Methods in Ecology and Evolution

British Ecological Society

Methods in Ecology and Evolution 2016, **7,** 1316–1324 doi: 10.1111/2041-210X.12605

Evaluation of statistical methods and sampling designs for the assessment of microhabitat selection based on point data

Irene L. Gorosito[1,2]*, Mariano Marziali Bermúdez[3,4], Richard J. Douglass[5] and María Busch[1,2]

[1]Departamento de Ecología, Genética y Evolución, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina; [2]Instituto de Ecología, Genética y Evolución de Buenos Aires, Consejo Nacional de Investigaciones Científicas y Técnicas, Intendente Güiraldes 2160 - Ciudad Universitaria, C1428EGA Buenos Aires, Argentina; [3]Departamento de Física, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina; [4]Instituto de Física de Buenos Aires, Consejo Nacional de Investigaciones Científicas y Técnicas, Intendente Güiraldes 2160 - Ciudad Universitaria, C1428EGA Buenos Aires, Argentina; and [5]Montana Tech of the University of Montana, Butte, MT 59701, USA

Summary

1. Information on resource selection by a species is essential for understanding the species' ecology, distribution and requirements for survival. Research on habitat selection frequently relies on animal detection at point locations to determine which resource units are used. A variety of approaches and statistical tools can be employed for assessing selection based on habitat variables measured in those units. The aim of this work was to evaluate the reliability of common sampling designs and statistical methods in detecting habitat selection at fine scales based on point data

2. We reviewed literature on microhabitat selection to determine characteristics of typical studies and analysed simulated small-mammal live-trapping data as a case study. We considered various scenarios differing in the number of sampled units and sampling duration. For each scenario, a set of simulated surveys was analysed through two univariate tests (Welch's *t*- and Mann–Whitney *U*-test), generalized linear models (GLMs), mixed-effect models (GLMMs) and occupancy models (OMs).

3. Analysis of simulated data revealed that overall performance of all statistical methods improved with increased trapping effort. Univariate tests were especially sensitive to the number of sampling units, while modelling methods took also advantage of longer trapping sessions. Univariate tests and GLMs provided partially correct information in most cases, whereas GLMMs and OMs offered higher probabilities of fully describing simulated habitat preferences.

4. With typical sampling efforts, appropriate statistical analysis of point data is able to provide a moderately accurate description of habitat selection at small scales, in spite of the violation of closure and independence assumptions of applied models. Modelling approaches are proliferating; we encourage using models that can deal with multiple sources of variability, such as GLMMs and OMs, when data are hierarchically structured. There is no *a priori* best survey design; it should be chosen according to the scope and goals of the study, environment heterogeneity, species characteristics and practical constraints. Researchers should realize that sampling design and statistical methods likely affect conclusions regarding habitat selection.

Key-words: generalized linear model, live trapping, mixed-effect model, occupancy model, trapping effort, univariate test

Introduction

Knowing how species use different resources is essential to understanding patterns of abundance and distribution. In particular, management of wildlife populations generally entails habitat management and it presupposes some understanding of species needs (Garshelis 2000; Manly *et al.* 2002). Researchers commonly infer species needs based on studies of habitat selection, which is determined on the basis of the ratio between use and availability (Manly *et al.* 2002). Therefore, how both use and availability are determined is crucial, especially for species that are not readily observed (Garshelis 2000).

A broad class of techniques employed in habitat selection studies is based on point data of animal presence or abundance: individuals are captured, observed or otherwise detected at fixed locations in different resource units, previously defined within the study area. Resource units are characterized by habitat variables related to food or shelter

---

*Correspondence author. E-mail: iree26@gmail.com availability, measured at each location. Statistical analyses are then performed in order to find significant relationships between environmental variables and animal presence or abundance.

The description of the pattern of habitat use of a species depends on sampling design and statistical methods employed. Key features of the sampling design are as follows:

1. Spatial scale and dimensionality should be chosen according to species characteristics (e.g. home range size, arboreality), resource distribution (heterogeneity and aggregation), as well as the scale of selection studied (ranging from microhabitat to geographical distribution).

2. The temporal scale of the study affects detection probabilities and the range of movements during the study. Whether a unit is classified as used/not used may depend on sampling length (Boyce *et al.* 2002). Patterns of use may change seasonally and differ among sexes, age classes, etc., the proportions of which could change temporally (Bilenca & Kravetz 1998).

3. Sampling effort is commonly reported as an important factor in most point data-based studies. However, the distribution of sampling effort between replicates and survey duration is highly variable throughout the literature and may also affect results (Jorgensen 2004).

4. Detection methods may convey different information. For example, radiotracking or spool-and-line tracking, tracks and live trapping may provide different conclusions about microhabitat selection (Douglass 1989; Prevedello, Garcia Rodrigues & Leite de Araujo Monteiro-Filho 2010).

A common issue in resource-use studies is that, while use can generally be detected, it is difficult to demonstrate non-use (Boyce *et al.* 2002). An increasing amount of literature describes the effect of imperfect detection on conclusions concerning resource use, especially when detection varies among resource units (MacKenzie 2006). Various approaches and statistical tools can be employed for assessing selection based on habitat variables: from comparing values of each variable between used and non-used/available units by univariate tests to modelling resource selection functions (Manly *et al.* 2002) using regressions or generalized linear models (GLMs). Emerging occupancy models (OMs) incorporate imperfect detection, preventing otherwise biased conclusions (MacKenzie *et al.* 2006). In particular, *N*-mixture OMs (Royle & Nichols 2003) additionally account for intensity of use instead of simply use/non-use.

Advantages of detailed models over simpler ones have been claimed by many authors, but such advantages cannot generally be taken for granted (Banks-Leite *et al.* 2014; Ives 2015). Moreover, studies dealing with microhabitat characteristics (i.e. factors that change within the home range of a single individual [Morris 1987]) require multiple sampling units within an area similar to the target species' typical home range size. As each individual is likely to be found at alternate sampling units during the observation period, sampling units are not closed (animals can enter or exit each unit) and spatially correlated (the same animal can be detected in neighbouring units). This sampling design explicitly violates unit-level closure and independence assumptions of many resource-use models. Is the violation of these assumptions a serious concern for the reliability of the results provided by various models of fine-scale habitat selection assessment?

