

Modeling Sage Grouse: Progressive Computational Methods for Linking a Complex Set of Local, Digital Biodiversity and Habitat Data Towards Global Conservation Statements and Decision-Making Systems

Anthonia Onyegahialam¹, Falk Huettmann², and Stefania Bertazzon¹

¹ Department of Geography, University of Calgary, Calgary Canada
t_onyeahialam@yahoo.com, bertazzs@ucalgary.ca

² EWHALE lab, University of Alaska-Fairbanks, Fairbanks, USA
fffh@uaf.edu

Abstract. Modern conservation management needs to link biological questions with computational approaches. As a global template, here we present such an approach from a local study on sage grouse breeding habitat, leks, in North Natrona County, Wyoming, using remote sensing imagery, digital datasets, spatial statistics, predictive modelling and a Geographic Information System (GIS). Four quantitative models that describe sage grouse breeding habitat selection were developed for multiple scales using logistic regression and multivariate adaptive regression splines (MARS-Salford Systems). Based on candidate models and AIC, important habitat predictor variables were elevation, distance to human development, slope, distance to roads, NDVI and distance to water, but not Sagebrush. Some predictors changed when using different scales and MARS. For the year 2011, a cumulative prediction index approach is presented on how the population viability of sage grouse can be assessed over time and space using Markov chain models for deriving future landscape scenarios and MARS for species predictions.

1 Introduction

Complex computations and advanced statistics play an important role for our daily lives. Biodiversity and habitats contribute to human well-being, but as well to economics and wealth [1]. Often, it is not clear to the general public that these complex subjects are linked, and how computing, quantitative methods, biodiversity and habitat data sets, biology and geography are connected towards a sustainable future [2], [3], [4]. We use the sage grouse (*Centrocercus urophasianus*) – a species of North American conservation concern - as an example how these research disciplines, with each method used here being at the forefront of the individual research discipline, can get merged at a local scale. As an outlook we show how such an approach can get applied globally. We believe that the methods presented here are of relevance to global management issues of biodiversity, wildlife and habitats, as well as to the well-being of humans.

2 Methods

Making informed decisions on the conservation of sage grouse requires the full understanding of its life history characteristics, preferences, limiting factors as well as its dependence on sagebrush vegetation. It also requires an evidence and an understanding of its habitat selection patterns and any observed underlying processes. This will aid in the identification and modelling of its habitats and how landscape change can affect this species with the goal of providing adequate and timely information for the species management. Approximately 736,000 km² of sagebrush vegetation types existed in North America [5], making it one of the most widespread habitats in North America. Unfortunately, much of this habitat has been lost or degraded over the last 100 years [6].

2.1 Biology Data

Sage grouse is an endangered species in North America. It can be found in sage (*Artemisia spp*) habitat, which usually is rangeland used by cattle but which can be increasingly diminished through other land uses such as real estate, industrial activities and road construction. Sage grouse mate at lek sites, which are crucial habitats for their reproduction; nests are usually found related to lek sites [7],[8],[9], [10]. For the study area, yearly lek site surveys were carried out during the years 1979-2001, resulting into a maximum of 11 detected and subsequently geo-referenced leks.

2.2 Remote Sensing and Other Habitat Data

A Landsat TM image was obtained 1985 for the study area. A second image, ETM+ Landsat 7, was obtained for summer July 2001. This image was processed and classified using approaches described in [11], [12], and resulting into an overall accuracy of 85%. Computations for deriving NDVI and Greenness were applied. Other data such as DEM (including derived slope and aspect), roads and water features were also available to the project (see [11] for earlier work).

2.3 Spatial Statistical Analysis

All data were imported into the Geographic Information System (GIS) ArcView 3.3. and ArcGIS 8.2 for further processing and analysis.

