International Journal of Earth, Energy and Environmental Sciences
ISSN: 2517-942X
Vol:1, No:1, 2007

Assessing Habitat-Suitability Models with a Virtual Species at Khao Nan National Park, Thailand

W. Srisang, K. Jaroensutasinee, and M. Jaroensutasinee

***Abstract*—**This study examined a habitat-suitability assessment method namely the Ecological Niche Factor Analysis (ENFA). A virtual species was created and then dispatched in a geographic information system model of a real landscape in three historic scenarios: (1) spreading, (2) equilibrium, and (3) overabundance. In each scenario, the virtual species was sampled and these simulated data sets were used as inputs for the ENFA to reconstruct the habitat suitability model. The 'equilibrium' scenario gives the highest quantity and quality among three scenarios. ENFA was sensitive to the distribution scenarios but not sensitive to sample sizes. The use of a virtual species proved to be a very efficient method, allowing one to fully control the quality of the input data as well as to accurately evaluate the predictive power of the analyses.

***Keywords*—**Habitat-Suitability Models, Ecological niche factor analysis, Climatic factors, Geographic information system.

I. Introduction

PREDICTION of species distribution is an important element of conservation biology. Management for endangered species [1], [2], ecosystem restoration [3], species re-introductions [4], population viability analyses [5], [6], and human–wildlife conflicts [7] often rely on habitat-suitability modeling. Multivariate models are commonly used to define habitat suitability and, combined with geographical information systems (GIS), allowing one to create potential distribution maps [8]. Numerous multivariate analyses are developed for building habitat suitability or abundance models in the past decade [9]-[12].

Ecological niche factor analysis (ENFA) [12] is a heuristic modeling approach recently developed to predict potential species distribution from presence-only data. This approach, based on Hutchinson ecological niche theory [13], creates habitat suitability maps that indirectly reveal potential species distribution [12], [14]. ENFA was originally created to predict fauna distributions that are especially susceptible to erroneous or 'false' absences due to an animal's ability to disperse or hide during field surveys. Hirzel et al. [12] suggest that inclusion of these types of 'false' absences in predictive modeling could substantially bias analysis. Therefore, ENFA is an alternative approach to model potential species distributions when there is no reliable absence data. Even though this approach does not incorporate species absence data, ENFA has potential for predicting plant species distributions from presence-only data sets.

ENFA compares the eco-geographical predictor distribution with a presence data set consisting of locations where the species had been detected with the predictor distribution of the whole area. Like the Principal Component Analysis, ENFA summarizes all predictors into a few uncorrelated factors retaining most of the information. But in this case, these factors have ecological meanings. The first factor is the 'marginality', and reflects the direction in which the species niche mostly differs from the available conditions in the global area. Subsequent factors represent the 'specialization'. They are extracted successively by computing the direction that maximizes the ratio of the variance of the global distribution to that of the species distribution. A large part of the information is accounted for by the first few factors. The species distribution base on these factors is used to compute a habitat suitability index for any set of descriptor values [12].

Modeling with ENFA is usually done by using the software *Biomapper* [12], [15], [16]. In this study, we rewrote the ENFA program with *Mathematica* [17], which is a mathematical and statistical package with visualization tools. We tested our ENFA program with virtual species data and real eco-geographical and climatic data at Khao Nan National Park, Southern Thailand.

II. Mathematical Procedures of ENFA

Eco-geographical and climatic parameters are first normalized through the Box-Cox transformation [18]. Though multinormality is theoretically needed for factor extraction through eigensystem computation [19], this method seems quite robust to deviate from normality [20]. The parameters are then standardized by retrieving means and dividing by standard deviations:

Manuscript received November 30, 2006. This work was supported in part by PTT Pubic Company Limited, the TRF/BIOTEC Special Program for Biodiversity Research and Training grant BRT T_549002, Complex System research unit, Institute of Research and Development, Walailak University, and DPST project, IPST, Thailand.

