

Observer effects and training in underwater visual surveys of reef fishes

A. A. Thompson^{1,*}, B. D. Mapstone²

¹Australian Institute of Marine Sciences, PMB No. 3, Townsville, Queensland 4810, Australia

²CRC Reef Research Centre, James Cook University, Townsville, Queensland 4811, Australia

ABSTRACT: Visual survey techniques are used widely to estimate abundances of target organisms in terrestrial and aquatic environments. There are a number of methodological 'errors' in almost all applications of visual surveys. Given the dependence of all visual survey data on the skill and technique of the observer, one potentially important source of imprecision and/or bias is variation among and within observers. In studies involving large amounts of fieldwork over great geographic range and many years, it is inevitable that observers will change from place to place and through time at any single site as they are replaced or gain experience. We present the results of 3 observer training/calibration exercises that indicate that observational studies in which multiple observers must be employed may be subject to considerable observer-related biases and imprecision. We found that careful training and calibration of observers ameliorated such effects for most taxa, but non-trivial levels of bias for some taxa and imprecision in estimates for several taxa remained even after thorough training. It is essential that the influence of observer bias and imprecision be well documented in multi-observer monitoring studies, so that (spurious) patterns related to differences among observers can be distinguished from real spatial or temporal patterns in the environment.

KEY WORDS: Observer training · Bias · Precision · Monitoring · Visual survey · Fish counts

INTRODUCTION

Estimation of abundances of organisms is central to many ecological field studies. Visual survey techniques are used widely to estimate abundances of target organisms in terrestrial (Caughley et al. 1976, Caughley 1977, Francis 1994) and aquatic (Keast & Harker 1977, Solazzi 1984) environments. In marine research and monitoring, visual survey techniques range from aerial surveys of large mammals (Estes & Gilbert 1978, Leatherwood 1979, 1982, Holt & Colongne 1987, Marsh & Sinclair 1989) to counts of fishes and benthos by SCUBA divers (Brock 1954, Bouchon-Navaro 1980, Williams 1982, Williams & Hatcher 1983, Eckert 1984, Sale et al. 1984, Harmelin-Vivien et al. 1985, Ayling & Ayling 1986, 1992, Doherty 1987, McCormick & Choat 1987, Mapstone & Ayling 1993), and even *in situ* counts of microscopic sessile organisms with underwater microscopes (Kennelly & Underwood 1984).

There are a number of methodological 'errors' in almost all applications of visual surveys, most of which result in underestimation of population densities (Jones & Chase 1975, Andrew & Mapstone 1987, Greene & Alevizon 1989, but see Brock 1982, Mapstone 1988). Estimation errors fall into 2 categories: those related to the visibility of organisms, which usually mean that some proportion of the population is not available to an observer; and those attributable to the observer, resulting in failure to either identify or count some organisms that are available (Caughley 1977, Fernandes 1990, Fernandes et al. 1990). Availability errors are generally considered to result in biased (under)estimates. They might also vary in magnitude from place to place and time to time, and contribute to imprecision in estimates. Observation errors, however, are often assumed to be roughly random and affect the precision of estimates but not involve bias (but see Andrew & Mapstone 1987, Marsh & Sinclair 1989).

In practice, it is often difficult or impossible to separate availability and observation errors, since both will

*E-mail: a.thompson@aims.gov.au

depend on the survey method, counting procedures, choice of sampling unit, time of sampling, etc. Their separation is especially difficult where the target organisms are highly mobile within the time taken to complete a count, such as with underwater surveys of fishes. Several authors have examined sources of methodological error in underwater visual surveys of fishes, including the counting method (Sale & Douglas 1981, DeMartini & Roberts 1982, Kimmel 1985, Sanderson & Solonsky 1986, Thresher & Gunn 1986, Greene & Alevizon 1989, Bortone et al. 1991, Samoilys 1992, Mapstone & Ayling 1993), the numbers of taxa counted simultaneously (Russell et al. 1978, Greene & Alevizon 1989, Lincoln-Smith 1989), the speed with which counts are done (Mapstone 1988, Lincoln-Smith 1989, St. John et al. 1990), and the shape and dimensions of sampling units (Sale & Sharp 1983, Fowler 1987, McCormick & Choat 1987, Mapstone 1988, Buckley & Hueckel 1989, Mapstone & Ayling 1993).

Whilst it is desirable that abundance estimates are unbiased and accurate, for the purposes of comparing between 2 or more surveys it is sufficient that biases remain constant. Spatial and temporal consistency in bias or the effects of survey method have been considered only infrequently (Bell et al. 1985, Thresher & Gunn 1986, Mapstone 1988, English & Bocking 1992, Mapstone & Ayling 1993). Given the dependence of all visual survey data on the skill and technique of the observer, one potentially important source of inconsistency in bias is variation among observers and/or within an observer as a result of changing experience and training. For example, Mapstone & Ayling (1993) and St. John et al. (1990) identified consistent biases between 2 'experienced' observers counting the same populations of reef fish.

Over recent years there has been growing demand for large-scale, long-term monitoring of environmental variables (such as abundances of organisms). In such studies involving large amounts of fieldwork over great geographic range and many years, it is inevitable that observers will change from place to place and through time at any single site. For the correct interpretation of results from such studies, it is essential that the influence of observer bias be well documented, so differences among observers can be distinguished from real spatial or temporal patterns. Observer biases should be minimised by thorough training of new observers and regular re-calibration of trained observers, but it is important to measure the progress of such training and the degree of residual observer-related error before interpreting data from multiple observers. To this end the Australian Institute of Marine Science (AIMS) has developed a 2 tiered training program to quantify and minimise observer bias in visual estimates of reef fish taken as part of their Long

Term Monitoring Program (LTMP). In this paper we present the results of 3 applications of the training procedure to address the questions: (1) Does training reduce variability in counts between observers? (2) Is training response similar across different taxa and on different occasions? (3) Is observer variability consistent across spatial scale? (4) What levels of inter-observer variation persist after training, and how will they influence the precision and/or bias of estimates? We then discuss the implications of these issues for visual survey methods in long-term and/or spatially extensive monitoring programmes where observers may change over both space and time.

