available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/ecolmodel

Mapping non-wood forest product (matsutake mushrooms) using logistic regression and a GIS expert system

Xuefei Yang[a], *Andrew K. Skidmore*[b], *David R. Melick*[a], *Zhekun Zhou*[a], *Jianchu Xu*[a,*]

[a] *Laboratory of Biogeography and Biodiversity, Kunming Institute of Botany, The Chinese Academy of Sciences, Long-quan Road 610, Kunming 650204, PR China*

[b] *International Institute for Geo-information Science and Earth Observation, Hengelosestraat 99, P.O. Box 6, 7500 AA Enschede, The Netherlands*

ARTICLE INFO

Article history:
Received 20 May 2005
Received in revised form
14 April 2006
Accepted 19 April 2006
Published on line 12 June 2006

Keywords:
Matsutake mushroom
Non-wood forest products
Spatial distribution
Logistic regression
GIS expert system
Northwest Yunnan
China

ABSTRACT

Matsutake (*Tricholoma* spp.) are a group of commercially important mushrooms that are increasingly threatened by over-collection. Ecologically sustainable management of matsutake has been hindered by the lack of essential information such as reliable distribution maps. Although a variety of spatial distribution models have been applied to map many different plants, this has rarely been attempted for mushrooms. In this study, we employed a logistic regression and a GIS expert system to model the fine-scale spatial distribution of matsutake in Yunnan, southwest China. Both models predicted mushroom habitat to an accuracy acceptable for resource management. The overall mapping accuracy of the GIS expert system was slightly better than the logistic regression model (70.37% versus 65.43%). Furthermore, unlike the logistic regression model, developing the GIS expert system required no field-based samples. This has important practical implications because it is very difficult to survey and sample mushrooms and other non-wood forest products (NWFP), which are usually inconspicuous species and/or lower plants. Therefore, when adequate samples are not available, incorporating local expert knowledge can help make better-informed management decisions and provide an affordable habitat identification tool.

1. Introduction

As a seasonal delicacy favored by the Japanese, matsutake have become a commercially important wild mushroom (Wang et al., 1997). Consumption in Japan is approximately 3000 tonnes per year, of which Japan produces 1000 tonnes in a good year (Van On, 1993); the remainder is imported mainly from Korea, China, and North America. "Matsu-take" translates literally as "pine-mushroom" from the Japanese. Originally, matsutake referred specifically to *Tricholoma matsutake*, but subsequently the name refers to a group of similar mushrooms (Hosford et al., 1997). While the taxonomy of these mushrooms is still under debate, it is generally accepted that there are 15 species (and one variety) distributed worldwide (Liu et al., 1999; Zang, 1990). Matsutake mushrooms are soil-borne and perennial mycorrhizal fungi. They develop a symbiotic association with the roots of specific trees (James, 1998; Ogawa, 1976).

Collection of matsutake can generate significant income, for example, in Canada, the British Columbian wild mushroom industry harvests 250–400 tonnes per year, with a value of US$ 25–45 million (Wills and Lipsey, 1999). Matsutake are also important to rural livelihoods in many parts of the world (Kranabetter et al., 2002; Pilz and Molina, 2002). Naturally

* *Corresponding author*. Tel.: +86 871 5212143; fax: +86 871 5223909.
E-mail addresses: xujianchu@mail.kib.ac.cn, xujianchu@cbik.ac.cn, jxu@icimod.org (J. Xu).

0304-3800/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecolmodel.2006.04.011

enough, species that command high prices, such as matsutake and truffles, are increasingly at risk from over-exploitation and/or improper ecosystem management (Larsen et al., 2000; Newton et al., 2003; Van On, 1993). This is certainly the case in northwest Yunnan (the subject of our study), where the production of matsutake declined from 530 tonnes in 1995 to 272 tonnes in 2000 (Xu and Ribot, 2004). Consequently, the *T. matsutake* has been given protected status (National Grade II) by CITES China (KIB-CAS, 2003). Hence, growing concerns have focused on how to properly use and protect the mushrooms. Development of sustainable harvest plans requires knowledge of mushroom distribution; however, this information is often scarce or of poor quality (FAO, 2002; Larsen et al., 2000; Lehmann et al., 2002; Vantomme et al., 2002; Zaniewski et al., 2002). Kranabetter et al. (2002) estimated the habitat of American matsutake (*Tricholoma magnivelare*) using aerial photos in collaboration with summarized ecological descriptions. Such studies are a seminal attempt to investigate mushroom distribution, but more work is required to develop better spatial models of these valuable resources in complex landscape mosaics.

Spatial models have been widely used to predict the distribution patterns of many species and plant communities. These models have been intensively reviewed by Franklin (1995), Guisan and Zimmermann (2000) and Austin (2002). Generally these studies have focused on dominant or overstorey species (Iverson et al., 1999), only a few of them considered shrubs (Franklin, 1998), ferns (Zaniewski et al., 2002) and cryptogams (Peltoniemi et al., 2005). To our knowledge, few have involved mushrooms. In the present study, we attempted to model the spatial distribution of matsutake at a fine scale in complex mountainous terrain in Shangri-La County, northwest Yunnan. In order to test the validity of the output, we compared two models: a logistic regression and a GIS expert system. The logistic regression was used because it is: (a) a widely used and proven statistical approach (Aspinall, 2002; Manel et al., 1999; Stephenson et al., 2006); (b) suitable for regression when the dependent variable is binary in nature; and (c) easily implemented with simple programmes in GIS software packages. The GIS expert system was used to integrate existing (local) knowledge into a model; it is then compared against the statistical model to estimate the robustness of this approach. To build the models it is critical to identify environmental factors that are likely to influence matsutake distribution.

2. Study area

2.1. *Northwest Yunnan*

Located in the southern mountain region (Hengduan Mountains) of the Eastern Himalayas, northwest Yunnan is in a transitional zone between the Qinghai-Tibet and Yunnan-Guizhou Plateaus. Three major rivers, the Lancang (Mekong), Jinsha (upper reaches of the Yangtze) and the Nu (Salween), run parallel in a southerly direction. High mountains and deep gorges dominate the regional landscape, with the elevation ranging from 6740 m at the summit of Kawagebo to about 500 m in the lower parts of the Nujiang valley. The variation of topography and latitude results in a high diversity of microclimates. Consequently, northwest Yunnan contains 40% of the province's 15,000 plant species and is recognized as a global biodiversity hotspot (Myers et al., 2002).

