

## Article (refereed) - postprint

---

Hill, Mark O.. 2012 Local frequency as a key to interpreting species occurrence data when recording effort is not known. *Methods in Ecology and Evolution*, 3 (1). 195-205. [10.1111/j.2041-210X.2011.00146.x](https://doi.org/10.1111/j.2041-210X.2011.00146.x)

© 2011 The Author. *Methods in Ecology and Evolution* © 2011 British Ecological Society

This version available <http://nora.nerc.ac.uk/16610/>

NERC has developed NORA to enable users to access research outputs wholly or partially funded by NERC. Copyright and other rights for material on this site are retained by the rights owners. Users should read the terms and conditions of use of this material at <http://nora.nerc.ac.uk/policies.html#access>

**This document is the author's final manuscript version of the journal article, incorporating any revisions agreed during the peer review process. Some differences between this and the publisher's version remain. You are advised to consult the publisher's version if you wish to cite from this article.**

Contact CEH NORA team at  
[noraceh@ceh.ac.uk](mailto:noraceh@ceh.ac.uk)

1 **Local frequency as a key to interpreting species occurrence data when**  
2 **recording effort is not known**

3

4 M. O. Hill<sup>1</sup>

5

6 Centre for Ecology and Hydrology, Wallingford OX10 8BB, England; moh@ceh.ac.uk

7 <sup>1</sup>present address 11 Chaucer Road, Cambridge CB2 7EB, England

8

9 *Word count: 7776*

10 **Keywords**

11 benchmark species; biodiversity; bryophyte; British Isles; data cleaning; discovery curve; epiphyte;  
12 neighbourhood; measurement of change; rank-frequency curve; recording visit; species richness

13 **Summary**

14

15 **1.** Data on the occurrence of species in grid cells are collected by biological recording schemes,  
16 typically with the intention of publishing an atlas. Interpretation of such data is often hampered by  
17 lack of information on the effort that went into collecting them. This is the ‘recorder effort problem’.

18

19 **2.** One measure of recorder effort is the proportion of a suite of common species (‘benchmark  
20 species’) found at a given location and time. Benchmark species have in the past been taken as a  
21 uniform set across a territory. However, if records are available from a neighbourhood surrounding a  
22 given location, then a local set benchmark species can be defined by pooling records from the  
23 neighbourhood and selecting the commonest species in the pooled set.

1

2 3. Neighbourhoods differ in species richness, so that the list of species that ‘ought’ to be found in one  
3 location may be longer than that for another. If the richness of a neighbourhood can be estimated,  
4 then a suite of benchmark species can be standardized to be the commonest of a fixed proportion of  
5 the total expected for the neighbourhood. Recording effort is then defined as the proportion of  
6 benchmark species that were found.

7

8 4. A method of estimating species richness is proposed here, based on the local frequencies  $f_j$  of  
9 species in neighbouring grid cells. Species discovery is modelled as a Poisson process. It is argued  
10 that when a neighbourhood is well sampled, the frequency-weighted mean frequency  $\sum f_j^2 / \sum f_j$  of  
11 species in the neighbourhood will assume a standard value.

12

13 5. The method was applied to a dataset of 2,000,000 records detailing the occurrence of bryophytes in  
14 3,695 out of the total 3,854 hectads (10-km squares) in Great Britain, Ireland, the Isle of Man and the  
15 Channel Islands.

16

17 6. Three main applications are outlined: estimation of recording effort, scanning data for unexpected  
18 presences or absences, and measurement of species trends over time. An explicit statistical model  
19 was used to estimate trends, modelling the probability of species  $j$  being found at location  $i$  and time  $t$   
20 as a the outcome of Poisson process with intensity  $Q_{ijt} x_{jt}$ , where  $x_{jt}$  is a time factor for species  $j$  and  
21  $Q_{ijt}$  depends on recording effort at location  $i$  and time  $t$  and on the time-independent probability of  
22 species  $j$  being found in hectad  $i$ .

## 23 **Introduction**

24

25 When the occurrence of species is mapped in grid cells, large-scale patterns of geographical  
26 distribution and long-term trends in frequency can be observed. Examples are the widespread decline

1 of arable weeds (Preston, Telfer *et al.*, 2002) and the disappearance of butterflies from parts of their  
2 former range (Asher *et al.*, 2001). Many changes are less obvious, and to measure these, various  
3 statistical methods have been devised. Several of these methods use the proportion of records of a  
4 given species relative to the total records of all species as a measure of frequency (Ball and Henshall,  
5 2006, Telfer *et al.*, 2002). Similar methods have been used with museum data (Hedenäs *et al.*, 2002,  
6 Hofmann *et al.*, 2007, Jeppsson *et al.*, 2010).

7

8 The method of Telfer *et al.* (2002) measures relative change only between two time-periods, but  
9 several of the others can measure trends at intervals. They do this by treating the total number of  
10 records as a measure of recording effort. A potential weakness of this approach is that different parts  
11 of the territory may be sampled unevenly at different time periods. If sampling at a particular period  
12 is concentrated on a region from which a species is absent or very scarce, then its frequency may  
13 appear to have declined without this actually being the case.

14

15 Unevenness of recording continues to be a problem. Prendergast *et al.* (1993) drew attention to it, and  
16 called it the ‘recorder effort problem’. They observed that virtually all large-scale floral and faunal  
17 surveys depend on volunteer recorders, creating datasets that may be biased in favour of recorder,  
18 rather than species, distributions. The role of volunteer recorders is increasing (Silvertown, 2009).  
19 Some volunteers can be persuaded to engage in standardized monitoring or recording programmes  
20 (Roy *et al.*, 2007, Newson *et al.*, 2005), but most cannot. As a consequence, there are large biases in  
21 many datasets (Boakes *et al.*, 2010, Dennis and Thomas, 2000, Hassall and Thompson, 2010, Petrik *et*  
22 *al.*, 2010).

23

24 The recorder effort problem can at least partly be solved if there is some measure of regional or local  
25 recording effort. One such measure is the number of recorded ‘benchmark species’, i.e. species that  
26 are thought to be ubiquitous and stable in their occurrence. Records of other species can be related to  
27 those of the benchmark species (Maes and Swaay, 1997). With a relatively small territory, it may be

1 possible to find such ubiquitous species, but with a larger territory such as Great Britain, this may be  
2 difficult. Even if there are ubiquitous species, they may differ widely in abundance across the  
3 territory, so that their presence in one district does not signify the same recording effort as in another.  
4 There is, however, no necessity for the same benchmark species to be used everywhere. Indeed, local  
5 recorders conducting systematic surveys often judge what ‘ought’ to be present in a particular location  
6 by comparing the records for that location with others in the neighbourhood. When an area is  
7 moderately well recorded, then a simple ‘wants list’ can be constructed for each recording unit by  
8 listing those species that are absent from the recording unit but present in most of its neighbours.  
9 Under-recorded units are those with long wants lists.

