DISCUSSION

In this report we have concentrated on the description of variation in abundances of several coral reef organisms in the Cairns Section of the Great Barrier Reef Marine Park. Our focus was on the implications of variation for the spatial design of sampling and monitoring programmes and the inference of spatial pattern, possibly arising from such effects as area-based management strategies or human impacts on the reef environment. The data we have presented indicate that existing spatial variation is large for most organisms, and that it is unlikely that small or even moderate effects of management strategies, human use, or natural perturbations will be reliably detectable as spatial pattern without considerable expense.

Scales of Variation

The consistently great heterogeneity at small scales (among transects) suggested that the transects we used were not integrating small-scale patchiness in distributions of organisms (Downing 1979, Elliott 1977, Green 1979). The genesis of such local variation probably lies in such bio-physical features as micro-topography, local inter- and intra-specific interactions, and local hydrodynamics, but will also include counting (methodological) variation (Link *et al.* 1994). It is tempting to suggest the use of larger sampling units to attenuate these large variances, but numerous studies have shown that sampling larger units generally is not the most cost-effective sampling strategy to minimise variances (see Andrew & Mapstone 1987 for review, Downing 1979, Downing & Anderson 1985, Downing & Cyr 1985, Downing *et al.* 1987, Fowler 1987, Mapstone & Ayling 1993, Pringle 1984). If these earlier results are accepted, then it will be necessary to sample many more sampling units than is often used (including in this study) to adequately sample the range of small scale variations and reduce sample variation.

Heterogeneity was reduced greatly at larger scales, even at scales of only small multiples of the size of sampling units (sites). Indeed, the strongest signal from these results was that the scale of our closely spaced sites is perhaps the least important scale to account for when sampling many reef taxa.

Similarity in coefficients of variation for location and reef means suggests as much heterogeneity among distant locations within the same habitats at one reef as among reefs, when measured in the same habitats. Our data suggest that neither of these sources of variation should be disregarded when sampling coral reef populations on the GBR with the intention of deriving results that are not peculiar to a very specific reef or location. We cannot infer what processes might be most influential at either of these scales from these simple descriptions of patterns. The similarity in magnitudes of variation among locations and among reefs over such a diversity of circumstances (regions, shelf positions, habitats), however, suggests that the key processes driving populations at each scale are either the same and/or produce the same magnitudes of effects on abundances.

The absence of clear or consistent pattern in CV's across larger-scale systematic effects such as Habitat, Shelf Position, or Region indicates that although such very large-scale factors might affect the abundances of some taxa, they generally neither attenuate or exacerbate apparently stochastic processes within such strata. Although some consistent changes in CV_t were observed with Habitat, they were taxon specific and involved few taxa. Accordingly, there is no clear advantage (on the basis of empirical sampling characteristics alone) to favouring particular habitats, shelf positions or regions for ecological, monitoring, or management studies of most taxa where reduced stochastic variation in sample data is desirable. It is noteworthy also that there was considerable consistency among taxa in the relative magnitudes of scale-related variation, and in the characteristics of predicted sampling requirements. Hence, whilst there were some taxa that were conspicuously poor candidates for monitoring studies, there were no clear candidates that would provide sensitive measures of impacts (based on their sampling characteristics alone).

Predictions of Future Sampling Requirements

Sub-sampling Reefs

Predictions of 'optimal' allocations of effort across sub-sampling hierarchies within reefs were highly variable, both among and within taxa. Predicted sample sizes needed to detect nominated effects, or the sizes of effects expected to be seen with limited sampling, were similarly variable. Cochran (1963), Winer (1971) and Winer *et al.* (1991) noted that such predictions would be uncertain because of their derivation from ratios of estimated (rather than known) variances. None of these authors, or others we know of, however, indicated the severity of such uncertainty, although McArdle (pers. com.) speculated that the predictions would be highly variable. Mapstone (1995, 1996) and Keough & Mapstone (1995) provide examples where the variations in cost-benefit procedures and sample size predictions are evident, but in most discussions of these procedures the uncertainty of the predicted sampling strategies are ignored (Andrew & Mapstone 1987, Bross & Cowell 1987, Cohen 1988, Downing 1979, Downing *et al.* 1987, Fryer & Nicholson 1993, Gerrodette 1987, Green 1989, Kenelly & Underwood 1984, 1985, Kennelly *et al.* 1993, Millard & Lettenmaier 1986, Peterman 1990, Prihoda 1983, Sokal & Rohlf 1981, Taylor & Gerrodette 1993, Underwood 1981, Zar 1984, Zedaker *et al.* 1993). It is clear from our analyses that if such procedures are to be interpreted realistically, and the strengths and limitations of proposed sampling schemes truthfully depicted, the uncertainty associated with the predictions cannot be ignored.