2.2. *Jidi village*

This study was centered on Jidi administrative village (27°43′24″–28°9′54″N, 99°32′12″–99°43′17″E) (Fig. 1), one of the most productive areas for *T. matsutake* in Shangri-La County, northwest Yunnan. The study area covers an area of 214.58 km^2 with elevations varying from 3100 to 4200 m

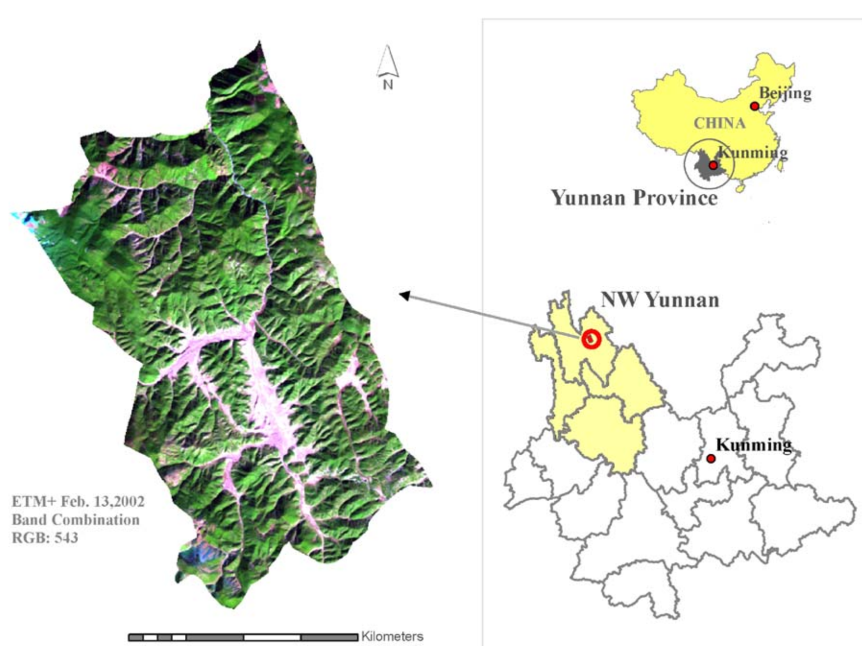

Fig. 1 – Location and satellite image of the study area.

Table 1 – Environmental variables used to analyse matsutake distribution

Variables	No. of observations	Abbreviation	Data type	Distribution
Matsutake P/A	222	M	Binary	Binomial
Forest type	216	F	Category	–
Elevation (m)	175	ELE	Numeric	Approximately normal
Slope (°)	180	S	Numeric	Approximately normal
Aspect (°)	174	A	Numeric	Approximately normal
Surface depth (cm)	120	SD	Numeric	Skewed
Litter cover (%)	108	LTC	Numeric	Skewed
Average tree height (m)	142	H	Numeric	Skewed
Tree cover (%)	162	TC	Numeric	Approximately normal
Tree basal area (m^2/ha)	119	BA	Numeric	Skewed
Shrub cover (%)	118	SHC	Numeric	Skewed

over the generally hilly terrain. Vegetation is predominantly forest, generally of pine, fir, oak, or combinations thereof. Flatter regions in lower elevations, which are relatively scarce, are used for agriculture and pasture. The climate is monsoon-influenced, clearly divided into a dry season (November–May) and a wet season (June–October). Data collected from the Meteorology Observation Station of Shangri-La County (located 20 km south of Jidi village, elevation 3200 m; data averaged from 1971 to 2002) shows an annual mean rainfall of 654 mm. Temperatures are generally mild, though winters are cool and snow may be persistent. The maximum monthly average of 13.5 °C occurs in July, while monthly average temperatures are below zero from December through February. This village includes 13 subordinate "natural villages" with a total population of approximately 1600. All the residents are Tibetans performing agriculture and animal husbandry. In recent decades, commercial harvest of matsutake has become the major income source for the community.

3. Methods

3.1. *Sampling design and field measurement*

The problems of field sampling for mushrooms are manifest. Firstly, the dissected landscape, aspect, soil conditions and vegetation cover may all greatly influence mushroom distribution. Moreover, even when probable mushroom sites have been identified it is difficult to be definite as the vegetative hyphae are embedded in the soil, and, even during the rare fruiting periods, the above ground mushrooms (sporocarps) are often covered by litter, herbs, or mosses. Without any preliminary knowledge or experience, finding sporocarps is not an easy task, thus our data collection (August–October, 2003) was assisted by local mushroom pickers. Sample sites were designated as either "presence" where mushrooms were found, or "absence" in sites identified by mushroom pickers as places where matsutake had never been found. Two measures were taken in order to mitigate subjective bias and keep the sampling as close to random as possible. First, six villages were randomly selected for the survey, in order to include the variance among villages. Secondly, seven mushroom pickers were selected as guides in order to minimise site preferences or bias introduced by individual guides.

In each field sample plot, topographic features (elevation, slope, aspect); floristic composition of different strata (tree layer, shrub layer); vegetation height; forest canopy cover; litter cover (large woody debris, leaves, needles, and humic material); substrate depth (including litter cover and top layer soil); and presence/absence of matsutake were recorded. Forest canopy cover and average tree height were visually estimated in a 25 m × 25 m plot. Shrub cover was estimated within a 5 m × 5 m subplot, and litter cover was estimated within a 2 m × 2 m subplot. A GPS connected with a hand-held computer (HP-Ipaq) was used to acquire the spatial location of sample points and record each measurement. Site position was identified by an average of 30 readings. In total, 222 field observations were recorded. Table 1 lists the environmental variables and their characteristics, where matsutake presence or absence (P/A) is the dependent variable and the remaining are independent variables.

3.2. *Data preparation*

Data used in this study were a forest type map, DEM (digital elevation model), slope, aspect, terrain position, litter cover map and 222 field measurements. The forest type map was produced using a Landsat 7 ETM + image (path/row: 132/41, passing date February 13, 2002) (Yang, unpublished data). The DEM was generated using digitized contour lines and high points from topographic map sheets, from which slope and aspect were derived. Further, "terrain position" (such as channels, ridges and planes), slope and aspect were calculated using the DEM and the "topographic feature" module in ENVI 4.0 software.

Although 222 samples were recorded in total, bad samples with missing values[1] were discarded. Only 163 samples were actually used, half of which (n = 82) were randomly selected for building the logistic regression model. The rest of the data set (n = 81) were used for model assessment and validation.

