The Quality of Animal Habitats Estimated from Track Activity and Remote Sensing Data

A. S. Zheltukhin^{*a*}, Yu. G. Puzachenko^{*b*}, and R. B. Sandlerskii^{*b*}

^aCentral Forest State Natural Biospheric Reserve, POB Zapovednik, Nelidovo Raion, Tver Oblast, 172513 Russia ^bSevertsov Institute of Problems of Ecology and Evolution, Russian Academy of Sciences, Leninskii pr. 33, Moscow, 119071 Russia E. maik. mmak@use.m.

E-mail: puzak@orc.ru

Abstract—A method is proposed for estimating the quality of animal habitats from field counts with positioning routes and tracks by means of GPS, multi-channel remote sensing Landsat data, digital elevation model, and discriminant analysis. The distribution of American and European minks is analyzed to demonstrate the principle of choosing an optimal method for analyzing the environmental characteristics that determine the distribution of species and for mapping and estimating the quality of habitats and the probability of track detection. Outlooks and some problems of implementation of the proposed approach are discussed.

DOI: 10.1134/S1995425509030035

Key words: habitat, track activity, remote sensing data, mustelids

The estimation of habitat quality is a traditional ecological problem. Habitat implies a joint action of ecological factors determining the level of population or the state of species population on some territory, i.e., ecological niche [1]. Solving this problem gives an insight into relations between a species and environmental conditions, favors the study of feasible interspecies relations and the management of populations. Usually, the problem is solved both by determining the relatively homogeneous types of habitat state on the basis of vegetation and, less frequently, relief, and by estimating their quality from either the species numbers or the character of their appearance in time. This approach was conceived in the 1930s [2]. In the 1960s, the scales used in hunting inventory allowed the evaluation of habitat quality depending on the stock of forage resources, protective conditions, suitability for nesting, etc. Habitats were commonly distinguished on the basis of taxation plans [3–5]. This logical scheme, based on the available concepts of species ecology, has recently been developed in terms of linguistic variables and qualimetry in the form of mathematical logic [6]. A similar scheme has been widely used in the USA in the last 10–15 years for solving problems of population management [7]. The present work reports a procedure for estimating habitat quality by means of remote multispectral data, digital elevation model, and animal track activity located by GPS. The program BIOCLIM can be considered an analog of this approach but on a regional scale [8].

A GENERAL SCHEME OF INVESTIGATIONS

The development of multichannel remote sensing data and technologies for constructing digital elevation

models opens new ways to study relationships between species and environmental conditions as well as to estimate habitat quality. The multichannel remote information transfers, either directly or indirectly, various environmental properties via relations of reflected solar radiation. It can be used to reproduce phytomass storages, biological productivity, the content of moisture in ecosystem, surface temperature, and energy balance. The reflected radiation spectrum contains information on plant composition, soil characteristics and other properties of the earth's surface, which either directly or indirectly determine environmental conditions for the species under study. In analysis, the environmental characteristics are described using both the original values of reflected radiation in various channels and the indices calculated mainly in terms of the differences and ratios between neighboring channels (Table 1).

It is desirable to use the scenes dated by different seasons of the year. Here, we employed the Landsat 7 images collected on 22 March 2001, 24 April 2000, 20 June 2002, and 27 September 2000. Using the scenes of different seasons makes it possible to take into account the explicit forms of unknown habitat characteristics expressed through the radiance of spectral channels, which affect the distribution of the species under study. The surveys performed in different years, as a rule, bring no bias into analysis because the habitat characteristics remain unchanged for long (except for the cases of natural disasters or significant economic activities).