Dynamic population level processes with spatial characteristics can create patterns relating locational attributes to them. These spatial patterns are the result of underlying processes and these processes can be described, measured, and evaluated using various spatial descriptive statistical methods. The observed patterns can then be related back to the ecology of the phenomenon in question and be used for predictive purposes. Point pattern analysis and spatial autocorrelation [13],[14],[15], were used to explore the distribution of observed sage grouse breeding habitats in North Natrona County, Wyoming. We examined if an underlying pattern exists in sage grouse breeding habitats.

The nearest neighbourhood index was estimated for describing the sage grouse breeding habitat patterns using a nearest neighbourhood index script in the Arcview

3.2 wGIS software. The analysis is based on the calculated average of all distances between each pair of points representing sage grouse breeding habitats using Pythagoras theorem. For the purposes of comparison, the calculated average nearest neighbourhood distance is compared to an expected average distance between nearest neighbours. The expected average distance between nearest neighbours is standardised to account for area coverage in a random point pattern.

According to [14] autocorrelation is a general statistical property of ecological variations observed across geographic space. [15] define spatial autocorrelation to measure the degree to which a spatial phenomenon is correlated to itself. Moran I and Geary's C indices are types of spatial autocorrelation measures. Moran's I represents the overall agglomerative patterns (i.e. are events clumped or dispersed ?) whereas Geary's C explains the similarity or dissimilarity of events [14]. The spatial autocorrelation of sage grouse breeding habitats was estimated using the above-mentioned metrics provided by the S-PLUS extension in the Arcview 3.2 GIS software.

2.4 Modeling Habitat Relationships

Based on known lek sites, we created 'presence/absence' locations. Following established methodology this was done using pseudo-absences, random locations. As outlined in [2] first we used Generalized Linear Models (GLM) to build a model [16],[17]. For predictors we choose Land use, Distance to water, Distance to roads, Distance to human development, NDVI, Greenness, Elevation, Slope and Aspect.

Model selection was done with a set of 83 candidate models and AIC; this method follows [18], [19] and was based on a modified S-PLUS code by [20].

Non-linear models are powerful inference and prediction tools. They provide great alternatives, and often improvements, when compared with GLMs. We decided to use the MARS-Salford (Multivariate Adaptive Regression Splines) algorithm due its speed, convenience and general accuracy [21]; similar algorithms such as Cart (Classification and Regression Trees), Neural Networks and TreeNet could also have been used [22], [23], [24].

We build models on two scales since it is known that the scale of study can affect the inference [25] and prediction. We choose the point scale, as well as the home range scale for each lek site in order to assess the scale effects. Home range size was determined from the literature and centered on lek sites.

2.5 Future Landscapes and Sage Grouse Predictions

Using the Markov [26], [27], [28] module in the IDRISI 32 software, multiple iterations were run on land use maps classified from Landsat 1985 and Landsat 2001 imageries in order to predict land cover changes for the next ten years, 2011 (see also [29] for other approaches to obtain cumulative effects landscape scenarios of the future). From the predicted distribution map of lek sites 2001 we computed a cumulative prediction index of lek site occurrence in the study area. This cumulative index for the study area was equaled with the number of known breeding sites. We used MARS since it allows for a convenient, fast and reliable modeling of the future distribution of sage grouse. Using the future landscape of 2011 as input into the 2001 MARS model, we then predicted lek sites spatially. Finally, we computed the cumula-

tive prediction index for 2011, as an indicator of how many breeding sites would exist in the future in the study area for the future landscape.

3 Results

3.1 Landuse Classification

Figure 1 shows a reclassified result of a hierarchical image classification approach. Overall, thirty seven vegetation and land use classes exist. The results of the reclassification with 11 classes were used for simplicity in the following model development.

3.2 Spatial Patterns

The results of the nearest neighbourhood and quadrat analysis, Geary's C and Moran's I indices indicate that the location of sage grouse breeding habitat followed distinct patterns. Thus, sage grouse does not seem to select the location of its breeding habitat randomly, but certain characteristics that are favourable to its breeding habitat are considered at every location. The results of the exploratory process set the stage for further analysis using other statistical techniques for modelling and predicting the state of its present and future habitats.

