Geography Department, University of Calgary

An automated method to derive habitat preferences of wildlife in GIS and telemetry studies: A flexible software tool and examples of its application[1]

By F. HUETTMANN and JULIA LINKE, Calgary

1 Introduction

In order to conserve wildlife species effectively, it is crucial to know their habitat preferences (e.g. AEBISCHER et al., 1993). However, in most cases it is very difficult to infer which habitat wildlife really needs. Although the application of telemetry and GIS have widely improved our knowledge in which habitat the animals are to be found (e.g. WHITE and GARROT, 1990; HOOGE and EICHENLAUB, 1997), it still presents a difficult task to know what the true habitat preferences of the studied wildlife species are (e.g. MORRISON, 2001). Only when the habitat preferences are known and quantified one can effectively manage these specific habitats, and, thus try to safeguard the habitat and its wildlife species of concern.

The international and statistical wildlife research literature has addressed these issues for a long time. One will find a wide array of terms and concepts that express the link between wildlife and habitat, such as habitat use, habitat choice, habitat selection and habitat preference (e.g. JOHNSON, 1980; AEBISCHER et al., 1993; JONES, 2000). The work by MANLY et al. (1993, 2002), MLADENOFF and SICKLEY (1998) and MLADENOFF et al. (1999) summarize well the recent approaches (reviewed e.g. in BOYCE and MCDONALD, 1999). One of the topics of debate in wildlife studies is centered around the assumption about (i) what habitat is available to the animal; other questions deal with (ii) the certainty of 'absences', (iii) the use of many predictors to explain animal distribution and abundance, and (iv) how to test in a sound fashion for significant predictors, or for the ones that are the most relevant ones in a parsimonious meaning. As outlined by BURNHAM and ANDERSON (1998) (see also CHERRY, 1998; ANDERSON et al., 2000) these points are crucial for a rigorous scientific approach since they affect hypothesis testing and parsimony to derive and infer valid knowledge for wildlife and their habitat link.

So far, no standardized and convenient software tool exist that deals with the implementation of Resource Selection Functions (RSF) and parsimonious model selections brought forward by MANLY et al. (1993, 2002) and BURNHAM and ANDERSON (1998). In the following, we present a method, summarized in a SPLUS software code, that allows relatively easily to derive habitat preferences for wildlife. In a flexible but robust way it also allows to apply several modeling algorithms and model selection techniques; it can test parallel numerous hypothesis in a sound way. The presented approach is based on a telemetry-GIS scenario, but it can easily be adjusted for other applications with 'presence only' data (see paragraph below for examples).

2 Materials and Methods

2.1 Wildlife data

Locating wildlife is a crucial component in many wildlife and field studies. Wildlife occurrences (presence locations) are often known from geo-referenced transect or aerial counts, opportunistic sightings, or records of hunting events. However, in many cases it is not known where these animals do not occur (absence); a crucial piece of information if one wants to differentiate presence habitats from absence habitats, and for inferring habitat preferences (KERY, 2002; SUITOR and HUETTMANN; 2002). In order to compensate for this situation, random data (pseudo-absence) is often used (e.g. MANLY et al., 1993, 2002; HUETTMANN et al., in review). This scenario applies specifically to telemetry studies, e.g. when only a subset of the overall population gets investigated but the presences of 'unmarked' animals in the overall population remain unknown. Similar situations are found in classical 'presence only' data sets such as geo-referenced hunting records, wildlife sightings, nesting records, museum collections and disease outbreaks even.

For comparing presence information with pseudo-absence information, random locations are easily obtained through a random draw of the X and Y coordinates within the habitat available to the individual or the studied population. Studies using telemetry and GIS can derive the available habitat by using Minimum Convex Polygons, Kernel Homeranges (e.g. HOOGE and EICHENLAUB, 1997), or other 'spatial filters' to exclude habitat from the investigation that is not available to the animals. Random draws and the selection of available habitat can nowadays be done in an automated fashion with a computer and GIS methods (e.g. HOOGE and EICHENLAUB, 1997; HUBER, 1999).

Although our GIS work is based on ArcGIS, ArcView and ArcView extensions/scripts this component of the presented approach can be easily implemented into most other standard GIS software packages.

2.2 Wildlife-habitat link

In order to investigate the (quantitative) link between wildlife and habitat, multiple regressions are widely used. Most often, this is implemented as Generalized Linear Models (GLM); they are robust, and their performance and behaviour are well known (e.g. McCULLAGH and NELDER, 1989; MENARD, 2001). They form the core of the widely accepted and applied method of Resource Selection Functions (MANLY et al., 1993, 2002). They quantify, and can test for, true habitat preferences by taking the availability fully into account using wildlife presence and absence (or random) scenarios and the proportion of the underlying habitat information (see also POTVIN et al., 2001).

