avoid unnecessary expense. When surveying a field site, it is important to move quickly and accurately, minimizing unnecessary breaks and ignoring minor discomfort to complete the site on schedule. Above all, the emphasis should be on safety. Surveys should not be continued if dangerous weather conditions develop, sampling equipment poses a risk, or other conditions associated with the sampling are compromised.

## 2.3 Prospectus

Proper planning of surveys and consideration of fundamental issues such as bias, accuracy, and precision can help ensure a useful, valid study (Box 2.2). Biologists who initiate surveys without considering the basics of survey design risk embarrassment and the outlay of considerable funds with few results. Those who carefully plan their surveys from the start will provide information that can advance science, influence politics, or shape laws and policies.

**Box 2.2 Guidelines for sampling**

- Develop clear, detailed objectives for the survey.
- Define the sampling frame—select what, where, and how to sample.
- Seek advice from statisticians when designing surveys.
- Consider how bias and sampling error will affect estimates. Correct for bias (calibrate sampling) if possible and maximize numbers of samples collected and allocate samples in time and space to reduce sampling error.
- Incorporate some form of randomization when selecting samples.
- Chose methods to sample that minimize impact to the organisms studied.
- When possible, sample using standard techniques so results among studies can be compared.
- Incorporate time and planning for obtaining sampling permits and conducting safe surveys, and develop contingency plans for when things go wrong in the field.
- When in doubt about survey design or logistics, refer back to survey objectives.

Comparison across large regions and communication among diverse researchers are becoming increasingly important, therefore the adoption and development of standard sampling techniques will play even a greater role in the future, similar to the standard methods already developed for climate science, water chemistry, geology, and medicine.

With the advent of technology, analysis procedures, and automation, many sampling procedures will become easier over time. However, the basic considerations will remain the same. A fundamental knowledge of basic sampling issues will help those surveying for biological diversity, no matter what tools are available in the future.

## 2.4 Key points

1. Set clear objectives for surveys, with a carefully defined sampling frame.
2. Design the survey, seeking help from a statistician when needed. Define when, where, and how to survey, using methods that minimize mortality to surveyed organisms and incorporate some form of randomness.
3. Precision and bias affect the utility of sampling estimates. Account for the effects of bias by using the same techniques over time to monitor trends, using a variety of gears with different bias to cancel species-related bias for point estimates, or, the best way, by validating sampling techniques with true population parameters. Maximize precision by increasing sample sizes and careful consideration of sample allocation over space and time.
4. Pilot surveys can help define the sampling effort required and identify the logistical challenges the main survey will face.
5. Sample using established standard sampling methods when possible to maximize the comparability of the data among studies. Ensure safe surveys and use field etiquette, including obtaining sampling permits in advance, to maximize survey efficiency.

CHAPTER 3

# Biodiversity monitoring: the relevance of detectability

**Stephen T. Buckland, Angelika C. Studeny, Anne E. Magurran, and Stuart E. Newson**

## 3.1 Introduction

Most studies that monitor biodiversity are deficient in an important respect: they assume that the captured or observed animals or plants form the community of interest, or at least are a representative sample from it. For single-site monitoring, if the site is small and individuals of the taxa of interest are easily detected, it might not be unreasonable to conclude that the counts correspond to the community of interest. However, for more cryptic species and/or larger sites, complete enumeration of individuals is seldom feasible (Coddington et al. 2009). Most diversity measures assume that the individuals we count are a random sample of the community (Magurran 2004, p136). It might be possible to ensure this if sampling is conducted on a sufficiently large random sample of plots within the site, where each plot is small enough to allow complete enumeration of individuals on the plot. If this is not achievable, we can expect heterogeneity among individuals in probability of detection.

Increasingly, in part due to the change of emphasis for biodiversity monitoring signalled by the 2010 Biodiversity Target of the Convention on Biological Diversity (which states that there should be 'a significant reduction of the current rate of biodiversity loss' by 2010), the focus is on quantifying biodiversity trends within regions or countries, in which case the community should be the entire biota of the region or country. Since we cannot record the entire biota directly, we must sample locations and species (Chapter 2), with the aim of drawing inference on the entire biota. If survey sites or plots are selected according to a randomized scheme, such as the stratified random scheme of the UK Breeding Bird Survey (Newson et al. 2005), the issue of representativeness of sites is addressed. Representativeness of species is more problematic, especially when complete species lists do not exist, and some species and taxa are much more difficult to survey than others. Hence, biodiversity monitoring will always be based on a reduced set of taxa. Within that set, a random subset of species might be an appropriate means of easing the task of monitoring.

Even after addressing these issues, we need to consider the issue of detectability, unless our plots are so small that all individuals can be detected and so numerous that all, or nearly all, species occur on at least one sample plot. We use the term 'probability of detection' for the probability that an individual is recorded. Recording of individuals might be achieved by visual or aural detection, or by capturing them in traps, nets, etc. A combination of methods might be used. Typically, we can expect detectability to vary according to a number of factors, but the one that interests us here is when detectability of individuals varies among species, as this will bias the species proportions used in many biodiversity measures (Yoccoz et al. 2001).

In this chapter, we consider the implications when detectability varies by species, with particular reference to monitoring regional biodiversity trends. We consider some measures of biodiversity that are potentially useful for quantifying trends in regional biodiversity and explore how these measures are affected when the issue of detectability is ignored.