W. Srisang is with School of Science, Walailak University, 222 Thaiburi, Thasala District, Nakhon Si Thammarat, 80161, Thailand (phone: +6675672005 ; Fax: +6675672038; e-mail: wsrisang@ gmail.com).

K. Jaroensutasinee is with School of Science, Walailak University, 222 Thaiburi, Thasala District, Nakhon Si Thammarat, 80161, Thailand (phone: +66 75672005; Fax: +6675672004; e-mail: krisanadej@ gmail.com).

M. Jaroensutasinee is with School of Science, Walailak University, 222 Thaiburi, Thasala District, Nakhon Si Thammarat, 80161, Thailand (phone: +6675672005; Fax: +6675672004; e-mail: jmullica@ wu.ac.th).

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_{x_j}} \quad (1)$$

where x_{ij} is the value of the variable x_j in cell i, $\bar{x}_j$ is the mean of this variable over all cells, and σ_{xj} is its standard deviation. Let Z be the $N\times V$ matrix of standardized measurements z_{ij}. The $V\times V$ covariance matrix among standardized variables is then computed as

$$R_G = \frac{1}{N} Z^T Z \quad (2)$$

where Z^T is the transposed matrix of Z. Because of standardization (3), R_G is also a correlation matrix.

The N_S lines of Z corresponding to the N_S cells where the focal species detected are then stored in a new $N_S\times V$ matrix (say S), from which the $V\times V$ species covariance matrix is calculated:

$$R_S = \frac{1}{N_S - 1} S^T S \quad (3)$$

Let u be a normed vector of the eco-geographical and climatic spaces. The variance of the global distribution on this vector is $u^T R_G u$, while that of the species distribution is $u^T R_S u$. The first specialization factor should thus maximize the ratio $\Theta(u) = u^T R_G u / u^T R_S u$. However, this vector must also be orthogonal to the marginality factor m, given that the vector of means over V columns of S:

$$m = \left\{ \frac{1}{N_S} \sum_{i=1}^{N_S} z_{ij} \right\} \quad (4)$$

The problem, therefore, becomes that of finding the vector u that maximizes $\Theta(u)$ under the constraint $m^T u = 0$. This is equivalent to finding u, such that

$$\begin{cases} u^T R_S u = 1 \\ u^T m = 0 \\ u^T R_G u \ max \end{cases} \quad (5)$$

A change in variables allows us to rewrite the problem

$$\begin{cases} v^T v = 1 \\ v^T y = 0 \\ v^T W v \ max \end{cases} \quad (6)$$

where $v = R_s^{½} u$, $y = z \Big/ \sqrt{z^T z}$, and $z = R_s^{-½} m$, $W = R_s^{-½} R_G \ R_s^{-½}$ is a symmetric matrix. It can be shown that the solution is given by the first eigenvector of

$$H = (I_v - yy^T) W (I_v - yy^T) \quad (7)$$

where I_v is a $V\times V$ identity matrix.

Indeed,

1) y is an eigenvector of H because $Hy = (I_v - yy^T)W(I_v - yy^T) = 0$;

2) H is symmetrical and thus admits a base of orthonormed eigenvectors so that $Hv = \lambda v \Rightarrow v^T y = 0$; and,

3) $v^T Hv$ is maximum for the first eigenvector, which also maximizes $v^T Wv$ since $v^T y = 0 \Rightarrow v^T Hv = v^T (I_v - yy^T) W (I_v - yy^T) v = v^T Wv$

The V eigenvectors of H are then back transformed, and the new eigenvectors ($u = R_s^{-½} v$) are stored in a matrix U. These vectors are R_s-orthogonal (all Su distributions have variance 1 and are uncorrelated). Furthermore, due to the constraint that u be orthogonal to m, this system has one null eigenvalue. The corresponding eigenvector is thus deleted from U, and m is substituted instead as the first column. It should be noted that, although all marginality is accounted for by the first factor, this factor is not ‘‘pure,’’ in that the niche of the focal species may also display some restriction on it, in addition to its departure from the mean. ‘Marginality’ (m) and ‘specialization’ (U) factors from this computation will be used to transform species distributions from an $N\times V$ matrix to an ecological niche factor value.