A digital elevation model estimates hydrothermal regime fluctuations from gradient shape of surface (curvature or Laplacian) and illumination [11]. Since moisture is redistributed at several hierarchical levels of

Table 1. Some spectral i	ndices ([9]) with additions)

Index	Calculation method for Landsat channels	Characteristic		
G/B	=b2/b1	Represent soils and rocks with a high content of iron		
R/G	=b3/b2	Various types of vegetation, aqueous objects,		
SWIR1/G	=b5/b2	wetland		
SWIR2/SWIR1	=b7/b5	Argillaceous deposits and rocks rich in clay		
SWIR2/R	=b7/b3	Roads, populated lands, fields, and other anthropogenic objects		
DVI	=b4-b3	Photosynthesis activity, net production,		
RVI (SR)	=b4/b3	transpiration, types of vegetation		
NDVI	=(b4-b3)/(b4+b3)			
GreenNDVI	=(b4-b2)/(b4+b3)			
TVI	$=100 \left[(b4 - b3)/(b4 + b3) + 0.5 \right]^{1/2}$			
SARVI2	=2.5((b4-b3)/(1+b4+6b3-7.5b1))	Productivity with atmospheric noise correction		
NDSI	=(b1-b4)/(b4+b1)	Sensitive to snow and ice thickness		
LMI	=b5/b4	Moisture content in green phytomass		
NDWI	=(b5-b4)/(b5+b4)	The same		
Kauth's Tasseled Ca	p transformation			
BR (radiance)	= 0.33183b1 + 0.33121b2 + 0.55177b3 + 0.42514b4 + 0.48087b5 + 0.25252b7	Overall radiance, albedo		
<i>GR</i> (green color)	= -0.24717b1 - 0.16263b2 - 0.40639b3 + 0.85468b4 + 0.05493b5 - 0.11749b7	Photosynthesis activity, net production		
WET (humidity)	= 0.13929b1 + 0.22490b2 + 0.40359b3 + 0.25178b4 - 0.70133b5 - 0.45732b7	Moisture content in green phytomass		
Energy characteristic	s of landscape [10]			
E_{λ}^{in}	$= ESUN_{i} \cos \theta_{s} / \pi d^{2}$	Input solar energy to the earth's surface		
	<i>ESUN</i> , mean solar exoatmospheric irradiances (solar constant)			
	<i>i</i> , spectrum zone			
	θ_s , solar zenith angle			
	<i>d</i> , earth–sun distance			
E_{λ}^{out}	At-sensor radiation	Reflected radiation		
R	$=\sum_{i=1}^{6} (E_{i}^{in} - E_{i}^{out})$	Absorbed radiation		
Т	$= \frac{K2}{\ln(K1/b6+1)}$.	Temperature of earth surface		
-	K^2 and K^1 , calibration constant			
A	$= (E^{in} / E^{out}) \times 100\%$	Albedo		
K	$= \sum_{i} p_i^{out} \log(p_i^{out} / p_i^{in}), \text{ energy fraction in the } i\text{-th channel of the total radiation at (reflected from) the surface}$	Kulbak's enthropy—the measure of structural complexity of the system which reflects (E^{out}) and absorbed energy (R)		
Ex	$= E^{out}(K + \log(E^{out}/E^{in})) + R$	Exergy, the part of energy which enters the system and can perform useful work to maintain the structure (organization) of the system		
ExNDVI	$=((b4-b3)/Ex)\times 100\%$	Efficiency of biological productivity (photosynthesis efficiency)		

Note: Latin letter b and figures denote the numbers of Landsat 7 survey channels.

relief (micro, meso, and macro), morphological parameters are estimated for each of them. The hierarchical relief levels are determined by two-dimensional spectral analysis [12].

As a result, the characteristics of each grid cell of the territory (pixel) are represented on a chosen scale through multispectral remote data and relief properties.

Animals are counted by their track activity detected by GPS at either regular or one-time routes. The tracks of each species are recorded as dots in geographical coordinates. The records are transferred into MAPINFO, where they are represented as routes with dots of "absence"(0) and "presence" of *i* tracks (i = 1, 2, 3, ..., n). The size of a dot is equal to a pixel area corresponding to resolution in the field used in remote sensing. The Landsat data with a resolution of 28.5 m satisfy the purposes of these investigations best of all. In a particular case, one dot can stand for several tracks, which is documented in field papers and MAPINFO database. Locating observation sites and the entire route becomes increasingly more precise with the appearance of new GPS models. However, in forests of high density, the accuracy decreases and GPS records can slightly deviate from the real route. The deviation does not usually exceed 30 m and corresponding corrections are introduced by MAPINFO. The Landsat data and relief characteristics provided by the digital elevation model are kept in the Access database in a geographical system of coordinates. The data organized in MAPINFO are copied into the database and added to the data characterizing environment by geographical coordinates. Thus, we obtain a sample with either the "presence" or "absence" of tracks at each point whose definition includes Landsat spectral radiance, calculated indices and relief characteristics.

