Predicting abundance from occupancy: a test for an aggregated insect assemblage

M. WARREN†, M. A. McGEOCH‡ and S. L. CHOWN§
†Department of Zoology and Entomology, University of Pretoria, Pretoria, 0002, South Africa, mwarren@zoology.up.ac.za, ‡Department of Conservation Ecology, University of Stellenbosch, Private Bag X1, Matieland 7602, South Africa, mcgeoch@sun.ac.za, §Spatial, Physiological and Conservation Ecology Group, Department of Zoology, University of Stellenbosch, Private Bag X1, Matieland 7602, South Africa, slchown@sun.ac.za

Summary

1. The ubiquitous, positive abundance-occupancy relationship is of potential value to conservation and pest management because of the possibility of using it to predict species abundance from occupancy measures.

2. He & Gaston (2000a) developed a model, and a parameterization method, for the prediction of abundance from occupancy based on the negative binomial distribution. There are to date few empirical tests of either the estimation method or model. Here we conduct such a test in a field-based mesocosm experiment using a Drosophilidae assemblage associated with decaying fruit.

3. With individual (and groups of) fruit as minimum mapping units, abundance estimates derived using the parameterization method of the He-Gaston model differed significantly from measured values, and were least accurate for the most abundant species.

4. Substitution of $k$-values corrected for species density in the model did not improve abundance predictions significantly. However, substitution of $k$-values calculated directly from the negative binomial distribution yielded highly accurate abundance predictions.

5. Although the distribution of fly species did not deviate significantly from the negative binomial distribution, and the finest possible minimum mapping units were used (individual fruit), the parameterization method in the He-Gaston model consistently underestimated the abundance of species in the assemblage because individuals were very highly aggregated within fruit.

6. Because of its potential importance, this model and parameterization method require further exploration at fine scales, commonly represented by individual habitat units, for highly aggregated species. The incorporation of spatially explicit information may provide a means of improving abundance predictions in this regard.

Key-words: abundance–occupancy relationship, aggregation, Drosophilidae, negative binomial distribution, spatial heterogeneity.

Introduction

The ubiquity of positive abundance-occupancy relationships has engendered considerable interest in the causes of this widespread relationship. Indeed, both theoretical and empirical explanations for the relationship abound (see Gaston, Blackburn & Lawton 1997; Holt, Gaston & He 2002a; for review). Ecological mechanisms include differences in resource availability, species resource breadth, population characteristics and biotic interactions (e.g. Brown 1984; Warren & Gaston 1997; Gaston et al. 2000; Holt, Warren & Gaston 2002b), while several mechanisms based on the spatial distribution of individuals have also been proposed (Hanski, Kouki & Halkka 1993; He & Gaston 2000a,b; Holt et al. 2002a). The relationship also has considerable intrinsic value because it can be used to predict species abundance levels from measures of occupancy (Kunin 1998; He & Gaston 2000a). Successful prediction of
Predicting abundance from occupancy

To predict species abundance from occupancy an appropriate description of the underlying spatial distribution of its individuals is required. The negative binomial distribution (NBD) has most frequently been used, because individuals commonly have contagious distributions (Pielou 1977; Taylor, Woiwood & Perry 1978; Legg et al. 1992). However, the relationship between abundance and occupancy is also well described by a range of other models, including poisson, power and logistic models, as well as variations thereof (Bliss & Fisher 1953; Pielou 1977; Nachman 1981; He, Gaston & Wu 2002; Holt et al. 2002a). Recently, Kunin (1998), Harte, Kinsig & Green (1999) and Harte, Blackburn & Ostling (2001) suggested that self-similarity, and the comparison of abundance–occupancy patterns across spatial scales, might be used to further understand the relationship. Taking Kunin’s (1998) rationale of using cross-scale occupancy comparisons a step further, He & Gaston (2000a) derived a model (hereafter the HG model) from the NBD to predict the abundances of species from measures of their occupancy. He & Gaston (2000a) parameterize the model using two spatial scales of occupancy data to estimate species abundance, simultaneously estimating the parameter k for the sampling extent of interest (hereafter the HG (parameterization) method). As the degree of aggregation of individuals increases, the parameter k of the NBD declines. Because k is estimated as part of the parameterization process, the model is generalisable to situations across the continuum of regular, random to aggregated species distributions (He & Gaston 2000a). In contrast to all previous models, the HG parameterization method requires no a priori estimates of aggregation, slope or intercept parameters for the abundance-occupancy relationship. The model is thus considered to be appropriate across a wide range of biological situations (He & Gaston 2000a,b). Indeed, the HG model was used to successfully predict abundance levels from occupancy for tree species (at a local scale) and passerine birds (at a regional scale) (He & Gaston 2000a,b; Kunin, Hartley & Lennon 2000).

Nonetheless, He & Gaston (2000b) have pointed out that the model has several limitations. For example, the efficiency of the model declines with increasing grain and extent, and model predictions are therefore expected to be best at fine scales. The value of the aggregation parameter, k, estimated by the HG method is thus sensitive to sampling extent (He & Gaston 2000b). The value tends to increase with spatial scale and this results in a simultaneous, disproportionate increase in the degree to which abundance is underestimated, i.e. at large scales species are predicted to be less aggregated, and abundances therefore lower, by the HG method than they are in reality (Kunin et al. 2000).