MATERIALS AND METHODS

Training procedure. The first stage of the AIMS training program involves familiarisation with both the theory of the technique to be employed and the identification of the target species. Proficiency in the accurate and rapid identification of the full list of species to be surveyed is generally achieved by referencing the numerous photographic texts available [e.g. Randall et al. (1990), Allen (1991), Myers (1991)], which combined give a good coverage of coral reef fish of the Indo-Pacific] coupled with underwater tuition by a skilled observer. The fine detail of the survey technique adopted by the LTMP is described to the trainee with reference to a Standard Operational Procedure (Halford & Thompson 1994) which was developed specifically to ensure a standard and unambiguous description of the technique. We do not discuss this basic preparation here.

The second tier of the training program involves a series of comparative surveys utilising the technique as described in Halford & Thompson (1994), and summarised below. This aspect of training is applied during the induction of new observers to the monitoring programme and annually thereafter to 'calibrate' co-existing observers against each other. The 3 training events we discuss include initial training of 2 new observers (March 1993 and March 1995) and 1 annual calibration of 2 experienced observers (March 1994). All field training was done at the same 3 sites on Myrmidon Reef, on the off-shore margin of the central Great Barrier Reef.

The sampling design for all 3 consisted of repeat surveys of 5 permanently marked transects at each of 3 sites along a section of continuous reef slope in depths of 6 to 10 m. Each transect was 50 m long and marked permanently with a steel picket at each end and 3 to 4 smaller steel guide rods at 10 to 15 m intervals between them. Training or calibration involved either a trainee and trained observer or 2 experienced

observers concurrently counting 196 species from 43 genera in 10 families during 2 passes along each transect. Larger and more mobile species were counted within a 10 m wide (first training event) or 5 m wide (second and third training events) belt centred along the line defined by the permanent markers. For these counts the 2 observers swam shoulder to shoulder to ensure the fish available to each observer during the period of survey were the same. A third diver swimming behind the observers laid a tape measure between the marker stakes.

After the observers had reached the end of all 5 transects in a site, they returned along each transect counting the remaining target species (in the family Pomacentridae) in a belt 2 m wide (first training event) or 1 m wide (second and third training events), with the tape measure as its deeper boundary. These narrower transects were surveyed in 'single file', with the observers separated by approximately 15 to 20 m, as they were too narrow for observers to swim side by side and retain the same perspective as a single observer would have in routine surveys. Counts from the 2 observers were compared in the field after each site was surveyed, and conspicuous discrepancies discussed before surveying the next site.

Transect dimensions altered between the first and second training events due to changes in the dimensions of transect used in the LTMP for which the observers were training. The changes stemmed from work by Mapstone & Ayling (1993) and for the purposes of this study it is important to note that the precision of both techniques was similar. We make the assumption that characteristics of observer training would not be altered substantially by those changes in the dimensions of the sampling unit.

An assumption underlying all comparisons between concurrent or sequential counts by the 2 observers was that the same assemblage of fishes was available for counting by both observers. This assumption rested on a second assumption for sequential counts of the Pomacentridae, namely that the separation between divers was sufficient to allow the (predominantly sedentary) pomacentrids to resume normal behaviour after disturbance by the passage of the first observer before being counted by the second observer.

The first training period involved 2 surveys of the 3 sites over 4 d. Site 1 was surveyed on Days 1 and 2, Site 2 was surveyed on Days 2 and 3, and Site 3 was surveyed on Days 2 and 4. The second and third training events involved 3 surveys of each site over 3 d periods, with all sites being surveyed (in the same order) on each day. Thus, in the first training event, effects of site and 'visits' to sites were both partially confounded with days. In the subsequent events, site effects were crossed with day effects, but 'visit' and

day effects were completely confounded. Some confounding between either sites or visits and days was unavoidable because of the time required to survey all sites. Our inferences of the effects of visit (= training) on relative bias between observers were thus based on the assumptions that: (1) daily variations in the abundances and behaviours of fish at each site were relatively trivial (which was generally the case; unpubl. data), and (2) that any day effects that were present were consistent across observers. Provided the second assumption was met, changes in the difference between observers could be attributed to the effects of training.

Data analysis. Because many species were uncommon or absent on many transects, data were generally aggregated taxonomically before analysis. Thus, we analysed data at family level for the 5 most abundant families counted in the larger transects (Acanthuridae, Chaetodontidae, Labridae, Scaridae, Zanclidae), and by genus for the 5 most abundant genera of the Pomacentridae (*Acanthochromis*, *Chromis*, *Chrysiptera*, *Plectroglyphidodon*, and *Pomacentrus*) counted in the narrower transects. Such taxonomic aggregation would be more likely to mask observer effects in the data than precipitate spurious effects, but was likely to be a feature of the routine analyses of many monitoring data.

Effects of training on observer bias: Effective training of observers would be expected to result in sequential modification of observer-specific bias and/or precision, most conspicuous as monotonic sequential changes in the differences between observers across sites and visits. Such incremental improvement would be expected because sites were surveyed sequentially within visits, and counts by the 2 observers were reviewed after each site had been surveyed. We examined the improvement in observer similarity by plotting the \log_2 -ratio of counts by Observer 1 and Observer 2 sequentially over visits and sites. The \log_2 transformation was used simply as a visual aid as it had the effect of standardising the scale of the proportional data presented so that a value of +1 indicated that Observer 1 counted twice as many fish as Observer 2, and a value of -1 indicated that Observer 2 counted twice as many fish as Observer 1.

Successful training should result in improvement in agreement between observers from the initial counts to the final counts during a training session. To assess the overall effectiveness of training, intra-class correlation coefficients (ICCs) were calculated for all taxonomically grouped data for each transect dimension from the first and last surveys of each training period. The ICCs were a measure of the agreement of the data sets arising from the 2 observers, so that as the observers' counts became more similar, the ICC approached 1. Thus an increase in ICC would indicate a general improvement in agreement between ob-

servers. To further examine changes in agreement between observers, structural regression equations were fitted to the data from the 2 observers. Structural regressions assume that both measurements are subject to error and as the resulting fitted line results from minimising deviations in both the x and y dimensions. As both observers were counting the same area, the expected 'best-case' line would have an intercept of zero and a slope of 1 (45°). Successful training would be indicated if this line was better approximated by the structural regression for the post training data than that for the data at the beginning of training.

Effects of training on precision: We considered 2 other effects of observer inexperience, both of which would affect the precision of estimates of population densities. Firstly, training inexperienced observers might be expected to reduce 'instability' in counting, and hence result in greater precision in estimates arising from observational errors. To assess whether observer training affected variation among counts, we plotted the \log_2 -ratio of coefficients of variation (CV) among the 5 transects surveyed on each visit to each site by the 2 observers. If training reduced variation among counts, then the \log_2 -ratios of CVs should approach 0 with observer training.

