IS MY MARINE PARK ACHIEVING ITS CONSERVATION GOAL? A STRAIGHTFORWARD ANALYTICAL APPROACH TO HELP MANAGERS ADDRESS THIS QUESTION

Anthony (Tony) Badih Rouphael

Email address: tony.rouphael@gmail.com

P.O Box 634, Como, Western Australia 6952, Australia

ABSTRACT

There is an increasing demand that managers of marine parks quantitatively demonstrate the achievement of their conservation goals. Monitoring is one tool that can help with this. One component of monitoring that is challenging for managers is the statistical treatment of monitoring data. Commonly used approaches, such as null hypothesis tests, are conceptually challenging and operationally complex, potentially leading to wrong conclusions and poor decisions. A more straightforward approach is parameter estimation with confidence intervals. Parameter estimation focuses on estimating the size of change or difference (an 'effect size') in a response variable and comparing this with a pre-defined effect size called a management threshold. Confidence intervals indicate the level of precision in estimates of change, which make for more balanced conclusions. Parameter estimation is also conducive to graphing, which can facilitate interpretation and communication to non-scientists. In this paper, I demonstrate three examples of parameter estimation and discuss their relative strengths and weaknesses. By presenting these examples, I hope to encourage managers to adopt statistical approaches that allow them to quantify environmental change in a way that will contribute to defensible conclusions, facilitate timely decision making and be understood by stakeholders.

Key words: Confidence interval, effect size, management threshold, marine park, monitoring, parameter estimation, statistical analysis

INTRODUCTION

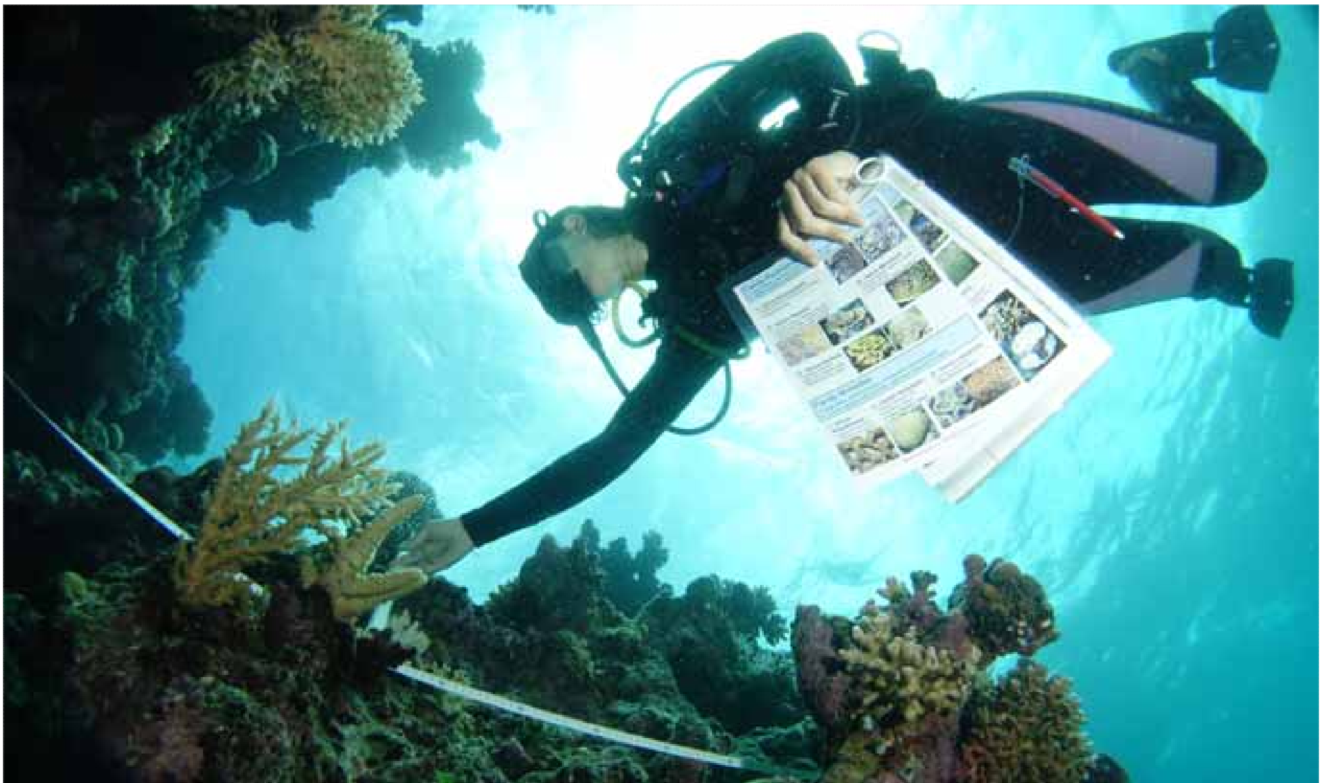
Most multiple-use marine parks (hereafter, 'marine parks') have several goals, including biodiversity conservation, facilitating tourism and supporting sustainable fisheries (Day et al., 2015). A zoning scheme and activity-specific regulations are among the management tools used to balance competing goals, but these are not always guaranteed to work. Therefore, monitoring is required to assess whether conservation goals are being achieved and to prompt changes in management strategies where there is evidence of an unacceptable level of environmental change (Addison et al., 2015a; Hockings et al., 2006; Pomeroy et al., 2004).

To directly compare monitoring data with a conservation goal, the latter needs to be quantitatively defined (Burgman et al., 2012). Quantitatively defining a conservation goal usually involves specifying a management threshold for the response variable of interest. A threshold is a value that if exceeded would trigger management intervention or further investigation. In some instances, a management threshold might be a fixed value, such as a water quality standard (ANZECC, 2001) or presented as a range of values called quantitative condition categories (Addison et al., 2015b). In other situations, a threshold can be defined in terms of a level of change or a mean difference, say between impact and control sites, which would be of management concern. Such change or difference, irrespective of its management importance, is also referred to as an 'effect size' (Cohen, 1990; Cumming, 2012). A management threshold can be considered a pre-defined effect size of management importance. A simple quantitative example of a threshold would be a 30 per cent decrease in the amount of live coral at a snorkel site relative to control sites. Another example could be when there is on average > 10 damaged coral colonies at a dive site compared to control sites.

Specifying a management threshold before the start of monitoring is important because it forces a manager to give due consideration to the level of environmental change that would be of ecological or social importance (Di Stefano, 2004). This is vital because it facilitates defensible conclusions and timely decision making. Also, from a philosophical perspective, specifying a management threshold before the start of monitoring lessens the risk of a monitoring programme deteriorating into a Baconian data gathering exercise (Underwood, 2000a).