10

11 The method that is outlined below is a generalization of this means of judging which species are likely  
12 to be present. It proceeds in four stages at each separate location.

- 13 1. Define the neighbourhood of location  $i$  as a set of weights for nearby locations, based on both  
14 spatial proximity and biological similarity.
- 15 2. Based on records of species presence, calculate weighted local species frequencies  $f_{ij}$  for all  
16 species in the neighbourhood, using weights defined at stage 1.
- 17 3. Calculate neighbourhood frequency as the frequency-weighted local mean frequency, i.e.  
18 
$$\frac{\sum_i f_{ij}^2}{\sum_i f_{ij}}.$$
- 19 4. Adjust local species frequencies by a ‘sampling-effort multiplier’  $\alpha_i$ , used to inflate (or  
20 occasionally deflate) them to the point where neighbourhood frequency assumes a standard  
21 value.

22 At stage 4, frequencies are assumed to result from a Poisson process, and the multiplier  $\alpha_i$  is applied to  
23 the intensity of the process, namely  $-\log(1-f_{ij})$ . This is converted back to a frequency. The resulting  
24 species frequencies  $f'_{ij}$  are treated as probabilities and are used to estimate trends or to identify  
25 locations with an excess of species (‘hotspots’).

# 1 **Data and methods**

## 2 DATASETS AND COMPUTER PROGRAMS

3 Two large datasets held by the Biological Records Centre (BRC) were used as the basis for the  
4 analysis. The immediate object of the study was to measure temporal trends in the probability of  
5 bryophyte species occurrence, based on the dataset of the British Bryological Society. These data  
6 were mapped as an atlas in the early 1990s (Hill *et al.*, 1991-1994) and have been steadily added to  
7 since then; in July 2010 they comprised 2,038,000 records of 1,196 taxa, of which 1,053 were  
8 species, in 3,695 hectads. The other dataset was the Vascular Plant Dataset, managed by BRC on  
9 behalf of the Botanical Society of the British Isles. In summer 2010 this had 12,675,000 records,  
10 from which lists of native and archaeophyte vascular plant species in 3829 out of the total 3,854  
11 hectads (10-km squares of the National Grids of Britain and Ireland) were abstracted. Hybrids and  
12 taxa not mapped in the most recent atlas (Preston, Pearman *et al.*, 2002) were excluded. The  
13 abstracted dataset comprised 1,666,000 records of 1,469 species.

14

15 Database manipulations were performed in Microsoft Access. Other calculations were done in  
16 Fortran, using the GNU Fortran G77 v0.5.25 compiler for Windows XP (Free Software Foundation,  
17 1999). Calculations other than those to used defined neighbourhoods were performed by a single  
18 program, Frescalo (FREquency SCAling LOcal). Source code and executables for Windows XP,  
19 together with a worked example, can be downloaded from the BRC website [www.brc.ac.uk](http://www.brc.ac.uk).

## 20 SELECTION OF HECTADS FOR DETAILED STUDY

21 Given the number of hectads and bryophyte species, the total dataset of observed neighbourhoods and  
22 species was substantial. For this reason, a systematic sample of one in a hundred hectads was selected  
23 for more detailed study. Hectads were numbered alphabetically in two series according to the grid  
24 under consideration, starting with the British grid and following it with the Irish grid. The systematic  
25 sample (Table 1), comprising hectads 100, 200, ... ,3800, includes localities in Scotland (10), England  
26 (15), Wales (2) and Ireland (10).

1 DEFINITION OF NEIGHBOURHOODS

2 The main use of the vascular plant data was to define neighbourhoods, i.e. sets of hectads that were  
3 both physically close to and floristically similar to a given hectad. For any given hectad (the ‘target  
4 hectad’), the procedure was first to take the 200 spatially closest hectads and then to select the 100  
5 floristically most similar hectads from among the 200. Floristic similarity of the vascular plant flora  
6 in two hectads was measured by Sørensen’s similarity coefficient, i.e. twice the number of species in  
7 common divided by the sum of the two individual species totals (Legendre and Legendre, 1998). The  
8 target hectad was included in the list. Weights were applied so that within the 100 selected  
9 neighbours, the species lists of those that were physically more distant or floristically less similar  
10 received lower weight than those from nearer and more similar hectads. Specifically

$$w_{i' i} = \left(1 - \frac{(k-1)^2}{200^2}\right)^4 \left(1 - \frac{(l-1)^2}{100^2}\right)^4$$

11 where  $w_{i' i}$  is the weight for hectad  $i'$  when considered as a neighbour of the target hectad  $i$  and  $i'$  is the  
12  $k^{\text{th}}$  nearest to  $i$  in order of distance and the  $l^{\text{th}}$  nearest in order of floristic similarity. In this scheme, the  
13 target hectad gets unit weight, and the others are weighted so that those halfway down the list receive  
14 weight about 0.33. For example if a hectad is 100<sup>th</sup> in order of spatial proximity ( $k = 100$ ) and 50<sup>th</sup> in  
15 order of floristic similarity ( $l = 50$ ) then

16  $w_{i' i} = (1 - 0.495^2)^4 \times (1 - 0.490^2)^4 = 0.325 \times 0.333 = 0.108$

17 Five neighbourhoods and their weighting are shown in Fig. 1. See also the supporting information.

18 STANDARDIZATION OF NEIGHBOURHOOD MEAN FREQUENCY

19 When a neighbourhood and its weights have been defined, then  $f_{ij}$ , the observed frequency of species  $j$   
20 in the neighbourhood of hectad  $i$  is defined as

21  $f_{ij} = \sum_{i'} w_{i' i} a_{i' j} / \sum_{i'} w_{i' i}$

22 where the summation is taken over the neighbourhood, and  $a_{i' j} = 1$  if species  $j$  is recorded in hectad  $i'$   
23 and  $a_{i' j} = 0$  otherwise. In neighbourhoods where there are numerous unsampled hectads, even the  
24 most frequent species may have local frequency much less than 1 (Fig. 2).

1 We assume that discovery is a Poisson process. Let

2  $\lambda_{ij}$  = discovery rate for species  $j$  in hectad  $i$

3  $s_i$  = number of searches made in hectad  $i$

4 Then the probability of finding species  $j$  in hectad  $i$  is given by

5 
$$p_{ij} = 1 - \exp(-\lambda_{ij} s_i)$$

6 Parameter  $s_i$  is a measure of sampling effort. In practice, we cannot measure  $\lambda_{ij}$  directly unless

7 separate recording parties are sent in to make independent surveys. We do, however, have

8 information on the frequencies  $f_{ij}$  for all species in the neighbourhood. On the assumption that the

9 target hectad is similar to those in its neighbourhood,

10 
$$\lambda_{ij} s_{i(N)} \approx -\log(1 - f_{ij}) \tag{1}$$

11 where  $s_{i(N)}$  is the weighted mean sampling effort for the neighbourhood, i.e.

