Application of Spatial Modeling Approaches and Techniques within an Ecological Framework

Ullrich von Bramann[1], Alicia V. Linzey[2], Antje Schwalb[1], Stephan van Gasselt[3]

Abstract—The characterization of the species distribution in the field of ecology is an important issue for the definition of habitats/niches and for the identification of impact of environmental factors (e.g., climate, pollution, natural relief). In this particular research field a number of stand-alone tools are used to model species distribution and ecological boundary conditions on the basis of environmental factors and remote-sensing terrain data (image and terrain models). In the context of overarching research projects, modeling of ecological boundaries and distributions are one of many research components integrated within a common framework. The implementation and improvement of ecological modeling approaches within state-of-the-art GI systems forms a crucial part for data exchange, traceability and, more importantly, validation and quality assessment. This work deals with the implementation of various models and their assessment.

Index Terms — Species distribution modeling, ArcGIS, Maximum entropy principle, Ecology

Introduction

Ecological applications in research fields such as invasive species management, epidemiology and sustainable nature conservation require the prediction of geographic distributions of individual species based on environmental boundary conditions at certain sample sites [1]. If data on positive occurrences (data of presence – DOP) and data on negative occurrence (data of absence – DOA) of species at sampling sites are available, common statistical methods can be used to perform species distribution modeling [2].Unfortunately, in most cases statistically significant data on the presence and absence of species are barely available or of questionable quality due to, e.g. sampling bias. Distribution modeling of occurrences of certain species within a geospatial context is therefore of utmost importance in order to characterize and quantify habitats and their ecological boundary conditions. Within an ongoing research framework, we are currently focusing on the distribution modeling of species whose spatial distributions are poorly defined. Two rodent species are currently investigated that provide a model system for the study of speciation mechanisms in the absence of species-morphological differentiation. Although their geographic ranges are largely non-overlapping, ecological characteristics that define their differing distributions are unknown. Environmental conditions are, however, well constrained at sampling sites where positively identified individuals of these species have been collected. In order to find a measure for their regional distribution an integration of environmental conditions needs to be assessed and correlated with sampling points. Such work is usually carried out outside the context of any Geographic Information System (GIS) although the identification of spatial dependences is required in order to constrain any modeling attempt. In this work we present conceptual and first-order approaches for GIS-integrated analyses to constrain and model species distributions.

- *1 Technische Universität Braunschweig, Institute of Environmental Geology, Langer Kamp 19c, D-38106 Braunschweig, Germany* ✉ *u.bramann@tu-bs.de, Fon:* ☎ *+49 531 391 7259*
- *2 Department of Biology, Indiana University of Pennsylvania, Indiana, PA 15705*
- *3 Freie Universität Berlin, Institute of Geological Sciences, Malteserstr. 74-100, D-12249 Berlin, Germany*

State of the Art

Several methods to perform presence-only modeling are currently in use. To predict suitable bioclimatic conditions and in order to predict the distribution the modeling package BIOCLIM in conjunction with the BIOMAP package developed at the Australian National University produces 35 parameters based on input data such as minimum and maximum temperatures, rainfall, solar radiation and evaporation [3]. The software makes use of bioclimatic parameters derived from mean monthly climate estimates, to approximate energy and water balances at a given location [4].
Another approach is realized by the DOMAIN software which delivers a suitability index by computing the minimum distance in environmental space to presence records [4]. It uses a point-to-point similarity metric to assign a classification value to a candidate site based on the proximity in environmental space of the most similar record site [7]. Generalized Linear Models (GLMs) and Generalized Additive Models (GALs) are general purpose statistical approaches for presence-absence modeling. Other approaches are realized by MaxEnt [5] which employs the maximum entropy approach and the tool *Genetic Algorithm for Rule-set Production* (GARP) which targets at iteratively searching for non-random correlations between species presence and absence [7]. All these model approaches are static and make use of different techniques to predict habitats/niches based on environmental input data. In the course of this work, an implementation of selected algorithms within Environmental Research Institute's commercial desktop GI system ArcGIS is envisaged in which also different modeling results processing requirements are compared after full implementation. Additionally, solutions are developed to allow for a dynamic approach by processing observations for given times in order to be able to make predictions on varying time scales. Though the time-dimension is currently not supported in ArcGIS 9.x, new releases will provide basic facilities. For the time being, a time-dependent (dynamic) approach must be modeled via a step-wise implementation and which can later be adapted for state-of-the-art functionality within a GIS-based environment.

1 Data

In order to perform distribution modeling using various approaches as described below and in order to compare results with in-situ data and other solutions, a number of variables need to be defined that are based on the environmental framework (climate and precipitation, terrain characteristics, vegetation and their derivatives), species sampling characteristics at given locations within a proper geospatial context, and the timing framework at which these data were collected.

STUDY SPECIES

Aethomys chrysophilus and *A. ineptus* are sibling rodent species, and cannot be distinguished using external morphological features. Identification requires determination of chromosome number, mitochondrial DNA analysis, or examination of sperm. The species are found in the southern regions of Africa in natural habitats that include dry and moist savannahs, dry forests, plantations and arable land. Specimens from a total of 35 localities in Botswana, Malawi, Namibia, Republic of South Africa, Tanzania and Zimbabwe have been identified between 1986 and 2006. *Aethomys* specimens from these localities were collected primarily by live-trapping for chromosome counts for positive identification. Subsequent to analyses that allowed identification using mitochondrial DNA, tissue samples were taken from live-trapped individuals that were released back into wild populations. Based on these identifications, the ranges of the two species were described by Linzey et al. [8].

