Identifying Biodiversity Hotspots by Predictive Models: A Case Study Using Taiwan's Endemic Bird Species

Chia-Ying Ko[1], Ruey-Shing Lin[2], Tzung-Su Ding[3], Chih-Hao Hsieh[4], and Pei-Fen Lee[1,5,]*

[1]*Institute of Ecology and Evolutionary Biology, National Taiwan University, Taipei 106, Taiwan*
[2]*Endemic Species Research Institute, 1 Ming-Shen East Road, Jiji, Nantou 552, Taiwan*
[3]*School of Forestry and Resource Conservation, National Taiwan University, Taipei 106, Taiwan*
[4]*Institute of Oceanography, National Taiwan University, Taipei 106, Taiwan*
[5]*Department of Life Science, National Taiwan University, Taipei 106, Taiwan*

Chia-Ying Ko, Ruey-Shing Lin, Tzung-Su Ding, Chih-Hao Hsieh, and Pei-Fen Lee (2009) Identifying biodiversity hotspots by predictive models: a case study using Taiwan's endemic bird species. *Zoological Studies* **48**(3): 418-431. Predicting species distributions and identifying biodiversity hotspots are essential in designing conservation strategies. Because of different spatial scales and/or species characteristics, uncertainty still exist as to which model is the best. Several models have been proposed to calculate the probability of species occurrences, predict biodiversity hotspots, and decide importance levels of those hotspots. We constructed predictive distribution models for 14 of 16 endemic bird species in Taiwan using a fine-resolution (1 × 1 km) breeding bird distribution dataset compiled over the past decade as well as environmental variables. We compared the performances of the 4 models: logistic regression (LR), multiple discriminant analysis (MDA), genetic algorithm for rule-set prediction (GARP), and artificial neural network (ANN). Maps for biodiversity hotspots were generated based on the species distributions from the 4 models. To account for potential uncertainty, we constructed hotspot maps using a frequency histogram and probability density function approaches. Based on the distribution maps and the area under the curve (AUC) of the receiver operating characteristic, all of our models made good predictions for each species (all AUC values were > 0.75). The non-linear models (GARP, ANN, and LR) provided better predictions than did the linear (MDA) model. GARP was the most consistent model when evaluated by it kappa, sensitivity, accuracy, and specificity values for each species and the 3 species categories (common, uncommon, and rare species). The prevalence of all species did not affect the final predictive performance. The 5 biodiversity hotspot maps derived from the frequency histogram approach showed a relatively similar pattern to maps generated by the probability density function, which indicated that of mid- to high-elevation areas had higher probabilities. In spite of some inconsistencies, the hotspot maps identified from these 2 approaches were fairly representative when evaluated against currently known hotspots. A GAP analysis indicated only 25% of the hotspots are currently protected by national parks. We concluded that the LR, GARP, ANN, and MDA models are all feasible to use for modeling bird species distributions. Although there were some limitations, we suggest using a combination approach to identify common features and conservation priorities of biodiversity hotspots. Comparing known and predicted hotspots can promote the reliability of the models as well as provide managers with greater confidence when planning conservation policies. Finally, this approach to identifying common features and conservation priorities of biodiversity hotspots can be applied to evaluate conservation efforts and provide a better tool to achieve efficient conservation. http://zoolstud.sinica.edu.tw/Journals/48.3/418.pdf

Key words: Conservation, Biogeography, Endemism, Model, Biodiversity.

U nderstanding species distributions and the environmental factors affecting those distributions is of fundamental importance in reserve planning (Corsi et al. 1999, Elith et al. 2006) and biodiversity

*To whom correspondence and reprint requests should be addressed. Tel: 886-2-33662469. Fax: 886-2-23623750.
E-mail:leepf@ntu.edu.tw

conservation (Gray et al. 2006, Lira-Noriega et al. 2007). To gain information on species distributions, many predictive models based on correlating presence data of species with environmental predictors are widely used (Manel et al. 1999a).

Research on species distribution predictive techniques has yet to produce a perfect modeling system (Guisan and Zimmermann 2000, Hirzel and Guisan 2002). Traditionally, models used in ecology to predict potential species distributions were multivariate in nature and based on linear functions such as multiple/linear regression, and multiple discriminant analysis (MDA; Austin and Meyers 1996, Jose and Fernando 1997). However, those methods led to some statistical and theoretical concerns, and methods which can model non-linear relationships were developed over time. Those methods include logistic regression (LR), genetic algorithm for rule-set prediction (GARP), ecological-niche factor analysis, maximum entropy, and artificial neural networks (ANN; Lek et al. 1996, Jose and Fernando 1997, Aitkenhead et al. 2004, Stockman et al. 2006, Phillips et al. 2006). These models give ecologists better tools to accurately predict species distributions (Stockwell 2007). However, increasing difficulties over which models to choose have also become a challenge. Although several studies compared model accuracies in predicting species distributions (Manel et al. 1999b, Moisen and Frescino 2002, Robertson et al. 2003), uncertainty still exists as to which model is the best to select under different spatial scales and/or species characteristics (Zaniewski et al. 2002).

Much attention has been focused on biodiversity hotspots (Myers et al. 2000, Kati et al. 2004). Sound biodiversity hotspots analyses and effective management of such hotspots are helpful for protecting endangered and endemic species (Myers et al. 2000, Bonn et al. 2002). Unfortunately, biodiversity hotspots tend to occur in areas with significant human impacts (Myers et al. 2000), leading to increase pressures on rare, threatened, and endemic species. At large geographic scales, a hotspot map can provide useful information for conservation planning (Reid 1998).

The development of species distribution prediction techniques and approaches has benefited from the application of biodiversity mapping in recent years (Rodríguez et al. 2007). Through collecting information from the field, modeling their distributions using various tools and

overlaying potential distributions with protected areas, ecologists developed the gap analysis program (GAP) for conserving biodiversity (Scott et al. 1993). However, using a single model to predict biodiversity hotspots might bear some uncertainties. Protecting all biodiversity hotspots in a situation of budget restrictions is almost impossible. Therefore, using several models to predict biodiversity hotspots and deciding the importance levels of those hotspots is a reliable approach (Araújo and New 2007). By analyzing the importance levels of biodiversity hotspots to determine befitting conservation priorities and conservation efficiency can be improved by paying more attention to higher-level hotspots. Analyzing the geographical and environmental constraints of higher-level biodiversity hotspots can also help understand the ecological characteristics of these hotspots.

The area of Taiwan is only about 36,000 km^2, but it contains diverse ecosystems and extensive forest cover. Due to high human density and intense economic development over the past several decades, conservation in Taiwan is challenging. Researchers have intensely investigated the distribution of animals in the past decade and have established a rich database of bird distributions (e.g., Lee et al. 2004). These long-term data are useful for predicting species distributions and comparing different predictive models. It is essential to determine the locations of biodiversity hotspots in Taiwan so that appropriate measures of ecosystem management can be proposed.