Secondly, by reducing differences in bias between observers, training should reduce the uncertainty in estimates because of potential observer effects. That is, as observers become more similar, there should be increased confidence that estimates of a static population would be the same (or very similar) irrespective of which (experienced) observer did the counting. To estimate the residual variation arising because of observer effects, we calculated the taxon-specific CVs between observers' counts on each transect, their means for each site, and their means over all sites at the end of training. The distributions of inter-observer CVs for each scale were plotted to indicate the degrees of uncertainty that should be attached to single estimates of abundance collected by different observers because of observer effects alone.

RESULTS

Effects of training on bias

Plots of sequential changes in \log_2 -ratios of observer-specific bias fell into 3 groups (Fig. 1).

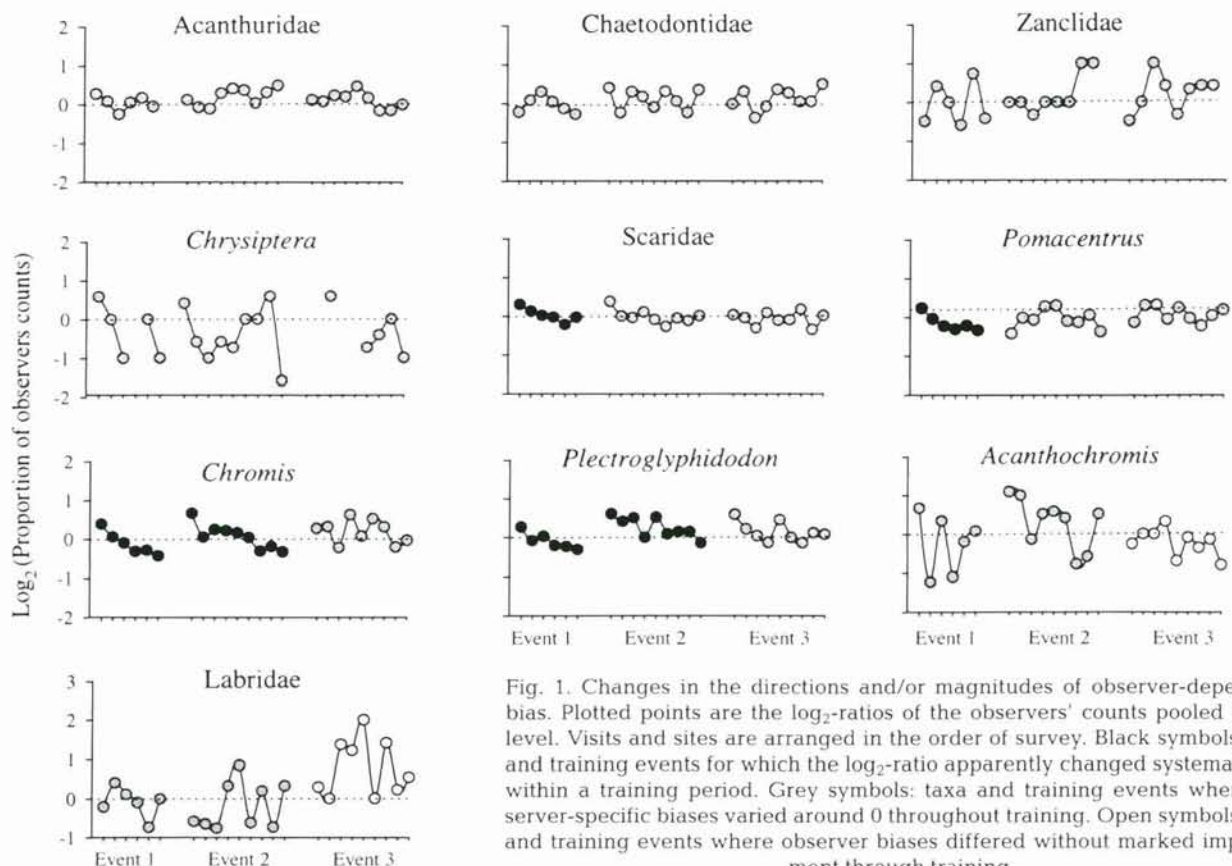


Fig. 1. Changes in the directions and/or magnitudes of observer-dependent bias. Plotted points are the \log_2 -ratios of the observers' counts pooled to site level. Visits and sites are arranged in the order of survey. Black symbols: taxa and training events for which the \log_2 -ratio apparently changed systematically within a training period. Grey symbols: taxa and training events where observer-specific biases varied around 0 throughout training. Open symbols: taxa and training events where observer biases differed without marked improvement through training

(1) Taxa for which the ratio was erratic or relatively stable around zero after the first 1 or 2 counts. This group included counts of Acanthuridae, Chaetodontidae, Zanclidae, and *Chrysiptera* on Training Events 1, 2 and 3 (E1 to E3); Scaridae and *Pomacentrus* on E2 and E3; Labridae and *Acanthochromis* on E1 and E2; and *Chromis* and *Plectroglyphidodon* on E3. For these taxa, either counting bias was not dependent on observer, or training rapidly corrected observer-dependent biases and resulted in similar biases for both observers after the first 1 or 2 counts.

(2) Taxa for which there was a consistent change in the \log_2 -ratio, usually resulting in reduced observer difference with training. These taxa included *Chromis* and *Plectroglyphidodon* on E1 and E2; and Scaridae and *Pomacentrus* on E1. These were taxa where training apparently resulted in correction, and sometimes slight over-correction, of differential bias by 1 or both observers. In all cases, the counts from the more experienced observer were relatively constant, but counts by the less experienced observer shifted from less than to equal or greater than those by the experienced observer.

(3) Taxa for which the ratio was mostly non-zero and showed no clear improvement through training (Labridae and *Acanthochromis* on E3). These taxa were those for which training was relatively ineffective in altering observer dependent bias. During the third training event, the trainee observer fairly consistently reported more *Acanthochromis* than the more experienced observer (average 2.8 vs 2.07 fish per transect), and fewer labrids than the experienced observer (average 1.33 vs 1.95 fish per transect). Each of these results suggested a failure of training to significantly attenuate or remove observer effects.

As \log_2 -ratios provide no account of the actual differences of abundance estimates between observers, Fig. 2 has been included to provide a reference to the observer-related differences in actual counts at the end of the third training period. Only 1 training period is included as absolute numbers were very similar for each taxonomic group on all training periods.