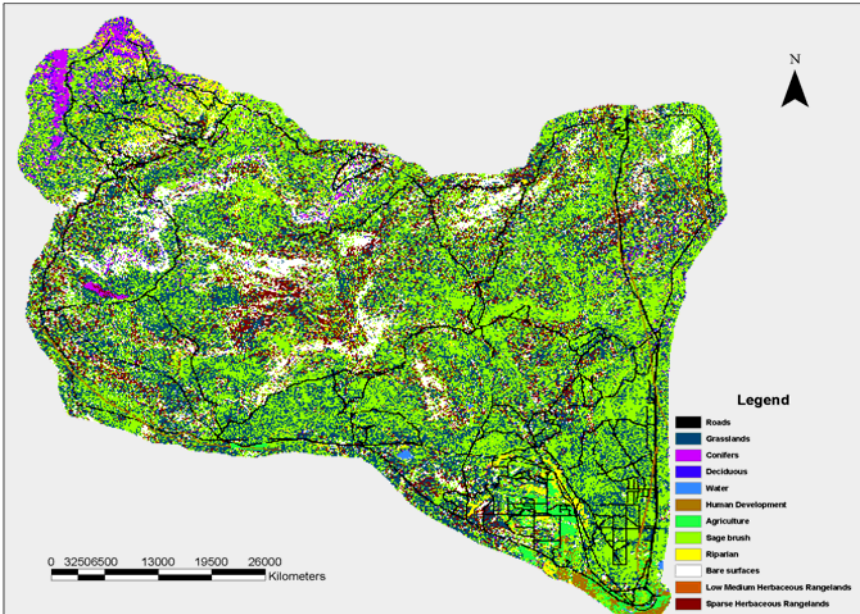


Fig. 1. Landcover classification from using 11 vegetation classes [7]

3.3 Spatial Prediction

For the GLMs, from the 83 candidate models we identified a final model which was within $5 \pm AIC$ units. This was done for two scales, lek site and home ranges. Based on expert knowledge, we identified the homerange model as more accurate due to better predictive accuracies and higher variances explained. Model coefficients are presented in Equation 1.

$$\ln(1/1-p) = -3.70503 - 0.56858 (\text{Slope}) + 0.00007 (\text{Dist. from Human development}) + 0.0030287 (\text{Elevation}) + 0.0000101 (\text{Dist. to water}) - 0.000901 (\text{Dist. to Roads}) - 4.96345 (\text{NDVI})$$

Equation 1

When using MARS on the home range scale, only three predictors are selected (Table 1). However, it should be remembered that same as with the GLM, these importance values are driven by the characteristics of the algorithm and GIS, but not necessarily by the true biological needs of the sage grouse [2].

Table 1. Importance of predictor variables in MARS (Home range scale)

Variable	Cost of Omission	Importance	Scale of Importance
Slope	0.078	100.000	
Distance to Human Develop.	0.077	82.111	
Distance to Roads	0.074	37.210	
Distance to water	0.074	0.000	
NDVI	0.074	0.000	
Vegetation Classes	0.074	0.000	
Greenness	0.074	0.000	
Elevation	0.074	0.000	
Aspect	0.074	0.000	

Sagebrush vegetation has been reported to be prominent in the life of the sage grouse. However, the results of the most appropriate model developed for predicting sage grouse breeding habitats using logistic regression as well as MARS and on multiple scales show that sage brush vegetation as such is not the only and key predictor variable that determined the location of sage grouse breeding habitats in North Natrona County. Terrain characteristics, proximity to sources of water, roads and human development play prominent roles in determining the occurrence of sage grouse breeding habitats in the area. This provides new insights into habitat features relevant for lek sites, and likely for related nest sites. NDVI, which describes the richness and vigour of the vegetation, was the only predictor variable in the model on multiple scales from logistic regression that gave information about the characteristics of the vegetation cover. Candidate models developed with logistic regression where sagebrush vegetation played a role as a predictor variable ranked between the 10th and 20th models across scales. This suggests that other identified habitat predictor variables should be more important in predicting the species habitat in North Natrona or other areas where sage brush is the major land cover type.