In addition, defining the most appropriate ‘finest’ sampling unit (also called minimum mapping units (MMUs)) for species remains problematic (He & Gaston 2000a; Kunin et al. 2000). The finest possible MMUs for a species would be those that are equivalent to the size of its individuals, where any occupancy measure would provide an exact measure of its abundance. Such an approach is sampling intensive and most often impracticable. In addition, MMUs defined in this way will also not necessarily permit comparisons between species of different sizes from the same assemblage (McGeoch & Gaston 2002). It is therefore not surprising that most species occupancy data are collected at scales far coarser than the size of individuals of the species concerned (McGeoch & Gaston 2002). Coarse-scale MMUs, however, fail to discern fine-scale distribution patterns (e.g. intraspecific aggregation), resulting in either over- or underestimates of abundance (Thomas & Abern 1995; He & Gaston 2000a; Kunin et al. 2000). Therefore, a compromise must be reached between sampling scales that are practical, and those that will provide a satisfactory estimate of species abundance.

The relative efficiency with which the HG method is able to predict abundance is also likely to differ for rare and common species. In butterflies, declines in the abundance of common species have been shown to go undetected at large scales, while the range sizes of rare species are overestimated at coarse scales (Thomas & Abern 1995; Cowley et al. 1999). Furthermore, k of the negative binomial distribution is known to be sensitive to species density and may need to be adjusted accordingly (Taylor, Woiwood & Perry 1979; but see Peng & Brewer 1994). Therefore, the HG method is unlikely to predict the abundance of rare and common species (Gaston 1994a) equally well at a particular scale.

Despite these potential drawbacks, the HG method is the only one available for predicting local species abundances in the absence of information on their aggregation or mean density (the model also appears to fit local scale data better than the fractal model of Kunin (Kunin et al. 2000)). In consequence, it has a wide range of potential applications in macroecology, conservation biology, and pest management (Gaston 1999). However, despite its potential importance, it remains to be empirically tested across a broad range of taxa and spatial scales. Indeed, with the exception of the tests based on the Pasoh tree and Bedfordshire passerine data, few other tests of this model and its parameterization method have been undertaken (He & Gaston 2000a). The objective here is therefore to provide an empirical test of the HG model and parameterization method using a Drosophilidae assemblage that is an exemplar of many systems characterized by a highly aggregated spatial distribution at local scales (Atkinson & Shorrocks 1984; Rosewell, Shorrocks & Edwards 1990). We first determine if the spatial
distributions of the species follow the NBD; providing a priori justification of the suitability of using the HG model. We then test the accuracy of the abundance predictions of the HG model for the species in this assemblage, defining the finest MMU as the single resource unit, i.e. a decaying fruit. Using accurate occupancy records determined in a mesocosm at three spatial scales (MMUs), we test if the abundances of rare and common species in the assemblage are predicted equally well. The effect of substituting k-values, calculated independently of the HG parameterization method, on model predictions are also assessed. Finally, we evaluate the effect of increasing MMU size on the accuracy of the abundance prediction by determining the scale at which the HG model prediction most closely matches the measured abundance of each species.

**Materials and methods**

A mesocosm was used to empirically test the abundance-occupancy model. Mesocosms have been found useful in community and macroecology because they are able to generate assemblages similar to those expected under non-experimental, field conditions (Warren & Gaston 1997; Relyea & Yurewicz 2002).

A Drosophilidae assemblage that inhabits decaying fruit (in this case nectarines (Prunus persicae Miller variety nectarina: Rosaceae)) was used. The fruit were washed and individually weighed before being placed in the field. Variation in fruit mass was small (Mean ± SE (g): 58.84 ± 0.05). Three small punctures were made in the skin of each fruit, because Drosophilidae do not lay on unbroken fruit surfaces (Feder & Krebs 1998). Five nectarines were randomly selected for insecticide residue tests and were found to have no detectable levels of residues of chemicals used in the local soft fruit industry (South African Bureau of Standards). A wire table (2.4 m × 3.6 m; 0.7 m high) was placed on the University of Pretoria’s Experimental Farm in Pretoria, South Africa (25°45’17.8”S, 28°15’29.3”E; Fig. 1). The table was placed inside a cage covered with pigeon wire to exclude birds, fruit-piercing moths and large wasps, while allowing the flies ready access to the fruit. Ants were also excluded by the application of grease to the Table 1 legs. The table was divided into six equal plots with each plot supporting 36 nectarines (n = 216 nectarines) (Fig. 1). Three of the plots in alternate rows of the two columns were artificially shaded with 80% shade netting to impose a level of microclimatic heterogeneity on the experiment (Fig. 1). The nectarines were placed 20 cm apart in a regular grid of 12 × 18 fruit, resembling a regular arrangement of plants in a field, where each plant would be the unit from which presence or absence (occupancy) of a species would be determined. Each fruit was placed on a plastic mesh in the centre of a round plastic container (15 cm diameter and 8 cm deep) containing washed, moist sand. Drosophilidae larvae drop down into the substrate under the fruit when they are ready to pupate after 4.5–8.5 days at 25 °C and 80% RH (Sevosten & van Alphen 1993). Larvae were therefore allowed to pupate in moist sand underneath the fruit. Starting from the fifth day after exposure of the fruit, the sand containing the fly pupae was removed and placed in 350 mL plastic jars every second day for 25 days in November 1998. The plastic jars were then taken to the laboratory and the emerging flies were identified (according to McEvey et al. 1988). Fresh, moist sand was immediately placed in the containers under the fruit at each collection.

