A guideline to monitoring populations

Version 1.0

This guide was written by Terry Greene in 2012.

Contents

Inventory and monitoring toolbox

DOCDM-870579

Department of Conservation
Te Papa Atawhai

Introduction

Why count plants and animals?

We need to count plants and animals for three main reasons:

- To understand what we have got in our area of interest
- To discover whether there has been any change in population size and, if so, what processes were driving that change
- Determine the effectiveness of management actions and whether any changes to those actions affected population size

Usually we count and measure the organisms or features that are of conservation interest (e.g. the abundance of a threatened plant, insect, bird or habitat feature). This is called direct monitoring. We can also monitor the threats (perceived or actual) to the object of interest, (e.g. the level of pollution or number or density of weeds and pests) or indicators of the presence of an organism (e.g. scat or burrows). This is indirect monitoring.

Without effective and efficient monitoring programmes, evaluation of the success or otherwise of management actions becomes extremely difficult and potentially misleading. It is critical, therefore, that the outcomes of management actions are able to be distinguished from the background noise or fluctuations (i.e. natural variability) found within all biological systems. For example, managers may want to know whether fencing to exclude stock, or the removal of predators, made a difference to the conservation outcome. They may well ask:

- Was the intervention effective?
- Did the management action result in an increase (or decrease) in the abundance of the plant or animal of interest?
- Was the desired response able to be distinguished from other factors affecting population abundance of the 'feature/organism' of interest?

Monitoring can also provide an early warning of threats to population abundance, such as the spread of an invasive weed that will compete with native plant species. An example of monitoring as an early warning system is surveillance monitoring used to assess rate of conifer spread within tussockland ecosystems (Raal et al. 2005).

The importance of monitoring is recognised in national and international legislation and treaties. Within New Zealand, government departments such as DOC must report annually to Parliament and Treasury about the effectiveness and efficiency of management actions. Monitoring is sometimes a condition of Resource Management Act consents. Monitoring is often an explicit requirement within international treaties and conventions. For example, the Convention on Biological Diversity requires that contracting parties (of which New Zealand is one) shall:

- Identify and 'monitor, through sampling and other techniques, the components of biological diversity identified [ecosystems and habitats, species and communities, and significant genetic resources]', and

- Identify and monitor the processes 'likely to have significant adverse impacts on the conservation of biological diversity', as well as
- Collect and maintain the data in good order.
  (Adapted from Article 7 and Annex 1, Secretariat of the Convention on Biological Diversity 2005)

During development of the Natural Heritage Management Programme, DOC recognised the need for 'a consistent set of standards for the design of freshwater and terrestrial monitoring programs, collection and handling of data and reporting on natural heritage outcomes' (Ross 2002). Improved and standardised monitoring practices, as promoted in this toolbox, will enable DOC and other interested parties to:

- Better understand what it is we do in terms of on-ground actions, by ensuring that monitoring objectives are clear, unambiguous and consistent with management and conservation needs.
- Understand how efficient and effective we are at meeting conservation objectives and outcomes, by accounting for variation with appropriate sampling design and data collection standards.
- Understand and improve the outcomes of our management on variables of interest, particularly the national status of key indicators, through the use of appropriate experimental design and predictive models.
- Enable comparison of data across a range of temporal and spatial scales (areas, conservancies, regions and nationally) through the use of consistent data collection, statistical analysis and reporting methods.
- Measure biodiversity status more effectively.
- Meet obligations under national legislation and policies, as well as those arising from international agreements.

Approaches to monitoring

'Targeted monitoring is defined by its integration into conservation practice, with monitoring design and implementation based on *a priori* hypotheses and associated [predictive] models of system responses to management' (Nichols & Williams 2006, p. 668).

In other words, there is a defined process for informed decision making that includes the development of explicit objectives, listing of potential management actions, a means for measuring confidence in predictive models, and a monitoring programme that is centred on providing parameter estimates against which competing models can be tested.

Targeted monitoring tends to be done to provide explicit information about past or current management actions. Sometimes, however, information is needed about an organism or feature before management begins. In such cases, targeted monitoring can be used to measure specific attributes (e.g. density or survival) of an organism or feature of interest. That information is then used to formulate hypotheses and testable, predictive models.

The second major approach to monitoring is general monitoring (Nichols & Williams 2006). Although this approach is in common use, it has often been criticised:

- At times, this type of monitoring is poorly targeted. It becomes a stand-alone activity, with little apparent purpose.
- Too often, a general monitoring programme fails to develop any explanatory hypotheses and associated models.
- It can result in continuous, but unnecessary, collection of baseline information.
- It often misuses statistical hypothesis testing. Weaknesses are compounded by inappropriate experimental and sampling designs.
- General monitoring can delay management intervention because of the time required to collect enough data to show a 'significant' decline.
- Too often it focuses on the cause of decline rather than the conservation remedy.
  (Nichols & Williams 2006)

This does not mean that general monitoring should be dismissed entirely. It can provide information useful for conservation. Indeed, in some situations it may be the only available information. General monitoring can be a cost-effective way to develop hypotheses and identify useful covariates to include in a more targeted monitoring programme. However, the value of information collected in this way depends on the original objectives of the monitoring programme. For example, general monitoring of a forest bird community is unlikely to provide useful information about species that are neither abundant nor widespread—the species that are usually in most need of conservation management (Nichols & Williams 2006). In contrast, a national bird survey scheme set up to look at large-scale trends in distribution and abundance over long timeframes (particularly when there is little current information) may well provide useful information on changing patterns or emerging downward trends. Once identified, these species trends can be investigated using a more targeted approach.

Whatever the approach, good sampling design allows data to be interpreted in a meaningful way. It is critical to effective and efficient monitoring.

What is the difference between inventory and monitoring?

Inventory and monitoring are often considered to be the same thing, with little distinction drawn between them (Morrison et al. 2001). Two things reinforce this view. Often, the type of information collected is the same, and the methods used to collect it may also be similar. However, there are fundamental differences, as summarised below:

An *inventory* is a stock take at a given point in time. It does not imply any future remeasurement. Usually the intent is to compile comprehensive information on the current state of an organism, such as the presence or absence of a species or group of species or ecosystem component. Examples are species lists under the New Zealand Bird Atlas Scheme, and high-country tenure review surveys. A major assumption of the method is that all significant species are detected.

Monitoring focuses on system dynamics (changes in state). It usually compares measurements at different places and times. Remeasurement is a key part of a monitoring programme. In population monitoring, the intent is usually to detect a trend and the rate at which change is occurring (i.e. whether a population is stable, decreasing or increasing, and whether that change is slowing or accelerating). The target may be a population of a single species, populations of numbers of species, or composition of selected ecosystems. The New Zealand Forest Service permanent plot networks which have formed the core of the National Vegetation Survey are a classic example of vegetation community monitoring (Lee et al. 2005). Others are the long-running seal-rookery and seabird monitoring sites, as documented by Bradshaw et al. (2000) and Moore (2004), and flowering and seed fall monitoring (Schauber et al. 2002; Richardson et al. 2005).

The principal differences are largely a function of time, the processes needed to meet study objectives, and the different study designs needed to assess an organism's (e.g. species') state (for inventory) as opposed to dynamics (for monitoring). A more detailed discussion of the differences between inventory and monitoring, including a list of generalised monitoring objectives, can be found in Lee et al. (2005).

DOC often distinguishes 'outcome' and 'result' monitoring, usually in relation to pest control and management. Result monitoring directly measures the results of management intervention on animal or plant pest populations, while outcome monitoring measures the benefits of management actions to the wider habitat, community or ecosystem. An example would be possum control in a particular reserve. Result monitoring would monitor the success of possum control by assessing possum populations before and after control whereas outcome monitoring would measure changes in the health of canopy tree species and/or native bird populations.

Measures used in monitoring

Demographic measures

These include assessment of survival, mortality, productivity and sex ratio. Such measures are often focused on a small number of individuals, then extrapolated to a wider population. They provide detailed information on change within a defined area, but are often costly to obtain. They can be used to develop predictive tools, such as population models and population viability estimates. Demographic measures are also commonly used as direct and immediate measures of the success of management actions (e.g. nesting success of bird species, survival through aerial 1080 poisoning operations, and changes in a population's sex ratio following extensive and prolonged pest control).

Population counting or numeric measures

These measures include census, indices, presence/absence, density and abundance. They are often applied to a much larger population of interest (compared with those used in demographic studies) and can provide an estimate (or index) of the overall population size and/or distribution.

Using these measures

These two types of monitoring measures should not be viewed as mutually exclusive. Depending on the monitoring objectives, a combination of the two is desirable; if not at the same time, then at least at intervals over the life of the project. This is particularly relevant, for example, when sexual dimorphism in a species is slight or non-existent, but the survival of one sex is thought to be distorted in some way (perhaps through dissimilar vulnerability to predation or disease). Simple numeric counts of such species are unlikely to reveal these distortions in the underlying population structure (other than perhaps identifying a trend of slow decline) until the population collapses suddenly following the demise of most of one of the sexes. In this case, periodic measurement of the sex ratio and sex-specific survival should be added to a more general population census programme. Good examples from New Zealand of uneven survival rates and population-scale impacts include those recorded for kokako (Innes et al. 1999) and kaka (Greene & Fraser 1998) where numbers of breeding females were severely reduced by predation. In the case of *Atriplex hollowayi* (Holloway's crystalwort)—an annual, succulent herb found on sandy beaches—a count of individual plants has been carried out since 1990 and has demonstrated significant year to year variation in numbers (de Lange et al. 2000). The processes underlying these fluctuations remain unclear, hampering interpretation and determination of appropriate conservation actions (de Lange et al. 2000). Incorporation of demographic measures and measurement of other factors that affect demographic parameters might help explain these fluctuations.

Clearly then, demographic and numeric measures of populations can be used independently and together to provide compelling evidence of the success or otherwise of management actions. This is especially so when done within an informed decision-making process, using well-defined objectives, robust sampling design, appropriate data collection methods and predictive models based on detailed hypotheses (Nichols & Williams 2006). Integration of both data-streams can only further improve our ability to make informed conservation and management decisions (Conroy & Noon 1996).

Importance of a project plan and technical specifications

A formal planning and approval process will ensure that realistic and appropriate objectives are established, an appropriate inventory or monitoring sampling design is developed and implemented, tasks and roles are identified and allocated, specifications are established and anticipated outputs set. A written project plan is also essential if you are to meet minimum technical standards. Ideally, the plan will be peer reviewed and include a timetable for periodic audit, developed prior to project approval. Such review processes should also be an integrated and compulsory component of all inventory and monitoring programmes.

DOC staff must complete a 'Standard inventory and monitoring project plan' (docdm-146272). To ensure sound planning occurs, approval is required from a line manager and appropriate technical reviewer before any new DOC programme begins.

A good technical specification document should provide enough instructions that practitioners will be able to 'go out and do it', without having to return for clarification or make their own decisions on how to apply a method as they go. Clear specifications will leave little room for individual interpretation, which ensures data are collected consistently among practitioners and from one sampling period to the next.

Technical specifications should therefore provide full and clear details for all aspects of a monitoring or inventory programme, including coverage of all the steps in the monitoring framework (see Fig.

1), information on the methodology being applied, plus details of the methods, data storage and analysis. Technical specifications should also describe the resources needed to undertake the chosen method, and include a checklist of equipment required, information on the approximate time it will take to reach the site and collect the data, recording tools required and skills needed by the members of the monitoring team.

Design and implementation framework

Designing and implementing a monitoring plan

A well-designed programme is essential for a successful monitoring campaign. Approaches to programme design are discussed here. Manipulative experimental designs and observational approaches are compared.

Once the need to monitor populations has been identified, substantial thought is required about the whole monitoring framework. This section presents a structured process to guide monitoring programme design, covering all the steps illustrated in Fig. 1. This section ends with a discussion on common failings seen in monitoring programmes.

Programme design is crucial if conservation managers are to obtain robust population abundance estimates and detect changes in number over time. Good sampling design, in particular, will maximise the value gained from the monitoring effort. There are a number of approaches to sampling designs: manipulative experiments, constrained designs and natural experiments, through to observational approaches. The choice of design will be the primary determinant of inferential strength—or how much confidence we can have in our conclusions (Nichols & Williams 2006). The two ends of the spectrum are:

1. *Manipulative experimental designs* provide strong inference, but are often under-utilised in conservation management because of perceived or real practical difficulties, such as possible higher cost and greater labour requirements, disturbance to threatened species, and limited options for creation of a control (non-treatment) group or site.
2. *Observational approaches* have much weaker inferential strength because of the greater number of potential explanatory hypotheses. Despite this, observational approaches can still play a significant role, provided the monitoring objectives are appropriate. Observational approaches can be set up with or without treatments. Passive (e.g. non-targeted and observation-based) monitoring of *in situ* change also has its place, but to ensure value for the monitoring effort expended, the programme has to be designed so that any detected change can be interpreted correctly.

Choosing a monitoring approach

Choice of monitoring method should also take account of how widespread and abundant the species of interest is to begin with (assuming this is known), the ease with which the organisms can be detected and counted, resources (funds, labour, equipment costs, time, etc.), and the inherent bias and precision of the various candidate methods. If plants or animals have to be disturbed, captured or handled in any way, appropriate permits and the relevant ethics approvals (particularly for animals) must be granted prior to commencement of a monitoring programme.

Although considerable effort has been put into developing methods that are able to extract useful information from monitoring data (Williams et al. 2002), there is no theoretical framework available to help managers choose the best monitoring design or data collection regime for a given situation and budget (Joseph et al. 2006). These choices are further complicated because the method with the least bias is unlikely to be the one with the most precision. So, monitoring design is not simple, but the framework set out here will clarify each step of the process.

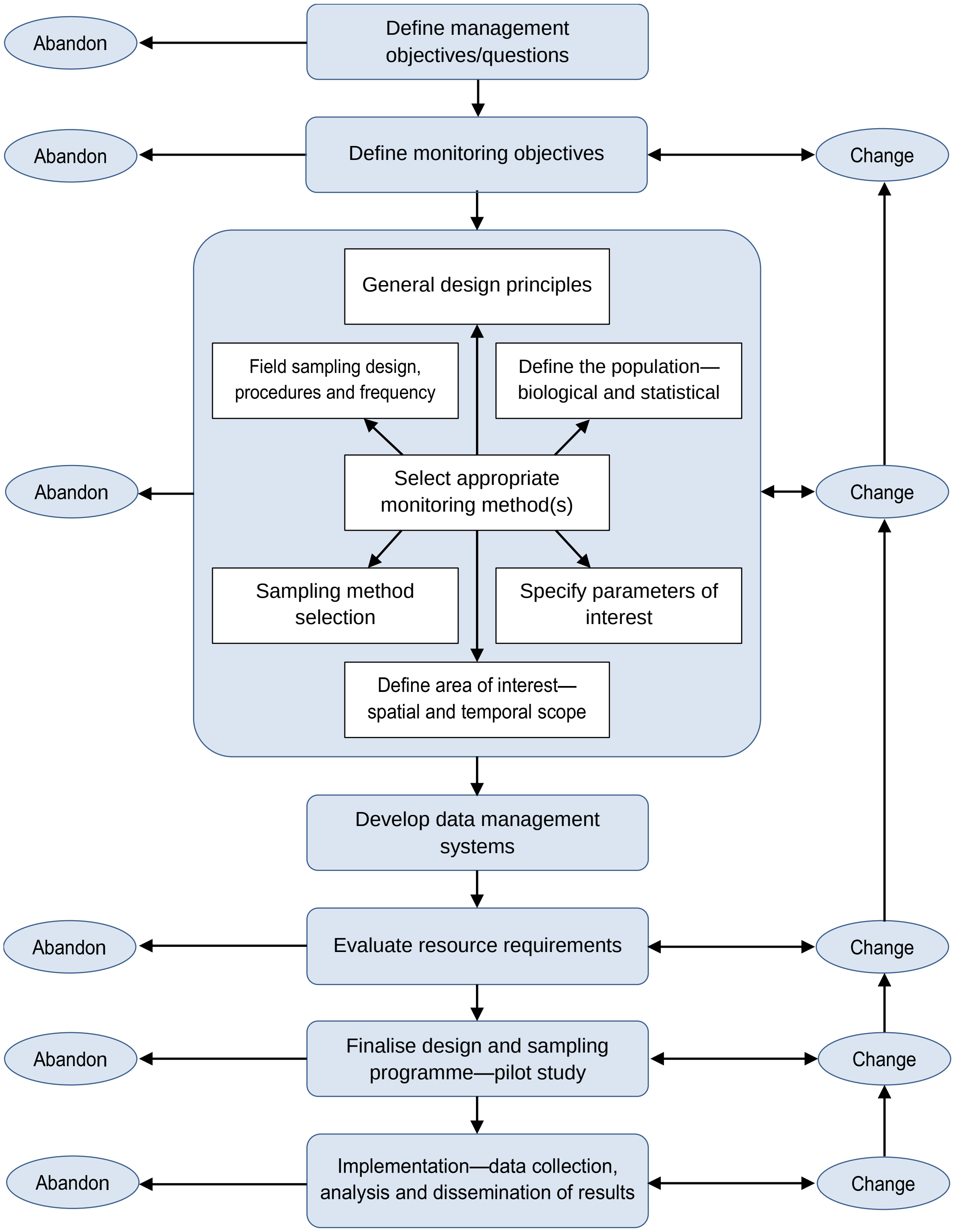

Figure 1. A recommended monitoring framework.