There are two variants of providing data for analysis: (i) each count, irrespective of its location, is introduced with its own ratio of the "presence" and "absence" of tracks; and (ii) in counts performed at regular routes the "absence" is assigned only to the points at which no tracks of a given species have been detected. When a species was recorded twice and more at one point, this point is multiplied in the data under analysis as many times as necessary. The second scheme assumes that if a track is recorded at one point at least once, this habitat is suitable for the given species. The more often a habitat is visited by animals, the better it is. The second scheme is good for estimating the quality of habitats, and the first one is useful for counting the numbers from the mean duration of a daily variation.

METHODS OF DATA ANALYSIS

Let us assume that, with some probability, tracks are related to the sites that are most favorable for animal activity. Of course, with few exceptions, the relationship between the detection of tracks and environmental conditions is not strictly deterministic. Thus, an optimal method for solving the problem is a method based on the verification of the difference between the actual distribution of track "presence-absence" data and the independence zero hypothesis. A stepwise method of discriminant analysis satisfies these requirements. It is conceptually close to the variance analysis, but can be used to construct a statistical model to describe relationships between the two states and environmental conditions. This allows us to interpolate the results obtained using a learning sample (field observations) and environment parameters estimated from remote sensing data and a digital elevation model. The method is given in detail in many handbooks of statistics as applied to problems of ecology [13]. A particular example should be given to consider details of the method's application.

MATERIAL AND AREA UNDER STUDY

The studies were performed in the southern part of the Central Forest State Natural Biospheric Reserve (56°26'-56°31'N and 32°29'-33°01'E) situated on the southern Valdai Hills within the boundaries of southern taiga and mixed forests. The area under study includes the upper part of the Mezha River, a tributary of the Western Dvina. The relief is composed of morainic ridges rising 40-60 m above intermorainal lows. Figure 1 shows a digital elevation model and Fig. 2 is a general scheme of major types of vegetation cover with regular routes. The present work considers the counts performed in January-March, 2006. The total length of regular routes is 42.135 km, and 441.75 km were covered during three months of counts. The workers of the reserve S. Topaly, S. A. Zheltukhin, and V. V. Kochetkov, took part in these counts.

DATA ANALYSIS

As an example, the quality of mink habitats is examined. The mink is strictly associated with a river net, which makes it possible to estimate the adequacy of the method. The American and European minks are distributed over the territory of the reserve. Their tracks do not discriminate between the species; therefore, they can be analyzed only simultaneously.

A stepwise discriminant analysis usually involves two methods, "forward" and "backward". The forward method implies that first a variable is taken which is best in identification of discriminated patterns. Next variable is added to attain the best discrimination after addition to the first one, etc. The background method estimates discrimination by all variables and the variables whose contribution to discrimination is doubtful are sequentially discarded. This method is good for in representing the multiplicative effects of variables in the discrimination of the pattern. Both versions can be realized in two ways: with the states introduced into the learning sample either in accordance with the observed occurrences or randomly (with equal probability). In the latter case, the rare events ("the presence of tracks")



Fig. 1. A digital elevation model based on a topographic map on a 1:10 000 scale.



Fig. 2. Main types of landscape cover.

ZHELTUKHIN et al.

				Fraction	The number of events	
Variables	Method	Condition	State	of correct distinguishing, %	absence	presence
All	Forward	Estimated	Absence	98.43	1451	23
			Presence	61.29	12	19
			On the whole	97.67	1463	42
		Equal	Absence	95.25	1404	70
			Presence	83.87	5	26
			On the whole	95.01	1409	96
	Backward	Estimated	Absence	98.37	1450	24
			Presence	58.06	13	18
			On the whole	97.54	1463	42
		Equal	Absence	96.74	1426	48
			Presence	83.87	5	26
			On the whole	96.47	1431	74
Relief	Forward	Estimated	Absence	97.28	1434	40
			Presence	61.29	12	19
			On the whole	96.54	1446	59
		Equal	Absence	92.26	1360	114
			Presence	90.32	3	28
			On the whole	92.22	1363	142
	Backward	Estimated	Absence	97.55	1438	36
			Presence	64.51	11	20
			On the whole	96.87	1449	56
		Equal	Absence	93.14	1373	101
			Presence	83.87	5	26
			On the whole	92.95	1378	127
Landsat	Forward	Estimated	Absence	99.32	1464	10
			Presence	51.61	15	16
			On the whole	98.33	1479	26
		Equal	Absence	90.70	1337	137
			Presence	74.19	8	23
			On the whole	90.36	1345	160
	Backward	Estimated	Absence	99.18	1462	12
			Presence	58.06	13	18
			On the whole	98.33	1475	30
		Equal	Absence	94.36	1391	83
			Presence	80.64	6	25
			On the whole	94.08	1397	108