Although our focus is on regional monitoring, there are parallels with site monitoring. For example, we might seek to quantify the biodiversity of a site by sampling a number of small plots within the site. In this case, we do not require that each plot is sufficiently large to reflect the entire biodiversity of the site, but we expect the plots collectively to achieve this. If it is possible to make each plot sufficiently small to allow complete enumeration of the plot, then the issue of detectability of individuals on the sample plots becomes irrelevant. If, however, there are too few plots, many species present in the site will not be detected because they do not occur on any of the plots. Thus we need a large sample of plots, but we do not need large volumes of data from each plot, to quantify reliably the biodiversity of the site. The same issues apply when we seek to quantify the biodiversity of a region by sampling many sites within a region. Note that, for a regional survey, we might adopt a hierarchical sampling scheme, with a representative set of sites through the region and a representative set of plots within each site.

We focus primarily on detectability of individuals in this chapter. However, as alluded to above, there is also an issue of detectability of species: if detectability of individuals of a species is very low, or if a species is so rare that few, if any, occur in the sampled sites or plots, then the species may be undetected (Chapter 4). In statistical terms, if there is an adequate sample of plots or sites, this is not a problem, as unrecorded species will represent a small proportion of individuals in the community; the estimated proportion will simply be zero if no individuals are detected. What constitutes an 'adequate sample' of plots or sites depends on the spatial heterogeneity of the environment and on heterogeneity in the detection probabilities (greater heterogeneity requires a larger sample size in both cases). It also depends on the size of sample plots/sites, especially if the site/region is spatially homogeneous with respect to its biodiversity. (In a very heterogeneous environment, we would prefer a large sample of small plots to a small sample of larger plots, as we need to sample more locations to quantify biodiversity of the site or region. In a more homogeneous environment, there would be less gain from sampling more locations and more gain from making plots larger.)

## 3.2 State of the field: which biodiversity measure?

In choosing our biodiversity measures, several considerations must be taken into account. First, we need to choose the component(s) of biodiversity we wish to measure. Buckland et al. (2005) list six criteria that we might like our measure to satisfy (Table 3.1). The non-statistical criteria are related to three components of biodiversity: species richness, species evenness, and abundance. The statistical criteria are that the measure should have good and measurable precision, and have an expectation that does not vary with sample size.

Another consideration is whether trends in biodiversity measures should be sensitive to trends in the rarest species of a community. Rare species are an important positive contributor to species richness, although they make it difficult to estimate (Colwell & Coddington 1994; Chapter 4). However, an increase in the number of rare species reduces species evenness and rare species generally have a negligible contribution to abundance.

If rare species are excluded from analyses, the geometric mean of relative abundance indices, as used by Gregory et al. (2003), has good properties,

**Table 3.1** The six criteria proposed by Buckland et al. (2005) for a biodiversity measure.

| Criterion | Description |
|---|---|
| 1 | For a system that has a constant number of species, overall abundance, and species evenness, but with varying abundance of individual species, the index should show no trend. |
| 2 | If overall abundance is decreasing, but number of species and species evenness are constant, the index should decrease. |
| 3 | If species evenness is decreasing, but number of species and overall abundance are constant, the index should decrease. |
| 4 | If number of species is decreasing, but overall abundance and species evenness are constant, the index should decrease. |
| 5 | The index should have an estimator whose expected value is not a function of sample size. |
| 6 | The estimator of the index should have good and measurable precision. |



meeting all the criteria of Buckland et al. (2005) except one: if the number of species is declining, the index fails to reflect this, as the rare species are excluded. If they are included, zero counts or abundances are problematic. They can be replaced by small positive values to allow the index to be calculated, but its variance is likely to be high when rare species are analysed (Buckland et al. 2005).

Note that 'rare' species are often defined pragmatically to be those species for which few if any individuals are detected by the sampling programme, so that estimation may be compromised, as for the above method using the geometric mean. Ecologically, it seems unsatisfactory for 'rare' to be determined by sampling effort, rather than by, for example, the proportion of individuals in the community that belong to a given species.

Whether rare species are included or not, Simpson's index performs well with respect to five of the six criteria of Buckland et al. (2005). The one criterion for which it fails relates to abundance: if all species are declining at the same rate, the index remains constant. Strictly, its expectation is a function of sample size. If the form of the index used is $1 - D$, where $D = \sum_i \hat{p}_i^2$ and $\hat{p}_i$ is the estimated proportion of the community that is of species $i$, it can be adjusted to ensure that it is an unbiased measure of $1 - \sum_i p_i^2$, where $p_i$ is the true proportion of the community that is of species $i$. The adjustment is to multiply the index by $n/(n-1)$, where $n$ is the total number of individuals detected (Lande et al. 2003). However, this assumes that the $n$ individuals detected are a simple random sample of the $N$ individuals (across all species) in the community for which we wish to draw inference. Even if we were interested only in local biodiversity, it is usually implausible that detectabilities or trappabilities will be the same across species. In the case of quantifying regional biodiversity, the correction has even less relevance. As $n$ is typically large, the issue is academic: the multiplier is so close to unity that it makes no practical difference.

The Shannon index also generally performs well with respect to five of the six criteria of Buckland et al. (2005). Like Simpson's index, it stays constant if all species are declining at the same rate. The Shannon index is also inferior to Simpson's index with respect to the dependence of its expectation on sample size. Its bias is $-(s-1)/n$, where $s$ is the total number of species in the community (and is generally unknown); hence, for large $n$, the bias can generally be considered negligible. However, if rare species are included in monitoring, and these account for a significant proportion of total abundance, then the Shannon index violates the criterion that its expectation should not be a function of sample size (Lande et al. 2003).

Buckland et al. (2005) proposed a modified form of the Shannon index that satisfies all six of their criteria, provided $n$ is appreciably larger than $s$. However, it has been constructed simply to satisfy the criteria, without any theoretical foundation, and its performance has not been rigorously assessed. In practice, it is generally preferable to use different indices to reflect the different elements of biodiversity, assumed here to be species richness, species evenness, and abundance. Only if a single headline index of biodiversity change is needed might the modified Shannon index be considered.