12 
$$s_{i(N)} = \sum_i w_{ii} s_i / \sum_i w_{ii}$$

13

14 The problem of estimating  $\lambda_{ij}$  is therefore solved if we can estimate  $s_{i(N)}$ . This is done by

15 standardizing the frequency-weighted mean local frequency

16 
$$\varphi_i = \sum_j f_{ij}^2 / \sum_j f_{ij}$$

17 This is a weighted mean frequency, because if weights  $v_{ij}$  are set to be equal to the frequencies  $f_{ij}$  then

18 
$$\varphi_i = \sum_j v_{ij} f_{ij} / \sum_j v_{ij}$$

19 Indeed,  $\varphi_i$  is analagous to the expected value of a frequency. Imagine that all the species are put into

20 a bag, in proportion to the frequencies. Then select a random species from the bag. The probability

21 that it is of species  $j$  is  $f_{ij} / \sum_j f_{ij}$ . Thus the expectation

22 
$$E(\text{frequency}) = E(f_{ij}) = \sum_j f_{ij}^2 / \sum_j f_{ij} = \varphi_i$$

23

24 An interesting property of  $\varphi_i$  is that it can be expressed as a ratio

25 
$$\varphi_i = \sum_j f_{ij} / ((\sum_j f_{ij})^2 / \sum_j f_{ij}^2) = \sum_j f_{ij} / N_2$$

26 where  $\sum_j f_{ij}$  is the average species richness of hectads in the neighbourhood and  $N_2$  is the ‘effective

27 species number’ (reciprocal of Simpson’s index) defined by Hill (1973). In other words,  $\varphi_i$  is the ratio

1 of mean species richness per hectad to a dimensionless number,  $N_2$ , which depends on the shape of the  
 2 rank-frequency curve but not on the absolute magnitude of the frequencies.

3

4 This property explains the rationale of the method that is described below. If, for example, half of the  
 5 records from the neighbourhood are deleted, so that the new value of  $f_{ij}$  is half the old value, then the  
 6 numerator goes down by a factor of two, but the denominator is unchanged. Thus  $\varphi_i$  is in some sense  
 7 a measure of sampling intensity. The basic assumption of the method is that in a well-sampled  
 8 neighbourhood,  $\varphi_i$  (hereafter called neighbourhood frequency) is roughly constant, say  $\Phi$ . This  
 9 assumption would certainly be correct if the rank-frequency curves were scaled to have the same  
 10 shape in all neighbourhoods. No such assumption is made. Indeed, the proposed method of rescaling  
 11 requires fitting only two parameters, one for sampling effort and the other for species richness. It  
 12 turns out that the resulting curves are remarkably similar.

13

14 Let  $f'_{ij}$  be the frequency of species  $j$  in neighbourhood  $i$  when the neighbourhood is sampled to a  
 15 standard extent, corresponding to a thorough search. Then the assumption of constant neighbourhood  
 16 frequency is that for all neighbourhoods  $i$

$$17 \quad \sum_j f'^2_{ij} / \sum_j f'_{ij} = \Phi$$

18 Without loss of generality, set the neighbourhood mean sampling effort corresponding to this extent  
 19 of sampling to 1. Setting  $s_{i(N)}=1$  in equation 1

$$20 \quad \lambda_{ij} \approx -\log(1 - f'_{ij})$$

21 Now consider the consequences of increasing the actual sampling effort  $s_{i(N)}$  by a sampling-effort  
 22 multiplier  $\alpha_i$ . On the assumption of a Poisson process, the frequency of species  $j$  for this multiplier is

$$23 \quad f_{ij}[\alpha_i] = 1 - \exp(-\alpha_i \log(1 - f'_{ij}))$$

24 The neighbourhood frequency is defined as

$$25 \quad \varphi_i[\alpha_i] = \sum_j f^2_{ij}[\alpha_i] / \sum_j f_{ij}[\alpha_i]$$

26 The problem of estimating  $s_{i(N)}$  is then that of finding a sampling-effort multiplier  $\alpha_i$  such that

$$27 \quad \varphi_i[\alpha_i] = \Phi$$

1 Then, by definition

$$2 \quad \alpha_i s_{i(N)} = 1$$

3 Note that  $\varphi_i[\alpha_i]$  is a property of the neighbourhood not the target hectad. Indeed the target hectad may  
4 be completely lacking in records, as is the case for the Coleraine hectad indicated in Figs 1 and 2.

5

6 Calculation of sampling-effort multipliers requires first that a standard neighbourhood frequency  $\Phi$  is  
7 selected and then that parameters  $\alpha_i$  are fitted so that  $\varphi_i[\alpha_i] = \Phi$ . In the example discussed here,  $\Phi$  was  
8 set to 0.74, which was the 98.5<sup>th</sup> percentile of observed values of  $\varphi_i[1]$ . This high percentile was  
9 chosen, because bryophytes are not very completely recorded. More detail for south-east England is  
10 given in the supporting information. In the applications, the precise value of  $\Phi$  is not critical, but it  
11 should correspond to a thorough search of the neighbourhood. Values of  $\alpha_i$  were fitted by successive  
12 approximation, multiplying trial values  $\alpha$  by  $\Phi/\varphi_i[\alpha]$  until convergence was achieved.

13

14 The curves in Fig. 2 differ not only because the sampling effort varies between neighbourhoods, but  
15 also because of variation in species richness. It is possible to correct for species richness to compare  
16 the shape of the rank-frequency curves after standardization. This is achieved by dividing ranks for  
17 frequency by the expected number of species (Fig. 3). In symbols, let  $R_{ij}$  be the rank of species  $j$  in  
18 neighbourhood  $i$ . Then  $R'_{ij}$ , the scaled rank is defined as

$$19 \quad R'_{ij} = R_{ij} / \sum_j f'_{ij} .$$

20 Note that  $\sum_j f'_{ij}$  is the expected number of species after rescaling to standard sampling effort.

## 21 ESTIMATING CHANGE IN SPECIES FREQUENCY

22 For reasons that are explained in the discussion, a simple Poisson model of species discovery on  
23 individual visits is not realistic. A general multiplicative model to estimate a time factor  $x_{jt}$  for species  
24  $j$  is

$$25 \quad P_{ijt} = 1 - \exp(-Q_{ijt} x_{jt}) \quad (2)$$

26 where  $P_{ijt}$  is the probability that species  $j$  is recorded in hectad  $i$  at time  $t$ , and  $Q_{ijt}$  depends on the  
27 recording effort in hectad  $i$  at time  $t$  and on a time-independent probability of species  $j$  being found in

1 hectad  $i$ . For the analysis presented here, a simple model of  $Q_{ijt}$  is used, based on the proportion of  
 2 common ('benchmark') species found in hectad  $i$  and the estimated probability  $f'_{ij}$  of finding species  $j$   
 3 there. Benchmark species are defined to be those for which

$$4 \quad R'_{ij} < R^*$$

5 where  $R^*$  is a standard value. In the following example,  $R^*$  was chosen as 0.27, which has the  
 6 interpretation that the top 27% of the expected number of species are treated as benchmark species. If  
 7 the rescaled curves have the average of the shapes shown in Fig. 3, then for a neighbourhood with  
 8 standard sampling effort, one would expect to 99.2% of benchmark species to be present, and the least  
 9 frequent benchmark species would have frequency  $f'_{ij}$  about 97.5%.