Applying the above home range formula with MARS for a prediction of lek sites to the overall study area for the year 2001 indicates that 28 of these breeding sites would exist.

3.4 Landcover Change and Future Scenario

Figure 2 represents the future distribution of lek sites in the landscape of North Natrona County, Wyoming, for the year 2011, based on the combined Cellular automata / Markov chain land cover prediction procedure. The results indicate a more fragmented landscape scenario for 2011, and the extent and distribution of sagebrush and herbaceous rangeland vegetation is lower as opposed to human development (table 2; map not shown here). Without ground-truthing information, the reliability of this classification cannot be statistically ascertained but likely presents the general trend.

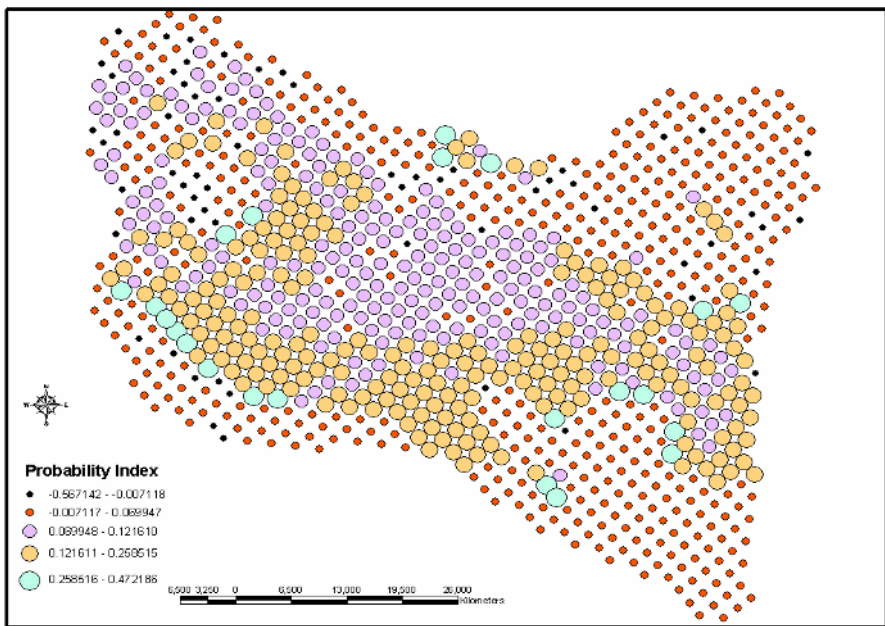


Fig. 2. Predicted future distribution of sage grouse lek sites for 2011

For the year 2011 a viability of the sage grouse population was estimated (compare [30]). Results suggest that an increase in the cumulative predicted probability index was experienced between 2001 (28 predicted lek sites) and 2011, increasing from 73.79 to 79.47. At a cumulative predicted index value of 79.47, an estimated value of 30 sage grouse breeding sites can be supported. However, the spatial distribution of these predicted sites in Figure 2 shows that they are mostly concentrated in the central portion of the study area, which coincides with predicted vegetation types such as sagebrush and rangelands reported to be favoured by sage grouse. Such a distribution might facilitate contact among individuals but also increases competition; it can pose

the threat that all relevant areas are concentrated into one single area making populations vulnerable if this one habitat would be majorly disturbed.