A related issue relevant to monitoring regional biodiversity is that of $\alpha$ and $\beta$ diversity (Chapter 6). Over large regions, much of the diversity will be due to $\beta$ diversity; that is it arises because species composition and relative abundances differ appreciably among locations within the region. Site-specific diversity measures will reflect $\alpha$ diversity only. By analysing data pooled across sampling locations (in which either species proportions $p_i$ are averaged across sites or counts or estimated abundances are summed across sites), $\gamma$ diversity is estimated. $\alpha$ diversity is estimated by averaging site-specific measures across sites, possibly weighting by number of individuals detected or trapped at each site, and $\beta$ diversity is estimated as either the difference between $\gamma$ and $\alpha$ diversity (additive $\beta$ diversity) or the ratio of $\gamma$ to $\alpha$ diversity (multiplicative $\beta$ diversity). If either the Shannon index or Simpson's index is used, estimated $\beta$ diversity is guaranteed to be positive using the additive definition and to be greater than unity using the multiplicative definition.

## 3.3 Detectability: are species counts relevant for monitoring biodiversity?

### 3.3.1 Individual detectability

All too often, numbers of individuals detected are substituted into equations for measuring diversity with little thought given to whether those numbers give biased estimates of the quantities of interest. Much of the theory underlying biodiversity measures assumes that all individuals in the community are equally likely to be sampled, irrespective of species. In real communities, this will seldom be the case. Some species will be inherently more easily detected or trapped than others. If detectability is not estimated, then biodiversity measures will be biased (see Box 3.1). The theoretical underpinnings claimed for some measures of biodiversity provide little reassurance in reality.

Let us explore the implications if issues of detectability are ignored. We consider the case where we are interested in biodiversity of a region, but the same arguments apply if we are simply interested in a single plot or site. Suppose for a given survey region we have:

- $s$ = number of species in the region
- $N_i$ = number of individuals of species $i$ in the region, $i = 1, \ldots, s$
- $n_i$ = number of individuals of species $i$ detected in the region
- $\pi_i$ = probability of detection of an individual of species $i$ with estimate $\hat{\pi}_i$
- $E(n_i) = N_i\pi_i$ = expected number of individuals of species $i$ detected
- $\hat{N}_i = n_i/\hat{\pi}_i$ = estimated number of individuals of species $i$ in the region
- $p_i = N_i/\sum_{i=1}^{s} N_i$, with corresponding estimate $\hat{p}_i = \hat{N}_i/\sum_{i=1}^{s} \hat{N}_i$
- $q_i = E(n_i)/\sum_{i=1}^{s} E(n_i)$, with corresponding estimate $\hat{q}_i = n_i/\sum_{i=1}^{s} n_i$.

When the true $N_i$ are not known, diversity indices such as Simpson's index ($D = \sum_{i=1}^{s} p_i^2$ or transformations of this) and the Shannon index ($H = -\sum_{i=1}^{s} p_i \log_e p_i$) should be estimated by substituting estimates $\hat{p}_i$ of $p_i$, which requires estimates $\hat{\pi}_i$. Instead, standard practice is to evaluate them by substituting $\hat{q}_i$ for $p_i$. If we can assume that $\pi_i = \pi$ for all $i$, which means that the probability of detection of an individual does not depend on species, then this strategy is justified. The assumption is seldom stated explicitly and almost never criticised, yet it seems unlikely that *any* community comprises individuals that are equally detectable irrespective of species, unless it is possible to enumerate all individuals on plots. If detection is by trapping, individuals of some species are more readily trapped than others; if it is by active searching, individuals of some species are more visible or audible than others. This issue is recognized, for example, in arthropod surveys, where different trapping methods are often used in an attempt to achieve a more representative sample of individuals (e.g. Colwell & Coddington 1994).

If we relax the assumption that all organisms are equally detectable, we can still use the measures of the previous section, but with counts replaced by abundance (or density) estimates. For $\alpha$ diversity, the $\hat{N}_i$ would be estimated for each plot, while for $\gamma$ diversity they would be estimated for the wider survey region, obtained from a combined analysis of the data from all sites in the region (with appropriate allowance for survey design, e.g. if the sites are a stratified random sample, analysis should be appropriately stratified).

### 3.3.2 Estimating individual detectability

The problem remains of how to estimate detectability and hence abundance. The preferred approach will depend on the taxa being sampled and the environment they occupy. The simplest solution is to have sample plots that are sufficiently small that all individuals on the plot are detected. In this case the counts are the plot abundance estimates. Equally, abundance values for the wider region from which the plots were sampled are readily estimated. In the case of simple random sampling, the $\hat{p}_i$ are then the same, whether based on counts or abundance estimates. This approach might be feasible with ground beetles, for example, or for large mammals in open habitat. However, the method is inefficient for most taxa because plots may have to be so small that most counts are zero. A further problem for mobile populations is that an instantaneous count of the plot is needed; if this cannot be achieved, then new individuals may move onto the plot while counting is being conducted, leading to overestimation of density. This overestimation will be greater for more mobile species, generating bias in the diversity measures.

Another solution is to use mark–recapture to estimate detectabilities. The approach tends to be costly and is prone to bias because even within a species, capture probability can be heterogeneous among individuals. This is in part because animals can move on and off the plot, and those with home ranges that are wholly in the plot are much more likely to be trapped than those whose home ranges are largely off the plot. This generates bias in detectability and abundance estimates.