10

11 Let  $s_{it}$  be the proportion of benchmark species found in hectad  $i$  at time  $t$ ; it is an approximate  
 12 measure of sampling intensity. Then for the simple model

$$13 \quad Q_{ijt} = -\log(1 - s_{it}f'_{ij}) \quad \text{if } s_{it}f'_{ij} < 0.98 \quad (3)$$

$$14 \quad = -\log(1 - 0.98) = 3.91 \quad \text{otherwise.}$$

15  $Q_{ijt}$  is truncated because for very common species in very well sampled hectads,  $s_{it}f'_{ij}$  can assume the  
 16 value 1. The time factor  $x_{jt}$  is estimated as the value of  $x_{jt}$  for which

$$17 \quad \sum_i P_{ijt} = \sum_i Q_{ijt}$$

18 It can be calculated iteratively, starting with a trial value of 1, and multiplying at each iteration by

$$19 \quad \frac{\sum_i P_{ijt}}{\sum_i Q_{ijt}}.$$

## 20 **Results**

### 21 ESTIMATION OF RECORDING EFFORT

22 According to the logic of rescaling frequencies, neighbourhood recording effort  $s_{i(N)}$  is measured by  
 23 the reciprocal of the sampling effort multiplier  $1/\alpha_i$  required to standardize  $\phi_i[\alpha_i]$  to  $\Phi$ . For the hectads  
 24 in Table 1, this varied from 0.08 for the Coleraine neighbourhood to 1.12 for the Ben Cruachan  
 25 neighbourhood. When local frequencies were calculated using the recorded species composition of  
 26 the target hectads, the picture was more complex (Fig. 4). Four of the target hectads had no records

1 and therefore no mean local frequency; they are omitted. The Crossmaglen hectad stands out as  
2 unusual, having only 19 recorded species but a relatively low mean frequency. The 19 species are  
3 from six separate localities, visited in 1929, 1949, 1950, 1952, 1980 and 1991. Most of them date  
4 from before 1960 when the British Bryological Society started systematic recording.

#### 5 STANDARDIZED RANK-FREQUENCY CURVES

6 Neighbourhood frequencies were mostly inflated by the standardization procedure, but in the case of  
7 the Ben Cruachan neighbourhood they were slightly deflated. After rescaling, the originally disparate  
8 rank-frequency curves of Fig. 2 appear much more similar (Fig. 3). Curves for the most poorly-  
9 sampled neighbourhoods had somewhat shorter tails because the rarest species were less likely to be  
10 encountered. If the average curve for relatively well-sampled neighbourhoods, i.e. those with  $\alpha_i < 2.0$ ,  
11 is calculated for the whole dataset, average standardized discovery rates can be calculated as

$$12 \quad \lambda(R') = -\log(1-f'(R'))$$

13 where  $f'$  is the mean standardized frequency for the average curve. For  $R'$  in the range 0.1 to 1.8 there  
14 was a nearly linear relation between  $R'$  and  $\log(\lambda)$ . This allows the average curve to be well  
15 approximated by an equation of the form

$$16 \quad f'(R') = 1 - \exp(-\exp(a+b R'))$$

17 The fitted constants  $a$  and  $b$  were in this case 2.005 and -2.545.

#### 18 RECOGNITION OF UNUSUAL SPECIES OR MISSING SPECIES

19 The hectad with the largest number of unusual species was SH62, in the Rhinog Mountains (Table 1).  
20 This hectad also had the greatest excess of observed species over the number expected. In the  
21 lowlands, the Boxworth hectad TL36 also had several locally scarce species, of which the three  
22 scarcest were *Racomitrium heterostichum*, *R. ericoides* and *Grimmia trichophylla*, with probabilities  
23 8%, 8% and 9%. These Grimmiaceae are normally plants of acid rock in the uplands. All were on  
24 artificial imported substrates, the *Racomitrium* species being on clinker of a sewage works and *G.*  
25 *trichophylla* on a memorial rock brought from the Cheviot to Boxworth churchyard.

26

1 In the other direction, two Scottish and two Irish hectads were completely unrecorded (Fig. 5).  
2 Inevitably they had a large number of omissions. As measured by omissions, the best-recorded hectad  
3 was that at Boxworth, with only four unrecorded species falling above the threshold  $f'_{ij} = 70\%$ .

#### 4 DOCUMENTING SPECIES CHANGE

5 The reappearance of epiphytes as a result of cleaner air provides a good example of change in  
6 frequency over time (Fig. 6). There is a suggestion of a slight decline over the period 1960-1975,  
7 followed by a strong recovery after 1990. Time factors were calculated for the whole of Britain and  
8 Ireland; if they were calculated for those districts that had experienced severe air pollution, the  
9 increase would appear even more marked.

## 10 **Discussion**

### 11 THE RECORDER EFFORT PROBLEM

12 The recorder effort problem has vexed many investigators, causing them to make ad hoc adjustments  
13 such as rejection of data from locations there were not adequately sampled at both an earlier and a  
14 later time. The novel feature of the neighbourhood-frequency method is that it pools data from  
15 neighbouring locations, so that the fact that a particular location was not sampled adequately at one  
16 date does not mean that the information from the other date is wasted. During the first recording  
17 scheme for bryophytes, it took 30 years to achieve adequate cover of Great Britain (Hill *et al.*, 1991-  
18 1994). During that period, an individual county might be recorded intensively only for a period of ten  
19 years. Recording was therefore very patchy in space and time. A method that can cope with such  
20 patchiness is needed.

21

22 Uneven recorder effort is not the only source of bias. Recorders differ widely in competence.  
23 Recording may be affected by the weather and by the seasonal apparency of species. The advice of  
24 Rich & Smith (1996) to note the time spent in recording is helpful but not sufficient. In the end, what

1 matters is not the effort that went in, but how successful it was. The methods advocated here depend  
2 on the outcome of recording, not the input of effort.

3

4 Most current methods for estimating species richness start from the assumption that there is a definite  
5 number of species at each location (Colwell and Coddington, 1994, Magnussen *et al.*, 2010).