Computer simulations have been employed to assess the performance of field and statistical methods used to study various ecological features (e.g. Alldredge & Ratti 1986; Efford 2004; Heithaus *et al.* 2006). Simulations are based on simplified models not reproducing all aspects of field surveys, but allow comparing posterior analyses of simulated data sets with known habitat preferences. The aim of our work was to evaluate the reliability of common sampling designs and statistical methods in detection of habitat selection at fine scales based on point data. For this evaluation, we simulated small-mammal live trapping as a case study and reviewed literature on microhabitat selection to discuss requirements and limitations of typical studies.

Materials and methods

LITERATURE SURVEY

We conducted a systematic search of literature describing studies on microhabitat selection by small mammals which relied primarily on point data obtained through trapping methods (live, snap and/or pitfall). We looked for articles containing *microhabitat*, *habitat use*, *habitat selection*, *small mammal*, *live trapping*, *resource use* or *resource selection* in either the abstract, title or keywords, using search engines in 38 journals. We discarded search results of studies not fitting in the adopted definition of microhabitat or based on other techniques than trapping.

For each article, we extracted sampling design characteristics and analysis methods applied. In some cases, part of the information was absent or not clearly stated and had to be deduced or estimated from available data. Further details on the survey are in Appendix S1, Supporting Information.

SIMULATED DATA

We simulated live-trapping surveys of small mammals in longitudinal habitats through a simple model specially developed for emulating movements and captures of small mammals on fine spatial and temporal scales. Ecologically important longitudinally shaped habitats are found along field edges in agroecosystems, along river margins and along railway embankments (Bennett 1990; Ellis *et al.* 1997; Ylönen *et al.* 2002). In these habitats, most movements are basically constrained to a single dimension and thus are easier to handle computationally than in two- and three-dimensional spaces. We based our simulations on available data for the Pampean grassland mouse *Akodon azarae* Fischer 1829, which inhabits mainly field borders (Bilenca & Kravetz 1998; Hodara *et al.* 2001; Gómez *et al.* 2011).

Simulations were run in GNU/OCTAVE 3.0 (Octave development community 2012) using *ad hoc* code for a modified version of an algorithm given by Efford (2004). Our simulation approach was more complex than assuming a stationary distribution, but accounted for drastic small-scale and short-time fluctuations that may be important in field surveys. A brief description of the methods used to simulate movements and captures of small mammals is provided below. Details are in Appendix S2.

We modelled hypothetical longitudinal habitats as arrays of 75 10-by-10-m cells representing adjacent resource units, with traps placed only in 25 central adjacent cells to avoid edge effects (Fig. 1). This

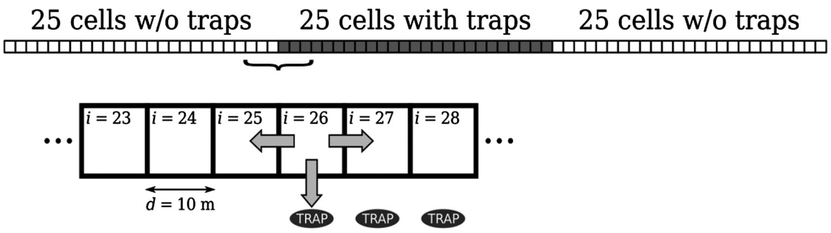

Fig. 1. (Top) Representation of the linear habitat by an array of 75 cells, 25 of which have traps. (Bottom) Detail of the section marked with a brace in top figure. Arrows indicate the three possible transitions associated with the 26th cell.

design sets ~5 traps within the extent of individual home ranges with each transect covering a broad range of microhabitats. The quality of each resource unit (cell) as perceived by animals was determined by a linear combination of three continuous variables quantifying local microhabitat features (such as *proportion of bare soil* or *mean vegetation height*). We named these variables *hp* (high preference, cells having high values of this variable are strongly preferred), *sp* (slight preference, cells having high values of this variable are slightly preferred) and *ma* (moderate avoidance, cells having high values of this variable are moderately avoided). We added a fourth variable *ci* (completely ignored) which had no effect on the preference of a cell to test for type I errors.

We first assigned random values to habitat variables and then introduced spatial correlation at two levels. For each longitudinal habitat, one of the three relevant variables (either *hp*, *sp* or *ma*) was randomly increased in every cell to represent variability between transects. Then, values obtained for each cell were averaged with both neighbouring cells to prevent sharp differences between adjacent cells, simulating small-scale spatial correlation.

We simulated a total of 4500 transects, each having mean animal densities between 0·2 and 3·2 rodents/cell. These densities are in agreement with estimates for *A. azarae*, based on home range size and overlap data (Priotto & Steinmann 1999) and an ongoing mark–recapture study (Gorosito, in preparation). Density dependence was introduced in the model as a penalty term (equal to minus the number of individuals in excess of the first individual in the cell) to perceived habitat quality, representing intraspecific competition. The inclusion of this term imitates setting local carrying capacities proportional to cell qualities.

Trapping simulations were run for 7 days. Captures at each cell were recorded for each day and animals were released at the cell of capture at the beginning of the next day. Movement and capture of animals were considered to be stochastic processes. At any time during the simulation, animals could remain in the current cell, move to a neighbour cell to the right/left or enter a trap (if there was one active in the cell). Movement rates were based on the penalized habitat quality so that animals stayed longer in higher-quality and less populated cells. The expected mean squared displacement of each individual was in the order of the home range size of small mammals (Mendel & Vieira 2003; Abramson *et al.* 2006). We adopted a capture rate per individual which led to an average number of captures per night at each 25-trap transect ranging from 0 to 4·57, in agreement with values reported in the literature for field studies (Hodara *et al.* 2001; Gomez *et al.* 2011) using live traps baited with rolled oats and peanut butter. For simplicity, we did not consider social behaviour and assumed uniform animals' tendency to enter traps, which may not apply in general (Jorgensen 2002; MacKenzie *et al.* 2006).

ANALYSIS

We applied commonly used statistical tools to our simulated data to find associations between capture locations and environmental variables, for which animals' habitat preferences assigned by us in the virtual system were known. In order to test the effect of the number of transects surveyed, we chose random subsets comprising various quantities of transects (N_T) from the 4500-replicate pool so that each of these subsets would correspond to a single microhabitat study with a certain number of replicates. We refer to each of these subsets as *a survey*. A total of 500 surveys were analysed for each N_T, with the number of transects ranging from 5 to 100 at intervals of 5 from 5 to 60 and at intervals of 10 from 60 to 100 (i.e. $N_T = 5, 10, 15,\ldots, 55, 60, 70,\ldots, 100$). Likewise, to test the effect of survey length, we evaluated capture data up to the third and the seventh day separately. In consequence, total trapping effort ranged between 375 and 7500 trap nights (tn) using three-day data, and between 875 and 17 500 tn using seven-day data. These figures are consistent with real field studies listed in our literature sample (Appendix S1).