Two other key components of monitoring are the approach to statistically treat data so these can be compared with a management threshold and the approach used to help interpret the cause of observed environmental change (Fabricius & De'ath, 2004). The latter relates to the monitoring design. Both factors are equally important, but it is probably the statistical treatment of data that often proves most challenging to managers of marine parks. This is because managers may not be trained in statistical methods, and some analytical approaches can be computationally and conceptually demanding (Walshe & Wintle, 2006). Such challenges may make it difficult for managers to successfully communicate or defend results to sceptical decision makers (e.g. politicians, funding agencies) and economic stakeholders (e.g. tourism operators).

It is the statistical treatment of data collected in a monitoring programme and how the outputs are compared to a management threshold that is the focus of this paper. I begin by highlighting some of the challenges for managers using null hypothesis tests and control charts to treat data. I then describe parameter estimation with confidence intervals (hereafter 'parameter estimation'), which offers a simple and practical alternative. To illustrate the utility of parameter estimation, I present three examples or variants of this analytical approach. The three vary in how an 'effect size' is quantified, how uncertainty about the effect size is interpreted and how the effect size is compared with a management threshold. Each example has its own strengths and weaknesses, which are all discussed. My intention is not to endorse one variant over another, nor do I recommend replacing superior analytical approaches if the technical expertise is available. Instead, I hope to encourage managers to adopt analytical approaches that allow environmental change to be quantified in a way that minimises misinterpretation, will contribute to defensible conclusions and will be understood by stakeholders who may not be scientists.



Monitoring is a critical element of MPA management © Commonwealth of Australia (GBRMPA)

NULL HYPOTHESIS TESTS AND CONTROL CHARTS

One of the most common approaches to statistically treat data is null hypothesis testing, typically using Analysis of Variance (ANOVA) or similar (e.g., Generalised Linear Models) (Benedetti-Cecchi, 2001; Downes et al., 2002; Green, 1979; Underwood, 2000a). This involves calculating a probability (p-value) indicating the strength of evidence against a null hypothesis (i.e. a hypothesis of no difference or change). Null hypothesis testing has a number of strengths for the assessment of environmental change (Underwood, 1997; Downes et al., 2002). It provides a logical basis for distinguishing between alternative hypotheses, such as whether a disturbance has or has not resulted in environmental change. When using ANOVA, objective decision rules can be used to reduce the risk of falsely concluding that there has been change (Type I error). When combined with power at the planning stage, the Type II error rate (risk of falsely concluding that there has not been a change) can also be predicted for the study (Mapstone, 1995).

Unfortunately, there are at least three reasons why null hypothesis testing may be unsuitable for managers of marine parks. First, a null hypothesis of no environmental change is usually, if not always, invalid when associated with activities permitted in marine parks. This is because the effects of tourists on benthic habitats (Marion & Rogers, 1994) and the effects of fishing (Young et al., 2014) cannot be mitigated entirely even with management. Therefore, the question to be addressed by monitoring in marine parks is not “has an activity caused environmental change?”, but “what is the size of that change?”, or, more pointedly, “how does the size of change relate to a management threshold?” Second, many managers are administrators, not scientists, and may lack the technical knowledge to use ANOVA or similar statistical tools. Consequently, they may struggle to accurately explain outputs such as p-values or recognise some of the assumptions that need to be considered when interpreting p-values (Stewart-Oaten, 1996; Walshe et al., 2007). Third, a p-value, when presented on its own, does not convey uncertainty in conclusions (Cumming, 2012; Cumming & Calin-Jageman, 2017).

Control charts have also been proposed as an alternative way to treat monitoring data (Burgman, 2005). Indeed, control charts share some advantages offered by parameter estimation. However, standard control charts do not normally illustrate uncertainty in the parameter being estimated and assume that the variable being monitored has little temporal variability (Morrison, 2008). Further, it is not straightforward to link management thresholds to control sites using standard control charts.

PARAMETER ESTIMATION

One analytical method that avoids some of the limitations mentioned above is parameter estimation (Di Stefano, 2004; Rouphael et al., 2011; Walshe & Wintle, 2006). Widely used in medicine and psychology (Altman et al., 2000; Cumming & Calin-Jageman, 2017), parameter estimation, especially in conjunction with management thresholds, appears to be largely ignored by managers of marine parks. This is unfortunate because parameter estimation focuses on the size of change, provides an intuitive way to quantify uncertainty in results, allows for a straightforward way to compare estimates of change or differences with management thresholds and, if summarised graphically, facilitates communication of results to laypersons. These issues are explored more fully below.

To quote Fowler et al. (1998, page 6), “the measures which describe a variable of a sample are called statistics. It is from the sample statistics that the parameters of a population are estimated”. Thus the means, medians and effect sizes that are calculated from monitoring data are imprecise estimates of the true population parameters. Imprecision is captured using confidence intervals. More precisely, a confidence interval includes a single value estimate, such as a sample mean or effect size, and a range of values around an estimate that are also considered plausible for the population under investigation (Gardner & Altman, 2000).

Confidence intervals can be used like null hypothesis tests to derive dichotomous conclusions based on the degree of overlap between pairs of confidence intervals or whether a confidence interval includes zero (Tryon, 2001). However, like null hypothesis testing, deriving conclusions in this way is, in part, a function of sample size and does not take into consideration the size of the effect (Cumming, 2012; Di Stefano et al., 2005).

Although it is desirable to have sufficient replication to provide a definitive answer to the question of whether a mean or an effect size and their associated confidence intervals are entirely above or below a threshold, this will rarely be the case for managers. Typically, there will be too few resources to provide precise estimates and thus the associated confidence intervals will be wide and overlap thresholds. Consequently, conclusions derived from monitoring data will need to be tempered by the width of a confidence interval, how much a confidence interval overlaps with a threshold and long-term trends in the variable being monitored (Masson & Loftus, 2003). Conclusions drawn in this way may be more subjective than those derived using null hypothesis tests. Nevertheless, this increased level of subjectivity may be acceptable for most stakeholders in the context of marine parks, especially if a precautionary approach is adopted where decisions favour environmental outcomes.

BACKGROUND TO THE VARIANTS

The three variants (or examples) of parameter estimation presented in this paper are modified versions from the literature. All examples relate to activities that are legally permitted in marine parks and, for the purposes of realism, to situations where a Before/After by Control/Impact (BACI) monitoring design is unattainable. More precisely, all relate to situations where a single impact site is monitored and compared with multiple control sites and where there are no baseline data. In terms of monitoring a single impact site, such a situation is not unrealistic in many marine parks. Managers are typically less concerned about an average effect among replicate management zones or sites, compared with understanding impacts to individual sites, especially those that are iconic, popular or unique. In terms of a baseline, legal activities within marine parks are often well established before the instigation of monitoring, rendering it impossible to obtain baseline data (Buckley et al., 2008). Nevertheless, managers should always attempt to incorporate baseline and other elements of experimental design into their monitoring design wherever possible (Underwood, 2000b).