With data from all 3 training periods, both ICCs and structural regression equations showed that training (or calibration) improved agreement slightly between observers' estimates for both non-pomacentrid families and pomacentrid genera (Fig. 3, Table 1). For non-pomacentrid families the ICC improved from 0.984 (95 % CI: 0.970 to 0.992) before training to 0.992 (95 % CI: 0.985 to 0.996) after training, whilst for pomacentrid genera the ICC improved from 0.974 (0.955 to 0.986) before training to 0.991 (0.982 to 0.995) after training (Fig. 3). Structural regression analysis for both taxonomic groupings showed improvement with training in both the fit to the expected 'best case' line of $\text{Estimate}_{\text{Obs1}} = \text{Estimate}_{\text{Obs2}}$ and a reduction in dispersion about that line, though for neither the non-pomacentrids nor the pomacentrid genera did the slopes of the regressions differ significantly from 1 before or after training. Training did improve the correspondence of the intercepts with 0, however (Table 1).

Effects of training on precision

There was very little effect of training on the relative precision ($\text{CV}_{\text{Obs1}}/\text{CV}_{\text{Obs2}}$) of estimated site means. In all training periods, the \log_2 -ratio of observer-specific CV varied considerably and apparently erratically for all taxa, except perhaps *Plectroglyphidodon* spp. and

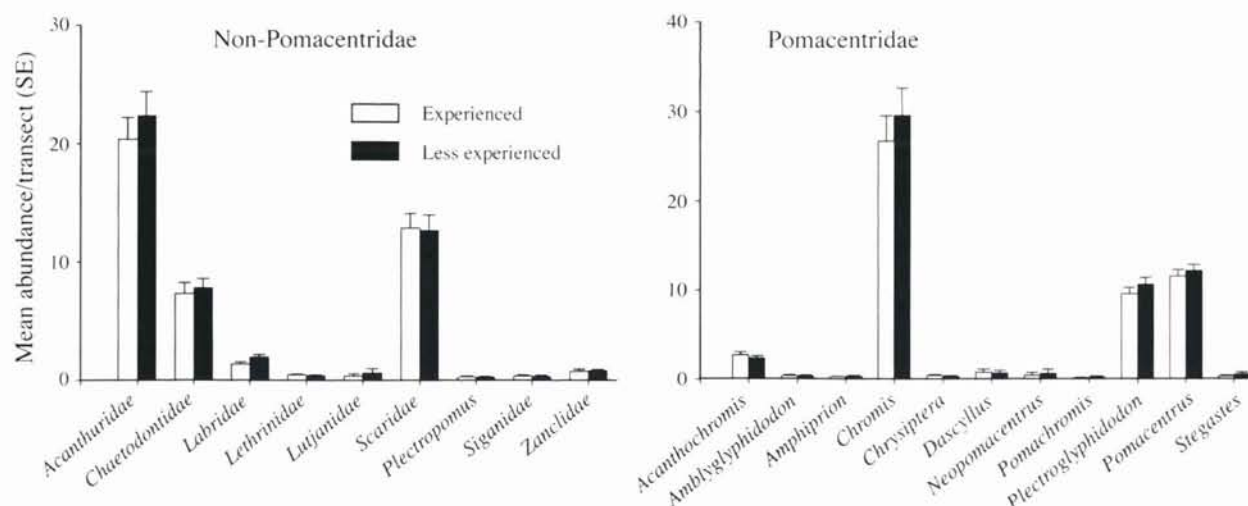


Fig. 2. Comparison of untransformed abundance estimates for each observer on the last day of the third training period

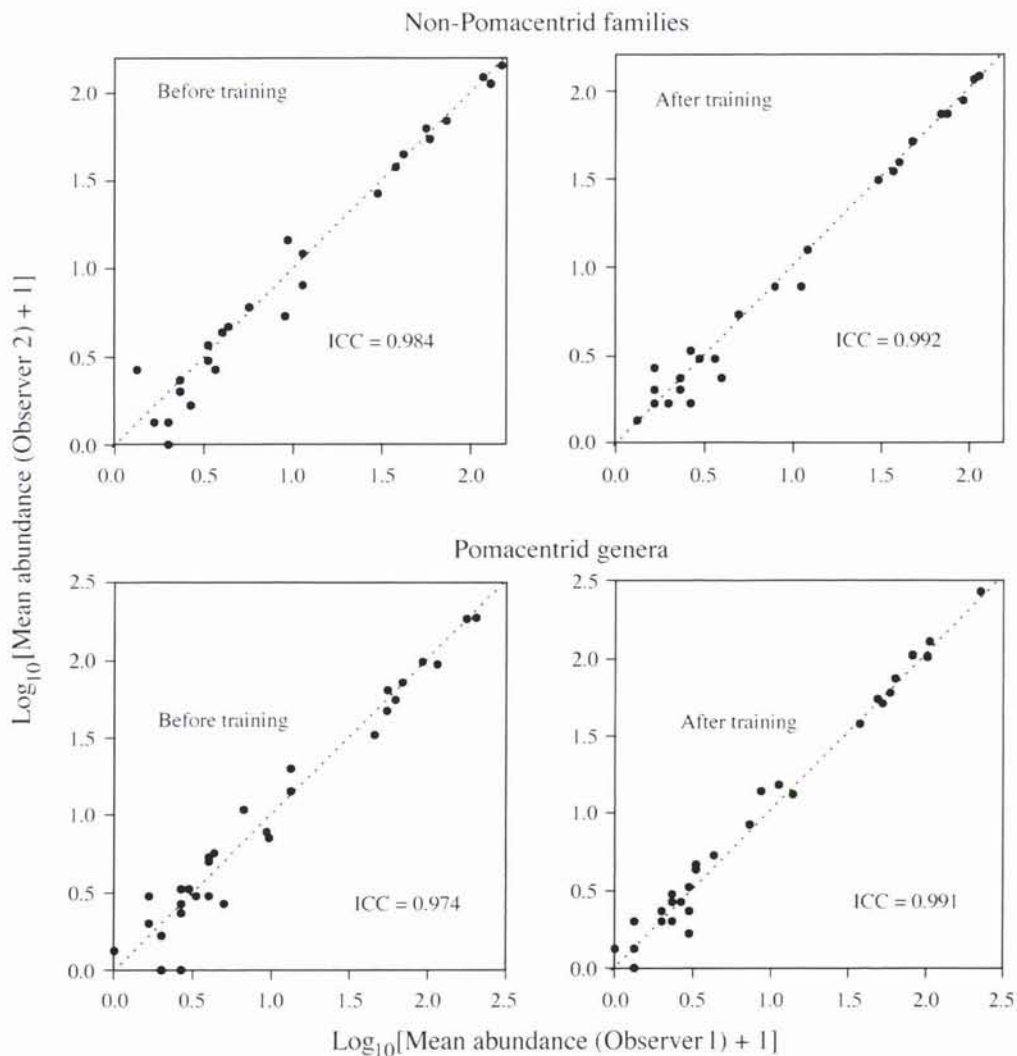


Fig. 3. Comparison of logged mean abundance estimates for observers before and after training. Dashed lines indicate lines of identity ($Obs_1 = Obs_2$).