Capture data from each survey were analysed by means of three types of statistical methods as described below. Characteristics of these methods and examples of use are provided in Appendix S3. All statistical analyses were conducted using the software R version 3.0.1 (R Core Team 2013).

Univariate tests

All sampled cells were classified in two groups: those with and without any capture. Distributions of each resource variable were compared between groups by means of two frequently used tests: Welch's unequal variances *t*-test (Ruxton 2006) and Mann–Whitney *U*-test (Zar 2010). We considered that an association between habitat variables and animal presence was found by the *t*-test (*U*-test) when the null hypothesis of equal means (medians) for the corresponding variable was rejected, that is $P < 0.05$. The sign of the difference between means (medians) was used to determine whether the variable was positively or negatively associated with animal presence.

Resource selection functions via generalized linear models (GLM/GLMM)

Capture frequencies per trap were related to environmental variables by means of a binomial logit-linked GLM and a binomial logit-linked mixed-effect GLMM (Zuur *et al.* 2009). In the latter, a random explanatory term was included to account for among-transect variability (e. g., sites placed closer or further from a water body, with northern or southern slopes, etc.) modelled in our simulations as varying average population densities and environmental variable profiles. GLMMs were evaluated using the LME4 package (Bolker 2013). For both GLMs and GLMMs, we followed two methods to estimate the coefficients that relate environmental variables with capture success. First, we performed a stepwise elimination of non-significant terms after Zuur *et al.* (2009), starting with a linear combination of the four environmental variables. At each step, the likelihood ratio method was used to test the significance of removing each variable from the model; the environmental variable for which its removal led to the highest *P*-value was discarded. The elimination process was repeated until the removal of any remaining variable led to a *P*-value <0·05. Remaining variables in the final model were considered associated with animal presence. The second method consisted of ranking all candidate models, that is the full model and all its nested models, based on Akaike information criterion (AIC), as described by Burnham & Anderson (2002), using the MUMIN package (Bartoń 2013). Models with ΔAIC < 4 with respect to the best (lowest AIC) model were considered to have enough support and were averaged. An environmental variable was considered associated with animal presence when scoring 0·7 or higher relative importance (i.e. the sum of Akaike weights of all models containing the corresponding regression coefficient). For both methods, the sign of regression coefficients indicated positive or negative associations.

Resource selection functions via OMs

We selected an N-mixture model among other OMs because it connects local abundance of unmarked animals (intensity of use) to detection probability at the corresponding unit (Royle & Nichols 2003), accounting for heterogeneous densities. We modelled the abundance at the j-th cell of the i-th transect as a Poisson-distributed random variable A_{ij}, with mean λ_{ij} log-linked to a linear combination of the environmental variables at that cell plus a normally distributed random effect $\tilde{a}_i$, with zero mean and variance $1/\tau$, that represented transect-to-transect variability. The resulting likelihood function was

$$L(a_k, r, \tau | y_{ij}) = \prod_i \int_{-\infty}^{+\infty} \left\{ \prod_j \frac{Y!}{y_{ij}!(Y - y_{ij})!} \right. \times \sum_{A_{ij}=0}^{\infty} p_{ij}^{y_{ij}} (1 - p_{ij})^{Y - y_{ij}} \frac{\lambda_{ij}^{A_{ij}} e^{-\lambda_{ij}}}{A_{ij}!} \left. \right\} \sqrt{\frac{\tau}{2\pi} e^{-\frac{\tilde{a}_i^2 \tau}{2}}} d\tilde{a}_i, \qquad \text{eqn 1}$$

where y_{ij} was the total number of captures at the j-th cell of the i-th transect during Y trapping occasions (nights) and

$$p_{ij} = p(A_{ij}, r) = 1 - (1 - r)^{A_{ij}} \qquad \text{eqn 2}$$

was the success probability at the corresponding trap, assuming homogeneous individual detection probability r. The integral over $\tilde{a}_i$ and the infinite sum over A_{ij} constitute mixture distributions for the mean abundance λ_{ij} and the probability of success p_{ij}, respectively. Truncating the infinite sum up to an arbitrary bound may lead to unrealistic estimations (Dénes, Silveira & Beissinger 2015). We avoided this issue by evaluating the sum analytically, which results in a finite sum

$$\sum_{A_{ij}=0}^{\infty} p_{ij}^{y_{ij}} (1 - p_{ij})^{Y - y_{ij}} \frac{\lambda_{ij}^{A_{ij}} e^{-\lambda_{ij}}}{A_{ij}!} = \times \sum_{x_{ij}=0}^{y_{ij}} \frac{(-1)^{x_{ij}} y_{ij}!}{x_{ij}!(x_{ij} - y_{ij})!} \exp\{-\lambda_{ij}[1 - (1 - r)^{Y - y_{ij} + x_{ij}}]\}. \qquad \text{eqn 3}$$

Parameters were estimated using Markov chain Monte Carlo method, as implemented in WINBUGS 1.4.3 (Lunn *et al.* 2000), with the aid of the package R2WINBUGS 2.1-19 (Sturtz, Ligges & Gelman 2005). To avoid numerically integrating on $\tilde{a}_i$, we modelled the random effects as stochastic nodes with their normal distribution as prior. For all other parameters, non-informative priors were assumed. Our non-standard likelihood function was introduced into the model using the *zeros trick* (WINBUGS help). Due to the higher computational cost of this method, 500 surveys were analysed using three-day data and 200 surveys were analysed using seven-day data, with $N_T = 25$ and 50 only.

We classified the results for each survey and by each statistical method as *perfect* when all relevant variables (*highly preferred*, *slightly preferred* and *moderately avoided*) were identified as significant and with the correct sign (no errors), *good* when at least one relevant but not the *completely ignored* variable was identified as significant and with the correct sign (type II error on some but not all relevant variables, but no type I error), *null* when no significant relationships were found (type II error on all relevant variables, but no type I error) and *bad* when either the *completely ignored* variable was identified as significant (type I error) or relevant variables were identified as significant but were given wrong signs. Then, we tallied the frequencies of perfect, good, null and bad results for each sampling design (number of transects and number of days) and statistical analysis (t-test, U-test, GLM, GLMM and OM).

