

Presence-only Habitat Suitability Modelling Using Unclassified Landsat ETM+ Imagery: Fine-resolution Maps for Common Small Mammal Species in Bulgaria

Vasil V. Popov

Institute of Biodiversity and Ecosystem Research, Bulgarian Academy of Sciences, Tsar Osvoboditel 1 Blvd. 1000 Sofia, Bulgaria;
E-mail: vasilvpopov@gmail.com

Abstract: The aims of the study were to generate habitat suitability models for 19 common small mammal species in order to 1) compare two approaches to eliminate the effect of sampling bias and spatial autocorrelation, 2) choose the best model and provide accurate high resolution habitat suitability maps, 3) examine the obtained relationships within the context of ecological realism of the models, and 4) evaluate the potential of satellite imagery in the modelling process. The models were fitted using three different sets: (1) all data (ID-models); (2) by using spatial bias file (MBD-models); and (3) spatial filtering (SF-models). Although MBD- and SF- models showed lower performance statistics they produced ecologically more reliable habitat suitability maps than ID-models, especially for widely distributed species. MBD-models produced an optimal output that provides the maximal explanation of input data while still maintaining generality. Models based on this approach were chosen to produce the final maps. According to the AUC values of test partition, 6 models were fair, 7 were good, and 6 were excellent. Across all fitted models, the highest percentage contribution was on average for altitude, which had a contribution of 54.3 %, followed by the PC1 of reflectance data (9.9%), topographic wetness index (7.1%), aspect (6.9%), PC2 (4.9%), slope (3.4%), and PC3 (2.0%). The species-environment relationships registered by the models corresponded very well to the known ecology of the studied species. It was shown that raw Landsat ETM+ data integrated with topographical variables have the potential to produce models that have a high degree of ecological realism and reflect both the general type of distribution of each species in the country, and peculiarities associated with fine scale local conditions. In this sense satellite data can be regarded as especially suitable for modelling small and less mobile animals.

Key words: small mammals, MAXENT, Landsat, modelling, bias, autocorrelation, SDM

Introduction

In recent years, habitat suitability modelling, HSM (or species distribution modelling, SDM) is playing an increasingly important role in studying distribution patterns of organisms (GUISAN and THUILLER 2005). These models are based on the analysis of the interrelationship between the known occurrences of a particular species and the state of environmental variables that are important for its existence in this place. They allow to extrapolate these relationships in unexplored regions, or into hypothetical scenarios of future or past climatic conditions (GUISAN and THUILLER

2005). Over the last years a large amount of environmental data has become available for analyzing species distributions in the form of interpolated digital maps with environmental measurements. These environmental layers include abiotic factors, such as temperature, precipitation, or their proxies such as elevation, slope, aspect, topographic wetness index, hill shading, etc. Climatic variables typically have a great range of continuous values, and are often used for predicting species distributions at large scales. In contrast, topographic data are often used on the re-