Distance sampling (Buckland 2001) can be an effective way of estimating detectability from sightings surveys (see Box 3.2). For biodiversity monitoring, a problem of distance sampling is that there are typically too few detections for many species to allow estimation of detectability. However, by pooling data across species and then using multiple-covariate distance sampling with species as a factor (Alldredge et al. 2007; Marques et al. 2007) this difficulty is largely avoided, although the detectability of the rarest species with, say, fewer than 10 detections might have to be assumed to be the same as for similar, more common, species (Section 3.4).

Box 3.3 discusses different methods of detectability and distinguishes between individual detectability and species detectability.

### 3.3.3 Species detectability

For rare species, $\hat{q}_i$ and $\hat{p}_i$ will be zero if $n_i = 0$. This too is a detectability issue, but relates to the detectability of a species, rather than to the detectability of individuals. Although many methods exist to estimate the number of



**Box 3.1 If we estimate species proportions by the proportions observed in our sample, are measures based on these proportions biased?**

Intuitively, it might seem that we would not introduce bias into biodiversity measures even when the $\pi_i$ vary among species, provided that $\pi_i$ is independent of $N_i$, that is that the detectability of individuals of a species is independent of species abundance. Thus, although $q_i$ may be smaller than $p_i$ for a less detectable species and larger for a more detectable species, we might hope that these biases compensate, so that the distributions of $q_i$ and $p_i$ are the same. Unfortunately this is not the case. We can see this by considering the distributions across species $i$ of $E(n_i)$ and $N_i$. We would like the distribution of $N_i$ across species to be identical to that of $kE(n_i)$ for some constant $k$ (which cancels top and bottom when $q_i$ is evaluated, giving $q_i = p_i$). Thus we would like the distributions of $\log_e(N_i)$ and $\log_e[E(n_i)]$ to be the same apart from location (which will differ by the factor $\log_e k$). However, $\log_e[E(n_i)] = \log_e(N_i) + \log_e(\pi_i)$, so that $\text{var}\{\log_e[E(n_i)]\} = \text{var}[\log_e(N_i)] + \text{var}[\log_e(\pi_i)] + 2\text{cov}[\log_e(N_i), \log_e(\pi_i)]$. When $\pi_i$ and $N_i$ are independent, the covariance term is zero, and $\text{var}\{\log_e[E(n_i)]\} > \text{var}[\log_e(N_i)]$ except when $\pi_i = \pi$ for all $i$ (in which case $\text{var}[\log_e(\pi_i)] = 0$). Thus if $\pi_i$ and $N_i$ are independent, the distributions are identical apart from location only when detectability does not vary by species. The two distributions have the same variance when $\text{var}[\log_e(\pi_i)] = -2\text{cov}[\log_e(N_i), \log_e(\pi_i)]$, giving $\text{cov}[\log_e(N_i), \log_e(\pi_i)] = -0.5\ \text{var}[\log_e(\pi_i)]$. If the $\pi_i$ are very variable, this corresponds to a strong negative correlation between detectability and abundance.

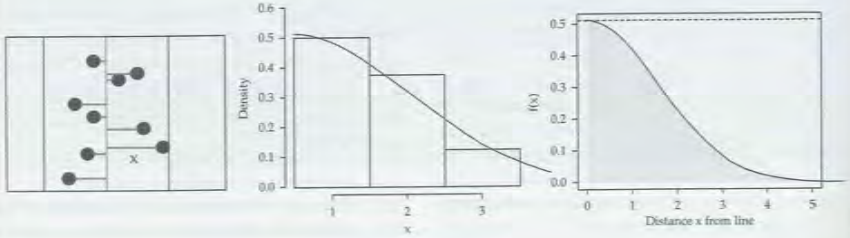
These results indicate that we can generally expect the $q_i$ to be more variable across species than the $p_i$ when detectabilities $\pi_i$ vary across species. Species evenness takes its maximum value when all the $p_i$ are the same, corresponding to zero variance. As variability increases, evenness decreases. Thus, if we estimate $p_i$ by $\hat{q}_i$, we expect to underestimate evenness and hence biodiversity.

**Conclusion**

Measures of biodiversity that are based on species proportions are biased when these are estimated from samples, unless individuals of different species are all equally detectable.

## Box 3.2 Distance sampling

Distance sampling (Buckland 2001, 2004) covers several related methods in which distances of detected individuals from a line or point are recorded, from which the effective area corresponding to a count is estimated. The most commonly used method is line transect sampling, for which lines are placed randomly in the survey region or, more commonly, a set of equally spaced parallel lines is randomly superimposed on the survey region. An observer traverses each line, recording any animals detected within a distance $w$ of the line, together with their distance from the line. These distances are used to estimate a detection function, which is the probability that an animal is detected, as a function of distance from the line. For the basic method, it is assumed that this probability is 1 at zero distance from the line, that is animals on the line are seen with certainty. Given an estimate of the detection function, we can estimate the proportion of animals detected within a strip extending a distance $w$ from the line on either side. This allows us to estimate animal density by adjusting encounter rates (i.e. the number of animals detected per unit length of line) to allow for animals missed in this strip. Given random placement of an adequate number of lines (or a grid of lines) through the survey region, this density estimate is representative of the whole survey region, allowing abundance within that region to be estimated. Probability of detection $\pi_i$ of individuals of species $i$ comprises two components, one reflecting the probability that an animal within a surveyed strip of half-width $w$ is detected and the other reflecting the probability that an animal is within a surveyed strip.




The left-hand plot shows part of a line transect with positions of detected animals shown. The distances $x$ are shown in the form of a histogram with three distance intervals in the middle plot, together with a half-normal detection function fitted to the distance data. For an animal that is within a survey strip, its probability of detection is represented in the right-hand plot by the proportion of the rectangle that is shaded, corresponding to the fitted half-normal function.