Flies pupated over a period of 19 days (day 7–day 25). Fly emergence was 2.1 times higher on the 23rd day than any other puation date and the distribution-abundance patterns of individuals pupating on this date are examined here. The data used thus represent the largest cohort of flies simultaneously present in the fruit. At least six Drosophilidae species were found. However, specimens belonging to the genus Zaprionus could not be identified to species level. The identified species were: Drosophila simulans Sturtevant, D. melanogaster Meigen, D. buzzatii Coquillett, D. buzzatii Patterson & Wheeler, Zaprionus morphospecies group 1 (which may include both Z. tuberculatus Malloch and Z. sepoides Duda) and Zaprionus morphospecies group 2 (which may include Z. vittiger Coquillett and Z. indiana Gupta) (McEvey et al. 1988). Damaged specimens (n = 55 of 3029 flies) that could not be identified to species level were removed from subsequent analyses. However, 50 of these specimens were either

| Species | U | V (U) | SE (U) |
|---------|---|------|--------|
| Dsim    | −150–64 | 3.98 × 10^9 | 1996–19 |
| Zap1    | −0.46 | 2.6 | 0.51 |
| Dmel    | −0.02 | 8.77 × 10^-4 | 0.03 |
| Zap2    | −0.01 | 9.60 × 10^-4 | 0.01 |

Table 1. The test statistic U, the variance of U(V (U)) and its SE (SE (U)) used to determine if species distributions deviated from the negative binomial distribution. The values of U fall within the range given by the SE (U) and therefore the species distributions follow the negative binomial distribution (Bliss & Fisher 1953) (Dsim = Drosophila simulans; Zap1 = Zaprionus morphospecies group 1; Dmel = D. melanogaster; Zap2 = Zaprionus morphospecies group 2).

© 2003 British Ecological Society, Journal of Animal Ecology, 72, 468–477
Predicting abundance from occupancy

*D. simulans* or *D. melanogaster*, and are most likely to have been the former because this species was overwhelmingly dominant. The remaining 5 individuals were all *Zaprionus* species. The total measured abundance of each species was calculated by summing the abundance values across all 216 fruit for each of the six species. Species constituting < 5% of all individuals caught were considered rare. Occupancy (number of MMUs occupied on the table) was calculated for each species as the number of MMUs occupied on the table divided by the number of grid cells (0·04 m², 0·16 m², 0·36 m² for each MMU size of 4 × 4 fruit) and occupancy was therefore not calculated for sampling areas larger than 0·36 m².

The distributions of the species were tested against the negative binomial distribution (Bliss & Fisher 1953), to determine the suitability of using the HG model to predict the abundances of species in this assemblage. *Drosophila busckii* and *D. buzzatii* were not tested because these species occupied < 1% of the total number of fruit available for colonization (n = 216). The significance of the difference between the observed and expected second moments calculated from

\[ U = s^2 - (\bar{X} + \bar{X}^2/\hat{k}_2) \]  

is determined by comparison with its standard error (Bliss & Fisher 1953). Values of *U* falling within the range of the standard error indicate that the distribution of individuals fits the negative binomial distribution.

The HG model was then used to estimate the abundance of each species, across all samples, using the formula

\[ \hat{N} = AK\alpha[(1 - A\hat{J}\alpha)^{-1\alpha} - 1] \]

where \( A \) is the extent of the sampling area (for this study 2·4 m × 3·6 m = 8·64 m²), \( a \) is the size of the sampling area or MMU (0·04 m², 0·16 m², 0·36 m² for each scale used) and \( A_\alpha \) is the area of occupancy (MMU size × number of occupied MMU) (He & Gaston 2000a). Here, \( A_\alpha \) is the measured occupancy (free of sampling error) for each fly species because all flies in each fruit were counted and identified (including those of the rare species). Therefore, complete abundance and occupancy records were used that were not subject to sampling artefacts. Abundance (\( \hat{N} \)) and \( k \) (aggregation parameter) are solved simultaneously in equation 2, using the Newton-Raphson iteration method (Mathews 1987), for each of two MMUs (spatial scales) for each species (the HG method). MMU pairs that were used to estimate abundance and \( k \) for each species were (i) 0·04 m² and 0·16 m² (ii) 0·04 m² and 0·36 m², and (iii) 0·16 m² and 0·36 m². To determine if the estimated abundances for each species were similar for the three sets of MMU pairs, they were compared using a Chi-square test (Sokal & Rohlf 1998). Predicted abundances were compared to real abundances using a Chi-square test (Sokal & Rohlf 1998). The accuracy of the HG method was calculated as predicted abundance/real abundance. Values closest to unity indicate predicted abundances that are equivalent to real abundances and are therefore the most accurate. Values below and above 1·0 represent under- and over-estimation of abundance, respectively. Spearman’s rank correlation was used to determine if there was a significant relationship between method prediction accuracy and real species abundance (Sokal & Rohlf 1998). The presence of a significant relationship therefore demonstrates that rare and common species are not predicted with equal accuracy by the method in question.

To adjust \( k \) for species density, \( k \) was also calculated for each species using the following equation modified from Taylor *et al.* (1979) (see Shorrocks & Rosewell 1986):

\[ \frac{1}{k} = a'm^{b'-2} - m^{-1} \]

where \( m \) is the mean number of animals per fruit and \( a' \) and \( b' \) are constants (\( a' = 8·48; b' = 1·51 \) for drosophilids obtained from Shorrocks & Rosewell 1986) (see also Taylor *et al.* 1978). These \( k \)-values (assuming a common \( k \) across scales; Kunin 1998; He & Gaston 2000a) were then substituted into equation 2 to obtain a second estimate of abundance at each of the three MMUs (0·04 m², 0·16 m² and 0·36 m²). These abundance estimates were then also compared to the real abundance values using a Chi-square test, and Spearman’s rank correlation to determine if there was a significant relationship between prediction accuracy (predicted abundance/measured abundance) and species abundance for each MMU (Sokal & Rohlf 1998).

Finally, a third value of \( k \) was calculated; using the NBD (\( \hat{k}_{\text{NBD}} \)) with a maximum likelihood solution (Bliss & Fisher 1953) for each of the species with > 1% occupancy. This was then substituted into equation 2 to obtain a third abundance prediction for each species. These abundance estimates were again compared to the real abundance values using a Chi-square test (Sokal & Rohlf 1998).