gional or nationwide scales (GUISAN and ZIMMERMANN 2000). At these scales the distribution of the organisms is largely affected also by land surface characteristics (forests, crops, water bodies, bare rock, etc.) and human impact on earth surfaces (agriculture, deforestation, urbanization, etc.). So, land-cover data are increasingly used as a predictor in species distribution modelling (PEARSON *et al.* 2007, FRANKLIN 2009). In European scale, CORINE land cover is widely used in this respect (LUOTO *et al.* 2007, THOMAES *et al.* 2008, MÜCHERA *et al.* 2009). It is based on satellite data (Landsat TM, MSS and SPOT XS) as well as auxiliary data in the form of national topographic and thematic maps, statistical land cover information and aerial photographs. These data were merged applying expert knowledge and following a standard procedure. CORINE is a hierarchical land cover classification with 44 classes at level 3 (EEA 2000). One of the drawbacks of CORINE and classified land cover maps in general concerns the fact that usually they are designed for a specific mapping purpose and as a result, they may be not representative and thematically detailed enough for the focal species (BRADLEY and FLEISHMAN 2008). Most nomenclatures used for mapping are designed for the purpose of compiling an inventory of human activities (land use). Beside this, these maps are based on discrete classes which do not capture gradual changes in the landscape (CORD *et al.* 2014). In this respect, the use of continuous unclassified remotely sensed data on reflectance as explanatory variables in species distribution is a promising alternative (LEYEQUIEN *et al.* 2007, PETTORELLI *et al.* 2011, SHIRLEY *et al.* 2013, CORD *et al.* 2014). Among the various products of this type Landsat TM imagery is especially useful for these purposes (TURNER *et al.* 2003, ZIMMERMANN *et al.* 2007, KERR and OSTROVSKY 2003). Landsat images are free to the public and easy to obtain (WOODCOCK *et al.* 2008). These images are available on an annual basis offering opportunities for ecologists to easily accessible multi-year data over large areas. Thus, they allow to register time changes in land cover and thus allows monitoring of habitats. Spectral values for different bands are associated with a variety of landscape characteristics (KERR and OSTROVSKY 2003) and the subjectivity of man-made classification in standard land cover maps and the loss of information associated with classifying an inherently continuous attribute is removed (ZIMMERMANN *et al.* 2007). In most cases, however, the data on reflectance as explanatory variables are used strictly for modelling purposes and the resulting outputs are assessed only in terms of their statistical accuracy (TATTONI *et al.* 2012, SHIRLEY *et al.* 2013, CORD *et al.* 2014). Not enough information as to what

extent the relationships between species occurrences and remote-sensing spectral variables reflect the ecological characteristics of species. The interpretations of these variables from the ecological point of view are important for understanding which features improved habitat quality for conservation priority species can help managers to define effective guidelines (TATTONI *et al.* 2012, SHIRLEY *et al.* 2013).

There are a number of modelling techniques and algorithms for exploring relationships between species and their environment and to map their spatial distribution (GUISAN and ZIMMERMANN 2000, GUISAN and THULLER 2005). Among the many alternative modelling methods, maximum entropy approach has recently gained much popularity, due to the availability of a quality software MAXENT that implements the method and is free for academic and research use (<http://www.cs.princeton.edu/schapiro/MAXENT>). Maximum entropy evaluates the ecological niche of the species by finding a probability distribution, which is based on the maximum entropy distribution using the restriction that the expected value of each variable in this estimated distribution corresponds to its empirical average (PHILLIPS *et al.* 2006). Recently it has been shown that MAXENT is equivalent to a Poisson regression model and hence is related to a Poisson point process model, differing only in the intercept term, which is scale-dependent in MAXENT. This equivalence will allow a number of improvements to MAXENT (RENNER and WARTON 2013). One the main drawback of MAXENT concerns its inability to make general predictions and as a result, it is prone to overfitting (PETERSON *et al.* 2007). Overfitting occurs when a model fits too tightly to calibration data, that leads to more omission errors. This in turn increases the effect of sampling bias and can lead to a model that reflects the spatial aspects of the research itself rather than the real distribution of a species. The sampling bias also increases spatial autocorrelation of the model residuals which worsens the quality of the model and seemingly enhances the its accuracy (VELOZ 2009), which leads to type I errors (DORMANN *et al.* 2007) and misleading estimates of the model parameters (KÜHN 2007). This means that greater significance may be attributed of environmental variables that are just typical of the region, subject to more intensive research leading to spatial omission or commission errors of extrapolation (RONDININI *et al.* 2006, KRAMER-SCHADT *et al.* 2013). These problems can be solved by reducing the number of occurrence points from areas in which they are concentrated, using spatial filtering (DORMANN *et al.* 2007, PHILLIPS *et al.* 2009, VELOZ 2009). This method, however, increases the risk of excessive reduction

of the number of calibration points, which prevents building a statistical robust model. Alternatively, manipulation of background data can be employed. Recent versions of MAXENT allow the inclusion of so-called bias file that allows the user to choose background data with the same bias as occurrence data (PHILLIPS *et al.* 2009, SYFERT *et al.* 2013). Model outputs not only differed among different modelling approaches, but also within methods used to evaluate different species, which makes it difficult to identify the 'best' modelling technique (SEGURADO and ARAÚ 2004, PEARSON *et al.* 2007). When the goal of study is to use predictive models of species distributions for prioritizing areas for conservation, it is crucial to understand situations where it is preferable to reduce sampling bias by manipulating occurrence data versus manipulating background data (KRAMER-SCHADT *et al.* 2013).