Results

LITERATURE SURVEY SUMMARY

We reviewed 152 articles on microhabitat selection based on live trapping (see Appendix S1). Total trapping effort varied in a wide range, with median 7740 ($Q_{25\%} = 3496$, $Q_{75\%} = 17\,660$), and showed no trend along years. Most studies collected data during 3 (= mode) consecutive nights, with 47·6% of the studies trapping for at least five consecutive nights per session. Typical designs involved 2–12 ($Q_{25\%}$–$Q_{75\%}$) trapping sessions and 2–12 ($Q_{25\%}$–$Q_{75\%}$) grids/transects.

Overall, in 35·5% of the cases, selection was assessed only by means of hypothesis tests (e.g. t-tests, analyses of variance), in 20·4% by ordination methods (e.g. discriminant analysis, canonical correspondence analysis) and in 23·7% by fitting regressions and generalized linear models (GLMs). In 13·2% of the cases, combinations of more than one of these methods were used. Methods based on detailed modelling, such as mixed-effect models (Zuur *et al.* 2009) and OMs (MacKenzie *et al.* 2006), were used in only eight articles in our survey. However, modelling approaches are steadily adopted: in 2000–2004, 41% of papers applied GLMs; in 2010–2015, 63% applied GLMs, GLMMs or OMs.

ANALYSIS OF SIMULATED DATA

The proportion of *perfect*, *good*, *null* or *bad* results varied depending on both the statistical method and the sampling design. A comparison among all statistical methods applied for $N_T = 50$ and seven-day data (8750 tn) is shown as an example (Fig. 2). Generally, both univariate tests (t-test and U-test) performed similarly and yielded *perfect* results less frequently than modelling methods. Among modelling methods, GLMMs and OMs yielded a higher proportion of *perfect* results than GLMs. *Bad* results were more frequent for GLMs than for univariate and other modelling methods. There were no practical differences in the outcomes of GLMs or GLMMs when model selection was based either on a stepwise regression (SW) or on model averaging (MA) using Akaike's information criterion weights.

SAMPLING DESIGN

Due to the strong similarities found throughout all scenarios between the performance of t- and U-tests, and between stepwise and model-averaging approaches, we next examined the effect of sampling design on the outcome of t-tests and model-averaged GLMs and GLMMs only. Note that the same

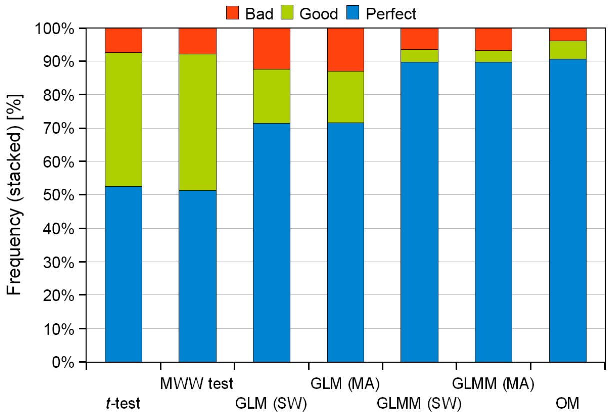

Fig. 2. Frequencies of perfect, good and bad results obtained using different statistical methods for $N_T = 50$ and seven-night data (8750 tn): *t*-test, *U*-test, generalized linear models (GLMs) and mixed-effects GLMs (GLMMs) using stepwise regressions (SWs) and model-averaging (MA), and occupancy models (OMs).

conclusions apply for their corresponding counterparts (i.e. *U*-tests and stepwise GLMs and GLMMs, respectively).

Frequencies of *perfect* results obtained by *t*-tests (Fig. 3a) increased with trapping effort from less than 2% to about 74% in the range from 375 to 17 500 tn. In addition, for a given trapping effort, *perfect* results were slightly more frequent with three-day data, indicating that increasing the number of transects yields better results than increasing trapping duration when data are analysed through univariate tests. *Null* results comprised 39% of the total at the lowest trapping effort, but decreased and eventually vanished at 4000 tn. *Bad* results appeared between 5% and 11% of the surveys and showed no apparent trend with respect to trapping effort. However, seven-day data seemed more prone to errors than three-day data.

Outcomes of model-averaged GLMs (Fig. 3b) were similar to those of *t*-tests for the lowest trapping efforts, but improved faster reaching about 87% *perfect* results at 17 500 tn. Seven-day data gave *perfect* results somewhat more often than three-day data for trapping efforts only below 3000 tn. The rate of *null* results, starting from 37% at 375 tn, was slightly lower than for the *t*-test. For a given trapping effort, three-day data from more transects yielded a lower frequency of *bad* results (5–10% of the surveys) than seven-day data from fewer transects (up to 14% of the surveys for some trapping efforts.)

Model-averaged GLMMs (Fig. 3c) had the best performance, with up to 95% *perfect* results at the highest trapping efforts and less than 8% of *bad* results for all trapping efforts. Approximately 31% of the surveys gave *null* results at the lowest trapping effort, and no *null* results appeared above 1750 tn. Seven-day data produced slightly better outcomes than three-day data, with a higher rate of *perfect* results and lower *null* results which mostly occurred at the lowest trapping efforts.

METHOD–VARIABLE SPECIFICITY

Frequencies of correct associations for each variable obtained by different methods were plotted as functions of trapping effort using three- and seven-day data. For the sake of clarity, *U*-tests and stepwise regressions were excluded as before. However, we included OMs in these comparisons to demonstrate similarities with GLMMs.