To distinguish between the three \( k \)-values and abundance estimates obtained we use (a) \( k_{\text{HG}} \) and \( k_{\text{NBD}} \)-abundance for those derived using the HG parameterization method, (b) \( k_{\text{adj}} \) and \( k_{\text{ab}} \)-abundance for those obtained by substituting \( k \) adjusted (Taylor *et al.* 1979) into the HG model, and (c) \( k_{\text{NBD}} \) and \( k_{\text{NBD}} \)-abundance for those obtained by substituting \( k \) of the NBD (Bliss & Fisher 1953) into the HG model.

**Results**

None of the species abundance distributions differed significantly from a negative binomial distribution (Table 1), and the HG model was thus considered appropriate for estimating the abundances of species in
the drosophilid assemblage. Species occupying fewer than 1% of the fruit available for colonization, namely D. buzzatii and D. busckii, were not considered because their occupancies were too low to assess the form of their spatial distribution. Of the six species found, Drosophila simulans was the only common species, whereas the remainder were rare (each contributing < 2% to the total number of sampled individuals) (Table 2).

Estimates of abundance and $k$ did not converge to a constant root using the HG method for D. melanogaster, D. buzzatii and D. busckii for any combination of scales. Although a range of appropriate initial $k$-values were used in the iteration process (1; 0.01; 0.001; based on previously reported values for drosophilid species, see Discussion), the estimated $k$-value fluctuated by up to 14 orders of magnitude depending on the initial value of $k$ used. The inability of the iteration process to converge to a constant root for these species results from the slope of the derivative of the function being small, and its tangent almost horizontal to the curve (Mathews 1987). One common and two rare species were thus used to test the accuracy of the HG model and parameterization method.

| Table 2. Measured occupancy (number of fruit occupied), abundance ($N$; % relative abundance) and mean abundance per fruit ($\pm$ SD) for each species in the assemblage (Dsim = Drosophila simulans; Zap1 = Zaprionus morphospecies group 1; Dmel = D. melanogaster; Zap2 = Zaprionus morphospecies group 2; Dbuzz = D. buzzatii; Dbusck = D. busckii). Predicted abundance using (a) the original method of He & Gaston (2000a) to solve for abundance and $k_{HG}$ simultaneously, (b) predicted HG $k_{adj}$ abundance†, and (c) predicted HG $k_{NBD}$ abundance‡ for three minimum mapping units (MMUs). Dashes indicate either no convergence at a solution for abundance and $k$, or abundances were not predicted for these species (see text). Values in parentheses indicate results of $\chi^2$ analyses for the differences between real and predicted abundances in the following order: d.f.; $\chi^2$-value; significance level. ***$P < 0.001$

| Species | MMU pair | Measured occup. | $N$ (% | Mean $N$ ($\pm$ SD) | 0.04 m$^2$; 0.16 m$^2$ | 0.04 m$^2$; 0.36 m$^2$ | 0.16 m$^2$; 0.36 m$^2$
| --- | --- | --- | --- | --- | (2; 2531.86; *** | (1; 2542.75; *** | (1; 2586.52; ***
| Dsim | 102 | 2869 | 13.28 $\pm$ 3.35 | 184.04 | 177.87 | 153.36
| Zap1 | 21 | 56 | 0.28 $\pm$ 0.06 | 23.46 | 23.83 | 26.01
| Dmel | 17 | 10 | 0.05 $\pm$ 0.05 | 8.40 | – | –
| Zap2 | 8 | 9 | 0.04 $\pm$ 0.01 | – | – | –
| Dbuzz | 1 | 9 | 0.04 $\pm$ 0.01 | – | – | –
| Dbusck | 2 | 2 | 0.01 $\pm$ 0.01 | – | – | –

| (b) Predicted HG $k_{adj}$ abundance†

| MMU | 0.04 m$^2$ | 0.16 m$^2$ | 0.36 m$^2$
| --- | (5; 2287.52; *** | (5; 35147.98; *** | (5; 11665 $\times$ 10$^5$; ***
| Dsim | 102 | 2869 | 13.28 $\pm$ 3.35 | 316.96 | 902.07 | 3239.60
| Zap1 | 21 | 56 | 0.28 $\pm$ 0.06 | 44.99 | 352.47 | 4262.51
| Dmel | 17 | 10 | 0.05 $\pm$ 0.05 | 29.97 | 902.91 | 1673229.00
| Zap2 | 8 | 9 | 0.04 $\pm$ 0.01 | 8.40 | – | –
| Dbuzz | 1 | 9 | 0.04 $\pm$ 0.01 | 1.04 | 2.81 | 4.59
| Dbusck | 2 | 2 | 0.01 $\pm$ 0.01 | 1.80 | 0.81 | 0.64

| (c) Predicted HG $k_{NBD}$ abundance‡

| MMU | 0.04 m$^2$ | 0.16 m$^2$ | 0.36 m$^2$
| --- | (3; 0.38; ns | (3; 1469 $\times$ 10$^5$; *** | (3; 103113 $\times$ 10$^8$; ***
| Dsim | 102 | 2869 | 13.28 $\pm$ 3.35 | 2874.27 | 6.52 $\times$ 10$^5$ | 1.72 $\times$ 10$^7$
| Zap1 | 21 | 56 | 0.28 $\pm$ 0.06 | 60.54 | 1336.64 | 54256.75
| Dmel | 17 | 24 | 0.11 $\pm$ 0.04 | 23.90 | 82.42 | 1561.03
| Zap2 | 8 | 10 | 0.05 $\pm$ 0.02 | 9.96 | 16.67 | 79.12
| Dbuzz | 1 | 9 | 0.04 $\pm$ 0.01 | – | – | –
| Dbusck | 2 | 2 | 0.01 $\pm$ 0.01 | – | – | –