The purpose of this study was to compare outputs of 4 predictive models and combine species distribution models to locate biodiversity hotspots. Using fine-resolution (1 × 1 km) endemic bird species data from Taiwan, we compared the results of 4 widely used species distribution models: LR, MDA, GARP, and ANN. We also examined alternative techniques of ensemble forecasting, i.e., the frequency histogram and probability density function approaches, to derive biodiversity hotspots. Finally, we evaluated the current protection status using the hotspot distribution map. Since the birds we examined are mostly forest-dwelling by nature, our hotspot map focused on forest bird species biodiversity in Taiwan.

Study area

The island of Taiwan lies at latitudes 22°-25°

18'N and longitudes 120°27'E-122°E (Fig. 1), and is located between the Taiwan Strait to the west and the Pacific Ocean to the east. Mountainous areas comprise almost 65% of the island and the Central Mountain Range passes north-south through the main axis of the island. In the western portion are mostly plains and hills. The climate ranges from tropical in the south to subtropical in the north and alpine in the high mountains (to 3952 m), with strong influences of alternating monsoons (Lee et al. 2004). The weather is generally hot and moist, with a mean annual temperature of 18.0°C and an average annual total precipitation of 2510 mm. Temperature decreases with increasing elevation, with a lapse rate of -5.43°C/km (Su 1984). Because of the highly variable topography, a great diversity of forest habitats can be found in Taiwan (Lee et al. 2004, Koh et al. 2006).

Taiwan has long been under high development pressure. A very high population density (of around 635 individuals/km² in 2007) and construction of roads have resulted in forest fragmentation and destruction that have threatened wildlife. Although we have no record of resident

bird species becoming extinct, many bird species have limited distributions, and some areas have shown clear decreases in species richness (Lee et al. 2004).

MATERIALS AND METHODS

Bird data

We compiled bird distribution data from 2 bird inventory projects conducted in 1999-2003 (Koh et al. 2006) and 1993-2004 (Hsu et al. 2004). We used point counts to conduct field sampling. Sampling sites were chosen based on 4 geo-information system (GIS) layers, i.e., ecoregion, vegetation type, road distribution, and elevation range. Sites were selected to best represent the habitat characteristics of a particular elevation and ecoregion type. Since many breeding birds in Taiwan live in forests, whenever possible, we chose forested areas to obtain potential occurrence records of all species. Each site was sampled, with sufficient duration (Shiu and Lee 2003), once a year or seasonally during the survey period.

We transformed the sampling sites to a 1 × 1 km grid system. All grids with a species richness of ≥ 2 were put into the models to ensure that the absence data of species were credible. In total, 4082 of 37,552 cells had either the complete presence or absence of bird data (Fig. 1).

Currently there are 16 endemic bird species in Taiwan (Table 1). We excluded Styan's Bulbul (*Pycnonotus taivanus*) and another relatively new addition, the Taiwanese Hwamei (*Garrulax taewanus*) (Dickinson 2003, Li et al. 2006), from the analyses due to possible confusions with the Chinese Bulbul (*P. sinensis*) and the introduced Hwamei (*G. canorus*), respectively. The prevalence of each species ranged from 1% to 18% (Table 1). Three distribution categories, based on the general consensus and prevalence values, were recognized: common (> 4%), uncommon (3%-4%), and rare species (< 3%).

Environmental variables

We used over 80 environmental variables to model species distributions. However, some variables did not provide a significant contribution to explaining the species distribution relationships. 'Important' variables were classified into 4 types: topography/geography, climate, vegetation/land use, and human activities (Table 2). Most of
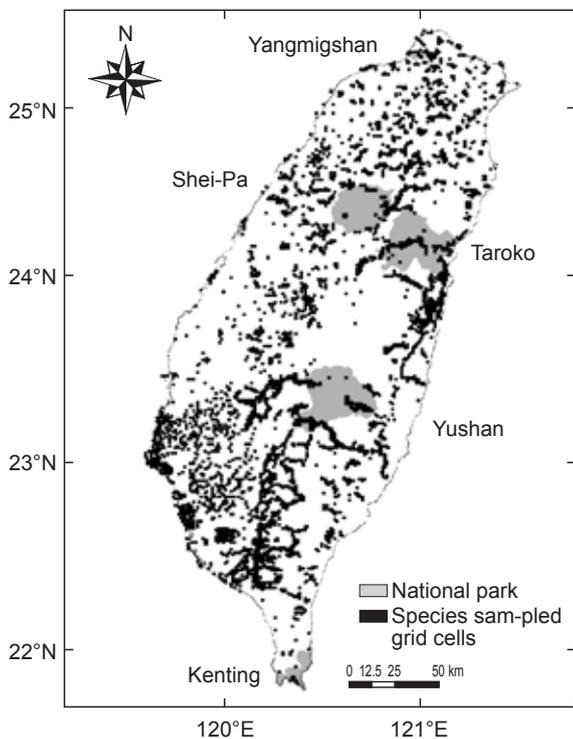
Fig. 1. Location of Taiwan showing the 5 national park boundaries and the 4082 1 × 1 km grid cells used for bird sampling.

these variables were described by Lee et al. (2004) and Koh et al. (2006). We generated elevation, elevation range, and ridgeness layers based on a digital elevation model (DEM, at a 40 m resolution) created by the Aerial Survey Office, Forestry Bureau, in Taiwan. Ridgeness is a variable designed to capture the complexity of the topography. We calculated ridgeness by applying the flow direction and accumulation functions in ArcGIS (Li et al. 2005). Climate variables included annual mean temperature, total monthly precipitation, and other related variables for the period of 1959-1985. Vegetation-related layers included a land-cover map generated by a supervised classification of SPOT mosaic images taken in the latter half of 2002 and provided by the Aerial Survey Office, Forestry Bureau. We also calculated a normalized difference vegetation index (NDVI) using these images. The forest cover type was extracted and used to calculate forest density (in percent (%), as the ratio of forest area to grid area) and number of forest patches within the grid. We used road density, percentage of built-up areas, and distances to the nearest city and road to represent the dimension of human activities. All of these indices of human activities were published by the Ministry of the Interior, Taipei, Taiwan in 2000. Road density was calculated as the ratio

Table 1. List of 16 endemic bird species in Taiwan. The prevalence was calculated based on 4082 1 × 1 km bird sampled grid cells

Family	Common English Name	Scientific Name*	Category	Prevalence (%)
Phasianidae	Taiwan Hill Partridge	*Arborophila crudigularis*	Common	10
	Swinhoe's Pheasant	*Lophura swinhoii*	Rare	1
	Mikado Pheasant	*Syrmaticus mikado*	Rare	2
Corvidae	Taiwan Blue Magpie	*Urocissa caerulea*	Uncommon	4
Paridae	Yellow Tit	*Parus holsti*	Uncommon	4
Pycnonotidae	Styan's Bulbul	*Pycnonotus taivanus*	Common	10
Sylviidae	Taiwan Bush Warbler	*Bradypterus alishanensis*	Uncommon	3
Timmallidae	Taiwanese Hwamei	*Garrulax taewanus*	Common	5
	White-whiskered Laughing-thrush	*Garrulax morrisonianus*	Common	5
	Steere's Liocichla	*Liocichla steerii*	Common	16
	Taiwan Barwing	*Actinodura morrisoniana*	Uncommon	3
	White-eared Sibia	*Heterophasia auricularis*	Common	18
	Taiwan Yuhina	*Yuhina brunneiceps*	Common	18
Reguliidae	Flamecrest	*Regulus goodfellowi*	Uncommon	3
Turdidae	Taiwan Whistling Thrush	*Myophonus insularis*	Common	11
Muscicapidae	Collared Bush Robin	*Luscinia johnstoniae*	Common	5

*Nomenclature and English names follow Dickinson (2003) and Li et al. (2006).