Define management objectives and questions

Precise definition of each management objective and question is vital. Such statements provide fundamental information on which to base a monitoring programme. Management objectives/questions and monitoring objectives are different but related subjects. A monitoring programme may be designed to answer all or just part of a management objective/question. For example, a management objective might be ‘Maintain forest structure and composition’; a management question might be ‘Does the goat control programme at Site A effectively protect the forest by permitting regeneration?’ The corresponding monitoring objective might be ‘Determine whether goat-palatable species are regenerating to the same extent in control and non-control areas of forest’. Another example of a management objective might be ‘To ensure protection of threatened species at priority sites in beech forests in the South Island’. The corresponding monitoring objectives might be ‘Determine the effectiveness of a broad-scale control regime in reducing possum abundance to less than 5% RTC’; ‘Determine the effectiveness of the possum control regime on mistletoe abundance’.

Questions for managers

When designing a monitoring programme, managers should also think carefully about specific uses to be made of the data, for example:

- What are the links between the management objectives and the proposed monitoring programme?
- Who will be involved in interpreting the data and communicating the results and conclusions through to decision makers?
- Are there adequate collaborative links between researchers who are designing monitoring approaches, practitioners, and the policy makers? (See Field et al. 2007.)
- Are the mechanisms in place for the monitoring outcomes to inform subsequent management actions?

Define monitoring objectives

The design of effective monitoring programmes for any species requires consideration of management objectives, how monitoring might be implemented in relation to those objectives, and how the results might be used in decision making (including when, to what effect and by whom). For example, the results might feed into testing of a predictive model that shows how a biological system might respond to management actions. Some monitoring programmes will contribute directly to decision making in the near future, especially if their timetables align with strategic planning timeframes. Other programmes will also contribute directly, but not until long-term records have been built up.

Monitoring objectives need to be *worthwhile*, *specific*, *unambiguous*, *realistic* and *measurable*. For example, is the objective to estimate total abundance or to estimate abundance in different habitats or parts of a locality? Objective development should be treated as an iterative process. Significant deviations from original ideas may be required to accommodate the practicalities of resource limitations or requirements of a certain design or method. However, it is essential to assess the effect of any changes on inferential strength (and thus potential value) of a monitoring programme.

Monitoring objectives generally fall into three main categories (Lee et al. 2005), as listed below (with typical questions noted for each):

1. Monitoring for changes in system state and integrity.
   - Are things changing and to what extent?
   - What is the system's state?
   - What timeframe are we interested in?

2. Monitoring for management action.
   - When should we intervene?
   - What might we need to do?
   - Have we been successful?
   - How can we do better?
   - Can we predict what the most appropriate management action is?

3. Monitoring for fundamental understanding (research monitoring).
   - Do we understand what is going on?
   - How might we develop scenarios or predict the future?

Select appropriate monitoring methods

Monitoring methods need to align with monitoring objectives, i.e. given an objective, what sort of results are necessary to answer it and how should the data be obtained? Thus, the more explicit the objectives, the easier the task of selecting appropriate methodologies. However, complexity (single or multi-species), validity of assumptions, variance, power, and cost should also be used as filters against which objectives, potential sampling methods and monitoring programme designs can be evaluated, then accepted, changed or abandoned.

General design principles

Monitoring programme designs must address two major sources of variation and uncertainty common to all species counts: (1) spatial and temporal variation and (2) detectability.

Spatial and temporal variation may mean investigators are unable to apply monitoring methods over the entire area or timescale of interest or relevance because of the resulting scale, cost, logistical limitations, sampling constraints and species characteristics, etc. In such situations, sample units or plots must be selected from the entire area of interest (which includes the target population).The area needs to be sampled in a way that permits inferences to be drawn about the entire area, e.g. by using random, systematic, or stratified sampling (see 'Concepts in survey design: bias and precision' and 'Probability sampling' sections below).

Temporal variation in abundance within (e.g. monthly or seasonal variation) and between years can obscure more subtle changes (such as management impacts). Studies done in a particular year or season can be unrepresentative of the population of interest. If the objective of the study is to derive unbiased point estimates or estimates of trend in population abundance, then the study must be conducted over an appropriate time period to account for this variation (e.g. response to masting cycles). That, in itself, requires a good knowledge of the target species and relevant environmental drivers.

Detectability refers to an almost universal problem in animal population monitoring (compared with plant monitoring). Very often, monitoring methods are unable to detect all animals present, even within the plots selected as part of the sampling frame (i.e. within the Mainland Island, Operation Ark site, or catchment, etc.). A well-designed monitoring programme will incorporate methods for estimating or removing effects of variable detectability. Estimated changes in abundance and density (particularly for animal populations) will therefore reflect true changes rather than differences in detectability. However, if it is not possible to account for these effects, the impact of such methodological bias, and the associated reduction in ability to identify trends, must be clearly acknowledged.

Define the population—biological and statistical

There are two definitions of a population; one statistical, one biological. Both need to be considered when designing a sampling programme. The statistical definition underlies all inference within inventory and monitoring programmes and is, therefore, more general. A *target population* includes all sampling or experimental units about which we would like to draw an inference. If some section of the target population is unable to be sampled (e.g. because of field methods or animal behaviour), the subset that is sampled is referred to as the *sampled population*. It is from this subset that a *representative sample* is taken. Provided the sample is collected in an appropriate manner, inferences can then be made about the sampled population. Any extension of inference to the wider target population assumes that the sampled population is representative of the target population (Morrison et al. 2001).

Simple random sampling, stratified sampling, systematic sampling (and less commonly cluster sampling and adaptive sampling) are all methods that can be used to ensure a sample is representative of a target population, vegetation community or habitat type. Subdivision of the target population into strata (homogeneous subgroups within the sampling frame, e.g. vegetation communities) can be used to improve the sampling efficiency, particularly if it is known that strata are related to variations in distribution, abundance and density. This approach ensures that sampling is spread over the entire area of interest rather than being clumped in non-representative areas that may occur by chance. It increases the precision of estimates compared with those derived from small random samples.

A target population for inventory and monitoring programmes can include a variety of biological entities or groupings (e.g. specific sexes, ages, cohorts, tagged groups or individuals). It is therefore critical to define what this entity is, in a biological sense, if we are to have any confidence in drawing inferences about the entire population of interest. This is particularly so if there is a chance that some part of the target population might be ignored. For example, if individuals are counted in only one habitat fragment within a metapopulation (spatially separated populations of same species which interact at some level), it would be unwise to assume the trends found in that one habitat fragment apply to the entire population. This limitation would be particularly pronounced if demographic parameters (such as sex ratio, age structure, productivity, etc.) or predator status, etc. varied markedly between habitats. Ideally, the sampling design should seek to match both the sampled population and the target population regardless of the defined biological grouping being considered, thus ensuring the inference will be valid (Morrison et al. 2001).

Specify parameters of interest

As with selection of methods, the selection of appropriate parameters or indicators to monitor needs to be closely linked with programme objectives. For example, if the objective is to measure change in density of a population or community of plants or animals following pest control, then certain monitoring methods will be more suitable than others.

Explicitly define the area of interest—spatial and temporal scope

It is essential to define the area of interest in terms of space and time, before initiating a monitoring programme. It is then possible to make definitive statements on how widely results can be extrapolated. In other words, are the results only applicable to a limited geographic area or can they be appropriately applied to other areas? Consider the following points in light of the programme's objectives:

- Identify the geographic locations where the target species or population of interest occurs.
- Assess which of these locations are subject to threats that the monitoring objectives are endeavouring to address.
- Evaluate whether the threatening processes are continuous, periodic, threshold dependent or of some other pattern, and identify the time scales over which these risks are apparent (e.g. seasonal, annual or episodic).
- Identify the spatial and temporal scales over which the target population varies, within and between certain areas (e.g. distribution and nature of metapopulations, dispersal, immigration, etc.).
- For each objective, determine whether definitive spatial and temporal boundaries can be established and the extent to which the sampling frame can be defined.

These points should all be summarised into a succinct, unambiguous and comprehensive explanation of the sampling frame (Miller & Allan 2002), i.e. make a clear statement of the monitoring programme's scope.

Sampling method selection

The taxa-specific sections of the toolbox include a number of simple decision tables and diagrams. These will assist selection of the most suitable population sampling method and they provide guidance on cost and achievability. These keys are only intended to be guides to the major categories of sampling methods. There is no standardised, prescriptive approach to method selection, no definitive list of monitoring protocols, and no single correct method, but clearly some methods are more suitable than others for a given objective, habitat or species.

Rather, the aim is to provide a variety of potential population monitoring pathways. These need to be assessed iteratively, taking account of the monitoring programme's objectives, inherent assumptions of each method, design ramifications of implementing a particular method, and biology of the species of interest. National monitoring protocols exist in some instances for reasons of national consistency, e.g. Residual Trap Catch (RTC), and these should be adhered to. Where standard protocols do not exist, the principal inventory and monitoring options are provided, along with appropriate design, data collection and analysis specifications to ensure consistency and comparability.

Once objectives, population of interest, appropriate measurement parameters and the spatial and temporal areas of interest have been defined, and issues regarding spatial variation and detectability are considered, the keys and decision tables can be used to select an appropriate monitoring method. Following selection of a potential method, its inherent assumptions—and all methods have them—should be examined rigorously to see whether the method is realistic when applied to a specific situation. If the assumptions appear sensible and can be met, development of a suitable field sampling regime can then proceed.

Field sampling design, procedures and frequency

Selection of a monitoring method has direct consequences for the design of a field sampling programme. For monitoring of plant populations, the plot size, sample size, design, layout, location, spatial scale and frequency of remeasurement all need to be appropriate to size, longevity and growth form of the species being monitored. For example, for long-lived plants such as the red-flowered mistletoe (*Peraxilla tetrapetala*) it may be more useful to measure reproductive effort (e.g. the number of flowers or fruit produced per year) or the reproductive success of the species (e.g. number of seedlings) rather than size, whereas for mat plants such as the pygmy button daisy (*Leptinella nana*) it may be more appropriate to measure cover, using quadrats, line intercept or point intercept methods. RECCE descriptions are suitable for monitoring long-term compositional changes in vegetation, but 20 × 20 m permanent plots are more suitable if data on mortality and recruitment rates are required (e.g. for canopy trees). Similarly, for monitoring of animal populations, spatial scale, distribution and number of point and line transects for counts using distance sampling must also be considered.

Essential to this process are assessments of desired size of effect and precision (e.g. confidence levels) required for the monitoring programme. Generally speaking, as sample size increases, precision and the level of certainty able to be inferred will also increase. Explicit and early consideration of the required precision will assist with the selection of appropriate sampling methods and sampling designs:

- Specify desired levels of precision and confidence in the ability to detect this level of precision before the fact (e.g. biodiversity managers might be seeking a 25% change in population size with 95% confidence following possum control).
- Assess the cost of achieving these levels (given the chosen sampling method).

- Consider the trade-off between sample intensity and precision (see discussion on analytical power in 'Choosing among designs and sample sizes' below).

Decisions on length of time over which a monitoring programme should operate, frequency and seasonal timing of remeasurement, and parameters to be monitored are clearly dependent on the monitoring objectives, species and system being studied. For example, the average individual life spans of species range from days to centuries; many organisms have a marked seasonality so can only be measured at certain times in their lifecycle; some organisms (e.g. rodents) are prone to eruptive outbreaks; some ecosystems undergo rapid unpredictable change (e.g. braided rivers), while others (e.g. forests) may show little change over decades (Lee et al. 2004). As well, populations of plants and animals are influenced by numerous factors, e.g. variation in climatic conditions; herbivory; predation. These influences on the populations of interest can be rare, common, predictable, unpredictable, local, widespread, and can act slowly or quickly. Monitoring programmes therefore need to be structured so they are able to partition out these sources of variation on the population of interest, where it is relevant to the management objectives. The obvious solution is to ensure monitoring is carried out over a timeframe that is long enough to include most temporal influences acting on a population (Morrison et al. 2001). If these processes are slow, subtle and complex, or are very rare events, then long-term monitoring programmes are called for.

Develop data management systems

The types of data gathered, and how these need to be integrated in the management programme will define the type of support needed for data handling, storage and analysis and the training and infrastructure needed to support these activities. The usefulness of a monitoring programme will depend on its ability to synthesise, interpret and present results in a form that can inform and guide action (Lee et al. 2004).

Depending on the monitoring objectives, sampling design and monitoring methods being used, the following actions are recommended:

- Ensure the collected data are promptly converted into an electronic format.
- Ensure the electronic data entry and storage format (e.g. spreadsheet) corresponds with the data collection format (e.g. the fields on the electronic form should be in the same order as those on the paper data collection sheet). This consistency increases data processing speed and accuracy. Deviations from the recommended format should not be tolerated.
- Use automated data validation rules (particularly useful at the data collection stage when using data loggers) or other data checks (e.g. drop down 'pick-lists' in spreadsheet columns) to improve data quality.
- Store data in a format suitable for importing them into any specialist software that is used for analysis. An appropriate default format would be a spreadsheet with data arranged in column variables.
- Ensure that data files can be converted to a variety of file types (e.g. .txt or .dbf).

Data should be stored securely in a specified electronic format (e.g. National Vegetation Survey (NVS) databank for vegetation data) with appropriate levels of metadata recorded. Duplicate copies should be held elsewhere and paper records/samples stored in an approved archive.

Evaluate resource requirements

Resource evaluation is important to ensure adequate resources can be committed to the monitoring programme for its intended duration including analysis and reporting. On far too many occasions monitoring programmes fail because of staff turnover and inadequate resources. The cost of a monitoring programme is dependent on current knowledge of the species being monitored (e.g. knowledge of the effects of aerial 1080 operations on rat populations at comparable sites), programme design (including frequency of remeasurement), spatial and temporal scales, equipment required and various overheads costs such as those for data entry, statistical advice (before, during and after initiation of monitoring) and analysis. Each monitoring programme should therefore be costed individually. There are many important considerations:

- Determine whether the necessary funds are available now, whether the programme can be sustained into the future, and whether the scale and complexity of the project are feasible within budgetary and other resource constraints.
- Practitioners should be aware that monitoring has hidden costs, such as field time lost due to bad weather and staff-training time. Such things should be planned for when allocating resources.
- Similarly, investment in sampling infrastructure (e.g. marking of tracks and plots, fencing of plots) can be time consuming and costly (at least initially), but is usually worthwhile in the long run.
- Once the total cost of the proposed programme is known, a decision can be made as to whether it should proceed, be changed or be abandoned, well before substantial resources are committed.
- Also consider whether it is likely that the question initially posed can be answered.
- Is cost the only resource issue, or are the skills required and degree of difficulty important factors?
- What is the cost of not proceeding with monitoring?
- Is there a cheaper robust alternative?
- Can monitoring be scaled down and still produce meaningful information?
- Should the work be done at all if monitoring won’t provide an answer or is too expensive?

Finalise design and sampling programme

At this point, the entire design should be reviewed for practicality and cost effectiveness, an approach for implementation considered, and any amendments made prior to initiating sampling. The sampling programme will benefit from critical peer review by those capable of commenting on objectives, design (including analytical methods and statistical power), practicality and management implications. There are at least three approaches to implementation: test, pilot study, or full scale implementation (Hill et al. 2005). The second approach—use of a pilot study—is recommended as the bare minimum.

Test

Undertake trials of one, some or all of the components of data collection and data analysis to determine whether or not the chosen methods are practical. The trials can be field or office based.