3.3. *Statistical analysis of environmental variables*

The relationships between environmental variables and mushroom distribution were examined with three standard statistical tests: (a) Spearman rank order test; (b) Kendall τ-

[1] At each sample plot, the values of all 11 variables (refer to Table 1) should be recorded. The so-called "bad samples" refers to those samples that have one or more variable values missing. In regression analysis, samples with missing values are automatically excluded.

test; and (c) discriminant function analysis. A Kruskal–Wallis test together with a Mann–Whitney *U* test were employed to test significant differences among sub-classes of each variable. The result was used for reference when formulating the GIS expert rules.

3.4. *Preparing the litter cover map*

The statistical analysis suggested that litter cover is important in predicting matsutake distribution. Hence a litter cover map was produced using a generalized linear regression model with the following input data: forest type map, elevation, slope and aspect. Backward stepwise procedure was used to select explanatory variables. Significance values for "enter" and "remove" were set at 0.05 and 0.06, respectively. A model with forest type and slope as input data was established ($R^2 = 0.83$, $n = 108$, $p < 0.001$). This was executed using the "model maker" module in the ERDAS IMAGINE 8.6 software.

3.5. *Modelling distribution of matsutake with logistic regression*

3.5.1. *The logistic regression*

Logistic regression describes the relationship between the response and the linear sum of the predictor variables. Using the equation below, the presence/absence of matsutake is transformed into a continuous probability *y* ranging from 0 to 1. Values close to 1 represent high probability of presence, whereas, values close to 0 represent high probability of absence.

$$y = \frac{\exp(b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_n x_n)}{1 + \exp(b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_n x_n)}$$

where *y* is the probability, x_n the explanatory variable, b_n the coefficient of x_n, and exp is an exponential function.

3.5.2. *Building the logistic regression model*

Preliminary variable selection was carried out with the assistance of the previous statistical analyses of environmental variables using 82 samples. The best model was selected based on two criteria: approximate variance explained (Nagelkerka R^2) and goodness of fit (Hosmer and Lemeshow test statistic; for details refer to SPSS Inc., 2001).

3.5.3. *Spatial implementation of logistic regression model*

Spatial implementation of the model was achieved using the "model maker" module in ERDAS IMAGINE 8.6 software. The forest type map was recoded to the numeric coefficients that represented the forest types; for instance, pixels with oak forest are recoded with its coefficient, 0.659. Coefficients of various forests are listed in Table 2. The probability surface of matsutake distribution was calculated as the output of the logistic regression.

3.5.4. *Presence/absence transformation and sensitivity analysis for the threshold levels*

Model performance was evaluated on the basis of presence/absence, thus probability values (0–1) were converted to presence or absence at set thresholds. Arbitrarily, presence of an individual is assumed at probability values of greater than or equal to 0.5. However, the 0.5 value may not be optimal in all cases (Manel et al., 1999), therefore, we expanded the sensitivity analysis to consider threshold levels from 0.4 to 0.8.

Table 2 – Coefficients for each forest type in the logistic regression model

Forest type	Coefficients
Non-forest	−7.918
Broadleaved	−8.443
Fir	−8.807
Oak	0.659
Oak-pine-mixed	0.204
Pine	0

Note that pine forest was used as the standard and hence the zero value.

3.6. *Modelling distribution of matsutake with a GIS expert system*

3.6.1. *A GIS expert system*

GIS expert systems are computer programs that simulate the behavior of human experts—they are designed to solve problems related to geographic information systems (Skidmore et al., 1996; Stock, 1987). The GIS expert system used in this study was developed under ENVI-IDL environment by the Natural Resource Department at the International Institute for Geo-informatics Science and Earth Observation, the Netherlands. Bayesian theory (Aspinall, 1992; Aspinall, 1993; Skidmore, 1989; Skidmore et al., 1996) was used as the inference engine. In the GIS expert system, we infer the presence of matsutake at a certain location (a hypothesis) based on available evidence.

3.6.2. *Input data and knowledge based rule formulation*

The selection of environmental variables and formulation of rules was based on the availability of the data and the integration of knowledge from several sources, including: (a) literature (Liu et al., 1999; Zang, 1990); (b) local knowledge acquired through discussion with mushroom pickers; (c) personal field observations; and (d) results from statistical analyses. Where there was disagreement between different sources, a subjective decision was made based on field knowledge. We used the following data layers: (a) litter cover; (b) terrain position; (c) slope; (d) DEM; (e) aspect; and (f) forest type. Table 3 shows the detailed probability estimates for these criteria. The system requires users to input a set of rules that link a hypothesis to the evidence. These rules include *a priori* estimates of probability and the initial conditional probability. The *a priori* probability can be assigned based on knowledge (Skidmore, 1989) or from the estimation of an expert. In this case, we assigned it based on field knowledge. For instance, *a priori* probability for presence is 0.60 because approximately 60% of the study area is potential matsutake habitat. The initial conditional probability is the probability (as estimated by the user) of presence or absence (hypothesis) S_a ($a = 1, \ldots, n$) at location (or pixel) X_{ij} by given a piece of evidence.

3.6.3. *Sensitivity analysis for a priori probabilities*

As is the case for the logistic regression, users are requested to input *a priori* probability values to initiate the GIS expert sys-

Table 3 – *A priori* probability estimates and the initial conditional probability of matsutake presence and absence with given evidence

	A priori	Evidence																							
		Litter cover		Topographic feature			Slope				DEM (m)					Aspect					Forest type					
		L1	L2	Tr	Tp	Tch	S1	S2	S3	S4	D1	D2	D3	D4	D5	A1	A2	A3	A4	A5	F1	F2	F3	F4	F5	F6
Absence	0.40	0.60	0.40	0.40	0.50	0.70	0.50	0.50	0.50	0.90	0.70	0.20	0.30	0.50	0.90	0.50	0.40	0.30	0.40	0.50	0.99	0.99	0.80	0.20	0.30	0.40
Presence	0.60	0.30	0.70	0.50	0.50	0.30	0.30	0.40	0.50	0.01	0.30	0.40	0.50	0.50	0.01	0.40	0.50	0.60	0.50	0.30	0.01	0.01	0.20	0.80	0.80	0.60

L1, <60%; L2, ≥60%; Tr, ridge; Tp, plane; Tch, channel; S1, 0–5°; S2, 6–15°; S3, 16–40°; S4, 41–90°; D1, ≤3300; D2, 3301–3400; D3, 3401–3600; D4, 3601–3700; D5, >3700; A1, 0–90°; A2, 91–180°; A3, 181–270°; A4, 271–360°; A5, no aspect; F1, non-forest; F2, broadleaved forest; F3, fir forest; F4, oak forest; F5, oak-pine-mixed forest; F6, pine forest.

tem process. Examining the sensitivity gives a better understanding of the model and output, therefore, sensitivity analysis was performed across a range of *a priori* probabilities from 0.01 to 0.99.