Table 2. List of environmental variables used in this study

Category	Environmental variables
Topography/geography	Slope, elevation, distance to an area above 3000 m, distance to a river, distance to the coastline, river density, and ridgeness.
Climate	Annual mean temperature, monthly mean temperature for 12 mo, warmth index, total annual precipitation, total precipitation for each month, total precipitation in the dry period (Oct.-Mar.), ratio of dry-period to total annual precipitation, and number of months short of precipitation.
Vegetation and land use	Forest density, number of forest patches, mean normalized difference vegetation index, and naturalness index
Human activities	Distance to the nearest road, distance to the nearest city, urban development index, population density, road density, and percentage of built-up areas

of the total length (m) of all levels of paved roads to the total area of the grid (ha). The percentage of built-up areas was calculated as the ratio of the total built-up area to the total area of the grid. We used ERDAS Imagine 8.7 and ArcGIS 9.2 to derive and analyze these variables.

Modeling methods

We applied 4 commonly used models to predict species distributions: the MDA, LR, ANNs, and GARP models. The 4 models differ in (1) the data used (GARP uses only presence records in contrast to the others which use both presence and absence records); (2) underlying function (MDA uses a linear function while the others apply non-linear functions), and (3) output characteristics (the MDA has outputs of either 0 or 1 while the others have outputs ranging from 0 to 1).

We followed standard procedures to derive the species distribution maps (Guisan and Zimmermann 2000, Moisen and Frescino 2002, Elith et al. 2006). We first analyzed the relationship between the occurrence of each species and each environmental variable and retained variables with significant correlations ($p < 0.01$) in the prediction models. After this elimination, there were approximately 10-20 environmental variables of different types for each species included in each model. No quadratic terms were used.

We used SAS 9.0 software to perform the MDA and LR calculations and backward elimination to select variables in the final model. Significant variables at each step had to significantly reduce the scaled deviance. The change in the scaled deviance was approximately distributed like χ^2 (McCullagh and Nelder 1989, Collett 1991). Although all explanatory variables were potential predictors, only those selected by the aforementioned criterion were used in the final interpretation. For the GARP calculations, we used a desktop implementation of GARP (Scachetti-Pereira 2001). We combined 4 types of rules to build the final model: atomic, logistic regression, bioclimatic envelope, and negated bioclimatic envelope rules. We set GARP to perform 100 runs per species with a maximum of 1000 iterations and a convergence limit of 0.01. With these settings, GARP produced 100 models per species in which all grid cells were predicted to be either present (1) or absent (0). We then summed these model results to derive a probability map for each species. We used MATLAB 7.0 to perform the ANN and determined the number of hidden layers

and neurons in the hidden layers through a series of iterations and model performance assessments (Manel et al. 1999b). A network had 1 hidden layer with 5 neurons being built for each species.

Model evaluation

The LR, ANN, and GARP models require an arbitrary threshold probability to determine a species' presence from the model prediction. Selections of the threshold involve evaluating omission and commission errors. Many studies (e.g., Fielding and Bell 1997, Liu et al. 2005) found that a species' prevalence may interact with the threshold, and it was suggested that thresholds be selected and model performance assessed.

To choose objective thresholds for each species' model, we first estimated the area under the curve (AUC) of the receiver operating characteristic (ROC) plots to ascertain a model's efficiency (DeLeo 1993). ROC curves were created by plotting the sensitivity (i.e., the true-positive rate) of a model on the ordinate against the value of (1 - specificity) (i.e., the false-positive rate) on the abscissa. This curve used every possible threshold value that could be chosen to predict the probability of the presence of a particular species. The AUC is independent of the prevalence of each species. A value of 1 represents a perfect discrimination between sites where a species is present versus those where it is absent, whereas a value of 0.5 indicates no significant difference between the 2 states (DeLeo 1993). We chose the threshold through comparing an original species' prevalence with the value of the least distance to the upper-left corner of the ROC curve. If the value of the ROC curve was greater than the species prevalence, we used the value of the ROC curve as the threshold, otherwise we used the prevalence. After threshold determination, we calculated the kappa value, sensitivity, specificity, and accuracy. Finally, we used kappa and sensitivity tests as a key to determine the best final model for each species based on the assumption that correct species presence locations are more crucial than absence locations.

Biodiversity hotspot analysis

After deciding the threshold, the distribution probability for each species was differentiated between 1 (presence) and 0 (absence). To derive the hotspot maps with the frequency histogram and probability density function approaches, we

used ensemble forecasting proposed by Araújo and New (2007). We overlaid 14 potential species' maps of each modeling method and defined the richest 5% of grids as hotspots of endemic bird species. To derive the frequency histogram hotspot maps, we first derived the hotspot map for each predictive model, combined the 4 hotspot maps, and recognized 5 hotspot classes (from 0 to 4 with 0 indicating no models predicted the grid as a hotspot; 1 indicating 1 model predicted the grid as a hotspot, and so forth). To obtain the probability density function hotspot map, we calculated the probability of each grid by averaging the predictive values of the 4 models. Known hotspots were extracted using the richest 5% of the survey grids. We evaluated the power of these ensemble forecasting approaches by comparing them with known hotspots. Finally, we evaluated the protection status of endemic bird hotspots by overlaying the boundaries of 5 national parks (Fig. 1).

RESULTS

Comparisons of models

The 3 models (ANN, GARP, and LR) for each species all had AUC values of > 0.75, and the accuracies predicted by MDA were all > 72%, indicating good discrimination. Examples of the predicted distributions of the Taiwan Yuhina (*Yuhina brunneiceps*), the Flamecrest (*Regulus goodfellowi*), and the Mikado Pheasant (*Syrmaticus mikado*), respectively representing common, uncommon, and rare species categories, using ANN, GARP, LR, and MDA were satisfactory compared to the actual data (Fig. 2). The GARP and MDA models were more optimistic than the LR and ANN models.

Each model contributed to the prediction when evaluated using the kappa, sensitivity, accuracy, and specificity values of each species (Fig. 3). The MDA model had lower assessment values among the 4 measures, except for sensitivity. The LR and ANN models had similar trends for kappa, accuracy, and specificity. Measures by GARP were not the highest, but they remained fairly consistent. The only exception was the case of the Taiwan Whistling Thrush (*Myiophoneus insularis*) which most models poorly predicted. Sensitivity and specificity showed higher variability among the 4 models than did the kappa and accuracy values ($p < 0.05$). In terms of

the magnitude, the kappa and sensitivity measures highly varied (kappa, 0.15-0.59; sensitivity, 0.10-0.97), whereas accuracy and specificity (all values > 0.5) were more stable. Overall, the non-linear models (GARP, LR and ANN) performed better than the linear model (MDA).