Usually, many predictors are used to build a wildlife-habitat model. However, since science is based on the parsimonial paradigm, it is critical to know which predictors perform best and which of the many predictors simplify the model to a desired degree.

2.3 Model selection

Habitat data have started to become freely available over the WWW and therefore many habitat studies use multiple predictors for explaining wildlife distribution and abundance. However, this approach can raise several methodological and statistical issues. Predictors could be correlated or interacting with each other; this situation can create a complex statistical research problem to solve (e.g. QUINN and KEOUGH, 2002). Often, it has hindered the progress in our understanding of how multivariate wildlife-habitat relationships act. So far, one problem was that for a parsimonious approach the best models had to be select-

ed. In order to achieve this, either the interactions had to be ignored or simplified, or the actual model selection process was unspecific, making valid inference difficult. This could result into the public dismissal of the modeling approach due to lack of trust in the results. Using many predictors in a model raises another problem about traditional hypothesis testing. Since classical wildlife-habitat research statistics are based on simple hypothesis tests and p-values (e.g. CHERRY, 1998; MORRISON, 2001), these were used to determine the significance of predictors and 'best' models to infer the distribution and abundances (see also ANDERSON et al., 2001, 2002; ROBINSON and WAINER, 2002). However, there has been increased use of the information criteria (e.g. BURNHAM and ANDERSON, 1998) to provide sound evidence of the best model to be selected from a pre-specified set of candidate models. These approaches diminish the traditional p-value approach and do not produce 'statistically significant' results in the classical sense, *sensu* hypothesis testing (ANDERSON et al., 2002). Most of these information criteria are centered around the Kullback-Leibler information, e.g. Akaike's Information Criteria (AIC, AKAIKE, 1974), which have received an increasing acceptance in recent times (e.g. ANDERSON et al., 2000). The best model is selected as an optimized tradeoff between accuracy and number of predictors.

2.4 The general approach

In order to implement the topics mentioned above towards an automated and convenient approach, the following procedures are pursued (see Fig. 1):

Steps 1 and 2
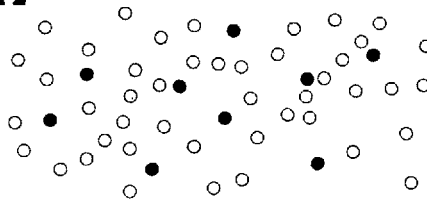
Presence ●
Absence/Random ○

Step 3
Random draw from pool of absence locations
(presence samples equal to absence/random)

Step 4

Response	Predictor 1	Predictor 2
E.g. 9 presences 9 absences		More predictors possible . .

Step 5
GLM Response ~ Predictor 1
GLM Response ~ Predictor 2
GLM Response ~ Predictor 1 + Predictor 2

Step 6
Select best model, e.g. using AIC

Step 7
Store details of selected model into file
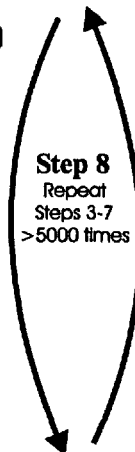
Step 8
Repeat
Steps 3-7
> 5000 times

Fig. 1. Methodological steps to derive habitat preference.

Step 1. The 'presence only' locations for the species of investigation are compiled, e.g. from a GIS.

Step 2. Random locations ('pseudo-absences') are created within the 'available' habitat, e.g. from a GIS.

Step 3. For a balanced statistical model the same number as 'presence only' locations are drawn randomly from a larger pool of random locations within the available habitat.

Step 4. The 'presence only' data are merged with the 'random locations' into one table.

Step 5. A hypothesis is formulated as an individual algorithm, e.g. GLM. It is also possible to have many hypothesis, which then get formulated as specific individual models.

Step 6. The best model is selected, e.g. using the lowest AIC value.

Step 7. The best model and all its parameters are stored, e.g. in an ASCII text file.

Step 8. Steps 3–7 get repeated 'infinite times' in order to obtain certainty (in most cases 1000–5000 repeats are sufficient; MANLY et al., 1993, 2002; HUETTMANN unpublished).

Step 9. Analyse the selected models, e.g. predictors, their coefficients and the AIC.

We developed a SPLUS 2000 (MathSoft 1999) script to run an automated loop for the Steps 3–7. Based on the initial GIS work, this script selects random samples from the overall pool of random locations, merges 'presence only' with random locations, selects automatically the best model for the GLM scenarios by using the lowest AIC, and saves the results into an ASCII file ready for (visual) exploration and for further analysis. However, the general approach can easily be implemented into other standard programming languages as well, also using different modeling algorithms, such as Classification and Regression Trees, Neural Networks, Multiple Adaptive Regression Splines, and many others, e.g. YEN et al. (in review). The script could also easily be modified to use different model selection criteria, e.g. p-values, BIC (Bayesian Information Criteria), AICc, QAIC (QUINN and KEOUGH, 2002).