This method is used to sample a wide variety of terrestrial, freshwater, and marine animals, as well as some plant populations (e.g. trees in rain forest).

species that are not detected (see Chapter 4), the resulting estimates tend to be particularly sensitive to the assumptions made. Species detectabilities will be even more heterogeneous than individual detectabilities (because individuals of rare species are much less likely to be encountered than individuals of common species), so estimators should accommodate this heterogeneity. Even so, it is a challenging problem to estimate the number of undetected species purely based on data from those species that are detected. Most measures of biodiversity are not unduly affected if some rare species are estimated to have zero abundance. Species richness is an exception to this.

For regional monitoring there are three reasons why a species might not be detected: it might be present on at least one of the sample plots but not detected, it might be present in the region but not on any of the sample plots, and it might be absent from the region at the time of the survey.



## Box 3.3 Which methods of estimating detectability might be appropriate for which taxa?

### Individual detectability

Distance sampling (Buckland 2001, 2004) is useful for species where individuals are readily detectable from a line or point, at least if they are close to the observer. It can be effective for estimating the detectability of medium and large terrestrial mammals, marine mammals, most birds, many reptiles, many fish and benthic fauna, and plants where individual plants are readily identifiable.

Mark–recapture (Borchers et al. 2002; Williams et al. 2002) can be effective when trapping methods or other methods that involve capture are used. Groups for which probability of capture might be quantified using mark–recapture include small mammals, birds in rain forests, reptiles and amphibians, and insects.

In rare circumstances when animals are trapped, a sufficiently large proportion of the population might be removed to allow removal and related catch methods (Borchers et al. 2002) to be used to estimate the trapping probability of individuals.

### Species detectability

The species abundance distribution summarizes the counts of each species observed in a survey, recording the number of species represented by one, two, three, ... individuals. To estimate the species richness of a community we need to estimate the number of species represented by no individuals in the sample. There are many possible methods for this (Magurran 2004; Chapter 4). The methods are applicable to any taxon, although estimation of the number of missing species tends to lack robustness when there is no replicate sampling and individuals of different species have different detectabilities.

When there are repeat visits to each of a number of sites, occupancy methods (MacKenzie et al. 2006; Chapter 4) allow estimation of species richness from simple species lists compiled on each visit to each site. Again, the methods are applicable to any taxon.

Rare species, which can account for a substantial fraction of the species richness in a community (Coddington et al. 2009), may be missed during sampling simply by chance. Undersampling is particularly problematical in tropical arthropod assemblages. Species richness estimators offer one solution (Chapter 4) as long as there is no inherent bias in the sampling methodology. However, rare and common species may differ in ecology (for example by occupying different habitats (Magurran & Henderson 2003)). Moreover, most sampling techniques violate the assumption that individuals are captured or recorded at random. A light trap is a classic example of this since not all moth species respond to the light source in the same way (Southwood & Henderson 2000). It is not just species that vary in their responses—male and female moths may react differently to the light stimulus (Altermatt et al. 2009). Heterogeneity in this sex difference amongst species will contribute to bias. Weather conditions can also affect sampling. Holyoak et al. (1997) found that the total number of moths captured by a light trap was positively correlated with temperature and negatively correlated with rainfall, and that the strength of this correlation varied with taxon. Preston (1948) argued that a moth trap is an acceptable method of sampling because, while species vary in their degree of phototropism, it is random with regard to commonness and rarity amongst those taxa that are equally attracted to the light. However, comparisons of diversity amongst sites, or over time periods, in which the detectability of species varies are challenging.

Species might be missed because they are not present at the time of the survey. Temporal turnover in the form of local colonizations and extinctions is an intrinsic feature of all ecological communities (MacArthur & Wilson 1967). Species are gained and lost by island assemblages while seed banks may contain many species that are only occasionally seen in the 'expressed community'. This raises the conceptual issue of who belongs to the community. Should species that have been recorded in the past, and which may re-occur at some point in the future, be included in a species list? In practice of course it is not always easy to determine which form of absence—methodological or ecological—is involved. These issues are discussed further in Magurran (Chapter 7). Box 3.4 looks at ways in which detectability might be addressed in surveys of butterflies.

**Box 3.4 How might detectability be addressed in surveys of butterflies?**

We use the example of butterflies to illustrate the issues that should be addressed.

In the UK's Butterfly Monitoring Scheme (Pollard 1979), observers walk along transects, placed subjectively through the best butterfly habitat in a site, and record all butterflies entering a box ahead of the observer. Observed species proportions may be biased in such surveys for three reasons: (1) smaller, less active species may pass undetected within the box, (2) conceptually, the method is a snapshot count of butterflies in the box, but for butterflies in flight there will be a flow of butterflies in and out of the box, which biases counts upwards, and this bias is greater for more active species, and (3) the species proportions along the transect may not be representative of the site because the transect is positioned in what is judged to be the best habitat.

The above biases might be removed or at least reduced by placing transects at random through the site and by recording distances from the line of detected butterflies when they are, say, 1 m ahead of the observer (or a distance that is sufficient to ensure that responsive movement of the butterfly does not occur before their position is recorded). By recording the butterfly when it crosses this imaginary line ahead of the observer, bias arising from butterflies in flight is avoided. Detected butterflies that do not cross this line while in sight are not recorded. Line transect methods may be used to analyse the recorded distances, to allow for varying detectability of species, on the assumption that any butterfly on the transect itself will be detected.

The above methods may be workable in relatively open habitats. In more closed habitats, such as rain forest, it may be necessary to use mark–recapture to estimate the catchability of individuals of different species. In this context it is worth noting that bias in the estimates of the probability of capture can be tolerated, provided relative estimates of probability are well estimated. That is, if individuals of one species are twice as likely to be caught as individuals of another, we need to be able to estimate that, but we can tolerate some bias in estimating the actual probabilities of capture.

