Towards an ecological framework for investigating pollution

A. J. Underwood¹, C. H. Peterson²

¹ Institute of Marine Ecology, Zoology Building, University of Sydney, NSW 2006, Australia
² Institute of Marine Sciences, University of North Carolina at Chapel Hill, Morehead City, North Carolina 28557, USA

ABSTRACT: Three aspects of the study of effects of pollution in marine systems are discussed. First is the evaluation of relative sensitivities and reliabilities of different methods of detecting pollution, including a brief contrast of processes operating in mesocosms and in the field. Second is the problem of interpretation of pollution, i.e. determining the importance of the observed effects of pollution to the biological system. Species selected for detecting pollution may not provide useful information about the economic effects on exploited parts of natural systems, nor about trophic structure of a community, nor about future sizes of populations of important species. The choice of appropriate species as indicators or detectors of pollution also requires determination of how representative they are of other species likely to be affected by pollution. Finally, there is the problem of prediction of future consequences of pollution. Some methods used to detect pollutants might be useful as early warnings of future deleterious effects, although the usefulness of these measures may be lessened by the decoupling of reproductive rates of many marine invertebrates from the eventual recruitment to adult populations. Other measures such as patterns in whole assemblages of species can usually only detect pollutants after sufficient time has elapsed for populations to have changed. Nevertheless, these offer more direct measurements of the importance of pollution to the continued functional well-being of the system. A mixture of different types of measures allows the best synthesis of predictive power while providing the most useful information for interpretation of the consequences of pollution to a marine system.

INTRODUCTION: THE RATIONALE FOR STUDIES OF POLLUTION IN THE SEA

The need to identify sensitive, robust and repeatable methods for detecting marine pollutants is widespread. Their practical use in field situations is usually stimulated by evidence that some putative pollutant has entered, or is entering, a biological system. The evidence is usually chemical, but it might be casual observation or public opinion. The first step in a study of pollution is therefore to test, formally, the null hypothesis that there is no pattern of difference among samples (using any appropriately chosen biological variable) that is correlated with the known amounts of putative pollutants. This requires demonstration of 2 different things. First, there must be differences among samples and, second, the patterns of difference must be correlated with concentrations of pollutants (and not some other environmental variables). The methods investigated in the GEEP Workshop were all designed to test this null hypothesis, with varying degrees of sophistication, precision, power and cost.

Formally, the correlation established in rejecting this null hypothesis should be demonstrably causal. Where possible, this will best be achieved by some comparison of experimentally polluted sites with control, unpolluted samples. Ideally, this would be done in a properly controlled and replicated field experiment to remove interference from naturally confounding variables (or in suitable laboratory conditions such as the Solbergstrand mesocosms; see Bakke et al. 1988, Warwick et al. 1988, and other papers in this MEPS SPECIAL). Alternatively, prior knowledge of well-established causal links between pollutants and their known effects would suffice to demonstrate that pollutants caused the observed differences. For most of this workshop, the causal relationship was considered as though it had been demonstrated.

One synthetic aim of the GEEP Workshop was to evaluate various techniques to determine which are the
most reliable, sensitive and cost-effective. This paper is primarily concerned with the subsequent procedures – those used after the initial null hypotheses have been rejected, and there is no doubt that pollution is causing some measured biological response.

Four different, but related, ecological research programmes become appropriate, although not all are directly relevant to the aims of the workshop, and not all will be discussed here. A suggested framework for investigations of existing pollution is illustrated in Fig. 1. One important area of research (EXPLANATION) is to determine why a particular method for detecting pollution works. What processes at the level of a community, population, individual organism, organ system, tissue, etc., actually cause the measured response? What is the action of polluting chemicals that leads to a detectable change in the biology of affected organisms? This area of research will not be directly discussed here. It is, however, important because, without a mechanistic understanding of the correlations that underlie the test for pollution, the pattern of response detected may not be caused by pollution, and any subsequent predictions or management based on it may fail.

The second area of research (EVALUATION) is one of the main purposes of the workshop – to determine which methods might be considered best for various purposes. The primary issue is to assess which techniques actually detect the pollution and at what level of sensitivity. If detection of pollution were the only goal, further evaluation would involve little more than a cost-benefit analysis of all the methods that actually succeeded in detecting some measurable biological response (i.e. that rejected the first null hypothesis). Evaluation of a technique should, however, also depend on the remaining 2 areas – INTERPRETATION and PREDICTION. (as in Fig. 1). One method may be preferred because it is also capable of detecting a potentially deleterious change at an earlier stage than other methods, i.e. the preferred method has greater sensitivity. Alternatively, if 2 different methods of detecting pollution are similar in sensitivity one will be preferable to the other if it also allows better PREDICTIONS to be made of future changes in numbers, sizes, types of organisms, etc., which will occur later because of pollution.

