reduce the natural variability of some community structure measures. Sometimes a response to pollution is detectable only at the species level[4]. For example, *Ampelisca sarsi* survived at low densities after the *Amoco Cadiz* oil spill, when all other species of the genus had disappeared completely[5]. Similarly some Cirratulidae (e.g. *Chaetosome setosa*) or Spionidae (e.g. *Scolelepis fuliginosa*) tend to increase in number in response to increased organic matter levels, while the other members of these families show no response to such perturbation. Some species appear to be useful bioindicators of particular types of perturbation (e.g. oil spill, organic matter, metallic contaminants) to which other species are more or less resistant. That way, species level identification is of immense use.

However during calamities like tsunami, El-Nino, oil-spill, earthquake, etc., when there is a need to find out the impact on the biota very quickly and to produce a report at the shortest possible time, is species level identification essential? In many instances, aggregations of the species data to higher taxonomic levels have been made and the resultant data matrices have been subjected to several forms of statistical analyses to see how much information has been lost compared with the full species-level analysis. The results of such analyses have made it clear that little information appears to be lost due to aggregation. Such an approach is called as 'Taxonomic Sufficiency (TS)'. It was defined[6] to mean, 'in any project, organisms must be identified to a level (species, genera, family, etc.) which balances the need to indicate the biology of organisms present with accuracy in making the identifications' or in other words TS is 'the identification of taxa to taxonomic levels higher than species without significant loss of information in detecting changes in assemblages exposed to environmental stress'. Many investigators have used TS and have suggested that this concept is appropriate for identifying the effects of impact on marine and freshwater communities[7].

This paper explains with appropriate examples how species could be eliminated; the species data could be aggregated to higher levels and where (tsunami notwithstanding) TS could be used. The advantages are also explained.

S. Ajmal Khan is in the Centre of Advanced Study in Marine Biology, Annamalai University, Parangipettai 608 502, India.
e-mail: s_ajmalkhan@rediffmail.com

Species elimination

When the number of species is more, rare species could be eliminated for the purpose of processing the data quickly. It will not lead to loss of information[1]. To cite an example, in Norway four replicate grab samples were taken at each of six sites (A–E and G) in a fjordic complex (Frierfjord/Langesundfjord) linked to Oslofjord[8]. Site A represented the control site, and others, the contaminated sites (Figure 1). The configurations produced from an ordination map involving all the 110 species recorded at six stations in Norway were compared with a similar one produced involving just 19 arbitrarily selected species. It could be seen in the figure that the ordinations are remarkably similar in the way in which they discriminate between the sites (only difference noticed was that the stations G and E were transposed in location). The above example proves that species elimination does not lead to any loss of information.

The way in which the species can be eliminated requires careful consideration. A commonly employed method is to remove those species that are rare in respect of their total abundance at all stations in the survey, for example those species comprising less than 1 or 2% of the total number of individuals. However this will be dangerous in situations where total abundance between stations varies greatly, as is often the case. Situations commonly seen are certain stations have a low overall abundance of organisms, but there may be many species which are absolutely characteristic of those stations. Using the above method of species reduction, all these species could be eliminated. It is also recommended that species accounting for >*p*% of the total score (abundance or biomass) in any one sample is retained (*p* is chosen to reduce species to the required number; typically $p = 3$ or 4).

Species aggregation

The painstaking work involved in sorting and identifying samples to the species level has resulted in community analysis for environmental impact studies being traditionally regarded as labour-intensive, time-consuming and therefore relatively expensive. One practical means of overcoming this problem is to exploit the redundancy in community data by analysing the samples to higher taxonomic levels, such as family, rather than to species. In the case of marine macro- and meio-benthos, aggregations of the species data to higher taxonomic levels have been made and the resultant data were subjected to several forms of statistical analysis to see how much information has been lost compared to the full species-level analysis. While the family level aggregation was effective for the macrobenthos, the generic level aggregation was useful in the case of meio-benthos. That way TS has to be found out for various groups of organisms and various situations.