Pomacentrus spp. (Fig. 4). In the case of *Pomacentrus* spp., training seemed to reduce the variability among counts by the less experienced observer relative to the variability in counts by the more experienced observer. For chaetodontids, labrids, and *Chromis* on E3, the less experienced observer's estimates were generally less precise than the more experienced observer's estimates.

Effects of variability between trained observers on abundance estimates

As would be expected for similarly biased observers whose differences in counts were effectively stochastic, imprecision in estimates decreased with increasing spatial scale (and averaging). CVs among transects were both larger on average and considerably more

Table 1. Results of structural regression analysis of observer estimates before and after training. CL: confidence limit

	Regression equation	90 % CL of intercept	90 % CL of slope
Non-pomacentrids			
Before training	$y = 0.059 + 0.974x$	0.022, 0.097	0.922, 1.027
After training	$y = 0.021 + 0.995x$	-0.008, 0.050	0.954, 1.036
Pomacentrids			
Before training	$y = 0.050 + 0.975x$	0.006, 0.095	0.913, 1.036
After training	$y = -0.025 + 0.985x$	-0.051, 0.002	0.949, 1.021

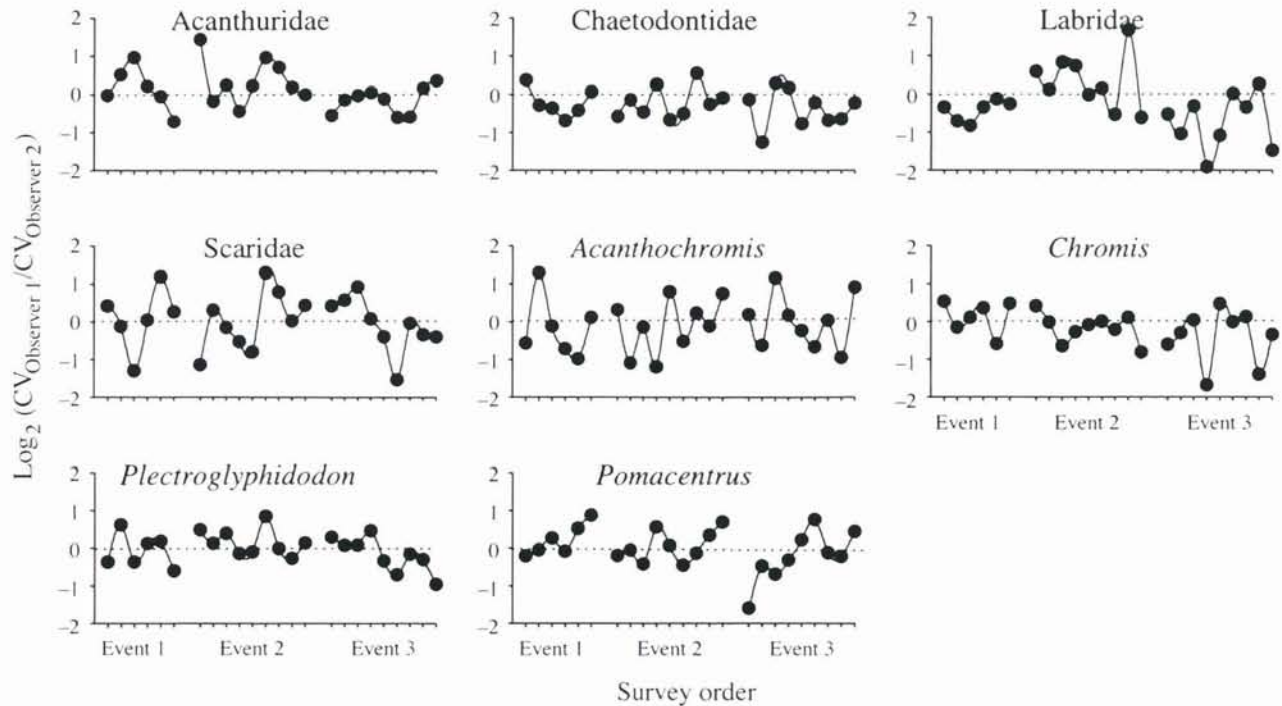


Fig. 4. Changes in the directions and magnitudes of observer-dependent (relative) precision of estimated mean abundances. Plotted points are the \log_2 -ratios of the Coefficients of Variation of observers' estimates of mean abundance at the site level. Visits and sites are arranged in the order of survey. Only plots for those taxa observed frequently on most transects are shown. Plots for taxa with lower abundances were considerably more variable than those shown. Breaks in the horizontal axis separate training periods

variable than CVs among site means or overall means (Fig. 5). Even when data were averaged over 3 sites (15 transects), however, the ratio of $SD_{\text{mean}}/\text{mean}$ (or SE/mean) was greater than 0.1 for several taxa (Fig. 5).

There was also an obvious positive relationship between abundance of target taxa and imprecision of estimates attributable to observer effects. That is, ob-

server effects were less for abundant taxa than for relatively uncommon taxa. Taxa with mean abundances of below 3 per transect exhibited highly erratic and often high estimates of inter-observer imprecision (Fig. 6). Above this threshold, however, precision was relatively constant over observed densities, with a mean of approximately 5.4% for non-pomacentrid

