CHAPTER 10

MINIMISING OBSERVER ERROR

Increasing the reliability of a monitoring project

CLIVE HURFORD

Countryside Council for Wales, Plas Penrhos, Ffordd Penrhos, Bangor, Gwynedd, LL57 2BQ
clive.hurford@serapias.net

1. INTRODUCTION

As the monitoring result has a direct influence on the way that a site is managed, a monitoring project must provide the same result regardless of who does the monitoring, and that result must be the right result.

To achieve this level of consistency, we must minimise the opportunities for observer bias during the data collection phase of the monitoring project. In practice, this means giving careful consideration to the respective merits of recording cover, structure, abundance and frequency: these are the measurements that we are most likely to use for habitat monitoring. The following sections use the results from sampling trials to show where observer bias is most likely to arise, and to illustrate ways of minimising it.

1.1 Background information on the sampling trial data

Over the period from 1996 to 2004, members of the CCW monitoring team carried out a series of multiple-observer sampling trials to assess the degree of observer bias attached to measures commonly used in vegetation survey and surveillance projects. These sampling trials were carried out in a variety of habitats, and by observers with varying degrees of field experience, e.g. university students, conservation site managers, habitat specialists and professional field surveyors.

During the sampling trials, a series of observers assessed the same attributes at fixed sample points. The number of observers involved in each sampling trial ranged from eight to 20. The main measures that we tested were 1) estimates of percentage cover, 2) recording cover / abundance scores using the Domin scale; 3) assessing the vegetation against cover targets, and 4) and recording species frequency.

C. Hurford & M. Schneider (eds.), Monitoring Nature Conservation in Cultural Habitats, 79–92.

These results from these trials build on information gathered during an exercise carried out by the English Field Unit (Leach & Doarks, 1991), which looked at the degree of variation between experienced vegetation surveyors when compiling a species list (and abundance scores) for two fixed quadrats (see 3.1).

The data sets that I have used to demonstrate observer bias are typical of the results obtained from the sampling trials: the only modification I have made is to remove one outlier data set from each set of results. In every case, the inclusion of the outlier data set would have increased the range of observer variation in the results, not reduced it. Over time we have come to realise that you can train three out of four people to be vegetation surveyors: the fourth person should be actively discouraged from pursing this career path.

2. RECORDING ESTIMATES OF VEGETATION COVER

If you decide to record estimates of vegetation cover in your monitoring project, then you must consider which form of cover assessment to use; there are several options. The commonest forms of cover assessment in current usage are subjective measures:

- Estimates of percentage cover;
- The Domin cover scale; and
- The Braun-Blanquet cover scale.

Two other options are pin-frame recording and estimating against cover targets. Some researchers favour pin-frame recording, because it is an objective measure with good statistical properties. The method is too time consuming, however, to be used for monitoring large areas of vegetation. Conversely, cover targets have been used in many CCW monitoring projects, and this form of assessment is covered in the sections below.

2.1 Estimates of percentage cover

Ecologists have been uncomfortable with the use of straightforward estimates of percentage cover for some time. This discomfort gave rise to the Braun-Blanquet and Domin cover scales, which are assessed in 2.2.

The main problems with recording percentage cover estimates arise simply because it is a subjective measure, and can be influenced by a number of variables, e.g.:

- The familiarity of the observer with the habitat or species being assessed;
- The size of the area of search;
- The complexity of the vegetation; and
- The structure of the vegetation.

To demonstrate the degree of observer bias and subjectivity attached to recording estimates of percentage cover, I have used data from a sampling trial where the attribute being assessed, ericoid cover, was easy to identify and easy to see. The habitat was