The suitability map for the focal species can be computed by building on a count of all cells from the species distribution that lays as far or farther apart from the median than the focal cell on a factor axis. This count is normalized in such a way that the suitability index ranges from zero to one. Practically, this is performed by dividing the species range on each selected factor into a series of classes, in such a way that the median would exactly separate into two classes (Fig. 1). For every cell from the global distribution, we count the number of cells from the species distribution that are either in the same class or in any class farther apart from the median on the same side (Fig. 1). Normalization is achieved by dividing twice this number by the total number of cells in the species distribution. Thus, a cell laying in one of the two classes directly adjacent to the median would score one, and a cell laying outside the species distribution would score zero.

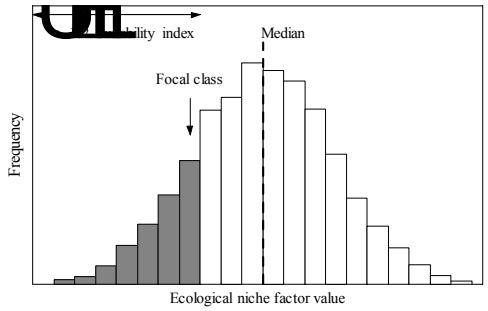


Fig. 1 the suitability of any cell from the global distribution is calculated from its situation (arrow) relative to the species distribution (histogram) on all selected niche factors. Specifically, it is calculated as twice the dashed area (the sum of all cells from the species distribution that lies as far or farther from the median dashed vertical line) divided by the total number of cells from the species distribution (surface of the histogram)

III. Methods

This study generated a virtual species completely characterized by its ecological niche, which would be modeled by a 'true' habitat suitability map. Three data sets were then generated to simulate three different scenarios. These data sets, in conjunction with environmental variables, were fed into the ENFA analysis, which produced 'predicted' habitat suitability maps. Finally, resulting models were evaluated by statistically comparing each 'predicted' map with the 'true' map. These steps (Fig. 2) will now be developed in full detail.

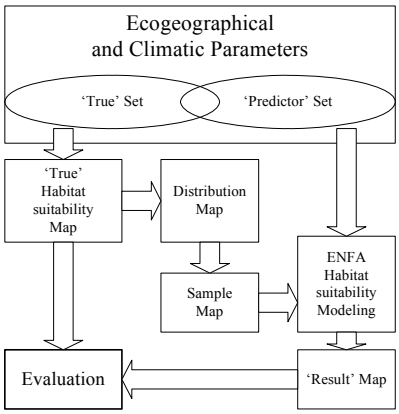


Fig. 2 Flow chart summarizing the steps involved in the study

A. Ecogeographical and Climatic Parameters

Although a virtual species is used, ecogeographical and climatic data were collected from a rectangular region of 35.1×30.1 km^2 located in Khao Nan National Park, Nakhon Si Thammarat, Southern Thailand (Fig. 3). This rectangular region is numerically modeled by GIS maps of 351×301 cells with 0.1×0.1 km^2/cell. Ecogeographical and climatic data were composed of three ecogeographical parameters and six climatic parameters. Ecogeographical parameters were computed from military maps (DEM) including elevation, slope, and aspect.

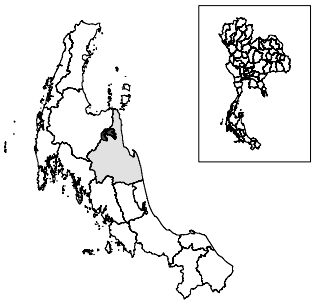


Fig. 3 Study area (35.1×30.1 km^2) at Khao Nan National Park (dark gray area), Nakhon Si Thammarat province, Thailand

We collected climatic data from eight study sites around Khao Nan National Park from April-May 2006 and uploaded this data to the Thailand National Biodiversity Database System (NBIDS) [21]. These climatic parameters were composed of solar noon temperature, maximum temperature, minimum temperature, relative humidity, %cloud cover, and amount of rainfall. The climatic data were then interpolated using invert distance weighted methods to generate our study area.