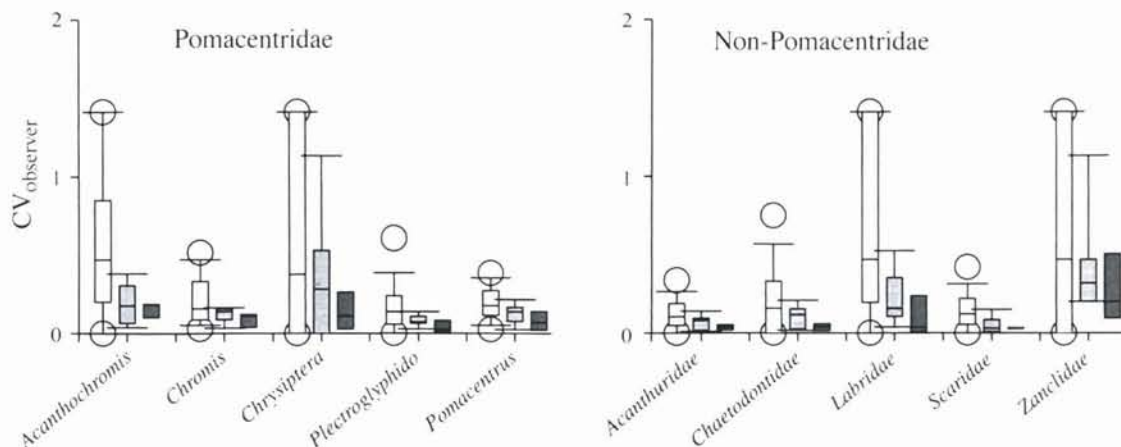


Fig. 5. Coefficients of variation between observers after training for counts along transects (open bars), in estimated site means (light grey bars), and estimated location means (dark grey bars). Central horizontal lines indicate mean CV, boxes show 25–75% quartiles, vertical lines indicate 10–90% quartiles, and circles show extreme values

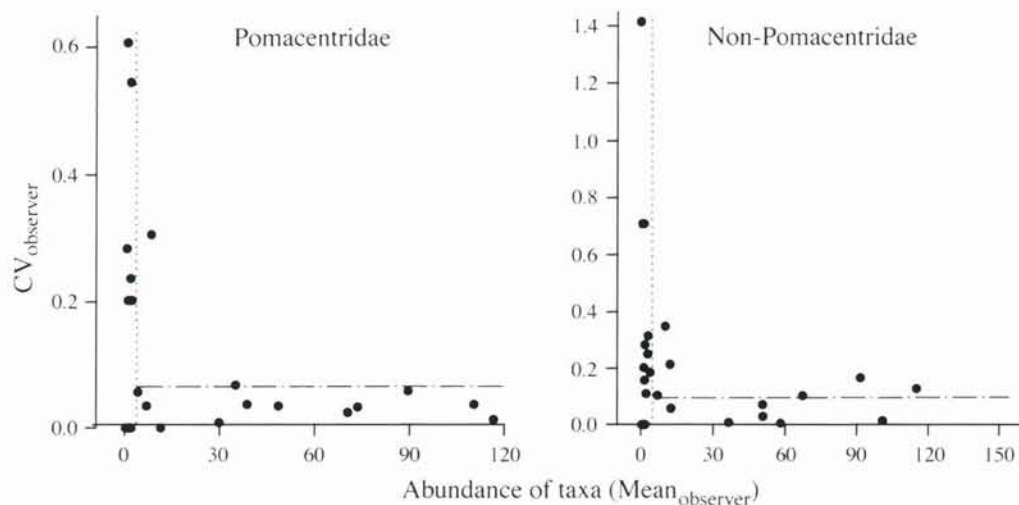


Fig. 6. Plots of inter-observer precision for trained observers against abundance of target taxa. Vertical dotted line divides the data at a mean abundance of 3 fish per transect. Horizontal dot/dash line shows the mean imprecision attributable to observer effects for the more abundant taxa

families and 9.8% for pomacentrid genera (with the omission of 2 outlying points that resulted from schools of fish moving into or out of a transect between the 2 observers' counts).

DISCUSSION

We found little evidence of consistent observer bias in our trials. In most cases, initial differences in bias were corrected (but sometimes over-corrected) by training. Where over-correction was apparent, training should probably have continued until that problem was resolved. Our data indicate that observational studies in which multiple observers must be employed may be subject to observer-related biases and imprecision, although that careful training and calibration of observers can ameliorate, but not eliminate, such effects.

The presence of observer bias in visual survey data has been noted previously (Christensen & Winterbottom 1981, Sale & Sharp 1983, Sanderson & Solonsky 1986, Mapstone & Ayling 1993), but in general little or no effort has been made to reduce such bias through training (but see Bell et al. 1985, St. John et al. 1990, Samoilys 1992). Mapstone & Ayling (1993) and St. John et al. (1990) noted that despite preliminary training and the use of previously 'experienced' observers, differential bias between observers persisted. We were able to reduce such effects on counts of reef fishes, however, through thorough training and formal evaluation of bias, although the success of training varied among taxa. At the conclusion of each of 3 training sessions, observer-specific biases were negligible for most taxa, but persisted for some. Imprecision arising from untrained or inconsistently biased observers, however, remained, and was substantial at the scales of individual transects and sites.

Reduction of differences in bias between observers might be attributable to several factors, aside from improved taxonomic skills and general competency gained from experience. Establishing the specific rules by which fish are counted is likely to be one of the most important mechanisms for standardising bias. Because most fish can move across the boundaries of transects quickly, and certainly within the time it takes an observer to scan an area, consistency in decisions whether to include or exclude fishes crossing transect boundaries is vital. In the LTMP these rules are explicitly defined in a Standard Operational Procedure (Halford & Thompson 1994) and designed so that the actual count best approximates a theoretical instantaneous count of the target species within the bounds of the transect. Different interpretation of such rules might account for initial differences in bias, but those effects should decrease with training. Judgements about individual fish by 2 observers should not produce differential bias, however, unless the rules are consistently interpreted differently.

A second mechanism that might have contributed to training 'success' in our work was increased familiarity with the sites that were being repeatedly surveyed over a short period. In this case, the apparent improvement in concurrence between observers would be misleading. To test whether apparent success of training during repeated surveys of the same sites was real or an artefact of such familiarisation, we recommend that in future training exercises a series of 'new' (unfamiliar) sites be surveyed by both observers and their counts compared. Strong similarity between observers when counting new sites would indicate that training had indeed been successful and provide the greatest assurance that counts from unfamiliar sites during routine monitoring by different observers would be comparable. If

observers were found to produce consistently different counts at the unfamiliar test sites, then further training would be required.

There was evidence in the first of our training events that correction of observer-specific bias continued throughout the training period. Under such circumstances, the training period should be extended until the observers are counting consistently and comparably. We can only speculate why such obvious trends continued throughout one training instance but not in the other two. It seems unlikely that changes in the behaviour of fishes during the training period would be responsible since such behaviours would be observed by both observers. We suggest 2 candidate explanations. Firstly, the trend from an initially negative bias in the inexperienced observer through to a positive bias may have reflected a competitive aspect of the training programme, with the trainee intent on 'out-counting' the trainer.