Across the 3 species categories, GARP showed consistently higher values for the 4 measures than did the other models (Table 3). The accuracy values of the 4 models were > 0.7, but other assessment measures showed differences. The GARP and MDA models offered the best abilities to predict a species' presence. In contrast, the ANN did not produce satisfactory results because the model had a significantly lower sensitivity value ($p < 0.05$). Based on kappa and sensitivity values, the models predicted common species better. Rare species appeared to be modeled with greater accuracy than common and uncommon species.

When all data, i.e., 14 species from 4 models, were pooled, the predictive performance did not consistently vary with the prevalence available for modeling (Fig. 4). We found no significant correlations between the prevalence and assessment measures (kappa value, $p = 0.42$; accuracy, $p = 0.62$; sensitivity, $p = 0.18$; and specificity, $p = 0.66$). The exception to this pattern occurred with the kappa value of common species, as the species prevalence and kappa values were significantly correlated ($r = 0.33$, $p < 0.05$).

Biodiversity hotspots

The predicted hotspots of Taiwan's endemic bird species showed similar patterns and were mostly located along the Central Mountain Range, judging by the frequency histogram and probability density function approaches of ensemble forecasting (Figs. 5, 6). All 5 maps derived from the frequency histogram approach showed relatively similar patterns (Fig. 5). The extent of hotspots by MDA (3517 km²) was the highest, followed by ANN (2951 km²), LR (2641 km²), and GARP (2155 km²). Based on the combined hotspot map, GARP contributed most predictions in class 4 (25.1%); followed by LR (20.5%), ANN (18.3%) and MDA (15.4%). For class 3, LR and GARP had higher contributions (34.6 and 34.5%, respectively) than MDA (23.2%) and ANN (22.1%).

The hotspot map generated by the probability density function indicated that areas at mid and high elevations had higher probabilities (Fig. 6). The cutoff probability to determine the hotspot

was 0.6. The probability classes included 0.6-0.7 (4.4%), 0.7-0.8 (1.1%), and 0.8-0.9 (0.16%).

Not all of the known hotspots were located in the higher class or probability region, although all known hotspots were located in the ensemble forecasting region predicted by the frequency

(a) Taiwan Yuhina (*Yuhina brunneiceps*), a common species

(b) Flamecrest (*Regulus goodfellow*), an uncommon species
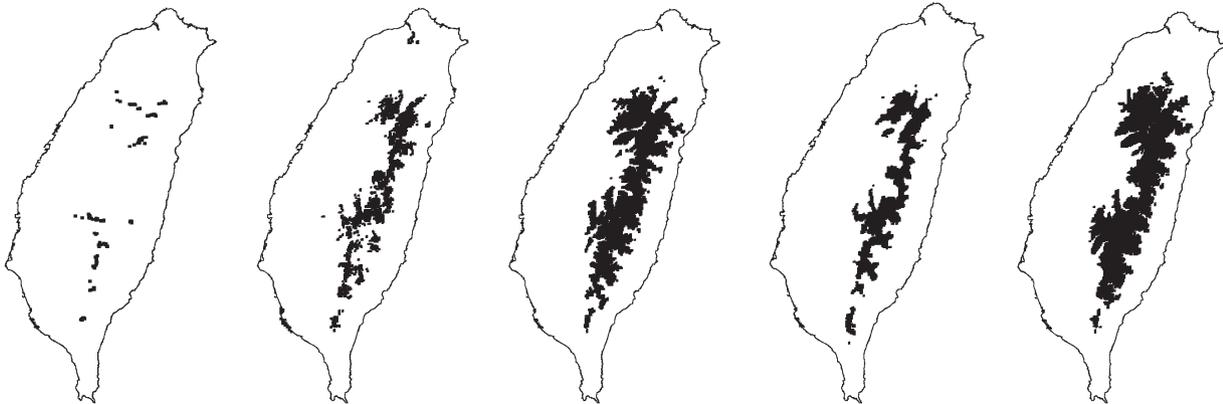
(c) Mikado Pheasant (*Syrmaticus mikado*), a rare species