3 Results

An example of the SPLUS script is presented in the Appendix. In the following, we present applications of the presented method, including its most important results, strengths and bottlenecks:

Example 1: Marbled Murrelet nesting habitat selection (HUETTMANN et al., in review)

The application: Eighty-five nesting locations of Marbled Murrelets *(Brachyramphus marmoratus)*, an endangered North American seabird of the northern Pacific, were obtained through telemetry from a helicopter. This presents one of the largest nest samples known for this species. The nest locations were plotted with a GIS, and random locations were created within the available nesting habitat. For nest and random locations, the elevation, slope and aspect were derived from a DEM (Digital Elevation Model) GIS layer. The approach shown above was used, based on the stepAIC command for the initial model of presence/random ~ elevation + slope + aspect and all its interactions and squared terms. In a stepwise model selection process based on AIC, it was tested which of the individual predictors were selected the most (Table).

Strengths: The model selection is done in a consistent fashion allowing for an objective conclusion. The model selection procedure is repeated 'infinite times' (here 1000 repeats), always with a new randomly drawn 'pseudo absence' data set. This allows to test the full

Table. Influence of slope (in 10° classes), elevation (in 100m classes) and aspect (N,E,S,W, Flat) on nesting habitat selection of Marbled Murrelets for the years 1998–2000 (only the 3 most frequently selected models are shown). Summary result from 1000 tests of Marbled Murrelet nest preferences for slope and aspect. Results are derived from Sample B: Locations that only fall in Old Forest polygons + 100 m bufferzone (see also HUETTMANN et al., in review).

Sample	Model Rank	Model Specification	Number Selected	Mean (SD) Coefficients for the most commonly recurring model	
B	1	Slope Elevation²	118	INTERCEPT Slope Elevation²	– 0.526 (0.231) 0.346 (0.084) -0.007 (0.001)
	2	Slope	114	INTERCEPT Slope	– 0.845 (0.209) 0.306 (0.828)
	3	Slope² Elevation²	77	INTERCEPT Slope² Elevation²	– 0.209 (0.199) 0.067 (0.020) – 0.006 (0.001)

landscape and is required when the tested predictors show a strong variation within the overall available habitat/landscape.

Bottleneck: One of the biggest constraints was to introduce the concept of AIC, rather than traditional significances, into the management process and build trust in the results. A second constraint is the stepwise approach to select predictors (QUINN and KEOUGH, 2002; HARRELL, 2002), which can be potentially biased by the order of the predictors in the individual model.

Example 2: Testing for forest edge preferences of Marbled Murrelet nest locations (HUETTMANN et al., in review)

A forest cover layer of Old-Growth Forest was used to derive proximities of Marbled Murrelet nest locations and random locations falling within this forest type (available habitat). The described SPLUS script was used to test the hypothesis how the proximity to forest edge affects nest locations. Only edge proximity was used as predictor (formula response ~ edge proximity); it was not tried to select predictors for the best model and thus, this component of the SPLUS script was not used. This approach allows to derive regression coefficients for each run. The regression coefficient was summarized from 1000 runs (see Fig. 2) to capture the relationship that Marbled Murrelet nests have with the edges of Old-Growth forests in the landscape; a question of major conservation interest.

Strengths: Proximities to habitat features rather than the directly underlying habitat is used. The 1000 repeats test in a robust fashion how edge proximities affect nest locations. The mean coefficients can be used to interpret the statistical relationship that nest locations have with linear edge features, e.g. a positive (preference) or negative (avoidance) association and its magnitude.

Bottleneck: Proximities to edge features were originally derived from aerial photography, which are subject to interpreter's perception and bias (FORTIN et al., 2000). On a very small scale, and due to GIS mapping errors, such proximities are often only accurate within the 'classification' distance, but the approach allows well to derive a general trend.

Example 3: Testing for preferences of seismic lines for Grizzly Bears (LINKE et al. in review)