Although I encourage the adoption of BACI style monitoring design where possible, the effect sizes and management thresholds illustrated in this paper represent spatial differences between impact and control sites rather than interactions. An interaction can be defined as “some pattern of difference from before to after a planned disturbance in the relationship between the mean of whatever variable is measured in the disturb location compared with that in the control” (Underwood, 1997). Although effect sizes and confidence intervals can be calculated for an interaction (Masson & Loftus, 2003), they are not straightforward to interpret (Di Stefano, 2004). However, the primary reason why effect sizes in this paper are based only on spatial differences is because of the absence of baseline data. When a statistical interaction cannot be used to evaluate whether there has been an impact, more caution is required when inferring the cause of change. A levels-of-evidence approach would be useful to help infer causation in such situations (Fabricius & De’ath, 2004).

Careful thought is required when choosing the appropriate source of variability to calculate confidence intervals especially if confidence intervals are to be compared between impact and control sites. When a monitoring design is characterised by replicate impact and control sites, it is reasonable to calculate confidence intervals for treatment means based on site level replication (i.e. variability among sites). However, when there is only one impact site there is no site level replication for the impact treatment. Instead, within-site level replication (e.g., transects or quadrats used to sub-sample an individual site) needs to be used to construct a confidence interval for the impact site in order to assess the precision of the mean estimate. For the control sites, there are two options to generate a confidence interval. The first is to generate a confidence interval using site level replication because there are multiple control sites. The second is to pool all within-site replicates from all control sites to generate a confidence interval, which has the advantage that confidence intervals are constructed for both the impact and control treatments using the same units. An issue with using within-site variability in this way is that one is making the assumption that there is no among site level variability. If this assumption is wrong, pooling replicates to generate a confidence interval in this way may lead to a misleadingly high level of precision for the mean of the control treatment. For this reason, when the impact mean is plotted on the same graph as the control mean, it is usually preferable that only the confidence interval for the control mean is shown. This is to avoid confusing different sources of variability used to construct the two confidence intervals, potentially resulting in misinterpretation. This is the approach illustrated in this paper. Nevertheless, there are at least two reasons why it is still helpful to plot the confidence interval of the impact site on a separate graph. First, it will allow managers to evaluate the level of precision for the mean used at the impact site. Second, change in variability at the impact site might be indicative of human disturbance and worthy of investigation (Warwick & Clarke, 1993).

Whatever approach is adopted to construct confidence intervals, the units used, the type of confidence interval (e.g. 90 per cent vs 95 per cent), degrees of freedom, assumptions and limitations need to be made explicit to stakeholders. The data should also be available to experts who might be required to undertake a more detailed assessment if uncertainty remains high and/or the results are challenged by stakeholders.

THE THREE VARIANTS

In the remaining section of this paper, I illustrate three variants or examples of parameter estimation. They differ in how the effect sizes are estimated and how confidence intervals are used to aid interpretation. Each variant presented is structured as follows: theory, scenario and strengths and weaknesses. The management thresholds used to illustrate the variants are fictional and thus have no ecological or social basis. In addition, the examples are based on hypothetical data chosen to help the reader better appreciate the application of parameter estimation.

All thresholds used in the examples are benchmarked against control sites, but parameter estimation is equally suitable using thresholds based on fixed values. Benchmarking thresholds to control sites is important when response variables being monitored are spatially and temporally dynamic even in the absence of human activities. This is usually the case with marine organisms and their habitats (Connell et al., 2004; Hatcher et al., 1989; Hughes et al., 1999). When linking a management threshold to control sites the following should be considered. That control sites are chosen to be as similar as possible to the impact site except for the presence of the activity (e.g. snorkelling) that is potentially contributing to change in the response variable (e.g. coral cover) being monitored (Downes et al., 2002). This is usually not a great challenge for activities with spatially discrete impact zones, such as snorkelling, scuba diving and reef walking. Another consideration is that control sites and impact sites are not already severely damaged as a result of a widespread disturbance event, such as declining water quality or increasing sea surface temperatures (Hughes, 1994). Under such circumstances, benchmarking a management threshold against control sites could be misleading. In this situation, linking a threshold to a restoration outcome might be more appropriate.

Variant 1: Effect size and its confidence interval compared with a threshold

Theory

The first variant illustrating how data from an impact site can be compared with control sites is based on an approach suggested by Rouphael et al. (2011) for managers of marine parks. They propose comparing the difference between the means of the impact and control sites (i.e., an 'effect size') and its confidence interval with a management threshold. In terms of calculating an effect size, Fowler et al. (1998) recommend subtracting the smaller mean (irrespective of whether it relates to the impact or control sites) from the larger mean to maintain positive differences. They also show how to calculate a confidence interval for an effect size, which involves pooling the two treatment sources of variance. In a situation when there is a single impact site but multiple control sites, Rouphael et al. (2011) suggest using the within-site sources of variances from each treatment to ensure the variances used to construct the confidence interval for the effect size are based on the same units. But as stated before, pooling subsamples (e.g. transects or quadrats) from all control sites would be valid only under the assumption that there is no among control site variability. An alternative would be to only use among control site variance and associated degrees of freedom to generate the confidence interval for the effect size. This is the approach taken in this example.

Fox (2001) and Di Stefano et al. (2005) graphically illustrate how the approach could be interpreted depending on where the effect size (± confidence interval) was in relation to the threshold. A modified version of their graphs is illustrated in Figure 1. Scenario 1 clearly indicates that the threshold is exceeded, which would trigger a site-specific investigation or management intervention. Scenario 2 is the opposite of Scenario 1 and would not require a response. Scenario 3 shows that the middle of the confidence interval is on the threshold and thus there is an equal chance that the threshold has or has not been exceeded. The broad width of the confidence interval suggests that the estimate is not very precise. Under such a scenario, a manager could immediately undertake a new survey, if resources were available, or instigate management as a

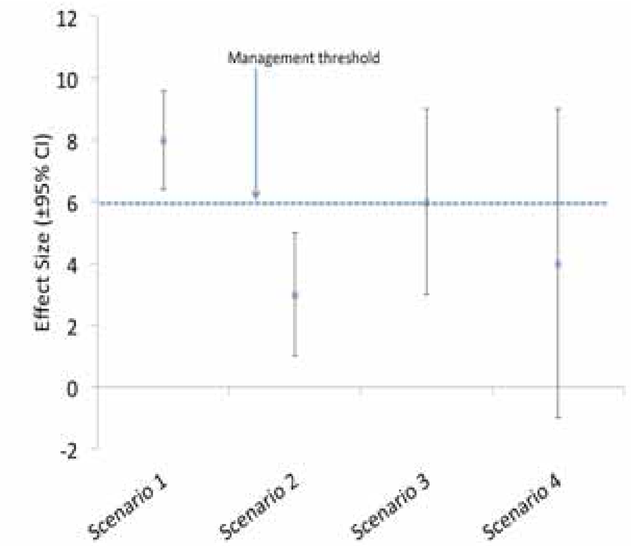


Figure 1: Four potential scenarios or outcomes following a survey that differ in terms of where the effect size lies in relation to the threshold and the width of the 95 per cent confidence interval (refer to text for interpretation).

precaution. In Scenario 4, the middle of the confidence interval is below the threshold, but the upper length of the confidence interval intercepts the threshold. In Scenario 4, the confidence interval is so wide that there is no sensible way to interpret the data. A confidence interval, such as that shown in Scenario 4, needs to be reduced in width through increased replication.