In the present study, the majority of species data comes from regional surveys of protected areas, reserves or potentially interesting from a biogeographical point of view, parts of the country. As a result, the information is geographically clustered. Within the areas, samples were often collected from the most perspective habitats, rather than systematically or randomly. As a result species occurrence points may be biased. These reserves, however, are valid for the dataset as a whole, but it is unclear to what extent the potential overfitting of data, due to sampling bias and spatial autocorrelation, affects the individual species data sets, given their significant differences in distributional patterns.

The aims of the study were to generate habitat suitability models for 19 common small mammal species in order to 1) compare two approaches to eliminate the effect of sampling bias and spatial autocorrelation, 2) chose the best model and provide accurate high resolution habitat suitability maps, 3) examine the obtained relationships within the context of the ecological realism of the models, and 4) evaluate the potential of satellite imagery in the modelling process.

Methods and Materials

Species data

Presence records of 19 small mammal species from field surveys conducted during a 10 year period (1995-2005) were compiled. The data set contains 279 unique point collections from georeferenced localities. Geographical coordinates were taken with a GPS with an accuracy of 2-5 m. Part of these data have been published (POPOV 1999, BERON *et al.* 2000a,b,

POPOV 2000, MINKOVA and POPOV 2002, SICHANOV *et al.* 2006, POPOV *et al.* 2006) and summarized in PESHEV *et al.* (2004), POPOV (2007), and POPOV *et al.* (2007).

Environmental predictors

Four topographic variables were used – Digital Elevation Model (DEM), aspect (ASP), slope (SLP), and topographic wetness index (TWI). DEM images, as GeoTIFF files with 30 m resolution, were acquired from NASA LP DAAC (2014). DEM derivatives (ASP, SLP, TWI) were calculated from the elevation layer by using SAGA GIS Version 2. 1. 4 (available at <http://www.saga-gis.org>). Aspect values were in degrees starting from 0° – N, 90° West, 180° South, and 270° East; for pixels on a flat surface a value of – 1 was attributed. To characterize land cover several spectral bands from the Enhanced Thematic Mapper Plus (ETM+), TM1 (Blue), TM2 (Green), TM3 (Red), TM4 (near-IR), TM5 (mid-IR) and TM7 (mid-IR), with a pixel resolution of 30 m and pixel values (reflectance) between 0 and 255, were used. Images had sub-pixel geolocation accuracies (corrected for terrain displacement and errors in image geometry), and no further geometric processing was applied. In order to minimize the noise not related to ground conditions and to make images comparable a correction method of histogram matching was employed using R 3.0.1 and the additional 'landsat' package (R Core Team 2013, GOSLEE 2011). Scenes from the years 1999 – 2000 were selected (Table 1), to temporally coincide with the middle of the period of collecting the species data. All images represented the growing season, however, due to the spatial extent of study, it was not possible to obtain all images for the same year and phenological stage. Therefore, some extraneous phenological variability was present, concerning primarily certain areas of cultivated fields in NE Bulgaria (paths 181-182, rows 30-31). On the other hand, a thorough analysis of the overlap areas showed that changes in surface reflectance between scenes were minimal and it can be assumed that this variability will have a negligible impact on the accuracy and predictive capability of the models. To ensure the correspondence of collecting localities and environmental conditions presented by satellite imagery each record from the dataset was visually inspected using recent satellite maps available on Google Earth to ensure that the recent situation matched the original description. Any major discrepancies were further evaluated in relation to the time during which they occurred. If these changes occurred before 1999 it was estimated that the records do not correspond with the satellite

data and were excluded from further consideration (5 cases).