Grizzly Bear locations were obtained from Satellite-GPS telemetry carried out in the Foothills Model Forest of Alberta, Canada. For each Grizzly Bear, individual Minimum
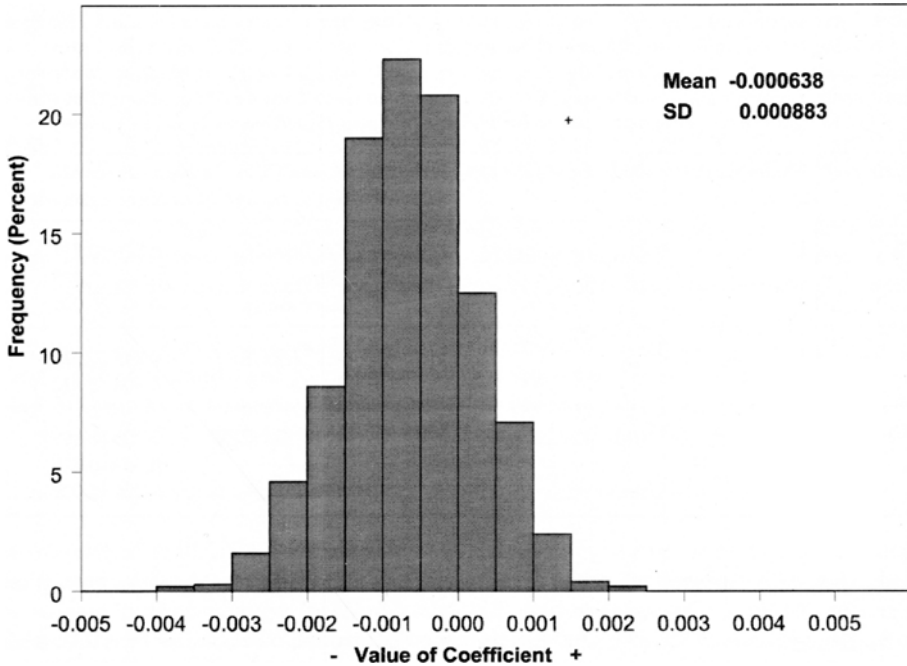
Fig. 2. Distribution of regression coefficients for Marbled Murrelet nests in relation to old-growth forest edges from 1000 tests.

Convex Polygons (MCP) were computed in order to delineate the habitat available to these animals. Random locations were computed in a GIS to derive 'pseudo-absence' locations for each individual MCP. For each of the Grizzly Bear and random locations, the direct proximities to Seismic Lines (derived from Indian Remote Sensing IRS imagery, Linke et al., in review) were computed in a GIS. Three hypothesis were tested whether proximity to seismic lines, a random distance or a constant distance would be determining Grizzly Bear presence (= GPS locations). The model selection was done using the extractAIC command (VENABLES and RIPLEY, 2002; MASS library), which computes the AIC value for each specified model; the model with the lowest AIC was selected.

Strengths: Several hypothesis were tested within the same model run for the same merged data set allowing for a solid result to draw inference on.

Bottleneck: Some Grizzly Bear GPS locations (presences) could be missing due to lack of satellite connections. The true model could still not be in the set of candidate models (e.g. ANDERSON and BURNHAM, 2002).

Other examples and applications:

For presence and random locations the presented approach can deal with underlying habitat types and/or proximities to habitat features; these can easily be derived from the GIS. Details of the specific statistical test are defined in the hypothesis, and captured in the modelling algorithm.

The presented approach is not limited to the three examples given above. Instead of a balanced sample as used here, e.g. same number of presence and absence/random locations, one could define a higher absence/random sample, drawn randomly from the overall pool of random locations. This approach was favored by MANLY et al. (1993, 2002) and poten-

tially can be more robust. Also, it is very easy to replace in the existing SPLUS script (see Appendix) the AIC model selection criteria with other criteria such as p-values, BIC, Schwartz-Criteria, AICc, QAIC etc. (BURNHAM and ANDERSON, 1998; QUINN and KEOUGH, 2002). It is possible to include these different model selection methods parallel and as individual hypothesis even. This allows also to evaluate how the individual model selection criteria perform for the same data and whether they could seriously affect conservation decisions (HUETTMANN, unpublished).

Instead of the classical GLM modeling algorithm, linear models can also be used, also different link functions, e.g. with poisson or gaussian distributions (e.g. HUETTMANN and LINKE, in press; QUINN and KEOUGH, 2002). Another highly effective option can be achieved when using non-linear modeling algorithms such as Neural Networks, Classification and Regression Trees, MARS and others (HUETTMANN and DIAMOND, 2001; YEN et al., in review). Most of these algorithms offer a (non-significant) ranking on the importance of individual predictors. We think that these approaches allow for a very effective and powerful test on the importance of individual predictors in a non-linear fashion.

The use of 'presence only' data strongly increased in recent times. This was made possible with freely accessible GIS-data over the WWW, and due to an increased awareness of existing data sets, such as digitized museum collections. We think that the approach shown here fits well with existing algorithms that deal with 'presence only' information such as GARP, Biomapper, Ecological-Niche Factor Analysis, DOMAIN, Mahalanobis Distance and others (e.g. PETERSON et al., 1999; DUNN and DUNCAN, 2000; HIRZEL, 2001; HIRZEL et al., 2002; SUTTOR and HUETTMANN, 2002).

4 Discussion

The outlined steps and general methods of the SPLUS script build, in parts, on existing theory (e.g. MANLY et al., 1993, 2002). However, a flexible and convenient software tool, and experiences from its application to real conservation questions were not available, so far. The results show that with the automated approach habitat preferences are relatively easy and quick to obtain. Model runs with approximately 1000 simulations take less than 6 h on a modern computer (PC IBM).