| Species Known Distributions | Artificial Neural Network (ANN) | Genetic Algorithm for Rule-set Prediction (GARP) | Logistic Regression (LR) | Multiple Discriminant Analysis (MDA) |
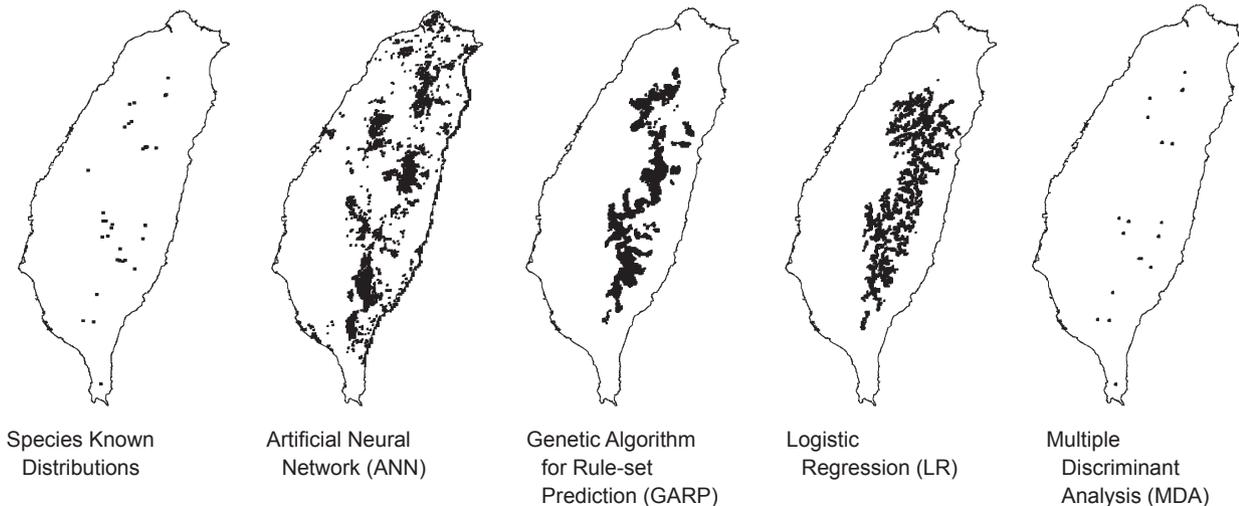
Fig. 2. Distribution maps for 3 selected species generated by 4 predictive models.
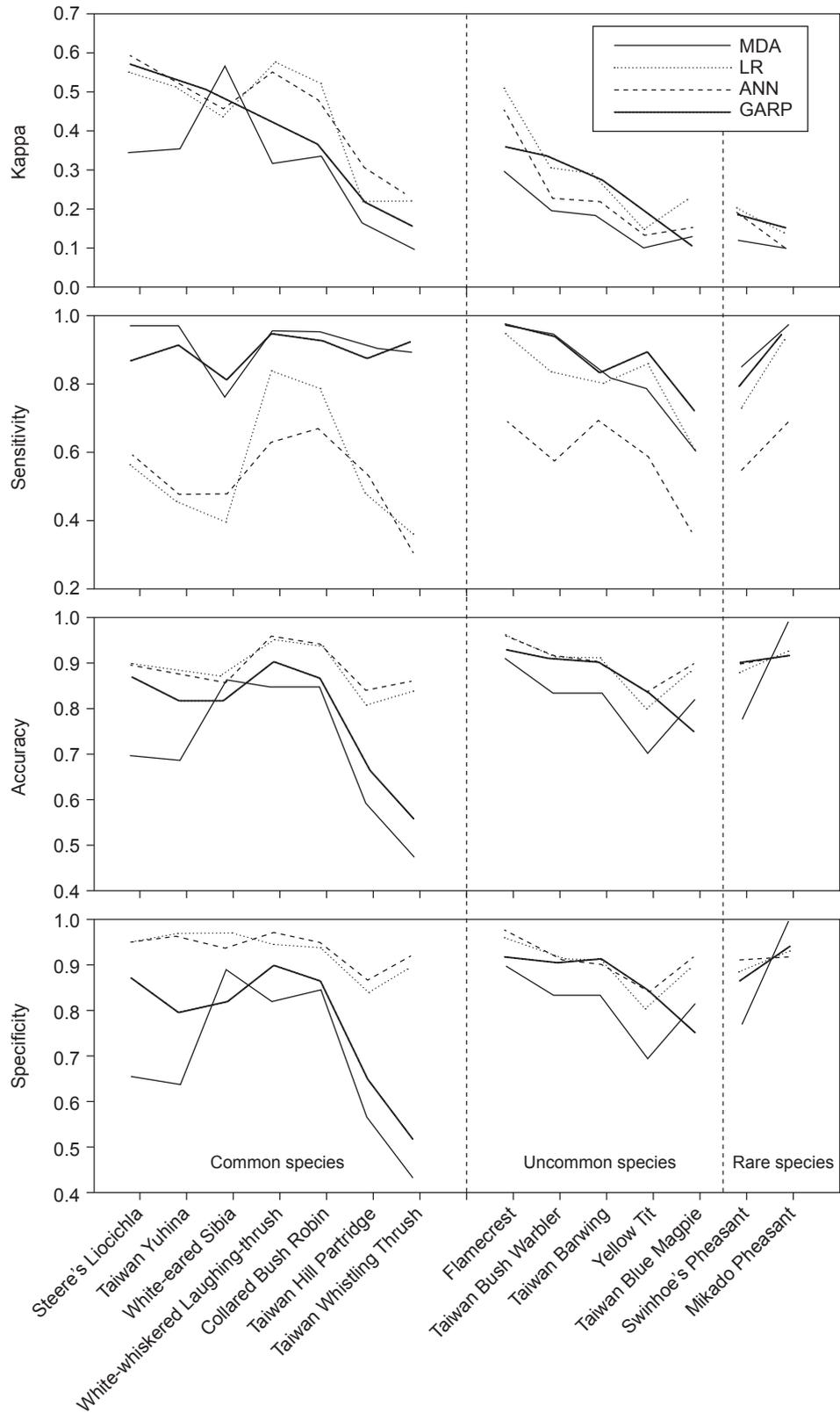
Fig. 3. Comparison of 4 model performances (kappa, accuracy, sensitivity, and specificity) from the 4 predictive models (multiple discriminant function (MDA), logistic regression (LR), artificial neural network (ANN), and genetic algorithm for rule-set prediction (GARP)) for 14 (of 16) endemic bird species in Taiwan.

histogram and probability density function. With the frequency histogram approach, 31% of the known hotspots were located in class 4, 27% in class 3, 24% in class 2, and 18% in class 1. Nonlinear models predicted 68.9% of the known hotpots as possible hotspots while the linear model predicted 63%. In the probability density function, the minimum probability of the known hotspots

was 0.43 and the maximum was 0.86. Eighty-two percent of the known hotspots had predicted probabilities of > 0.6 (0.6-0.7, 42.3%; 0.7-0.8, 32%; 0.8-0.9, 7.7%). The known hotspots belonging to classes 3 and 4 all had higher predictive probabilities (0.6-0.9) while those of classes 2 and 1 were 0.5-0.8 and 0.4-0.7 respectively.

Based on the frequency histogram approach, we identified classes 3 and 4 (i.e., probabilities of > 0.6 in the probability density function), with a total area of 1583 km^2 in area, as hotspots. Only 25% of the hotspots are currently under protection by national parks. Shei-Pa and Yushan National Parks contributed more than the others. No hotspots occurred in Yangmingshan or Kenting National Parks.

The location of hotspots showed distinct environmental characteristics (Table 4). Among the environmental variables we examined, elevation, mean NDVI, annual mean temperature, road density, and forest density were significant factors determining species distributions. Hotspots were located at mid-elevations (with peaks at around 1844-2308 m), and had moderate mean temperatures (11-14°C), higher mean NDVI (0.4-0.6), little human disturbance, and high forest
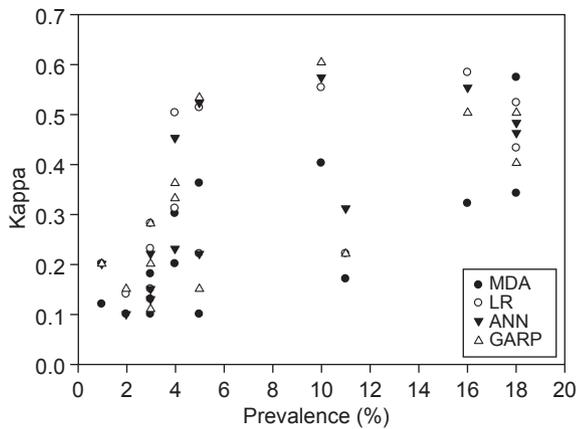

Fig. 4. Relation between kappa and data prevalence (%) in 4 predictive models for 14 (of 16) endemic bird species in Taiwan.