Since adjacent bands in multispectral images were correlated, Principal Component Analysis was applied. The first three principal components (PC1, PC2, and PC3), accounting for more than 95 % of initial spectral information were used as explanatory variables in modelling.

Three species (*Apodemus epimelas*, *Mus macedonicus* and *M. spicilegus*) show very peculiar distributional pattern throughout the country. *A. epimelas* occurs only in the southwestern part, *M. macedonicus* is known solely from the lowlands of S Bulgaria, while *M. spicilegus* occurs only in the lowlands of N Bulgaria. Most probably these patterns reflect their dispersion from southwest, south and northeast, respectively, following appropriate environmental conditions. In order to account for the spatial dispersal constraints that probably partly determines their distributional patterns two dummy variables, representing Euclidean distances from southeastern (ESW) and southwestern (ESW) corners of the country were created and added to the other variables once modelling the distribution of those species. For *A. epimelas* and *M. spicilegus* only ESW was employed since it represents a spatial gradient that have an opposite effect of each species. For *M. macedonicus* both variables were used since they describe the spread of the species from several lowland areas along the southern border of the country. By adding these variables, the model was prevented to incorrectly predict species in areas that are environmentally suitable but are outside their colonizable range.

GIS analyses were performed with GRASS GIS 6.4 (GRASS Development Team 2008), figures and map layouts were produced using Quantum GIS 1.6 (Quantum GIS development team 2011). The whole geographic data set was projected according to WGS 1984 UTM Zone 35N reference system; the spatial resolution for all maps is 30 m.

Habitat suitability modelling and validation

MAXENT was used with its default settings and a logistic output, with suitability values ranging from 0 (unsuitable habitat) to 1 (optimal habitat) (PHILLIPS and DUDIK 2008). The use of the standard settings and therefore the auto feature selection, means that the program automatically adds modelling features depending on the number of the points used in the training set: below 10 points linear functions are used; between 10 and 14 points quadratic features are added; between 15 and 79 points hinge features are added and above 79 points product and

threshold features are allowed. The relative variable importance was assessed based on MAXENT's built-in Jackknife functionality. In order to test to what extent the potential overfitting of data, due to sampling bias and spatial autocorrelation, affects the individual species data sets, three types of models were made and compared. The first type of models was based on the initial data (ID-models). The second type was based on manipulation of background data (MBD) by introducing a bias file that allows one to choose background data with the same bias as occurrence data (DUDIK *et al.* 2005, PHILLIPS *et al.* 2009, SYFERT *et al.* 2013). The third type was based on spatial filtering (SF). For MBD-models, a biased prior was constructed using MAXENT – all unique species locations were provided to MAXENT as if they represented a single species and the resulting prediction, reflecting the sampling effort was used as a bias file. For SF-models, the localities that are within 30 km of one another were randomly removed, keeping the most localities possible. The 30 km distance was chosen on the basis of autocorrelations as detected in the semivariograms and Moran's I-distance plots of ecogeographical variables. In order to test the models based on a large number of points (more than 30) of the first two types crossvalidation was used. The partition of occurrence localities into testing and training bins was done by using 15-fold cross-validation ('crossvalidate' partitioning scheme available in the MAXENT software) which partitions occurrences randomly into 15 bins. Each of these 15 bins was used for testing once, while all others were used for training in that iteration. A total of 15 models were run, and evaluation metrics were summarized across these iterations. For models based on the low number of occurrences a jackknife approach (PEARSON *et al.* 2007) was used as it is a good validation method for small sample sizes. In the jackknife approach, the model was tested using one record of presence data and calibrated with the rest. In SF-models, the fraction of the original dataset which was excluded on the basis of the spatial filtering was used as a test sample, provided directly in the MAXENT software. The three model types were compared on the basis of two threshold-independent metrics (AUC_evaluation and AUC_test). These metrics derive from Area Under the Curve of the Receiver Operating Characteristic plot, each providing a single rank-based measure of model performance across all possible thresholds (levels of strength of the prediction). AUC_evaluation is a measure of overall model performance, while AUC_test describes the capability of a model to predict a fraction of the species records omitted during model train-