B. Virtual Ecological Niche: The 'True' Habitat Suitability Map

On this spatial canvas, the virtual species was generated by creating a simulated ecological niche in an n-dimensional space [13]. It was modeled by a niche coefficient H ($H \in [0,1]$), which can be viewed as the probability that each cell belongs to the niche; note that H is de facto a habitat suitability index. This value was built as summarized in (8).

$$H = \frac{1}{\sum w_i} \sum w_i H_i + \varepsilon \qquad (8)$$

where H is the habitat suitability of the focal cell, H_i is the value of the i^{th} partial niche coefficient, w_i is the weight assigned to the i^{th} partial niche coefficient, and ε is a random value.

Global habitat suitability is composed of a weighted average of partial niche coefficients (H_i) and a stochastic coefficient (ε). The partial niche coefficients are the habitat suitability engendered by each predictor value. They are computed from four predictors that are picked out of the nine available predictors by four niche functions (i.e. elevation with Gaussian function, aspect with Gaussian function, the amount of rainfall with truncated linear, and minimum air temperature with decreasing linear function). Three types of functions are used to model three types of environmental optimum: 1) a Gaussian function to model a median optimum, 2) a linear function to model an extreme optimum, and 3) a truncated linear function to model a buffer zone effect. Each of these H_i values is then weighted by a w_i factor and the global niche coefficient is calculated as their weighted average. Finally, a random term ε, generated from a uniform distribution in the range [−0.05, 0.05], is added. The niche-function parameters and the weights are arbitrarily tuned in order to generate about 50% of cells with $H \geq 0.5$.

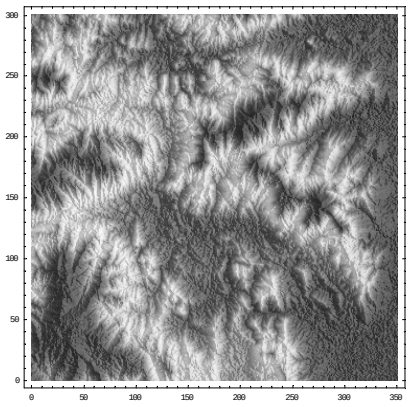


Fig. 4 The 'true' habitat suitability map generated to model the ecological niche of the virtual species. High suitability areas are indicated by white pixels

This produces the 'true' habitat suitability map (Fig. 4), representing the 'real' intrinsic preferences of our virtual species. By 'true' map, we mean that it represents the kind of information usually unreachable by ecologists, the information they are trying to reveal through field sampling and statistical analysis. The 'true' map will be constantly used as a basis to generate data and as a reference to assess the accuracy of habitat suitability analyses.

C. Distribution Map

Distribution maps are computed on the basis of the 'true' map, the distribution maps give the 'true' presence/absence of the virtual species, information usually unavailable to field ecologists. Three distribution scenarios are addressed in order to determine the advantages and drawbacks of each habitat suitability analysis. They can be viewed as three historical phases of colonization—the fundamental niche does not change but the realized one does: 1) a 'spreading phase' showing a density gradient from the north-west corner of the map to the south-east corner, 2) an 'equilibrium phase' where the species are abundant enough to occupy all the available suitable areas, and 3) an 'overabundance phase' where the species are so numerous that it has to spread in less suitable areas (Fig. 5).

The 'equilibrium' distribution map is computed as follows. To each cell of the 'true' habitat suitability map is added a random value taken in the range [−0.2, 0.2] (uniform distribution); this is made in order to introduce some stochasticity into the model. If the resulting habitat suitability coefficient is larger than 0.7, the cell is marked as occupied.

