5. ASSESSMENT OF SPATIAL AND TEMPORAL VARIABILITY

5.1 Introduction

Environmental impacts on water quality caused by the activities of man can be conveniently considered in two classes: the pollution of the water column with substances not normally found in the water column, or found in very small quantities(such as detergents, refined petroleum, chemical wastes); and the perturbation of normal levels of naturally occurring dissolved and suspended components of the water column (such as nutrients, organic solids, suspended sediments). In the former case, the assessment of environmental impact takes the form of a determining whether those substances have reached some critical concentration. In the latter category, however, assertion of an impact is more difficult because it rests on a probabilistic assessment of whether levels of naturally occurring substances have risen to levels beyond those within the natural, but often very variable, range. In both instances, the decision of whether a perceived impact is cause for concern - for example, with respect to its effect on biological systems - is made only after the detection of an effect. In this pilot study, we are not concerned with either the detection of non-natural pollutants or the rules for deciding whether an impact is cause for management action. We deal here with the optimisation of the procedures for detecting perturbations to normal levels of naturally occurring components of the water column and the assigning of such perturbations to a specific source, vis the Magnetic Quay Development.

The assessment of changes in the levels of naturally occurring nutrients, solids and turbidity in the water column as a result of any development is likely to be complicated by the inherent variability over a variety of spatial and temporal scales. The correct interpretation of data collected during an impact assessment study rests on one's ability to discriminate natural variability from 'abnormal' changes likely to be caused by the development of interest (in this case, Magnetic Quay). Such distinctions can be made only with knowledge of the magnitudes of natural variability present prior to the commencement of activities likely to cause impact, and the temporal and spatial scales at which they occurred. This information should be obtained from a soundly designed baseline study, including pilot studies designed to facilitate the projection of optimal sampling designs

by which impact can be identified during construction and operational phases of a development. The design of impact assessment studies based on such pilot studies will ensure the most powerful and economic tests of the effects of development and minimise, within sensible logistic constraints, the chances of making erroneous decisions about the presence or absence of environmental impacts.

We conducted the following two pilot studies to estimate the variability in water quality at local spatial scales and short-term temporal scales in Nelly Bay. This information provided the basis for the design of an impact assessment programme that could meet the requirements of detecting as small a perturbation to water quality as logistically possible with the smallest feasible probabilities of either falsely asserting that an impact had occurred (Type I error) or failing to detect an impact that had occurred (Type II error; see Box 2, Benthic Baseline Study).

5.2 Materials and Methods

Variables Considered

The following components of water quality were assessed in the pilot studies: nitrate; nitrite; ammonia; ortho-phosphates; suspended solids; turbidity; coliform bacteria; total heterotrophic bacteria. Various physico-chemical properties of the water column (dissolved oxygen, temperature, salinity, pH) were also measured when each water sample was collected as were a range of environmental variables (wind speed and direction, wave height and direction, etc.). Correlations between significant patterns in water quality and environmental factors were considered to see if any significant changes in water quality were conspicuously related to such parameters. The sampling and analytical procedures used to quantify all variables have been described elsewhere in this report.

Field Work and Sampling Design

Spatial Variability

Spatial patterns in the above variables were measured on December 8, 1988. The pilot study was not repeated on other days because of cost constraints, and it must therefore be assumed that the results obtained on December 8

were not atypical. This assumption was to some degree verified by the results of the pilot study of temporal variability which was repeated on two days and also contained a spatial component (see below). All water samples were collected from a moored vessel, as described elsewhere. Note that dissolved oxygen, temperature, salinity, and pH could not be measured for many of these samples because of equipment failure.

Five components of spatial variability in the composition of surface waters were considered: variation between the north and south ends of Nelly Bay; variation between the shallow, inshore, reef-flat environment and the deeper, offshore, reef-slope environment (locations); variation with depth over the reef slope; variation among sites separated by approximately 75m; and variation between replicate samples taken about 5-10m apart. At each end of Nelly Bay, three haphazardly selected sites were sampled over the reef flat and reef slope. At each site, two 1l samples of water were taken from a depth of 0.2m and about 5m apart. At the reef slope sites, two samples were also taken from about 1m above the bottom, a depth of 4-5m. Water over the reef flat was too shallow to consider a depth component of variability. Note that coliform and total bacteria were not cultured from the samples taken from near the bottom during this pilot.