If species richness is the favoured measure of biodiversity, then species detectability rather than individual detectability becomes relevant. Species-specific counts of butterflies in traps may be used to estimate the number of species present that are not represented in the sample. However, different methods can lead to quite different estimates of this number, that is estimation is not robust.

Given data from repeat visits to multiple sites, occupancy methods are useful for estimating species detectability, and hence species richness (Chapter 11). To apply these methods, we need record only what species were detected at each visit to each site. No counts are needed.

## 3.4 Case study: the UK Breeding Bird Survey

We consider here the Scottish data for the years 1994–2007 from the UK Breeding Bird Survey (BBS) (Freeman et al. 2007; Newson et al. 2008). Plots of size 1 km$^2$ were selected according to a stratified random sampling scheme, with sampling rate proportional to the number of volunteer observers available. In 2007, 405 plots were surveyed.

The survey of each 1 km square is conducted by line transect sampling. Two parallel transects, each of length 1 km, are defined, and the observer walks along them, recording all detected birds. These are recorded in one of four categories: within 25 m of the line, between 25 and 100 m of the line, more than 100 m from the line, or flying over the site. In common with Newson et al. (2008), we use here data for the first two categories only.

If it were possible to detect all birds within 100 m of the line, then the combined count for a square out to 100 m would represent all birds in a plot of size 0.4 km$^2$ (width of 200 m (100 m either side of the line) and length of 2 km, corresponding to the two lines at each site). This would give us a direct estimate of density, from which we can estimate the size of the Scottish population of each recorded species, taking account of the stratified random sampling scheme. For a complete count, we would expect three times as many birds of a species between 25 and 100 m of the line as between 0 and 25 m of the line. The observed difference in counts tends to be less than this. We use the two counts to fit a detection function, which assumes that all birds on the line (distance of zero) are detected, and that the probability of detection falls as distance from the line increases. In our analyses, we assumed that the form of this detection function was half-normal.



For each species, a single half-normal model was fitted across years and sites. The dependence of probability of detection on habitat and year was assessed using the multiple-covariate distance sampling (mcds) engine of Distance (Thomas et al. 2010). The size of the Scottish population in a given year of each species was estimated as

$$\hat{N}_i = \sum_r \frac{A_r}{m_r a}\left[\sum_s \sum_j \frac{1}{\hat{\pi}_{ijsr}}\right]$$

where $\hat{\pi}_{ijsr}$ is the estimated probability of detection of the $j$th detected bird of species $i$ at site $s$ of region (stratum) $r$ for that year, $a = 0.4\,\text{km}^2$ is the size of the covered region in a single site, $m_r$ is the number of sites surveyed in region $r$ in that year, and $A_r$ is the size of region $r$.

We conducted these analyses on species recorded in at least one year in Scotland on BBS squares and classified as 'farmland' species. These classifications are based on Jacobs preference indices, calculated according to Newson et al. (2008). For the rarest species, the detectability was assumed to be the same as for a similar, more common species. The species' common and scientific names are listed in Table 3.2, together with counts and corresponding estimates of the size of the Scottish population in 2007.

For Simpson's index, we define $1 - D_a = 1 - \sum_{i=1}^{s} \hat{p}_i^2$ and $1 - D_c = 1 - \sum_{i=1}^{s} \hat{q}_i^2$. Hence the first estimate is obtained from abundance estimates $\hat{N}_i$, while the second is obtained from mean counts $\bar{n}_i$, where $\bar{n}_i = \frac{\sum_r A_r \bar{n}_{ir}}{\sum_r A_r}$ is a weighted mean count per site of species $i$ to allow for different sampling rates by stratum and $\bar{n}_{ir} = \frac{\sum_s n_{isr}}{m_r}$ is the mean count per site of species $i$ in region $r$, where $n_{isr}$ is the count of species $i$ in site $s$ of region $r$. (We consider here $\gamma$ diversity only; to estimate $\alpha$ diversity, and hence $\beta$ diversity, we would need to estimate the index for each 1 km square, and average these across squares, weighting by the stratum sampling rates.) Similarly for the Shannon index, we define $H_a = -\sum_{i=1}^{s} \hat{p}_i \log_e \hat{p}_i$ and $H_c = -\sum_{i=1}^{s} \hat{q}_i \log_e \hat{q}_i$. We also evaluated the geometric mean of relative abundances estimates, using either abundance estimates $\hat{N}_i$ (in which case relative abundance in a given year for species $i$ is taken as the abundance estimate for that year divided by the abundance estimate for 1994) or mean counts $\bar{n}_i$ (so that the relative abundance for species $i$ in a given year is estimated as the mean count for that year divided by the mean count for 1994). To avoid the problem of calculating a geometric mean when any estimates

**Table 3.2** List of species $i$ comprising the farmland breeding bird community in Scotland. Also shown for each species are the following quantities for 2007 only: estimated mean probability of detection in the surveyed strips $\hat{\pi}_i$, total count $n_i$, mean of counts per site $\bar{n}_i$ (weighted by stratum area), and estimated abundance $\hat{N}_i$ (in thousands). 'UK' indicates that the estimate $\hat{\pi}_i$ was obtained using UK data, as there were too few detections in Scotland alone.