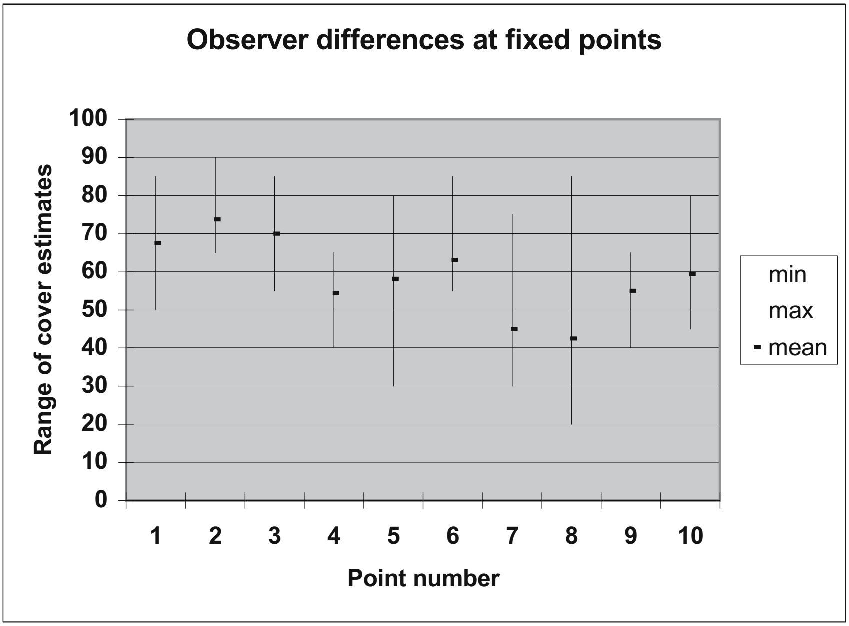

Figure 10-1. The results from a sampling trial to test the range of variation between observers estimating the percentage cover (at intervals of 5%) of ericoids at fixed points in blanket bog vegetation. The mean range of uncertainty was 36%.

blanket bog, which is naturally species-poor, and the species-group comprising the attribute (i.e. *Calluna vulgaris*, *Erica tetralix* and *Empetrum nigrum*) could not be confused with anything else in the sample area. In addition, the recorders were all experienced field surveyors or habitat specialists with a professional interest in that habitat type. Fig. 10-1 shows the results from this sampling trial, where the seven observers assessed the percentage cover of ericoids within a 1m-radius area of search at the same ten sample points. We marked the location of each sample point with a numbered cane.

The data set in Fig. 10-1 shows that, on average, the difference between the lowest and highest cover estimate at each sample point varied from 15% to 65%, with a mean difference of 35%. Bearing in mind that I have already removed an outlier data set, these results should be a concern for anybody recording percentage cover estimates for surveillance or monitoring purposes. In my experience, these results are not atypical of the results from multiple-observer field trials. In fact, I have seen many that are less well aligned, particularly when the attribute being assessed was either difficult to measure, such as bryophyte cover or bare ground, or difficult to separate from similar species in the search area, such as a species of grass or sedge.

The range of uncertainty associated with recording percentage cover estimates has been known, or suspected, for many years, which raises the question of why we persist with the measure. From a monitoring perspective, we cannot base habitat condition assessments on a measure where we have to ignore cover shifts of ca. 35% (either side of our estimate) to accommodate observer error: changes of that magnitude would have a dramatic impact both on the conservation value of a habitat and on the species that are associated with it

2.2 Domin and Braun-Blanquet cover scales

Strictly speaking, the Domin scale is a cover and abundance scale, but this is a moot point as, in practice, both the Domin and Braun-Blanquet scales (Table 10-1) are used primarily for recording vegetation cover.
Both of these cover scales have been used extensively in vegetation survey and surveillance projects, the Domin scale mostly in the UK, Braun-Blanquet more in Europe. The initial impression is that these scales reduce the scope for observer error, with the Braun-Blanquet scale being more robust than Domin. Both scales were developed to describe vegetation for survey purposes and, within an appropriate classification system, both can do that to good effect.

However, problems arise for monitoring from two areas: the first is related to the process that the surveyors use to arrive at their cover class, the second is related to the implications attached to making an error. I will deal with the problems that arise through the recording process first.

When a vegetation surveyor is using either of these cover scales, they initially estimate the percentage cover of the species in question, and then translate that estimate into the appropriate cover class. The problem here, as we saw in the sampling trial results for percentage cover estimates (2.1), stems from the accuracy of the original cover estimate, which varies as a result of observer bias.

Table 10-1. The Domin and Braun-Blanquet scales.