Table 1. Landsat ETM+ images used in the present study, acquired from the United States Geological Survey (USGS) Global Visualization Viewer (<http://earthexplorer.usgs.gov/>)

| Scene | Year | Month | Date | Path | Row |
|---------------------------------|------|-------|------|------|-----|
| p181r030_7dk19990723_z35_61.tif | 1999 | 7 | 23 | 181 | 30 |
| p181r031_7dk20000725_z35_61.tif | 2000 | 7 | 25 | 181 | 31 |
| p182r029_7dk20000614_z35_61.tif | 2000 | 6 | 14 | 182 | 29 |
| p182r030_7dk20000614_z35_61.tif | 2000 | 6 | 14 | 182 | 30 |
| p182r031_7dk20000614_z35_61.tif | 2000 | 6 | 14 | 182 | 31 |
| p183r030_7dk20000621_z35_61.tif | 2000 | 6 | 21 | 183 | 30 |
| p183r031_7dk20000621_z35_61.tif | 2000 | 6 | 21 | 183 | 31 |
| p184r030_7dk20000628_z34_61.tif | 2000 | 6 | 28 | 184 | 30 |
| p184r031_7dk20000628_z34_61.tif | 2000 | 6 | 28 | 184 | 31 |
| p185r029_7dk20000822_z34_61.tif | 2000 | 8 | 22 | 185 | 29 |

ing (PHILLIPS *et al.* 2006). The values of these metrics range from 0.5 (random model) to 1.0 (perfect discriminatory abilities). Model performance was classified according to AUC scores based on SWETS (1988) as 0.90–1.00 = excellent, 0.80–0.90 = good, 0.70–0.80 = fair, 0.60–0.70 = poor, and 0.50–0.60 = fail. These non-parametric measures vary according to the proportion of the study region that is suitable for the species and, therefore, should not be compared between species (PHILLIPS *et al.* 2006). They are however appropriate for comparisons of the relative ranking ability of model types for the same species in the same study region (LOBO *et al.* 2008, PETERSON *et al.* 2008). The difference in AUC_{test} scores between model types was tested using paired Wilcoxon signed-rank tests.

In addition, model outputs were evaluated by qualitative visual examination of the maps of the species' predicted potential distribution in term of known natural history information for the study species and the existing knowledge of where the species might occur, including those probable areas inhabited by the species for which no presence information exist (POPOV 2007, POPOV *et al.* 2007). As indications of overfitting, small regions of high prediction, lying close to calibration localities, that do not correspond to recognized environmental conditions that the species is known to inhabit were searched on the maps and examined on the basis of expert knowledge.

Results

In order to check whether there is a potential sampling bias within the data set a MAXENT Model was

created on the basis of all unique occurrence points. The resulting model showed a relatively high value of AUC (0.8) indicating the occurrence of sampling bias. The inspection of the output map indicated that the sampling probability was higher in mountainous and forested regions, large river valleys, as well as along the Black Sea coast. Although this result indicates that the overall dataset shows a spatial bias, it is not clear to what extent it affects particular species models. In order to obtain an information in this respect, three types of models were run and compared. It can be expected that if the ID-models suffer of spatial bias and autocorrelation they will have AUC_{test} values higher than those made by corrections (MBD- and SF-models), indicating greater overfitting. Model performance was assessed on a per species basis. The comparisons showed that the three model types did not differ in terms of overall performance (AUC_{train}) but AUC_{test} values of MBD- and SF-models were significantly lower across species (Tables 2, 3). Fig. 1 shows the variability of these differences across species.