Alternatively, counting very many taxa may have influenced the capacity of the trainee to 'correct' for bias in one species whilst continuing to count all others efficiently. The LTMP field training of observers entailed a 'quick scan' of the data after each site was surveyed and identification of species for which counts differed consistently. Differences were discussed to reduce the difference in bias in subsequent counts. The main focus was on developing consistency in searching techniques and the rules by which individuals were included or excluded from counts (points explicitly outlined in the Standard Operational Procedure developed after this first training event). This process almost certainly had the effect at the next site of increasing the attention given to counting the species that previously differed most between observers. Ultimately, the focus on those species where differences were greatest may have led to over-correction for the initial biases. Lincoln-Smith (1989) documented the effects of counting increasing numbers of species on the reliability of counts for each species, and recommended that diverse assemblages should be counted in sub-sets to minimise species-specific bias in estimated abundances. Sub-setting of diverse assemblages for counting purposes should be considered, but the benefits gained in reduced observer-related bias or imprecision must be weighed against the increase in time taken for such surveys, and hence their cost.

Results of the second and third training sessions were closer to those desired from a training programme. In general, observer-specific biases were relatively slight at the beginning, of training and had essentially disappeared by the end of training. This may reflect the use of 3 (rather than 2) sets of surveys of each site, with a corresponding increase in discussion during training.

The obvious effect of abundance on the precision of estimates between observers is of particular concern. There are 2 plausible explanations for the imprecision at low density. The first is simply a mathematical artefact of whole numbers in that the omission or inclusion of one fish when abundances are low creates a proportionately large variation to the mean. Certainly some of the very high CVs we saw in our data occurred when 1 observer scored 1 fish whilst the second scored 2 or 3 on a particular site. The second factor is again related to counting numerous species, in that the more abundant taxa are likely to create a more pronounced search image in the observer's mind, leading to the rarer species being overlooked. This is the type of error which can only be countered by experience and could account for some of the 'saw tooth' shape in comparison profiles seen in our training plots.

Persistent observer-dependent biases have the potential to precipitate spurious patterns in data, especially when observers are confounded with other factors (e.g. locations, times) in the design of sampling programmes. Correction of such biases, preferably through training, is thus central to the correct interpretation of data from multi-observer programmes. Training exercises of the sort we describe are important not only for the training of new observers, but also for the periodic comparison of experienced observers to flag developing differences in bias resulting from different responses to experience. Such training or calibration is important also because it forces observers to consider explicitly the details of standardised survey techniques, a factor crucial to long-term consistency in data collection for monitoring. Even after training, however, considerable imprecision attributable solely to differences between observers' counts can remain, particularly in non-aggregated data. Estimates from different observers were relatively precise only when data were aggregated, apparently smoothing out inconsistencies in counts of individual species. Such imprecision will reduce the potential to detect pattern in survey data and should be considered explicitly when planning multi-observer programmes and interpreting their results, especially when it is not expected that data will be aggregated during analyses. If the objectives of a monitoring programme include monitoring the dynamics of assemblage structure or species diversity, the relatively large imprecision in counts of less abundant species may be particularly influential on the potential for erroneous inferences of changes in diversity from such monitoring. It is essential that such considerations be incorporated into the design of monitoring programmes, especially where multiple observers are involved in multi-species surveys over large spatial scales or long periods.

Acknowledgements. We thank Andrew Halford and Alistair Cheal for being willing trainees, numerous volunteers who acted as field assistants, and the crew of the RV 'Sirius' for field support. Hugh Sweatman and 3 anonymous reviewers provided helpful comments on drafts, and Ray Muller provided useful statistical advice. Fieldwork and salary for A.A.T. were funded by the Australian Institute of Marine Science, and B.D.M. was supported by the Australian Institute of Marine Science, the Great Barrier Reef Marine Park Authority, and the Cooperative Research Centre for the Ecologically Sustainable Development of the Great Barrier Reef. This article is Australian Institute of Marine Science contribution number 880.