3.7. *Model evaluation and comparision*

Overall mapping accuracy, sensitivity and specificity, Kappa coefficient and *Z*-test were employed to evaluate and compare the performance of the models. An independent sample set ($n = 81$) was used for model validation and an error matrix was generated. Overall accuracy is defined as the proportion of the total number of correctly predicted sites to the total number of testing samples (Fielding and Bell, 1997). For the binary error matrix, sensitivity and specificity are used as the measurements of accuracy in predicting presence and absence, respectively (Fielding and Bell, 1997). Sensitivity is defined as the proportion of correctly predicted presence to the total number of presence in testing samples; specificity is the proportion of correctly predicted absence to the total number of absence in testing samples (Fielding and Bell, 1997). The Kappa coefficient and its variance (Cohen, 1960; Congalton, 1991; Skidmore et al., 1996) were computed. Differences between classifications were tested through a z-statistic using Kappa coefficients (Cohen, 1960; Congalton, 1991; Skidmore et al., 1996). Apart from statistical evaluation, a further visual assessment based on field knowledge was undertaken to investigate how well a produced map indicated the actual ground situation.

4. Results

4.1. *Environmental variables related to matsutake distribution*

All three statistical tests show that forest type, litter cover and elevation are important environmental variables for matsutake distribution (Table 4). Comparatively weaker evidence (with support of only one or two tests) suggests that aspect, slope, tree height and shrub cover also influence matsutake distribution. The results of Kruskal–Wallis ANOVA test (Appendix 1), and Mann–Whitney *U* test (Appendix B) showed elevation, slope, litter cover, tree height and tree cover to be significant. These results were further used to develop the expert rules.

4.2. *The logistic regression model*

Based on above analysis, we used forest type, litter cover, and elevation to build the logistic regression model. The individual explanatory variables ranked as follows: forest type was the most important, explaining 68% of the variance; litter cover, explained 17.3%; and elevation, explained 6.8%. All kinds of data combinations of the selected explanatory variables were fed into the model and the results were compared. By examining the variance explained and goodness of fit, the model with all these three explanatory variables as input (see below) was selected as the best, since it explained the highest variance

Table 4 – Correlation coefficients and p-value of discriminant function analysis of matsutake and environmental variables

	ELE	S	A	SD	LTC	H	BA	SHC	TC	F
Spearman rank order correlations	0.22*	−0.13	0.13	0.06	0.21*	−0.02	0.17	−0.15	0.06	−0.46*
Kendall τ correlations	0.18*	−0.11*	0.11*	0.05	0.18*	−0.02	0.14*	−0.12*	0.05	−0.42*
p-Value of discriminant function analysis	0.008*	0.416	0.001*	0.853	0.000*	0.006*	0.150	0.271	0.690	0.000*

The full names of the variables can be found in Table 1.
* Significant ($p < 0.05$).

Table 5 – Nagelkerka R^2 and Hosmer and Lemeshow statistic for logistic regression with different input data ($n = 82$)

Variables used	F	LTC	E	F, LTC	F, LTC, E
Nagelkerka R^2	0.68	0.173	0.068	0.726	0.732
Hosmer and Lemeshow statistic	0.999*	0.023	0.382*	0.337*	0.94*

The full names of the variables can be found in Table 1.
* Significant goodness of fit ($P < 0.05$).

(73.2%) and adequately fitted the data (0.94) (Table 5).

$$M = \frac{\exp(-12.616 + F + 0.043\mathrm{LTC} + 0.003\mathrm{ELE})}{1 + \exp(-12.616 + F + 0.043\mathrm{LTC} + 0.003\mathrm{ELE})}$$

where M is the probability of matsutake presence, F the land forest type, LTC the litter cover, and ELE is the elevation. The coefficients for each forest type are listed in Table 2.

4.3. *Model performance and sensitivity*

The maps produced by the two models showed some variation (Fig. 2). In terms of accuracy, the GIS expert system yielded a slightly better result in overall mapping accuracy than the logistic regression model (70.37% versus 65.43%; see Tables 6 and 7). However, the Z-test (for Kappa) statistic shows that there is no significant difference in the accuracy level of both models (Table 7). Moreover, both models have the same sensitivity, i.e., 87.23%; with specificities of 47.05% (the GIS expert system) and 35.29% (the logistic regression), respectively (Table 6). This reveals that they were equally good at predicting presence, but not very accurate for absence. Although not particularly impressive, the GIS expert system was better at predicting absence. Note that the logistic regression predicted matsutake to be present in 76.9% of the study area, compared with 61.4% presence predicted by the GIS expert system.

The results of the sensitivity analyses are shown in Figs. 3 and 4. For the logistic regression model, the accuracy level is relatively stabile (around 65.43%) when the threshold is less than 0.65, but above this value there is an abrupt drop in accuracy as the consequence of the rapid increase of false presence (Fig. 3). Thus it appears that adopting a threshold

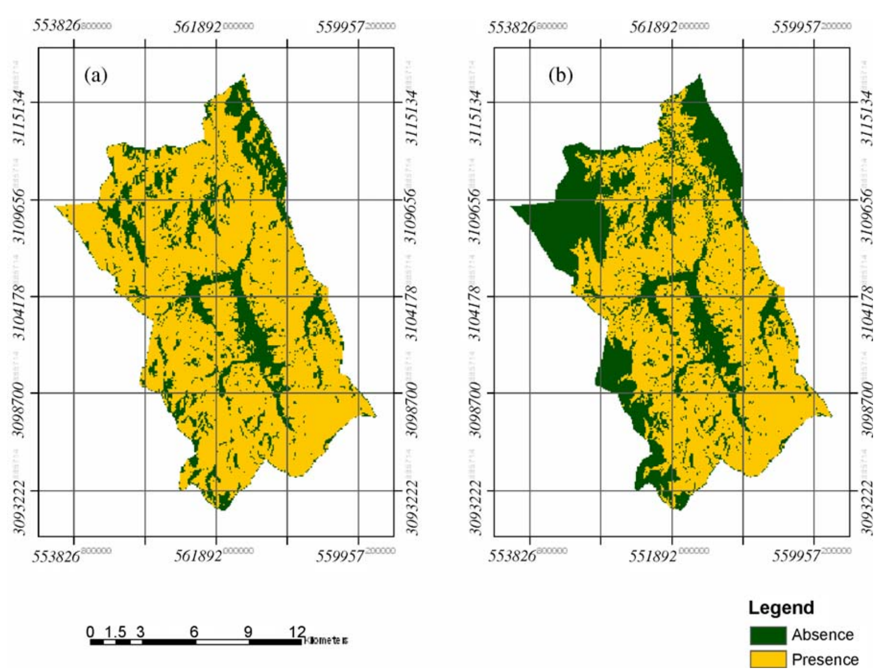

Fig. 2 – Presence/absence of matsutake predicted by: (a) the logistic regression (threshold = 0.5); (b) the GIS expert system (*a priori* for absence = 0.4).