With this variant, the monitoring data are summarised using two figures. One figure is used to determine if the effect size (± confidence interval) is above or below the threshold and to indicate the level of precision of the effect size via the widths of the confidence intervals. The other figure is used to show the means for the impact and control sites separately, thus indicating which had the larger mean at the time of sampling. This figure would also be used to illustrate temporal trends, which is helpful when interpreting data. An example is illustrated below.

Scenario

In this scenario, a manager is concerned about the level of browsing by domestic animals in a mangrove stand (the 'impact site') situated in a resource-use zone of a marine park. The conservation goal for this zone is to maintain the structural integrity of the stand, while still permitting livestock access to feed. However, rangers have reported dead seedlings in the mangrove stand and are concerned about recruitment failure. Nevertheless, they have also observed dead seedlings in mangrove stands in other areas of the marine park where livestock and other human activities are not permitted. It is therefore apparent that seedling mortality may occur as a result of natural processes. Rather than prohibiting livestock in the resource-use zone, the manager decides to monitor seedling mortality to ensure the level of mortality there does not greatly exceed an average level observed among three control stands (i.e. the control sites). To operationalise the conservation goal, the manager links it to a quantitative management threshold. If the threshold is exceeded the manager will implement management to reduce the risk of further seedling mortality. The manager suggests that, on average, more than six dead seedlings (per 50 m^2) in the mangrove stand in the resource-use zone, relative to the average number in the unbrowsed stands, would be worthy of management concern. Obviously, seedling mortality is not the only variable that could be used to monitor the structure of a mangrove stand, but a single variable is used here for illustrative purposes only.

Hypothetical data, representing four consecutive surveys, are graphically summarised in two ways to aid interpretation. The two ways are shown in Figures 2 and 3. Figure 2 illustrates the effect size and its 95 per cent confidence interval for each survey. Each of the four effect sizes and confidence intervals are also shown in relation to the management threshold. To reiterate, in this example, the effect size is the difference between the

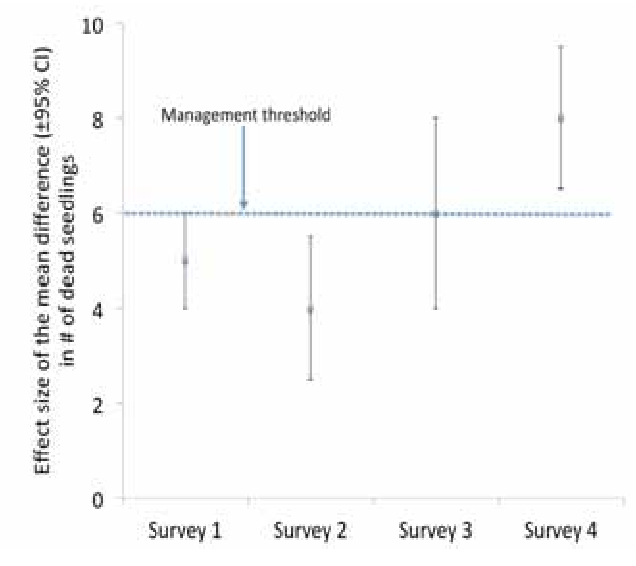


Figure 2: Shows the effect sizes and the associated 95 per cent confidence intervals for the number of dead mangrove seedlings over four surveys. These are shown in relation to a management threshold. See text for detail.

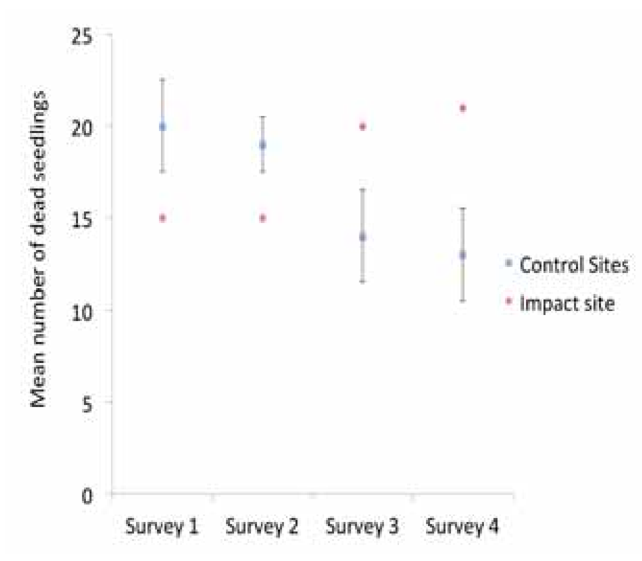


Figure 3: Means of the impact and control sites for four surveys. Note that this figure is used to complement the previous figure by illustrating which mean is larger at each survey and shows trends over time. Also note there are no confidence intervals for the impact means or a threshold in this figure (refer to text for details).

mean of the single impact site and the mean of the control sites. The middle value of a confidence interval is the most plausible estimate of the effect size while values at the extremities of the confidence interval are less plausible. Recall that the confidence intervals for the effect sizes in Figure 2 are based only on among site variability for the control treatment because there is no site level variability for the impact treatment. Figure 3 illustrates the same data, but shows the impact and control site means individually. Note that the mean number of dead seedlings at the control sites had fallen following Survey 2. As stated earlier, most environmental variables that will be monitored in a marine park will vary naturally through time independent of human influence. Note also that there is no confidence interval for the impact site in Figure 3 for reasons explained earlier. But on a separate graph (not illustrated) it would be desirable to calculate a confidence interval for the impact site to assess the level of precision and how the width of the confidence interval changes through time.

Eye on the Reef © Commonwealth of Australia (GBRMPA)

Figure 2 indicates that at the time of Survey 3 the threshold may have been exceeded while at Survey 4 the threshold is clearly exceeded. Figure 3 confirms that during Survey 3 and Survey 4 the mean number of dead seedlings is higher at the impact site compared with the control sites. This suggests that browsing, rather than natural processes is the cause of the increased number of dead seedlings at the impact site. In Figure 2, Survey 3 shows that the threshold is potentially exceeded because a large proportion of the confidence interval overlaps the threshold. Such a result might prompt the manager to increase replication to reduce uncertainty or to instigate management as a precaution. Survey 4 is unambiguous in terms of exceeding the threshold because the entire confidence interval is above the threshold (Figure 2). This result would warrant management action.