| Species | $\hat{\pi}_i$ | $n_i$ | $\bar{n}_i$ | $\hat{N}_i$ |
|---|---|---|---|---|
| Red-legged partridge, *Alectoris rufa* | 0.51 | 29 | 0.053 | 21.8 |
| Grey partridge, *Perdix perdix* | 0.61 | 35 | 0.049 | 16.6 |
| Common quail, *Coturnix coturnix* | 0.61 | 1 | 0.001 | 0.2 |
| Common pheasant, *Phasianus colchicus* | 0.49 | 314 | 0.562 | 237.3 |
| Common kestrel, *Falco tinnunculus* | 0.44 | 8 | 0.018 | 8.5 |
| Corncrake, *Crex crex* | (0.61) | 0 | 0 | 0 |
| Stock dove, *Columba oenas* | 0.47 | 29 | 0.021 | 9.2 |
| European turtle dove, *Streptopelia turtur* | (0.59)uk | 0 | 0 | 0 |
| Sky lark, *Alauda arvensis* | 0.65 | 1140 | 2.818 | 891.8 |
| Yellow wagtail, *Motacilla flava* | 0.37uk | 2 | 0.005 | 3.1 |
| Common whitethroat, *Sylvia communis* | 0.32 | 194 | 0.438 | 285.9 |
| Rook, *Corvus frugilegus* | 0.78 | 958 | 2.000 | 543.1 |
| Carrion crow, *Corvus corone* | 0.64 | 695 | 1.219 | 398.7 |
| Hooded crow, *Corvus cornix* | 0.53 | 43 | 0.181 | 69.4 |
| Tree sparrow, *Passer montanus* | 0.27 | 72 | 0.084 | 62.4 |
| Brambling, *Fringilla montifringilla* | 0.38 | 1 | 0.001 | 0.5 |
| Common linnet, *Carduelis cannabina* | 0.38 | 359 | 0.869 | 452.3 |
| Twite, *Carduelis flavirostris* | 0.45 | 17 | 0.070 | 33.0 |
| Yellowhammer, *Emberiza citronella* | 0.41 | 411 | 0.935 | 482.0 |
| Corn bunting, *Emberiza calandra* | 0.43 | 11 | 0.005 | 2.6 |

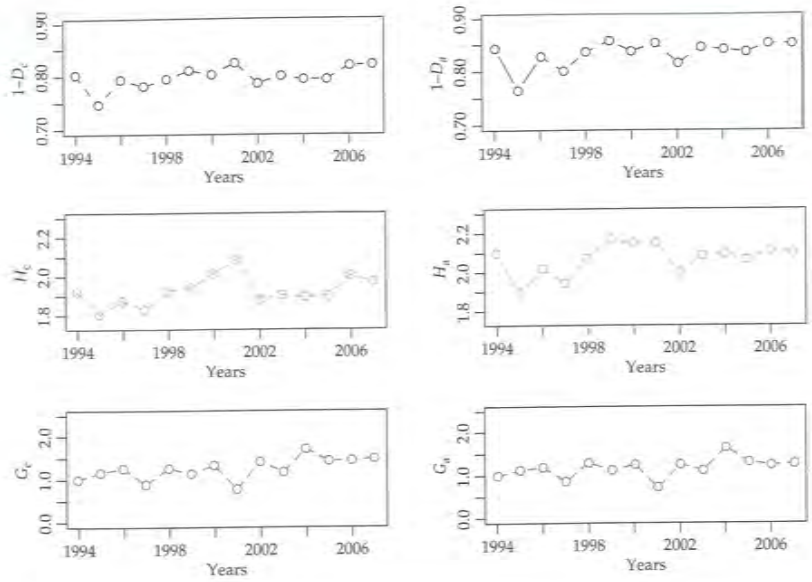



**Figure 3.1** Estimated trends in breeding bird biodiversity in Scottish farmland, 1994–2007, assessed using Simpson's index $1-D$, the Shannon index $H$ and the geometric mean $G$. The indices derived from counts $\hat{n}_i$ uncorrected for detectability are shown on the left-hand side, while those derived from abundance estimates $\hat{N}_i$ are shown on the right-hand side.

are zero, we excluded any species for which there was at least one year during 1994–2007 in which no birds were recorded. These species were red-legged partridge, common quail, corncrake, turtle dove, yellow wagtail, brambling, and corn bunting.

The above indices were evaluated for each year 1994 to 2007 (Fig. 3.1). Tests for linear trend were also conducted (Table 3.3).

As anticipated, diversity as measured by Simpson's or the Shannon index is estimated to be greater when we allow for different detectabilities between species than when we do not (Fig. 3.1). However, the evidence for trend in these indices is largely unaffected (Table 3.3). For the geometric mean based on counts, we obtain evidence of an increase in biodiversity over the time period of the BBS ($p = 0.024$), but if we allow for detectability, the test for linear trend becomes non-significant ($p = 0.126$, Table 3.3).

**Table 3.3** Estimated slope of relationship between each of the following indices and year: Simpson's index ($1-D$), the Shannon index ($H$), and the geometric mean of relative abundance estimates ($G$). For each index, we show estimate (slope), standard error (s.e.) and the $p$ value of the test of no slope for two cases: estimates calculated from counts uncorrected for detectability (subscript $c$) and estimates calculated from abundance estimates (subscript $a$).

| Index | Slope | s.e. | $p$ | Index | Slope | s.e. | $p$ |
|---|---|---|---|---|---|---|---|
| $1-D_c$ | 0.0024 | 0.0011 | 0.058 | $1-D_a$ | 0.0031 | 0.0015 | 0.059 |
| $H_c$ | 0.0072 | 0.0047 | 0.153 | $H_a$ | 0.0076 | 0.0050 | 0.153 |
| $G_c$ | 0.0359 | 0.0139 | 0.024 | $G_a$ | 0.0223 | 0.0136 | 0.126 |

## 3.5 Discussion

Ignoring detectability might not be a major problem if the bias is consistent across time or space. However, if effort varies between times or locations, then bias arising from variable detectability among species will itself be variable, generating bias in temporal or spatial trend estimates in biodiversity. We found clear evidence of bias in estimates of the Shannon and Simpson's indices from the Scottish farmland BBS data, when counts were uncorrected for detectability. In addition, we found that evidence for a trend in relative abundance, as measured by the geometric mean across species, vanished when the relative abundance estimates were corrected for detectability.