Table 6 – Error matrices of the two models: (a) logistic regression (threshold = 0.5) and (b) the GIS expert system (*a priori* for absence = 0.4)

	No. of pixels from reference		
	Absence	Presence	Total
(a) Logistic regression			
No. of pixels from model			
Absence	11	5	16
Presence	23	42	65
Total	34	47	81
	Overall accuracy		65.43%
(b) Expert			
No. of pixels from model			
Absence	16	6	22
Presence	18	41	59
Total	34	47	81
	Overall accuracy		70.37%

Table 7 – Statistics for evaluation of model performance

	Overall accuracy (%)	Sensitivity (%)	Specificity (%)	K	K variance	Z
Logistic regression	65.43	87.23	35.29	0.2343	0.00223	1.89 (NS)
Expert	70.37	87.23	47.05	0.3605	0.00225	

NS, not significant ($p < 0.05$).

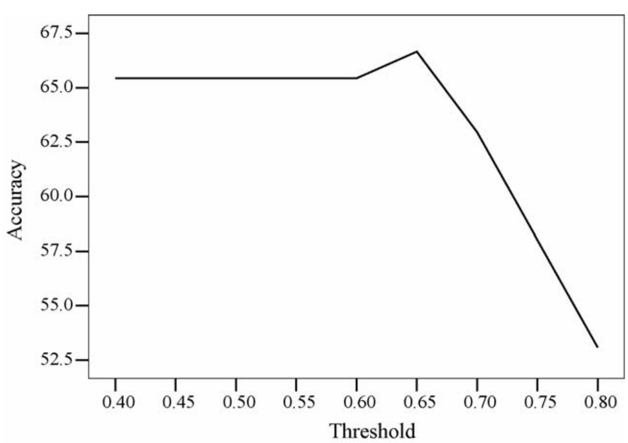

Fig. 3 – The accuracy changes with different thresholds that divide “presence” and “absence” for the logistic regression. Decreasing accuracy at levels above 0.65 reflect an increase of false “presence” results.

value of 0.5 is appropriate. For the GIS expert system, when the *a priori* probabilities for absence range from 0.01 and 0.8 the overall accuracy remains relatively stable. However, the specificity and sensitivity change considerably; these changes are inversely related with specificity increasing as sensitivity decreases (Fig. 4).

5. Discussion

5.1. *Ecology of matsutake in northwest Yunnan*

Forest type, litter cover and elevation were consistently shown to be the most significant environmental factors influencing matsutake distribution. That forest type is important is unsurprising since the presence of matsutake globally is associated with specific tree species from the Pinaceae and Fagaceae (Cao and Yao, 2004). In the study area, *T. matsutake* is generally associated with *Pinus densata* and *Quercus pannosa*. The role of litter cover on matsutake presence is not yet clear. Tree litter components, such as oak leaves, pine needles or bark may act as a carbon source for matsutake development (Vaario et al., 2002), and they are also likely to be important as mulch, maintaining a moist, dark microhabitat. The importance of litter cover to matsutake concurs with statements by

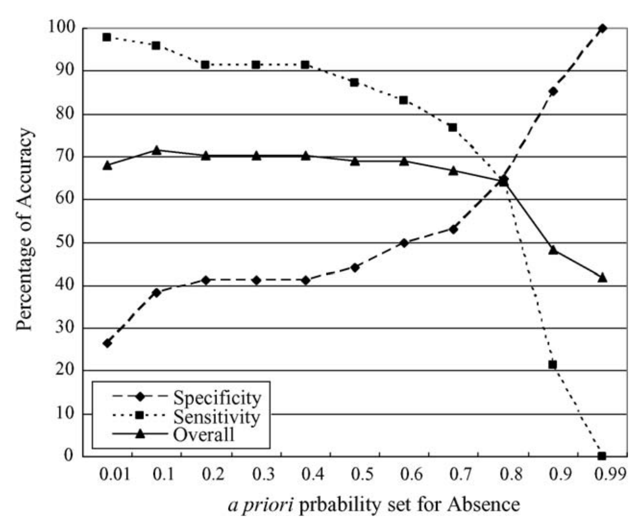

Fig. 4 – The change of specificity, sensitivity and overall accuracy with increase of *a priori* probability set for absence input for GIS expert system (it also implies the decrease of *a priori* probability set for presence, since the sum of *a prior* for absence and presence is always equal to 1).

mushroom pickers that areas where litter has been removed by raking (caused by less experienced mushroom pickers) are relatively unproductive.

Elevation is unlikely to directly influence the distribution of matsutake, but greatly modifies other environmental variables such as temperature and habitat disturbance. Air temperature declines with increasing elevation at a rate of 6 °C per 1000 m (Lookingbill and Urban, 2003)—this will be directly reflected in soil temperatures. There must be an optimum temperature and moisture balance for matsutake, and temperature thresholds could be critical for the development and fructification of the mushroom (Hosford et al., 1997; Kinugawa, 1963). To some extent elevation may also reflect the degree of human disturbance on matsutake habitats. Lower areas are readily accessible and heavily utilized by local people; for instance, villagers traditionally use oak branches and leaves as fuel, animal bedding and farmyard compost. Many oak forests around the villages are noticeably stunted and have formed bushy habits. Similarly, pine forests at low elevation are conspicuously thinned due to timber utilization for house construction. Disturbance does not favor matsutake; this was shown in the statistical analyses where thinner forest canopy was associated with a lower probability of matsutake presence. Despite the obvious importance of forest type, litter cover and elevation, however, there is clearly a complex interaction of variables influencing matsutake, such as the age of host trees, microbial competition (Ogawa, 1976), and dynamic processes including management practices and wild life interaction (Amaranthus et al., 1996).