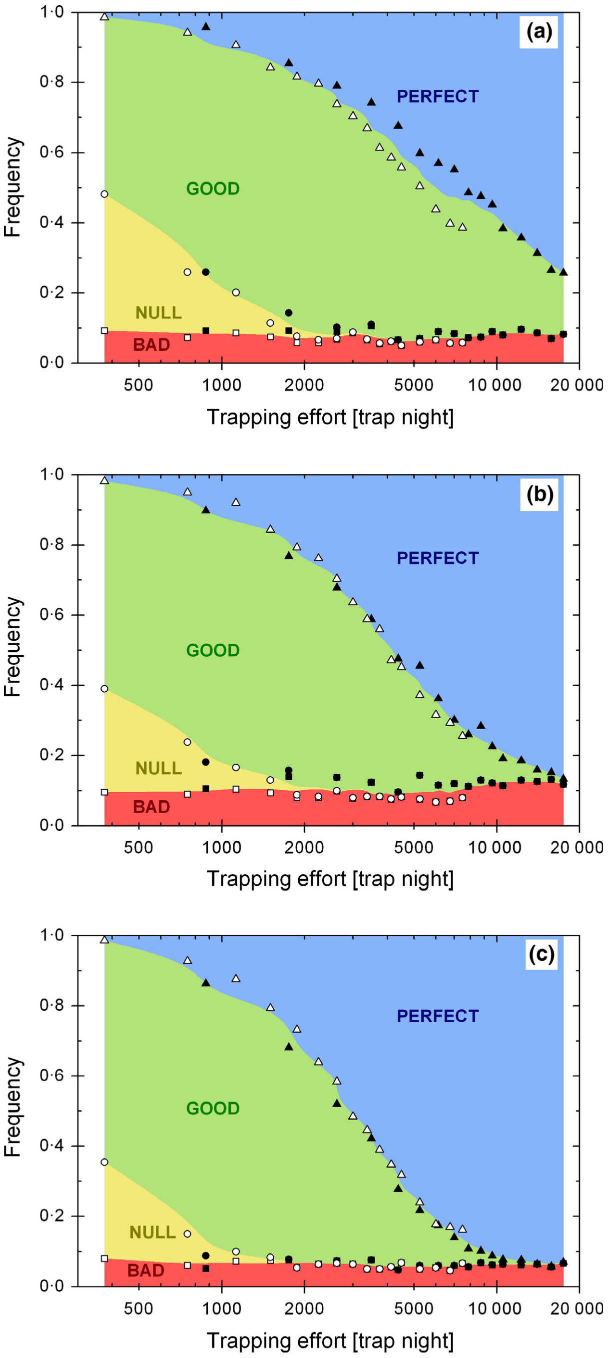

Fig. 3. Stacked frequencies of perfect, good, null and bad results obtained using *t*-test (a), GLMs with model averaging (b) and GLMMs with model averaging (c), for different trapping efforts. White symbols correspond to three-day data and black symbols correspond to seven-day data. Colour areas are visual aids.

The *highly preferred* variable (*hp*) was correctly associated with animal presence for all data sets with trapping efforts above 1800 tn by GLMMs and OMs, and above 4000 tn by *t*-tests and GLMs (Fig. 4a). Frequencies of correct

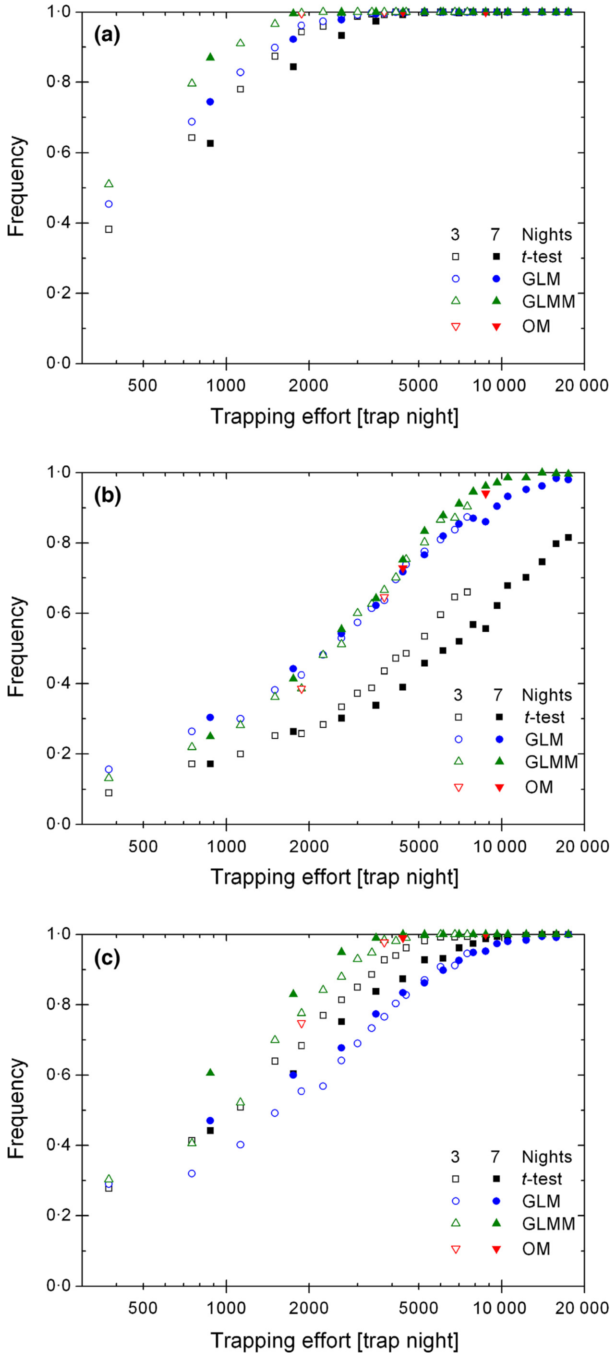

Fig. 4. Frequencies of correct significant associations between rodent presence and environmental variables: *hp* = highly preferred (a), *sp* = slightly preferred (b) and *ma* = moderately avoided (c), using different statistical methods and trapping efforts.

associations decreased at lower trapping efforts, with GLMMs always performing better than GLMs and *t*-tests. In particular, for a given trapping effort, *t*-tests were more effective with three-day data from more transects than with seven-day data from less transects.

Frequencies of correct associations for the *slightly preferred* variable (*sp*) were lower than for *hp* and increased faster when using GLMs, GLMMs or OMs than *t*-tests (Fig. 4b). No differences between GLMMs and OMs were apparent, and

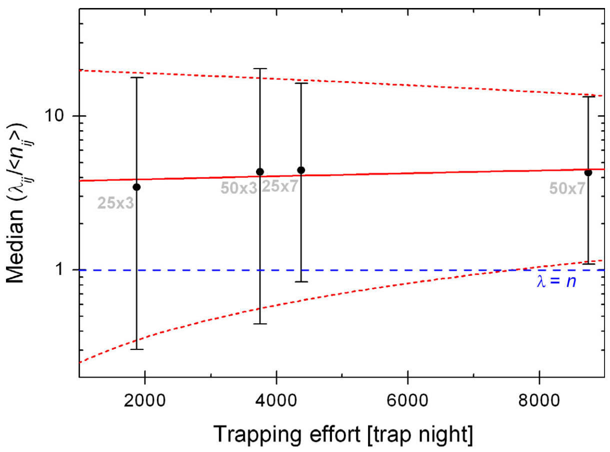

Fig. 5. Cell abundance estimates λ_{ij} (mean and 95% CI) provided by occupancy models normalized by corresponding mean cell abundances at each cell $\langle n_{ij}\rangle$. The median was calculated over all simulated cells with $\langle n_{ij}\rangle > 0$. Colour lines are visual aids.