Despite such uncertainty, however, some generalisations can be inferred from our data. Firstly, sites were often not included in projected sub-sampling within habitats & reefs. This result is consistent with relatively small variation among sites discussed above. Secondly, sampling multiple locations and transects was consistently important. These results indicate that strategies such as those adopted by Sale *et al.* (1984), Doherty (1987), and in the AIMS Long Term Monitoring Programme (AIMS 1992, Oliver *et al.* 1995), where several closely spaced sites were sampled with multiple transects but no 'location scale' sampling was done, are almost certainly inefficient.

To the extent that our results can be extended to general recommendations, we recommend that at least 5-6 locations within habitats or reefs be sampled in future studies for most organisms. Doing so would not only ensure adequate coverage of the sampling space with which many studies (such as those just cited) are concerned (Hurlbert 1984), but also be likely to efficiently estimate the variances within reefs and/or habitats. The remaining available effort should then be put into sampling several transects within each location. It is likely that sampling multiple sites in close proximity will be useful only for studies where the experimental units are site-scale or location-scale. In the case of location-scale effects, sub-sampling at multiple sites will be likely to provide a well behaved error term in analyses, whereas sampling only transects, even many of them, might not (see also McArdle *et al.* 1990).

It is clear also from our data that optimising sub-sampling of experimental units by cost-benefit procedures is unlikely to affect substantially the expected power of statistical tests. This is to be expected from the algebra of the non-centrality parameters for tests since the numbers of sub-samples at all levels appear in both the numerator and denominator of the noncentrality parameters from which power is calculated, but we know of no empirical investigations of this subject. Very poor representation of experimental units (*i.e.*, poor precision of estimates for each unit) through highly inadequate sub-sampling might be expected to influence power, but in our results both optimised and non-optimised results provided relatively similar results. This may reflect the fact that both sets of calculations were based on fairly similar cost-constraints (1 *vs* 1.5-2 days per reef), and relatively large total numbers of sub-sampling units (~15).

These results have important implications for the design and interpretation of future studies, especially with respect to the role and scope of pilot studies. Our data do not provide the sought-after prescription of a 'best' allocation of sampling effort across different spatial scales, or a clear and unequivocal guide to the replication needed to assess either management strategies or human impacts on the GBR environment (Mapstone *et al.* 1989). Indeed, the analyses we present demonstrate that

such messages are likely to be unavailable or flawed in ecological field studies. At best, we can provide some guidelines on the scales that are (empirically) likely to require least emphasis in future sampling programmes, and insights into the reliability of predictions of required sample sizes to detect nominated effects.

It is clear also that the hitherto recommended approach of doing small pilot studies to fine-tune sampling strategies for larger programmes should be reconsidered. We do not suggest that prior information is unnecessary for designing major sampling programmes, but we suggest that pilot estimates should be treated more cautiously than they have been previously. For such pilot estimates of variance to be sufficiently robust to provide sound predictions of future sampling (Andrew & Mapstone 1987, Elliott 1977, Fairweather 1991, Green 1979, Kennelly *et al.* 1993, Keough & Mapstone 1995, Mapstone 1988, Underwood 1981, 1991, 1993, 1996), they will almost certainly need to be larger and more specifically targeted at deriving multiple estimates of variance (rather than precise estimates of means) than is now considered appropriate (McArdle *et al.* 1990, Underwood 1991). The tendency to view the predictive results of such pilot studies as definite also is clearly misguided. We have demonstrated that predictions of 'optimum' allocations of effort, sample sizes, and statistical power are highly variable. The careful design of field studies from pilot data will require explicit consideration of that uncertainty (Keough & Mapstone 1995, Mapstone 1995, 1996).