Pilot study

Conduct a small-scale implementation of the entire sampling programme in the field to check whether a useful answer is likely to be produced. Pilot studies are essential if high variance is suspected, standard methods are not being used, significant resources for the monitoring programme are required, and the methods are being applied to species and habitats for the first time. Rather than question the need for a pilot study, it is more useful to ask whether there are any exceptional conditions that negate the need for a pilot study. A partial exception might include the existence of a well-established methodology or standard operating procedure (SOP). Data from the pilot study can then be used in a power analysis to determine the number of samples required to achieve the desired effect size, precision and confidence levels, and to evaluate the validity of a sampling method's assumptions. *Ignoring the need for a small scale pilot study can result in the final sampling design becoming a very expensive pilot study by default.*

Full scale

Here, a sampling design is produced and implemented immediately. This approach should only be adopted with extreme caution. The only potential exceptions are where well-established methodologies, such as an SOP or established best practice guidelines, already exist and they can be implemented with minimal adaptation, or for small projects where the risk of failure is small.

If deficiencies in the proposed sampling design emerge during this review phase (e.g. poor precision or estimates, or the data requirements are too expensive to answer the objective or question being posed), then we have three options:

1. Increase the amount of data being collected (sampling effort) to improve precision.
2. Accept a lower confidence interval—assuming that this new level can be achieved.
3. Abandon the current design and consider redefining the question.

The last phase of the planning process is to allocate responsibility for each of the tasks. This minimises the potential for wasted resources and effort through human error. It also ensures individuals have a clear understanding of the tasks and roles assigned to them (Miller & Allan 2002).

Implementing the sampling programme

When all the previous steps have been carried out, the inventory or monitoring project can be put into practice.

Data collection

Once the design of the sampling programme has been finalised, data collection for the monitoring programme can proceed. This process will be straightforward if field staff are adequately trained and they have access to all necessary resources, data are collected according to the standards and procedures identified previously, data quality and storage protocols are followed, and any problems are quickly identified and dealt with.

Field staff must have read the technical specifications, including the method protocol, before heading off into the field. Some 'official' monitoring protocols (e.g. Residual Trap Catch) are revised periodically. Check which version was used when the data were last collected. Identify variations and how those might affect monitoring in future. Use the latest version for new projects.

Data collection tools are also vital because they help to streamline the data handling process and, if well designed, they will ease the data analysis and interpretation process. There are many different forms of data collection tools. Some common ones are:

Data recording tools

These are tools for use in the field to physically record the data as it is observed or collected. These will be tools like record sheets, electronic data loggers and field note books. The use of paper data sheets or electronic data entry formats reduces the likelihood that important data are missed and ensures that only analysable data are collected. Without a data template, you risk ending up with a pile of unsystematic and uninterpretable observations.

Considerable effort should be spent on the design of data sheets and, depending on the circumstances and complexity of the monitoring programme, several forms may have to be designed (where they are not supplied with the toolbox). Note: minimum attributes are required for all methods described in the toolbox. Sutherland (2006) lists three main types of data sheets:

- Single event sheets where a form is completed for each occasion (e.g. a single survey).
- Continuous data sheets where a new observation is recorded usually in association with a date, time and location (e.g. when a new animal is caught and marked).
- Updated record sheets which are often based around a site, nest or individual. Much of the data will only be recorded once (e.g. location data, band number) but the sheet can be added to at future visits.

Regardless of the type of data sheet to be used, the following tips for creating data sheets (after Sutherland 2006, p. 9) are worth examining:

- Place boxes around everything that has to be filled in, especially if other people are filling the forms in.
- Make the box size appropriate to the amount of detail to be captured.
- Arrange the sheet to maximise efficient use of space. It may be possible to fit multiple records onto one sheet.
- Order the data fields in a logical sequence. Include space for observer's name, date and time (including start and finish times if necessary).

- Consider how the data will be entered into a spreadsheet or database. The data sheet sequence should mirror that of the spreadsheet or database.
- Consider how the data will be analysed. Should the data be continuous or categorical?
- Get the data sheet peer reviewed for ease of use, especially by those likely to be using it. Test it in the field and modify if necessary.
- Think about data entry rules, particularly if ambiguous and unusual cases are likely. Would the inclusion of a worked example help?
- Leave space for notes in which unusual observations can be recorded.

Data curation

Information should be collected, consolidated and securely stored as soon as practical, preferably immediately on return from the field. The key steps involved are data entry, storage, and maintenance for later analysis. Before storing data, check for missing information and errors, and ensure metadata are recorded.

Metadata records hold information such as the name of the survey or monitoring programme, when the survey was conducted, its purpose, objectives, methods, sampling design, names of the team members, localities covered including full grid references or latitude and longitude, information on the location of raw data, data access limitations, conditions surrounding data use, location of back-up copies of data and maps, etc. (see ‘Standard inventory and monitoring project plan’—docdm-146272).

Storage tools can be either manual or electronic systems. They will usually be in the form of spreadsheets, databases, summary sheets or other filing systems. All data, whether they be data sheets, metadata or site access descriptions, should be clearly labelled, copied physically and/or backed up electronically, with the copy stored at a separate location (ideally a fire- and flood-proof archive) for security purposes. Losing data (particularly prior to analysis) is an all too common disaster. There are many stories of data and sampling gear being lost during helicopter trips, river crossings, vehicle theft, office relocations and over-zealous spring cleaning operations.

Seek advice from experts about the best data storage systems. Use official secure repositories where these exist; e.g. the NVS databank maintained by Landcare Research holds physical and electronic data from vegetation plots throughout New Zealand. An approved Departmental standard for vegetation data curation is in place (‘NVS data entry, archiving and retrieval SOP’—docdm-39000).

Samples

Collecting, processing, identifying and storing samples can also be part of data collection during inventory and monitoring projects. Evaluate whether it is necessary to take a sample in the first place, where to take it from (e.g. avoid taking a plant sample from within a permanent plot if that species is uncommon within the plot), and whether you need a collecting permit. Don’t make the assumption that someone else will identify and process all your samples for you. It is a common mistake to fail to plan for sample identification and processing. These activities can add considerable cost and time to monitoring programmes.

Collecting quality data

Attention to data quality reduces the likelihood of making poor management decisions based on flawed data sets. Quality data are those collected consistently, using standardised techniques in a comparable and repeatable way. For data to be useable and of good quality, they must address the objectives of the study, be collected and recorded in accordance with the methods specified in the design, and be complete and accurate.

Poor quality data waste time, resources and effort. They lead to poor inferences, poor management decisions and contribute little to answering the monitoring question (Hill et al. 2005). Poor quality data have both top-down and bottom-up effects. At a management level, poor quality data may lead to a string of problems and impacts:

- Manager may not be able to report consistently over time or at different scales (e.g. nationally or locally). There will be little confidence in the results and a lack of evidence to know whether objectives have been met or questions answered.
- Consequently, very little will be learnt about the effects of management actions.
- Managers might believe their actions are achieving certain outcomes, while the reality could be quite the opposite.
- Management decisions might then be based on poor or inadequate information, with inappropriate actions taken.

From the field workers' perspective, poor quality data are also problematic—and disheartening. Workers do not know whether all their hard work in the field is valued and leading to measurable changes.

A well-written technical specification is essential to ensure high standards of data collection are set and maintained.

Quality assurance/quality control

Quality assurance/quality control (QA/QC) is an essential, but often overlooked, component of the data collection and data processing phases of a monitoring programme (Morrison et al. 2001). Quality assurance and control can be maximised by ensuring the following:

- The chosen observers have skills appropriate to the monitoring task (e.g. they can identify plants, birds, have good hearing and are fit enough for the terrain being covered).
- Observers have access to appropriate equipment and data collection tools (e.g. data sheets or data loggers).
- All observers are well trained in the use of this equipment and the measurement techniques set out in the sampling protocol (see 'The importance of training' below). This is particularly important when relatively new or complex methods, such as distance sampling, are being used and where subjective visual estimates (e.g. canopy cover) are being collected. Use group training sessions to calibrate all observers' practices, measurements and interpretations, to ensure data collection is consistent across the group.

- All observers know and understand the chosen field methodology (including potential violations of assumptions). They follow the specifications consistently over time and space. Variations in a method that are perceived at the time to be inconsequential, e.g. deviating from transect routes or sample points in order to find more of the target species, can have a substantial impact on data analysis, data interpretation and comparability with previous samples.
- Measurement techniques are as rigorous as possible and able to be repeated, thus minimising sampling error and observer bias.
- Data are collected according to the technical specifications and defined procedures.
- Data collection occurs over reasonable timeframes so that fatigue, attentiveness and other behavioural lapses do not jeopardise data quality.
- Observers know where they are. Population overestimates can result from incorrectly recording the same individual as occupying very different locations or by monitoring a larger block than intended.
- Observers comprehensively document the location of sampling lines, plots and access routes.
- All data are checked for obvious errors prior to and during data entry (e.g. through the use of validation rules).
- Those collecting and entering the data are given responsibility for data quality.
- Data entry and analysis should proceed as soon as data are collected.
- Quality assurance and control procedures should continue for the life of the monitoring programme.

The importance of training

It should be mandatory for all observers to participate in relevant training prior to the start of a monitoring programme. Intensity of training should reflect the complexity of the sampling protocol. The more complex methods and designs should include practical training, such as: field exercises that include species identification and distance estimation; examination of the theoretical information relevant to the methods being used; discussion of likely biases; potential problems of implementation in the field; and a clear explanation of each method’s assumptions (Scott et al. 1986). A training period of several days or even weeks (e.g. on-the-job supervision) may be required to bring all observers up to an equivalent standard. Without this investment, observers are unlikely to have the skills and knowledge to make good decisions (Kissling & Garton 2006). At worst, untrained observers may compromise data quality, precision and inferential power.

Data analysis

Data analysis options must also be considered before starting any monitoring programme. It is strongly recommended that advice is sought on experimental design, sampling design and statistical procedures before data are collected. Failure to do so can mean waste of significant amounts of time and money, through collection of inadequate data or data that are unable to address the objectives of the monitoring programme. For these reasons, anticipate the need for advice on data analysis and budget for this.

Selection of analysis tools should be consistent with the monitoring design and methods being employed, e.g. distance sampling data requires the use of the program DISTANCE, reconnaissance plots can be analysed using the program PC-RECCE (useful for an initial examination of the data). Initial data exploration and analysis can be investigated with relatively simple tools (e.g. graphs, pivot tables, univariate statistics, etc.). The following self-teaching resources are available:

- 'Using Excel to enter, manage and explore data' (docdm-426700)
- 'Basic statistics using Excel' (olddm-644074)
- 'Designing science graphs for data analysis and presentation'[1]

Specialist analysis tools are often complex, both to use and to interpret the output, and they require statistical skills and intensive use of computers. It is therefore unrealistic to expect everyone to be able to use complex software or analyse complex ecological data. Data management and analysis routines and statistical skills should be identified at the beginning of a project and appropriate advice or training sought on their use. Alternatively, people with the necessary skills should be contracted to do these jobs.

Uncertainty over the type of analysis that should be conducted is usually a symptom of poorly defined objectives and sampling design. These shortcomings should be addressed as soon as possible. If there are a number of legitimate options for analysis, choice should be driven by the need to answer the monitoring question or objective that was posed initially. Regardless of the analysis tools used, all conclusions reached should be supported by the collected data, some level of certainty must be reported (e.g. confidence intervals), the degree of change detected stated, and some conclusion drawn about the question posed (Hill et al. 2005). It is useful to remember that the overall objective is to obtain an understanding of the system in which you are interested, as efficiently as possible; the data analysis is only a tool.

Reporting

A monitoring or inventory programme remains incomplete if field data are collected, stored and analysed but not reported on. Reporting completes the feedback loop, allowing evaluation of results and outcomes against the original management objectives and questions. Informed management decisions can then be made. Reports may also include a critique and re-evaluation of the current monitoring programme (objectives, design, field methods, etc.) and recommendations for improvements.

To be most effective, results must be communicated to all interested parties throughout the monitoring programme and at its completion. The way in which this is done depends on the intended audience and information required. In order to disseminate results effectively, data have to have been explored and summarised first (Hill et al. 2005). Unfortunately this is rarely done well, and often there is a long delay between data collection and reporting.

Reporting must begin with a thorough assessment of the target audiences and their respective needs. This assessment will determine what results should be presented and how they are best communicated. Reporting can be carried out in many different ways. A written report is the most common approach and it should be regarded as an essential part of any monitoring programme.

[1] http://www.doc.govt.nz/upload/documents/science-and-technical/docts32.pdf

The report should present the objectives and major questions, methods used, results, and a discussion that includes the major conclusions and recommendations. Visual presentation of information (e.g. graphs, tables, diagrams and maps) is a particularly effective means of communicating information. All written reports should go through a peer review process and be lodged in the library or filing system to facilitate access by others. Reporting methods that complement written reports include talks with affected or interested groups, one-on-one discussions, and displays, e.g. GIS-based presentations, use of maps, videos, posters and interactive media.

There are four main ways of presenting data (ref. Sutherland 2006):

1. *Graphs*. Visual representations of numerical or spatial information are often easier to interpret than lists of numbers or complex tables. Provided they are drawn and used appropriately, graphs can communicate a large amount of information very quickly, often at multiple levels of detail. They can show complex relationships among multivariate data. Although there are a vast number of graphical formats to choose from, these have been categorised into several main types. The use of different types has been summarised by Kelly et al. (2005).
2. *Tables* present the exact values of the data for specific cases. They are particularly useful when numerous measurements have been made and are too complicated to be graphed.
3. *Maps* are the best way of presenting distribution data. They can be produced at a variety of scales and can be very useful in describing patterns, e.g. the number and size of a bird's home range in a given area in relation to habitat variables.
4. *Text*. Data can also be presented as facts within text. These facts are often simple measures that may not fit particularly well into graphs and tables.

Common sins of monitoring programmes

Common errors and failings can beset any inventory and monitoring programme. Always consider the potential to fall into these traps and make conscious efforts to either eliminate or minimise their influence.

Common monitoring sins (adapted from Sutherland 2006) are noted below:

1. *Failure to use probability-based sampling designs.*
   Subjective selection of sample units is usually not recommended. What are considered to be 'representative' sample units are often not. If non-probability sampling is required (e.g. if topographic constraints mean only ridges can be sampled), inference must be restricted to the area sampled (e.g. to ridges only).

2. *Collecting too many or too few samples.*
   Potentially, it is a waste of time to collect too many or too few samples. Collect enough for a useful analysis and a suitable level of inference. In the case of too many samples, over-collection may raise ethical and conservation issues.

3. *Changing monitoring methods during an inventory or monitoring programme.*
   This will prevent useful comparison between places and across years. Avoid changes in the

monitoring method and the sampling protocol, e.g. be consistent about which transects or points need to be covered and how often, (but see point 14 below: Field et al. (2007) emphasise the value of an adaptive approach to monitoring design.).

4. *Counting the same individual in two locations as different individuals.*
   This is a violation of sampling independence. Potentially, it inflates parameter estimates. Uncertainty over the observer's location increases the risk of individuals being counted twice.

5. *Not being familiar with the species being monitored.*
   A good understanding of the target species is essential when considering sampling method assumptions, potential biases and interpretation of results. For example, males of a species may be more vocal and visible during breeding season compared with females; a threatened plant might be confined to a certain substrate or altitude band.

6. *Poor experimental design when conducting management experiments.*
   Poor sampling design, insufficient replication and the lack of adequate controls (non-treatments) will make interpretation of management actions extremely difficult.

7. *Failure to store information in an accessible, secure manner.*
   Poor data storage protocols for data and metadata can result in loss of data or incomplete data sets. This can make retrospective comparisons difficult or impossible.

8. *Not providing precise information on sampling dates and locations.*
   Records of dates and locations (GPS point data and/or polygons) are vital to future interpretation of data. 'Site A', 'Pureora' or 'Rowallan 1' might be sufficient at the time, but these names will probably mean little later.

9. *Not being honest about the application of the inventory and monitoring methods used.*
   This may seem self-evident, but if sampling equipment, such as a line of traps or a camera, is placed only in locations most likely to catch animals and this sampling is done only during fine weather, these arrangements need to be documented. Further surveys using slightly different methods (e.g. random trap placement with traps set in all weather) may lead to completely different results and an incorrect conclusion that the target species has declined.

10. *Believing the density (or abundance) of a sampled population is the same as the absolute density (or abundance).*
    Every sampling method has its inherent assumptions, biases and inaccuracies. The trick is being able to assess the potential impact of these on the point estimate or trend.