GLMs only showed slightly lower performance at the highest trapping efforts. For a given frequency of correct associations of *sp*, univariate tests would require about twice the trapping effort than modelling methods. *t*-tests were also less effective with seven-day data from fewer transects than three-day data from more transects for the *slightly preferred* variable.

Overall, the *moderately avoided* variable (*ma*) was correctly associated more often than *sp*, but less than *hp* (Fig. 4c). Using GLMMs or OMs led to correct results more frequently than univariate tests, while GLMs had the poorest performance. For a given trapping effort, both GLMMs and GLMs seemed to work better employing seven-day than three-day data. On the contrary, *t*-tests were more effective with three-day data, as was the case for *sp* and *hp*.

ABUNDANCE AND DETECTABILITY ESTIMATION

Occupancy models allow the estimation of both detectability and the relation between abundance and environmental variables. The fraction of cells where fitted abundance was overestimated, that is $\langle n_{ij}\rangle$ laid below the 95% confidence interval of the corresponding λ_{ij}, increased with trapping effort. To illustrate this effect, we plotted medians of $\lambda_{ij}/\langle n_{ij}\rangle$ (mean and 95% confidence interval) evaluated over all cells with $\langle n_{ij}\rangle > 0$ in function of trapping effort (Fig. 5). Confidence intervals shrink with more data, as expected, but do not converge to actual abundances, indicating that abundance estimates are biased.

Discussion

According to our review, an increasing proportion of studies employed modelling methods in recent years. Our study suggests that modelling approaches are expected to yield accurate results regarding habitat selection at small scales more often than univariate tests. This may not be a novel statement (Garamszegi *et al.* 2009), but we hope illustrating this point will encourage further development and adoption of detailed modelling methods.

Performance of all statistical methods improved with higher trapping effort, as more transects were included in the analysis. This result implies that no bias occurred in detected preferences although the assumption of independence was not met because animals moved between neighbouring sample units. Stronger simulated animal preferences were more frequently detected, while weaker preferences required larger samples to be detected. Thus, while correct partial information was obtained in most cases, a complete picture of animal preferences (i.e. what we called *perfect* in our analysis) was uncommon unless very large samples were analysed (about 10 000 tn). Jorgensen (2004) already noted the need for larger sampling efforts based on a literature survey up to year 2000, having a median effort of 5000 tn. Although our survey median is slightly larger, only about 47% of the studies conducted in the period 2001–2015 reached efforts above 10000 tn. Frequencies of correct and wrong associations are intimately related to the model used to generate the data sets and cannot be extrapolated directly to field situations. Yet, our results suggest that typical trapping efforts might not suffice for properly assessing microhabitat selection. Selection in two- and three-dimensional habitats was not addressed in our study and might be even more challenging: extra degrees of freedom for animal movements could increase probabilities of detecting animals in not selected places as a consequence of spillover from good sites (Morris 1997).

Model selection through stepwise and model-averaging approaches led to similar frequencies of correct/wrong results. Nevertheless, we must note that we classified hundreds of simulated data sets automatically while a unique data set is available in real cases. Therefore, a single model obtained through stepwise selection may leave important associations out, especially weaker ones. Conversely, AIC-based selection of multiple candidate models offers a broader perspective (Burnham & Anderson 2002).

For a given trapping effort, univariate methods performed better with three-day data than seven-day data because the latter reduce the number of transects (and so presence/absence points) in exchange for more trapping nights. In addition, long surveys of abundant and highly detectable species could lead to too few absence points and lack of contrast between preferred/not-preferred habitats for univariate tests. On the contrary, modelling methods, which account for use intensity, were less sensitive to how total trapping effort was distributed among days and transects. Yet, for a given trapping effort, seven-day data yielded slightly better results likely because including less transects reduced the variability in population abundance and distribution of habitat variables. Despite statistical methods, effects of sampling duration and number of sites may vary according to species abundance and behaviour: rare or trap-shy species may be captured only after several days while others are likely to be captured earlier (Webb 1965).

Variability between transects was detrimental for the performance of GLMs, which did not include random effects. On the other hand, OMs and GLMMs provided better and similar results, likely due to their having the same hierarchical structure to accommodate replication (Zuur *et al.* 2009) and the adoption of constant detectability in simulations (MacKenzie 2006). Non-uniform detectability, which may be caused by terrain slope, weather, illumination, trap odour, etc., could blur the connection between captures at a site and use of that site, leading to biased results (MacKenzie 2006; Perea *et al.* 2011; Kajin & Grelle 2012). In such situations, OMs including covariates for detectability could offer advantages over GLMMs.

OMs overestimated abundance, with the subsequent underestimation of detectability. This bias may be due to the violation of the model's closure assumption which, however, did not prevent microhabitat use to be properly assessed as long as abundance estimates are treated as relative intensities of use (Kajin & Grelle 2012). Spatially explicit models incorporating marked animals to assess resource selection accounting for imperfect and correlated detection have been recently developed (Royle *et al.* 2013) and are worth being explored for small-mammal microhabitat use.

EXAMPLES FROM REAL SURVEYS

To exemplify some points discussed above, we picked four papers from our literature review that studied microhabitat selection by the rodent *Akodon montensis* in subtropical South American forests using different approaches. The comments below are hypothetical and not meant to criticize cited authors.