	Domin scale		**Braun-Blanquet scale**
+	A single individual. No measurable cover	+	Less than 1% cover
1	1-2 individuals. No measurable cover.	**1**	1-5% cover
2	Several individuals, but less than 1% cover	**2**	5-25% cover
3	1-4% cover	**3**	26-50% cover
4	5-10% cover	**4**	51-75% cover
5	11-25% cover	**5**	76-100% cover
6	26-33% cover		
7	34-50% cover		
8	51-75% cover		
9	76-90% cover		
10	91-100%cover		

The second problem comes in to play when the original cover estimate is at or near the boundary between two cover classes, as then the observer is forced to reconsider whether the cover is above or below that boundary before allocating a cover class to the species. In our experience, in marginal situations, where the cover of a species is close to a boundary between two cover classes, the chance of two observers allocating the species to the same cover class is no better than 50:50. Unfortunately, if you are using the Domin scale, you are never far away from a boundary between cover classes. In one respect, this suggests that the Braun-Blanquet scale is better adapted for monitoring than the Domin scale, because there are less cover classes and therefore fewer boundaries. This is not necessarily true, however, because the implications attached to making an error are greater. Table 10-2 shows the results of multi-observer sampling trials designed to look at observer differences using the Domin scale to assess vegetation cover.

For both sampling trials, we fixed the location of four quadrats (two 1 x 1 m quadrats and two 50 x 50 cm quadrats) and asked the participants to record a Domin score for each of the attributes listed in the table. A different set of observers was used to record the Domin scores at each quadrat.

The results from these trials show that all three surveyors agreed on the Domin score in only two (10%) of the 20 assessments (and in one of those the attribute was absent). By contrast, the surveyors recorded three different Domin scores in seven (35%) of the 20 assessments. Also of significance, 50% of the assessments differed by more than one Domin score, with a maximum difference of five Domin scores recorded for one assessment.

If these results are an accurate reflection of observer variation using the Domin scale, and the results from other sampling trials suggest that they are, then even if there was no change in the vegetation cover at the sample points, during a repeat monitoring exercise 50% of the assessments would differ by more than one Domin score as a result of observer error. So if we want to use Domin for recording vegetation cover, we have to accept that there is a 50% chance at every sample point that there will be observer error of at least two points on the Domin scale. What are the implications of this for conservation management?

Table 10-2. A comparison of Domin scores from two different sampling trials. Trial 1 involved professional ecologists and Trial 2 involved university students. Observers 1, 2 and 3 recorded the Domin estimates in fixed quadrats 1-4. Trial 1 was carried out in mire vegetation and Trial 2 was in dune grassland. The blocks of data highlighted in red draw attention to assessments that differed by more than one Domin score.

Attribute	**Quadrat 1**			**Quadrat 2**			**Quadrat 3**			**Quadrat 4**		
	1	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**
Trial 1												
Sphagnum cover	**4**	**6**	**8**	**10**	**8**	**9**	**5**	**5**	**5**	7	8	8
Grass cover	**4**	**5**	**7**	**0**	**0**	**0**	4	4	5	0	1	0
Trial 2												
Moss cover	6	7	7	**7**	**5**	**6**	**7**	**5**	**6**	5	4	4
Grass cover	4	5	4	4	4	5	7	5	5	**9**	**10**	**8**
Bare sand	7	6	7	**8**	**5**	**6**	2	0	0	2	0	0

Most of the differences occurred within the range of Domins 4 to 8. So at the bottom end of that range, Domins 4-6, we would have to ignore any changes of cover from 4% to 33% because there would be 50% chance that there had been no change at all. Similarly, in the range from Domins 6 to 8, we would have to ignore changes in cover from 33% to 80% for the same reason. Cover changes of this magnitude are not an early warning of change: again, they would have a dramatic impact on the conservation value of any habitat. Unless we are prepared to accept this level of uncertainty, we cannot use the

Domin scale for monitoring the condition of vegetation. Furthermore, it is worth noting that the data set collected by the professional ecologists in Trial 1 was no better aligned than the data set collected by university students in Trial 2.