The presented SPLUS script is flexible enough to be adjusted for a whole variety of wildlife management, research and conservation questions, e.g. proximity preferences to spatial features, preferences of underlying habitat, testing many hypothesis, using many predictors, testing for specific interactions among predictors, and using 'presence only' data. Its advantages are the flexibility and automation of the actual testing procedure.

Model selection results derived from classical p-values do not necessarily match with the ones derived from AIC values (e.g. ANDERSON and BURNHAM, 2002). It was found that results from AIC selections are still difficult to communicate and to present to a general audience, to the public and to wildlife managers, which are mostly trained to infer from significances derived from p-values. However, the AIC approach allows for a robust inference and is much more accepted in statistical and scientific circles nowadays (e.g. ANDERSON and BURNHAM, 1998). Potential constraints with the presented approach are caused by the required skill to run the SPLUS script, and by the need to work-up data in a GIS and in SPLUS. If telemetry data are used, obtaining valid and unbiased geo-referenced empirical data could present another problem; for instance, the investigated animals need to be representative for the population and their behaviour not to be affected by autocorrelation, handling and gear.

The SPLUS script presented here was made possible due to free and accessible statistical software algorithms, e.g. AIC commands in the MASS library (VENABLES and RIPLEY, 2002) from the WWW. Together with free (habitat/GIS) data over the WWW, such approaches are likely to improve many unresolved questions in the field of wildlife management and research; we suggest they are further pursued.

Since the use of GIS and telemetry is still increasing, it is hoped that the presented approach finds a wide consideration and appreciation. Further improvements of the presented approach, e.g. data fitting tests, web-availability, developing a user-friendly Graphical User Interface (GUI) and a direct GIS-connection, are suggested so that the wildlife-habitat link can be better studied and understood, and that both, wildlife and habitat, can be effectively managed and conserved for the benefit of mankind.

Acknowledgements

Summary

The conservation of wildlife is closely linked with its habitat. Knowing the preferred habitat is crucial for most managers and researchers to conserve wildlife effectively. Here we present an software approach to 'presence only' (wildlife location) data in order to derive habitat preferences in an automated but flexible fashion. The approach is based on presence and random locations, e.g. derived from a telemetry and GIS study, and uses for instance a GLM (Generalized Linear Model) to develop Resource Selection Functions (RSF). The presented approach is very flexible and computationally powerful and can easily be adjusted for various applications. We show how this approach is used for conservation applications by using endangered and charismatic wildlife species as examples. In addition, we give an outlook of future applications for the presented method.

Key words: Wildlife habitat preferences, GIS, wildlife telemetry, GLM (Generalized Linear Models), Resource Selection Functions (RSF), SPLUS software tool

Zusammenfassung

Eine automatisierte Methode zur Bestimmung der Habitatpräferenzen von Wildtieren in GIS-und Telemetrie-Studien: Ein flexibles Software-Werkzeug und Beispiele seiner Anwendung

Die Erhaltung von Wildtieren ist nur möglich, wenn auch deren Habitat vollkommen mitberücksichtigt wird. Die Kenntnis der Wildtier-Habitatpräferenz ist meist grundlegend dafür, dass Manager und Wissenschaftler Wildtiere dann auch in effektiver Weise erhalten können. Hier wird ein Software-Ansatz vorgestellt, welcher auf „Presence-Only"-Daten (Lediglich Wildtier-Lokationen) basiert, um in einer automatisierten aber flexiblen Weise die Habitatpräferenzen zu erhalten. Dieser Ansatz basiert auf Wildtierpräsenz- und Zufalls-Lokationen, z. B. erhoben durch eine Telemetrie- und GIS-Studie, und benutzt beispielsweise ein GLM (Generalized Linear Model) um dann Resource Selection Functions (RSF) zu entwickeln. Der vorgestellte Ansatz ist sehr flexibel und computer-effizient; er kann

relativ leicht an diverse Anwendungen in der Wildltier-Biologie angepasst werden. Dieser Beitrag zeigt, wie dieses Konzept für allgemeine Naturschutzfragen eingesetzt werden kann, indem Anwendungsbeispiele aus der Praxis für bedrohte und charismatische Wildtier-Arten vorgesetllt werden. Des weiteren wird ein Ausblick dieser Methodik vorgestellt.

Schlüsselwörter: Wildtier Habitatpräferenzen, GIS (Geografische Informations Systeme), Wildtier-Telemetrie, GLM (Generalized Linear Model), Resource Selection Functions (RSF), SPLUS Software Werkzeug

Résumé

Une méthode automatisée pour déterminer les habitats d'élection de la faune sauvage dans des études de SIG et de télémétrie: un outil logiciel flexible et exemples de ses applications

La conservation de la faune sauvage est intimement liée à son habitat. La connaissance des habitats préférentiels est cruciale pour la plupart des gestionnaires et chercheurs pour procéder à une conservation effective. Nous présentons ici un logiciel axé sur des données de «seule présence» afin d'en déduire les habitats de prédilection selon un processus automatisé mais souple. L'approche est basée sur la présence et sur des localisations aléatoires, telles qu'elles résultent d'une étude télémétrique et au moyen du SIG. Elle fait appel, par exemple, à un modèle linéaire généralisé (MLG) pour développer des fonctions de sélection de ressources (FSR). L'approche présentée se montre très souple et performante et peut aisément être ajustée pour des applications variées. Nous montrons comment cette approche s'applique à des objectifs de conservation en faisant usage d'espèces menacées ou emblématiques à titre d'exemple. En complément, nous donnons les perspectives d'applications futures de la méthode présentée.