Yet again, one method might be deemed to be superior to other methods if it also enabled us to INTERPRET the potential consequences of the pollution to the biological system (i.e. whether the measured differences between polluted and unpolluted samples indicate the existence of any important deleterious change in the polluted system). If the identified levels of a contaminant have no effect on important processes operating within the ecosystem, it does not benefit anyone to have identified the existence of the supposed problem. As an example of this, there are several natural cases of chemical influences in marine and freshwater ecosystems that are not man-made and may not be deleterious e.g. natural seepages of oil (Milgram 1974, Cowell 1976, Clark 1986) and uranium in freshwater habitats in Australia (Conway et al. 1974, Jeffrey & Simpson 1986). Yet, these same chemicals are identified as pollutants in other biological systems. This has been debated in attempts to define pollution as something other than the existence of a potentially damaging chemical (see the contrasting views on 'pollution' and 'contamination' in Nelson-Smith 1972 and Clark 1966). Detection of contamination is just the first step. Identification of the possible consequences of pollutants is also an important research goal for applied ecologists.

In this paper we address some of the problems of evaluation, interpretation and prediction based on
methods used to detect the presence of pollution in marine ecosystems. Our aim is to discuss some areas of concern from the point of view of experimental population biologists. We have chosen the focus of populations because they provide a potential link between the various methods using individual organisms, or sub-organismic techniques of biochemistry, histology, cell pathology, etc., and those broad-spectrum indices of detection that involve community variables.

**EVALUATION: COMPARISONS OF TECHNIQUES**

Experimental mesocosms are a relatively recent initiative that will allow increasingly sophisticated tests of whether pollutants cause the observed differences in biological variables between contaminated and uncontaminated sites. In a properly designed experiment using mesocosms, the application of pollutants can be controlled so that causation can be identified without confounding due to other sources of variation between polluted and control samples (see the discussion of this in Warwick et al. 1988).

As Connell (1974) suggested, the aim of a laboratory experiment is to hold constant as many confounding variables as possible and then to vary those of interest. The field experiment, in contrast, permits uncontrolled natural variations in all variables except the one that is being manipulated. The very nature of a laboratory experiment, such as that using the mesocosms for this workshop, is a compromise between gain in precision of experimental treatments and loss of potentially important variables that affect the organisms (see Mertz & McCauley 1982). For studies of sustained or chronic pollution, the rationale of comparisons from one site to another is that of the field experiment. Pollutants presumably enter a biological system in consistent combinations and concentrations, which form spatial patterns, and the effects of these can be perceived despite background fluctuations in every other environmental variable.

Because the mesocosms are a laboratory experiment (as defined by the above dichotomy) it is important to evaluate how closely the results might be applicable to the usual sampling programme and problems in the field. Without suitable contrast of these 2 approaches in such extrapolation, the evaluation of techniques to detect the biological effects of pollutants is likely to be, at best, difficult and, at worst, erroneous.

The first evaluation is to determine how well controlled were the background variables (such as waterfall, sedimentation, etc.) that potentially confound comparisons of different experimental treatments. In the existing experimental design, 4 basins were used, each of which contained one single experimental treatment (control or one of three concentrations of pollutants), Bakke et al. (1988). A better experimental design, though one logistically impossible for the workshop, would be to use replicate basins of each treatment, on the premise that confounding sources of variation will be randomly allocated over each set of replicates. Thus, every treatment will be sampled with the same average value of each background intrinsic variable (Hurlbert 1984). The only experimental designs that enable proper evaluation of the effects of experimental treatments are those that allow statistical comparisons among experimental treatments with the variation within each treatment at all nested levels of replication necessary to complete the experiment (see also Underwood 1981).

The second point of evaluation is the likely similarity between the results from experimental mesocosms and field samples. There are many potential differences between experimental mesocosms and natural environments, some of which may cause stresses on the organisms (e.g. different rates of flow, lack of a water-column over the sediments, etc.). If the organisms in communities in the mesocosms or, for that matter, in field situations, are already stressed by natural processes, the extra imposition of a pollutant may have larger effects than would otherwise occur. Such synergisms under natural conditions have been described in experimental analyses of rocky intertidal and soft-bottom subtidal systems (Peterson & Black 1988). Detailed analysis has previously demonstrated that maintenance of the mesocosms in the laboratory causes some physiological stresses to the communities (Berge et al. 1987). Treatments with pollutants may therefore cause more exaggerated effects than would occur under field conditions (where stresses are generally less). Alternatively, if the animals or communities suffer very great levels of stress just from being kept in the laboratory, it may be impossible to detect differences between controls and treated groups, because the former would show depressed values of any chosen variable, and the latter are not likely to be very much more affected.