11. *Assuming sampling efficiency is the same in different habitats.*
    Differences in topography and vegetation structure (e.g. forest, woodland, grassland) will influence every sampling method. For example, the assumptions of distance sampling are often very difficult to meet in a densely forested area, but are usually less challenging in open habitats.

12. *Not knowing why you are monitoring.*
    If you don't know the objectives of the study you are unlikely to understand the type of data

required to provide an appropriate answer. Collection of irrelevant data can compromise the efficiency and effectiveness of a monitoring programme.

13. *Assuming others will collect data in the same way with the same enthusiasm and attention to detail.*
    This failing highlights the need for detailed and effective monitoring specifications that leave little room for interpretation. Periodic audits of performance should be built into the project plan to ensure data quality.

14. *Failing to conduct, and learn from, early-stage data analysis.*
    Field et al. (2007) note how important it is to plan, fund and execute sophisticated analyses of monitoring data at the first available opportunity, then to use those results to improve methods.

Statistical concepts

What is sampling?

It is very rare that all of a population can be measured. More commonly, a selection from the population is chosen and only these selected items are measured (*N*). A population could be all the birds of a particular species in a forest, all the plots that could be established in a study-site, or all the people who walk a particular track. The selection of items from a population is called a sample. In these examples, the samples would be samples of birds, plots and people, respectively.

Samples are selected according to a sampling design. The size of the sample to be selected, and the way it is to be selected, are defined in the sample design. Summary statistics are calculated from the sample and are used to estimate population parameters (e.g. population size $\hat{N}$). In the examples above, the population parameters of interest might be the total number of birds (i.e. the population size), the average height of seedlings in the plots, or average length of time walkers take on the track. Choosing the best sample scheme and the most appropriate method to estimate population parameters are very important steps in inventory and monitoring (see 'Sampling approaches' for more details).

Concepts in survey design: bias and precision

An important aspect when designing a sample is to use some form of random selection of items. In sampling, these items are called sample units. Simple random sampling, with random placement of plots within a study site, means population parameters can be estimated without bias.

Bias is quite a technical term and full understanding of it involves knowledge of mathematical concepts such as expected values. Loosely speaking, no bias means that if the same sample design were used many times and simultaneously on the same population (although obviously this is impossible) then, on average, the value of the summary statistic would equal the value of the population parameter. If you were interested in the average number of seedlings/m^2 and randomly located 1-m^2 plots within the study site, the average of the plot counts of seedlings would be an unbiased estimate of the true average number of seedlings/m^2.

Continuing with the example of the seedling counts from plots; while you could feel assured that by using simple random sampling the average of many repeated, simultaneous surveys would give an unbiased estimate of the true average number of seedlings/m^2, you would in fact have done only one survey.

But, how close to the true value is your sample estimate? This question introduces the concept of sample variance. If all the individual sample estimates are similar then the survey has small sample variance, and if all are very different then the survey has large sample variance. Sample variance is often referred to as precision—a survey with good precision has small sample variance and a survey with poor precision has large sample variance.

The ideas of bias and precision are summarised in Fig. 2.

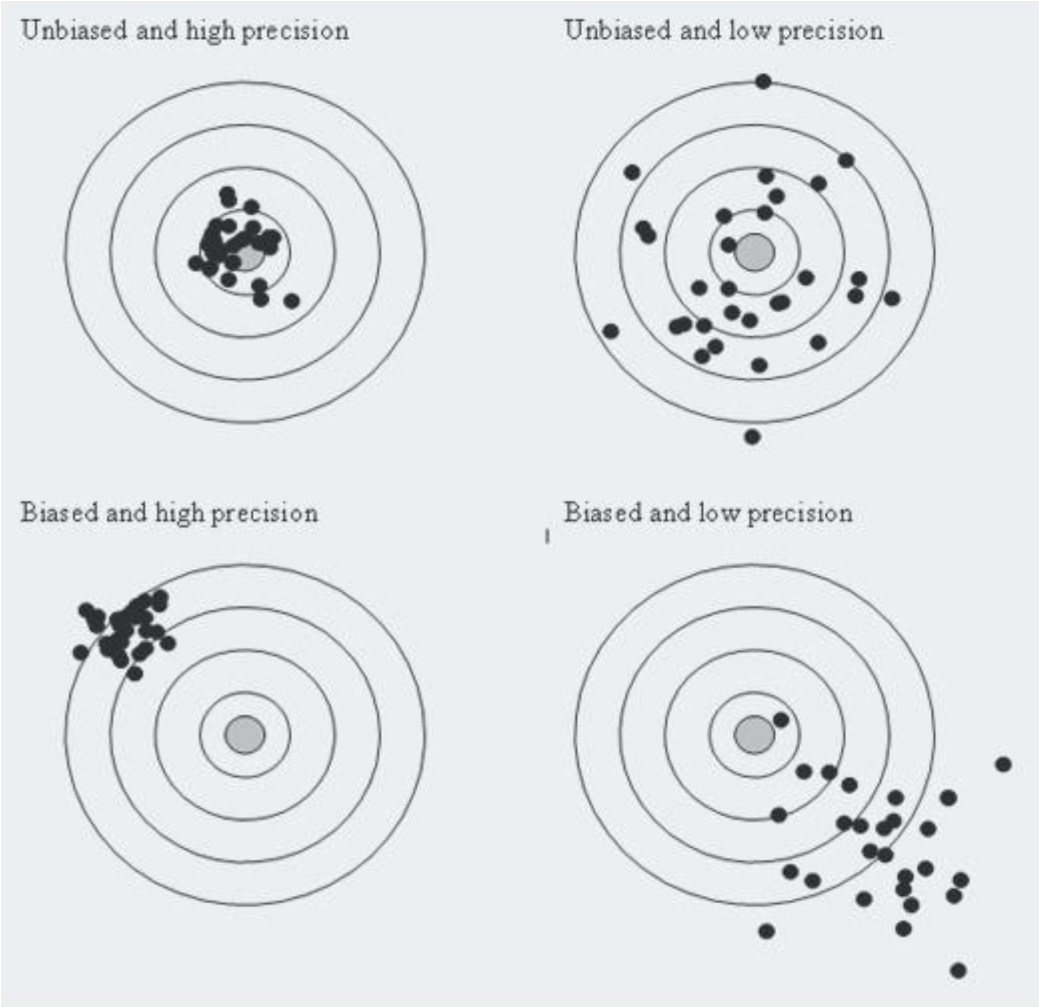

Figure 2. Different states of bias and precision.

Each dot represents a possible value of the sample estimate (from the many repeated, simultaneous samples), whereas the bullseye represents the true value of the population parameter. Putting this back into the context of sampling, you have done just one survey and calculated just one sample estimate.

- The display in the top left corner of the figure shows an ideal situation: an unbiased design and high precision.
- The display in the bottom right corner shows the worst situation: bias and low precision.
- The top right display illustrates that on the average the estimate will be correct, but it may be quite wrong.
- In the bottom left, the survey estimate in every case will tend to be wrong, but it will be out by roughly the same amount each time.

If the survey was designed so that the scatter of potential sample estimates was as in the worst case (bottom right), then chances are your single estimate will be a long way from the bullseye, i.e. it will be quite wrong. For example, your sample estimate might indicate there are 25 seedlings/m^2 when in fact there are 12 seedlings/m^2 (or 55 seedlings/m^2). If the survey was designed so the scatter of potential sample estimates was as in the best case (top left), then chances are your single estimate will be close to the bullseye—it may not be the correct value but it won't be wrong by much.

The most important issue here is how bias and precision can be managed. Bias is often the result of how the data were collected, e.g. through inadvertent use of a stretched tape when measuring tree diameters. Bias might also be introduced by measuring the wrong characteristic, e.g. by counting seedlings in early spring when few have emerged. It might also exist for statistical reasons to do with the mathematical formulae used to calculate the sample estimate, e.g. through use of regression estimators. The bias to do with the mathematical calculation is not a concern for the most of the commonly used survey designs, such as simple random sampling, systematic sampling and stratified sampling.

Precision is most readily controlled by sample size. Generally, a survey with a large sample size will be more precise than one with a small sample size. In other words, the more effort you put into collecting data the better the likely result, but this is true only up to a certain point. Typically, the gain in precision is not linear with increasing sample size, and certainly the difference in precision between sampling 60% of all the units in a population and sampling 70% of the units will be minimal. However, the gain in precision between sampling 5% of the units and 15% will usually be quite large. Establishing the best sample size requires a decision on the desired level of precision. This is one of the considerations in power analysis, as discussed later ('Choosing among designs and sample sizes').

The other way to control precision begins with thinking about what sample unit to use and the actual sample design. Simple random sampling, stratified sampling and systematic sampling are discussed below, but there are many other useful designs, such as cluster sampling, adaptive sampling and generalised random tessellation stratified (GRTS) sampling. Although those designs tend to be more complicated to use in the field, and they have more complicated estimation formulae, they can result in substantial gains in precision when used in appropriate situations. Helpful texts on those and other designs are Thompson (1992) and Manly (2001). Note that each sample design usually has its own mean and variance estimators, i.e. its own mathematical equations for estimating the population mean and its precision. It is very important to use the correct equation.

Probability sampling

Simple random sampling

The most elementary sample design is simple random sampling, where a selection of units is drawn randomly from the population. The important principles here are that each unit in the sample is selected randomly and the probability of any unit appearing in the sample is known. This is where the term probability sampling comes from. For simple random sampling, the probability of a sample unit appearing in the sample is the same for each unit. There are two variations on simple random sampling: sampling with replacement and sampling without replacement. Most commonly, sampling is without replacement, so, unless stated otherwise, assume simple random sampling will be of that type.

The probability of a unit appearing in the sample is used in theoretical statistics, from which we get the standard equations needed to estimate sample variance. If the probability of a unit appearing in the sample is not known, it is not probability sampling and the standard equations cannot be used to estimate sample variance and confidence intervals. This is the problem with non-probability sampling and is one reason why non-probability sampling is not recommended.

Stratified sampling

Stratified sampling is a very efficient sample design. Here the population units are divided into groups called strata. Sample units are selected from within each stratum, e.g. by simple random sampling or systematic sampling (as discussed below). The idea behind stratified sampling is that the groupings are made so that the population units within a group are similar. The sample variance is then calculated as the (weighted) sum of the within-stratum variances. Because the groupings have been made so that units within a stratum are similar, strata should be less variable than the population as a whole. How the population is grouped will be important in determining the overall sample variance. If there is good knowledge about the population, it can be grouped so that units are very similar and within-stratum variances will be low. This will result in a very precise sample estimate. On the other hand, if there is limited knowledge, the groupings may be such that the units within the strata are not very similar and there will be smaller gains in precision from the use of stratification.

As a general rule, the survey effort within each stratum should reflect the size of the stratum and how variable it is and, if known, how much it costs to sample a unit within the stratum. Maximum gains in precision for a given stratified design and fixed amount of survey effort will be achieved by putting more survey effort into the strata that are large, highly variable and relatively cheap to survey. Mathematical equations set out in Thompson (1992) and Manly (2001) can be used to decide on the optimal allocation of effort among strata.

Returning to the seedling count example; being an observant biologist, you might notice that there are areas within a site of high, medium and low light levels. The boundaries of these areas could be used to divide the study area into three strata (as in Fig. 3). Any number of strata could have been delineated, but the better the grouping is at ensuring your objects of interest (seedling counts in this case) are similar within each stratum, the greater the gain in precision.

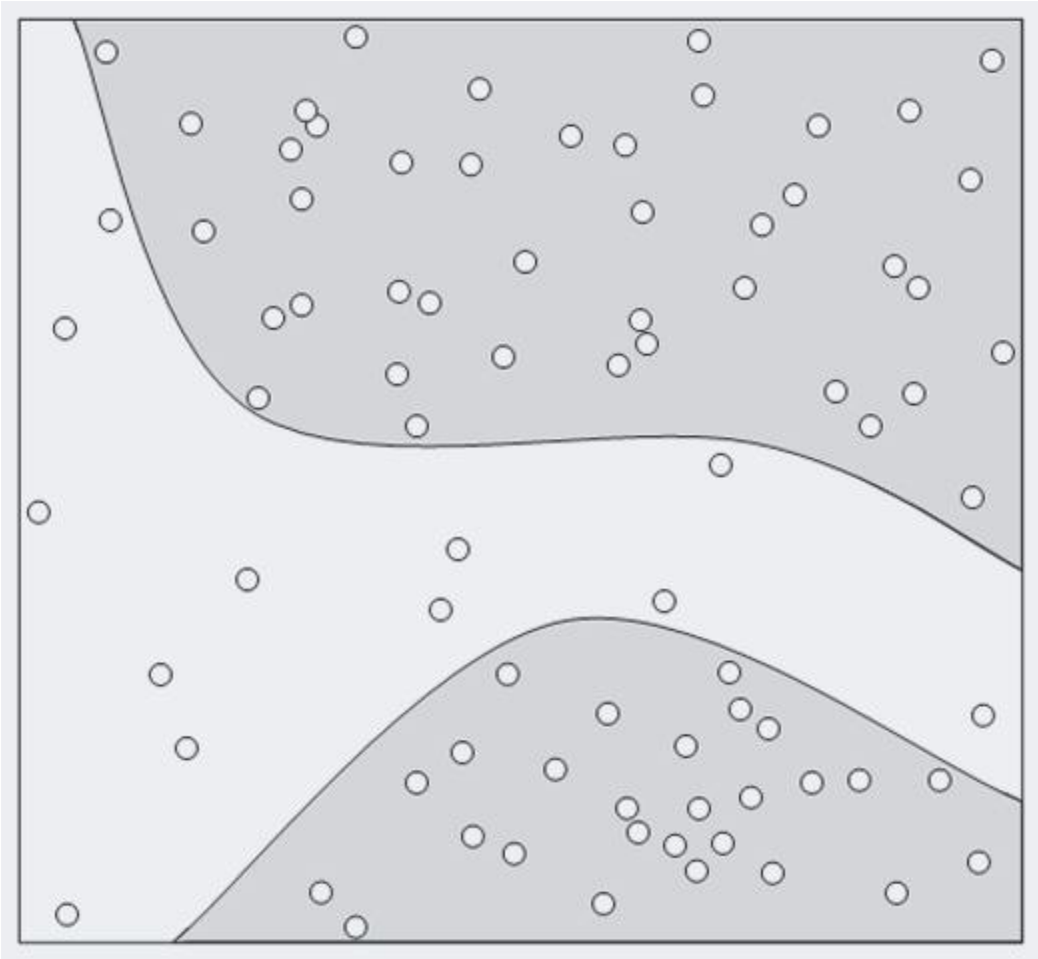

Figure 3. Stratified sampling.

The study area depicted above is divided into two strata; one of high light (in grey) which covers two separate areas, and one of low light (in white). Sampling has occurred more intensively within the high light area where there is a greater density of seedlings.

Systematic sampling

A good alternative to simple random sampling or stratified sampling is systematic sampling. Here, every *k*th unit on a line or grid is sampled. This tends to ease the job of collecting data in the field, and generally it gives a more precise estimate than simple random sampling because the population is covered more evenly. However, systematic sampling will give misleading results, usually in the form of bias, if there is a pattern in the population that is consistent with the pattern used for sampling, e.g. if every 4th plant was planted within the furrow of a ploughed paddock and you happened to sample every 4th plant. Although such regular patterns are fairly rare in natural ecosystems, it is important to check for them. There may be sufficient tendency toward sampling ridges, valleys or streams to cause problems. If there is a regular pattern in the landscape or population, at a scale relevant to your monitoring, you should not use systematic sampling.

Theoretically, estimation of sample variance is difficult with systematic sampling. It is common to use the simple random sampling formula. This will tend to give a conservative estimate of sample variance and precision. In other words, the sample estimate may be more precise than what is estimated.

Non-probability sampling

All the designs described in the previous section involve probability sampling. Probability—or an element of chance—has been used when selecting the sample units. The alternative approach is called non-probability sampling. Sample selection is based on subjective judgment. Non-probability sampling can come in various forms, e.g. haphazard, snowball, judgment and convenience sampling. All should be avoided because it is impossible to make valid statements about precision, and inferential power is either poor or non-existent.

A typical non-probability survey (using the seedling example) would involve walking into a study site and placing plots in locations that were considered representative of that site. The average of the plot counts (of seedlings/m^2) might give an answer very close to the true number, but is no way of assessing this. The danger of the subjective representative sample described here is that the chosen plot locations might not be representative after all.