Table 3. Statistics of 4 model evaluation measures (kappa, accuracy, sensitivity, and specificity) and the area under the receiver operating characteristic curve (AUC) for 3 different species categories (common, uncommon, and rare species) by 4 predictive models: multiple discriminant function (MDA), logistic regression (LR), artificial neural network (ANN), and genetic algorithm for rule-set prediction (GARP). Averages for each species category are presented

Species category and model	Mean value				
	Kappa	Accuracy	Sensitivity	Specificity	AUC
Common species (*n* = 7)					
MDA	0.32	0.72	0.91	0.70	
LR	0.43	0.88	0.55	0.93	0.90
ANN	0.44	0.89	0.53	0.94	0.88
GARP	0.42	0.79	0.89	0.80	0.88
Uncommon species (*n* = 5)					
MDA	0.21	0.82	0.83	0.82	
LR	0.30	0.90	0.81	0.90	0.92
ANN	0.25	0.90	0.58	0.91	0.87
GARP	0.30	0.87	0.87	0.87	0.90
Rare species (*n* = 2)					
MDA	0.10	0.89	0.43	0.89	
LR	0.20	0.91	0.81	0.91	0.93
ANN	0.15	0.91	0.44	0.92	0.79
GARP	0.20	0.91	0.84	0.91	0.93
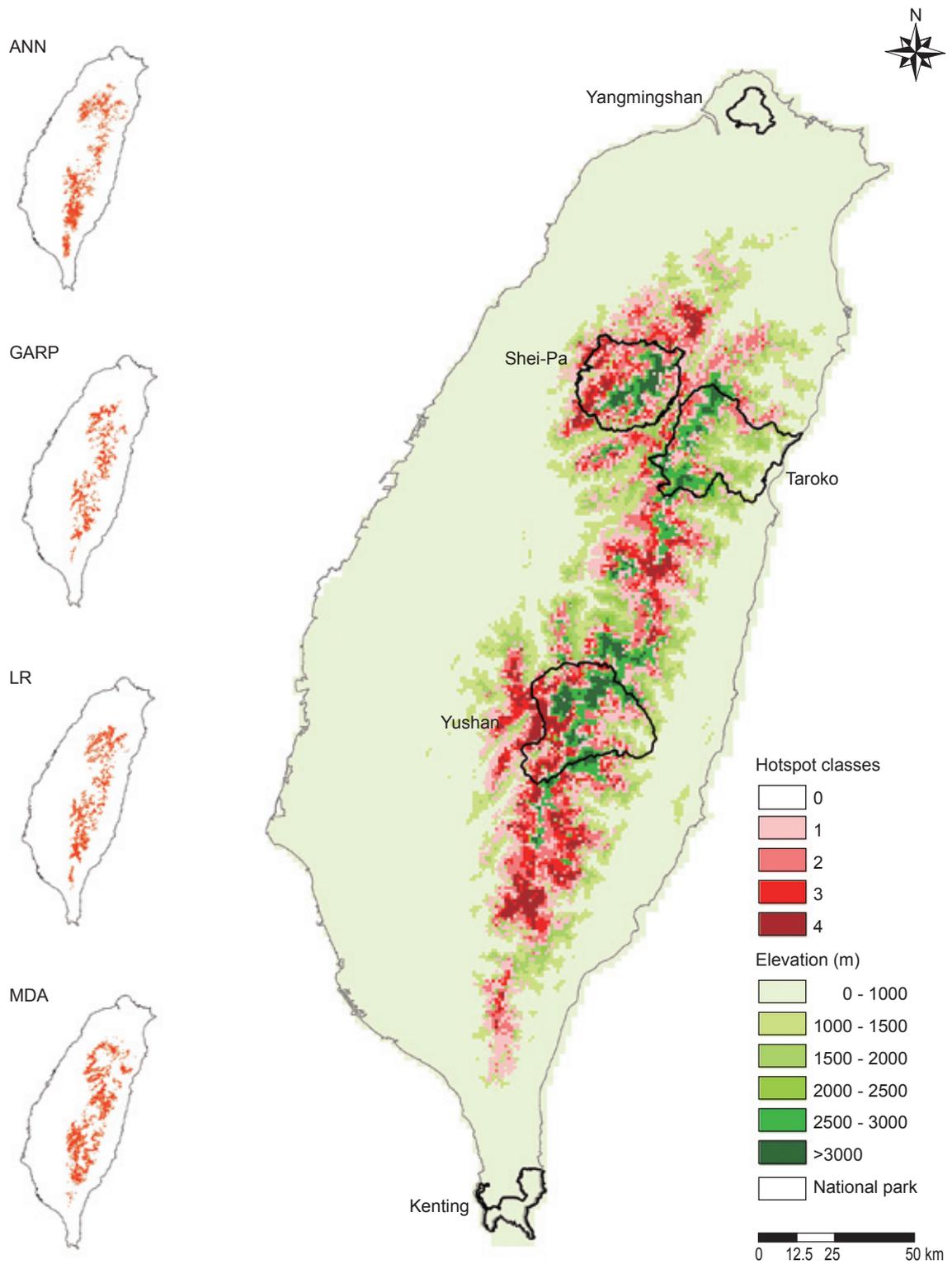
Fig. 5. Biodiversity hotspot maps for 14 (of 16) endemic bird species in Taiwan. Maps in the left column are hotspots generated by each predictive model. On the right shows the overall hotspot map by the frequency histogram approach, i.e., overlaying all 4 hotspot maps. Five classes were identified: 0, no hotspot predicted; 1, predicted by only 1 model; 2, predicted by 2 models; 3, predicted by 3 models; and 4, predicted by 4 models.

cover. When comparing different hotspot classes, all variables, except for road density, had identical patterns of decreasing standard deviations from no model to all 4 models predicting hotspots.
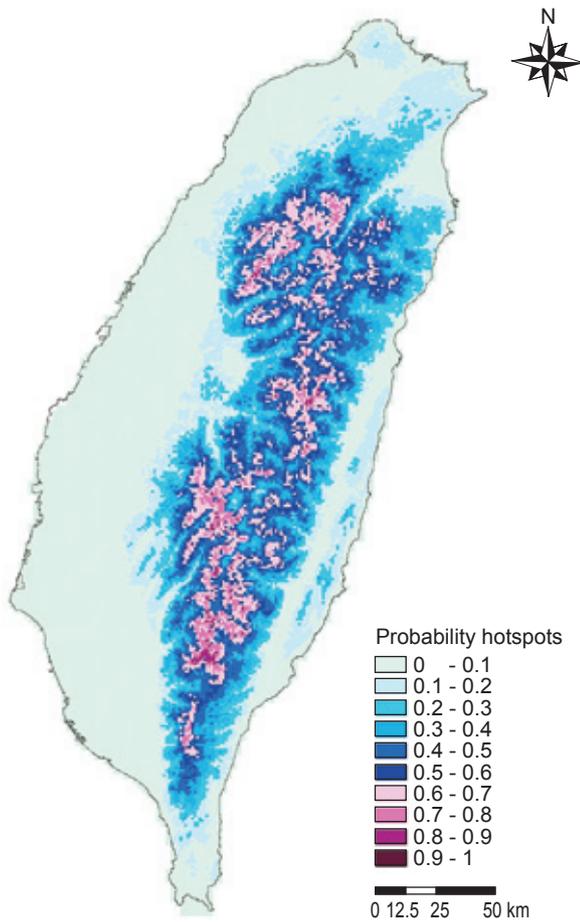
Fig. 6. Biodiversity hotspot maps for 14 (of 16) endemic bird species in Taiwan created by the probability density function approach.