Assemblages of organisms in mesocosms differ from those in the field because of differences in rates of reproduction and recruitment. Previous experimental work has shown that some species and types of organisms do not recruit into the mesocosms in the same numbers as in control boxes of sediment at natural depths (Berge et al. 1987). Comparisons of some measures of community structure might therefore fail to detect any effects of pollution in mesocosms, but would not fail when used in the field. This will occur if the mesocosms fail to recruit those widespread opportunistic species that respond to disturbances by rapid invasion and multiplication in disturbed areas (Grassle
& Grassle 1974, Pearson & Rosenberg 1978). Such species are apparently lacking from the mesocosms. Thus potential for differences between polluted and unpolluted sites is not mimicked in the mesocosms (see also the discussion of this point in Warwick et al. 1988).

As a final consideration, some species may potentially have been reduced in numbers, or disappeared altogether from the mesocosms during the course of the experiment, as a result of the processes of emigration, mortality and reduced recruitment, regardless of the effects of pollutants. Such changes in the composition of the communities could have been evaluated by comparisons of the control boxes at the end of the experiment with similar boxes sampled at the very beginning. Any major differences in composition of the communities would almost certainly reflect deletions from the fauna (Berge et al. 1987), notably of those individuals and species most sensitive to stressful conditions. This would inevitably make it more difficult to detect the effects of pollutants than should normally be the case in the field sampling.

Good accounts of some of the problems and advantages of these experimental mesocosms can be found in Warwick et al. (1986) and Berge et al. (1987). Clearly, a detailed comparison of results from laboratory mesocosms and events in the field must await further considerations of these points. Consequently, the results from experimental mesocosms have been interpreted cautiously in these proceedings.

**INTERPRETATION: CONSEQUENCES OF POLLUTION TO POPULATIONS**

Once the null hypothesis that there is no biological effect of a putative pollutant has been rejected, we are then faced with determining whether the observed level of pollution matters. Are there important deleterious consequences to the biological system that can be determined from the observed changes in the organisms and systems? One way of providing further advice to what will be a political decision about the tolerable levels of pollution is to relate the observed biological effects to consequences for the populations in the system. Community-level effects have certain implications for the populations that make up the community, while physiological, cellular and subcellular effects may also relate to the populations of the organisms in which the effects were identified. Relating observed effects of pollution at a population level provides a potential for comparing effects and implications in a common currency despite the widely disparate levels at which the effects of pollution were actually detected.

Furthermore, political decisions setting the acceptable levels of pollution will be commonly based on the degree to which the pollution affects economically important populations, such as those supporting commercial and recreational fisheries, although some decisions on pollution will be based on risks to human health, or in some instances on maintaining species diversity. Nevertheless, the issue of how a given effect of pollution relates to production of economically important marine populations represents an ecological question central to the issue of whether pollution at a given measured level actually matters.

Specific community measures have implications of change for various processes in populations (Fig. 2). For example, changes in species diversity (however defined) imply changes in abundance of at least one population in that community. Similarly, other measures of structure of communities imply that change has occurred in at least one population-level parameter. Most community measures relate directly to abundances of the populations, and somewhat less directly to contributions to food-webs (Fig. 2). If levels of pollution of the marine benthos are to be evaluated by the potential impact upon production of organisms higher in the food-chain, then measures on the whole community have direct relevance to the issue of whether a given level of pollution matters. This is one justification for inclusion of community-level studies in assessment of marine benthic pollution. However, a paucity of knowledge of the contributions to food-webs by various benthic organisms and a lack of information about the degree of interchangeability of various prey species in the diets of important consumers weaken the argument that measures on the whole community relate directly to the value of any marine system.

Furthermore, if a subset of data about communities could suffice to present most of the information contained in the whole set of data, the less expensive and time-consuming assessment of that subset may represent a preferable method for analysing the importance of pollution in that system. The subset that best detects the presence of pollution may, however, not also be the subset that best serves to indicate the impact on higher trophic levels. Determining which benthic species are important in those food-webs leading to economically valuable fishes and invertebrates requires more research to improve the assessment of the consequences of pollution.