The 'overabundance' distribution map is computed in a similar way but with a 0.5 habitat suitability threshold to simulate the overflowing density.

The 'spreading' distribution needs an additional operation: each cell of the 'true' habitat suitability map is beforehand multiplied by a value decreasing in $1/d^2$, d being the distance to a point arbitrarily placed north-westward to south-eastward corner of the map. This gradient function is tuned to produce values ranging from 0 to 1, 0.5 lying approximately in the middle of the map. This new gradient map is then submitted to the same operations as the 'equilibrium' scenario (habitat suitability threshold = 0.7).

This generating method allow us obtain distribution maps with a presence density correlated with area suitability.

D. Sample Maps

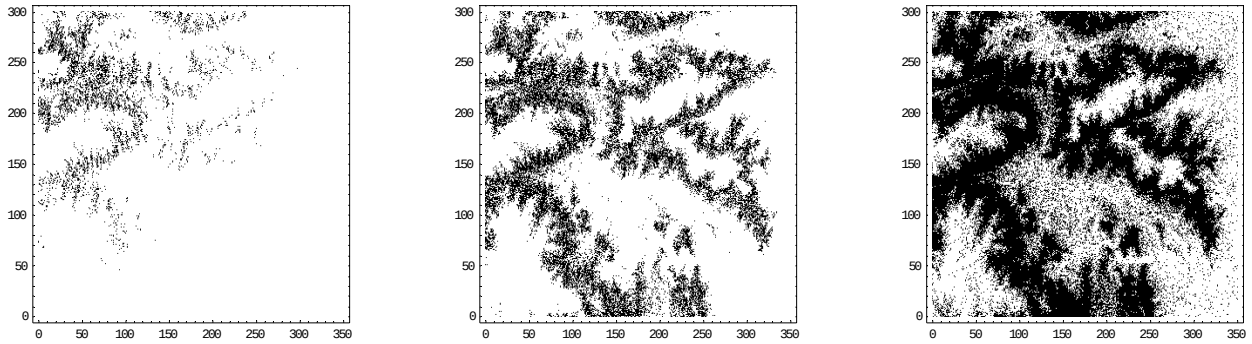
These distribution maps are then used to simulate 'field' sampling data usually resulting from the trapping, detecting or observation activities of field biologists. Since ENFA needs only presence data, ENFA data sets are generated by randomly sampling points in the distribution maps in order to obtain the targeted sample size.

E. Result habitat suitability maps

The simulated data sets are then submitted as dependent variables to the ENFA. The independent variables are a set of five predictors out of the nine available; four of them are arbitrarily taken among those used to generate the 'true' habitat suitability map and other is new. In order to evaluate quality and quantity of the three scenarios, sampling sizes of 250 to 5000 points are simulated ten times for all scenarios and analyses, except for the spreading scenario in which only 250 to 4000 points are simulated. This is because the spreading scenario has only 4000 occupied cells and therefore, it is impossible to get more than 4000 points sample size in this case

F. Evaluation

The accuracy of the 'result' habitat suitability maps has finally to be assessed. With a real species, we would have used independent evaluation data and calculated various statistics to assess the accuracy of the classification [22]. But here, with a virtual species, the 'true' habitat suitability that the models are supposed to reproduce is perfectly known. More adapted statistics based on the Pearson correlation coefficient between the two maps could thus be used. In order to get round the pseudo-replication, engendered by spatial



(a) (b) (c)

Fig. 5 Distribution maps of the virtual species for three colonization scenarios. Black points are the cells where the species is present and the white ones are those where the species is absent. Map (a) represents the 'spreading' scenario: the species entered the area from the northwest and is currently propagating in all directions, settling down in the most suitable areas. Map (b) represents the 'equilibrium' scenario in which the species occupies uniformly all the suitable areas. Map (c) represents the 'overabundance' scenario in which very high densities force the species to occupy less adequate areas

auto-correlation between cells, we proceed as follows: 1000 cells are picked randomly and a determination coefficient R^2 (proportion of variance explained by the model) is computed between the values of these cells in the 'result' map and the 'true' map [23]. Each scenario is replicated ten times and the mean and standard deviation of R^2 are computed. The mean R^2 ($<R^2>$) is used to assess the accuracy of the models. Their sensitivities to distribution scenario and sample size are also assessed with a Student *t*-test for each method. In order to compare results, two sampling sizes are addressed: 500 and 2000 points for all scenarios and analyses.