Püttker *et al.* (2008) compared habitat variables between used and not used units in six plots through ANOVAs and Goodin *et al.* (2009) applied a stepwise GLM to find associations between captures and habitat variables in four plots, both in the Atlantic Forest. Captures were negatively related to canopy cover in both works, but Goodin *et al.* (2009) also found significant relations with forbs, shrubs, bromeliads and bamboo. Püttker *et al.* (2008) measured bamboo too, but did not make conclusions about it due to significant bamboo × plot interactions in their ANOVA. Perhaps, a GLMM might have allowed Püttker *et al.* to account for variability among plots explicitly and tell whether there was an overall preference for bamboo-rich sites or not. On the other hand, Melo *et al.* (2013) conducted a correspondence analysis that associated *A. montensis* with cover-providing ferns, which Goodin *et al.* excluded in their stepwise approach; a model-averaging analysis might have considered ferns as an alternative predictor of use. Other variables quantifying cover-providing plants in Goodin *et al.*' work had *weaker effects* on selection ($|\text{estimate}| < 1{\cdot}4$) than the *strong avoidance* of sites having high proportion of bare ground ($\text{estimate} = -3{\cdot}1$). Ferns are likely to have also a weak effect, therefore being hard to detect.

Dalmagro & Vieira (2005) studied a single plot in the Araucaria Forest where *A. montensis* was positively related to canopy cover, contrary to above studies. Although microhabitat requirements may be specific to each macrohabitat and cannot be generalized (Jorgensen 2004), this contrast may also result from different detectability instead of different selection. An explicit analysis of detectability, which could have distinguished between these two possibilities, was likely possible since all cited studies trapped for ≥6 consecutive days. For instance, Kajin & Grelle (2012) applied OMs to their 5-day data of an arboreal species in the Atlantic Forest, which comprised similar total trapping effort, and successfully found that terrain slope affects detectability.

RECOMMENDATIONS

- We encourage using modelling approaches instead of univariate tests. Models should incorporate hierarchical structure if present in sampled data (as we did in GLMMs and OMs).
- Strength of selection affects the probability of detecting differential use. Model-averaging approaches may offer a broader perspective than stepwise selection, which may discard biologically important variables.
- Non-uniform detectability should be dealt with OMs including appropriate covariates, which in turn require more trapping nights at each site, especially for rare or cryptic species. However, if detectability is uniform, GLMMs are a simpler alternative. Whenever possible, sources of variability on detectability should be controlled through careful sampling design (Banks-Leite *et al.* 2014).
- Our simulation results support Jorgensen's (2004) call for larger sampling efforts. Using modelling approaches, there is no *a priori* best distribution of sampling effort between number of transects/grids vs. number of trap nights in each transect/grids. Survey design should be chosen according to the scope and goals of the study, heterogeneity of the study area (which determines the number of replicates) and characteristics of studied species (abundance, behaviour, detectability, seasonality). Logistical and practical constraints such as accessibility, funds and manpower are also to be taken into account when planning a research study (Banks-Leite *et al.* 2014).
- Researchers should be aware of the variability in conclusions induced by the choice of sampling design and statistical methods when discussing results from various sources regarding habitat selection by a species.

Acknowledgements

We thank the reviewers and the associated editor for their insightful comments which helped improving the manuscript. This work made use of free software packages GNU/OCTAVE and R, and we are grateful for the support of their respective developers. This work was supported by CONICET under grant PIP 112-200801-01410 and University of Buenos Aires under grant UBACyT 2011-2014 20020100100512 and UBACYT 2014-2017 20020130100282BA.

Data accessibility

This manuscript includes no data. OCTAVE script used to simulate trapping data and R script used to evaluate N-mixture models are provided as online supporting information (Appendices S4 and S5).

References

Abramson, G., Giuggioli, L., Kenkre, V., Dragoo, J., Parmenter, R., Parmenter, C. & Yates, T. (2006) Diffusion and home range parameters for rodents: *Peromyscus maniculatus* in New Mexico. *Ecological Complexity*, **3**, 64–70.

Alldredge, J.R. & Ratti, J.T. (1986) Comparison of some techniques for analysis of resource selection. *The Journal of Wildlife Management*, **50**, 157–165.

Banks-Leite, C., Pardini, R., Boscolo, D., Cassano, C.R., Puttker, T., Barros, C.S. & Barlow, J. (2014) Assessing the utility of statistical adjustements for imperfect detection in tropical conservation science. *Journal of Applied Ecology*, **51**, 849–859.

Bartoń, K. (2013) MuMIn: Model selection and model averaging based on information criteria. R package version 1.9.5. URL: http://CRAN.R-project.org/package = MuMIn.

Bennett, A. (1990) Habitat corridors and the conservation of small mammals in fragmented forest environment. *Landscape Ecology*, **4**, 109–122.

Bilenca, D. & Kravetz, F. (1998) Seasonal variations in microhabitat use and feeding habits of the pampas mouse *Akodon azarae* in agroecosystems of central Argentina. *Acta Theriologica*, **43**, 195–203.

Bolker, B.M. (2013) lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0-4. URL: http://CRAN.R-project.org/package = lme4.

Boyce, M.S., Vernier, P.R., Nielsen, S.E. & Schmiegelow, F.K.A. (2002) Evaluating resource selection functions. *Ecological Modelling*, **157**, 281–300.

Burnham, K.P. & Anderson, D.R. (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd edn. Springer-Verlag, New York, New York, USA.

Dalmagro, A.D. & Vieira, E.M. (2005) Patterns of habitat utilization of small rodents in an area of Araucaria forest in Southern Brazil. *Austral Ecology*, **30**, 353–362.

Dénes, F.V., Silveira, F.L. & Beissinger, S.R. (2015) Estimating abundance of unmarked animal populations: accounting for imperfect detection and other sources of zero inflation. *Methods in Ecology and Evolution*, **6**, 543–556.

Douglass, R.J. (1989) The use of radio-telemetry to evaluate microhabitat selection by deer mice. *Journal of Mammalogy*, **70**, 648–652.

Efford, M. (2004) Density estimation in live-trapping studies. *Oikos*, **106**, 598–610.

Ellis, B.A., Mills, J.N., Childs, J.E., Muzzini, M.C., McKee, K.T. Jr, Enria, D.A. & Glass, G.E. (1997) Structure and floristics of habitats associated with five rodent species in an agroecosystem in Central Argentina. *Journal of Zoology*, **243**, 437–460.

Garamszegi, L.Z., Calhim, S., Dochtermann, N., Hegyi, G., Hurd, P.L., Jørgensen, C. *et al.* (2009) Changing philosophies and tools for statistical inferences in behavioral ecology. *Behavioral Ecology*, **20**, 1363–1376.