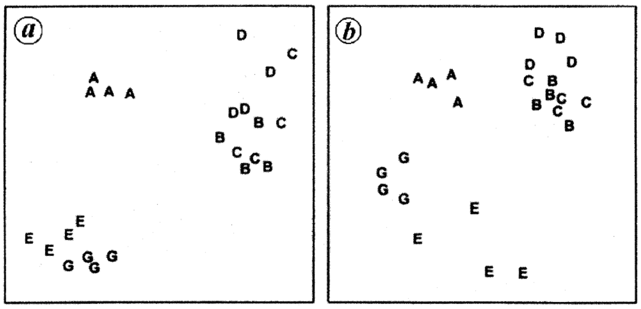

Figure 1. MDS plots for data of Frierfjord macrofauna. ***a***, Using all 110 species; ***b***, Using 19 arbitrarily selected species.

Methods for treating aggregation data

Taxonomic levels higher than species are most appropriate for multivariate methods of studies. All ordination (principal component analysis and multi-dimensional scaling)/clustering techniques are amenable to aggregation. It has been reported that aggregation to higher taxa reflects well-defined pollution/impact gradients more closely than species. Data after aggregation were also found to be effective for the distributional method such as Abundance-Biomass Comparison Curve and for univariate methods. Diversity indices defined at hierarchical taxonomic levels for comparative purposes were also useful.

Analysis of aggregation data with univariate method

In South Tikus Island[9], Thousand Islands Indonesia 10 replicate transects across a single coral-reef site were sampled over six years during 1981, 1983, 1984, 1985, 1987 and 1988. The number of taxa so also Shannon diversity (H') at species and genus levels not only showed immediate post-El Nino drop (1982–83) but also subsequent suggestion of partial recovery (Figure 2). What is noteworthy is the admirable similarity in the trend at the species and genus levels treatment of data.

Analysis of aggregation data with graphical method

Time series samples of macrobenthos were collected over the period 1963–1973 inclusive, in a sea loch system on the west coast of Scotland[10]. Since 1966, the pulp-mill effluent was discharged to the sea lochs with the rate increasing in 1970 and a significant reduction in 1972. Shannon diversity and ABC (Abundance Biomass Comparison) curve were used to find out the impact of effluent on benthos. While Figure 3 shows the species level results, Figure 4 shows the results of species data aggregated to family level. It is evident that the curves are identical, indicating that there would have been no loss of information had the samples only been sorted originally into families.

Analysis of aggregation data with multivariate methods

Soft-bottom mesocosms

In the soft-bottom mesocosms at Solbergstrand, Norway, box-corers of sublittoral sediment were subjected to three levels of particulate organic enrichment (L = low dose, H = high dose and C = control), there being four replicates from each treatment. After 56 days, the meiobenthic communities were analysed. Figure 5 [multi-dimensional scaling (MDS)] shows clear differences in community structure of copepods between treatments at species level, which are also equally evident when the species data were aggregated into genera and families.

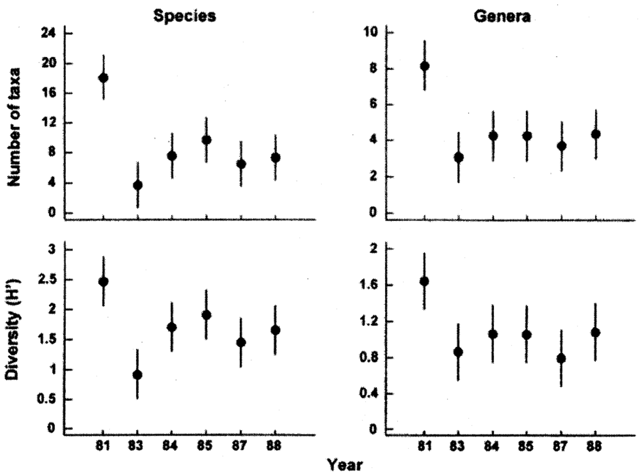

Figure 2. Impact of El Nino on Indonesian reef corals-means with 95% confidence intervals for number of taxa and Shannon diversity for species data (left) and data aggregated into genera (right).