5.2. *Model performance and its factors*

Strictly speaking the same input data set should be used for model comparison. However, in addition to the three most important layers (forest type, litter and elevation) identified in the initial statistical analysis, the mushroom pickers and local experts insisted that topographic features, slope and aspect also play crucial roles in matsutake distribution; hence, they are also included into the GIS expert system. To some degree this could contribute to the difference of the two outputs.

Both models predict the matsutake distribution with moderate accuracy. Statistically, the GIS expert system gives a slightly better result (approximately 5% improvement) than the logistic regression in overall mapping accuracy, while the Z-test showed that there is no significant difference in the accuracy level of the maps produced by the two methods. However, the spatial distribution of matsutake, as predicted by the two methods, produced somewhat visually different maps (see Fig. 2). The GIS expert system predicts more areas of matsutake absence than the logistic regression model. Which one is correct? The expert system seems subjectively to be more reasonable based on field observation and experience—the areas of difference between the two models (i.e., where logistic regression predicts matsutake presence, but GIS expert does not) are located at elevations greater than 3700 m, where there is little probability of matsutake occurring.

Model performance can be affected by various factors, such as quality of the input data, limitations of the model itself, and the sampling techniques. In our study, errors can be introduced to both models by input layers such as forest type and litter cover mapping. They may also be introduced by the derivation of topographic data sets, for instance, interpolating the DEM from contour line and high points; calculating the slope and aspect from the DEM; and geometric corrections of satellite images.

In terms of model limitations, the GIS expert system depends greatly on the validity of different sources of expert knowledge. When different sources disagree, some form of subjective assessment is necessary. In Jidi, for example, mushroom pickers stated that matsutake prefers a southerly aspect, but the Mann–Whitney *U* test suggested a slightly higher probability on the northerly aspect. The local expert opinion was supported by findings in North America (Hosford et al., 1997) that suggested a southwest aspect is optimal. After weighing the data and using personal field knowledge, we eventually assigned a slightly higher probability to the southerly aspects. However, the integration of knowledge from different sources can be somewhat arbitrary, hence a careful process should be taken into consideration.

The relatively poor performance of the logistic regression could be, in part, caused by the form of the algorithm employed. As used in this study, the algorithm is a transformed linear regression, in which the probability of matsutake presence becomes higher with increasing litter cover and elevation. However, the relationship between matsutake presence and elevation is not linear in nature. From lower to middle elevations, matsutake presence increases with higher elevation, but, above 3700 m, the chances of finding matsutake diminish. While logistic regression overestimates mushroom presence in these high areas, the GIS expert system's "constraint rules" avoid this kind of error. The introduction of a more complex form of logistic regression model with polynomial, quadratic, or spline surface functions would likely reduce this problem. Similarly, there may also be value in exploring the usefulness of a rule based logistic regression such as the GARP model (Anderson et al., 2003). The efficacy of using these techniques is clearly a subject for future work.

The poor performance of both models to predict absence could be due to two reasons: (1) matsutake is particularly prevalent in the study area; (2) identifying absence is much more difficult than presence (see Section 5.4)—thus false information about matsutake absence would contribute to the low specificity.

5.3. *Choosing the best model*

A wide range of methods to model species distribution are available, but, whatever method is employed they all depend upon the relationships between environmental variables and the species considered (Corsi et al., 2000). These relationships can be defined inductively by statistical approaches or deductively by human experts and/or empirical experience. Choosing the most appropriate model can be difficult. Often we select the model with higher mapping accuracy; however, the selection of a model must balance mapping accuracy against user requirements, availability of information, and cost (Smits and Dellepiane, 1999). If the purpose of the study is to provide basic information for matsutake management, then fast, easy and low cost methods will be preferred. In this study, we applied both inductive and deductive approaches. The well established logistic regression model quantified the relationship first, and then extrapolated spatial distribution over the study area using this relationship; by contrast, the GIS expert system used pre-existing expert knowledge to infer the spatial distribution of species. We showed that the GIS expert system produces reliable results and higher mapping accuracy than the logistic regression model.

Not only did the GIS expert system outperform the logistic regression, but the former also requires relatively few samples. For the logistic regression model, two independent sample sets are required, one for developing the model and the other for evaluation, whereas, the GIS expert system only requires an evaluation (or test) sample set. Furthermore, a statistical approach using the regression model requires relatively large quantities of good quality data. First, the number of samples for building the model has to be reasonably high; Huisman et al. (1993) suggested a minimum of 50 samples and 250 samples for a more accurate model. Secondly, a complete data set is essential as missing values hamper the logistic regression model development. Obviously, parsimonious sampling is important as fieldwork is always time consuming and costly, for this reason there is a growing trend to try and predict species distributions from limited datasets without compromising accuracy (Stockwell and Peterson, 2002). This has an important practical implication for management. When adequate samples are not available, incorporating local expert knowledge can help make better-informed management decisions. However, this does not necessarily indicate that the logistic regression is inferior to the GIS expert system. As discussed earlier, the GIS expert system relies on the valid habitat knowledge and is susceptible to human subjectivity of information input.

5.4. *Difficulty of predicting mushroom presence*

Although the logic and methods of modelling mushroom and vascular plant distributions are similar, working with mushrooms is far more problematic. Firstly, the reflectance as measured by remote sensing gives little or no direct indication of the nature of the understorey and ground cover; rather identification of understorey plants or plant communities is based on the deduced relationships with overstorey and topography. There are also sampling limitations with mushrooms because: (a) they are small; (b) vegetative growth is underground; and (c) the duration of the above ground sporocarp is limited. Moreover, sampling only the presence or absence of sporocarps is likely to reject many sites in which hyphae are present. Refining a sampling strategy specifically designed for organisms like mushrooms is desirable, but this will require a better understanding of mushroom ecology.

5.5. *Species distribution modelling for NWFP management*

The difficulties of mapping matsutake are indicative of the challenges of modelling non-wood forest products (NWFP) generally. NWFP tend to be unobtrusive species and/or lower plants. However, there is a need to monitor NWFP as they provide important natural resources (Arnold and Perez, 2001; FAO, 1995). About 80% of the population of the developing world use NWFP for health and nutritional needs, and several million households worldwide depend heavily on NWFP for subsistence and/or income (FAO, 2002). At the same time, the threat to NWFP resources and their rate of extinction are being driven by increasing use, especially for commercial purposes (Schippmann et al., 2002). Distribution maps of NWFP habitat can be integrated with other knowledge (for instance site productivity) to estimate overall production or to assist policy makers to formulate management planning and balanced land access for these natural resources. In the Jidi area, for example, mastutake habitat maps can be used to help design a rotational harvesting plan.