The order in which ends of the bay and locations (inshore/offshore) were sampled was haphazard, but for logistic reasons, the order in which sites were sampled was not randomised over ends of the bay and location. This may have resulted in some confounding of any apparent systematic spatial pattern with the time of day at which sites were sampled, although all sampling was confined to the period between 1000 and 1600 hours. The extent to which temporal variability may have determined apparent spatial patterns was qualitatively examined, however, by considering the time of day at which groups of sites that differed significantly were sampled in the light of the results of the pilot study of temporal variability.

Temporal Variability

Diel variability in water quality was assessed at two locations over two periods of 24 hours in December 1988 (9-10/12/88 and 19-20/12/88). At each location on each day, two 1l samples of water were taken from within 1m of the surface every three hours from midday one day to midday on the following day. Locations could not be sampled simultaneously but were sampled within the same hour. Replicate samples were taken 15 minutes

apart and within 5m of each other. Thus, this pilot study assessed variation between days, variation between locations (one inshore and one offshore), among times of the day, and between replicates. Variation between replicates necessarily contained components of small scale spatial and short term temporal variability.

Samples from the reef flat location could not be taken during night low tides because of navigation hazards and absence of flowing water. Consequently, the analyses were unbalanced. To compensate for these missing data and balance analyses, data from the similar time at the offshore location were deleted on each day. Deletion of data from some other cells was also occasionally necessary because data were lost through equipment failure or sample contamination.

Statistical Analyses

Data from both of the above pilot studies were analysed by multi-factorial, mixed model analyses of variance. The spatial variability study constituted separate three factor designs for the surface water samples (End of Bay x Location x Sites (EoB, L)) and the analysis of depth effects on the reef slope (EoB x Depth x Site). Ends of the bay, location, and depth were considered fixed effects and 'sites' was considered a random variable.

The study of temporal variability was also a three factor design, comprising Days (random) x Locations (fixed) x time of day (fixed). Because the time of day at which samples were taken differed slightly between locations, and the relation of time of day to tidal phase and local whether conditions etc. varied between days, time of day was considered nested within days and locations for analysis.

A factor was considered a significant source of variation if the probability of that assertion being wrong was less than 5%, and was considered potentially significant for error (Type I) probabilities of 5-10%. Cochran's statistic was used prior to analyses of variance to assess whether variances were likely to be heterogeneous, and data were transformed to normalise variances where appropriate. Where necessary, a posteriori comparisons among means were made by Ryan's Test. Components of variation were calculated as the ratios of the (unbiased) estimate of variation among levels of each factor (derived from the mean-square

estimates) to the sum of all such estimates in an analysis. Such ratios are biased (but consistent within each analysis), but give approximate indications of the distribution of variation among multiple sources.

Analyses of the statistical power (= compliment of Type II error, or probability that a difference of specified magnitude would be detected if it existed) of the pilot studies followed the procedures recommended by Cohen (1977). When analyses indicated that sites did not constitute a significant source of variation, the 'sites' and 'residual' sources of variation were pooled and used as the estimate of residual variation for calculation of the power of tests of other terms in the spatial analyses. Similarly, in the analyses of temporal variability, when the days x location interaction was not significant (with $P > 0.25$) and accounted for very little of the variation (< 10%), that term was pooled with the residual and the power of tests of location effects based on the pooled residual variances and degrees of freedom.

5.3 Results

Spatial Variability in Surface Waters

The surface waters of Nelly Bay were spatially relatively homogeneous and apparently well mixed with respect to nutrients (nitrate, nitrite, ammonia, ortho- phosphate). There were no significant differences among ends of the bay, locations, or sites for any of these variables ($P > 0.25$ in all cases) and almost all variability was among replicates.

Turbidity varied significantly among sites within locations and ends of the bay ($F = 5.07$, 8,12 df, $P = 0.006$), but did not vary systematically in any respect. The only factors to account for any variation were the random variables sites and replicates.

For suspended solids, the interaction between location and end of the bay was significant ($F = 8.66$, 1,8 df, $P = 0.019$). The interaction occurred because at the north end of Nelly Bay the concentration of suspended solids inshore (4.2 mg/l) was greater than offshore (2.6 mg/l), whereas at the southern end of the bay the concentrations of suspended solids did not differ significantly between locations (inshore, 1.2 mg/l; offshore, 2.0 mg/l). Inshore, the north end of the bay had more suspended solids than the south, but offshore the ends of the bay did not differ significantly.