Loch Linnhe macrobenthos

MDS ordinations of the Loch Linnhe macrobenthos are given in Figure 6 using both double square root and untransformed abundance data. The ordinations have been performed separately using all 115 species, 45 families and phyla. In all ordinations (species, family and phyla), the separation to the right of the years 1970, 1971 and 1972 associated with increasing pollution levels and community stress and a return to the left in 1973 associated with reduced pollution levels and community stress are evident. This pattern is equally clear at all levels of taxonomic aggregation. Again, the separation of the most polluted years is most distinct at the phylum level, at least for the double root transformed data (and the configuration is more linear with respect to the pollution gradient at the phylum level for the untransformed data).

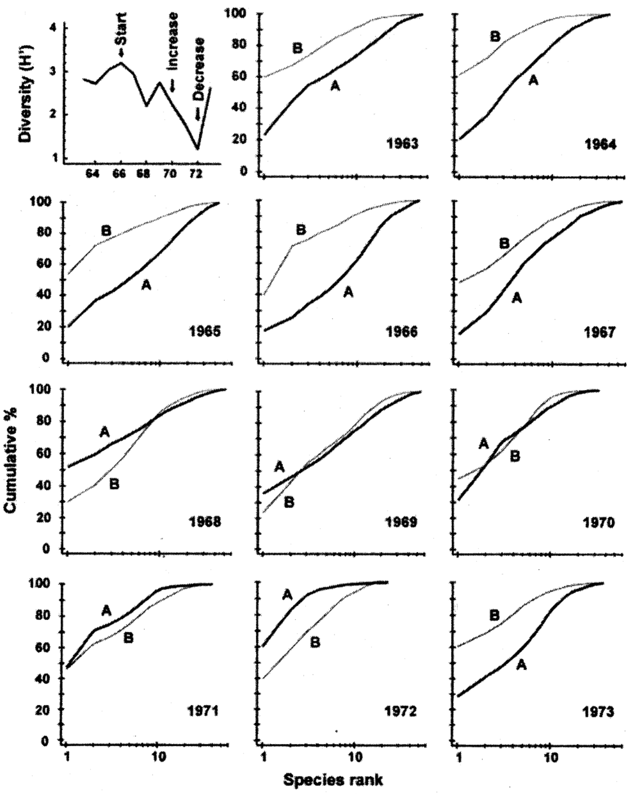

Figure 3. Shannon diversity and ABC (abundance – thick line and biomass – thin line) plots for the species level data of Loch Linnhe macrofauna collected over 11 years (1963–1973).