IV. Results and Discussion

$<R^2>$ ± S.D. of 'spreading', 'equilibrium', and 'overabundance' of 500 points scenarios are similar to 2000 points scenarios (Table I). This result confirms the previous work of Hirzel et al. [24]. They showed that $<R^2>$ of 'spreading', 'equilibrium', and 'overabundance' of 300 and 1200 points scenarios were ranging from 0.55-0.60 with the SD of 0.03-0.05. Due to the stochasticity added in the process of building the 'true' habitat suitability map, it is impossible to obtain $R^2 = 1$. The best model would have been the map computed just before the addition of stochasticity and this one gives the maximum $R^2 = 0.59$. The 'equilibrium' scenario had the highest $<R^2>$ among three scenarios; there were no differences among 16 sample sizes and no interaction term between scenario and sample size (Two-way ANOVA: scenarios: $F_{2, 432} = 133.262$, $P<0.001$; sample size: $F_{15, 432} = 1.469$, NS; scenario x sample size interaction: $F_{30, 432} = 1.337$, NS, Fig. 6). The 'equilibrium' scenario had the fewest S.D. among three scenarios (One-way ANOVA: $F_{2, 45} = 15.780$, $P<0.001$) and post-Hoc tests showed that the S.D. of the 'equilibrium' scenario was lower than both the 'spreading' scenario ($P<0.01$, Fig. 6) and the 'overabundance' scenario ($P<0.001$, Fig. 6). Our results suggest that the 'equilibrium' scenario had the highest quantity and quality among the three scenarios. Our results also indicate that ENFA was sensitive to the distribution scenarios but not sensitive to sample sizes (Table II, Fig. 6). The reason that the 'overabundance' scenario was not the best scenario was because most individuals in this species tend to occupy a low suitability area. This might lead to a low R^2. The 'Spreading' scenario was not the best scenario because if a species starts spreading from an area that has other uncorrelated parameters, then the 'spreading' scenario may interpret these uncorrelated parameters as essential parameters. Therefore, this also could cause a low R^2.

Our results suggest that 1) ENFA will be most suitable to model plants or native species that occur in an area for

TABLE I
$<R^2>$ AND (S.D.) OF THE PROPORTION OF EXPLAINED VARIANCE OBTAINED BY COMPARING 'RESULT' MAP WITH 'TRUE' MAP

Scenarios	$<R^2>$	S.D.
Spreading, 500 points	0.47	0.04
Equilibrium, 500 points	0.53	0.03
Overabundance, 500 points	0.46	0.05
Spreading, 2000 points	0.47	0.04
Equilibrium, 2000 points	0.55	0.02
Overabundance, 2000 points	0.49	0.03

The greater the value of $<R^2>$ shows the higher the predictive power of the 'result' map.