Table 2. Changes in the Class Area Landscape Metric of Vegetation and Land Use Classes from 2001 to 2011

Vegetation and Land Use Classes	Class Area (2001) (ha)	Class Area (2011) (ha)	Change (ha)
Roads	20522.97	23719.05	3196.08
Grasslands	115747.1	68422.86	-47324.3
Conifers	6773.31	29805.75	23032.44
Deciduous	4166.28	18944.55	14778.27
Herbaceous Rangelands	40155.57	25304.13	-14851.4
Water	399.42	36790.65	36391.23
Human Development	6234.21	32495.4	26261.19
Agriculture	2495.25	8562.87	6067.62
Sage brush	148235	100193.58	-48041.5
Riparian	17374.14	22941.36	5567.22
Bare surfaces	42060.78	36994.95	-5065.83

4 Discussion

Assessing biodiversity components and carrying out progressive and pro-active conservation research and management can require intensive digital data preparation and computationally demanding applications. Usually, this is done with the extensive use of modeling and statistical soft- and hardware and advanced (spatial) datasets; see [31], [32],[33] for change detection modeling and [29] for a future Landscape example. Much more multidisciplinary expertise and international projects are needed to address issues relevant for sustainable landscape management on larger scale. This localized project and its methodology can be applied on a larger, continental and global scale, as well as with other species. It is made possible through the advent of large scale data (Sagemap USGS website: <http://sagemap.wr.usgs.gov.htm>, Wyoming State clearing house: <http://www.wygisc.uwyo.edu/clearinghouse/county.html>) as well as through global data sets such as Landcover and Human Populations (SAGE and HYDE [34]), Climate Data (CRU[35]), Digital Elevation (etopo2 [36]), and others. These global aspects to species and biodiversity modeling should be enforced much stronger. First approaches are coming forward for the globe as well as for world-wide marine and terrestrial applications (e.g. Neubauer and Huettmann unpublished, Meyer and Huettmann unpublished).

It is well known that habitats are modified, transferred and lost globally at an increasing rate [37]. For the study area, table 2 shows that roads, conifers, deciduous, human development, agriculture, riparian cover classes, which are not favorable for the preservation of sage grouse breeding habitats, increased. This trend is in contrast

to herbaceous rangelands, grasslands, sagebrush and bare surface classes which sage grouse is more associated with and favorable to the species [38].

The presented methods are affected by error propagation and accuracy issues [39]. In this study, the classification result from the Landsat 1985 image could not statistically get assessed for accuracy and reliability. Secondly, the Markov chain model used in predicting future land use scenario is based on a large assumption that the rate and type of change in a land use at a given time will be the same for another time period [26], [27]. Also, the reliability of spatial models developed using logistic regression and MARS have not been statistically ground-truthed and its use and success in predicting sage grouse breeding habitats has not been quantified. Unfortunately, future models cannot be ground-truthed with real data neither since still in the future. More research is needed to overcome these problems across scales, but the overall methodology is established now and gets further developed [29]. We think the future will see an increased use of wildlife habitat resources, but also the data availability and computer power will increase making methods like the ones presented here more feasible providing opportunities for decisions support systems on a large and global scale (see for instance [40]). We suggest to further emphasize these biology-based computer models, their data and applications in governments, for environmental impact studies and in policy towards a sustainable future of the globe (for more details on Biodiversity Informatics see also <http://jbi.nhm.ku.edu/viewissue.php?id=2004>).

Acknowledgments

This international project was supported by the University of Calgary, Ford Foundation International Fellowship Fund, German Academic Exchange Service known as Deutscher Akademischer Austausch Dienst (DAAD), Wyoming Fish and Game Department, Wyoming, US, Digital Environmental Management Canada and Stuttgart University of Applied Sciences, Germany. We wish to thank M. Hall-Beyer, all committee members, J. and S. Linke, G.C and E.C. Onyehialam and family, D. Jacobson, D.J. Hagan, L. Strecker, J. and C. Finnigan and all other project contributors and supporters. This is EHWALE publication no 15.