6 However, if records are kept over a long period, this assumption is not realistic. Species come and go.  
7 They may be temporarily resident or may opportunistically immigrate from surrounding favourable  
8 habitat (Owen, 2010). Thus a probabilistic model of species occurrence and discovery may be just as  
9 realistic as a deterministic one.

10

11 A strength of the neighbourhood-frequency method is that it requires no extra information beyond  
12 species lists from neighbouring localities. It compensates for data incompleteness by using a species-  
13 richness estimator (Hortal, 2008). There are two key assumptions, namely that the probability of  
14 finding a species at a locality can be estimated by its frequency in the neighbourhood, and that in a  
15 well-sampled neighbourhood the weighted mean frequency is a constant, independent of species  
16 richness. A potential weakness of the method is that some neighbourhoods may be so poorly recorded  
17 that they provide no information about local species frequency. If such neighbourhoods exist, then  
18 they should be omitted or neighbourhood size should be increased.

#### 19 NEIGHBOURHOODS AND THE PROBABILITY OF OCCURRENCE

20 Neighbouring hectads can give an indication of how likely a species is to be found in the target hectad  
21 only if the target is reasonably similar to its neighbours. If the target is not similar to its neighbours,  
22 then the key approximation of equation (1) will not hold. The reason for using multiplicative weights  
23 here is simply that purely spatial smoothing may include very dissimilar entities such as a sandy coast  
24 and an inland mountain, whereas purely floristic weights may result in a ‘neighbourhood’ scattered  
25 over several hundred kilometres.

1

2 The use of neighbourhoods to define a smoothing window is well established, both in physical space  
3 to show local frequency (Gibbons *et al.*, 1993) and in environmental space to define climate response  
4 surfaces (Huntley *et al.*, 1995). Neighbourhoods do not have to have low weight near their  
5 boundaries. Prendergast *et al.* (1993) used unweighted blocks of 25 hectads to define their  
6 neighbourhoods. The smoothing kernel  $(1-x^2)^4$  used here is hardly distinguishable in practice from a  
7 Gaussian function. It uses ranks rather than actual distances in order to ensure that remote islands  
8 also have a good supply of neighbours. For floristic similarity, the Sørensen coefficient was chosen  
9 because it measures overall similarity, so that if the flora of one hectad is a small subset of that of  
10 another, then the two hectads have low floristic similarity.

11

12 Hectads can be exceptional in various ways. We have already seen that a hectad in the Rhinog  
13 Mountains was markedly richer in species than its neighbours (Fig. 5). In the other direction, hectads  
14 with little land are likely to have fewer species than their neighbours would indicate. These factors  
15 can be mitigated by using floristic similarity as well as spatial proximity to define the neighbourhood.  
16 Thus the hectads that are similar to Boxworth lie mainly to the west of it (Fig. 1); those to the north  
17 and east of it are in the East Anglian fen country, which is very flat and lacks ancient woodland.  
18 Likewise the neighbourhood of Blackpool lies mostly along the coast and not in the Lancashire  
19 uplands.

20

21 There is undoubtedly scope for improving the definition of neighbourhoods, especially in the  
22 mountains, where maximum altitude and the number of wet days per annum (Ratcliffe, 1968) can give  
23 useful information. There is, of course, no necessity to use floristic similarity in defining  
24 neighbourhoods. Similarity of physical attributes might be preferable, but a suitable dataset for both  
25 Britain and Ireland was not readily available.

## 1 STANDARDIZATION OF LOCAL FREQUENCIES

2 The method of standardizing local frequencies depends on three assumptions: that the target hectad  
3 resembles its neighbours; that species discovery is basically a Poisson process; and that in a well-  
4 worked neighbourhood there is a characteristic mean species frequency. The first of these  
5 assumptions has already been discussed. The second is not true in detail but is plausible enough at the  
6 relatively large scale of recording visits to the neighbourhood. Recorders visiting an area can search it  
7 in various ways, but they are always constrained by their inability to make a random search of the  
8 whole hectad. Thus, on any one visit, a recorder would not ordinarily expect to cover the whole range  
9 of habitats in the hectad, let alone search its whole area. This will apply to visits to neighbouring  
10 hectads just as much as to the target hectad. A more thorough study of recorder behaviour would no  
11 doubt reveal that visits come in many kinds, ranging from the quick examination of a wall to a full  
12 day's systematic recording of a range of habitats. Thus the chance of a given species being  
13 discovered is the outcome of a two-stage stochastic process. The first stage concerns the type and  
14 duration of the visit, while the second concerns the frequency with which a given species is  
15 encountered during a visit of a certain type. When these processes are combined, each species will  
16 have a standard probability of being recorded on a visit. Under most assumptions about the nature of  
17 this two-stage process, the discovery of less common species will be a rare event, so that their records  
18 can indeed be taken as the outcome of a Poisson process.

19

20 The third assumption is at present based on limited data, namely that the untransformed  
21 neighbourhood frequency in well-recorded neighbourhoods was in the range  $\phi_i[1]=0.67$  to  $0.75$ . On  
22 closer inspection, many of these neighbourhoods were found to include some hectads that were not  
23 particularly well recorded. Thus the standard value  $\Phi=0.74$  was selected. In most applications, the  
24 exact choice of  $\Phi$  is not critical. If frequencies follow the average curve found here, then lowering  $\Phi$   
25 to  $0.7$  reduces the expected number of species by  $2.2\%$ , while raising it to  $0.9$  increases the expected  
26 number by  $7.6\%$ . However, the question of whether the proportion of locally rare to locally common  
27 species is the same in different parts of the country deserves further investigation. The fact that  
28 observed species totals for well-sampled hectads Ben Cruachan in western Scotland, Fordingbridge in

1 southern England and Boxworth in eastern England are in good agreement with predictions (Fig. 5) is  
2 encouraging but amounts to rather limited evidence.

3

4 Multipliers  $\alpha_i$  for sampling effort are closely related to neighbourhood frequency  $\varphi_i[1]$  (Fig. 7). The  
5 scale for  $\varphi_i[1]$  is shown as a reciprocal to emphasize the discrepancy between  $1/\varphi_i[1]$ , which is the  
6 naïve estimate of  $\alpha_i$ , and the relatively large values of  $\alpha_i$  that apply in poorly-recorded  
7 neighbourhoods. By definition, a sampling-effort multiplier of 1 corresponds to a neighbourhood  
8 frequency of  $\Phi$ . The trend line in Fig. 7 is calculated from data for all 3,854 neighbourhoods. It  
9 crosses the x-axis ( $\alpha_i = 0$ ) at  $1/\varphi_i[1] = 1.164$ . This implies a value  $\varphi_i[1]=0.86$  if the neighbourhood  
10 were exhaustively sampled. This is not intrinsically unlikely. When rank-frequency curves and  
11 corresponding statistics were calculated for vascular plants, which are much better recorded than  
12 bryophytes, the average value of  $\varphi_i[1]$  was 0.80 and the largest value, for Stoke Mandeville, was 0.86.  
13 This neighbourhood included 293 vascular plant species with local frequency exceeding 0.99, at  
14 which point  $R'_{ij} = 0.45$ . Bryophytes with standardized local frequency exceeding 0.99 were relatively  
15 few (Fig. 3), and the boundary for such species on the average curve was  $R'_{ij} = 0.18$ .