Replication of Experimental Units

Our data illustrate that only large magnitude spatial patterns in ecological effects are likely to be detectable with good certainty by most sampling programmes, given current conventional standards for critical Type I error rates. For tests of large scale phenomena such as management actions (zoning), effects of fishing, or gross geographical effects, visual surveys of four reefs within a given category are likely to detect median effects of only about 50-100% of existing abundances at best. In several cases, particularly for small fishes counted within small transects and abundances of sessile benthos, our methods would be sensitive to only gross changes in abundances of about 75-150% of standing stocks if only four replicate reefs were sampled in each 'treatment' condition, even when the test criterion was set larger than that used by convention (α=0.1 *vs* α=0.05). Considered in terms of the expected numbers of reefs needed in a sample to detect more moderate effects, equivalent to 50% of standing abundances, our data indicate that replication will have to be great in order to ensure good confidence (Power = 0.9) of detecting systematic effects on abundances. Even with relatively liberal significance criteria (by conventional standards α=0.1 would be considered liberal), in excess of 10 reefs or locations would need to be sampled to realise 90% power in 50% or more cases. These results generally held whether sampling was only within selected habitats or over whole reefs. Although it would clearly be more efficient to sample some taxa in one habitat than in the other, there was no consistent evidence that measuring reef-wide effects would be more economic in either front reef or back reef environments for all or most groups.

A corollary of these results is that, with current approaches, looking for subtle spatial effects on reef organisms will be expensive. If relatively moderate effects are considered important, then either: i) very many reefs or locations will have to be sampled; ii) it will be necessary to reconsider our dogmatic adherence to low Type I error rates in the interests of constructing more balanced inferences about results, whatever they might be; or iii) alternative approaches to sampling and monitoring studies will be required. Such alternatives will be discussed at the end of the document.

Finally, it is clear from these data that the use of predictive power analyses must be regarded with greater caution than so far suggested in the literature (Andrew & Mapstone 1987, Bernstein & Zalinski 1983, Fairweather 1991, Peterman 1990, but see Mapstone 1995, 1996, Keough & Mapstone 1995). Single (point) estimates of sample sizes, detectable effect sizes, or statistical power (or β=1-power) should be interpreted cautiously unless accompanied by statements of confidence. The highly skewed distributions of such estimates we observed indicate that single estimates have a high likelihood of underestimating the true means (of power, effect size, or sample size calculations) and thus may result in inadequate sampling. Adjustment for such potential errors can only be made by

considering the uncertainty around estimated sample sizes etc., which means that multiple estimates of variances will be required (see also Underwood 1991, 1993, 1996). Again, this suggests the need for larger pilot studies that deliver better, multiple estimates of variance.

The implications of these conclusions are two fold. Firstly, the conventional approaches to sampling or funding strategies may need re-thinking, particularly where strong inferences will be made from either 'positive' or 'negative' results (Hayes 1987, Millard 1987). It may be better in future studies to do (and fund) large 'pilot' studies to gain sound impressions of the merits of proceeding with subsequent studies, given that those subsequent studies are likely to be constrained by *reduced* funding. If the substantive pilot studies indicate that the proposed future project is weak, then funding should be refused or the approach modified. Secondly, it is likely to be inefficient to adopt an approach for assessing management strategies in which the effects of management are compared only periodically, and where inferences of success or failure rely on the detection of spatial pattern, unless dramatic effects of management are expected (Alcala 1988, Russ 1989, 1984a). In so far as there has been a 'strategy' for assessing the effectiveness of the management of the Great Barrier Reef Marine Park to date, it seems to be based on just such an approach (Mapstone *et al*, 1990). This is likely to be uninformative in the GBR region because the spatial effects of different zoning strategies seem likely to be slight relative to background variation (*e.g.*, see Ayling *et al.* 1991, Ayling & Ayling 1991, 1992a).

Sampling to Represent Reef Status and Large Scale Pattern

The existence of strong interactions between effects of shelf position and/or habitat and/or region (see also Mapstone *et al.* 1995) emphasise the need to sample comprehensively around reefs and across gross geographic clines when an objective of sampling is to monitor the status of the GBR or sections of it, or to examine the effects of any one of these factors on abundances. Further, it was clear from our data that several of the habitat, shelf position, or regional patterns evident in data from entire reefs were not consistent across locations within reefs. It apparently has been assumed in a number of past studies that standardising the location of restricted sampling within reefs provided security for the inference of among reef patterns (AIMS 1992, Dinesen 1983, Done 1982, Doherty 1987, Mapstone 1988, Sale *et al.* 1986, Williams 1982). For such an argument to provide a legitimate basis for inference of cross-shelf, habitat, regional, or (probably) temporal patterns among reefs, the effects of each of these factors would have to be consistent across each of the others, and among reefs. This is clearly not so, at least in the Cairns section of the GBR Marine Park.