We have not tested the Braun-Blanquet scale, but there is no reason to believe that we would not see a similar pattern. If anything, the effects of observer error using the Braun-Blanquet scale would be even greater, because the cover bands are wider. If we had to accommodate observer differences of only one point on the Braun-Blanquet scale, for example from 2 to 3, then we would have to write-off a change from 5% to 50% cover when there may have been no change at all. Consequently, neither of these cover scales are appropriate for monitoring habitat condition.

2.3 Cover targets

The last estimate of vegetation cover that we have tested extensively is the assessment of cover targets, also known as 'cover pseudospecies'. These are used in multi-variate statistical analyses, such as TWINSPAN, to help to separate vegetation types with a similar species composition. To the best of my knowledge, however, cover targets had not been used in vegetation monitoring projects before 1996, when we began to look at measurability issues as part of EU/CCW Life Project (Brown, 2000; Hurford & Perry, 2000).

The concept underpinning the use of cover targets is that if we can identify the point at which the cover of the competitive or dominant species starts to impact on the occurrence of the more sensitive stress tolerating species, we only have to assess whether the cover of the competitors is greater or smaller than that value.

Table 10-3. The results from sampling trials to assess the scale of observer variation when recording cover targets. Observers 1, 2 and 3 assessed whether the vegetation cover of the attributes was greater than or less than 20% in fixed quadrats 1-4. If the cover of the attribute was borderline (at or around 20%), the recorders were told to record it as greater than 20%. Trial 1 was carried out in mire vegetation: Trial 2 was in dune grassland. The blocks of data highlighted in red draw attention to inconsistent assessments.

Attribute	Quadrat 1			Quadrat 2			Quadrat 3			Quadrat 4		
	1	2	3	1	2	3	1	2	3	1	2	3
<20% / >20%												
Sphagnum cover - Trial 1	>	>	>	>	>	>	>	<	>	>	>	>
Grass cover - Trial 1	>	>	>	<	<	<	<	>	<	<	<	<
Moss cover -Trial 2	>	>	>	>	>	>	>	>	>	<	<	<
Grass cover -Trial 2	<	<	<	<	<	<	>	>	>	>	>	>
Bare sand -Trial 2	>	>	>	>	>	>	<	<	<	<	<	<

The pitfall to this approach, of course, is that you cannot just pull this value out of the air. You must have a good understanding of your site and the habitat that you are monitoring before you can decide what this cover value should be. However, an experienced field recorder should be able to obtain this value from a relatively quick survey exercise (see Chapter 11).

Table 10-3 shows the results of cover target assessments carried out by the same groups of recorders that carried out the Domin sampling trial (Table 2). I have used this data set to illustrate that, although the surveyors struggled to achieve any level of consistency recording Domin values, the same set of recorders could achieve a high level of consistency when assessing cover targets.

The results in Table 10-3 show that the surveyors provided consistent assessments in 18 of the 20 assessments (90%). This level of consistency compares favourably with the trial results for percentage cover estimates and Domin scores. These results are in keeping with those from other sampling trials, though it is not unusual for surveyors to achieve 100% consistency. This level of consistency has a price, however, because the results do not give any indication of the actual cover of the attribute, they only tell you whether it is obviously more than, or less than, the cover target. This is not a particularly high price, however, if you take account of potential impact of the observer error associated with recording estimates of percentage cover.

By focusing attention on a single boundary, cover target assessments are relatively straightforward, and will deliver consistent results, until you are on a site where the cover of the attribute is consistently close to the cover target. There are two ways to avoid an inconsistent monitoring result in this situation. ? You can:

1. Introduce a decision rule stating that, for example, 'the grass cover at the sample point must be obviously less than 20% cover: if you have to stop to think about it, then the point must fail'; and
2. Use the cover target as one of a suite of co-occurring attributes (see Chapter 11) that must all pass before the sampling point can pass. For example, you can state that before a sample point can pass, 'four positive indicator species must be present, the grass must form <20% cover, and all of the negative indicator species must be absent'.

We regularly employ both of these precautionary measures in projects developed to monitor Natura 2000 habitats. The most damaging error that we can make in a monitoring project is to say that a habitat is in good condition when it isn't. Both of these measures discriminate against making that mistake. If we are going to make an error, we will err on the side of caution, which is how it should be on sites of high conservation value.