†$k_{adj}$ calculated from Taylor et al. (1979).
‡Estimate of $k$ of the negative binomial distribution (Bliss & Fisher 1953).
An abundance estimate could also only be obtained for the smallest combination of MMUs (0.04 m² and 0.16 m²) for *Zaprionus* species group 2 (Table 2a). The predicted *k*HG-abundances obtained for the three MMU pairs did not differ significantly from one another for either *D. simulans* or *Zaprionus* species group 1 (Sign tests for both species: $Z = -0.707; P > 0.05$). However, all predicted abundance estimates derived using the HG method differed significantly from the measured abundance values (Table 2a). The HG method consistently underestimated (all values are below one) the abundances of species at all MMU combinations (Fig. 2a). The greatest disparity occurred for the most abundant species (*D. simulans*), which was underestimated by approximately 94% for all MMU combinations (Table 2a). The accuracy of abundance estimates also declined significantly with an increase in real abundance ($r_s = -0.93; t(5) = -5.48; P < 0.01$).

Substituting the *k*_adj values into the equation for the HG model yielded abundance estimates for the species that were closer to reality than the *k*_HG-abundance estimates, though they remained significantly different.
from the measured abundance values for all species (Table 2b, Fig. 2b). In all cases, \( k_{adj} \) values were an order of magnitude smaller than \( k_{HG} \) (i.e., the species were more aggregated than the original method predicted) (Table 3). As expected, when \( k_{adj} \) was substituted into equation 2 the method accuracy was not sensitive to the abundance of the species. That is, the abundances of rare and common species were less accurately predicted (underestimated) than those of the rare species. When using \( k_{adj} \) in the model the abundance prediction improved. Although these predictions remained significantly different from measured abundance, rare and common species were predicted with equal accuracy.

Substituting \( k_{NBD} \) into equation 2 yielded highly accurate abundance predictions that were not significantly different from the real abundance values at the finest MMU (Table 2c; Fig. 2b). Predicted \( k_{NBD} \) abundances for the larger MMUs (0·16 m\(^2\) and 0·36 m\(^2\)) were, however, significant overestimates of abundances for all four species (Table 2c).

**Discussion**

The Drosophilidae assemblage structure found here was very similar to the structure that has been found for other dipteran assemblages associated with ephemeral resources (e.g., Atkinson & Shorrocks 1984; Sevenster & van Alphen 1993). The high variability in relative abundance between species (\( D. \) simulans constituted 94% of all emerged flies) and in occupancy of resources across species appears typical of such assemblages (Beaver 1977; Atkinson 1985; Shorrocks & Rosewell 1987).

The distribution patterns of individuals of each species did not deviate significantly from the NBD. Nonetheless, the HG method consistently underestimated the abundances of species in the assemblage. Estimated \( k_{HG} \) values were always larger than both \( k_{adj} \) and \( k_{NBD} \). All the species in the assemblage were thus more aggregated than the HG method estimated; \( k_{HG} \) lay between 0·58 and 1·49, while \( k_{NBD} \) and \( k_{adj} \) ranged between 0·058 and 0·432 with the lower range of the latter an order of magnitude smaller than \( k_{HG} \). In dipteran assemblages, \( k \) is usually less than one, and the \( k_{adj} \) and \( k_{NBD} \) values calculated here are closer to those reported in the literature than to those estimated as \( k_{HG} \) (Atkinson 1985; Shorrocks & Rosewell 1987; Rosewell et al. 1990; Shorrocks, Rosewell & Edwards 1990). The \( k_{adj} \) and \( k_{NBD} \) values therefore better reflected the highly aggregated nature of fly individuals (\( k << 1 \)), whereas \( k_{HG} \) was less sensitive to extreme aggregation. Substitution of \( k_{NBD} \) into the HG model thus improved estimated abundances and, at the smallest MMU, the measured abundances of the species were accurately predicted.

Although He & Gaston (2000a) and Kunin et al. (2000) also found that the HG method underestimates abundances, the magnitude of this underestimation was not as large as that found here. In the absence of sampling error, such underestimation is thought to arise in two ways: (1) deviation of the species spatial distribution pattern from the NBD, and (2) MMUs that are too large to detect sufficient variation in species abundance across the sampling extent (He & Gaston 2000a; Falster, Murray & Lepschi 2001). We demonstrated that the abundance distributions of the Drosophilidae used here did not deviate significantly from the NBD. Rosewell et al. (1990) also found that the NBD adequately described the distribution of *Drosophila* assemblages, and the NBD distribution is not unusual for invertebrate assemblages (Sevenster 1996). However, the distributions of the data we use must deviate to some extent from a perfect NBD. This deviation, although not significant, may be sufficient to result in substantial underestimation of species abundances. Any deviation from the NBD by species that are more strongly aggregated than this distribution predicts, thus results in significant underestimation of abundance values. The limiting values of \( k \) lead to other mathematical distributions, and as \( k \to 0 \) (strong aggregation) the distribution converges to the logarithmic series (Quenouille 1949). Therefore, although the maximum likelihood solution approach to calculating \( k \) of