References

1. Cogan C.B.: Biodiversity Conflict Analysis at Multiple Spatial Scales : In Scott J. Michael, Patricia J. Heglund, Michael L. Morrison, Jonathan B. Haufler, Martin G. Raphael, William A. Wall, Fred B. Samson (eds) Predicting Species Occurrences Issues of Accuracy and Scale. (2002) 229-239.
2. Manly B., F.J, Lyman L. McDonald, D.L. Thomas, T.L. McDonald, W.P. Erickson: Resource Selection by Animals: Statistical Design and Analysis for Field Studies. Kluwer Academic Publishers. (2002).
3. Smallwood, K.S., and Morrison, M.L.: Using the best scientific data for endangered species conservation. - Environ. Manage. (1999):421-435
4. Huettmann, F.: Databases and science-based management in the context of wildlife and habitat: towards a certified ISO standard for objective decision-making for the global community by using the internet. - J. of Wildl. Manag. (in press)

5. McArthur E.D. and Ott, J.E.: Potential natural vegetation in the 17 conterminous western United States. in Barrow, J.R., McArthur, E.D., Sosebee, R.E. and Tausch, R.J. (compilers) Proceedings: Shrubland ecosystem dynamics in a changing environment. - General Technical Report INT-GTR-338, (Odgen, UT: Intermountain Research Station) (1996) 16-28
6. Connelly J. W., Schroeder, M. A. Schroeder, Sands, A.R., and Braun, C.E.: Sage Grouse Management Habitat And Management: Guidelines to manage sage grouse populations and their habitats. - Wildlife Soc. Bull. (2000) 967–985.
7. Bent A.C: Life histories of North American gallinaceous birds. - Smithsonian Inst., U.S. Nat. Mus. Bull.162. (1932)
8. Beck T. D. I., and C. E. Braun: The strutting ground count, variation, traditionalism, management needs. - Proc. Ann. Conf. West. Assoc. Fish Wildl. Agencies 60 (1980) 558-566.
9. Schroeder M. A., Young, J.R., and Braun, C.E.: Sage grouse (*Centrocercus urophasianus*). In A. Poole and F. Gill, editors. - The Birds of North America, No. 425. The Birds of North America, Philadelphia, Pennsylvania, USA.(1999) 1-28
10. Hupp J. W. and Braun, C.E.: Topographic distribution of sage grouse foraging in winter. - J. of Wildl. Manage.(1989) 823-829.
11. Onyehialam A. I.: Integration of Classification Techniques for Vegetation and Land Use Classes using Landsat 7 ETM+ and GIS data. M.Sc Thesis submitted at Stuttgart University of Applied Sciences, Stuttgart. Germany (2003).
12. Fisher C., Gustafson, W. and Redmond, R.: Mapping Sagebrush/Grasslands FromLandsat TM-7 Imagery: A Comparison Of Methods. Prepared for the USDI, Bu-reau of Land Management Montana/Dakota State Office (2002).
13. Goodchild M.F.: Spatial Autocorrelation. Geo Books Norwich, CATMOG (1986): 47
14. Legendre, P.: Spatial Autocorrelation: Trouble or New Paradigm? - Ecology (1993) 1659-1673.
15. Cliff A. D. and Ord, J.K: Spatial Autocorrelation, Pion (1973).
16. Hosmer D.W. and Lemeshow, S.: Applied Logistic Regression. John Wiley and Sons Inc. New York. (1989)
17. McCullagh, P. and Nelder, J.A.: Generalized Linear Models. Chapman and Hall, London (1983).
18. Burnham K.P. and D.R. Anderson: Model Selection and Inference: a practical information-theoretic approach. Springer-Verlag, New York (1998).
19. Frank E. Harrell, Jr.: Regression modelling strategies with application to linear models, Logistic regression and survival analysis. Springer (2001).
20. Huettmann, F. and J. Linke (2003): An automated method to derive habitat preferences of wildlife in GIS and telemetry studies: A flexible software tool and examples of its application. - European J. of Wildl. Res. (2003) 219-232.
21. Yen, P., F. Huettmann, and F. Cooke: Modelling abundance and distribution of Marbled Murrelets (*Brachyramphus marmoratus*) using GIS, marine data and advanced multivariate statistics. - Ecol. Mod. (2004) 395-413
22. Hastie.T, R.Tibshirani and J. Friedman: The elements of statistical learning Data Mining, Inference, and Prediction. - Springer Series in Statistics (2001).
23. Lek, S., A. Belaud, P. Baran, I. Dimopoulos, J. Lauga and S. Aulagnier : Application of neural network to modelling non-linear relationships in ecology. - Ecol. Mod. (1996) 39-52.
24. Lusk J., Guthery, F.S., and DeMase, S. J.: A Neural Network for Predicting Northern Bobwhite Abundance in the Rolling Red Plains of Oklahoma. In: Scott J. M., P. J. Heglund, M. L. Morrison, J.B. Haufler, M.G. Raphael, W. A. Wall, F.B. Samson (eds): Predicting Species Occurrences Issues of Accuracy and Scale. (2002) 345-356.