MBD- and SF- models revealed significantly lower AUC_{test} scores than the initial models (Table 3). On the other hand SF-models showed a considerable variation among species of AUC_{test} values (Fig. 1). With some exceptions, for the species presented with a large number of occurrences, corrected models (MBD- and SF-models) showed lower AUC_{test} values compared to the ID-models. In most cases, these differences affected species with wide ecological tolerance and no preference for certain areas characterized by a specific combination of environmental conditions (*A. flavicollis*, *M. arvalis*, *C. suaveolens*, *C. leucodon*, *A. sylvaticus*). Conversely, some of the species presented with many points, but confined to specific areas of the country such as *Clethrionomys glareolus*, *S. araneus*, *M. subterraneus* (attached to forests in mountainous areas) showed no significant differences between the three model types. Species, represented by a small number of points, which most often are associated with ecologically and spatially distinct regions (*Neomys fodiens*, *Chionomys nivalis*) or specific combination of environmental variables (*Dryomys nitedula*, *Muscardinus avellanarius*, *Micromys minutus*) showed little differences among model types. The visual inspection of the output maps showed that models with adjustments (MBD- and SF-models) in most cases were more diffuse – give more territory and less contrast in the suitability values within limited territories, making them more plausible from an ecological point of view. This relaxation leads to the fact that MBD- and SF- out-

Table 2. Descriptive statistics of model performance characteristics for all species (n=19). AUC_train - evaluation of performance based on all points; AUC_test - evaluation of performance based on 15-fold cross-validation or jackknife approach. ID – models based on initial data; MBD – models based on background manipulation (bias file), SF - models based on spatial filtering

| Parameter | Mean | Min | Max | SD |
|---------------|------|------|------|------|
| AUC_train_ID | 0.90 | 0.80 | 0.99 | 0.06 |
| AUC_train_MBD | 0.89 | 0.80 | 0.99 | 0.06 |
| AUC_train_SF | 0.87 | 0.71 | 0.97 | 0.07 |
| AUC_test_ID | 0.84 | 0.72 | 0.98 | 0.09 |
| AUC_test_MBD | 0.81 | 0.66 | 0.98 | 0.10 |
| AUC_test_SF | 0.78 | 0.40 | 0.99 | 0.15 |

puts often include some extreme occurrence points that fall outside of the minimum presence threshold of ID-models. Furthermore, inspection of the maps showed that many of the ID-model outputs predicted low suitability for some species in certain areas, which does not correspond with the knowledge of their ecology. Often these areas coincided with parts of the country which were not been the subject of an intensive research, eg. parts of Central and Western Danubian plain, upper parts of Thracian plain.

The comparison between MBD- and SF-models indicated that most often the results were similar. In some cases, however, SF-models sharply diverged and showed features that can be interpreted in favor

Table 3. Wilcoxon Matched Pairs Test. Marked tests are significant at $p < .05000$

| Comparison | Z | p-level |
|---------------------------|-----------------|----------|
| AUC_train & AUC_train_MBD | 1.649916 | 0.098960 |
| AUC_train & AUC_train_SF | 0.707107 | 0.479500 |
| AUC_test & AUC_test_MBD | 3.535534 | 0.000407 |
| AUC_test & AUC_test_SF | 2.460595 | 0.013871 |

of the their lower ecological reality – high contrasts within small territories, unjustified environmentally absences from large areas.

These results suggested that overall predictive performance was higher in modelling scenarios that correct for sampling bias (MBD-models). They produce a more optimal output that provides the maximal explanation of calibration data while still maintaining generality. Models based on this approach were chosen to produce the final maps, using the full set of data for each species and were subject to further analysis (Fig. 3). Their performance characteristics are shown in Table 4.

For the 19 MAXENT models the mean AUC for the training data was 0.90 (range: 0.8 to 0.99) while for the test data, mean AUC is 0.84 (range: 0.72 to 0.98). According to AUC_train 10 models were good and 9 excellent, while according to AUC_test 6 were fair, 7 were good, and 6 – excellent. However, as it may be expected these evaluations largely depended on species prevalence.