Strengths and weaknesses

A strength of this variant is that a confidence interval is generated for the actual effect size, not just the individual mean estimates. When an effect size is combined with a confidence interval, as shown in Figure 2, a manager can intuitively assess how likely a monitoring programme is able to clearly distinguish whether a threshold has been exceeded (Walshe & Wintle, 2006; Walshe et al., 2007). For instance, a wide confidence interval overlapping a threshold makes it difficult to derive clear-cut conclusions; in this case, managers should consider increasing the level of replication. Andrew and Mapstone (1987) and Green (1979) show how precision of an effect size can be quantified objectively. Andrew and Mapstone (1987), Di Stefano et al. (2005) and Cumming (2012) also show how to determine the sample size for a desired level of precision.

One challenge with this and the other variants of parameter estimation described in this paper is the need to minimise subjectivity when drawing conclusions about whether a threshold has been exceeded or not. Subjectivity will be high when a large amount of the confidence interval overlaps the threshold making it impossible to conclusively state whether the threshold has or has not been exceeded. Reducing the width of the confidence interval by increasing replication is the most direct way to reduce uncertainty, but this may be impossible if managers have limited resources. Consequently, in some situations, it might be prudent to assume that a threshold has been reached even if the most plausible estimate of the effect size (i.e., the middle value of the confidence interval) is below it, but a large amount of the confidence interval overlaps the threshold. As this could have costly ramifications, managers and stakeholders should agree beforehand on the type of action to be triggered in such a scenario.

Although the focus of this paper is on data analysis, it is important to reiterate the weaknesses of the monitoring design illustrated in this and the other examples. The weaknesses are the absence of baseline data and only having one impact site. A sub-optimal design such as this can limit the ability to reliably infer causation (Downes et al., 2002; Green, 1979; Underwood, 2000a).

That is, one cannot state categorically that the threshold was exceeded due to livestock browsing, as opposed to natural processes acting at that site. In this hypothetical example, the lack of a baseline and having only a single impact site was unavoidable because livestock browsing was occurring before monitoring was initiated and only one mangrove stand, in the resource-use zone, was exposed to browsing by livestock. When a monitoring design is sub-optimal, other tools, such as levels-of-evidence approach, should be used to facilitate inference (Downes et al., 2002; Fabricius & De'ath, 2004; Rouphael et al., 2011).

Variant 2: Difference between extremities of confidence intervals compared with thresholds

Theory

Variant 2 is based on an approach proposed by Walshe and Wintle (2006) who show how an effect size can be quantified by comparing the difference between the extremities of confidence intervals rather than between the means of impact and control treatments. Their approach is useful when a more conservative effect size estimate is preferred. However, assessing uncertainty is not as straightforward as in the previous variant because there is no single confidence interval generated for an effect size.

The Walshe and Wintle (2006) approach is modified here in two ways. First, a management threshold is stated before the start of monitoring. Second, in the absence of site level variability for the impact treatment, the effect size is calculated as the difference between the mean of the impact site and an extremity of the confidence interval for the control treatment. The reason why the confidence interval for the impact site is not used for estimating the effect size in this situation is explained in the section 'Background to Variants'.

Scenario

Variant 2 is based on a scenario where a manager of a marine park is concerned about the effect of resort guests walking on seagrasses on a reef flat in a tourism-use zone. The conservation goal for seagrass meadows is to maintain their structural integrity, defined in part by the density (i.e. number per unit area) of seagrass stems. An environmental awareness campaign at the resort has greatly reduced the number of guests walking on the reef flat. The manager believes the current intensity of walking on the reef flat should not cause seagrass stem density to drop below a level of management concern. The manager recognises that stem density changes seasonally and that a management threshold should be linked to the control condition. The manager proposes that if seagrass stem density at the reef flat (hereafter 'impact site') is less than the control sites by an absolute value of 10 or more stems per m^2, then further investigation would be required. To reiterate, the effect size is measured as the difference between the mean of the impact site and the upper confidence interval of the control treatment. As in the previous example, there was no baseline period and only one impact site.

Figure 4 shows hypothetical data for four consecutive surveys. For each survey, stem density data from the impact site and control sites are collected and their means and 95 per cent confidence intervals plotted. Note that the confidence intervals for the impact site are not shown in Figure 4. Although a confidence interval can be generated for the impact site based on within level variability, this should not be compared directly with the confidence intervals for the control treatment, which are based on site level variability.

The first three surveys suggest that the management threshold has not been exceeded. However, Survey 4 indicates the mean density of stems is less at the impact site and that the difference in the density of seagrass stems between the mean of the impact site and the upper confidence interval of the control treatment is greater than 10 stems. Based on this result the management threshold has been exceeded (Figure 4). As with all examples given in this paper, interpretation focuses on effect sizes, not on whether the means are statistically different.

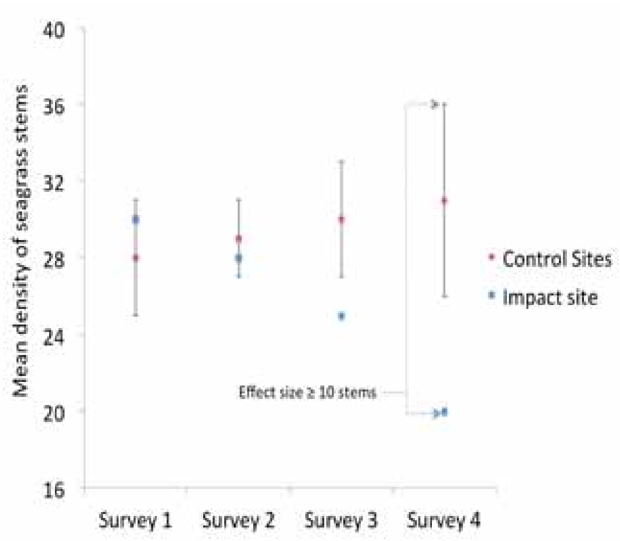


Figure 4: At each survey an effect size is calculated as the difference between the upper 95 per cent confidence interval of the control treatment and the mean for the impact site. In Survey 4, the effect size is greater than 10 stems and thus exceeds the management threshold.

Parameter estimation techniques can be used to summarise social data. Here Dr. Salwa Elhalawani collects data from a fisher to understand fishing intensity and the level of dugong bycatch in Elba National Park © Dr. S Elhalawani

Strengths and weaknesses

When using this variant, it is worth noting a second definition of a confidence interval that states that the central value of a confidence interval is about seven times more plausible than values at the limits (Cumming, 2012). Therefore, using the extremity of a confidence interval to define a parameter of interest, as opposed to the central value, might be considered conservative. Another strength of this approach is that it is straightforward to quantify effect sizes and to compare these with a management threshold.