3. RECORDING SPECIES PRESENCE AND ABUNDANCE

The section looks at the observer variation associated with estimating species diversity and abundance. Management plans often include an aim to 'maintain or increase the diversity (or biodiversity) of a habitat or site'. This section looks at the options that are available for measuring this, and at the levels of observer variation associated with some of the more common recording methods that are used. The sections below include assessments of species diversity, species frequency, and species abundance. Where available, we have used sample trial data to inform our recommendations.

3.1 Assessing species diversity

During the 1980s, the English Field Unit carried out a sampling trial designed to assess observer variation in recording species diversity and abundance in grassland vegetation (Leach & Doarks, 1991). This trial involved 14 experienced grassland surveyors, who were asked to independently record all of the species in two fixed quadrats, one 1 x 1 m quadrat and one 10 x 10 m quadrat.

The results from this trial showed that, in the 1 x 1 m quadrat, the most successful surveyor found 73% of the species recorded in the quadrat, while on average the surveyors recorded only 63% of the species. As might be expected, the detection rate in the 10 x 10 m quadrat was considerably lower, with the most successful surveyor recording only 63% of the species and the average detection rate falling to 55%.

These results suggest that if we are interested in recording changes in species diversity, then we have to be prepared to live with observer variation of □ 30%. I fear that we have to accept that we will never have this level of information for our sites (particularly if we take account of the diversity of other species groups associated with the habitats, e.g. invertebrates) and that we must learn to live without it.

To some degree, information on species diversity is ‘nice to know’, in as much as we have never had this level of information before, and this has not stopped us managing habitats. I have known it prevent conservation managers making management decisions, but the habitat was still being managed of course, albeit passively. In truth, we will probably never know (except maybe in very species-poor habitats) the true diversity of species on a site. The only practical alternative is to record reliable indicators of diversity.

3.2 Assessments of frequency

Frequency is an objective measure that uses ‘presence or absence’ data to assess how frequently an attribute is present in a set of samples: this figure is typically expressed as a percentage. For example, you could assess the frequency of otter spraints along a stretch of river by dividing it into 20 sections and simply recording whether otter spraints were found in each section. If otter spraints were found in 15 sections, they occurred at a frequency of 75%.

Table 10-4. The results from sampling trials to test the effects of observer bias on species frequency data. Trials 1 and 2 involved students from University of Wales Swansea; Trial 3 involved professional ecologists. The blocks of data highlighted in red draw attention to inconsistent assessments.

Attribute	Quadrat 1			Quadrat 2			Quadrat 3			Quadrat 4		
	1	2	3	1	2	3	1	2	3	1	2	3
Trial 1												
Calluna vulgaris	+	+	+	+	+	+	+	+	+	+	+	+
Erica tetralix	+	+	+	+	+	+	+	+	+	+	+	+
Trichophorum cespitosum	+	+	+	+	+	+	+	+	+	+	+	+
Juncus squarrosus	+	+	+	+	+	+	+	+	+	+	+	+
Sphagnum sp.	-	-	-	-	-	-	-	-	-	-	-	-
Trial 2												
Cerastium semidecandrum	+	+	+	+	+	+	+	+	+	+	+	+
Erophila verna	+	+	+	+	+	+	-	+	-	+	+	+
Hornungia petraea	-	-	-	-	-	-	-	-	-	+	+	+
Saxifraga tridactylites	+	+	+	+	+	+	+	+	+	+	+	+
Peltigera canina	+	+	+	+	+	+	+	+	+	+	+	+
Trial 3												
Molinia caerulea	+	+	+	-	-	-	+	+	+	-	+	+
Potentilla erecta	+	+	+	-	-	-	+	+	+	-	-	-
Andromeda polifolia	-	-	-	+	+	+	-	-	-	+	+	+
Narthecium ossifragum	+	+	+	+	+	+	-	-	-	+	+	+
Rhynchospora alba	-	-	-	+	+	+	-	-	-	+	+	+

The quality of a habitat can be assessed in much the same way. When we have defined how to recognise good quality habitat, we can monitor to assess how frequently the vegetation meets the definition: there are several examples of this in the habitats case studies. The reliability of the result, however, will depend on the measurability of the attributes that we assess at each monitoring point.