Mots clefs: habitats de prédilection pour la faune sauvage, SIG (système d'information géographique), télémétrie de la faune sauvage, MLG (modèles linéaires généralisés), FSE (fonctions de sélection de ressources), logiciel SPLUS.

Trad.: S. A. DE CROMBRUGGHE

References

AEBISCHER, N. J.; ROBERTSON, P. A., 1994: Testing for resource use and selection by marine birds: a comment. Journal for Field Ornithology 65, 210–213.

AEBISCHER, N. J.; ROBERTSON, P. A.; KENWARD, R. E., 1993: Compositional analysis of habitat use from animal radio-tracking. Ecology 74, 1313–1325.

AKAIKE, H., 1974: A new look at statistical model identification. IEEE Transactions of Automatic Control AU 19, 716–722.

ANDERSON, D. R., BURNHAM, K. P., 2002: Avoiding pitfalls when using information-theoretic methods. Journal of Wildlife Management 66, 912–918.

ANDERSON, D. R.; LINK, W.; JOHNSON, D. H.; BURNHAM, K. P.; 2001: Suggestions for presenting the results of data analyses. Journal of Wildlife Management 65, 373–378.

ANDERSON, D. R.; BURNHAM, K. P.; THOMPSON, W. L., 2000: Null hypothesis testing: problems, prevalence, and an alternative. Journal of Wildlife Management 64, 912–923.

BOYCE, M. S.; MCDONALD, L. L., 1999: Relating populations to habitats using resource selection functions. Trends in Ecology and Evolution 14, 268–272.

BURNHAM, K. P.; ANDERSON, D. R., 1998: Model selection and inference: A practical information-theoretic approach. Springer-Verlag, New York, New York, USA.

CHERRY, S., 1998: Statistical tests in publications of the Wildlife Society. Wildlife Society Bulletin 26, 947–953.

DUNN, J. E., DUNCAN, L., 2000. Partioning Mahalanobis D² to sharpen GIS classification. University of Arkansas Statistical Laboratory Technical Report No. 29.

FORTIN, M.-J.; OLSON, R. J.; FERSON, S.; IVERSON, L.; HUNSAKER, C.; EDWARDS, C.; LEVINE, D.; BOTERA, K.; CLEMAS, W. V., 2000. Issues related to the detection of boundaries. Landscape Ecology 15, 453–466.

HARREL, F., 2002: Regression Modelling Statistics. New York: Springer.

HIRZEL, A., 2001: When GIS come to life. Linking landscape- and population ecology for large pop-ulation management modelling: the case of Ibex *(Capra ibex)* in Switzerland. PhD thesis. University of Lausanne.

HIRZEL, A. H.; HAUSSER, J.; CHESSEL, D.; PERRIN, N., 2002: Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data? Ecology 83, 2027–2036.

HOOGE, P. N.; EICHENLAUB, B., 1997: Animal movement extension to Arcview. ver. 1.1. Alaska Science Center – Biological Science Office, U.S. Geological Survey, Anchorage, AK, USA.

HUBER, W. A., 1999: ArcView Random Numbers. URL:http://www.quantdec.com/arcview.htm. Quantitative Decisions, 539 Valley View Road, Suite 300, Merion Station, PA 19066-1317.

HUETTMANN, F.; DIAMOND, A.W., 2001: Seabird colony locations and environmental determination of seabird distribution: A spatially explicit seabird breeding model in the Northwest Atlantic. Ecological Modelling 141, 261–298.

HUETTMANN, F.; LINKE, J., in press: Assessment of different link functions for modeling binary data to derive sound inferences and predictions. International Conference on Computational Science and its Applications (ICCSA 2003); Computational and methodological developments in spatial analysis in GIS, Proceedings Montreal May 2003.

HUETTMANN, F.; CAM, E.; BRADLEY, R.; LOUGHEED, L.; TRANQUILLA MCFARLANE, L.; LOUGHEED, C.; COOKE, F.; ZHARIKOV, Y.; YEN, P.; LANK, D., in review: Breeding habitat selectivity by Marbled Murrelets in a fragmented old-growth forest landscape. Wildlife Monographs.

JOHNSON, D. H., 1980: The comparison of usage and availability measurements for evaluating resource preference. Ecology 61, 65–71.