DISCUSSION

During the past decade, species predictive models have matured and become a useful tool for many applications (e.g., Corsi et al. 1999, Margulesand Pressey 2000, 2007, Miller et al. 2004, Rodríguez et al. 2007). We followed the ensemble forecasting approach (Araújo and New 2007) to define biodiversity hotspots by combining 4 hotspot maps generated by each predictive model. As a general rule, an area identified as a biodiversity hotspot by several models may have greater credibility than that identified by a single model. However, our results only partially supported this statement. Our results indicated that areas with higher frequencies or higher probabilities were correlated with known hotspot locations, but a few of the hotspots were located in low-frequency or -probability regions. Although comparing different predictive models can be a useful conservation approach to determine biodiversity hotspots, ensemble forecasting approaches should be applied with caution when locating biodiversity hotspots. We agree with the warning made by Araújo and New (2007) that ensemble forecasting approaches still require strong support of the better individual predictive models.

Despite variations, the 4 predictive models were all workable, with the non-linear models having better predictive power than the linear one. Based on kappa and sensitivity values, and the consistency of these measurements, GARP was a better choice in our situation using a fine resolution (1 × 1 km) grid. Several studies (e.g., Feria and Peterson 2002, Stockman et al. 2006) also reported that GARP had higher predictive capabilities under diverse circumstances.

Taiwan's endemic bird species richness patterns might be a proxy for the overall breeding

Table 4. Mean (± 1 standard deviation; SD), minimum, and maximum of 5 environmental variables characterizing Taiwan's endemic species hotspots generated by the frequency histogram approach

Variable	Classes 1-4 (*n* = 5894)			Classes 2-4 (*n* = 3246)			Classes 3 and 4 (*n* = 1583)			Class 4 (*n* = 541)		
	Mean (SD)	Min	Max	Mean (SD)	Min	Max	Mean (SD)	Min	Max	Mean (SD)	Min	Max
Annual mean temperature (°C)	13.6 (2.2)	7	20.3	13.1 (1.7)	8.3	18.7	12.9 (1.3)	9.3	16.7	12.8 (1.1)	10.2	16.2
Elevation (m)	1917 (437.8)	417.3	3063.6	2037 (300.4)	922.9	2714.6	2076 (232.2)	1477.8	2617.7	2081 (202)	1569.7	2588.6
Forest density (%)	98 (5.1)	3.6	100	98.2 (3.8)	22.9	100	98.5 (2.9)	67.2	100	98.7 (2.2)	84.5	100
Road density (m/ha)	1.8 (5.5)	0	61.9	1.9 (5.6)	0	39.5	2.3 (6.1)	0	39.5	2.74 (6.62)	0	36.0
Mean NDVI	0.44 (0.075)	0.16	0.63	0.45 (0.074)	0.21	0.62	0.45 (0.073)	0.20	0.62	0.45 (0.072)	0.27	0.63

bird species richness pattern (Nieh 1999, Koh et al. 2006). The significant environmental variables in our hotspot analysis were similar to variables correlated with breeding bird species richness in other studies (Lee et al. 2004, Koh et al. 2006). Both breeding bird and endemic species hotspots showed that elevations around 2000 m had the highest species richness. These results suggest that distribution patterns found for endemic species might be a surrogate for the entire 150 breeding bird species of Taiwan, thus reducing the possible time and energy of collecting distribution data of all breeding bird species. Finally, Nieh (1999), who used 2 km grid data and a regression approach to derive the breeding bird hotspots in Taiwan, showed similar hotspot locations to our predictions. Thus, we hypothesize that Taiwan's endemic bird species hotspots might be a good indicator of avian biodiversity representing all breeding bird species.

The issue of using presence-absence data in ecology remains a concern. Our findings indicated that models using presence-only data provided better predictions than those using both presence and absence data. One of the possible reasons is that the presence of a target species can often be confirmed at a location, but it is generally impossible to confirm the absence even with our 1 km^2 data. Due to the highly complex topography in Taiwan, there are many areas that are inaccessible. If an observed absence was 'presence-not detected' datum which has sometimes been used to acknowledge that 'non-detection' does not equate to species absence, this kind of absence data can also confuse the biological and sampling processes (Field et al. 2005, MacKenzie 2005a b). Thus, determining a suitable scale and checking survey data quality are necessary for predicting species distributions.

Although presence-absence data can provide more information for model building, our results showed that these models provided little additional improvement. When neighboring grid cells had similar environments but different results for species presence/absence, these presence-absence models could not address these differences. Such a condition of neighboring grid cells creates a falsehood in model building and decreases the accuracy.

One important caveat of constructing environmental variables is their correlations to species habitat requirements. If the variables do not represent a species' habitat requirements, the prediction might not be adequate. For example,

the Taiwan Whistling Thrush is a common and widespread bird with sufficient data for model building, but our model performance was not satisfactory. One of the possibilities is that the current environmental variables do not represent the key predictors for this species. The Taiwan Whistling Thrush occurs in riparian zones, but we had few environmental variables to represent this habitat type, and even with finer-scale data, we could not resolve the issue. To avoid similar conditions, future predictive model need to be designed with appropriate environmental variables.

We simultaneously used 2 threshold-determining approaches for determining the presence/absence of a species once the models generated the probabilities. Our rules of defining thresholds referred to the ROC plot-based and prevalence approaches. Liu et al. (2005) showed that these 2 threshold-determining approaches used independently were relatively better than other approaches. Although the combination of 2 threshold-determining approaches used here resulted in lower assessment measures due to more-serious restrictions, we still expected that the predicted presence sites should be more accurate.

Finally, the LR, GARP, ANN, and MDA models are feasible for use in modeling species distributions. Although they have some limitations, we suggest that using a combination approach to identify common features and conservation priorities of biodiversity hotspots is still a reliable and better tool for the efficient use of conservation management.

REFERENCES

Aitkenhead MJ, MJ Mustard, AJS McDonald. 2004. Using neural networks to predict spatial structure in ecological systems. Ecol. Model. **179:** 393-403.

Araújo MB, M New. 2007. Ensemble forecasting of species distributions. Trends Ecol. Evol. **22:** 42-47.

Austin MP, JA Meyers. 1996. Current approaches to modeling the environmental niche of eucalypts: implication for management of forest biodiversity. For. Ecol. Manage.

85: 95-106.

Bonn A, ASL Rodriguez, KJ Gaston. 2002. Threatened and endemic species: Are they good indicators of patterns of biodiversity on a national scale? Ecol. Lett. **5:** 733-741.

Collett D. 1991. Modeling binary data. London: Chapman and Hall.

Corsi F, E Dupre, L Boitan. 1999. A large-scale model of wolf distribution in Italy for conservation planning. Conserv. Biol. **13:** 150-159.

DeLeo JM. 1993. Receiver operating characteristic laboratory (ROCLAB): software for developing decision strategies that account for uncertainty. Proceeding of the Second International Symposium on Uncertainty Modeling and Analysis. College Park, MD: IEEE Computer Society Press, pp. 318-325.

Dickinson EC. 2003. The Howard and Moore complete checklist of the birds of the world, 3rd ed. Princeton, NJ: Princeton Univ. Press.