LITERATURE CITED

- Allen GR (1991) Damselfishes of the world. Mergus, Melle
- Andrew NL, Mapstone BD (1987) Sampling and the description of spatial scale in marine ecology. *Oceanogr Mar Biol Annu Rev* 25:39–90
- Ayling AM, Ayling AL (1986) A biological survey of selected reefs in the Capricorn Section of the Great Barrier Reef Marine Park. Report to the Great Barrier Reef Marine Park Authority, Townsville
- Ayling AM, Ayling AL (1992) Bramble reef replenishment area—baseline survey. Report to the Great Barrier Reef Marine Park Authority, Townsville
- Bell JD, Craik GJS, Pollard DA, Russell BC (1985) Estimating length frequency distributions of large reef fish underwater. *Coral Reefs* 4:41–44
- Bortone SA, Martin T, Bundrick CM (1991) Visual census of reef fish assemblages: a comparison of slate audio and video recording devices. *Northeast Gulf Sci* 12:17–23
- Bouchon-Navaro Y (1980) Quantitative distribution of the Chaetodontidae on a reef of Moorea Island (French Polynesia). *J Exp Mar Biol Ecol* 55:145–157
- Brock RE (1982) A critique of the visual census method for assessing coral reef fish populations. *Bull Mar Sci* 44:893–898
- Brock VE (1954) A preliminary report on a method of estimating reef fish populations. *J Wildl Manage* 18:297–308
- Buckley RM, Hueckel GJ (1989) Analysis of visual transects for fish assessment on artificial reefs. *Bull Mar Sci* 44: 893–898
- Caughley G (1977) Analysis of vertebrate populations. Wiley, New York
- Caughley G, Sinclair R, Scott-Kimms D (1976) Experiments in aerial survey. *J Wildl Manage* 40:290–300
- Christensen MS, Winterbottom R (1981) A correction factor for, and its application to, visual censuses of littoral fish. *S Afr J Zool* 16:73–79
- DeMartini EE, Roberts D (1982) An empirical test of the biases in the rapid visual technique for species-time census of reef fish assemblages. *Mar Biol* 70:129–134
- Doherty PJ (1987) The replenishment of populations of coral reef fishes, recruitment surveys, and the problems of variability manifest on multiple scales. *Bull Mar Sci* 41(2): 411–422
- Eckert GJ (1984) Annual and spatial variation in recruitment of labroid fishes among seven reefs in the Capricorn/Bunker Group, Great Barrier Reef. *Mar Biol* 78:123–127
- English KK, Bocking RC (1992) A robust procedure for estimating salmon escapement based on the area-under-the-curve method. *Can J Fish Aquat Sci* 49:1982–1989
- Estes JA, Gilbert JR (1978) Evaluation of an aerial survey of Pacific Walrus (*Odobenus rosmarus divergens*). *J Fish Res Bd Can* 35(8):1130–1140
- Fernandes L (1990) Effect of the distribution and density of benthic target organisms on manta tow estimates of their abundance. *Coral Reefs* 9:161–165
- Fernandes L, Marsh H, Sinclair D (1990) Bias in the manta tow technique surveys of *Acanthaster planci*. *Coral Reefs* 9: 155–160
- Fowler AJ (1987) The development of sampling strategies for population studies of coral reef fishes. A case study. *Coral Reefs* 6:49–58
- Francis CM (1994) Population changes of boreal forest ducks—A comment. *J Wildl Manage* 58:582–586
- Greene LE, Alevizon WS (1989) Comparative accuracies of visual assessment methods for coral reef fishes. *Bull Mar Sci* 44:899–912
- Halford AR, Thompson AA (1994) Visual census surveys of reef fish. Long-term monitoring of the Great Barrier Reef Standard Operating Procedure Number 3. Australian Institute of Marine Science, Townsville
- Harmelin-Vivien ML, Harmelin JG, Chauvet C, Duval C, Galzin R, Lejeune P, Barnabe G, Blanc F, Chevalier R, Duclerc J, Lassere G (1985) Evaluation visuelle des peuplements et populations de poissons: methodes et problemes. *Rev Ecol Terre Vie* 40:467–539
- Holt RS, Colongne J (1987) Factors affecting line transect estimates of dolphin school density. *J Wildl Manage* 51: 836–843
- Jones RS, Chase JA (1975) Community structure and distribution of fishes in an enclosed high island lagoon in Guam. *Micronesica* 11:127–148
- Keast A, Harker J (1977) Strip counts as a means of determining densities and habitat utilization patterns in lake fishes. *Environ Biol Fish* 1:181–188
- Kennelly SJ, Underwood AJ (1984) Underwater microscopic sampling of a sublittoral kelp community. *J Exp Mar Biol Ecol* 76:67–78
- Kimmel JJ (1985) A new species-time method for visual assessment of fishes and its comparison with established methods. *Environ Biol Fish* 12(1):23–32
- Leatherwood S (1979) Aerial surveys of bottlenose dolphin, *Tursiops truncatus*, and the West Indian manatee, *Trichechus manatus*, in the Indian and Banana rivers Fla. *Fish Bull US* 77:47–59
- Leatherwood S (1982) Effect of varying altitude on aerial surveys of bottlenose dolphins. *Rep Int Whaling Comm* 32: 569–575
- Lincoln-Smith MP (1989) Improving multispecies rocky reef fish censuses by counting different groups of species using different procedures. *Environ Biol Fish* 26:29–37
- Mapstone BD (1988) The determination of patterns in the abundance of *Pomacentrus moluccensis* Bleeker on the Southern Great Barrier Reef. PhD thesis, University of Sydney
- Mapstone BD, Ayling AM (1993) An investigation of optimum methods and unit sizes for the visual estimation of abundances of some coral reef organisms. Report to the Great Barrier Reef Marine Park Authority, Townsville
- Marsh H, Sinclair DF (1989) Correcting for visibility bias in strip transect aerial surveys of aquatic fauna. *J Wildl Manage* 53:1017–1024
- McCormick MI, Choat JH (1987) Estimating total abundance of a large temperate-reef fish using visual strip-transects. *Mar Biol* 96:469–478
- Myers RF (1991) Micronesian reef fishes, 2nd edn. Coral Graphics, Guam
- Randall JE, Allen GR, Steene RC (1990) The complete diver's and fishermen's guide to fishes of the Great Barrier Reef and Coral Sea. Crawford House Press, Bathurst

- Russell BC, Talbot FH, Anderson GRV, Goldman B (1978) Collection and sampling of reef fishes. In: Stoddart DR, Johannes RE (eds) Monographs on oceanographic methodology 5: Coral reefs: Research methods. UNESCO, Norwich, p 329–345
- Sale PF, Doherty PJ, Eckert GJ, Douglas WA, Ferrell DJ (1984) Large-scale spatial and temporal variation in recruitment to fish populations on coral reefs. *Oecologia* 64:191–198
- Sale PF, Douglas WA (1981) Precision and accuracy of visual census technique for fish assemblages on coral patch reefs. *Environ Biol Fish* 6:333–339
- Sale PF, Sharp BJ (1983) Correction for bias in visual transect censuses of coral reef fishes. *Coral Reefs* 2:37–42
- Samoilys MA (1992) Review of the underwater visual census method developed by DPI/ACIAR project: visual assessment of reef fish stocks. Conference Workshop Series, Queensland Department of Primary Industries, No. QC92006
- Sanderson SL, Solonsky AC (1986) Comparison of a rapid visual and a strip transect techniques for censusing reef fish assemblages. *Bull Mar Sci* 39:119–129
- Solazzi MF (1984) Relationship between visual counts of coho salmon (*Oncorhynchus kisutch*) from spawning fish surveys and the actual number of fish present. In: Symons PEK, Waldichuck M (eds) Proceedings of the workshop on stream indexing for salmon escapement estimation. *Can Fish Aquat Sci Tech Rep* 1326:175–186
- St. John J, Russ GR, Gladstone W (1990) Accuracy and bias of visual estimates of numbers, size structure and biomass of a coral reef fish. *Mar Ecol Prog Ser* 64:253–262
- Thresher RE, Gunn JS (1986) Comparative analysis of visual census techniques for highly mobile reef associated piscivores (Carangidae). *Environ Biol Fish* 17:93–116
- Williams DMcB (1982) Patterns in the distribution of fish communities across the central Great Barrier Reef. *Coral Reefs* 1:35–43
- Williams DMcB, Hatcher AI (1983) Structure of fish communities on outer slopes of inshore, midshelf and outer shelf reefs of the Great Barrier Reef. *Mar Ecol Prog Ser* 10: 239–250

This article was presented by Daniel Alongi (Senior Editorial Advisor), Townsville, Australia

*Manuscript received: October 5, 1996
Revised version accepted: May 9, 1997*