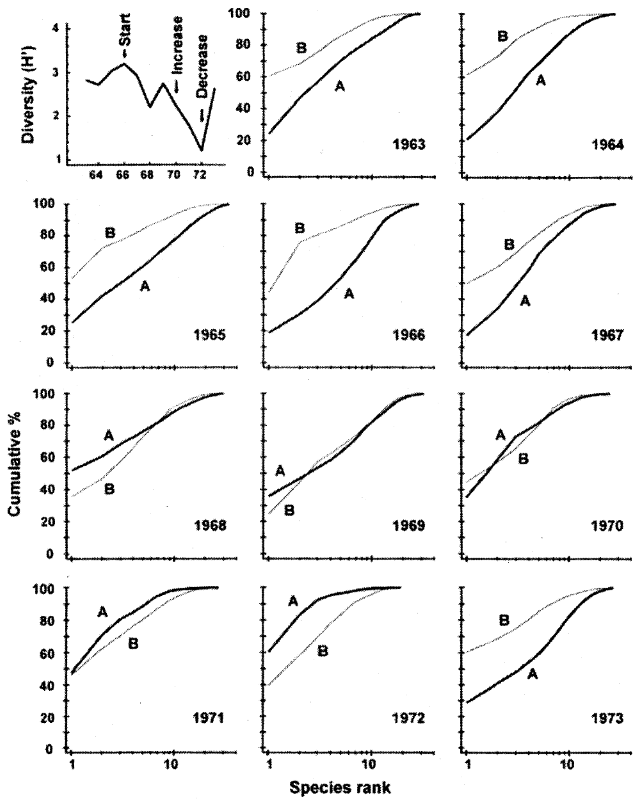

Figure 4. Shannon diversity and ABC (abundance – thick line and biomass – thin line) plots for the family level data of Loch Linnhe macrofauna collected over 11 years (1963–1973).

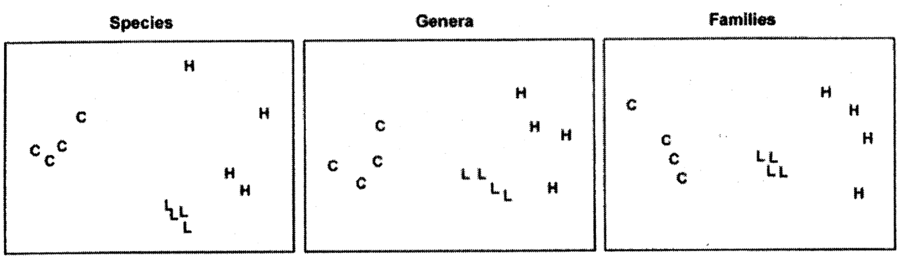

Figure 5. MDS plots for data of nutrient-enrichment experiment at Solbergstrand analysed at species, genera and family levels.

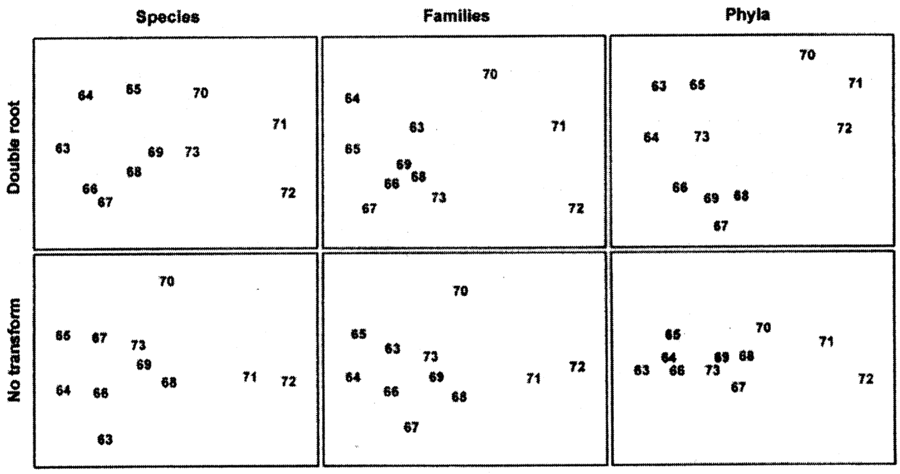

Figure 6. MDS plots for data of Loch Linnhe macrofauna analysed without transformation and with transformation at species, family and phylum levels.

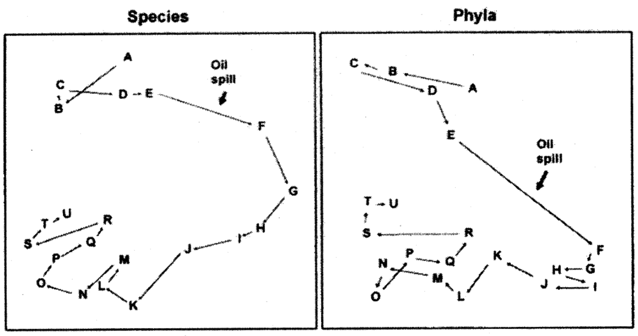

Figure 7. MDS for species and phylum level data of macrobenthos collected on 21 occasions (A–U) in the Bay of Morlaix spanning the period of wreck of *Amoco-Cadiz* in March 1978.