We have tested the ability of surveyors to record species frequency on many occasions, and Table 10-4 shows the results from three sampling trials. A minimum of 12 surveyors participated in each trial, including many with little or no previous experience of vegetation recording or of the species that we selected for the exercise.

To help to overcome these problems, before each trial we spent c. 30 minutes training the surveyors to identify the relevant species. To ensure that the surveyors searched intensively for the species, we subdivided each of the fixed quadrats into 16 cells, and asked the surveyors to search for the presence of each species in each cell. The main purpose of the exercise, however, was to test the ability of the surveyors to detect the presence of each species in each of the four fixed quadrats.

The sampling trial results in Table 10-4 show that, overall, the surveyors agreed on the presence or absence of the selected species in 58 of the 60 assessments (97%), and that the level of consistency within each trial never dropped below 95%. The results of Trial 2 were particularly impressive, as most of the participants were students with backgrounds in either marine biology or zoology, and four of the five species being assessed were diminutive spring annuals.

These results suggest that, given a short period of training in species identification, even inexperienced surveyors can achieve consistent results when assessing the presence or absence of a small number of species at a sample point.

3.3 Assessments of abundance

Abundance differs from frequency by referring to the number of individuals of a species present, as opposed to the mere presence or absence of the species (see Fig. 10-2).

The commonest scale for assessing species abundance, in the UK at least, is the DAFOR scale, where D = Dominant, A = Abundant, F = Frequent, O = Occasional and R = Rare. Unfortunately, there does not appear to be a standard definition of these categories, so you will come across various interpretations of the DAFOR scale in different publications.

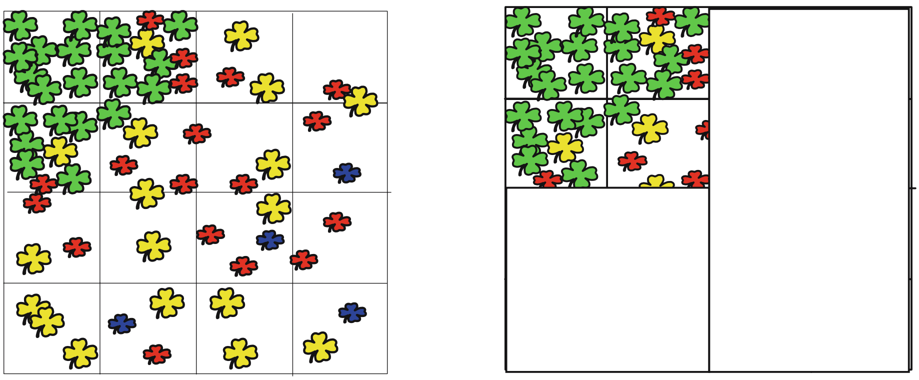

Figure 10-2. Frequency and abundance: the green plant (clustered in the top left corner of the 1 x 1 m quadrat on the left) is present in only four of the 25 x 25 cm cells, giving a cell frequency of 25%, but it has an overall abundance of 21 in the quadrat. The scarcer blue plant is also present in four of the cells and shares the same cell frequency score as the green plant (25%), but the blue plant has an overall abundance of only four. Note how, in the smaller quadrat on the right, it is no longer possible to see an increase in the frequency of the green plant as it now has frequency of 100%, nor could we record a decline in the blue plant as it is no longer present in the sample area. The size of the area of search is critical if we are using frequency measures for monitoring.

A sampling trial carried out by the English Field Unit (Leach & Doarks, 1991) found wide variation between experienced grassland surveyors using a form of the DAFOR scale. For example, the DAFOR assessments for *Lotus corniculatus* (a relatively common and easily identified species) ranged from absent to abundant between surveyors recording the same patch of vegetation. Leach and Doarks concluded that while compiling a species list with DAFOR may be a good way for a surveyor to become familiar with a habitat, it is probably of little value as a monitoring method.