#### 16 IMPROBABLE AND MISSING SPECIES

17 The estimated local frequency of the rarest species is raised by the inclusion of target hectads in their  
18 own neighbourhoods. Prendergast *et al.* (1993) excluded target hectads (which they called reference  
19 hectads) from their neighbourhoods. For the present study, target hectads were included because the  
20 resulting smoothed frequencies are used for other purposes, notably interpreting trends. It would be  
21 paradoxical for locally rare species such as *Amblystegium radicale* in the Rhinog Mountains to have a  
22 positive probability in nearby hectads but a zero probability in the target.

23

24 Calculated probability values are used for routine screening of bryophyte records received by the  
25 recording scheme of the British Bryological Society. Records that are new for their vice-county (Hill  
26 *et al.*, 2008) are automatically queried. Records of locally scarce species are now also scrutinized and  
27 may be queried unless the species is already known from the hectad.

1 ESTIMATION OF TRENDS

2 The big advantage of the neighbourhood-frequency method in estimating trends is that it corrects for  
3 uneven spatial recording over time. Methods that are based on global frequency as a proportion of all  
4 records can be misleading if sampling is spatially uneven. Indeed, species can apparently decrease  
5 when they are actually increasing (an example is given in the supporting information). The advantage  
6 of the method over ones that use the same benchmark species throughout the territory is that local  
7 benchmark species are selected to be suitable for each neighbourhood.

8

9 Trend estimation depends on the choice of benchmark species, and therefore on the parameter  $R^*$ , the  
10 benchmark limit. In a sensitivity analysis,  $R^*$  was varied from 0.14 to 0.41. Time factors  $x_{jt}$  increased  
11 when  $R^*$  was larger. This is because sampling intensity  $s_{it}$  is measured by the proportion of  
12 benchmark species found in location  $i$  at time  $t$ . If  $R^*$  is increased, then relatively rarer species are  
13 added to the set of benchmark species, and  $s_{it}$  accordingly decreases. To make up for the fall in  $s_{it}$ ,  
14 time factors  $x_{jt}$  need to be increased, roughly in proportion. However, the pattern of proportional  
15 change was almost completely unaffected by variation in  $R^*$ , with time factors merely being  
16 multiplied by a constant (see supporting information).

17

18 In the introduction, it was asserted that benchmark species should be ones that are fairly stable over  
19 time. When the data were reanalysed after exclusion of strongly increasing or decreasing species  
20 from the list of possible benchmarks, the new time factors hardly differed from the old ones. This  
21 revised analysis is therefore not reported here, but with other datasets the exclusion of such species  
22 might make a difference.

23

24 The basic model (2)

25 
$$P_{ijt} = 1 - \exp(-Q_{ijt} x_{jt})$$

26 can be varied in many ways depending on the definition of  $Q_{ijt}$ . One variant is

27 
$$P_{ijt} = 1 - \exp(-\lambda_{ij} s_{it} x_{jt}) \tag{4}$$

1 which is based on multiplication of three parameters.  $\lambda_{ij}$  is the encounter rate for species  $j$  in hectad  $i$ ;  
2  $s_{it}$  is the recording effort in hectad  $i$  at time  $t$ ; and  $x_{jt}$  is a time factor for species  $j$ . If species  $j$  is  
3 ubiquitous in the neighbourhood of hectad  $i$ , then  $f_{ij} = f'_{ij} = 1$ , and the conversion

$$4 \quad \lambda_{ij} = -\log(1 - f'_{ij})$$

5 makes  $\lambda_{ij}$  infinite, so that  $P_{ijt} = 1$ , regardless of the values of the other parameters. In practice, no  
6 species is so common that it is sure to be found on a very short visit. Moreover, recorders do not  
7 always make complete lists and sometimes omit the commonest species. Thus for analysing trends,  
8 the assumption of ubiquity should not be made, and model (3) may in fact be preferable to model (4).

#### 9 RECORDING EFFORT AND DISCOVERY CURVES

10 In the background of much of the above discussion are three important concepts: visits, discovery  
11 curves and recording effort. Prendergast *et al.* (1993) corrected for recording effort by means of  
12 discovery curves for individual visits. These were obtained by rarefaction. In the present study,  
13 recording intensity for an individual locality and time period is measured by the proportion of  
14 benchmark species found at that time. Discovery is assumed to be a random process, derived by  
15 sampling a neighbourhood frequency distribution.

16

17 Species lists from individual visits provide information at a higher resolution than the summarized  
18 hectad data analysed here. When data from individual visits are available, the total number of records  
19 can be treated as a measure of recording effort (Ball & Henshall, 2006). This measure is rarely  
20 available for historical datasets. Until about 1990, many vascular plant and bryophyte records were  
21 summarized on master cards, which did not distinguish visits. Even in more recent years, botanists  
22 have sometimes summarized their local flora using tetrads (2-km squares) or quadrants (5-km  
23 squares), and have not necessarily distinguished individual visits within these units. Therefore, the  
24 summary of occurrences by hectad and quinquennium to underpin the analysis for Fig. 6 may give the  
25 best detail that could be recovered without serious loss of data.

26

1 Except in comprehensive and systematic surveys, the problem of variable habitat coverage cannot be  
2 ignored. For example, there may be a targeted habitat survey such as the Scottish Loch Survey 1984-  
3 1994, data from which were incorporated in the BRC database in the late 1990s (Preston and Croft,  
4 1997). This resulted in aquatic plants seeming to increase in Scotland (Preston *et al.*, 2003).  
5 Likewise, the Survey of Bryophytes of Arable Land (Preston *et al.*, 2010) resulted in a large increase  
6 of bryophyte records from this habitat during the period 2001-5. Most such biases are less obvious,  
7 with recorders during a particular period tending, for example, to avoid towns and villages or being  
8 reluctant to record non-native plants.

9

10 A fruitful topic for future study will be to understand and correct for such changes in recorder  
11 behaviour. This will require individual visits to be analysed in more detail, distinguishing the depth  
12 of recording within visits and the coverage achieved by the overall spread of visits. This should allow  
13 the probability of species being found to be modelled better than by model 3 above. There is no  
14 necessity to use benchmark species. For example, if a list of ten rare species is received from a visit,  
15 then the recorder has clearly omitted to make a full list or has selected a subset of what was found for  
16 a museum or private herbarium. Such partial lists can provide information. Given the habitat and  
17 local frequency of the species in such a list, the probability of a particular species being included  
18 could in principle be estimated, but that would require a much more complicated model than the one  
19 used here.