The interaction between ends of the bay and location was also significant for counts of total bacteria on plates from the collected water samples (F = 104.7, 1,8 df, P < 0.0001). The pattern of variation was the same as for suspended solids - that is, the samples from the inshore location at the north end of the bay contained more bacteria (2051.7 colonies per plate culture) than all other locations, which did not differ significantly (offshore-north, 226.7; offshore-south, 246.7; inshore-south, 448.3).

Coliform bacteria were found only in samples from the north end of the bay, a pattern that was also statistically significant (F = 15.28, 1,8 df, P = 0.005). The difference between ends of the bay accounted for approximately as much variation (47%) as all other sources combined.

Effects of Depth

Depth was not a significant source of variation in any of the water quality variables measured (P > 0.25 in all cases). The only significant terms in any analysis were the effects of end of the bay for suspended solids (F = 12.99, 1,4 df, P = 0.02), and the effect of random sites for turbidity (F = 15.45, 4,12 df, P = 0.0001). The difference between ends of the bay (North > South) when averaged over depth is suggestive that the slightly greater (though not significantly so) concentration of suspended solids in surface waters at the north end of the bay (see above) was reinforced by a similar difference at depth. As before, the major source of variation in all analyses was variation among replicate samples.

Temporal Variability

As with spatial patterns in variability, in most analyses of temporal variability the majority of variation occurred among replicate samples taken in close proximity. There were no significant effects of day or time of day on the concentrations of nitrite, ammonia, ortho-phosphate, suspended solids, or coliform bacteria (P > 0.1 in all cases). None of the variables measured differed significantly with location on either day.

Both turbidity and total counts of bacteria differed significantly between days (day 1 < day 2 in both cases) and among times of day within days and

locations. Diel variations in turbidity were not consistently related to tidal phase or wind or sea conditions, but was significantly negatively related to salinity (day 1, $r = -0.376$, 20df, $P < 0.1$; day 2, $r = -0.617$, 28df, $P < 0.001$). Although bacterial content of the water differed among times only on day 2, trends in abundance were similar at both locations on both days: bacteria tended to be more abundant nocturnally than diurnally. There were no conspicuous correlates of bacterial abundance.

Nitrate also varied in concentration with time of day ($F = 3.37$, 20,24 df, $P = 0.003$), but differences among times were significant only on the second day. There was no consistent correspondence between nitrate concentration and tidal phase or day-night cycle, or physico-chemical properties of the water on either day. The interaction between day and location was also significant for nitrate concentration ($F = 5.02$, 1,20 df, $P = 0.035$), but the interaction reflected only differences between days at the inshore location (day 1, 3.25 ugN/l < day2, 5.83 ugN/l). Concentrations did not differ significantly between days at the offshore location (4.9 ugN/l, 4.0 ugN/l) and locations did not differ significantly on either day.

Power of Tests in Pilot Studies

In almost all analyses, there was very low power (Power < 0.5) to detect relatively small (<= 25% of existing average levels) spatial or temporal differences in the measured variables. For turbidity, and concentrations of ortho- phosphates, nitrite and suspended solids, however, the analyses had great power to detect moderate differences (> half of average levels) between locations, ends of the bay, and days (Power > 0.9 in all cases). With the exception of turbidity (for which the a posteriori calculation of power was not appropriate - see below), the same was true for detecting differences among sites and times of day. Thus we are reasonably confident that the apparent absence of moderate differences between locations, ends of the bay, sites, times or days were not simply the result of high rates of Type II error.

The power to detect even moderate spatial or temporal differences in the concentration of nitrates, ammonia, and coliform bacteria was poor (Power < 0.4). Thus, even had large (> the average existing levels) differences in these variables occurred, we would have been unlikely to detect them with the above sampling programmes, even when using pooled estimates of residual variation. Note that it was inappropriate to calculate a posteriori the

power of tests for which F-ratios were significant, since in those cases the only error that could have been made was in asserting a difference that had occurred by chance alone.

5.4 Discussion

With respect to most of the variables measured in these pilot studies, Nelly Bay seemed a relatively homogeneous environment. The major spatial patterns in water quality indicated that for some variables (coliform and total bacteria and suspended solids), the north end of the bay was subject to slightly different conditions of water quality than the southern end.