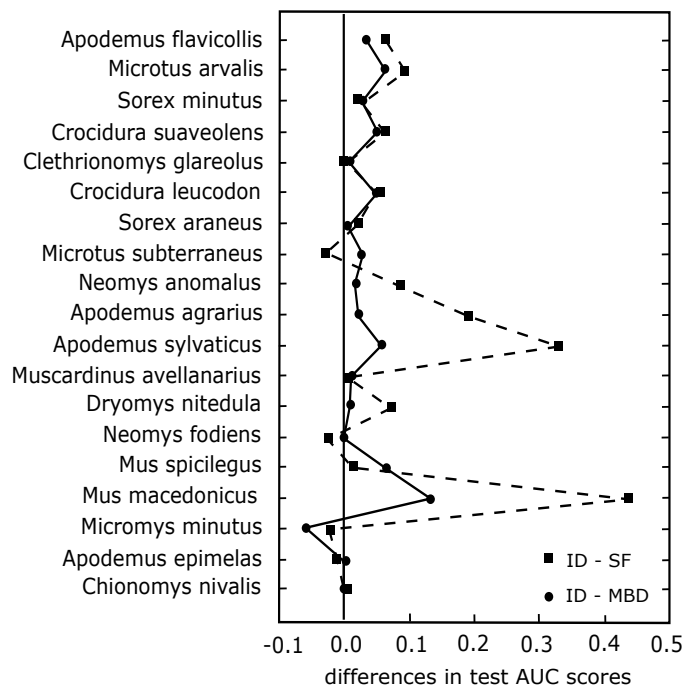


Fig. 1. Comparison of differences in test AUC scores between model types. Species are arranged according to their number of occurrence points in the initial data set from low (bottom) to high (top)

Table 4. Performance characteristics of MBD-models and relative contributions of the environmental variables. Abbreviations: N – number of training occurrence points; DEM – digital elevation model; PC1 – PC3 principal components of Landsat reflectance data; TWI - topographic wetness index; ASP – topographic aspect; SLP – slope

| Species | Model performance characteristics | | | Relative contributions of the environmental variables | | | | | | |
|---|-----------------------------------|-----------|----------|---|------|------|------|------|------|------|
| | N | AUC train | AUC test | DEM | PC1 | TWI | ASP | PC2 | SLP | PC3 |
| <i>Chionomys nivalis</i> | 10 | 0.97 | 0.96 | 87.6 | 0 | 5.7 | 1 | 5.1 | 0.5 | 0 |
| <i>Apodemus epimelas</i> | 14 | 0.99 | 0.98 | 3.8 | 8 | 1.2 | 9.1 | 1.2 | 1.5 | 0 |
| <i>Micromys minutus</i> | 17 | 0.94 | 0.81 | 75.9 | 1.8 | 5.6 | 15.6 | 1 | 0.2 | 0 |
| <i>Mus macedonicus</i> | 25 | 0.97 | 0.93 | 30.9 | 1 | 0 | 0.7 | 0.4 | 3 | 0.2 |
| <i>M. spicilegus</i> | 26 | 0.99 | 0.94 | 16.9 | 0.2 | 0 | 1.7 | 0.5 | 0.5 | 0.9 |
| <i>Neomys fodiens</i> | 26 | 0.97 | 0.95 | 68.7 | 1.5 | 18.2 | 5.6 | 0.9 | 5.1 | 0 |
| <i>Dryomys nitedula</i> | 28 | 0.89 | 0.8 | 41.6 | 27.3 | 6.1 | 17.6 | 3.2 | 2.4 | 1.6 |
| <i>Muscardinus avellanarius</i> | 33 | 0.93 | 0.89 | 73.4 | 8.9 | 12.9 | 2.3 | 0.5 | 1.8 | 0.3 |
| <i>Apodemus sylvaticus</i> | 43 | 0.82 | 0.73 | 69.9 | 0.4 | 0.9 | 9.5 | 12.5 | 3.1 | 3.7 |
| <i>Apodemus agrarius</i> | 52 | 0.89 | 0.86 | 62.7 | 11.3 | 17.1 | 0.4 | 4.7 | 3.5 | 0.3 |
| <i>Neomys anomalus</i> | 61 | 0.85 | 0.81 | 46.6 | 22.9 | 6.3 | 9.5 | 11.8 | 2.7 | 0.1 |
| <i>Microtus subterraneus</i> | 71 | 0.86 | 0.8 | 80.1 | 2.2 | 1.9 | 12.4 | 0.5 | 0.5 | 2.5 |
| <i>Sorex araneus</i> | 75 | 0.91 | 0.88 | 82.6 | 3.8 | 4.8 | 7.2 | 0.9 | 0.2 | 0.6 |
| <i>Crociodura leucodon</i> | 79 | 0.8 | 0.72 | 30.1 | 29.5 | 4.3 | 10.2 | 21 | 4 | 0.9 |
| <i>Clethrionomys (Myodes) glareolus</i> | 88 | 0.95 | 0.92 | 85.9 | 0.6 | 7.1 | 1.6 | 1.1 | 1.8 | 1.9 |
| <i>Crociodura suaveolens</i> | 92 | 0.84 | 0.73 | 57.6 | 17 | 4.7 | 6.6 | 3.8 | 7.6 | 2.6 |
| <i>Sorex minutus</i> | 93 | 0.86 | 0.79 | 37.2 | 15.8 | 14.2 | 12.3 | 2.5 | 13.2 | 4.8 |
| <i>Microtus arvalis</i> (s. l.) | 97 | 0.83 | 0.72 | 53.5 | 7.2 | 9.7 | 2.8 | 13.1 | 6.7 | 7 |
| <i>Apodemus flavicollis</i> | 112 | 0.85 | 0.78 | 26.5 | 29.1 | 14.6 | 4.5 | 8.1 | 6.4 | 10.8 |
| Mean | | 0.90 | 0.84 | 54.3 | 9.9 | 7.1 | 6.9 | 4.9 | 3.4 | 2.0 |