Amoco-Cadiz oil-spill

Macrofauna species were sampled at station 'Pierre Noire' in the Bay of Morlaix on 21 occasions between April 1977 and February 1982 spanning the period of the wreck of the *Amoco-Cadiz* in March 1978. The sampling site was 40 km away from the initial tanker disaster but substantial coastal oil slicks resulted. The species were aggregated into five phyla: Annelida, Mollusca, Arthropoda, Echinodermata and 'others'.

Species level MDS and MDS for the aggregated data are shown in Figure 7. The MDS for the aggregated data closely reflects the timing of pollution events, the configuration being slightly more linear than in the species level analysis. All pre-spill samples (A–E) are in the top left of the configuration, the immediate post-spill sample (F) shifts abruptly to the bottom right after which there is a gradual recovery in the pre-spill direction. It is of interest to note that in the species level analysis, although the results are similar, the immediate post-spill response is rather more gradual and the community response at the phylum level is remarkably clear.

Indonesian reef corals

Due to the El Nino of 1982–83, extensive bleaching of reef corals was noticed throughout the Pacific. Figure 8 shows the coral community response at South Pari Island over six years in the period 1981–1988 based on ten replicate line transects along which coral species cover was determined. The immediate post-El Nino shift on the species MDS and a circuitous return towards the pre-El Nino conditions are clear. This is closely reflected in the genus level analysis also.

Ekofish oil-platform macrobenthos

Changes in the community structure of the soft-bottom benthic macrofauna in relation to oil drilling activity at

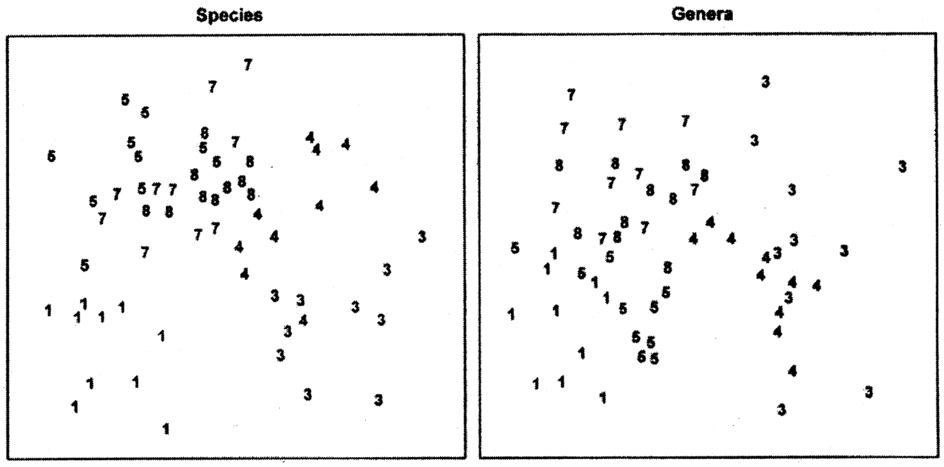

Figure 8. MDS for species and genus levels data collected at South Pari Island in Indonesia over six years 1981–1988 with El Nino occurring in 1982–83 (1 = 1981 and 3 = 1983).