JONES, J., 2001: Habitat selection studies in avian ecology: A critical review. Auk 118, 557–562.

KERY, M., 2002: Inferring the absence of a species – a case study of snakes. Journal of Wildlife Management 66, 330–338.

LINKE, J.; FRANKLIN, S. E.; STENHOUSE, G. B.; HUETTMANN, F., in review: An approach for relating landscape structure with grizzly bear population use of the Alberta foothills. Ursus.

LINKE, J.; HUETTMANN, F., unpublished: Testing for preferences of linear landscape features for Grizzly Bears. Geography Dept. University of Calgary.

MANLY, B. F. J.; MCDONALD, L. L.; THOMAS, D. L., 1993: Resource selection by animals: Statistical design and analysis for field studies. London: Chapman and Hall.

MANLY, B. J.; MCDONALD, L. L.; THOMAS, D. L.; MCDONALD, T. L.; ERICKSON, W. P., 2002. Resource selection by animals. Dordrecht: Kluwer Academic Publishers.

MATHSOFTT. 1999: S-PLUS 2000, Professional release 2. Seattle, Washington.

MCCULLAGH, P.; NELDER, J. A., 1989: Generalized linear models. Second edition. London: Chapman and Hall.

MENARD, S., 2001: Applied logistic regression analysis. Sage, University paper series on quantitative applications in the social sciences, 07–106. Thousand Oaks,California: Sage.

MLADENOFF, D. J.; SICKLEY, T. A., 1998: Assessing potential gray wolf restoration in the northeastern United States: a spatial prediction of favorable habitat and potential population levels. Journal for Wildlife Management 62, 1–10.

MLADENOFF, D. J.; SICKLEY, T. A.; WEYDEVEN, A. P., 1999: Predicting gray wolf landscape recolonisation: logistic regression vs new field data. Ecological Applications 9, 37–44.

MORRISON, M. L., 2001: A proposed research emphasis to overcome the limits of wildlife-habitat relationship studies. Journal of Wildlife Management 65, 613–623.

PETERSON, A. T.; SOBERON, J.; SANCHEZ-CORDERO, V., 1999: Conservatism of ecological niches in evolutionary time. Science 285, 1265–1267.

POTVIN, F.; LOWELL, K.; FORTIN, M.-J.; BELANGER, L., 2001: How to test habitat selection at the home range scale: A resampling random windows technique. Ecoscience 8, 399–406.

QUINN, G. P.; KEOUGH, M. J., 2002: Experimental design and data analysis for biologists. Cambridge: Cambridge University Press.

ROBINSON, D. H.; WAINER, H., 2002: On the past and future of null hypothesis significance testing. Journal of Wildlife Management 66, 263–271.

SUITOR, M.; HUETTMANN, F., 2002: Modelling absence habitat for Grizzly Bears in Kluane National Park with different modeling algorithms. Final 535 Project Report. University of Calgary, Canada.

VENABLES, W. N.; RIPLEY, B. D., 2002: Modern applied statistics with S-Plus (fourth edition). Springer Verlag, New York.

WHITE, G. C.; GARROTT, R. A., 1990: Analysis of wildlife radio tracking data. San Diego: Academic Press.
YEN, P; HUETTMANN, F.; COOKE, F., in review: Modelling abundance and distribution of Marbled Murrelets *(Brachyramphus marmoratus)* using GIS, marine data and advanced multivariate statistics. Ecological Modelling.