Choosing among designs and sample sizes

Determining the appropriate design and sample size to achieve your objectives is a critical aspect of project planning. The choices become questions of precision. Various designs and sample sizes can be assessed to estimate likely precision, then the options can be compared. Consider the challenge of designing a long-term monitoring programme with a set of objectives that necessitates the need for replicate sites, replicate plots within sites and replicate visits to plots within seasons, among seasons and over a number of years. You need to decide on how much survey effort is needed and how best to allocate that effort among sites, plots and visits. Given the monitoring objectives, is it better to have a few sites and many plots within each site, or many sites and a few plots within each? Is it better to visit all sites once a year or a few six times a year? What would be the effect of doubling, or halving, the total survey effort? The best design will have taken account of the cost of sampling, where the most variation is likely to be, and estimates of how to achieve the desired precision. Obviously, this can get very complicated, very quickly.

The traditional approach to design and sample size involves statistical power analysis. Statistical power is a concept often used when considering how to design a survey. Power has its base in statistical theory. Without going into detail, imagine an experiment designed to examine differences in plant growth with and without fertilizer. The power of the experiment is the chance that, given there is a difference in growth, it is detected. Other terms used in relation to power are type I and type II errors. A type I error is the chance of deciding that there is a difference in growth when in fact there isn't, and a type II error is the chance of failing to detect a true difference. Type I and II errors are usually denoted as a and b. Power is therefore $1–b$.

Power relates to statistical hypothesis testing. Although hypothesis testing may not always be the best decision-making framework for conservation management, there are a number of power analysis software tools that can be useful when assessing alternative designs and sample sizes. Simple estimates of precision or power can be done using a spreadsheet program like Microsoft Excel. For more complex designs, the best estimates are likely to come from specialist power analysis software, such as the SPSS module Sample Power (see 'Getting the sample size right'—olddm-318638), specialist packages such as 'pwr' in the statistical software R (R Core Development Team, 2011), or Program MONITOR[2]. See Gerrodette (1987) for a useful early paper on power and Steidl et al. (1997) for a good summary. O'Donnell & Langton (2003), Peltzer et al. (2005) and Haigh et al. (2007) provide relevant New Zealand examples of power analysis.

Field et al. (2007, p. 488) comment that: 'Perhaps the most obvious and widely known method of increasing statistical power is simply to increase the sample size. In long-term monitoring studies, this can correspond to extending monitoring over a longer period. We suggest that it can be very useful to all concerned—researchers, managers and funding agencies—to know in advance how rapidly statistical power is likely to increase over time, and thus exactly how long-term an investment will be required in order to achieve the objective of the programme.' They recommend 'an assessment of the future trajectory of statistical power should be built into the early stages of any monitoring programme' (p. 488).

Observational v. experimental studies

The distinction between observational and experimental studies is based on whether there is some intervention. When a population is simply being observed (e.g. to estimate the size or change in size of a bird population), it is an observational study. When the population is being measured after some kind of treatment, it is an experimental study. Here, one population, or a part of that population, is subject to a treatment, such as supplementary feeding, while another population, or part of it, is not. The point of the experiment is to determine the effect of the treatment.

There are three important concepts to address when designing a true experiment:

1. Random allocation of units to treatments

This may be relatively easy for some studies. For example, to study the effect of fertiliser on plant growth, plants growing in pots will be either fertilised or not. The experimental unit here is a plant in a pot. Each unit is randomly assigned to receive one of two treatments (fertiliser and no fertiliser). In ecological studies, random allocation of units is often more difficult because the treatments are typically large-scale and entire areas receive treatments (e.g. when possum control is being done primarily for management purposes rather than experimental ones).

2. Controls (non-treatment)

Controls are where some of the experimental units are not subject to the treatment manipulation. Experimental controls are best thought of as non-treatments to avoid confusion with other uses of

[2] http://www.esf.edu/efb/gibbs/monitor/

the term control (e.g. pest control). In ecological studies, this usually means there is at least one non-treatment area. The purpose of a non-treatment area is to allow comparison with the area that received the treatment. Statistical analysis is used to identify any significant variation over and above that seen in the non-treatment area. Ideally then, both the treatment and non-treatment areas should be as similar as possible, to maximise the chances of detecting differences that are attributable to the treatment.

3. Replication

Replication enables measurement of the intrinsic variability of the experimental units that has nothing to do with the treatment, i.e. the experimental units are physically separable and this allows treatment to be assigned independently (Williams et al. 2002). There needs to be sufficient replication to allow reasonable estimation of this variation—the challenge is deciding how much is sufficient. Without replication, error cannot be estimated and statistical tests cannot be applied.

It is important to avoid confusing replication with *pseudoreplication*. This occurs when multiple measurements are taken on the same experimental unit but they are treated as independent data points. Pseudoreplication should always be avoided because the results are not scientifically valid. Thus, it is important to replicate observational studies to understand patterns (over time or space) and to define appropriate sample units for experimental studies.

Studies conducted over time, either observational or experimental, need to be carefully planned and designed. Issues to consider are the allocation of effort in ways that balance spatial and temporal replication, adequacy of field protocols so any observed change in the population is not confounded by differences in data collection methods or observer ability, and choice of suitable statistical analyses.

The statistical analysis of data collected on the same sample units over time is called longitudinal data analysis or repeated data analysis, or analysis of repeated measures (ARMS). This specialised area of statistics is based on analysis methods for data with temporal correlation, where the observations made in one time period are not independent of what was observed in previous time periods. Training in repeated measures analysis has been identified as a prerequisite for these types of analysis and is currently available to DOC staff.

Sampling approaches

Introduction to sampling approaches

This section of the Inventory and Monitoring Toolbox provides a number of starting points from which researchers and managers can assess the objectives of their intended population analysis against the biology of the target species and the variety of available methods. It is not the aim of this document, nor is it practical, to address the many variations on general methods that are discussed below or make strict recommendations. However, where there are obvious advantages (e.g. where a method that is particularly suitable for threatened plants) or where there are established national protocols (as there are for monitoring of vertebrate pests) recommendations are made as to the most appropriate sampling and survey methods. Users are encouraged to understand and be critical of all monitoring methods and their application.

Figure 4 (after Thompson 2002) illustrates the hierarchical relationships between the different approaches to population abundance monitoring. This framework is relevant to all species groups discussed in the Inventory and Monitoring Toolbox, although obviously some approaches will be of more relevance to some species than others. A brief, generic description of each approach is provided below.

It is also worth pointing out here that monitoring approaches can also be classified as either direct or indirect. Direct monitoring entails counting actual plants and animals, whereas indirect monitoring involves inferring population size and trend from counts of (usually) animal sign (such as scats or tracks) or from information on animals taken by hunters, anglers and the like. Indirect monitoring has been developed extensively for fisheries management and, overseas, for harvest of some ungulates.

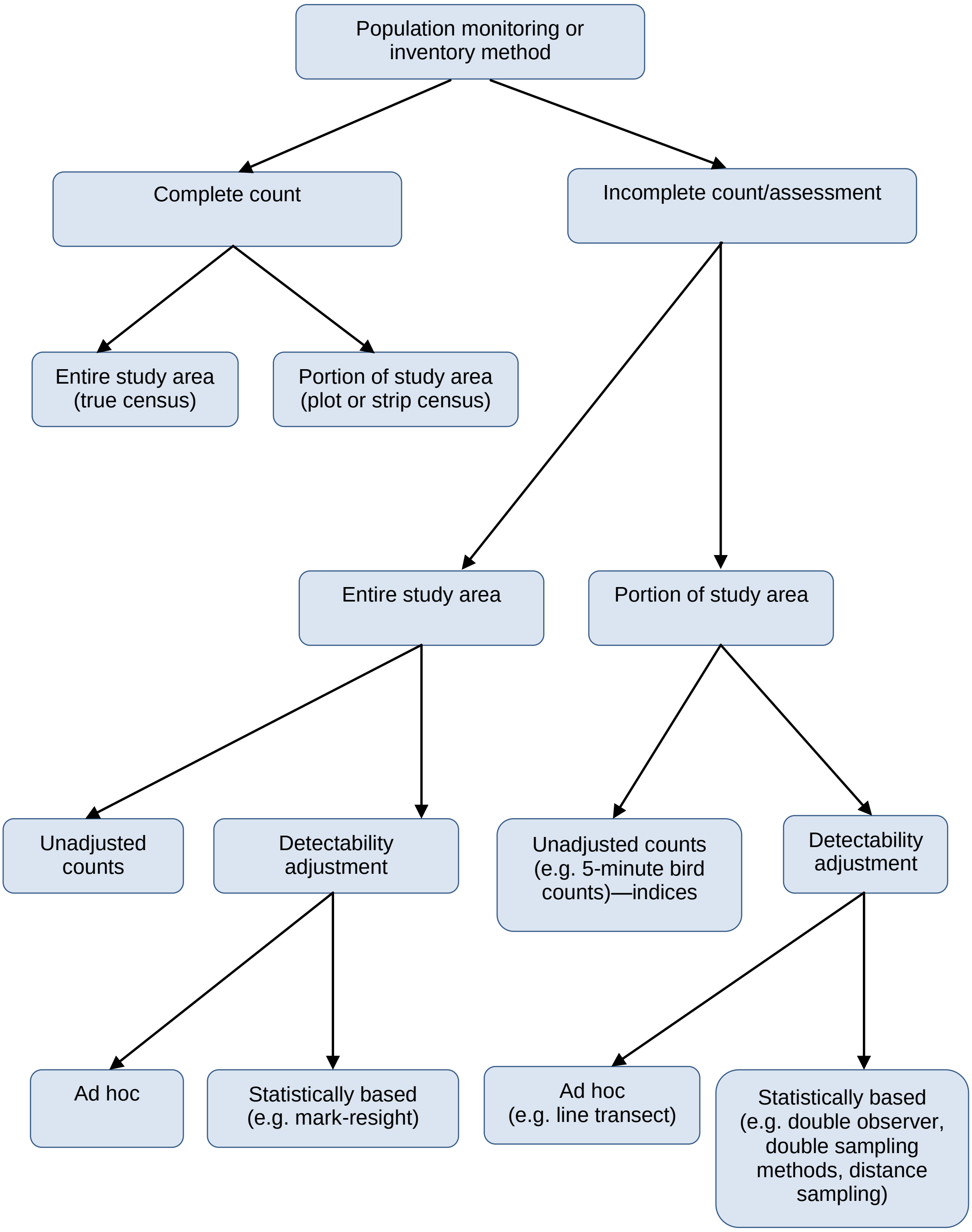

Figure 4. Count methods.

Complete counts—covering the entire population

The fundamental assumption of all complete count methods is that all objects of interest within the covered region or plot are detected (Borchers et al. 2002). For example, permanent plots assume that every tree and sapling is tagged and/or counted within a 20 × 20 m (400 m^2) plot and every seedling is counted within 24 seedling subplots with a 0.49 m radius.

Total count

A total count is a complete census of individuals within a sampling unit. That is, the area searched either covers the entire survey region or sampling frame (covered region = surveyed region) and therefore you require no statistical analysis, or the covered region is less than the surveyed region (as in plot surveys) and abundance is estimated using design-based or model-based methods (Borchers et al. 2002).

In its strictest sense, a total count of all organisms of interest will only be possible if all the following conditions are met:

- All plants and animals within a particular population, roost, colony, lek or group, etc. are able to be counted.
- The individuals are highly visible.
- The roost, colony, lek, etc. is fully occupied (e.g. all members of a shag colony are present and available to be counted).
- Population is demographically closed for the duration of survey (i.e. no births or deaths and no inter-site movements—individuals are only counted once).

These conditions are rarely met in an inventory or monitoring situation for populations of mobile species. Even remote census techniques (Best & Fowler 1981) and methods where the intent is to count all members of roosts, nesting colonies (Moore 2004) or large visible flocks of species such as river birds or waders (Maloney et al. 1997) rarely, if ever, achieve this goal. Conditions affecting detectability, visibility and distribution—and most animal populations tend to either cluster or avoid one another—can, therefore, cause observers to severely underestimate abundance.

However, exceptions can occur at small spatial scales, for birds in particular, or in areas or habitats where individual plants and animals or other objects of interest, such as seabird burrows, are readily detectable (Thompson 2002). For example, total counts may be possible for species that occur in restricted habitats, or are rare and living in isolated and well-defined areas. Examples of such species are the New Zealand dotterel, shore plover and Chatham Island oystercatcher (see Schmechel & O'Connor 1999, Dowding & Davis 2007).

There are many examples from the plant world where a total count has been used to determine the size of a population. In Northland, a complete count of *Atriplex hollowayi* has been undertaken since 1990 (de Lange et al. 2000). In Bay of Plenty, the rare orchid *Calochilus roberstonii* has been subject to one of the longest-running monitoring programmes for plants in New Zealand (Merrett et al. 2001). Complete counts have been carried out since 1985, with an annual count since 1993. Large shrubs, trees, and other long-lived plants are often subject to total counts. *Muehlenbeckia astonii*, *Pittosporum patulum*, *Ileostylus micranthus* and *Euphorbia glauca* are, amongst many others, all subjects of repeated counts at sites throughout New Zealand.

Great care is needed to avoid being complacent about the possibility of missing some individuals, even when one is familiar with the target organism and its habitat. Even large animals can be surprisingly easy to overlook, particularly if some individuals take evasive action as you enter their habitat. Similarly, some plants can be missed if they occur in areas of dense vegetation or they do not produce above-ground parts every year (Sutherland 2006). Providing these limitations are understood (i.e. providing you do not believe that abundance derived from a total count is necessarily the same as the true abundance) and relative bias and precision of the counts remain constant over time, total counts remain a legitimate and useful sampling option.

In many situations, it may be good enough and better than nothing to get close to a complete count, thereby reducing the sampling variance (compared with other sampling methods) for between-year comparisons. It is also possible that such a sampling approach will be more efficient (or more accurate) for a given expenditure. However, this is dependent on being able to demonstrate that a consistent standard of completeness (or detectability) has been met. If this cannot be done, interpretation of estimates should only proceed with a great deal of caution. 'Minimum number known to be alive' (MNA) is often the measure obtained from counts that are not a true census but are as complete as possible. Despite these concerns, counts of this type are used extensively in New Zealand to measure temporal and spatial trends—these may be better than no knowledge at all. Prominent among these is the National Wader Count scheme initiated by the Ornithological Society of New Zealand in 1983 (Sagar et al. 1999). Other so-called complete counts of this type include lake and waterfowl counts, multi-species riverbed surveys (Maloney 1999), albatross colony counts (Moore 2004) and gannet colony counts (Greene 2003).

Total mapping

Total mapping is simply a variation of a total count that reduces the chances of counting the same individual twice (i.e. individuals become identifiable). While this method is commonly used for estimating numbers of territorial birds, it can also be used for local endemic plants, e.g. *Clematis marmoraria*; and invertebrates, such as *Placostylus* snails (Sherley et al. 1998) whose entire population occurs within a small, discrete area. When monitoring birds, if the density of birds is low enough, breeding territories of uniquely banded individuals can be mapped within a defined area and used as a complete count (i.e. total mapping) of the breeding population within that area. Call playback and provision of food can be used to enhance detectability, particularly for those species that are relatively cryptic. Bear in mind, however, that such tools might unduly influence the dynamics or health of the population being observed.

A major problem when 'total mapping' of mobile species (and for territory mapping) is being able to determine whether or not marked individuals are resident within an area (e.g. within a territory) given the variety of behaviours and densities (e.g. range, territoriality, seasonal flocking and conspicuousness) displayed by species. With total mapping, the observer usually only attempts to estimate the breeding population and does not include other members of the population (either marked or unmarked) nor those that do not hold territories, such as non-breeders and transients. However, an advantage of this is that it allows the observer to efficiently target any unmarked individuals for inclusion in the marked population.

As this method requires significant effort, with repeat visits to large proportions of a given site, it is extremely expensive in terms of both time and cost. For this reason, it is within a class of methods that is relatively inefficient in terms of results per unit of fieldwork effort. However, for rare territorial species, counts of this nature may be the only option for obtaining reliable abundance estimates, especially if there are too few individuals for use of methods such as mark-recapture and distance sampling.

In New Zealand, total mapping counts have been used to good effect to measure population changes of small, highly visible forest passerines following the aerial application of toxins used to control possums (Powlesland et al. 1999).