## 20 **Acknowledgements**

21 This study was funded jointly by the Joint Nature Conservancy Committee (JNCC) and the Natural  
22 Environment Research Council as a part of the Biological Records Centre (BRC) work programme  
23 for 2010. I am grateful to Chris Cheffings (JNCC), David Roy (BRC) and Lawrence Way (JNCC) for  
24 their encouragement, and to the British Bryological Society and the Botanical Society of the British  
25 Isles for assembling such a wealth of data. Colin Harrower of BRC helped with database queries and  
26 prepared the map of neighbourhood weights, which was plotted using the DMAP program written by

1 Dr Alan Morton. My colleagues Nick Isaac, Tom Oliver and Chris Preston read over and commented  
2 on draft reports as the work progressed. Finally I would like to thank five referees and an associate  
3 editor for valuable and constructive comments on earlier drafts.

#### 4 **References**

- 5 Asher, J., Warren, M., Fox, R., Harding, P., Jeffcoate, G. & Jeffcoate, S. (2001) *The millennium atlas*  
6 *of butterflies in Britain and Ireland*. Oxford University Press, Oxford.
- 7 Ball, S. & Henshall, M. (2006) Using data to interpret changes in the UK's biodiversity. *The Bulletin*  
8 *of the British Ecological Society*, **37**, 51-54.
- 9 Boakes, E.H., McGowan, P.J.K., Fuller, R.A., Ding, C.Q., Clark, N.E., O'Connor, K. & Mace, G.M.  
10 (2010) Distorted views of biodiversity: spatial and temporal bias in species occurrence data.  
11 *Plos Biology*, **8**.
- 12 Colwell, R.K. & Coddington, J.A. (1994) Estimating terrestrial biodiversity through extrapolation.  
13 *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **345**,  
14 101-118.
- 15 Dennis, R.L.H. & Thomas, C.D. (2000) Bias in butterfly distribution maps: the Influence of hot spots  
16 and recorder's home range. *Journal of Insect Conservation*, **4**, 73-77.
- 17 Free Software Foundation (1999). *GNU Fortran G77 v0.5.25*. Downloaded 2010 from  
18 <http://kkourakis.tripod.com/g77.htm>.
- 19 Gibbons, D. W., Reid, J. B. & Chapman, R. A. (1993) *The New Atlas of Breeding Birds in Britain and*  
20 *Ireland: 1988-1991*. T. & A.D. Poyser, London.
- 21 Hassall, C. & Thompson, D.J. (2010) Accounting for recorder effort in the detection of range shifts  
22 from historical data. *Methods in Ecology and Evolution*, **1**, 343-350.
- 23 Hedenäs, L., Bisang, I., Tehler, A., Hamnede, M., Jaederfelt, K. & Odelvik, G. (2002) A herbarium-  
24 based method for estimates of temporal frequency changes: mosses in Sweden. *Biological*  
25 *Conservation*, **105**, 321-331.

- 1 Hill, M.O. (1973) Diversity and evenness: a unifying notation and its consequences. *Ecology*, **54**, 427-  
2 432.
- 3 Hill, M. O., Blackstock, T. H., Long, D. G. & Rothero, G. P. (2008) *A checklist and census catalogue*  
4 *of British and Irish bryophytes updated 2008* British Bryological Society.
- 5 Hill, M. O., Preston, C. D. & Smith, A. J. E. (1991-1994) *Atlas of the bryophytes of Britain and*  
6 *Ireland, Vol. 1 (1991), Vol. 2 (1992), Vol. 3 (1994)*. Harley Books, Colchester.
- 7 Hofmann, H., Urmi, E., Bisang, I., Müller, N., Küchler, M., Schnyder, N. & Schubiger, C. (2007)  
8 Retrospective assessment of frequency changes in Swiss bryophytes over the last two  
9 centuries. *Lindbergia*, **32**, 18-32.
- 10 Hortal, J. (2008) Uncertainty and the measurement of terrestrial biodiversity gradients. *Journal of*  
11 *Biogeography*, **35**, 1335-1336.
- 12 Huntley, B., Berry, P. M., Cramer, W. & McDonald, A. P. (1995) Modelling present and potential  
13 future ranges of some European higher plants using climate response surfaces. *Journal of*  
14 *Biogeography*, **22**, 967-1001.
- 15 Jeppsson, T., Lindhe, A., Gärdenfors, U. & Forslund, P. (2010) The use of historical collections to  
16 estimate population trends: a case study using Swedish longhorn beetles (Coleoptera:  
17 Cerambycidae). *Biological Conservation*, **143**, 1940-1950.
- 18 Legendre, P. & Legendre, L. (1998) *Numerical ecology, 2nd English edition*. Elsevier, Amsterdam.
- 19 Maes, D. & Swaay, C. A. M. v. (1997) A new methodology for compiling national Red Lists applied  
20 to butterflies (Lepidoptera, Rhopalocera) in Flanders (N-Belgium) and the Netherlands.  
21 *Journal of Insect Conservation*, **1**, 113-124.
- 22 Magnussen, S., Smith, B., Kleinn, C. & Sun, I.F. (2010) An urn model for species richness estimation  
23 in quadrat sampling from fixed-area populations. *Forestry*, **83**, 293-306.
- 24 Newson, S.E., Woodburn, R.J.W., Noble, D.G., Baillie, S.R. & Gregory, R.D. (2005) Evaluating the  
25 Breeding Bird Survey for producing national population size and density estimates. *Bird*  
26 *Study*, **52**, 42-54.
- 27 Owen, J. (2010) *Wildlife of a garden: A thirty-year study*. Royal Horticultural Society, London.