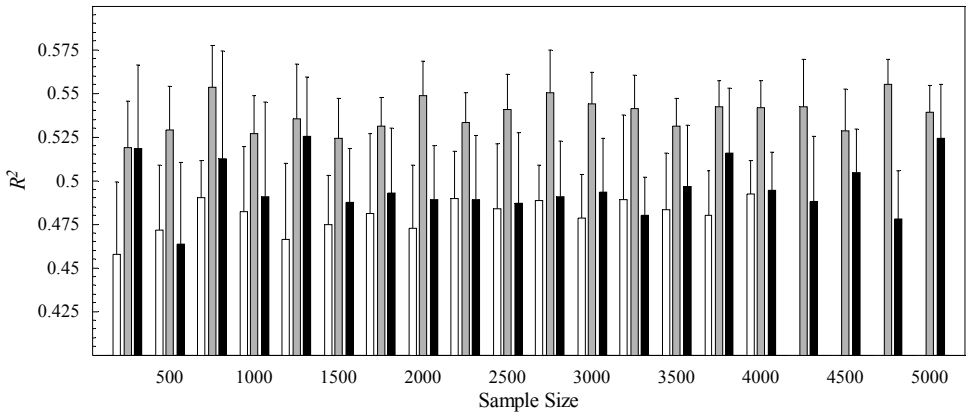


Fig. 6 $<R^2>$ ± S.D. of three scenarios 'spreading' (white), 'equilibrium' (gray), and 'overabundance' (black)

TABLE II
SENSITIVITY ANALYSES OF THE METHODS

	Spreading, 500 points	Equilibrium, 500 points	Overabundance, 500 points	Spreading, 2000 points	Equilibrium, 2000 points
Equilibrium, 500 points	*				
Overabundance, 500 points	NS(0.67)	*			
Spreading, 2000 points	NS(0.95)	*	NS(0.62)		
Equilibrium, 2000 points	**	NS(0.07)	**	**	
Overabundance, 2000 points	NS(0.26)	*	NS(0.16)	NS(0.28)	**

Significant results: NS non Significant, *P<0.01, ** $P<0.001$

sometime and might reach their equilibrium stage, 2) ENFA might not best model invasive species and 3) since ENFA not affect by the number of sample sizes in all three scenarios; therefore, ENFA may be suitable to model habitat-suitability for rare species.

The virtual species approach proved to be the most serviceable. When comparing models on a real data basis, it is only possible to make assumptions about what is the true habitat suitability by using various expert and statistical evaluation methods. Many factors are out of reach and may introduce a bias that cannot be accurately assessed [8], [10], [24]-[25]. Therefore, real data are only a snapshot of a dynamical situation and can only give a partial and instantaneous comprehension of the fundamental ecological niche. By generating a virtual species, the 'true' is now completely reachable and resulting models can be accurately compared to it.

ACKNOWLEDGMENTS

We thank Asst. Prof. Dr. David Harding, Mr. Anun Charoensuk, Mr. Pattanabhorn Rintachak, and Khao Nan National Park rangers for kindly help. This work was supported in part by PTT Pubic Company Limited, the TRF/BIOTEC Special Program for Biodiversity Research and Training grant BRT T_549002, Complex System research unit, the Institute of Research and Development, Walailak University, and DPST project, IPST, Thailand.

REFERENCES

[1] L. Palma, P. Beja, and M. Rodgrigues, "The use of sighting data to analyse Iberian lynx habitat and distribution," *J. Appl. Ecol.*, vol.36, pp. 812–824, 1999.

[2] J. A. Sanchez-Zapata, J. F. Calvo, "Raptor distribution in relation to landscape composition in semi-arid Mediterranean habitats," *J. Appl. Ecol.*, vol. 36, pp. 254–262, 1999.

[3] D. J. Mladenoff, R. C. Haight, T. A. Sickley, and A. P. Wydeven, "Causes and implications of species restoration in altered ecosystems. A spatial landscape projection of wolf population recovery," *Bioscience*, vol. 47, pp. 21–23, 1997.

[4] U. Breitenmoser, F. Zimmermann, P. Olsson, A. Ryser, C. Angst, A. Jobin, and C. Breitenmoser-Würsten, Beurteilung des Kantons St. Gallen als Habitat für den Luchs, KORA, Bern, 1999.

[5] H. R. Akçakaya, and J. L. Atwood, "A habitat-based metapopulation model of the California Gnatcatcher," *Conservat. Biol.*, vol. 11, pp. 422–434, 1997.