Garshelis, D.L. (2000) Delusions in habitat evaluation: measuring use, selection, and importance. *Research Techniques in Animal Ecology: Controversies and Consequences* (ed. M. Pearl, L. Boitani & Fuller T), pp. 111–164. Columbia University Press, New York.

Gomez, D., Sommaro, L., Steinmann, A., Chiappero, M. & Priotto, J. (2011) Movement distances of two species of sympatric rodents in linear habitats of Central Argentine agro-ecosystems. *Mammalian Biology*, **76**, 58–63.

Goodin, D.G., Paige, R., Owen, R.D., Ghimire, K., Koch, D.E., Chu, Y.K. & Jonsson, C.B. (2009) Microhabitat characteristics of *Akodon montensis*, a reservoir for hantavirus, and hantaviral seroprevalence in an Atlantic forest site in eastern Paraguay. *Journal of Vector Ecology*, **34**, 104–113.

Heithaus, M.R., Hamilton, I.M., Wirsing, A.J. & Dill, L.M. (2006) Validation of a randomization procedure to assess animal habitat preferences: microhabitat use of tiger sharks in a seagrass ecosystem. *Journal of Animal Ecology*, **75**, 666–676.

Hodara, K., Busch, M., Kittlein, M.J. & Kravetz, F.O. (2001) Density-dependent habitat selection between maize cropfields and their borders in two rodent species (*Akodon azarae* and *Calomys laucha*) of Pampean agroecosystems. *Evolutionary Ecology*, **14**, 571–593.

Ives, A.R. (2015) For testing the significance of regression coefficients, go ahead and log-transform count data. *Methods in Ecology and Evolution*, **6**, 828–835.

Jorgensen, E.E. (2002) Small mammals: consequences of stochastic data variation for modelling indicators of habitat suitability for a well-studied resource. *Ecological Indicators*, **1**, 313–321.

Jorgensen, E.E. (2004) Small mammal use of microhabitat reviewed. *Journal of Mammalogy*, **85**, 531–539.

Kajin, M. & Grelle, C.E.V. (2012) Microhabitat selection when detection is imperfect: the case of an endemic Atlantic Forest mammal. *Ecological Research*, **27**, 1005–1013.

Lunn, D.J., Thomas, A., Best, N. & Spiegelhalter, D. (2000) WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, **10**, 325–337.

MacKenzie, D.I. (2006) Modeling the probability of resource use: the effect of, and dealing with, detecting a species imperfectly. *The Journal of Wildlife Management*, **70**, 367–374.

MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L.L. & Hines, J.E. (2006) *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence*. Academic Press, Burlington.

Manly, B.F., McDonald, L.L., Thomas, D., McDonald, T.L. & Erickson, W.P. (2002) *Resource Selection by Animals: Statistical Design and Analysis for Field Studies*, 2nd edn. Springer, London.

Melo, G.L., Miotto, B., Peres, B. & Cáceres, N.C. (2013) Microhabitat of small mammals at ground and understorey levels in a deciduous, southern Atlantic Forest. *Anais de Academia Brasileira de Ciências*, **85**, 727–736.

Mendel, S.M. & Vieira, M.V. (2003) Movement distances and density estimation of small mammals using the spool-and-line technique. *Acta Theriologica*, **48**, 289–300.

Morris, D.W. (1987) Ecological scale and habitat use. *Ecology*, **68**, 362–369.

Morris, D.W. (1997) Optimally foraging deer mice in prairie mosaics: a test of habitat theory and absence of landscape effects. *Oikos*, **80**, 31–42.

Octave development community (2012) GNU/Octave. URL: www.gnu.org/software/octave

Perea, R., González, R., San Miguel, A. & Gil, L. (2011) Moonlight and shelter cause differential seed removal and selection by rodents. *Animal Behaviour*, **84**, 717–723.

Prevedello, J.A., Garcia Rodrigues, R. & Leite de Araujo Monteiro-Filho, E. (2010) Habitat selection by two species of small mammals in the Atlantic Forest, Brazil: Comparing results from live-trapping and spool-and-line tracking. *Mammalian Biology*, **75**, 106–114.

Priotto, J.W. & Steinmann, A.R. (1999) Factors affecting home range size and overlap in *Akodon azarae* (Muridae: Sigmodontinae) in natural pasture of Argentina. *Acta Theriologica*, **44**, 37–44.

Püttker, T., Pardini, R., Meyer-Lucht, Y. & Sommer, S. (2008) Responses of five small mammal species to micro-scale variations in vegetation structure in secondary Atlantic Forest remnants, Brazil. *BMC Ecology*, **8**, 9.

R Core Team (2013) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna. URL: http://www.R-project.org.

Royle, J.A. & Nichols, J.D. (2003) Estimating abundance from repeated presence-absence data or point counts. *Ecology*, **84**, 777–790.

Royle, J.A., Chandler, R.B., Sun, C.C. & Fuller, A.K. (2013) Integrating resource selection information with spatial capture-recapture. *Methods in Ecology and Evolution*, **4**, 520–530.

Ruxton, G.D. (2006) The unequal variance t-test is an underused alternative to Student's t-test and the Mann-Whitney U test. *Behavioral Ecology*, **17**, 688–690.

Sturtz, S., Ligges, U. & Gelman, A. (2005) R2WinBUGS: a package for running WinBUGS from R. *Journal of Statistical Software*, **12**, 1–16.

Webb, W.L. (1965) Small mammal populations on islands. *Ecology*, **44**, 479–488.

Ylönen, H., Jacob, J., Davies, M.H. & Singleton, G.R. (2002) Predation risk and habitat selection of Australian house mice *Mus domesticus* during an incipient plague: desperate behaviour due to food depletion. *Oikos*, **99**, 284–289.

Zar, J.H. (2010) *Biostatistical Analysis*, 5th edn. Prentice Hall, Upper Saddle River, New Jersey, USA.

Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A. & Smith, G.M. (2009) *Mixed Effects Models and Extensions in Ecology With R*. Springer Science+Business Media, New York.

Received 3 April 2016; accepted 1 June 2016
Handling Editor: David Murrell

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Appendix S1. Literature survey on small mammal microhabitat selection studied through the live-trapping technique.

Appendix S2. Details about the model and simulations of small-mammal movements and trapping procedures.

Appendix S3. Characteristics and application examples of evaluated statistical methods.

Appendix S4. OMs.R: R script for estimating OMs parameters.

Appendix S5. SIMULATION.M: Octave script for simulating live-trapping data.