The possibility exists, however, that these results were attributable to temporal confounding. The inshore-north location was the first sampled, though this did not correspond to any particular environmental conditions except tidal phase: these samples were collected during flood tide, whilst all others were collected between high and low tide. There was no significant diel cycle in concentration of suspended solids, and diel patterns in the abundances of bacteria did not correspond to tidal phase or indicate differences in abundance between mornings and afternoons. It thus seems unlikely that the above spatial patterns can be attributed to specific temporal or tidal characteristics.

Suggested Impact Assessment Programme

The design of an impact assessment programme where the variables of interests are potentially both spatially and temporally labile even at small scales presents several problems. Both spatial and temporal scales must be taken into account when designing the sampling protocol if it is expected (on the basis of prior information) that both constitute important sources of variation. Balanced against this, the sampling design must be affordable, but powerful enough to detect any important environmental impact that may occur (see also discussion in report on benthic biota).

Repeating spatially comprehensive sampling several times within a short interval (e.g. over several days within one or two weeks) is likely to be extremely costly. Further, with random components of both spatial and temporal sources of variation (such as sites and days respectively) included in a single analysis, tests of the effects of a development are

often low in power. Unless sites and/or days do not constitute significant effects, and can be pooled legitimately with residual variation, the power of the tests can only be improved by either sampling on many days and at several sites,

In the case of the Magnetic Quay development, we have demonstrated that for most components of water quality sites and days do not constitute important sources of variation. Cost benefit analyses of the data from the study of spatial variability indicated, further, that the most efficient allocation of sampling effort was to dispense with sampling sites and concentrate on replicates. This strategy would be satisfactory provided that the replicates were well dispersed within locations and thus effectively integrated variation at the scales of 5-10m and 50- 75m. Although some components of water quality varied with time of day, none of the patterns of variation suggested that a particular time of day or tidal phase should be favoured when sampling, given that sampling will be logistically constrained to daylight hours.

The suggested programme for the estimation of environmental impacts during the construction phase of the Magnetic Quay development is necessarily a compromise between logistics and the need to cater to both small scale spatial and short term temporal variability. We suggest that within any day of sampling, samples be collected at four stations near to the development (and expected to suffer any effects of construction), and at four stations sufficiently removed from the development to be insulated from any perturbations caused by the development. Here, 'station' is used to describe a tract of fringing reef stretching from the coast to the sandy bottom beyond the reef slope, consistent with its usage in the description of studies of benthic organisms in this report. At each station, three replicate samples should be taken in inshore waters over the reef flat and three from offshore waters over the reef slope. There is no indication from the pilot studies that depth is likely to be an important source of variability, but this may change in the event of an impact and so samples should be taken from near the substratum as well as near the surface where possible. Replicate samples should be well dispersed over an area of approximately 100m (longshore) x 75m (perpendicular to the shore) in each location.

It is highly desirable that the stations at which water quality is assessed correspond to those at which benthic biota are sampled, so that any perturbation to nutrient levels etc. in the water column can be related to the condition of the benthic organisms at that location. We therefore suggest that the four impact stations correspond to the Nelly Bay stations 1, 2, and 5 and Geoffrey Bay station 4 described in the report of the baseline study of benthic organisms (Figure 2 that document). We suggest that Florence Bay, Arthur Bay, Geoffrey Bay station 1, and Picnic Bay station 2 be used as control stations. It may also be considered important to sample at specific other locations, such as in Gustav Creek and off Bright Point.

Estimates of within cell and among sites variation obtained from the pilot studies were pooled and used to estimate the power and sample size characteristics of the above suggested impact assessment programme, based on the power/sample size tables in Cohen (1977). An arbitrary effect size of 50% of existing levels of components of water quality was used in these calculations. The results of these calculations indicated that the above sampling programme should prove a powerful method of detecting moderate perturbations to water quality on any given day (Power > 0.8 for most of the above variables, Type I error = 0.1; worst cases: coliform bacteria, Power = 0.15; ammonia, Power = 0.34). Detection of much smaller effects (say 25% of means) with the same power is unlikely to be viable for most variables.

The steps involved in deciding whether an impact has occurred during construction of Magnetic Quay are discussed in the report of the study of benthic organisms (Box 8). Consistent with that protocol, we suggest that if a development-related perturbation to water quality is detected on a given day, the above sampling programme be repeated on one or more days shortly after the impact was detected to assess whether it persisted. In this way, the potential for erroneous management action to arise from what was really a chance event resulting from daily fluctuations in water quality will be minimised and the prohibitive expense of routinely sampling every few days averted.