---

**Table 3.** Estimated \( k_{HG} \) (HG method), \( k_{adj} \) (Taylor et al. 1979) and \( k_{NBD} \) (± SE) (Bliss & Fisher 1953) values for each of the identified species for each minimum mapping unit pair (\( D. \) simulans; \( Z. \) riparius morphospecies group 1; \( D. \) melanogaster; \( Z. \) riparius morphospecies group 2; \( D. \) buzzatii; \( D. \) busckii)

| Species         | Estimated \( k_{HG} \) | \( k_{adj} \) | \( k_{NBD} \) (± SE) |
|-----------------|------------------------|---------------|---------------------|
|                 | 0·04 m\(^2\); 0·16 m\(^2\) | 0·04 m\(^2\); 0·36 m\(^2\) | 0·16 m\(^2\); 0·36 m\(^2\) |
| \( D. \) simulans | 1·162; 1·311           | 1·494; 0·432  | 0·140 ± 0·211       |
| \( Z. \) riparius| 0·855; 0·683           | 0·585; 0·079  | 0·058 ± 0·015       |
| \( D. \) melanogaster | –; –              | –; 0·063           | 0·143 ± 0·076       |
| \( Z. \) riparius morphospecies group 2 | –; – | –; 0·060 | 0·097 ± 0·087 |
| \( D. \) buzzatii | –; 0·618             | –; 0·062           | –                   |
| \( D. \) busckii  | –; –                 | –; –                | –                   |
the NBD is definitive (Anscombe 1950; Bliss & Fisher 1953), the test of significance for the fit of data to a NBD is clearly not sufficiently sensitive to be used to identify assemblages with distributions for which the HG model would be appropriate. Distribution-fitting tests are known to have low power (May 1975; Buzas et al. 1982; McGeoch & Gaston 2002) and are therefore not an effective means of determining the appropriateness of the HG model. A statistic more sensitive to departures from the NBD as $k \to 0$ would be a useful tool for testing the suitability of assemblages for application of the HG model.

A second possible explanation for the underestimation of abundance is the use of minimum mapping units that are too large. In this study the finest possible minimum mapping unit with realistic habitat boundaries was used, i.e. the fruit. Predicted abundances were indeed closest to reality at this smallest MMU for all species. Nonetheless, aggregation still occurred beyond the resolution achievable with this finest MMU, resulting in an overestimation of $k$ by the HG method ($k_{HG}$). In most highly aggregated assemblages a MMU that is the size of the single resource unit may contain numerous individuals (Atkinson 1985; Shorrocks & Rosewell 1987). Consequently, resource units containing either one or 335 individuals (such as D. simulans in this study) will yield identical occurrence maps. A single fruit represents a MMU that is far ‘too large’ to achieve an accurate estimate of abundance, using the HG method, for assemblages that are highly aggregated. However, it is not only impractical but also unrealistic to use a MMU smaller than a single fruit for assemblages such as this. Therefore, although the use of coarse MMUs is known to reduce the accuracy of abundances predicted (He & Gaston 2000b), even the use of the finest, possible, and biologically realistic, MMUs may be inadequate to predict abundances accurately. There are many such assemblages where the resource unit is generally considered the smallest practicable MMU, e.g. those associated with tree holes, dung, carcasses, fruit and flower heads. Although habitat units are usually larger than the size of the individual utilizing them, these habitats are (i) a standardized unit for all the species in the assemblage (this is required in order to make cross-species comparisons, McGeoch & Gaston 2002) (ii) readily measured in the field and (iii) represent a habitat with a functional boundary unlike the artificial boundaries of grid cells, i.e. representation-based sample units, sensu Gaston (1994b). As a result, habitat units are likely to remain the MMU used for such assemblages.

The HG parameterization method and its application thus require further exploration at fine scales, often represented by the individual habitat units of species, and for species that are highly aggregated at this scale. Based on published $k$-values (estimated using the NBD) for a variety of taxa, strong levels of aggregation are apparently widespread; for example, in weevils (Peng & Brewer 1994), eight species of Drosophila (Shorrocks & Rosewell 1987; Rosewell et al. 1990), various other Diptera (Atkinson & Shorrocks 1984; Renshaw et al. 1995), eriophyid mites (Hall, Childers & Eger 1991) and a variety of solanaceous and other, annual, plant species (Timmer et al. 1989; Johnson et al. 1995). Although the $k$-values for these taxa ranged between 7·15 and < 0·005, $k$-values in the range of 0·01 to 0·6 were most common, demonstrating very high levels of aggregation. For species such as these, the only solution for abundance estimation unfortunately remains the use of a reasonably accurate estimate of aggregation ($k$) prior to application of the model, which requires an $a$ priori abundance estimate (e.g. the Nachman model, Nachman 1981; see also Holt et al. 2002a) and thus defeats the purpose of the He & Gaston (2000a) approach.

There are obvious advantages to being able to apply the HG model under conditions of strong aggregation at fine scales. For example, scouting for pest populations in fields, orchards and vineyards results in presence-absence (occupancy) data for crop units, such as trees, leaves or fruit (Wilson & Morton 1993; Overholt, Ogedah & Lammers 1994). Were it possible to derive accurate predictions of pest abundance from these data alone, more reliable damage estimates and economic injury levels could be calculated. Other than the insensitivity of the parameterization method to very high degrees of aggregation, the HG model appears highly suitable for such application. Pest abundance estimates are generally required for scales in the order of single or multiple fields or orchards, scales much finer than geographical mapping data readily available for many plant and vertebrate taxa. Mapping units and sampling extent are thus comparatively fine. Furthermore, predictions of abundance and $k$ could be derived using a short range of scales (small spatial scale differences between the two minimum mapping units used), which is another of the conditions for optimal performance of the HG method (He & Gaston 2000b; Kunin et al. 2000).