- 1 Petrik, P., Pergl, J. & Wild, J. (2010) Recording effort biases the species richness cited in plant  
2 distribution atlases. *Perspectives in Plant Ecology Evolution and Systematics*, **12**, 57-65.
- 3 Prendergast, J. R., Wood, S. N., Lawton, J. H. & Eversham, B. C. (1993) Correcting for variation in  
4 recording effort in analyses of diversity hotspots. *Biodiversity Letters*, **1**, 39-53.
- 5 Preston, C. D. & Croft, J. M. (1997) *Aquatic Plants in Britain and Ireland*. Harley Books, Colchester.
- 6 Preston, C. D., Hill, M. O., Porley, R. D. & Bosanquet, S. D. S. (2010) Survey of the bryophytes of  
7 arable land in Britain and Ireland 1: a classification of arable field assemblages. *Journal of*  
8 *Bryology*, **32**, 61-79.
- 9 Preston, C. D., Pearman, D. A. & Dines, T. D. (2002) *New atlas of the British and Irish Flora*. Oxford  
10 University Press, Oxford.
- 11 Preston, C. D., Telfer, M. G., Arnold, H. R., Carey, P. D., Cooper, J. M., Dines, T. D., Hill, M. O.,  
12 Pearman, D. A., Roy, D. B. & Smart, S. M. (2002) *The changing flora of the UK*. DEFRA,  
13 London.
- 14 Preston, C. D., Telfer, M. G., Roy, D. B., Carey, P. D., Hill, M. O., Meek, W. R., Rothery, P., Smart,  
15 S. M., Smith, G. M., Walker, K. J. & Pearman, D. A. (2003) The changing distribution of the  
16 flora of the United Kingdom: technical report. CEH Monks Wood, Huntingdon.
- 17 Ratcliffe, D. A. (1968) An ecological account of Atlantic bryophytes in the British Isles. *New*  
18 *Phytologist*, **67**, 365-439.
- 19 Rich, T.C.G. & Smith, P.A. (1996) Botanical recording, distribution maps and species frequency.  
20 *Watsonia*, **21**, 155-167.
- 21 Roy, D.B., Rothery, P. & Brereton, T. (2007) Reduced-effort schemes for monitoring butterfly  
22 populations. *Journal of Applied Ecology*, **44**, 993–1000.
- 23 Silvertown, J. (2009) A new dawn for citizen science. *Trends in Ecology & Evolution*, **24**, 467-471.
- 24 Telfer, M. G., Preston, C. D. & Rothery, P. (2002) A general method for the calculation of relative  
25 change in range size from biological atlas data. *Biological Conservation*, **107**, 99-109.

Hectad	Location	Sea	$\phi_i[1]$	$\alpha_i$	Obs	Exp	Absent	Improb
NB13	NW Lewis	39	0.53	3.5	90	292	116	
ND04	Flow country		0.41	5.7	32	295	181	
NG63	Scalpay nr Skye	88	0.63	1.9	-	301	223	
NH66	Cromarty Firth	18	0.47	4.3	100	340	147	1
NJ93	Ellon		0.44	6.7	86	289	124	
NN03	Ben Cruachan	14	0.75	0.9	337	329	10	21
NO03	Bankfoot		0.59	3.0	129	327	117	3
NR59	Jura	88	0.56	2.6	-	331	238	
NS67	nr Glasgow		0.53	3.6	164	349	113	6
NT68	nr Dunbar	86	0.55	3.4	151	334	98	
NX87	W of Dumfries		0.51	4.0	69	320	169	
NY88	Bellingham		0.69	1.3	312	285	10	17
SD33	Blackpool	1	0.51	4.0	155	312	77	1
SE33	E Leeds urban		0.51	4.2	63	293	143	
SH62	Rhinog Mts		0.74	1.0	402	340	10	36
SJ68	Warrington		0.60	3.7	114	264	88	5
SK68	Worksop		0.48	5.2	85	235	90	
SN71	Black Mts		0.69	1.4	384	342	14	17
SO71	nr Gloucester	1	0.62	2.4	172	265	46	2
SP71	Stoke Mandeville		0.60	2.8	72	199	71	1
ST10	Honiton		0.70	1.5	195	269	28	6
SU11	Fordingbridge		0.71	1.3	263	268	8	1
SW54	St Ives	89	0.72	1.1	225	301	51	4
SY67	Portland	77	0.63	2.3	156	282	78	5
TF37	E of Lincoln		0.58	3.3	130	203	33	3
TL36	Boxworth		0.68	1.5	184	187	4	10
TM45	Aldeburgh	39	0.67	1.9	166	205	19	7
TQ93	Weald nr Ashford		0.67	1.8	185	251	38	2
C92	SE of Coleraine		0.17	12.6	-	303	215	
G80	nr Carrick		0.41	6.0	102	300	126	
H80	nr Crossmaglen		0.29	10.2	19	246	159	
L82	Galway Bay	60	0.34	6.7	84	368	179	
M84	W of Athlone		0.27	9.9	70	201	85	
N84	Kilcock, Meath		0.30	9.2	42	204	110	
R21	Mullaghareirk Mts		0.20	12.1	153	265	71	7
S21	Comeragh Mts		0.34	6.3	266	329	28	
T25	Gorey, SE coast	93	0.25	10.3	35	207	112	
W58	nr Cork		0.21	12.6	-	268	196	

**Table 1.** Systematic sample of one in a hundred hectads in Britain (4-character hectads) and Ireland (3-character hectads). Column headings are Hectad – grid reference with Ordnance Survey lettering for 100-km squares; Sea – Sea area km<sup>2</sup>;  $\phi_i[1]$  – neighbourhood mean frequency before multiplication;  $\alpha_i$  – sampling-effort multiplier (see text); Obs – Observed number of bryophyte taxa; Exp – Predicted number; Absent – taxa predicted to be present with at least 70% probability that were not recorded; Improb – taxa with less than 20% probability that were actually recorded.

## CAPTIONS FOR THE FIGURES

**Figure 1.** Neighbourhoods of selected hectads (shown with crosses) using both floristic similarity and spatial proximity. Circles of decreasing size represent neighbouring hectads, with weights 0.7-1.0, 0.2-0.7 and 0.05-0.2. The selected hectads are, from north to south, Flow country, Ben Cruachan, Coleraine, Blackpool and Boxworth.

**Figure 2.** Rank-frequency curves for bryophytes in the five neighbourhoods shown in Figure 1

**Figure 3.** Rank-frequency curves for bryophytes in the five neighbourhoods shown in Figures 1 and 2 after adjustment of neighbourhood frequency by sampling-effort multipliers and of species rank by division by the expected species number  $\sum_i f'_{ij}$ .

**Figure 4.** Mean local frequency of species actually found in target hectads in relation to total number recorded, expressed as a proportion of the expected total. The trend line shows the mean frequencies that would be expected if the probability of discovery followed the average rank-frequency curves after scaling as in Figure 3.

**Figure 5.** Observed species numbers in hectads in relation to numbers predicted. The diagonal line shows where observed and predicted are equal. Geographical regions are distinguished by symbols: GB1 – Scotland and Northumberland; GB2 – Wales, southern and western England; GB3 – eastern England; IRL – Ireland.

**Figure 6.** Change in frequency of three epiphytic mosses over period 1960-2010. Time factors for each species are calculated for quinquennia 1960-64, 1965-69 etc.

**Figure 7.** Sampling-effort multipliers  $\alpha_i$  in relation to reciprocal mean neighbourhood frequencies  $1/\phi_i[1]$ . The trend line is a quartic curve calculated for the full hectad dataset, not just the 38 selected as a sample.