Measures of pollution at the physiological, cellular and subcellular levels also have implications for various population processes (Fig. 2). For example, measures of 'scope for growth' relate directly to individual rates of growth and body-size, which in turn imply changes to reproductive output because this varies with body-size in many marine invertebrates (Spight & Emlen 1976, Hughes & Roberts 1980, Peterson 1983). It is not clear, however, that reduced growth
and reduced reproductive output necessarily have any impact on abundances of populations nor on contributions to food-webs of those populations. The high rate of mortality of larvae of benthic invertebrates (Thorson 1950, Mileikovsky 1971), combined with density-dependent regulation of numbers of recruits (Connell 1985), or density-independent decoupling of numbers settling from the plankton and numbers entering a population (Caffey 1985), may render negligible the impact of reduced reproductive output on abundance of a population. In addition, many invertebrates recruit from areas far removed from the adult population, and only widespread episodes of pollution will cause major reductions in reproductive output over a geographical scale sufficient to affect local populations. As a result of the large-scale dispersal of many marine species, the continued chronic pollution of any relatively small geographical locality will have negligible consequences for most local populations, however long the pollutants continue to enter the system (provided only that pollutants do not kill the newly-recruited larvae as they arrive in the populations). In addition, rates of predation on benthic invertebrates are often greater on smaller individuals within a population (Connell 1970, Paine 1974) such that pollutants that act to reduce rates of growth and thereby maintain individuals at small sizes for longer periods may actually enhance the contributions to food-webs by the affected populations.

In general (Fig. 2), measures of pollution made on individuals (or at smaller scales) relate more to growth, reproductive output or body-size than to abundance and contribution to food-webs. One of the major reasons for this is that the choice of organisms for individual, or sub-organismic, methods of detecting pollution in marine systems has apparently been dictated by the requirement that the animals actually
survive throughout the period of pollution, so that they can be assayed at some subsequent time. Such organisms may therefore often be those that show the most resistance to, or are simply unaffected by, the pollution. There are exceptions, for example Wyers et al. (1986) used pilot studies to determine appropriate concentrations of oil for examining pollution on a hermatypic coral. However, the majority of the physiological and biochemical techniques for recognizing the existence of pollution are not as closely related to society's concerns as are the alternative measures based on communities.

Nevertheless, these physiological and biochemical measures have other potential advantages over measures of structure of communities. Firstly, they can be better used to assess the specific toxicity of pollutants that may pose health risks for humans. One helpful trend is the development of effects measures with some degree of specificity to particular classes of pollutants (e.g. Stegeman et al. 1988). Secondly, these sub-organismic measures may allow predictions (see later) of future responses by populations and communities. If these techniques provide reliable advance warning of problems due to pollutants, they have important roles to play in monitoring, where rapid managerial responses are required to the appearances of, or changes in concentration of, pollutants. Therefore, one basis for the choice of appropriate species for these organismic and suborganismic measures is that they should be 'ecological canaries' capable of providing an early-warning of impending deterioration within the system.

Finally, the physiological and biochemical measures are made on carefully selected species. Species could be chosen for their direct importance, either as a 'keystone' species in organizing the community, or because they form an important link in food-webs leading to economically important consumers. For example, Mytilus edulis, a species commonly used in analyses of 'scope for growth', is often a 'keystone' species (sensu Paine 1966, Dayton 1971) in rocky shore communities, where it displaces other species that use the surfaces of the rock, but provides habitat and shelter for a wide range of additional invertebrates that would otherwise be rare or absent (Suchanek 1985).

A major problem with measures of pollution at levels of organization lower than the population is that they are usually done on only a single species (Kimball & Levin 1985), although there are exceptions (Klumpp & Burdon-Jones 1982). Each species is not necessarily chosen to be in any way representative of the entire community, nor because it is ecologically significant. The strategy for selection of these 'target' species usually requires that the organism be common, large enough to handle, and tolerant of a wide range of environmental conditions, so that it will be present over a broad gradient of concentrations of putative pollutants. These criteria constrain the choice of species for tests of the existence of pollution. Consequently, the ecological implications of such non-random selection of species must be considered. It would be revealing to test the relationship between 'scope for growth' or other physiological and biochemical changes in the chosen species and rates of mortality of the other species in the same community. In choosing an appropriate species to serve as a 'target' organism, consideration should be given to its ecological or economic significance, its representativeness of other species in the system, or its suitability as an indicator of the current status of ecological processes or as a predictor of the future health of the system. Otherwise, instead of choosing a 'keystone' species as an object of study, physiologists or biochemists may be shackled to a 'milestone' species, i.e. one which is chosen for study without consideration of alternatives because of the great weight and past history of invested time and effort that has gone into the development of techniques using it.