One weakness of this variant is that a confidence interval is not generated for the effect size, as it was in the previous example. Instead, examining the confidence interval for each mean assesses precision. Therefore, the ability to determine how well the monitoring programme is capable of assessing whether a threshold has been reached is not as clear-cut as in the previous example.

Variant 3: Comparing percentiles between impact and control sites

Theory

A third variant of parameter estimation that can be used to compare data from an impact site with a control site is based on that proposed by ANZECC (2001) and Fox (2001). This approach was proposed for water quality monitoring, but has application for monitoring other variables. This approach is not as straightforward as the previous two variants, but has some advantages when data are counts and highly skewed, and when stakeholders cannot initially agree on a threshold in terms of absolute values.

With this approach, the median, or 50th-percentile (50thP), of the data from an impact site is compared with the 80th-percentile (80thP) of the control treatment. A median is the middle value of a data set, while an 80thP represents a value that partitions a data set into 80 per cent and 20 per cent of all values, respectively. The choice of using a median, rather than a mean, is often desirable when a data set does not conform to a normal distribution or when outliers have a disproportionate influence on a measure of central tendency.

Rouphael and Hanafy (2007) show how the ANZECC (2001) approach can be simplified in order to monitor change in the amount of broken coral at a dive site. Instead of using a 'rolling' percentile, Rouphael and Hanafy (2007) propose estimating the median and the 80thP based only on the most recent survey data. This is unavoidable when baseline data are absent. They also discuss the advantages of this approach for managers of marine parks who lack the technical skills to use more complex statistical approaches. Walshe and Wintle (2006) expand on ANZECC (2001) by recommending that confidence intervals be placed on the median and 80thP. Although ANZECC (2001) describes its approach as a 'process control chart', I refer to the version presented here as parameter estimation because confidence intervals are estimated for the median and 80thP, and because a rolling percentile is not used to calculate the 80thP. Instead, the median and 80thP are calculated and compared based only on the most recent survey data. For reasons given in the previous example, a confidence interval is not shown for the impact site median when it is juxtaposed with the median and confidence interval for the 80thP.

Scenario
In the following scenario, a manager is concerned about a temporary decline in water quality associated with the deepening of a marina adjacent to her marine park. The marine environment adjacent to the marina is zoned general-use, which permits a range of human activities, such as shipping. Given that the general-use zone borders the marina, the manager acknowledges that a temporary decline in water quality near the marina is acceptable provided it does not lead to long-term and widespread environmental damage.

The manager learns that over the next six months, excavators will remove sediment from the marina. This will result in the re-suspension of sediment, leading to turbid water plumes moving down current from the marina to a bay that supports coral assemblages. The manager is concerned that excessive levels of sediment in the water column may lead to an unacceptable level of impact to the assemblages.

The marina authority agrees to limit the frequency of plumes contacting coral assemblages in the bay by controlling the intensity of excavation and the timing in relation to tidal cycles. The marine park manager is still concerned that plumes from the marina will increase turbidity to a point where sediment may lead to coral colony mortality. Consequently, the marina authority instigates water quality monitoring in the bay to ascertain when the amount of sediment in the water column is regularly exceeding background level. However, the manager and the marina authority cannot agree on a threshold expressed in absolute values above background level nor is there a suitable water quality standard because the marine environment is naturally turbid anyway. As a compromise, they decide that if the median (50thP) suspended sediment concentration (SSC) in the bay was above the 80thP of the control sites for three consecutive daily surveys, then they would assess the condition of the coral assemblages directly. The water quality control sites are located well away from the plumes. Figure 5 shows hypothetical data for seven water quality surveys following the start of excavation. It also shows that the median values of SSC at the bay (i.e. the impact site) for the last three surveys are above the 80thP threshold for the control sites. Indeed, since Survey 3, the median SSC for the bay has steadily increased relative to the 80thP, providing additional evidence that the threshold was not exceeded due to a random natural event.

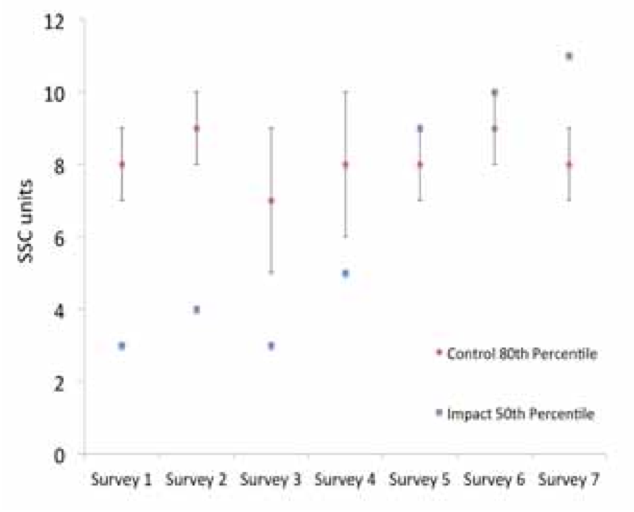


Figure 5: With this approach, the 50thP SSC value from the bay (the impact site) is compared with the 80thP value (±95 per cent confidence interval) from all control sites. The 50thP has exceeded the 80thP in three consecutive surveys (Surveys 5, 6 and 7), which is the trigger for action.

Strengths and weaknesses
An advantage with this approach is its flexibility. For instance, in this example, the 80thP is used as the threshold, but there is no reason why another percentile could not have been chosen. Similarly, a different number of consecutive times the 80thP was exceeded could also have been chosen. This would depend, in part, on how frequently monitoring could be undertaken and on an improved understanding of the relationship between levels of SSC and coral colony mortality.

Another advantage of this approach is that it avoids the need to state a management threshold defined in terms of an absolute value (ANZECC, 2001). This is helpful when water quality standards are unavailable for the area of interest or where stakeholders cannot immediately agree on a threshold defined in terms of absolute values. Other advantages of this approach include the ease of interpretation and the flexibility in terms of statistical assumptions (ANZECC, 2001).

Fox (2001) highlighted one limitation. He warned against assuming that a shift from the 50thP to the 80thP represented an ecologically significant effect. Thus, although a threshold in absolute values need not be defined up-front, at some stage, the manager will need to assess the ecological relevance of this threshold.