Complete counts—covering a portion of the study area

Plot sampling

Plot sampling attempts to count individuals within a defined subset of plots inside the survey area. The essential assumption underlying plot sampling methods is that all individuals in the searched plots within the overall survey area are detected with certainty. Plots are chosen according to a probability-based sampling design, e.g. a simple random sample, stratified sample or a systematic sample (see 'Probability sampling' above). Plots can take a variety of shapes (quadrats, strips, circular plots, etc.). Estimation of the total number of individuals within the survey area can then be derived by extrapolation from the numbers counted within searched plots. Uncertainty in this estimated abundance is simply a result of the entire survey area not having been searched and the variation observed between plots. Two different approaches can be used to deal with this uncertainty.

1. *Design-based methods* (of which there are many) use the survey design to introduce randomness to plot selection when estimating abundance from survey data. Such design-based estimation approaches are valid whatever the spatial distribution of plants and animals within the survey region.

2. *Model-based methods* provide a flexible and useful alternative, principally for mobile organisms. These methods use a statistical model of the distribution of organisms within the survey region to estimate abundance: The average density of the animals of interest (for example) is estimated using assumptions about the randomness in their location, size, sex, etc. (a state model), and randomness in whether a particular animal is detected (an observation model). This approach can improve precision, but it can be biased if assumptions cannot be met.

Although plot surveys are common when monitoring plant populations, they are uncommon in wildlife abundance estimation. The assumption that all animals are detected within a plot is often unrealistic, particularly if the animals are mobile and moving about at speed, or not all animals are available to be counted at a point in time. The methods may also be applicable to some populations of large surface-nesting birds such as albatrosses, to colonial mammals such as seals, and to counts (indirect rather than direct) of the burrows made by seabirds such as petrels.

Incomplete counts—unadjusted counts, simple counts and indices

In cases where it is not possible to take a complete count of a population, an incomplete count must be taken instead. Three approaches to incomplete counts are covered here: presence/absence indices, territory mapping and indices of relative abundance.

Given the difficulty of conducting complete counts, data from incomplete counts (where not all individuals are counted within a sampling unit) often have to be used for monitoring purposes. Incomplete count methods can be classified into two broad groups: ‘those that do not account for, or do not properly account for, incomplete and (very often) unequal detectability of individuals (index methods, e.g. 5-minute bird count, Residual Trap Catch) and those that do (plot counts, distance sampling, mark-recapture, etc.). Index methods may be further classified as presence/absence and relative abundance index techniques’ (Thompson et al. 1998).

Incomplete or partial counts of plants and animals can be represented by a simple relationship between the observed count (*n*), the probability of detection (*p*) and the true number of animals (N), within a defined area and time period.

Thus the population estimate ($\hat{N}$) is $\hat{N} = \frac{n}{p}$

Our ability to estimate abundance with any degree of accuracy is therefore entirely dependent on our ability to estimate *P* or detectability (Thompson et al. 1998). We can either assume that the proportion of individuals detected is constant across all plots and times (as index methods do) or we can adjust abundance estimates using the *P* derived from the sampling process (or, in the case of plot sampling, from the design process). Provided that (1) our method of estimating *P* is statistically valid (and there are a number of improvised methods that are not) and (2) we have assurance that the assumptions underlying the method are biologically realistic, then we can have some confidence in our estimates. Ideally, the assumptions underlying any count method must be tested for validity before the method is adopted as part of a monitoring programme for any species. In practice, assumptions are not often satisfied, but we use a method anyway because we believe it is robust to the deviation from assumptions (M. Efford, pers. comm.).

Methods adjusting for incomplete detectability are invariably more time consuming and costly than indices of abundance, and these factors must be taken into account when formulating objectives and planning the logistics of a monitoring programme. Such methods are more likely to be used on particular species where (1) an unbiased estimate of abundance and density is required, (2) there is a need to compare these estimates across time and/or space (particularly if detection probabilities are known or thought to vary) and (3) the assumptions of the chosen method can be met. Often these conditions are unable to be met and an index of some sort will have to be used. However, counts at or on a sampling unit must bear some consistent relationship (either known or unknown) to actual abundance or density in the survey area or the sample-based estimates will be unreliable measures of abundance.

Presence/absence indices

Presence/absence surveys are commonly used to evaluate the spatial distribution of a species or a number of species, e.g. as in the New Zealand Bird Atlas (Robertson et al. 2007). Although this approach is not strictly a counting method (occupancy being a different population state variable), the percentage of sample units containing the species can be used as an index of distribution, a surrogate for abundance, or both (Thompson et al. 1998, MacKenzie et al. 2006). (Ways of dealing with occupancy are discussed in the next subsection.)

At its most basic level, records of occurrence are accumulated but no attempt is made to quantify absence. Such records are often derived from a variety of sources. They also tend to vary widely in quality. This can result in extremely biased coverage and a data set from which little inference about populations can be derived. However, 'rough and ready' inventory data such as this can be used as a basis for subsequent, more detailed surveys. Simply relying on records of the presence of a species within a defined area can easily mask a decline (or increase) in numbers over time (see Fig. 5).

Occupancy—adjusting for detection probability

The detection of population trends using presence/absence methods is complicated by the need to estimate occupancy (the proportion of plots in which the species is present), done by fitting a model that allows for a measured rate of non-detection, assuming independence of successive surveys. This will require repeated surveys of plots. However, if the population is rare and/or the monitoring objective is mainly concerned with defining the proportion of sites and area occupied (i.e. spatial distribution) over time or resource selection relationships, then estimation of site occupancy rates, where the probability of detection (P) is < 1, may be a useful approach (MacKenzie et al. 2002). That is, greater value is derived from presence/absence surveys if the probability of failing to detect target species within surveyed areas is also estimated.

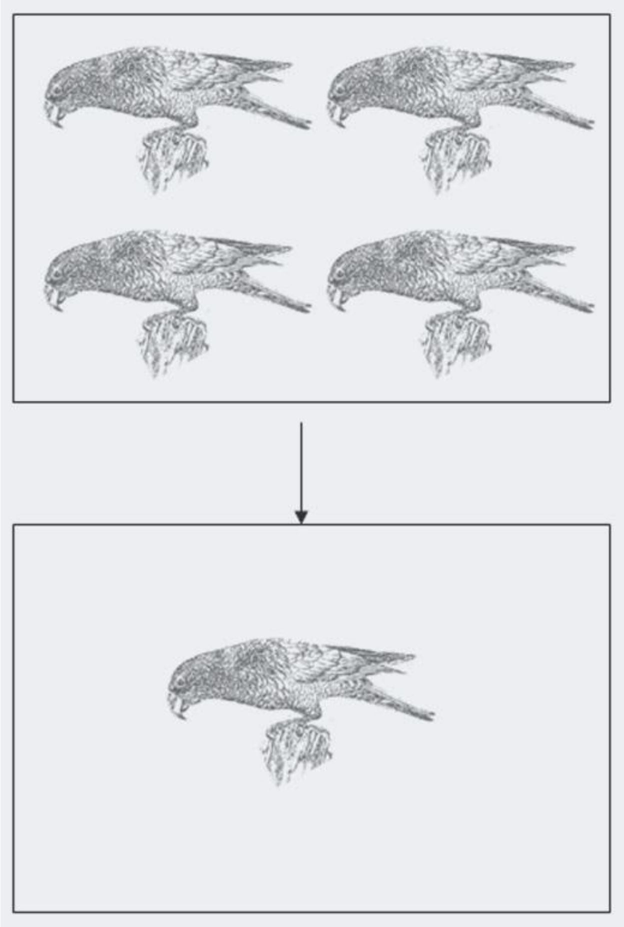

Figure 5. An illustration of a limitation of presence/absence data.

Figure 5 represents the actual number of kaka on the same plot over two successive time periods. If we assume complete detectability, presence/absence data only demonstrates kaka were present during both sampling periods. It would not detect the 75% decline that had occurred (Thompson et al. 1998).

Territory mapping

Territory mapping involves the standardised mapping of locations of unmarked animals. It is usually applied to monitoring of birds, but can be used for any animal inhabiting a defined territory within an area over a number of visits. (Plants do not have territories so are excluded from this discussion.) The method has been used widely overseas (Thompson et al. 1998; Bibby et al. 2000). Distinct sighting clusters are counted as the core locations of a territory for a single pair of animals (as per the British Trust for Ornithology's Common Bird Census, see Fuller et al. 1989; Marchant et al. 1990; Bibby et al. 2000). The major problem when applying this method is how to define the sighting clusters consistently, given the range of density and behaviours (e.g. territoriality and conspicuousness) displayed by different animal species. In addition, this method only estimates the breeding population. It does not include other members of the population (e.g. non-breeders or transients).

As for total mapping, territory mapping requires significant effort (repeat visits to a large proportion of a given site, followed by intensive mapping and analysis). It is therefore extremely expensive, relatively inefficient in terms of results per unit of fieldwork effort, and the calculation of population 'indexes' derived from such data is potentially significantly biased (Verner 1985). Although these flaws are usually acknowledged, proponents of the method argue that even though annual estimates of 'territories in any one year may be poor estimates of the real population, the results still reflect changes over time as long as the analysis guidelines are adhered to' (Bibby et al. 2000).

Indices of relative abundance

Indices, such as 5-minute bird counts, are measures or count statistics that contain information about the abundance or density *relative* to the absolute (or actual) abundance or density of a population (Williams et al. 2002). The association between a simple count or an index to actual abundance or density is usually assumed to be positive (although there are exceptions). An increase in an index, for example, implies some proportional increase in actual abundance.

A huge number of methods can be found in the literature that seek to convert counts of animals and plants into indices of abundance or density and trends in population dynamics over time (Williams et al. 2002). Most of these techniques rely on sighting and catching of individuals, or inferred evidence (particularly for animals) of their presence (e.g. calls, footprints, pellets, nests). Playback of taped vocalisations in order to draw out a response is also a commonly used technique to obtain information on distribution and abundance of animals (Flux & Innes 2001; Ralph & Dunn 2004), as are hybrid methods that attempt to utilise unsolicited records and elicited responses (both sightings and calls).

Provided appropriate sampling design and analysis principles are followed (i.e. there is robust definition of sample populations, objectives, desired precision, power, sample units and cost constraints, etc.), indices can provide much useful information about the *relative abundance* of populations. This is particularly so for species that are difficult to observe or capture, such as mammalian predators and forest birds. In many cases, relative measures may be the only practical survey method available and they tend to require less effort and expense than more formal estimation methods such as distance sampling or mark recapture, etc. (Williams et al. 2002).

The validity of simple counts and indices as count methods rests largely on two assumptions:

1. The number of birds, fish, plants, etc. that are counted is consistently *and linearly* correlated with actual abundance or density.[3] That is, the detection probability (*P*) of the animals or plants remains constant so that comparisons between, for example, habitats, times or management treatments, are not confounded.
2. If the index is to be used to estimate a parameter (e.g. absolute abundance v. relative abundance), then the index must be calibrated so that an unbiased absolute estimate can be calculated (Williams et al. 2002).

[3] Non-linear relationships can also be catered for. For example, users of trap catch indices in New Zealand adjust for non-linearity when they apply a correction for sprung traps (a frequency-density transformation).

Failure to calibrate an index, test for variation in detection probabilities and use of the index cautiously even when detection probabilities are thought to be constant (or almost so) can result in highly misleading comparisons (Thompson et al. 1998; Williams et al. 2002). Unfortunately, many indices have been neither calibrated nor validated, but they are used as if they have been. In addition, indices generally yield weaker inferences than more formal estimation methods (Williams et al. 2002).

For an index to be useful, it is essential that either the sampling variance of the index is small or the measure can be easily obtained so that the sampling variance can be reduced by having large sample sizes. If high levels of effort (resources) are required to obtain a precise estimate of the index with sufficient power to detect significant change, then the index is not a cheap alternative. The observer would do better to (a) attempt to measure the actual population by accounting for the detectability of individuals (*p*), (b) correct for bias using methods such as double sampling (Bart et al. 2004), or (c) abandon the study entirely and revisit the objectives.

Although a lot can be done to control for variations in detectability by applying sound sampling design principles, e.g. by incorporating differences in detection probabilities among observers as analysis covariates (Link & Sauer 1997, 2002), some influences, such as habitat and topography, are impossible to control. On the other hand, many of the recommended alternatives (often absolute measures) have their own problems, such as restrictive assumptions or increased sampling variance, that make them unsuitable for various field applications (Hutto & Young 2003).

In New Zealand, counts in forested habitats are particularly problematic when the assumption is made that all birds have been detected within some (often unspecified) distance from the observer, or worse, when this assumption is noted as unrealistic but ignored anyway, with no attempt made to adjust counts for detectability (Dawson & Bull 1975). As only the largest changes (50% or more, depending on the power of the design) are likely to be detected using uncalibrated index counts, such methods remain completely unsuitable for monitoring trends in abundance of those species already occurring in low numbers. Unfortunately, the less-abundant species are likely to be of most concern, precisely because of their low numbers (Williams et al. 2002; Purcell et al. 2005). Monitoring schemes that assess all species within an area will be of little value (with the exception of the distribution data that arises from the survey) for these rarer species. Therefore, affordable and useful monitoring programmes using indices are only likely to be possible for the most abundant, least-variable species, or those species that are most responsive to alteration of key habitat attributes, such as seed availability after beech mast and predator levels (O’Donnell 1996). It is worth noting that species that fit these criteria (e.g. keystone species) will differ between sites and, in some instances, over time (Purcell et al. 2005).

Evidence from North American bird count data suggests that trends are only likely to be detected following continuous data collection for at least 15 years, especially if sample units are visited less than six times per season. Studies designed to run for 20 years or more will be necessary to interpret trends from a range of species that includes those that are rarer, more vulnerable or less responsive to key habitat attributes than common species. The number and frequency of visits to sampling stations also needs to be adequate. Often the cost of meeting all these requirements will be prohibitively high (Thompson & Schwalbach 1995; Purcell et al. 2005).

Despite considerable debate within the literature (e.g. Dawson 1981; Thompson et al. 1998; Thompson 2002; Hutto & Young 2003; Bart et al. 2004) indices are still useful (e.g. O'Donnell & Dilks 1986; Dilks et al. 2003), *provided* their use is appropriate to the objectives of the study and the importance of cost versus inferential strength (Williams et al. 2002). Uncalibrated indices of relative abundance should be considered an option of last resort—particularly over short time periods. However, if the objective is simply to examine the relationship of species to habitat types, simple frequencies of occurrence derived from counts within a fixed-radius from a sample point or distance from a line may be sufficient to describe general biological patterns (Hutto & Young 2002). Funding constraints, general ease of use and problems in the application of alternatives, such as species rarity or inability to meet method assumptions, are likely to see indices used for some time yet. If indices of relative abundance are to be used appropriately, a deeper understanding is required of their limitations (particularly cost), the need to correct for bias, the need to gain sufficient sampling power, potential meaningful applications and appropriate analysis.

Recent studies by Royle (2004) and Royle et al. (2004) have highlighted the relationship between occupancy (repeated measures of presence/absence data) and abundance. Of particular interest is the development of models that seek to estimate abundance from site occupancy and count data derived from indices (5-minute bird counts), distance sampling, capture-recapture and trapping webs (MacKenzie et al. 2006).The potential benefits in terms of ease, efficiency, cost and utility from this approach are significant and deserve further investigation. Future developments in this field should be watched with interest.

Commonly encountered methods from which indices of relative abundance are derived include:

- Fixed-width strip transects (detectability less than certain)
- Line transects (unadjusted for detectability)
- Point counts (unadjusted for detectability)
- Area counts
- Vantage point counts
- Standardised mist netting
- Indirect counts (signs of animal activity, e.g. droppings, burrows, nests, feeding sign, foot prints)
- Call counts
- Playback response
- Colony, lek, roost counts (detectability not certain)
- Dog counts (i.e. counts using dogs that point)
- Shooting and trapping counts
- Driving/flushing counts
- Spotlight counts
- Counts derived from remote sensing imagery
- Harvest counts (e.g. CPUE: Catch per Unit Effort)
- Foliar Browse Index (FBI)
- Residual Trap Catch (RTC)
- Tracking tunnels

Incomplete counts—adjusting for incomplete detectability

Earlier in this chapter the distinction was made between those methods that provide a relative measure of abundance and those that provide an estimate of actual abundance ($\hat{N}$). Unbiased abundance or density estimates are only possible using count methods that account for incomplete detectability of plants or animals. Such methods will only work when applied correctly, i.e. when the assumptions of the methods are met and survey design is appropriate. Even if calculating the abundance of a population is not the principal aim of the monitoring programme, these methods will allow direct comparison between species and between the same species in different habitats. Such comparisons may not be possible using unadjusted counts such as indices. Four main groups of methods are summarised:

- Double-observer approach
- Double sampling (calibration of an index with measures of detectability)
- Distance sampling
- Mark-recapture/resight techniques

Point counts are traditionally conducted by a single observer. That person derives a count statistic (e.g. the number of birds detected at a point count location). Double-observer point counts are one of a number of survey approaches that attempt to calculate a detection probability that accounts for birds (or other animals) that are present but not detected (Nichols et al. 2000; Williams et al. 2002). This method uses two observers operating either independently of, or dependently on, one another. A model-based joint detection probability can then be calculated for each observer (using program DOBSERV) which is then used to compute a corrected abundance estimate (Nichols et al. 2000).