Table 2. The quality of identification of mink	"presence-absence" trac	cks in four variants of	discriminant analysis
---	-------------------------	-------------------------	-----------------------

are better distinguished. However, the number of "no track" events erroneously assigned to the "presence" may increase. As a result, the interpolative prediction extends the area inhabited by a species.

As follows from Table 2, only 31 mink tracks were recorded over the entire period of counting with a total number of spots without tracks of 1474. According to the Landsat resolution, a linear size of the spot is 28.5 m. A spot is taken as an elementary unit of count-

Variable	Method	Weight (influence)	Correlation	χ^2	Number of degrees of freedom	Ratio	F-criteria
All	Forward	0.23	0.4693	318.67	13	24.51	27.21865
	Backward	0.28	0.4371	370.33	30	12.34	13.8847
Relief	Forward	0.15	0.3665	216.15	9	24.01	25.77577
	Backward	0.16	0.381	234.52	16	14.65	15.79594
Landsat	Forward	0.12	0.3323	175.6	6	29.16	31.00887
	Backward	0.18	0.3931	250.56	22	11.36	12.31818

Table 3. Estimation of the quality of various variants of the model

ing. Seemingly, with the "presence" events being so rare, statistical analysis is improper. However, the chi-squared test is 318.6749 for the "forward" model with the number of degrees of freedom of 13 and the *F*-criterion for class discrimination of 27.22, and is 370.3382 for the "backward" model with the number of degrees of freedom of 30 and F of 13.88 (Table 3). Thus, both models are statistically valid. Although the number of detected tracks is not large, their location is well determined by environmental conditions. It is worth noting that these relationships are characteristic for stenotopic species. The number of the degrees of freedom indicates that the "backward" model includes 31 determining variables and the "forward" one includes 14. A chi-squared mathematical expectation equals the number of degrees of freedom; the qualities of the models are convenient to be compared by their ratio. The difference in identifying the states by discriminating variable is negligible for both models. However, the "forward" model is somewhat better. Combining various variables, we may study the role of various types of habitat characteristics in the distribution of species. Thus, for minks, the relief characteristics, with the Landsat data not taken into account, are good for describing the distribution of animals but, compared with a general model, they assign the "presence" state more frequently. As the Landsat spectral channels reflect the properties of vegetation cover rather than those of relief, we conclude that the effect of vegetation on minks is weaker than that of relief which increases the probability of the "absence" state.

It is worth noting that criteria for quality are not all identical in various models. The "backward" model based on remote sensing data is better than the same model based on relief characteristics, but is worse in terms of F criterion and ratio of chi-squared criterion to the number of degrees of freedom. In general, the relief characteristics provide more correctly discriminated states than remote sensing data. A close relation between the mink and the relief requires no comment and, more importantly, it agrees with results of formal analysis.

The discriminant analysis provides two variants of the predicted values for the classes of states (conditions) for the model interpolation over the entire territory, the probability of correct assignment of each pixel to a corresponding class and the relevant factors. The factor can be considered a habitat quality as perceived by the species. Appropriate maps can be drawn by transforming the table file format into a raster by means of MAPINFO. Figure 3 is a scheme of habitat quality (a virtual factor). The scheme indicates that the best mink habitats are related to erosional features. The details of the relationship between minks and environment can be estimated from the correlation of the factor of habitat quality with the given variables describing the environment (Table 4). It is well known that the mink habitats are associated mainly with brook and river valleys. Therefore, the factor of habitat quality correlates negatively with the height of the relief at all levels. In general, the distribution of minks is determined by the character of macro- and mesorelief and, in all cases, the mink prefers slightly inclined and convex slopes, i.e., terrace-like valleys. The mink prefers the relief covered by the vegetation whose productivity is low in spring and fairly high in summer and microclimate which is coolest in spring and fall. These conditions are characteristic for forest valleys covered by firs where humidification is in excess.