Straightforward counts are an alternative form of abundance recording, though these are generally reserved for species monitoring projects. However, counts can also be used to good effect in habitat monitoring projects, particularly for dealing with negative indicator species. For example, the Heath Rush *Juncus squarrosus* is a regular component of wet heath vegetation, but it can also be an indicator of over-grazing. So while we might expect to find the occasional plant of *Juncus squarrosus* in a stand of wet heath, we might be concerned if we were seeing small clusters of plants scattered throughout it. One way to deal with this is to set an upper limit for the density of plants that we are prepared to tolerate within the area of search at your monitoring points. For example, we could set an upper limit of no more than five plants of *Juncus squarrosus* within a 1 m radius of a monitoring point.

The approach could be applied equally well to aggressive species such as Bracken *Pteridium aquilinum* in dry heath, or Common Reed *Phragmites australis* in transition mires. Counts can also be used in habitat monitoring projects as an alternative to subjective vegetation cover estimates. This is particularly true in broad-leaved woodland, where we can use densities and ratios to provide reliable monitoring results for most of the important structural attributes (Chapter 27).

Photograph by Clive Hurford

Figure 10-3. Heath Rush *Juncus squarrosus*, here with Ling *Calluna vulgaris* and Mat Grass *Nardus stricta*, responds positively to over-grazing in heath vegetation.

As a general rule, if we can assess the attributes using simple objective measures, then we should use them, because they are much less prone to observer bias than subjective assessments.

4. ASSESSING VEGETATION HEIGHT

Vegetation height can be an important attribute of a habitat, particularly in grassland vegetation. Many of the species associated with grassland have specific structural requirements: several species of butterfly, small mammals, and fungi spring readily to mind. If vegetation height is an important attribute of the habitat, then we should include upper and / or lower limits, as appropriate, in the condition indicator table. There are three methods commonly used for measuring vegetation height in ecological studies: direct measures; sward sticks; and drop discs.

All of these can be used to good effect, and the methods and their relative strengths and weaknesses are outlined below. Stewart *et al.* (2001), carried out an evaluation of these methods, and concluded that all three were easy to use and delivered consistent results with negligible observer bias.

4.1 Direct measures

To record vegetation height using direct measures, we place a card (or hand) lightly on the vegetation at the point where ca. 80% of the vegetation is growing at or below that height. We then take a reading of this height on a ruler (Hodgson *et al*., 1971). This is the most subjective of the three methods, as the observer has to decide where to place the card. The qualifier 'ignoring tall stalks' improves the likelihood of consistency between recorders.

Direct measures are well adapted for recording the fine scale 'micro-heterogeneity' that some invertebrates require in short swards. The design of sward sticks and drop discs rule them out as practical alternatives if micro-heterogeneity is an issue.

4.2 Sward sticks

For the sward stick method we use a 45 cm metal rule, with 0.5 cm graduations, with a sleeve supporting a 2 x 1 cm piece of clear Perspex. The rule is held vertically, and the sleeve lowered until the Perspex touches the first piece of green non-flowering vegetation: we read the measurement from the rule at this point (Barthram, 1986).

The sward stick samples the smallest area of vegetation of the three methods and therefore gives the most variable results. Its advantage is that, by taking several measurements at each sample point, we can detect structural heterogeneity at each sample point. On the negative side, the sward stick is less well adapted for measuring variation in short vegetation than direct measures (Stewart *et al*., 2001).

4.3 Drop discs

The drop disc method (Holmes, 1974) is simple but effective: we simply let a disc (which is a standard size and weight) slide down the measuring stick from a height of 1.5 m until it rests on the vegetation at the sample point. We then take the reading of the vegetation height from where the disc is resting against the measuring stick. The measuring pole is marked at 1 cm intervals.

In medium to tall swards, the drop disc is recommended as the best method for measuring productivity and the effects of herbivory: it is also recommended for use in agri-environment schemes. The disadvantages are associated with the relatively large surface area of the disc, which makes it less well suited for detecting fine-scale heterogeneity, particularly in low swards. However, if the vegetation height data from a drop disc is combined with species frequency data, then it may well be possible to assume micro-heterogeneity at a monitoring point.