[6] H. R. Akçakaya, M. A. McCarthy, and J. L. Pearce, "Linking landscape data with population viability analysis: management options for the helmeted honeyeater *Lichenostomus melanops cassidix*," *Biol. Conservat.*, vol. 73, pp. 169–176, 1995.

[7] G. Le Lay, P. Clergeau and L. Hubert-Moy, "Computerized map of risk to manage wildlife species in urban areas," *Environ. Manage.*, vol. 27, pp. 451–461, 2001

[8] A. Guisan and N. E. Zimmermann, "Predictive habitat distribution models in ecology," *Ecol. Model.*, vol. 135, pp. 147–186, 2000.

[9] S. Lek, M. Delacoste, P. Baran, I. Dimopoulos, J. Lauga, and S. Aulagnier, "Application of neural networks to modelling nonlinear relationships in ecology," *Ecol. Model.*, vol. 90, pp. 39–52, 1996.

[10] S. Manel, J. M. Dias, S. T. Buckton, and S. J. Ormerod, "Alternative methods for predicting species distribution: an illustration with Himalayan river birds," *J. Appl. Ecol.*, vol. 36, pp. 734–747, 1999.

[11] S. L. Özesmi and U. Özesmi, "An artificial neural network approach to spatial habitat modelling with interspecific interaction." *Ecol. Model.*, vol. 116, pp. 15–31, 1999.

[12] A. H. Hirzel, J. Hausser, D. Chessel, and N. Perrin, "Ecological-niche factor analysis: how to compute habitat suitability maps without absence data?" *Ecology*, vol. 83(7), pp. 2027–2036, 2002.

[13] G. E. Hutchinson, Concluding remarks, Cold Spring Harbor Symposium, *Quantitative Biol*., vol. 22, pp. 415-427, 1957.

[14] J. Hausser, Mammifères de la Suisse: Répartition * Biologie * Ecologie. Commission des Mémoires de l'Académie Suisse des Sciences Naturelles, Birkhäuser Verlag, Basel, 1995.

[15] R. M. Chefaoui, J. Hortal, and J. M. Lobo, "Potential distribution modelling, niche characterization and conservation status assessment using GIS tools: a case study of Iberian *Copris* species," *Biol. Conservat.*, vol. 122, pp. 327–338, 2005.

[16] A. H. Hirzel, J. Hausser, and N. Perrin, (2004) Biomapper 3.1, Lab. of Conservation Biology, Department of Ecology and Evolution, University of Lausanne. Available: http://www.unil.ch/biomapper

[17] S. Wolfram, *The Mathematica Book*, 5th ed., Wolfram Media, 2003.

[18] R. R. Sokal, and F. J. Rohlf, *Biometry: The Principles and Practice of Statistics in Biological Research*. New York: W. H. Freeman, 1981.

[19] L. Legendre and P. Legendre, *Numerical Ecology*, 2nd English ed., Amsterdam, The Netherlands: Elsevier Science BV, 1998.

[20] G. V. Glass and K. D. Hopkins, *Statistical methods in education and psychology*, 2nd ed., London, UK: Prentice Hall, 1984.

[21] Thailand National Biodiversity Database System. (2006). Available: http://www.nbids.org

[22] A. H. Fielding and J. F. Bell, "A review of methods for the assessment of prediction errors in conservation presence/absence models," *Environ. Conservat.*, vol. 24, pp. 38–49, 1997.

[23] F. Mesplé, M. Troussellier, C. Casellas, and P. Legendre, "Evaluation of simple statistical criteria to qualify a simulation," *Ecol. Model*., vol. 88, pp. 9–18, 1996.

[24] A. H. Hirzel, V. Helfer, and F. Metral, "Assessing habitat-suitability models with a virtual species," *Ecol. Model.*, vol. 145, pp. 111–121, 2001.

[25] J. R. Alldredge, and J. T. Ratti, "Further comparison of some statistical techniques for analysis of resource selection," *J. Wildl. Manag*., vol. 56, pp. 1–9, 1992.