Developing ways to model NWFP is of particular importance in China, which is undergoing rapid large-scale changes in land-use. China is emerging from periods of enforced state directed agricultural land-use to a system of individual responsibility that has given greater freedom and incentive for local people to develop industries. Similarly, the commercial logging ban has forced a redirection of economic and agricultural resources (Yeh, 2000). In many of the poorest regions of southwest China, agriculture is primarily for self-subsistence and unlikely to ever produce income; for these rural communities, the sustainable use of NWFP is essential (particularly matsutake, which is now the most profitable). This sustainability is important, not only for the ongoing livelihoods of these people, but also in terms of preserving many NWFP that are facing rapid depletion and/or extinction.

6. Conclusions

Although the models presented here have clear limitations, this study has demonstrated that it is possible to model mushroom habitat in complex terrain with a reasonable level of accuracy. Models used for predicting distribution of other organism can be applied to mushroom. However, a pragmatic approach is important because modelling something as elusive as fungi is laced with difficulties. Not only is the ground habitat invisible from aerial views, but even ground-truthing with local experts requires a measure of good fortune to find the mushrooms (on the rare occasions when they are above ground). This is where local knowledge and judicious subjective decision making is invaluable. Incorporation of this type of information into a GIS expert system is the first step in creating a parsimonious model of mushroom habitats. Although we found the GIS expert system outperformed the logistic regression model, both in terms of accuracy and sampling efficiency, future refinements with multiple regressions are likely to lead to ever more accurate mapping. These preliminary habitat models provide a platform on which future ecophysiological research can be based that will enable the construction of more refined models, and more importantly, better conservation management.

Acknowledgements

This study was jointly sponsored by the Netherlands Fellowship Program, Knowledge Innovation Program of the Chinese Academy of Sciences (project grant no. KSCX2-1-09-06) and

CEPF as part of a MSc research project undertaken at the International Institute for Geo-information Science and Earth Observation, the Netherlands. Sincere thanks are due to many people for their immense help during the research and in editing the present paper. They are: Mr. Nawang Norbu, Mr. Henk van Oosten, Dr. Jan de Leeuw, Dr. Iris van Duren, Dr. P. S. Roy, Dr. P.K. Joshi, Mr. Andrew Willson, Dr. Horst Weyerhaeuser, Mr. Jun He, Ms. Ming Deng, Ms. Yan Cheng, Mr. Rong-ri Tang and the many local villagers with whom field data was successfully collected. Special thank also goes to Ms. Petra Budde for her distant help in solving problems of GIS expert system; Prof. Steve Walsh, Prof. George Malanson and Mr. Tiejun Wang for their critical comments and inputs during revision.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ecolmodel.2006.04.011.

REFERENCES

Amaranthus, M.P., Pilz, D., Moore, A., Abbott, R., Luoma, D., 1996. American Matsutake (*Tricholoma magnivelare*) across spatial and tempora scales. In: Paper presented at the Forest Soils Biology and Forest Management Symposia, Sacramento, CA.

Anderson, R.P., Lew, D., Peterson, A.T., 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. Ecol. Model. 162 (3), 211–232.

Arnold, J.E.M., Perez, M.R., 2001. Can non-timber forest products match tropical forest conservation and development objectives? Ecol. Econom. 39 (3), 437–447.

Aspinall, R., 1992. An inductive modelling procedure based on Bayes' theorem for analysis of pattern in spatial data. Int. J. Geograph. Inform. Syst. 6 (2), 105–121.

Aspinall, R., 1993. Habitat mapping from satellite imagery and wildlife survey data suing a bayesian modeling procedure in a GIS. Photogrammetric Eng. Remote Sens. 59 (4), 537–543.

Aspinall, R.J., 2002. Use of logistic regression for validation of maps of the spatial distribution of vegetation species derived from high spatial resolution hyperspectral remotely sensed data. Ecol. Model. 157 (2–3), 301–312.

Austin, M.P., 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecol. Model. 157 (2/3), 101–118.

Cao, Z., Yao, Y., 2004. Morphological and biogeographic study on *Tricholoma matsutake*-complex. Mycosystem 23 (1), 43–55 (in Chinese).

Cohen, J., 1960. A coefficient of agreement for nominal scales. Edu. Psychol. Measure. 20, 37–46.

Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sens. Environ. 37, 35–46.

Corsi, F., Leeuw, J.d., Skidmore, A., 2000. Modeling species distribution with GIS. In: Biotani, L., Fuller, T.K. (Eds.), Research Techniques in Animal Ecology: Controversies and Consequences. Columbia University Press, New York, pp. 389–434.

FAO, 1995. Non-wood forest products for rural income and sustainable forestry [on line source]. FAO. Retrieved 8th, July 2003. From World Wide Web: http://www.fao.org/docrep/v9480e/v9480e04.htm#why%20are%20non%20wood%20forest%20products%20important.

FAO, 2002. Impact of Cultivation and Gathering of Medicinal Plants on Biodiversity: Global Trends and Issues [on line source]. Inter-Departmental Working Group on Biological Diversity for Food and Agriculture. Retrieved 8th, July 2003. From World Wide Web: ftp://ftp.fao.org/docrep/fao/005/aa010e/AA010E00.pdf.

Fielding, A.H., Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ. Conserv. 24 (1), 38–49.

Franklin, J., 1995. Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. Progr. Phys. Geogr. 19 (4), 474–499.

Franklin, J., 1998. Predicting the distribution of shrub species in southern California from climate and terrain-derived variables. J. Vegetat. Sci. 9 (5), 733–748.

Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecol. Model. 135 (2/3), 147–186.

Hosford, D., Pilz, D., Molina, R., Amaranthus, M., 1997. Ecology and Management of the Commercially Harvested American Matsutake Mushroom (PNW-GTR-412): United States Department of Agriculture Ecology and Management of Forest Service Pacific Northwest Research Station.

Huisman, J., Olff, H., Fresco, L.F.M., 1993. A hierarchical set of models for species response analysis. J. Vegetat. Sci. 4, 37–46.

Iverson, L.R., Prasad, A., Schwartz, M.W., 1999. Modeling potential future individual tree-species distributions in the eastern United States under a climate change scenario: a case study with Pinus virginiana. Ecol. Model. 115 (1), 77–93.

James, F.W., 1998. Management Experiments for High-Elevation Agroforestry Systems Jointly Producing Matsutake Mushrooms and High-Quality Timber in the Cascade Range of Southern Oregon (General Technical Report PNW-GTR-424). U.S. Department of Agriculture, Pacific Northwest Research Station, Portland.