A possible future solution for improving abundance predictions for highly aggregated, abundant species may involve the incorporation of information on the relative spatial positions of occupancy measures, i.e. the degree to which nearby points have the same state (occupied or unoccupied) compared to distant points. Here, the distinction is made between aggregation in the form of statistical vs. spatial heterogeneity (Perry 1998). Statistical heterogeneity arises from kurtosis in the frequency distribution of records, and the degree to which this fits a particular distribution (such as the NBD), whereas spatial heterogeneity results from the clumping of records, or their deviation from a regular spatial arrangement (Perry 1998). Aggregation estimates based on $k$ of the NBD, including those estimated by the HG method, quantify statistical heterogeneity, but do not consider the spatial heterogeneity of abundance-occupancy records. However, for abundant species, deviation from spatial regularity is
likely to be large when the level of aggregation by individuals within habitat units (or MMUs) is high. Indeed, the abundances of rare species in this Drosophilidae assemblage were more accurately predicted by the HG method than the abundance of the common species (although the HG model performs equally well for rare and common species when provided with an accurate estimate of \( k \)). Therefore, for species known to be abundant (e.g. eruptive or outbreak species), estimates may be improved by incorporating this additional dimension of spatial information into abundance-occupancy models (see approaches by Perry 1998; Perry et al. 1999). Such spatially explicit exploration of the approach pioneered by He & Gaston (2000a) would clearly repay the effort because the latter provides a useful and much-needed means of predicting abundance from occupancy data, with considerable value in applied ecology.

Acknowledgements

We gratefully acknowledge Nordi Fruits for supplying the nectarines, I. Smit, M.J.W. Warren and R. Veldman for invaluable assistance and discussion, and F. He and an anonymous referee for helpful comments on the manuscript. This work was supported by NRF grant NRF4890 (MAM) and a Mellon Foundation Mentoring Scheme grant (MW).

References

Anscombe, F.J. (1950) Sampling theory of the negative binomial and logarithmic series distributions. *Biometrika*, **37**, 358–382.

Atkinson, W.D. (1985) Coexistence of Australian rainforest Diptera breeding in fallen fruit. *Journal of Animal Ecology*, **54**, 507–518.

Atkinson, W.D. & Shorrocks, B. (1984) Aggregation of larval Diptera over discrete and ephemeral breeding sites: the implications for coexistence. *American Naturalist*, **124**, 336–351.

Beaver, R.A. (1977) Non-equilibrium ‘island’ communities: Diptera breeding in dead snails. *Journal of Animal Ecology*, **46**, 783–798.

Bliss, C.I. & Fisher, R.A. (1953) Fitting the negative binomial distribution to biological data. *Biometrics*, **9**, 176–200.

Brown, J.H. (1984) On the relationship between abundance and distribution. *American Naturalist*, **124**, 255–279.

Buzas, M.A., Koch, C.F., Culver, S.J. & Solh, N.F. (1982) On the distribution of species occurrence. *Paleobiology*, **8**, 143–150.

Cowley, M.J.R., Thomas, C.D., Thomas, J.A. & Warren, M.S. (1999) Flight areas of British butterflies: assessing species status and decline. *Proceedings of the Royal Society of London B*, **266**, 1587–1592.

Falster, D.S., Murray, B.R. & Lepschi, B.J. (2001) Linking abundance, occupancy and spatial structure: an empirical test of a neutral model in an open-forest woody plant community in eastern Australia. *Journal of Biogeography*, **28**, 317–323.

Feder, M.E. & Krebs, R.A. (1998) Natural and genetic engineering of the heat-shock protein Hsp70 in *Drosophila melanogaster*: Consequences for thermotolerance. *American Zoologist*, **38**, 503–517.

Gaston, K.J. (1994a) *Rarity*. Chapman & Hall, London.

Gaston, K.J. (1994b) Measuring geographic range sizes. *Ecography*, **17**, 198–205.

Gaston, K.J. (1999) Implications of interspecific and intraspecific abundance-occupancy relationships. *Oikos*, **86**, 195–207.

Gaston, K.J., Blackburn, T.M., Greenwood, J.D., Gregory, R.D., Quinn, R.M. & Lawton, J.H. (2000) Abundance-occupancy relationships. *Journal of Applied Ecology*, **37**, 39–59.

Harte, J., Blackburn, T.M. & Ostling, A. (2001) Self-similarity and the relationship between abundance and range size. *American Naturalist*, **157**, 374–386.

Harte, J., Kinsig, A. & Green, J. (1999) Self-similarity in the distribution and abundance of species. *Science*, **284**, 334–336.

He, F. & Gaston, K.J. (2000a) Estimating species abundance from occurrence. *American Naturalist*, **156**, 553–559.

He, F. & Gaston, K.J. (2000b) Occupancy-abundance relationships and sampling scales. *Ecography*, **23**, 503–511.

He, F., Gaston, K.J. & Wu, J. (2002) On species occupancy-abundance models. *Ecocience*, **9**, 119–126.

Holt, A.R., Gaston, K.J. & He, F.H. (2002a) Occupancy-abundance relationships and spatial distribution: a review. *Basic and Applied Ecology*, **3**, 1–13.

Holt, A.R., Warren, P.H. & Gaston, K.J. (2002b) The importance of biotic interactions in abundance-occupancy relationships. *Journal of Animal Ecology*, **71**, 846–854.

Johnson, G.A., Mortensen, D.A., Young, L.J. & Martin, A.R. (1995) The stability of weed seedling population models and parameters in eastern Nebraska corn (*Zea mays*) and soybean (*Glycine max*) fields. *Weed Science*, **43**, 604–611.

Kunin, W.E. (1998) Extrapolating species abundance across spatial scales. *Science*, **281**, 1513–1515.

Kunin, W.E., Hartley, S. & Lennon, J.J. (2000) Scaling down: on the challenge of estimating abundance from occurrence patterns. *American Naturalist*, **156**, 560–566.