Oliver *et al.* (1995) clearly identify this limitation in the AIMS Long Term Monitoring Programme, in which only a restricted (standardised) location is sampled on each reef. Throughout their text, however, they refer to the data by reefs ("for brevity") and the conclusions they reached after the first year of monitoring refer mainly to cross-shelf and regional patterns in abundances. Given the data we have presented, some caveats should be considered when interpreting the results of such studies. Most importantly, it should be specified exactly what the within-reef sampling space was and conclusions about larger scale pattern should be restricted to those within-reef strata (at the expense of brevity, if necessary). For the future monitoring of reef organisms, therefore, we recommend stratification across both habitat and shelf position to depict accurately effects of either factor on abundances of most organisms.

Future Directions

Our results indicate that the high levels of existing (natural?) spatial heterogeneity in the GBR system mean that even large differences in abundances associated with human impacts cannot be taken to signal unequivocally important environmental impacts. This does not mean, however, that human impacts of smaller magnitude than natural variation are unimportant, though they may be difficult to detect (Bence *et al.* 1996, Kingsford & Gray 1996, Nisbet *et al.* 1996, Raimondi & Reed 1996, Stewart-Oaten 1996). Whilst the importance of localised, low frequency impacts might be assessed sensibly in relation to natural spatial variability, chronic or large scale impacts of relatively small

magnitudes might be considerably more important than natural disturbances of large magnitude but low frequency. Our data suggest, however, that unless human activities, including management strategies, generate very strong signals, they are unlikely to be recognisable in space over the great existing spatial variations in abundances.

For assessments of human impacts and management strategies to be fruitful, alternative approaches to monitoring will be necessary, which allow separation of the spatial variations inherent in the system from temporal changes in abundances that might arise from management strategies or low-level human impacts. It might be expected, therefore, that frequent (annual or sub-annual) repeated measures of different management units will provide more sensitive tests of the effects of management regimes on reef-associated organisms (Green 1989, Keough & Mapstone 1995). Such an approach seems likely to be more productive than occasional spatial comparisons because of the long-lived and sessile or relatively sedentary characteristics of many reef organisms, and recent developments in analyses of such temporal data. A key assumption of such an approach, however, is that sampling and observation biases are relatively stable through time (Thompson & Mapstone in press, Mapstone, Neale & Christie in prep). A repeated measures approach to monitoring is being taken in the AIMS LTM Programme, although with severe restrictions on the spatial coverage of sampling within reefs (Oliver *et al.* 1995), and has been recommended for impact assessment studies for some time (Green 1989, Keough & Mapstone 1995, Mapstone 1990, Mapstone *et al.* 1989, 1992).

We have not considered here the potential for sequential data from the same units (*e.g.*, reefs) analysed as repeated measures to detect temporal shifts in abundances and thus test the effects of management strategies or human impacts more sensitively than simple spatial analyses. Smaller scale empirical studies (Kaly *et al.* 1993a,b, Mapstone 1990, Mapstone *et al.* 1989, 1992), and theoretical work (Mapstone *et al.* 1994), however, suggest that such an approach will provide far more sensitive tests of impacts and/or management. Additional data concerning temporal variation (diel, tidal, lunar, and longer term) in abundances of reef associated fish have been collected by W. Richards and Reef Biosearch (in 1988-89), Choat (1982-present), and within the AIMS LTM Project. These data would provide a reasonable basis for investigating the merits of repeated measures analyses of fish abundances at local scales, where abundances of fish would be expected to vary relative to times taken to survey sampling units and because of short-term movement. There are few data, however, that would allow for repeated measures analyses at the scale of whole reefs or substantive strata of them, which would be the scales at which most management strategies should be assessed (but see Ayling & Ayling 1992b, 1993, 1994, 1995). Again, such investigations would imply a substantial shift from historical approaches to assessing the effectiveness of management strategies & protected areas, along the lines of the work done over the last decade in impact assessment studies. In view of the result of this study, we suggest that consideration of such a shift is necessary for the robust and informative assessment of the effectiveness of the GBRMPA strategy of managing the GBR Marine Park by zoning.