The drop disc is certainly the least subjective and simplest of the methods for measuring vegetation height, and would be more than adequate for assessing vegetation height in most situations. Unfortunately, drop discs are not produced commercially and have to be constructed by the surveyor. The disc itself should have a diameter of 30 cm and weigh 200 g, with a central slot or hole for sliding down the measuring stick. As the critical features of the disc are surface area and weight, the disc can be made from various materials, even cardboard secured with sticky tape. The main difficulty in making these discs is achieving the correct weight, as cardboard tends to be too light and plywood too heavy. However, with cardboard discs we can keep adding tape until we reach a weight of 200 g (this has the added bonus of waterproofing the cardboard), while with plywood discs you can drill holes in the disc to reduce the weight (it is best to varnish the disc first though, as a layer of varnish will push the weight back up). A 1.5 m or 2 m rule will suffice for a measuring pole in most grassland habitats.

Perhaps the most important point from a monitoring perspective is that these methods are not interchangeable - they provide different results. The golden rule therefore is not to change methods during the course of a monitoring project: choose the method and persist with it.

5. IN SUMMARY

In this chapter we have discussed various forms of field assessment available for monitoring the quality of a habitat, concentrating on three principal components: vegetation cover, species composition, and vegetation height.

The results from multiple-observer sampling trials have indicated that the most reliable measures for monitoring habitats are presence and absence data; simple counts of abundance; and using a drop disc to record vegetation height.

As a general rule, we should try to avoid using estimates of vegetation cover in a monitoring project unless absolutely necessary. If we decide that it is essential, then we should monitor against cover targets. The results from sampling trials suggest that if we set up a monitoring project where the result can depend solely on estimates of vegetation cover, then the reliability of the monitoring result will be compromised by unacceptable levels of observer bias.

For this reason, we should think carefully about what we need to know about the vegetation that we are monitoring before deciding how to monitor it. If we consider, within any broad habitat type, which examples of a habitat we regard to be of high conservation interest, and why, we will probably begin to focus on those with a good representation of stress tolerating species (Chapter 8). These species will become scarcer as the more competitive species achieve dominance. This suggests that, in most cases at least, it is actually the presence of the stress tolerators (and associated species) that dictates the conservation value of the habitat, rather than the cover of the potentially dominant competitors. If we accept this, then the most efficient and reliable approach to monitoring the condition of a habitat is to focus on the frequency (or abundance) of the stress tolerating associate species, and not the cover of the dominants.

6. REFERENCES

Stewart, K.E.J., Bourne, N.A.D. & Thomas, J.A. (2001). An evaluation of three quick methods commonly used to assess sward height in ecology. *Journal of Applied Ecology 2001*: **38**: 1148-1154.

Barthram, G.T. (1986). Experimental techniques – the HFRO sward stick. *Biennial Report of the Hill Farming Research Organisation 1984-85* (ed. M.M. Allcock), pp. 29 – 30. Hill Farming Research Organisation, Penicuik, Midlothian, UK.

Brown, A. (2000). *Habitat Monitoring for Conservation Management and Reporting. 3: Technical Guide*. Life-Nature Project no LIFE95 NAT/UK/000821. Integrating monitoring with management planning: a demonstration of good practice in Wales. Countryside Council for Wales, Bangor.

Holmes, C.W. (1974). The Massey grass meter. *Dairy Farming Annual*, pp.26 – 30. Massey University, Palmerston North, New Zealand.

Hodgson, J., Taylor, J.C. & Lonsdale, C.R. (1971). The relationship between intensity of grazing and the herbage consumption and growth of calves. *Journal of the British Grassland Society*, **26**: 231 – 237.

Hurford, C. & Perry, K (2000). *Habitat Monitoring for Conservation Management and Reporting. 1: Case studies*. Life-Nature Project no LIFE95 NAT/UK/000821. Integrating monitoring with management planning: a demonstration of good practice in Wales. Countryside Council for Wales, Bangor.

Leach, S.J. & Doarks, C. (1991). *Site Quality Monitoring methods and approaches (with particular reference to grasslands)*. Project No.135. English Field Unit, Nature Conservancy Council, Peterborough.