Authors generally recommend that the detection of individual birds by the primary observer be independent of detection by the secondary observer (Nichols et al. 2000). Achieving this in a field situation can be difficult, particularly when observers are able to pick up on detection cues from one another. However, even though there may be failure to achieve truly independent detection probability, estimates (resulting in detection probabilities biased high and associated abundance estimates biased low), from double-observer methods are still likely to be more robust than those from unadjusted counts (Nichols et al. 2000). Recent work appears to confirm this. Using an entirely dependent observer sampling approach (where detection probability was derived from the combined abilities of two observers) Forcey et al. (2006) were able to estimate higher detection probabilities with improved precision and fewer logistical constraints than they could with independent observers.

Forcey et al. (2006) also point out that the method (using dependent observers) is reasonably robust to small sample sizes (they recommend ≥ 10 individuals). There are also fewer adverse effects (relative to other methods) from misidentification of individuals and the method performs well in densely forested habitats. Problems with model-selection uncertainty and variation in observer ability can be addressed by maximising the number of animal detections (in this case, bird detections), employing a relatively small pool of observers highly skilled at visual and aural identification, and by restricting counts to a fixed radius (a radius sufficiently short so that all observers are able to detect birds or other animals at that distance) (Nichols et al. 2000; Forcey et al. 2006). Thus, double-observer counts have considerable potential as a monitoring method and are worthy of further investigation.

Double sampling

Demonstration of a relationship between an index of relative abundance and the actual population abundance would significantly boost confidence in the index. The process of doing this is known as calibration. To do this, data on both quantities are collected simultaneously, from which an equation is derived to convert index counts into estimates of population density (Conroy & Carroll 2001; Morrison et al. 2001; Bart & Earnst 2002). Although the process sounds straightforward, the practicalities are often extremely difficult. For each species, the true population size would have to be determined (using some sort of absolute complete count method) and compared with the proposed index at the same site—a daunting and often prohibitively expensive task as is apparent in the small number of such studies in the literature (Thompson 2002). Small sample sizes and high sampling variation in either (or both) actual abundance or index methods can exacerbate the situation.

These problems can be relieved by using double sampling. This approach only uses the more intensive quantitative methods on a sub-sample of the study area, whereas the less intensive method (the index) is used for all the samples (Bart et al. 2004). The index can then be calibrated against the more intensive method using a ratio (Bart et al. 2004) or regression estimator (Thompson et al. 1998)—it is essentially 'calibration on the fly'. The two key assumptions of this approach are that complete counts are achievable (or estimates are unbiased) in the sub-sample of units counted this way and that detectability is constant for incomplete counts. As we have already seen, both of these assumptions are often extremely difficult to meet in animal population studies. Although rigorous evaluation of these assumptions (and the double sampling method generally) is required for different populations and habitats before double sampling is incorporated into any population monitoring programme (Thompson et al. 1998; Thompson 2002; Bart et al. 2004), it does provide a possible way forward.

Distance sampling

Distance sampling encompasses a family of methods (line transects, point transects, cue counting, indirect sampling) designed to provide reliable, comparable (temporal and spatial) and unbiased absolute estimates of population density. Distance sampling has been widely advocated as a solution to the problem of incomplete detection (e.g. Buckland et al. 2001; Rosenstock et al. 2002; Ellingson & Lukacs 2003; Buckland 2006). Distance measurements, usually from a line (perpendicular) or point (radial) to the object of interest, are used to address incomplete and unequal detectability of individuals. These distances are then modelled using various forms of detection function and the best-fitting model used to generate density and abundance estimates for a given area. Robust, comparable and unbiased density estimates are possible provided model assumptions are met. Distance sampling methods have been applied to a diverse array of taxa and objects of interest, including rainforest trees, wilding pines, reptiles, invertebrates, marine mammals, terrestrial mammals, birds, animal dung and nests (Thompson et al. 1998; Barraclough 2000; Buckland et al. 2001, 2004). A comprehensive explanation of the methodology is provided by

Buckland et al. (2001, 2004) and Barraclough (2000). Specialised computer software for the analysis of distance data is freely available (such as program DISTANCE[4]).

Whilst static objects of interest such as plants, burrows or dung are relatively easy to deal with using distance sampling methods (indirect distance sampling), mobile objects of interest such as mammals, lizards and birds can be problematic. Robust use of the method therefore requires close attention to survey design (see Thompson et al. 1998; Buckland et al. 2001; Borchers et al. 2002) and satisfaction of critical model assumptions. Ideally these aspects should be evaluated within a pilot study before beginning any comprehensive monitoring programme. In addition to addressing survey design, potential bias and critical assumptions of the method (see below), estimates of transect length or number of points required to provide predetermined levels of precision (i.e. power) can also be calculated once a pilot study has been done.

Three critical assumptions must be met (or at least the potential impact of their failure evaluated) before valid density estimates can be assumed. In order of greater to lesser importance these assumptions (as well as the effect of failure to meet them) are:

1. All objects of interest present on the transect line or point are detected (P = 1). Failure to detect all objects will significantly underestimate density.
2. Objects of interest do not move prior to detection during a count. Undetected movements toward the line or point will overestimate density whereas those movements away from the line or point will underestimate density.
3. Distances from a transect line or point to objects of interest are accurately measured or recorded within the correct distance interval. Systematic overestimates or underestimates will produce biased results.

Difficulties in meeting these assumptions (especially the first two) should not be underestimated, particularly in forested environments where the objects of interest such as birds are often highly mobile (Dawson & Bull 1975; Dawson 1981; Hutto & Young 2002, 2003; Bart et al. 2004). In all likelihood, the ability to meet these assumptions will be compromised to some degree. Assessment of the potential for assumption violation on a case-by-case basis along with the establishment of specific sampling protocols will therefore be required. A number of approaches have been devised to account for failures of the first two assumptions, but these are generally specific to marine or aerial surveys (Borchers et al. 1998; Buckland et al. 2001, 2004) and as such may be difficult to apply to surveys of terrestrial mammals and birds (Thompson 2002).

The situation is further complicated if the intent is to monitor multiple species, particularly if the behaviour of individual species varies markedly and/or the status of the various species being monitored ranges from rare to common. If too few individuals are detected, the construction of an accurate detection function will be compromised and the density estimates derived from it inaccurate. For example, distance sampling may be appropriate for species such as kaka and kereru that generally remain stationary during the count period and are often quite noisy (when perched or in flight), provided they occur in reasonable numbers. Smaller forest birds, particularly those that are highly mobile (e.g. tui and silvereyes) and/or move toward or away from observers

[4] http://www.ruwpa.st-and.ac.uk/distance/

prior to detection (e.g. robins and blackbirds), are problematic and may seriously bias (negatively or positively) calculated density estimates.

Choice of line transects or points (also called point transects and variable circular plots) is largely dependent on their appropriateness given the topography and vegetation communities in the sampling area, and the species or object of interest being counted (Thompson et al. 1998; Borchers et al. 2002). Generally speaking, line transects are considered more efficient than point counts, particularly in more open habitats, as information is being collected continually and a smaller number of detection distances is required to estimate density. Surveys based on points are more effective in rugged, densely vegetated and fragmented or heterogeneous landscapes. They lead to improved detectability and distance estimation. Regardless of the sampling method chosen, it is critical that a legitimate spatial sampling framework (random, systematic, stratified, etc.) is utilised. We strongly urge readers to review the comprehensive material relating to survey design of distance sampling programmes found in Buckland et al. (2001, 2004).

Mark-recapture/resight/removal

Mark-recapture studies can be used whenever animals or plants can be individually marked or identified, a complete count is not possible, or when less expensive indices of relative abundance or density are not desirable or cannot be calculated. Individuals are marked, recaptured either once or several times after a short interval, then mark-recapture analyses are used to estimate population size. Marks must always be applied ethically and humanely. The marked object of interest may not need to be physically recaptured each time, just sighted, provided it can be identified accurately. Marks are generally specific to individuals (i.e. they are a unique identifier for each individual). Examples of individual marks are metal bands with unique numbers for birds and bats, unique colour combination bands for birds, transponders and radio-transmitters.

There are various methods available for estimating population size (see below). Factors that influence choice of method are whether (a) temporary and permanent marks are to be used (b) individually recognisable marks can be applied, (c) the population is open or closed during the sampling period, (d) study objectives can be met under a proposed method and (e) adequate resources and time are available for the study using that method.

In all cases, the advice of a statistician with expertise in modelling populations should be sought before undertaking mark-recapture studies to ensure an appropriate approach is used, determine how the most critical assumptions can be met, and assess what impact assumption violation might have on the precision of population estimates and the conclusions drawn from analyses (see table 1 in Lettink & Armstrong 2003).

Regardless of the chosen method, all mark-recapture models have assumptions that must be satisfied if biased estimates are to be avoided. Lettink & Armstrong (2003) provide a useful description of those assumptions common to all methods. These can be summarised as follows:

- Marked animals are representative of the population being studied.
- Marks do not influence behaviour or survival of marked animals.
- Loss of contact with a marked animal is random and independent of death.

Additional assumptions are required depending on the design of the mark-recapture study and will determine what the results can be used for. Four design approaches are possible:

1. Calculating very simple estimates (minimum number of animals or plants alive (MNA) following a recapture session.
2. Deriving population estimates from short-term sampling of closed populations (i.e. where there is no birth, death, immigration and emigration during the study) and obtaining an estimate of absolute abundance (Lincoln-Petersen estimators, program CAPTURE, mark-resight).
3. Using the recapture probabilities calculated in open-population models to estimate population size (Cormack-Jolly-Seber estimators).
4. Using robust design models that combine elements of both open- and closed-population modelling.

The following text provides brief summaries of the most commonly used approaches—the second approach (items a and b below) and third approach (item c below).

(a) Closed population mark-recapture

Population size can be estimated whenever the ratio of marked to unmarked objects of interest within a demographically and spatially closed population can be calculated. A closed population remains constant in size and composition and is not subject to individuals entering and leaving the population through births, deaths, emigration and immigration. Simple estimators derived from two sampling occasions and more complex multiple occasion estimators usually involve the capture and release of unmarked animals on each sampling occasion (Williams et al. 2002). Release of animals marked on successive sampling occasions increases the number of marked animals within a population and increases the accuracy and precision of the population estimates.

Closed population mark-recapture (capture-recapture) methods are recommended if a single estimate of abundance is required, whereas the 'robust model' provides more precise abundance estimates when calculating multiple estimates of abundance (*cf.* Schnabel estimator, Burnham & Overton method as described in Sutherland 2006). Choosing between the various methods of analysis is highly dependent on the researcher's knowledge of the animal or plant being studied and which of the various assumptions are most likely to be violated. A pilot study will provide significant insight about assumptions and reveal most practical issues of capturing, marking and resighting a given species. Such a study will also indicate the expected precision of a proposed method and a point at which a choice can be made to abandon the study if the cost of obtaining the desired precision is beyond the available resources (Sutherland 2006).

If there is only one marking session and one recapture session, and the population is closed, then the Petersen estimator or Lincoln index is applicable. This Lincoln-Petersen estimator assumes that:

- The population is geographically and demographically closed, with no births, deaths, immigration or emigration during the study, i.e. the population is constant in size and composition during the study period.

- All animals have the same probability of being caught.
- Marks are not lost.

If there are multiple recapture sessions and the population is still considered to be closed (i.e. sampling is conducted over relatively short time periods), then the second assumption above need not apply. Program CAPTURE[5] (Otis et al. 1978) or program MARK[6] (White & Burnham 1999) can be used to fit these more complex and versatile models, compare their performance (Burnham & Anderson 2002), and estimate population size using the most appropriate model.

(b) Closed population mark-resight estimators

If marks can be observed without recapturing the animals, a second capture session is unnecessary. The initial capture event might also be unnecessary if animals can be identified from their appearance or DNA profiles. A sample of animals can simply be observed and the number of marked and unmarked individuals counted. For this method to work, the marks have to be conspicuous enough to be readily observable. Marking systems suitable for this method include coloured leg bands and radio-transmitters. Estimates of precision can be improved by counting marked and unmarked animals on a number of occasions, as long as the assumption of population closure holds over each survey period (covering all sighting occasions). The assumption the population is closed can be checked by testing for a declining trend in the proportion of marked individuals—an increasing number of unmarked animals suggests immigration, emigration, births or deaths are occurring and that the population is open rather than closed (Sutherland 2006).

Program NOREMARK calculates population estimates based on resightings of individually marked individuals. Four estimators of abundance are provided along with simulation routines to assist with the design of mark-resight sampling programmes. The main limitation of this method (compared with other mark-recapture studies in which newly marked animals are released on each occasion) is that unmarked animals are not marked on subsequent occasions, thereby constraining sample size and reducing estimate precision. However, the advantage of this procedure is that 'resights' are cheaper to acquire than physically catching and handling animals (Thompson et al. 1998). Mark-resight procedures are only practical in situations where objects of interest are reasonably sedentary, readily identifiable, immobile objects like nests, where they occur in discrete habitats, or when radio-telemetry is a feasible option (Thompson et al. 1998).

(c) Open population models

Open population models (those subject to births, deaths, immigration and emigration during a study) are usually used for rigorous analysis of survival in animals. Because it is necessary to estimate a large number of parameters (the key ones being survival probability Φ and capture probability P) the precision of population estimates is often poor. Improving precision is usually reliant on catching a substantial proportion of the population on each sampling occasion and doing this regularly enough to ensure a high recapture rate. This can be extremely difficult. However, if recapture probabilities are high, and the assumptions of the model are met, then the recapture

[5] http://137.227.242.23/software/capture.html
[6] http://www.phidot.org/software/mark/index.html

probabilities calculated from these models can be inserted into the equations for estimating population size. The Jolly-Seber model (and its variant, Cormack-Jolly-Seber) is the most commonly used. This model has the following main assumptions in addition to those for closed populations:

- Every animal (of the same type) has the same probability of recapture (‘equal catchability’ or ‘capture heterogeneity’).
- Every animal (of the same type) has the same probability of survival from one sample to the next.
- Marks are not lost or missed.
- All samples are instantaneous and each release is made immediately after the sample.

As multiple parameters are estimated from single data sets, care must be taken when interpreting calculated population estimates. Variance estimates are typically positively correlated with the parameter estimates (i.e. small estimates have small variances, whereas large estimates have large variances) thus making underestimates look better than they really are (Thompson et al. 1998). The Jolly-Seber estimator is also sensitive to capture heterogeneity (i.e. differences in capture probability between individuals) and changes in animal behaviour (trap ‘shyness’ and trap ‘happiness’ or trap attraction). Although, in theory, capture heterogeneity can be minimised by increasing capture probabilities to 0.5 or more, in practice this can be extremely difficult or impossible in many instances. Changes in animal behaviour remain problematic (Thompson et al. 1998).

Williams et al. 2002 provide alternatives to the Jolly-Seber approach, including robust models and a useful summary discussion of assumptions and their implications for the design of mark-recapture studies.

Sampling methods not covered

- Removal, catch effort and change in ratio.

References

Barraclough, R.K. 2000: Distance sampling: a discussion document produced for the Department of Conservation. *Science and Research Internal Report 175*. Department of Conservation, Wellington. 26 p.

Bart, J.; Droege, S.; Geissler, P.; Peterjohn, B.; Ralph, C.J. 2004: Density estimation in wildlife surveys. *Wildlife Society Bulletin 32*: 1242–1247.

Bart, J.; Earnst, S.L. 2002: Double sampling to estimate bird density and population trends. *Auk 119*: 36–45.

Best, L.B.; Fowler, R. 1981: Infrared emissivity and radiant surface temperatures of Canada and snow geese. *Journal of Wildlife Management 45*: 1026–1029.