CONCLUSION

There is an increasing demand for managers of marine parks to demonstrate the achievement of conservation goals, often defined in terms of quantitative management thresholds. However, analysing monitoring data and comparing these with a

Eye on the Reef diver assessing impact of Tropical Cyclone Ita © Commonwealth of Australia (GBRMPA)

management threshold is not straightforward for laypersons. Some statistical approaches are complicated or may be invalid in the context of marine parks. Further, stakeholders easily misunderstand the outputs of some approaches. Parameter estimation offers a number of advantages for managers of marine parks, but there are few practical examples of how the approach could be applied in this context. In this paper, three variants of parameter estimation are presented. All three variants focus on the size of environmental change that is compared with an a-priori defined management threshold. However, the variants differ in how effect sizes and associated confidence intervals are estimated. The first variant calculates an effect size as the difference between the mean of the impact site and mean of the control sites. A confidence interval for the effect size is also calculated. The second compares the difference between the mean of the impact site and the upper confidence for the mean of the control sites. The degree of uncertainty is ascertained by examining the confidence for the mean of the control treatment. The third variant compares the median value of the impact site with the 80thP of the control sites. For this variant, confidence intervals are also generated for the 80thP. Each variant has its relative strengths and weaknesses that need to be considered carefully prior to adoption.

ACKNOWLEDGEMENTS

The following people are thanked for their informative and helpful suggestions to earlier drafts of this paper: Travis Elsdon, Michael Marnane, Ameer Abdulla, Paul Marshall and Prue Addison. I thank Bruce Mapstone and Bob Clarke for advice in relation to choosing the source of variance when calculating confidence intervals for effect sizes when there is a single impact site but multiple control sites. Thanks also to two anonymous reviewers whose comments also greatly improved the paper.

ABOUT THE AUTHOR

Anthony ('Tony') Badih Rouphael is a marine scientist based in Perth, Western Australia. Tony has worked in conservation management in Australia and the Red Sea

region. He has been privileged to work with conservation agencies in Yemen, Saudi Arabia and Egypt. His research interests include recreation ecology, monitoring approaches for protected areas and evaluating human impacts on marine and coastal ecosystems.

REFERENCES

Addison, P. F. E., Flander, L. B. and Cook, C. N. (2015a). Are we missing the boat? Current uses of long-term biological monitoring data in the evaluation and management of marine protected areas. Journal of Environmental Management, 149: 148–156. DOI: 10.1016/j.jenvman.2014.10.023

Addison, P. F. E., De Bie, K. and Rumpff, L. (2015b). Setting conservation management thresholds using a novel participatory modeling approach: Conservation management thresholds. Conservation Biology, 29(5): 1411–1422. DOI: 10.1111/cobi.12544

Altman, D.G., Machin, D., Bryant, T.N. and Gardner, M.J. (2000). Statistics with Confidence: Confidence Intervals and Statistical Guidelines. Second Edition. London: British Medical Journal.

Andrew, N. E. and Mapstone, B. D. (1987). Sampling and the description of spatial pattern in marine ecology. Oceanography and Marine Biology Annual Review, 25: 39–90.

ANZECC (2001). Australian guidelines for water quality monitoring and reporting: National water quality management strategy paper No. 7. Australian and New Zealand Environment and Conservation Council and Agriculture Resource Management Council of Australia and New Zealand. Canberra: Environment Australia.

Benedetti-Cecchi, L. (2001). Beyond BACI: Optimization of environmental sampling designs through monitoring and simulations. Ecological Applications, 13: 783–799. doi.org/10.1890/1051-0761(2001)011[0783:BBOOES]2.0.CO;2

Buckley, R., Robinson, J., Carmody, J. and King, N. (2008). Monitoring for management of conservation and recreation in Australian protected areas. Biodiversity and Conservation, 17: 3589–3606. doi.org/10.1007/s10531-008-9448-7

Burgman, M. (2005). Risks and Decisions for Conservation and Environmental Management. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511614279

Burgman, M., Lowell, K., Woodgate, P., Jones, S., Richards, G. and Addison, P. (2012). An endpoint hierarchy and process control charts for ecological monitoring. In: D. Lindenmayer and P. Gibbons (eds.) Biodiversity Monitoring in Australia, pp. 71–78. Australia: CSIRO Publishing.

Cohen, J. (1990). Things I have learned (so far). American Psychologist, 45: 1304-1312. http://dx.doi.org/10.1037/0003-066X.45.12.1304

Connell J. H., Hughes T. P., Wallace C. C., Tanner J. E., Harms K. E. and Kerr A. M. (2004). A long-term study of competition and diversity of corals. Ecological Monographs, 74: 179–210.

Cumming, G. (2012). Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis. New York: Routledge.

Cumming, G. and Calin-Jageman, R. (2017). Introduction to the New Statistics: Estimation, Open Science and Beyond. New York: Routledge.

Day J. C., Laffoley, D. and Zischka, K. (2015). Marine protected area management. In: G. L. Worboys, M. Lockwood, A. Kothari, S. Feary and I. Pulsford (eds.) Protected Area Governance and Management, pp. 609–650. Canberra: ANU Press.

Di Stefano, J. (2004). A confidence interval approach to data analysis. Forest Ecology and Management, 187: 173–183. doi:10.1016/S0378-1127(03)00331-1

Di Stefano, J., Fidler, F. and Cumming, G. (2005). Effect size estimates and confidence intervals: An alternative focus for the presentation and interpretation of ecological data. In: A. R. Burk (ed.) New Trends in Ecological Research, pp. 71–102. New York, USA: Nova Science Publishing.

Downes, B. J., Barmuta, L. A., Fairweather, P. G., Faith, D. P., Keough, M. J., Lake, P. S., et al. (2002). Monitoring ecological impacts: Concepts and practice in flowing waters. Cambridge: Cambridge University Press. ISBN 0511049560

Fabricius, K. E. and De'ath, G. (2004). Identifying ecological change and its causes: A case study on coral reefs. Ecological Applications, 14: 1448–1465. doi.org/10.1890/03-5320

Fowler, J., Cohen, L. and Jarvis, P. (1998). Practical Statistics for Field Biology. Second edition. Chichester, UK: Wiley.

Fox, D. (2001). The new ANZECC/ARMCANZ environmental water quality guidelines. Journal of Australian Water Association, 28: 46–50.

Gardner, M. J. and Altman, D. G. (2000). Confidence intervals rather than p values. In: D. G. Altman, D. Machin, T. N. Bryant and M. J. Gardner (eds.) Statistics with Confidence. Second edition, pp. 15-27. UK: BMJ Books. ISBN: 978-0-727-91375-3

Green, R. H. (1979). Sampling design and statistical methods for environmental biologists. New York, USA: John Wiley and Sons. ISBN: 978-0-471-03901-3

Hatcher B. G., Johannes R. E. and Robertson A. I. (1989). Review of research relevant to the conservation of shallow tropical marine ecosystems. Oceanography and Marine Biology: An Annual Review, 27: 337–414.

Hockings, M., Stolton, S., Leverington, F., Dudley, N. and Courrau, J. (2006). Evaluating effectiveness: A framework for assessing management effectiveness of protected areas. Second edition. Gland, Switzerland/Cambridge, UK: IUCN.