Appendix

A version of the SPLUS code to derive habitat preferences from presence-only and random location data (The latest version of this software is available from the author).

```
####################################################################
    ########
#Class Name   :scriptoseislpref3botao.ssc    version2              #
#Function     :Select the 'best' Model by finding the minimum AIC from
              2 models                                             #
#Author       :Dr. Falk Huettmann, supported by Botao Zheng        #
#Input Data   :Bear00499jjb Rndmpts00499jjb                        #
#Output Data  :bear004jj991000modelresults.txt                     #
#Date         :October 29, 2002                                    #
#Note: This version saves only the parameters of the first predictor; #
#      an adjustment is required if a model consists of many
       predictors                                                  #
#
#      The MASS library needs to be loaded before so that the
       extractAIC                                                  #
#      command  is available
####################################################################

{
  NumOfLoop <- 5001
#preferably a loop of > 5000 to simulate infinity

  NumOfModel <- 2
#needs to be adjusted for the number of hypothesis/models to be tested

  ii <- NumOfLoop
            while(ii > 0) {
      menuRandomSample(data = Rndmpts00499jjb, size = 294, replace =
F, save.name
            = "sampleout", show.p = T)
#rndmpts00499jjb is the file name that was imported into SPLUS and has
  the random locations to be drawn from
# file structure needs to be: column with response variable (absence
  or random=0), and columns with predictors as specified in the GLM

      ii <- ii - 1
      mergedbeartest<-copy.row(
            Bear00499jjb, "@END", sampleout,
            "@ALL")
#bear00499jjb is the file name that was imported into SPLUS and has
  the presence locations to be merged with the random draw from the
  random locations
# file structure needs to be: column with response variable (presence
  =1), and columns with predictors as specified in the GLM
```

```
table <- rep(1:5, rep(NumOfModel,5))

dim(table) <- c(NumOfModel,5)
dimXY <- dim(table)
jj<-dimXY[1]

while (jj != 0){
   table[jj,1] <- " "
   table[jj,2] <- " "
   table[jj,3] <- " "
   table[jj,4] <- " "
   table[jj,5] <- " "
   jj <- jj-1
}

# this is GLM and hypothesis 1#
testglm1<-glm(formula = Presrand ~ Slinprox, family = binomial(link =
   logit), data = mergedbeartest, na.action = na.exclude, control =
   list(epsilon = 0.0001,
   maxit = 50, trace = F))
```
CONT'D APPENDIX
```
      print('test1')
      table[1,1]<-"1"
      print(testglm1$coefficients)
      resultstestglm1<-testglm1$coefficients

      table[1,3]<-resultstestglm1[1]
      table[1,4]<-resultstestglm1[2]
      table[1,5]<-'seis'
      print(labels(resultstestglm1))

      aicglm1<-extractAIC(testglm1)
      print(aicglm1)
      table[1,2]<-aicglm1[2]

# this is GLM and hypothesis 2#
testglm2<-glm(formula = Presrand ~ 0.5, family = binomial(link =
   logit), data = mergedbeartest, na.action = na.exclude, control =
   list(epsilon = 0.0001,
   maxit = 50, trace = F))

      print('test2')
      table[2,1]<-"2"
      print(testglm2$coefficients)
      resultstestglm2<-testglm2$coefficients
      table[2,3]<-resultstestglm5[1]
      table[2,4]<-resultstestglm5[2]
      table[2,5]<-'const'
      print(labels(resultstestglm2))

      aicglm2<-extractAIC(testglm2)
      print(aicglm2)
      table[2,2]<-aicglm2[2]


# the following is an example how an additional hypothesis could be
   tested
```

```
# additional models can just be added by copy-past and then adjusted
   for the GLM to be tested

#testglm3<-glm(formula = Presrand ~ Rlowprox, family = binomial(link =
   logit), data = mergedbeartest, na.action = na.exclude, control =
   list(epsilon = 0.0001,
# maxit = 50, trace = F))

#       print('test3')
#       table[3,1]<-"3"
#       print(testglm3$coefficients)
#       resultstestglm3<-testglm3$coefficients
#       table[3,3]<-resultstestglm5[1]
#       table[3,4]<-resultstestglm5[2]
#       table[3,5]<-'rlow'
#       print(labels(resultstestglm3))

#       aicglm3<-extractAIC(testglm3)
#       print(aicglm3)
#       table[3,2]<-aicglm3[2]

   MinNumOfResult <- table[1,2]
   i <- 2
   while( i <= NumOfModel) {
        if(table[i,2] < MinNumOfResult) MinNumOfResult <- table[i,2]
        i <- i+1
   }
```

CONT'D APPENDIX

```
    i <- 1
   while( i <= NumOfModel) {
        if(table[i,2] == MinNumOfResult) result <- c(table[i,1],
   table[i,2], table[i,3], table[i,4], table[i,5])
        i <- i+1
   }


   cat(ii," ",result[1]," ",result[2]," ",result[3]," ",result[4],"
   ",result[5],
   file="bear004jj991000modelresults.txt",fill=TRUE,append=TRUE)


# kept for diagnostic control purposes
# cat(i," ",writeout1," ",writeout2," ",writeout3,
   file="bear004jj991000modelresults.txt",fill=TRUE,append=TRUE)

# kept for diagnostic control purposes
# export.data(DataSet = "writeout2",
# Columns = "ALL",
# Rows = "ALL",
# Delimiter = ",",
# ColumnNames = T,
# RowNames = T,
# Quotes = T,
# LineLength = "",
# FileName = paste("writeout2_bear004jj_", i,".txt",sep=""),
# FileType = "ASCII",
# OdbcConnection = "",
```

```
#  OdbcTable = "",
#  FormatString = "")

rm(result)
rm(table)
rm(sampleout)
rm(mergedbeartest)
rm(testglm1)
rm(testglm2)
rm(testglm3)
rm(testglm4)
rm(testglm5)

rm(aicglm1)
rm(aicglm2)
rm(aicglm3)
rm(aicglm4)
rm(aicglm5)

# kept for diagnostic control purposes
#rm(writeout1)
#rm(writeout2)
#rm(writeout3)

  }
}
```

Authors' address: F. HUETTMANN, PhD, JULIA LINKE, M.Sc. cand., Geography Dept., University of Calgary, Calgary AB T2N 1N4, Canada, email: falk@ucalgary.ca