Legg, D.E., Hein, G.L., Nowierski, R.M., Feng, M.G., Peaieres, F.B., Karner, M. & Cupers, G.W. (1992) Binomial regression models for spring and summer infestations of the Russian wheat aphid (*Homoptera: Aphididae*) in the Southern and Western plains states and Rocky Mountain region of the United States. *Journal of Economic Entomology*, **85**, 1779–1790.

Mathews, J.H. (1987) *Numerical Methods for Computer Science, Engineering, and Mathematics*. Prentice Hall International, Englewood Cliffs, N.J.

May, R.M. (1975) Patterns of species abundance and diversity. *Ecology and Evolution of Communities* (eds M.L. Cody & J.M. Diamond), pp. 81–120. Harvard University Press, Cambridge, Massachusetts.

McEvey, S., Potts, A., Rogers, G. & Wallis, S.J. (1988) A key to *Drosophilidae* (*Insecta: Diptera*) collected in areas of human settlement in southern Africa. *Journal of the Entomological Society of Southern Africa*, **51**, 171–182.

McGeoch, M.A. & Gaston, K.J. (2002) Occupancy frequency distributions: patterns, artefacts and mechanisms. *Biological Reviews*, **77**, 311–331.

Nachman, G. (1981) A mathematical model of the functional relationship between density and spatial distribution of a population. *Journal of Animal Ecology*, **50**, 453–460.
Overholt, W.A., Ogedah, K. & Lammers, P.M. (1994) Distribution and sampling of Chilo partellus (Lepidoptera: Pyralidae) in maize and sorghum on the Kenya coast. Bulletin of Entomological Research, 84, 367–378.

Peng, C. & Brewer, G.I. (1994) Spatial distribution of the red sunflower seed weevil (Coleoptera: Curculionidae) on sunflower. Environmental Entomology, 23, 1101–1105.

Perry, J.N. (1995) Spatial aspects of animal and plant distributions in patchy farmland habitats. Ecology and Integrated Farming Systems (eds D.M. Glen, M.P. Greaves & H.M. Anderson), pp. 221–242. Wiley, Chichester.

Perry, J.N. (1998) Measures of spatial pattern for counts. Ecology, 79, 1008–1017.

Perry, J.N., Winder, L., Holland, J.M. & Alston, R.D. (1999) Red-blue plots for detecting clusters in count data. Ecology Letters, 2, 106–113.

Pielou, E.C. (1977) Mathematical Ecology. Wiley, New York.

Quenouille, M.H. (1949) A relation between the logarithmic, Poisson, and negative binomial series. Biometrics, 5, 162–164.

Relyea, R.A. & Yurewicz, K.L. (2002) Predicting community outcomes from pairwise interactions: integrating density- and trait-mediated effects. Oecologia, 131, 569–579.

Renshaw, M., Silver, J.B., Service, M.W. & Birley, M.H. (1995) Spatial dispersion patterns of larval Aedes cantans (Diptera: Culicidae) in temporary woodland pools. Bulletin of Entomological Research, 85, 125–133.

Rosewell, J., Shorrock, B. & Edwards, K. (1990) Competition on a divided and ephemeral resource: testing the assumptions I. Aggregation. Journal of Animal Ecology, 59, 977–1001.

Sevenster, J.G. (1996) Aggregation and coexistence. I. Theory and analysis. Journal of Animal Ecology, 65, 297–307.

Sevenster, J.G. & Van Alphen, J.M. (1993) A life history trade-off in Drosophila species and community structure in variable environments. Journal of Animal Ecology, 62, 720–736.

Shorrock, B. & Rosewell, J. (1986) Guild size in Drosophilids: a simulation model. Journal of Animal Ecology, 55, 527–541.

Shorrock, B. & Rosewell, J. (1987) Spatial patchiness and community structure: coexistence and guild size of Drosophilids on ephemeral resources. Organization of Communities Past and Present (eds J.H.R. Gee & P.S. Giller), pp. 29–51. Oxford: Blackwell Scientific.

Shorrock, B., Rosewell, J. & Edwards, K. (1990) Competition on a divided and ephemeral resource: testing the assumptions II. Association. Journal of Animal Ecology, 59, 1003–1017.

Sokal, R.R. & Rohlf, F.J. (1998) Biometry: the Principles and Practice of Statistics in Biology Research, 3rd edn. W.H. Freeman & Co, New York.

Taylor, L.R., Woiwood, I.P. & Perry, J.N. (1978) The density dependence of spatial behaviour and the rarity of randomness. Journal of Animal Ecology, 47, 383–406.

Taylor, L.R., Woiwood, I.P. & Perry, J.N. (1979) The negative binomial as a dynamic ecological model for aggregation, and the density dependence of k. Journal of Animal Ecology, 48, 289–304.

Thomas, C.D. & Abery, J.C.G. (1995) Estimating rates of butterfly decline from distribution maps: the effect of scale. Biological Conservation, 73, 59–65.

Timmer, L.W., Zitko, S.E., Sandler, H.A. & Graham, J.H. (1989) Seasonal and spatial analysis of populations of Phytophthora parasitica. Citrus orchards in Florida. Plant Disease, 73, 810–813.

Warren, P.H. & Gaston, K.J. (1997) Interspecific abundance-occupancy relationships: a test of mechanisms using microcosms. Journal of Animal Ecology, 66, 730–742.

Wilson, L.J. & Morton, R. (1993) Seasonal abundance and distribution of Tetranychus urticae (Acari: Tetranychidae), the two spotted spider mite, on cotton in Australia and implications for management. Bulletin of Entomological Research, 83, 291–303.

Received 9 August 2002; accepted 5 February 2003