25. Huettmann F. and A.W. Diamond: Seabirds and Ecological Scale Questions for the North Atlantic. *Landsc. Ecol.* (in review).
26. Urban D.L and Wallin, D.O.: Introduction to Markov models in Learning Landscape Ecology. A practical guide to concepts and techniques by Gergel S.E and M.G Turner. Chapter 4 Springer Verlag NY. Inc (2002) 35-47.
27. Hogweg P.: Cellular automata as a paradigm for ecological modeling. *Appl. Math. Comput.* (1988) 81–100.
28. Drton, M, Marzban C. Guttorp .P and Schaefer, J.T.: A Markov Chain Model of Tornadic Activity. Department of Statistics, University of Washington, Seattle. June 8, (2003).
29. Huettmann, F., Franklin, S. and Stenhouse, G. : Deriving Future Landscape Scenarios of Grizzly Bear Habitat. - *Forestry Chronicle* (in press).
30. LaMontagne J.M., Irvine, R.L. and Crone,E.E.: Spatial Patterns of population regulation in sage grouse (*Centrocercus* spp.) population viability analysis. -*J. of Anim. Ecol.* (2002) 672-682.
31. Riebsame.W. E., Parton, W. J., Galvin, K. A., Burke, I. C., Bohren, L., Young, R. and Knop, E. : Integrated modeling of land use and cover change. -*Bioscience* (1994) 350–356.
32. Jenerette G. D and Jianguo Wu: Analysis and simulation of land-use change in the central Arizona – Phoenix region, USA. -*Landsc. Ecol.* (2001): 611–626.
33. Brondizio E., E. Moran, P. Mausel, and Y.Wu: Land cover in the Amazon Estuary: Linking of the Thematic Mapper with botanical and historical data. -*Photogrammetric Engineering and Remote Sensing*, (1996) 921-929.
34. Ramankutty, N., and Foley, J.A.: Estimating historical changes in global land cover: crop lands from 1700 to 1992. – *Global Biogeochemical Cycles* (1999) 997-1027.
35. New, M., Lister, D., Hulme, M. and Makin, I.: A high-resolution data set of surface climate over global land areas. – *Clim. Res.* (2002) 1-25.
36. National Geophysical Data Center (NGDC): ETOPO2. NOAA/NGDC E/GC3, 325 Broadway, Boulder, CO USA <http://www.ngdc.noaa.gov/mgg/fliers/01mgg04.html>
37. Forman, R.T.T.: *Land Mosaics: The Ecology of Landscapes and Regions*. Cambridge University Press (1995)
38. Braun C. E.: Sage grouse declines in Western North America: what are the problems? - *Proceedings, Western Association of Fish and Wildlife Agencies* (1998) 139-156.
39. Wisdom M.J., Wales, B.C., Rowland, M.M., Raphael, M.G., Holthausen, R.S., Rich, T.D. and Saab, V.A.: Performance of Greater sage grouse models for Conservation Assessment in the Interior Columbia Basin, USA. - *Conservation Biology* (2002) 1232-1242.
40. Refish, J., Kone, I. and Huettmann, F.: Review article: Hunting monkeys in tropical forests and its implications for biodiversity and human health. *Frontiers in Ecology and the Environment*. (submitted).