Bibby, C.J.; Burgess, N.D.; Hill, D.A.; Mustoe, S. 2000: Bird Census Techniques (second edition). Academic Press, London. 302 p.

Borchers, D.L.; Buckland, S.T.; Goedhart, P.W.; Clarke, E.D.; Hedley, S.L. 1998: Horvitz-Thompson estimators for doubleplatform line transect surveys. Biometrics 54: 1221–1237.

Borchers, D.L.; Buckland, S.T.; Zucchini, W. 2002: Estimating animal abundance: closed populations. Springer-Verlag, London. 314 p.

Bradshaw, C.J.; Lalas, C.; Thompson, C.M. 2000: Clustering of colonies in an expanding population of New Zealand fur seals (*Arctocephalus forsteri*). *Journal of Zoology 250*: 105–112.

Buckland, S.T. 2006: Point transect surveys for songbirds: robust methodologies. *Auk 123*: 345–357.

Buckland, S.T.; Anderson, D.R.; Burnham, K.P.; Laake, J.L.; Borchers, D.L.; Thomas, L. 2001: Introduction to distance sampling: estimating abundance of biological populations. Oxford University Press, Oxford. 432 p.

Buckland, S.T.; Anderson, D.R.; Burnham, K.P.; Laake, J.L.; Borchers, D.L.; Thomas, L. 2004: Advanced distance sampling: estimating abundance of biological populations. Oxford University Press, Oxford. 41 p.

Burnham K.P.; Anderson D.R. 2002: Model selection and multimodel inference: a practical information-theoretic approach. 2nd edition. Springer-Verlag, New York. 488 p.

Conroy, M.J.; Carroll, J.P. 2001: Estimating abundance of Galliformes: tools and their application. Pp 204–215 in Woodburn, M.I.A.; McGowan, P.J.K.; Carroll, J.P.; Musavi, A.H.; Zhang Zheng-wang (Eds): Galliformes 2000: Proceedings of the 2nd International Galliformes Symposium. King Mahendra Trust for Nature Conservation, Kathmandu, Nepal and World Pheasant Association, Reading, UK.

Conroy, M.J.; Noon, B.R. 1996: Mapping of species richness for conservation of biological diversity: conceptual and methodological issues. *Ecological Applications 6*: 763–773.

Dawson, D.G. 1981: The usefulness of absolute (census) and relative (sampling or index) measures of abundance. Pp. 554–558 in Ralph, C.J.; Scott, J.M. (Eds): Estimating numbers of terrestrial birds. *Studies in Avian Biology No. 6*. Cooper Ornithological Society, Allan Press, Kansas.

Dawson, D.G.; Bull, P.C. 1975: Counting birds in New Zealand forests. *Notornis 22*: 101–109.

de Lange, P.J.; Norton, D.A.; Crowcroft, G.M. 2000: Taxonomy, ecology and conservation of *Atriplex billardierei* and *A. hollowayi* sp. nov. (Chenopodiacae) in Australasia. *New Zealand Journal of Botany 38*: 551–567.

Dilks, P.; Willans, M.; Pryde, M.; Fraser, I. 2003: Large scale stoat control to protect mohua (*Mohoua ochrocephala*) and kaka (*Nestor meridionalis*) in the Eglinton Valley, Fiordland, New Zealand. *New Zealand Journal of Ecology 27*: 1–9.

Dowding, J.E.; Davis, A.M. 2007. New Zealand dotterel (*Charadrius obscurus*) recovery plan 2004–2014. Department of Conservation, Wellington. 28 p.

Ellingson, A.R.; Lukacs, P.M. 2003: Improving methods for regional landbird monitoring: a reply to Hutto and Young. *Wildlife Society Bulletin 31:* 896–902.

Field, S.A; O'Connor, P.J.; Tyre, A.J.; Possingham, H.P. 2007: Making monitoring meaningful. *Austral Ecology 32*: 485–491.

Flux, I.; Innes, J. 2001: Kokako management folder. *Threatened Species Occasional Publication 19*. Department of Conservation, Wellington. 73 p.

Forcey, G.M.; Anderson, J.T.; Ammer, F.K.; Whitmore, R.C. 2006: Comparison of two double-observer point-count approaches for estimating breeding bird abundance. *The Journal of Wildlife Management 70*: 1674–1681.

Fuller, R.J.; Stuttard, P.; Ray, C.M. 1989: The distribution of breeding songbirds, within mixed coppiced woodland in Kent, England in relation to vegetation age and structure. *Annales Zoologici Fennici 26*: 265–275.

Gerrodette, T. 1987: A power analysis for detecting trends. *Ecology 68*: 1364–1372.

Greene, B.S. 2003: Fence removal benefits gannets (*Morus serrator*) at Muriwai, northwest Auckland, New Zealand. *Notornis 50*: 100–102.

Greene, T.C.; Fraser, J.R. 1998: Sex ratio of North Island kaka (*Nestor meridionalis septentrionalis*), Waihaha Ecological Area, Pureora Forest Park. *New Zealand Journal of* Ecology *22*: 11–16.

Haigh, A.; Pledger, S.; Holzapfel, S. 2007: Population monitoring programme for Archey's frog (*Leiopelma archeyi*): pilot studies, monitoring design and data analysis. *DOC Research and Development Series 278*. Department of Conservation, Wellington. 25 p.

Hill, D.; Fasham, M.; Tucker, G.; Shrewry, M.; Shaw, P. 2005: *Handbook of biodiversity methods. Survey, evaluation and monitoring*. Cambridge University Press, Cambridge.

Hutto, R.L.; Young, J.S. 2002: Regional landbird monitoring: perspectives from the Northern Rocky Mountains. *Wildlife Society Bulletin 30*: 738–750.

Hutto, R.L.; Young, J.S. 2003: On the design of monitoring programs and the use of population indices: a reply to Ellingson and Lukacs. *Wildlife Society Bulletin 31*: 903–910.

Innes, J.; Hay, R.; Flux, I.; Bradfield, P.; Speed, H.; Jansen, P. 1999: Successful recovery of North Island kokako *Callaeas cinerea wilsoni* populations, by adaptive management. *Biological Conservation 87*: 201–214.

Joseph, L.N.; Field, S.A.; Wilcox, C.; Possingham, H.P. 2006: Presence-absence versus abundance data for monitoring threatened species. *Conservation Biology 20*: 1679–1687.

Kelly, D.; Jasperse, J.; Westbrooke, I. 2005: Designing science graphs for data analysis and presentation—the bad the good and the better. *Department of Conservation Technical Series 32*. Department of Conservation, Wellington. 68 p. http://www.doc.govt.nz/upload/documents/science-and-technical/docts32.pdf

Kissling, M.L.; Garton, E.O. 2006: Estimating detection probability and density from point-count surveys: a combination of distance and double-observer sampling. *The Auk 123*: 735–752.

Lee, W.; McGlone, M.; Norbury, G.; Overton, J.; Rutledge, D.; Wiser, S.; Wright, E.; Stephens, T.; O'Donnell, C. 2004: Department of Conservation, national biodiversity inventory and monitoring programme: scoping document. Landcare Research Contract Report (LC0405/014), Landcare Research Ltd, Lincoln. 33 p.

Lee, W.; McGlone, M.; Wright, E. 2005: Biodiversity inventory and monitoring: A review of national and international systems and a proposed framework for future biodiversity monitoring by the Department of Conservation. Landcare Research Contract Report (LC0405/122), Landcare Research Ltd, Lincoln. 216 p.

Lettink, M., Armstrong, D.P. 2003: An introduction to using mark-recapture analysis for monitoring threatened species. Pp. 5–32 in: Mark-recapture analysis for monitoring threatened species: introduction and case study. *Department of Conservation Technical Series 28*. Department of Conservation, Wellington.

Link, W.A.; Sauer, J.R. 1997: New approaches to the analysis of population trends in land birds: comment. *Ecology 78*: 2632–2634.

Link, W.A.; Sauer, J.R. 2002. A hierarchical analysis of population change with application to cerulean warbler. *Ecology 83*: 2832–2840.

MacKenzie, D.I.; Nichols, J.D.; Lachman, G.B.; Droege, S.; Royle, J.A.; Langtimm, C.A. 2002: Estimating site occupancy rates when detection probabilities are less than one. *Ecology 83*: 2248–2255.

MacKenzie, D.I.; Nichols, J.D.; Royle, J.A.; Pollock, K.H.; Bailey, L.L.; Hines, J.E. 2006: Occupancy estimation and modelling: Inferring patterns and dynamics of species occurrence. Elsevier Inc., San Diego. 324 p.

Maloney, R.F. 1999: Bird populations in nine braided rivers in the upper Waitaki Basin, South Island, New Zealand: changes after 30 years. *Notornis 46*: 243–256.

Maloney, R.F.; Rebergen, A.L.; Nilsson, R.J.; Wells, N.J. 1997: Bird density and diversity in braided river beds in the Upper Waitaki Basin, South Island, New Zealand. *Notornis 44*: 219–232.

Manly, B.F.J. 2001: Statistics for environmental science and management. Chapman and Hall, Florida.

Marchant, J.H.; Hudson, R.; Carter, S.P.; Whittington, P. 1990: Population Trends in British Breeding Birds. BTO, Tring, Herts, UK.

Merrett, M.F.; Ecroyd, C.E.; Cashmore, P.B. 2001: Ecology of *Calochilus robertsonii* (Orchidaceae) from Rotorua, New Zealand. *The Orchadian 13*(8): 375–381.

Miller, C.; Allan, C. 2002: Parks Victoria's draft monitoring guide: environmental monitoring in the Parks Victoria estate. Parks Victoria, Melbourne.

Moore, P.J. 2004: Abundance and population trends of mollymawks on Campbell Island. *Science for Conservation* 242. Department of Conservation, Wellington. 62 p.

Morrison, M.L.; Block, W.M.; Strickland, M.D.; Kendall, W.L. 2001: Wildlife study design. Springer-Verlag, New York. 211 p.

Nichols, J.D.; Hines, J.E.; Sauer, J.R.; Fallon, F.W.; Fallon, J.E.; Heglund; P.J. 2000: A double-observer approach for estimating detection probability and abundance from point counts. *Auk 117*: 393–408.

Nichols, J.D.; Williams, B.K. 2006: Monitoring for conservation. *Trends in Ecology and Evolution 21*: 668–673.

O'Donnell, C.F.J. 1996: Predators and the decline of New Zealand birds: an introduction to the hole-nesting bird and predator programme. *New Zealand Journal of Zoology 23*: 213–220.

O'Donnell, C.F.J.; Dilks, P.J. 1986: Forest birds in South Westland: status, distribution and habitat use. *Occasional Publication No. 10*, New Zealand Wildlife Service, Department of Internal Affairs, Wellington. 179 p.

O'Donnell, C.F.J.; Langton, S. 2003: Power to detect trends in abundance of long-tailed bats (*Chalinolobus tuberculatus*) using counts on line transects. *Science for Conservation 224*. Department of Conservation, Wellington.

Otis, D.L.; Burnham, K.P.; White, G.C.; Anderson, D.R. 1978: Statistical inference for capture data from closed populations. *Wildlife Monographs 62*. 135 p.

Peltzer, D.A ; Allen, R.B; Rogers, G.M. 2005: Dieback and recruitment of the forest dominants *Nothofagus fusca* and *Libocedrus bidwillii*, central North Island, New Zealand. *Science for Conservation 255*. 33 p.

Powlesland, R.G.; Knegtmans, J.J.W.; Marshall, I.S.J. 1999: Costs and benefits of aerial 1080 possum control operations using carrot baits to North Island robins (*Petroica australis longipes*), Pureora Forest Park. *New Zealand Journal of Ecology 23*: 149–159.

Purcell, K.L.; Mori, S.R.; Chase, M.K. 2005: Design considerations for examining trends in avian abundance using point counts: Examples from oak woodlands. *The Condor 107*: 305–320.

R Development Core Team. 2011: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.

Raal, P.; Pearce, J.; Pearce, S. 2005: Scattered wilding conifer surveillance, control and monitoring in the Otago Conservancy. Unpublished report, Department of Conservation, Otago Conservancy, Dunedin.

Ralph C.J.; Dunn E.H. (Eds). 2004: Monitoring bird populations using mist nets. *Studies in Avian Biology No. 29*.

Richardson, S.J.; Allen, R.B.; Whitehead, D.; Carswell, F.E.; Ruscoe, W.A.; Platt, K.H. 2005: Climate and net carbon availability determine temporal patterns of seed production by *Nothofagus*. *Ecology 86*: 972–981.

Robertson, C.J.R.; Hyvönen, P.; Fraser, M.J.; Pickard, R. 2007: Atlas of bird distribution of New Zealand. Ornithogical Society of New Zealand. 533 p.

Rosenstock, S.S; Anderson, D.R.; Giesen, K.M.; Leukering, T.; Carter, M.F. 2002: Landbird counting techniques: Current best practice and an alternative. *Auk 119*: 46–53.

Ross, P. 2002. Natural heritage inventory and monitoring project—business case. Department of Conservation, Christchurch. 56 p.

Royle, J.A. 2004: N-mixture models for estimating population size from spatially replicated counts. *Biometrics 60*: 108–115.

Royle, J.A.; Dawson, D.K.; Bates, S. 2004: Modeling abundance effects in distance sampling. *Ecology 85*: 1591–1597.

Sagar, P.M.; Shankar, U.; Brown, S. 1999: Distribution and numbers of waders in New Zealand, 1983–1994. *Notornis* 46: 1–43.

Schauber, E.M.; Kelly, D.; Turchin, P.; Simon, C.; Lee, W.G.; Allen, R.B.; Payton, I.J.; Wilson, P.R.; Cowan, P.E.; Brockie, R.E. 2002: Synchronous and asynchronous masting by 18 New Zealand plant species: the role of temperature cues and implications for climate change. *Ecology 83*: 1214–1225.

Schmechel, F.A.; O'Connor, S. 1999: Distribution and abundance of the Chatham Island Oystercatcher (*Haematopus chathamensis*). *Notornis 46*: 155–165.

Scott, J.M.; Mountainspring, S.; Ramsey, F.L.; Kepler, C.B. 1986: Forest bird communities of the Hawaiian Islands: their dynamics, ecology, and conservation. *Studies in Avian Biology No. 9*.

Secretariat of the Convention on Biological Diversity. 2005: Handbook of the Convention on Biological Diversity Including its Cartagena Protocol on Biosafety, 3rd edition, Montreal. 1493 p.

Sherley, G.H.; Stringer, I.A.N.; Parrish, G.R.; Flux, I. 1998: Demography of two land snail populations (*Placostylus ambagiosus*, Pulmonata: Bulimulidae) in relation to predator control in the Far North of New Zealand. *Biological Conservation 84*: 83–88.

Steidl, R.J.; Hayes, J.P.; Schauber, E. 1997: Statistical power analysis in wildlife research. *Journal of Wildlife Management 61*: 270–279.

Sutherland, W.J. (Ed.). 2006: Ecological census techniques; a handbook (second edition). Cambridge University Press, Cambridge. 336 p.

Thomson, F.R,; Schwalbach, M.R. 1995: Analysis of sample size, counting time, and plot size from an avian point count survey on Hoosier National Forest, Indiana. Pp. 45–48 in Ralph, C.J.; Sauer, J.R.; Droege, S. (Eds): Monitoring bird populations by point counts. USDA For. Ser. Gen. Tech. Rep. PSW-GTR-149, Pacific Southwest Res. Stn., Albany, California.

Thompson, S.K. 1992: Sampling. Wiley, New York.

Thompson, W.L. 2002: Towards reliable bird surveys: accounting for individuals present but not detected. *Auk 119*: 18–25.

Thompson; W.L., White; G.C., Gowan, C. 1998: Monitoring vertebrate populations. Academic Press, Inc., San Diego. 365 p.

Verner, J. 1985: Assessment of counting techniques. *Current Ornithology 2*: 247–302.

White, G.C.; Burnham, K.P. 1999: Program MARK: survival estimation from populations of marked animals. *Bird Study 46 Supplement*: 120–138.

Williams, B.K.; Nichols, J.D.; Conroy, M.J. 2002: Analysis and management of animal populations: modelling, estimation and decision making. Academic Press, San Diego. 817 p.

Appendix A

The following Department of Conservation documents are referred to in this method:

olddm-644074	Basic statistics using Excel
olddm-318638	Getting the sample size right
docdm-39000	NVS data entry, archiving and retrieval SOP
docdm-146272	Standard inventory and monitoring project plan
docdm-426700	Using Excel to enter, manage and explore data