Hughes, T. P. (1994). Catastrophes, phase shifts, and large-scale degradation of a Caribbean coral reef. Science, 265: 1547–1551. DOI: 10.1126/science.265.5178.1547

Hughes, T. P., Baird, A. H., Dinsdale, E. M., Moltschaniwskyj, N. A., Pratchett, M. S., Tanner, J. E. and Willis, B. L. (1999). Patterns of recruitment and abundance of corals along the Great Barrier Reef. Nature, 397: 59–63. doi.org/10.1038/16237

Mapstone, B. D. (1995). Scalable decision rules for environmental impact studies: Effect size, type I, and type II errors. Ecological Applications, 5: 401–410. doi.org/10.2307/1942031

Marion, J. L. and Rogers, C. S. (1994). The applicability of terrestrial visitor impact management strategies to the protection of coral reefs. Ocean and Coastal Management, 22: 153–163. doi.org/10.1016/0964-5691(94)90017

Masson, M. E. J. and Loftus, G. F. (2003). Using confidence intervals for graphically based data interpretation. Canadian Journal of Experimental Psychology, 57: 203–220.

Morrison, L. (2008). The use of control charts to interpret environmental monitoring data. Natural Areas Journal, 28: 66–73. doi.org/10.3375/0885-8608(2008)28[66:TUOCCT]2.0.CO;2

Pomeroy, R. S., Parks, J. E. and Watson, L. M. (2004). How is your MPA? A guidebook of natural and social indicators for evaluating marine protected area management effectiveness. Gland, Switzerland/Cambridge, UK: IUCN.

Rouphael, A. B. and Hanafy, M. (2007). An alternative management framework to limit the impact of SCUBA divers on coral assemblages. Journal of Sustainable Tourism, 15: 91–103. doi.org/10.2167/jost611.0

Rouphael, A. B., Abdulla, A. and Said, Y. (2011). A framework for practical and rigorous impact monitoring by field managers of marine protected areas. Environmental Monitoring and Assessment, 180: 557–572. doi: 10.1007/s10661-010-1805-9

Stewart-Oaten, A. (1996). Goals in environmental monitoring. In: R. J. Schmitt, and C. W. Osenberg (eds.) Detecting ecological impacts: Concepts and applications in coastal habitats, pp. 17–27. San Diego: Academic Press. DOI: 10.1016/B978-012627255-0/50004-5

Tryon, W. W. (2001). Evaluating statistical difference, equivalence, and indeterminacy using inferential confidence intervals: An integrated alternative method of conducting null hypothesis statistical tests. Psychological Methods, 6: 371–386. DOT:10.1037//1082-989X.6.4.371

Underwood, A. J. (1997). Experiments in ecology: Their logical design and interpretation using analysis of variance. Cambridge: Cambridge University Press. ISBN 9781107088917

Underwood, A. J. (2000a). Importance of experimental design in detecting and measuring stresses in marine populations. Journal of Aquatic Ecosystem Stress and Recovery, 7: 3–24. doi.org/10.1023/A:1009983229076

Underwood, A. J. (2000b). Trying to detect impacts in marine habitats: Comparisons with suitable reference areas. In: T. Sparks (ed.) Statistics in Ecotoxicology, pp. 279–308. Chichester, UK: John Wiley and Sons.

Walshe, T. and Wintle, B. (2006). Guidelines for communicating performance against standards in forest management. Prepared for the Forest and Wood Products Research and Development Cooperation.

Walshe, T., Wintle, B., Fidler, F. and Burgman, M. (2007). Use of confidence intervals to demonstrate performance against forest management standards. Forest Ecology and Management, 247: 237–245. doi:10.1016/j.foreco.2007.04.048

Warwick R. M. and Clarke K. R. (1993) Increased variability as a symptom of stress in marine communities. Journal of Experimental Marine Biology and Ecology, 172: 215–226. doi.org/10.1016/0022-0981(93)90098-9

Young, M. A. L., Foale, S. and Bellwood, D. R. (2014). Impacts of recreational fishing in Australia: Historical declines, self-regulation and evidence of an early warning system. Environmental Conservation: 1–7. doi: 10.1017/S0376892914000046

RESUMEN

Cada vez es más importante que los administradores de los parques marinos demuestren cuantitativamente el logro de sus objetivos de conservación. El monitoreo es una herramienta de gran utilidad para ello. Un componente del monitoreo que resulta complejo para los administradores es el tratamiento estadístico de los datos de monitoreo. Los enfoques comúnmente utilizados, tales como las pruebas de hipótesis nulas, son conceptualmente desafiantes y operacionalmente complejas, lo que puede llevar a conclusiones erróneas y malas decisiones. Un enfoque más directo es la estimación de parámetros con intervalos de confianza. La estimación de parámetros se centra en la estimación del tamaño del cambio o diferencia (un "tamaño de efecto") en una variable de respuesta y la comparación de esta con un tamaño de efecto predefinido denominado umbral de gestión. Los intervalos de confianza indican el nivel de precisión en las estimaciones de los cambios, lo que se traduce en conclusiones más equilibradas. La estimación de parámetros también es propicia para la representación gráfica, que puede facilitar la interpretación y la comunicación para un público no científico. En este artículo, se demuestran tres ejemplos de estimación de parámetros y se analizan sus fortalezas y debilidades relativas, con lo que se espera alentar a los administradores a adoptar enfoques estadísticos que les permitan cuantificar el cambio ambiental de una manera que contribuya a conclusiones defendibles y a facilitar la toma de decisiones oportunas y de fácil comprensión para los interesados.

RÉSUMÉ

Il y a une exigence accrue auprès des gestionnaires de parcs marins pour qu'ils fournissent une évaluation quantitative de la réalisation de leurs objectifs de conservation. La surveillance est un outil qui peut leur venir en aide dans ce processus. Cependant, le traitement statistique des données de surveillance constitue l'une des difficultés de cette méthode. Les approches couramment utilisées, telles les tests d'hypothèse nulle, sont conceptuellement exigeantes et complexes sur le plan opérationnel, et peuvent ainsi mener à des conclusions erronées et à de mauvaises décisions. L'estimation des paramètres avec des intervalles de confiance constitue une approche plus directe. L'estimation des paramètres consiste à estimer la taille du changement ou de la différence (une «taille d'effet») dans une variable-réponse, puis à la comparer avec une taille d'effet prédéfinie appelée seuil de gestion. Les intervalles de confiance indiquent le niveau de précision des estimations de changement, ce qui permet d'obtenir des conclusions plus équilibrées. L'estimation des paramètres est également adaptée à la représentation graphique, ce qui peut faciliter l'interprétation et la communication aux non-scientifiques. Dans cet article, je présente trois exemples d'estimation des paramètres et passe en revue leurs forces et faiblesses relatives. En présentant ces exemples, j'espère encourager les gestionnaires à adopter des approches statistiques qui leur permettent de quantifier les changements environnementaux de manière à soutenir efficacement leurs conclusions, à faciliter la prise de décision en temps opportun et à être compris par les intervenants.