Elith J, CH Graham, RP Anderson, M Dudík, S Ferrier, A Guisan, RJ Hijmans, F Huettmann, JR Leathwick, A Lehmann, J Li, LG Lohmann, BA Loiselle, G Manion et al. 2006. Novel methods improve prediction of species' distribution from occurrence data. Ecography **29:** 129-151.

Feria ATP, AT Peterson. 2002. Prediction of bird community composition based on point-occurrence data and inferential algorithms: a valuable tool in biodiversity assessments. Divers. Distrib. **8:** 49-56.

Field SA, AJ Tyre, KH Thorn, PJ O'Connor, HP Possingham. 2005. Improving the efficiency of wildlife monitoring by estimating detectability: a case study of foxes (*Vulpes vulpes*) on the Eyre Peninsula, South Australia. Wildlife Res. **32:** 253-258.

Fielding AH, JF Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/ absence models. Environ. Conserv. **24:** 39-49.

Fuller T, V Sánchez-Cordero, P Illoldi-Rangel, M Linaje, S Sarkar. 2007. The cost of postponing biodiversity conservation in Mexico. Biol. Conserv. **134:** 593-600.

Gray D, MP Scarsbrook, JS Harding. 2006. Spatial biodiversity patterns in a large New Zealand braided river. NZ J. Mar. Freshw. Res. **40:** 631-642.

Guisan A, NE Zimmermann. 2000. Predictive habitat distribution models in ecology. Ecol. Model. **135:** 147-186.

Hirzel A, A Guisan. 2002. Which is the optimal sampling strategy for habitat suitability modeling. Ecol. Model. **157:** 331-341.

Hsu FH, CT Yao, RS Lin, CC Yang, SJ Lai. 2004. Avian species composition and distribution along elevation gradient in the southern Taiwan. Endem. Species Res. **6:** 41-66. (in Chinese)

Jose MP, T Fernando. 1997. Prediction of functional characteristics of ecosystem: a comparison of artificial neural networks and regression models. Ecol. Model. **98:** 173-186.

Kati V, P Devillers, M Dufrêne, A Legakis, D Vokou, P Lebrun. 2004. Hotspots, complementarity or representativeness? Designing optimal small-scale reserves for biodiversity conservation. Biol. Conserv. **120:** 471-480.

Koh CN, PF Lee, RS Lin. 2006. Bird species richness patterns of northern Taiwan: primary productivity, human population density, and habitat heterogeneity. Divers. Distrib. **12:** 546-554.

Lee PF, TS Ding, FH Hsu, S Geng. 2004. Breeding bird species richness in Taiwan: distribution on gradients of elevation, primary productivity and urbanization. J. Biogeogr. **31:** 307-314.

Lek S, M Delacoste, P Baran, I Dimopoulos, J Lauga, S Aulagnier. 1996. Application of neural networks to modeling nonlinear relationships in ecology. Ecol. Model. **90:** 39-52.

Li SH, JW Li, LX Han, CT Yao, H Shi, FM Lei, C Yen. 2006. Species delimitation in the Hwamei *Garrulax canorus*. Ibis **148:** 698-706.

Li Z, Q Zhu, C Gold. 2005. Digital terrain modeling – principles and methodology. Boca Raton, FL: CRC Press.

Lira-Noriega A, J Soberón, AG Navarro-Sigüenza, Y Nakazawa, AT Peterson. 2007. Scale dependency of diversity components estimated from primary biodiversity data and distribution maps. Divers. Distrib. **13:** 185-195.

Liu C, PM Berry, TP Dawson, RG Pearson. 2005. Selecting thresholds of occurrence in the prediction of species distribution. Ecography **28:** 385-393.

MacKenzie DI. 2005a. Was it there? Dealing with imperfect detection for species presence/absence data. Aust. NZ J. Statistics **47:** 65-74.

MacKenzie DI. 2005b. What are the issues with presence-absence data for wildlife managers? J. Wildlife Manage. **69:** 849.

Manel S, JM Dias, ST Buckton, SJ Ormerod. 1999a. Alternative methods for predicting species distribution: an illustration with Himalayan river birds. J. Appl. Ecol. **36:** 734-747.

Manel S, JM Dias, SJ Ormerod. 1999b. Comparing discriminant analysis, neural networks and logistic regression for predicting species distribution: a case study with a Himalayan river bird. Ecol. Model. **120:** 337-347.

Margules C, RL Pressey. 2000. Systematic conservation planning. Nature **405:** 243-253.

McCullagh P, JA Nelder. 1989. Generalized linear models. London: Chapman and Hall.

Miller JR, MG Turner, EAH Smithwick, CL Dent, EH Stanley. 2004. Spatial extrapolation: the science of predicting ecological patterns and processes. BioScience **54:** 310-320.

Moisen GG, TS Frescino. 2002. Comparing five modeling techniques for predicting forest characteristics. Ecol. Model. **157:** 209-225.

Myers N, RA Mittermeier, CG Mittermeier, AB da Fonseca, J Kent. 2000. Biodiversity hotspots for conservation priorities. Nature **403:** 853-858.

Nieh CH. 1999. Species richness pattern of breeding birds in Taiwan. MS thesis. National Taiwan Univ., Taipei, Taiwan. (in Chinese)

Phillips SJ, RP Anderson, RE Schapire. 2006. Maximum entropy modeling of species geographic distribution. Ecol. Model. **190:** 231-259.

Reid WV. 1998. Biodiversity hotspots. Trends Ecol. Evol. **13:** 275-280.

Robertson MP, CI Peter, MH Villet, BS Ripley. 2003. Comparing models for predicting species' potential distribution: a case study using correlative and mechanistic predictive modeling techniques. Ecol. Model. **164:** 153-167.

Rodríguez JP, L Brotons, J Bustamante, J Seoane. 2007. The application of predictive modeling of species distribution to biodiversity conservation. Divers. Distrib. **13:** 243-251.

Scachetti-Pereira R. 2001. Desktop GARP. Available at htp:// www.lifemapper.org/deskgarp

Scott JM, F Davis, B Csuti, R Noss, B Butterfield, C Groves et al. 1993. Gap analysis: a geographic approach to protection of biological diversity. Wildlife Monogr. **123:** 1-41.

Shiu HJ, PF Lee. 2003. Assessing avian point-count duration and sample size using species accumulation functions. Zool. Stud. **42:** 357-367.

Stockman AK, DA Beamer, JE Bond. 2006. An evaluation of a GARP model as an approach to predicting the spatial distribution of non-vagile invertebrate species. Divers. Distrib. **12:** 81-89.

Stockwell D. 2007. Niche modeling. Boca Raton, FL: Chapman and Hall/CRC Press.

Su HJ. 1984. Studies on the climate and vegetation types of the natural forests in Taiwan (I): analysis of the variations on climate factors. Q. J. Chinese For. **17:** 1-14.

Zaniewski AE, A Lehmann, JM Overton. 2002. Predicting species spatial distribution using presence-only data: a case study of native New Zealand ferns. Ecol. Model. **157:** 261-280.