The probability of track detection is unambiguously related to the influencing factor (Fig. 4) and can be depicted in the map which estimates the habitat quality and represents equally a probable distribution of minks over the territory and their potential numbers (Fig. 5). In the given scheme, this probability applies to about four visits of the same spot. Thus, analysis of tracks reveals environmental preferences of minks and their possible distribution.

DISCUSSION

The proposed method of counting by means of GPS and by representing habitat characteristics via relief properties and remote sensing multispectral data with subsequent estimation of occurrence through discriminant analysis allows one to:

 interpolate observation results into the territory with similar landscape conditions where the relationship between a species and environmental conditions remains almost constant;

Туре	Variant	Variable	Correlation	
Relief	On the whole	Height	-0.55	
		Concavity	-0.21	
	Macro	Height	-0.41	
		Slope	-0.21	
		Concavity	-0.41	
	Meso	Height	-0.52	
		Slope	-0.16	
		Concavity	-0.42	
	Micro	Height	-0.51	
Landsat	Temperature	April	-0.34	
		September	-0.20	
	Biological productivity (<i>NDVI</i>)	April	-0.35	
		May	-0.24	
		June	+0.15	

Table 4. The main environmental properties determining the mink habitat quality

- estimate the main properties of the environment that determine species distribution over the territory;

- quantitatively estimate the environment for particular species in terms of a generalized factor;

- estimate the probability of species occurrence within the entire interpolation area.

In each case, corresponding maps can be drawn at a scale determined by satellite resolution.

In this report our approach is exemplified by the species for which the efficiency of the method can be estimated by any ecologist. Fairly reliable results have been obtained for other species (squirrel, white hare, forest marten, fox, ermine, weasel, and hazel grouse).



Fig. 3. Estimation of environmental quality by factor obtained from discriminant analysis. Black, habitats unsuitable for minks; white, the best habitats.

Errors can be caused by inaccuracy both of the geospatial reference of remote sensing data and relief and of the GPS location of spots. Using known procedures, we can eliminate these errors, but often it is not necessary, as with a proper sample volume their influence is actually negligible. An essential source of bias is the method of discriminant analysis based on models of linear algebra and hypothesized normal distribution. By means of statistical estimation, we can determine a potential scale of bias and reveal observed bias that contradict the hypothesis of normality and linearity. For ecologists, the observed bias of the "normal" behavior of the subject under study is of particular interest.

On the other hand, the relationship between a species and environment recorded in various seasons of the year, in different years, and under various macrogeographical conditions can be different and are to be described by different statistical models. Clearly, various seasons of the year and in different years should be analyzed to study a time-dependent relationship between a species and environmental conditions. It is more difficult to identify the territories with different relationships between a species and apparently similar physiognomic environmental conditions, particularly where they are part of the area under study. The combination of populations with two different types of relationships between species and the same environmental conditions is a combination of various parent communities (entities). The effect of this joining can be studied in terms of statistics. However, in the general case, one must be careful when interpolating occurrence and estimating the habitat quality obtained from a local sample for vast territories omitted from counts. More powerful methods of analysis, e.g., the method of neural networks, as compared with a relatively simple method of discriminant analysis, fail to provide any principally new results.



Fig. 4. Relationship between the probability of track detection and the "Habitat quality" factor. The model is given at a frequency of track "presence-absence" in a learning sample; *I*, from observations; *2*, with equal probability.