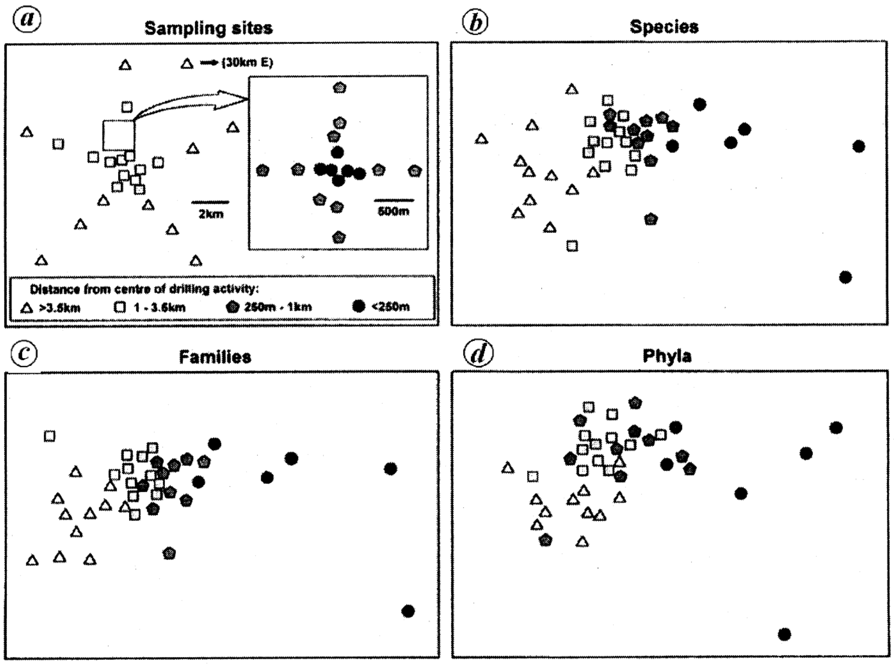

Figure 9. MDS for data of Ekofisk oil-platform macrobenthos analysed at species, family and phylum levels.

the Ekofish platform in the North Sea were studied[11,12]. The positions of the 39 sampling stations around the rig are coded by different symbol and shading conventions (Figure 9) according to their distance from the centre of drilling activity. In the species level MDS, community composition in all the zones is distinct, and there is a clear gradation of change from the inner (filled circle) to the outer (open triangle) zones. The MDS was repeated with the species data aggregated into family and phylum levels. The separation of sites is still clear. Thus phylum level analyses are surprisingly sensitive in detecting impacts/pollution-induced community change and little information is lost by working at the family level.

As elaborated above, the impact of tsunami on biota could also be found out at higher levels of taxonomy using the univariate, graphical and multivariate tools.

Advantages of aggregation

Laboratory advantages

Aggregation implies reduction in the costs associated with sorting and identification of organisms. It is quicker and easier to train personnel to sort higher taxonomic levels than species, and the risk of potential taxonomic classification error is lower at a higher level of identification. As fewer and fewer taxonomic experts are employed in marine laboratories, personnel without a high-level taxonomic expertise will be required to do the sorting work. Of course, sampling, sorting and identification are labour-intensive, and benthic infaunal monitoring has been omitted from some monitoring programmes for reasons of cost alone[7]. It was found that the estimated costs of genus, family, order and phylum-level identifications were respectively 23%, 55%, 80% and 95% less than that of species-level identification[13]. However the relationship between cost saving and taxonomic level is not straightforward. For example, the time spent dividing organisms into families will depend on (i) the number of species within each family, (ii) whether the numerically dominant species belong to several taxonomically complicated families or to a few taxonomically uncomplicated families and (iii) the taxonomic expertise available[4]. The potential for saving time and thereby money will therefore vary greatly from case to case.

Data analysis advantages

Analysis of higher-level taxonomic groups might more clearly reflect pollution gradients and be less affected by natural nuisance variables than species analyses. Aggregation of species to family level reduces the effect of single dominant species. Community responses to pollution should be more easily detected above the natural noise at higher taxonomic levels. A number of hypotheses have been put forward to explain the sufficiency of higher taxonomic levels compared to species in marine pollution studies.

One of the hypotheses is that natural environmental variables typically influence community structure at lower taxonomic levels than pollution, natural variables acting by species replacement and pollution by changing proportion of major taxa in the community. It was hypothesized that when a given stress increases, it successively affects individuals, species, genera and families[4]. Consequently, the community's response to increasing stress is manifested at higher and higher taxonomic levels. According to this hypothesis, the taxonomic level required to detect changes in the community could depend on the degree of environmental alteration.