One other potentially useful guideline in choosing species that might make appropriately sensitive indicators of the effects of pollution is to consider those species that are likely to be suffering stresses under natural conditions in the ecosystem of interest and that are most likely to succumb quickly to new stresses imposed by pollutants. These should therefore provide the most sensitive organisms in the system (Peterson & Black 1988). As an example, bivalves suffering from crowding are much more susceptible to smothering by sediment than those at smaller densities (Peterson & Black 1988). Where research has indicated the relative sensitivities of different species to natural sources of stress, the probable synergisms of extra stresses allow better choices of species for detecting the presence of pollutants.

PREDICTION: INTER-RELATIONSHIPS AMONG MEASURES OF POLLUTION

A central question in pollution research is to what degree various measures used to detect pollution also predict future change in other properties of the system. As an example, biochemical changes at a sub-cellular level in a target organism presumably precede cellular and tissue alterations, which in turn should later affect physiological function and ultimately, perhaps, some population parameters. Exposing the links that tie together various potential measures of pollution at different levels of biological organization is important for at least 2 reasons.
Firstly, such linkages require a better understanding of how pollution alters biological processes. Such an understanding would allow more certainty in prediction, because mere correlations between levels of pollutants and biological indices represent a risky basis for making predictions until causation has been established. Secondly, understanding the mechanistic inter-relationships between measures at different levels of biological organization is necessary for proper tests of whether any measure in fact has predictive capacity.

The usual assumption is that pollution acts first at the biochemical level and then later is reflected in tissues, physiology and, finally, processes in populations. For example, comparisons of lysosomes in polluted and unpolluted individuals should indicate the general well-being of the cells comprising the target species. The activities of enzymes in lysosomes within the digestive cells of molluscs are assumed to reflect the capacity of the animals to digest their food efficiently. If such assumptions are correct there is reason to believe that measurements of deterioration of lysosomal function in polluted individuals will anticipate later effects in the reproductive tissues and, eventually, reproductive output of the population (Fig. 2).

If such sequences of effects at increasing scales of organization are true, it follows that measures of pollution at the smallest scale (i.e. biochemical indices) are likely to detect episodes of pollution well in advance of any of the measures at higher levels of the hierarchical scale. Although this seems to represent a realistic contention, the suggested predictive capacities of measures of pollution have rarely, if ever, been tested.

In this workshop, the comparisons of techniques at different levels of organization on the same set of concentrations of pollutants provide a start to the process of inter-calibration of techniques and to the task of testing the predictability of those measures that respond more rapidly. Even if predictive links can be established between measures of pollution, we suggest (Fig. 2) that measures at sub-organismic levels will lead to predictions concerning growth, body-size and reproductive output in populations, but may not serve to predict future abundance in many populations of marine invertebrates because of the substantial degree of decoupling between reproductive output and subsequent recruitment (Clark 1986).

Consequently, measures made on sub-organismal processes will be unlikely to relate to changes in ecological value of the benthic system. Only where there is known to be some relation between reproductive output and subsequent numbers in a population is it likely that predictors of the former will provide any advance information about the latter. In contrast, virtually all of the measures of pollution using whole communities relate directly to abundances of populations and to benthic contributions to marine food-webs (Fig. 2). There are, of course, measures of pollution made on samples of individual whole organisms (rather than on tissues or cells on the one hand or communities on the other). Examples are the behavioural assays described by Dicks (1976) and Pearson et al. (1981) which demonstrate various consequences to populations and foodwebs of alterations in behaviour caused by oil pollution. Such behavioural assays were not considered in the workshop, but are included for completeness in Fig. 2.

We conclude that utilizing measures of pollution at different levels of biological organization represents a sensible strategy for environmental scientists because these measures serve different purposes. The organismic and sub-organismic measures potentially provide the earliest warning of possible future deterioration and may also be the most sensitive measures of pollution. In contrast, measures on communities may provide a better indication of the consequences of that pollution to processes of economic and societal value vested in the marine ecosystem. A complete assessment of any episode of pollution must include accurate measures of the biological effects of that pollution at a number of different scales. No one measure can satisfy all the requirements of those who must make decisions about potential environmental, economic and social impacts of pollution.

Acknowledgements. The preparation of this paper was aided by funds from the Australian Research Grants Committee, Caltex (Australia) Pty. Ltd., the University of Sydney Research Grant and the Institute of Marine Ecology (University of Sydney) to AJU and by NSF Grant OCE 84–15696 to CHP. We are grateful to P. Scanes in the preparation of the paper, to Professor D.T. Anderson and Dr P.A. Underwood for advice, and to Drs R. Black, P.G. Fairweather, K.A. McGuinness, Mr N.L. Andrew and Mrs M.G. Chapman for comments on the manuscript.

LITERATURE CITED