When the aforementioned methods are used to analyze vegetation, soil, and other features of landscape, an abstract representation of environmental characteristics



Fig. 5. The probability of mink track detection after four visits. The distribution of "presence-absence" tracks is given: *1*, from observations; *2*, with equal probability.

via multispectral survey may be substituted by such characteristics as forest height and density, the species composition of vegetation, soil properties, etc., which is sure to offer a more detailed description of the properties of the habitat of the species under study. However, a special complex of appropriate ecological investigations should be performed. The proposed approach requires, of course, a special knowledge and software. At present, fairly reliable digital elevation models with a resolution of 90 m for any territory in Russia are available through the Internet (a digital elevation model with a 90 m resolution-The NASA Shuttle Radar Topographic Mission (SRTM) http://srtm.csi. cgiar.org/). Most of the Landsat scenes are also accessible (Catalogue of Landsat scenes in free access and digital relief models SRTM-The Global Land Cover Facility (GLCF) http://glcf.umiacs.umd.edu/index.shtml). However, they can be located with reference to the territory and transferred into a format convenient for statistical analysis using either the package of the programs ERDAS, ENVI or their analogs. Cartographic information and GPS data can be treated by means of the geoinformation system MAPINFO. The hierarchical levels of relief can be distinguished by programs involving a two-dimensional spectral analysis and constructions of surfaces in a given band of spatial occurrences (Olaf Conrad, the program for computing morphometric characteristics of relief DiGem 2.0, 2002, http://134.76.76.30/SAGA/DiGeM/download/ digem.zip; the program of spectral analysis of images ImageJ 1.37 (National Institutes of Helath, USA, http:/rsb.info.nih.gov/ij/). In addition, the programs of statistical analysis either Statistica 7 or SPSS should be available. Besides, a good professional level is needed to take advantage of these technical means. This procedure is quite a challenge, but all problems can be solved. At present, a wide application of remote sensing data and the methods of spatial analysis is one of the most important lines of ecology development.

The work was financially supported by the Russian Foundation for Basic Research (grant no. 06-04-48536).

REFERENCES

- J. Grinnell, "The Niche-Relationships of the California Thashet," in *Foundations of ecology. Classic papers with Commentaries*, Ed. by L. A. Real and J. H. Brown (The University of Chicago Press, Ltd., London, 1991) pp. 118–125.
- A. A. Knize and V. A. Leont'ev, *The Basic Problems of Taxation* (Moscow–Leningrad, 1934) [in Russian].
- 3. Yu. G. Puzachenko, in *Hunting and Hunts* (1964), pp. 10–11 [in Russian].
- 4. D. N. Danilov, in *Foundations of Hunting* (Izd. Tzentrosoyuza, Moscow, 1966) [in Russian].
- 5. V. A. Kuzyakin, in *Hunting Taxation* (Lesnaya Promyshlennost', Moscow, 1979) [in Russian].
- Yu. G. Puzachenko, in *Ecoinformatics*, Ed. by V. E. Sokolov, (Gidrometeoizdat, St. Petersburg, 1992), pp. 7– 131.

- R. T. Milhous, M. A. Updike, and D. M. Schneider, "Physical Habitat Simulation System," in *Reference Manual—Version II. Fish and Wildlife Service Biologi cal Report, 1989, 16* (Washington, DC, 1989), pp. 1– 403.
- J. R. Busby, C. R. Margules, and M. P. Austin (Eds.), in Nature Conservation: Cost Effective Biological Surveys and Data Analysis (CSIRO, Canberra, 1991), pp. 64–68.
- F. Moreno Jose, M.-C. Gonzalez, L. Alonso, et al., Analysis, Investigation and Monitoring of Water Resources for the Management of Multi-Purpose Reservoirs. Survey of Remote Sensing Data Analysis Methods (Contract no ENV4-CT98-0740 version 1.0, WP 3100-EO Signal Analysis Methods), http://dataserv.cetp.ipsl.fr/ AIMWATER/reports/EOsurvey report.pdf.
- S. V. Jorgensen and Y. M. Svirezhev, *Towards a Thermodynamic Theory for Ecological Systems* (Elsevier Ltd., Oxford, UK, 2004).
- 11. V. V. Sysuev, Izv. RAN. Ser. Geogr. No. 4, 36 (2003).
- 12. Yu. G. Puzachenko, I. A. Onufrenya, and G. M. Aleshchenko, Izv. RAN, No. 4, 29 (2003).
- Yu. G. Puzachenko. Mathematical Methods in Ecological and Geographical Investigations (ACADEMA, Moscow, 2004) [in Russian].