Maurer[2] defended marine taxonomists and considered that the generalization of TS will tend to discourage interest in taxonomy in the context of biodiversity. In my opinion, a need for experts in marine invertebrate taxonomy in teaching and research roles in universities and research organisations is not incompatible with the need for technical professionals specialized in marine environmental management and pollution monitoring.

In conclusion, fast and cost-effective and accurate littoral pollution monitoring and impact studies due to natural calamities require only family-level identification of taxa, although in some cases it may be of interest to identify particular species whose presence or abundance is indicative of the impacts. The usefulness of TS in finding out the impact of various natural disasters including tsunami on marine organisms should be tested and made available in India.

1. Clarke, K. R. and Warwick, R. M., *Changes in Marine Communities: An Approach to Statistical Analysis and Interpretations*, Playmouth Marine Laboratory, Plymouth, UK, 2001, 2nd edn.
2. Maurer, D., The dark side of taxonomic sufficiency. *Mar. Pollut. Bull.*, 2000, **40**, 98–101.
3. Resh, D. J. and Unzicker, T. C., Water quality monitoring and aquatic organisms: importance of species identification. *J. Water Pollut. Control Fed.*, 1975, **47**, 9–19.
4. Ferraro, S. P. and Cole, F. A., Taxonomic level and sample size sufficient for assessing pollution impacts on the Southern Califironia Bight macrobenthos. *Mar. Ecol. Prog. Ser.*, 1990, **67**, 251–262.
5. Dauvin, J. C., The fine sand *Abra alba* community of the Bay of Morlaix twenty years after the Amoco Cadiz oil spill. *Mar. Pollut. Bull.*, 1998, **38**, 669–676.
6. Ellis, D., Taxonomic sufficiency in pollution assessment. *Mar. Pollut. Bull.*, 1985, **16**, 459.
7. Dauvin, J. C., Gomez Gesteira, J. L. and Salvande Fraga, M., Taxonomic sufficiency: an overview of its use in the monitoring of sublittoral benthic communities after oil spills. *Mar. Pollut. Bull.*, 2003, **46**, 552–555.
8. Gray, J. S. *et al.*, Analysis of community attributes of the benthic macrofauna of Frierfjord/Langensundfjord and in a mesocosm experiment. *Mar. Ecol. Prog. Ser.*, 1988, **46**, 151–165.
9. Warwick, R. M., Clark, K. R. and Suharsono, A statistical analysis of coral community responses to the 1982–3 El Nino in the Thousands Islands, Indonesia. *Coral Reefs*, 1990, **8**, 171–179.
10. Pearson, T. H., The benthic ecology of Loch Linnhe and Loch Eil, a sea-loch system on the west coast of Scotland. IV. Changes in the benthic fauna attributable to organic enrichment. *J. Exp. Mar. Biol. Ecol.*, 1975, **20**, 1–41.
11. Gray, J. S., Clark, K. R., Warwick, R. M. and Hobbs, G., Detection of initial effects of pollution on marine benthos: an example from the Ekofish and Elfish oilfields, North Sea. *Mar. Ecol. Prog. Ser.*, 1990, **66**, 285–299.
12. Warwick, R. M. and Clark, K. R., A comparison of methods for analyzing changes in benthic community. *J. Mar. Biol. Assoc. UK*, 1991, **71**, 225–244.
13. Ferraro, S. P. and Cole, F. A., Taxonomic level sufficient for assessing pollution impacts on the southern California Bight macrobenthos-revisited. *Environ. Toxicol. Chem.*, 1995, **14**, 1031–1040.

ACKNOWLEDGEMENTS. I thank Prof. T. Balasubramanian, CAS in Marine Biology, Parangipettai for encouragement and the authorities of Annamalai University for facilities. I also thank the Centre for Marine Living Resources and Ecology, Ministry of Ocean Development, Government of India, Kochi for support.

Received 3 January 2006; revised accepted 30 May 2006