KIB-CAS, 2003. The list of protected wild plants in Yunnan, P.R. China (in Chinese) [On line source]. Kunming Institute of Botany, The Chinese Academy of Sciences. Retrieved 23th, July 2003. From World Wide Web: http://www.kib.ac.cn/kmbg/yn_list.htm.

Kinugawa, K., 1963. Biological studies related to fruiting of matsutake (Series B). University of Osaka Prefecture, Osaka.

Kranabetter, J.M., Trowbridge, R., Macadam, A., McLennan, D., Friesen, J., 2002. Ecological descriptions of pine mushroom (*Tricholoma magnivelare*) habitat and estimates of its extent in northwestern British Columbia. Forest Ecol. Manage. 158 (1–3), 249–261.

Larsen, H.O., Olsen, C.S., Boon, T.E., 2000. The non-timber forest policy process in Nepal: actors, objectives and power. Forest Policy Econom. 1 (3/4), 267–281.

Lehmann, A., Leathwick, J.R., Overton, J.M., 2002. Assessing New Zealand fern diversity from spatial predictions of species assemblages. Biodiver. Conserv. 11 (12), 2217–2238.

Liu, P., Yan, M., Wang, X., Sun, P., Yang, X., 1999. Notes on the resources of Matsutake-group and their reasonable utilization as well as effective conservation in China. J. Nat. Resources 14 (3), 245–252 (in Chinese).

Lookingbill, T.R., Urban, D.L., 2003. Spatial estimation of air temperature differences for landscape-scale studies in montane environments. Agric. Forest Meteorol. 114 (3/4), 141–151.

Manel, S., Dias, J.-M., Ormerod, S.J., 1999. Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. Ecol. Model. 120 (2/3), 337–347.

Myers, N., Mittermeier, R.A., Mittermeier, C.G., Faseca, G.A.B.d., Kent, J., 2002. Biodiversity hotspots for conservation priorities. Nature 403, 853–858.

Newton, A.C., Davy, L.M., Holden, E., Silverside, A., Watling, R., Ward, S.D., 2003. Status, distribution and definition of mycologically important grasslands in Scotland. Biol. Conserv. 111 (1), 11–23.
Ogawa, M., 1976. Microbial Ecology of Mycorrhizal Fungus-*Tricholoma matsutake* (Ito et Imai) Sing. in Pine Forest. III. Fungal Flora in Shiro Soil and on the Mycorrhizae. The Government Forest Experiment Station, Tokyo, Japan.
Peltoniemi, J.I., Kaasalainen, S., Naranen, J., Rautiainen, M., Stenberg, P., Smolander, H., Smolander, S., Voipio, P., 2005. BRDF measurement of understory vegetation in pine forests: dwarf shrubs, lichen, and moss. Remote Sens. Environ. 94 (3), 343–354.
Pilz, D., Molina, R., 2002. Commercial harvests of edible mushrooms from the forests of the Pacific Northwest United States: issues, management, and monitoring for sustainability. Forest Ecol. Manage. 155 (1–3), 3–16.
Schippmann, U., Danna, J.L., Cunningham, A.B., 2002. Impact of Cultivation and Gathering of Medicinal Plants on Biodiversity: Global Trends and Issues. FAO, Rome.
Skidmore, A.K., 1989. An expert system clasifies eucalypt forest types using Thematic Mapper data and a digital terrain model. Photogrammetric Eng. Remote Sens. 55, 1449–1464.
Skidmore, A.K., Watford, F., Luckananurug, P., Ryan, P.J., 1996. An operational GIS expert system for mapping forest soils. Photogrammetric Eng. Remote Sens. 62 (5), 501–511.
Smits, P.C., Dellepiane, S.G., 1999. Quality assessment of image classification algorithms for land-cover mapping: a review and proposal cost-based approach. Int. J. Remote Sens. 20 (8), 1461–1486.
SPSS Inc., 2001. SPSS for Windows (Version 11.0.1). Chicago, Illinois.
Stephenson, C.M., MacKenzie, M.L., Edwards, C., Travis, J.M.J., 2006. Modelling establishment probabilities of an exotic plant, Rhododendron ponticum, invading a heterogeneous, woodland landscape using logistic regression with spatial autocorrelation. Ecol. Model. 193 (3–4), 747–758.
Stock, M., 1987. AI and expert systems: an overview. AI Applic. 1 (1), 9–17.
Stockwell, D.R.B., Peterson, A.T., 2002. Effects of sample size on accuracy of species distribution models. Ecol. Model. 148 (1), 1–13.
Vaario, L.-M., Guerin-Laguette, A., Matsushita, N., Suzuki, K., Lapeyrie, F., 2002. Saprobic potential of *Tricholoma matsutake*: growth over pine bark treated with surfactants. Mycorrhiza 12 (1), 1–5.
Van On, T., 1993. *Tricholoma matsutake*-matsutake (on line source). New Zealand Institute for Crop & Food Research Ltd. Retrieved 10th, July 2003. From World Wide Web: http://www.crop.cri.nz/psp/broadshe/matsutak.htm.
Vantomme, P., Markkula, A., Leslie, R.N., 2002. Non-wood Forest Products in 15 Countries of Tropical Asia—An Overview. FAO Regional Office for Asia and the Pacific, Bangkok.
Wang, Y., Hall, I.R., Evans, L.A., 1997. Ectomycorrhizal fungi with edible fruiting bodies. 1. *Tricholoma matsutake* and related fungi. Econom. Bot. 51 (3), 311–327.
Wills, R.M., Lipsey, R.G., 1999. An Economic Strategy to Develop Non-timber Forest Products and Services in British Columbia (PA97538-ORE). MOF Research Branch, Victoria, BC.
Xu, J., Ribot, J., 2004. Decentralization and accountability in forest management: case from Yunnan, southwest China. Eur. J. Develop. Res. 14 (1), 153–173.
Yeh, E.T., 2000. Forest claims, conflicts and commodification: the political ecology of Tibetan mushroom-harvesting villages in Yunnan province, China. China Quarter. 161, 264–278.
Zang, M., 1990. A taxonomic and geogrphic study on the Song Rong (matsutake) group and its allied species. Acta Mycol. Sin. 9 (2), 113–127 (in Chinese).
Zaniewski, A.E., Lehmann, A., Overton, J.M., 2002. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. Ecol. Model. 157 (2/3), 261–280.