The goal of many biodiversity studies is to compare sites or communities that reflect natural variation (such as a successional or altitudinal gradient) or that have been subjected to anthropogenic impacts such as pollution or disturbance (Chapter 17). The localities or samples in such a comparison may differ in a number of ways. For example, the body size of organisms often changes when a site is polluted or a system is over-harvested—indeed the shifting relationship between biomass and numerical abundance is used as a measure of impact (Warwick & Clarke 1994). Body size is likely to influence detectability. Many sampling devices such as gill nets or plankton nets trap individuals within a particular size range so, although it may seem sensible to deploy the same sampling gear when assessing biodiversity at different sites, variation in detectability could generate misleading answers. Behavioural changes in harvested or hunted species (Casas et al. 2009) may also affect detectability.

If detectability cannot readily be estimated as an integral part of a monitoring scheme, it may be possible to carry out experiments at sample locations to estimate detection probabilities, although in some communities, such as speciose tropical arthropod assemblages, this will not be feasible.

In our example, although we have corrected for detectability, we anticipate that some bias will remain. Our analyses assume that birds that are on or very close to the transect line will be detected with certainty, and this is unlikely to be the case for some species, especially nocturnal species. In addition, it is not always possible to follow the ideal transect line, and the actual line covered may be biased towards or away from habitats preferred by certain species. Nevertheless, bias in estimated biodiversity trends can be expected to have been reduced.

Detectability can be particularly difficult to accommodate when most species are represented by very few individuals, so that even if the survey design allows detectability to be estimated, there may be insufficient data for reliable estimation. For single-site monitoring in tropical environments, typically around a third of the species are represented by just a single individual (Coddington et al. 2009). If the focus is on estimating regional biodiversity trends and data are gathered from a number of representative sites through the region, the proportion of species for which there are too few data to estimate detectability is likely to be smaller. This is related to the issue of 'repetitive sampling': repeated sampling of the same units results in more (apparently) rare species than when sampling occurs at the same intensity but where each area is sampled just once (Dobyns 1997; McGill 2003a). However, for highly diverse communities, the number of species with very few detections will remain large. Another strategy to reduce the number of apparently rare species in the sample is to use different trapping methods at each sample location, as a species that is unlikely to be caught by one method may be more likely to be caught by another (Longino et al. 2002). Of course, if the different methods sample different species entirely then, while the community is better represented, the proportion of species with very few detections for the data pooled across the trapping methods will be within the range observed for the separate methods.

## 3.6 Prospectus

The rising interest in monitoring regional biodiversity, sparked in large measure by the 2010 target, is starting to focus attention on the communities that are to be monitored. In statistical terminology, it is important to define the population (which typically comprises individuals of a number of species within a region in our case) about which we wish to draw inference. In this context, it is important either to ensure that individuals of different species are equally likely to be in our sample or to gather data that allow estimation of detectability by species. Thus we anticipate that the design and field protocols of surveys that are to be used to monitor

biodiversity trends will receive closer scrutiny than has occurred to date.

We have shown how to adjust diversity indices when estimates of detectability are available from sample data. With respect to statistical inference, we expect further developments related to estimators and their properties, for when detectability varies among species. A thorough analysis would, for example, estimate uncertainty in the diversity estimates.

Especially when we are interested in monitoring regional biodiversity, our focus is more on a rate of change of diversity (and change in that rate of change) than in a single index value. While only a linear regression model was fitted to the adjusted diversity estimates here to look at trends over time, more sophisticated models such as the generalized additive models used by Buckland et al. (2005) provide greater flexibility, and thus potentially a better fit to the time series generated by the underlying ecological processes.

However, the issue of missing species will inevitably remain a focus of concern and of research, in particular when we are looking at a regional scale. Species might be missing because of seasonal effects at the time of the study or because they are rare and therefore missed in the sampling process. In the latter case, when a large number of plots is sampled repeatedly over time, occupancy methods are likely to prove useful for addressing this difficulty. It remains to be studied more closely how strongly the adjusted indices are affected by this uncertainty about the total number of species. Survey design as well as data analysis should try to take seasonal or other ecological effects into account as far as possible to allow identification of the different effects which lead to missed species in sample data.

## 3.7 Key points

1. When using diversity measures such as the Shannon and Simpson's indices, researchers are making the implicit assumption that individuals have been sampled at random from the community.
2. In most cases this assumption does not hold as there will be heterogeneity amongst individuals in the probability of detection.
3. If probability of detection of individuals can be calculated for each species, indices can be corrected for detectability.
4. Unseen species are also a detectability issue—in this case the species are too rare to have been detected by the sampling protocols or the protocols are ineffective for detecting the species.
5. The Scottish farmland bird community was used as a case study and detection functions fitted for each species. Estimates of biodiversity made using the Shannon and Simpson's indices were higher when the data were corrected for detectability, and the geometric mean of relative abundance estimates showed evidence of a positive trend for the uncorrected data, but no evidence of a trend for the corrected data.
6. A drawback of this approach is that detection functions cannot be fitted to rare species, so that either such species must be excluded from analysis or similar species must be identified that are assumed to have the same detectability.
7. If not estimated, variable detectability may introduce bias in many comparative investigations of biodiversity, such as contrasts between polluted and unpolluted assemblages and estimation of temporal or spatial trends in diversity.

## Acknowledgements

We are very grateful for review comments from Rob Colwell, which led to a much improved manuscript.

# PART II

# Diversity