ENVIRONMENTAL VARIABLES

The term environmental variables comprise three different sets of variable classes:
(1) climatic variables in a broader sense, i.e. interpolated bioclimatic variables as predictor variables. A total of 19 bioclimatic variables are available from Worldclim [http://www.worldclim.org/] from which we will use annual mean temperature, maximum temperature of warmest month, minimum temperature of coldest month, annual precipitation, precipitation of wettest month, precipitation of driest month and precipitation seasonality. The data is available at a spatial resolution of approximately 1 km^2.
(2) Vegetation data as observed in the field and inferred from multispectral remote-sensing data interpretation on the level of satellite data (Landsat MSS/TM/ETM+) and auxiliary airborne data. The geometric resolution is in the range meters to decameters/pixel.
(3) Terrain-model data as obtained indirectly from Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) stereophotogrammetry with a resolution of 30 m/pixel as well as Shuttle Radar Topography Mission (SRTM) data at a scale of 90 m/pixel. These data are required to identify natural boundaries such as rivers or high mountain ranges, watersheds as well as simple terrain characteristics such as aspects, slopes and topographic heights and relief as well as their derivatives.

2 APPROACH AND METHODOLOGY

In this work we employ (1) multispectral remote sensing data of various sources and times for vegetation mapping and change detection, (2) terrain model data such as SRTMs and ASTER data to identify terrain characteristics and natural barriers and (3) environmental data of various sources and scales to perform species distribution analysis in order to identify ecological characteristics that define differing distributions. The modeling will be performed in MaxEnt as well as ArcGIS in order to identify and analyze possible disparities of the results as a function of input data quality.
The MaxEnt modeling approach is a method to make predictions from incomplete information. It estimates a target probability distribution by finding the probability distribution of maximum entropy. For this purpose it employs efficient deterministic algorithms for optimal distribution modeling and can be applied to DOP-modeling as well as DOP/DOA-modeling if conditional models are applied for the latter [5-6]. MaxEnt is a general-purpose method to predict the environmental suitability for a species based on known environmental parameters. Thus it is employed to (1) estimate the target probability distribution for non-sampled areas and (2) allows conclusions on the impact of environmental parameters on differing species distributions.
Input as well as output is raster-based geospatially located data that are well suited for any GIS implementation. The implementation of MaxEnt and derivates as well as other modeling approaches allow to compare different workflows and results more conveniently in order to assess the suitability of a given model attempt. Solutions using MaxEnt are based on parallel or iterative processing which cannot be handled satisfactory using out-of-the-box tools and model-buildertools within ESRI's ArcGIS suite. Therefore, MaxEnt approaches are currently implemented using ArcObjects via VBA.NET and based on Microsoft's Component Object Model (COM) in order to be able to implement any routines within the new ArcGIS suite as well. Testing/verification is done against stand-alone tools. The emphasis is on the implementation of different model routines and on the user-definable selection/de-selection of parameters within a GIS context. As, in particular, the model approach's results rely on the quality of data, an integration of different terrain models and environmental datasets must be guaranteed.Time-dependence is established via step-wise solutions.

3 FUTURE WORK AND OUTLOOK

Two major issues need to be solved for improving the current workflow. One issue deals with the adaptation of the implementation using a dynamic model, the other issue deals with merging different approaches to species distribution modeling (SDM) in order to increase the resilience of the results. Furthermore it is planned to implement data on contaminations such as Acid Mining Drainage (AMD) and airborne contaminants such as particulate matter to aid the understanding of the impact of anthropogenic influence on ecosystems.

REFERENCES

[1] Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G.,Wall,W.A., Samson, F.B. (Eds.), 2002: Predicting SpeciesOccurrences: Issues of Accuracy and Scale. Island Press, Washington,DC

[2] Elith, J., 2002: Quantitative methods for modeling species habitat: Comparative performance and an application to Australian plants. In: Ferson, S., Burgman, M. (Eds.), Quantitative Methods for Conservation Biology. Springer-Verlag, New York, 39–58.

[3] Nix, H., 1986: A biogeographic analysis of Australian Elapid Snakes. In: Atlas of Elapid Snakes of Australia. (ed.) R. Longmore pp. 415. Australian Flora and Fauna Series Number 7. Australian Government Publishing Service: Canberra.

[4] Carpenter, G., Gillison, A.N.,Winter, J., 1993: DOMAIN: a flexible modeling procedure for mapping potential distributions of plants and animals. Biodivers. Conserv. 2, 667–680.

[5] Steven J. Phillips, Robert P. Anderson, Robert E. Schapire (2006): Maximum entropy modeling of species geographic distributions**.** In: *Ecological Modelling*, 190:231-259, 2006.

[6] Carpenter, G., Gillison, A.N.,Winter, J., 1993: DOMAIN: a flexible modeling procedure for mapping potential distributions of plants, animals. Biodivers. Conserv. 2, 667–680.

[7] Anderson, R.P., 2003: Real vs. artefactual absences in species distributions:tests for Oryzomys albigularis (Rodentia: Muridae) in Venezuela. J. Biogeogr. 30, 591–605.

[8] Linzey, A. V., M. H. Kesner, C. T. Chimimba, and C. Newbery. 2003: Distribution of veld rat sibling species Aethomys chrysophilus and Aethomys ineptus (Rodentia: Muridae) in southern Africa. African Zoology 38: 169-174.