Nevertheless, as a whole, these numbers indicated a good overall modelling performance.

The highest percentage contribution towards a model fitting was from DEM which had a contribution of 54.3 %, followed by the PC1 of reflectance data (9.9%), topographic wetness index (7.1%), aspect (6.9%), PC2 (4.9%), slope (3.4%), and PC3 (2.0%). MAXENT's response curves show how the model's predictions changed as the environmental variables varied (Fig. 2). These plots reflect the dependence of predicted suitability both on the selected variable and on dependencies induced by correlations between the selected variable and other variables. Altitude showed a high percent contribution for many species (Table 4). For some species, the logistic output increased with altitude (*Ch. nivalis*, *M. subterraneus*, *S. araneus*, *C. glareolus*), while for others it decreased (*A. agrarius*, *M. minutus*, *C. suaveolens*), (Fig. 2). Species, relatively little influenced by altitude, showed a strong dependence on PC1 – it can be assumed that the land cover (habitat type) is important for such species as *A. flavicollis*, *N. anomalus*, *D. nitedula* (Fig. 2). TWI had a posi-

tive impact on *N. fodiens*, *A. agrarius*, *M. minutus* (Fig. 2). Aspect had a significant effect on *M. minutus*, *D. nitedula*, *M. subterraneus*, *S. minutus*, *C. leucodon* (Table 4). The shape of the curves indicated a preference of these species for different parts of this gradient (Fig. 2). For *A. sylvaticus*, *M. subterraneus*, *N. anomalus*, *N. fodiens*, *S. araneus*, *S. minutus*, *C. glareolus* habitat suitability increased to low and high values of the gradient, i. e. in predominantly north facing slopes. *C. suaveolens*, *D. nitedula*, *M. arvalis* showed a preference to the lowest values of the gradient (flat places). High suitability at the middle part of the gradient (south and west facing slopes) was manifested in models for *A. epimelas*, *M. spicilegus*. PC2 had a high percent contribution in *A. sylvaticus* and *M. arvalis* (Table 4), showing a preference for high values of the gradient (Fig. 2). The percent contribution of slope was relatively low. It was greatest in the model for *S. minutus* with increasing suitability to the steeper slopes. PC3 had a relatively strong influence on *A. flavicollis* with a characteristic peak in the low values of this gradient (Fig. 2).

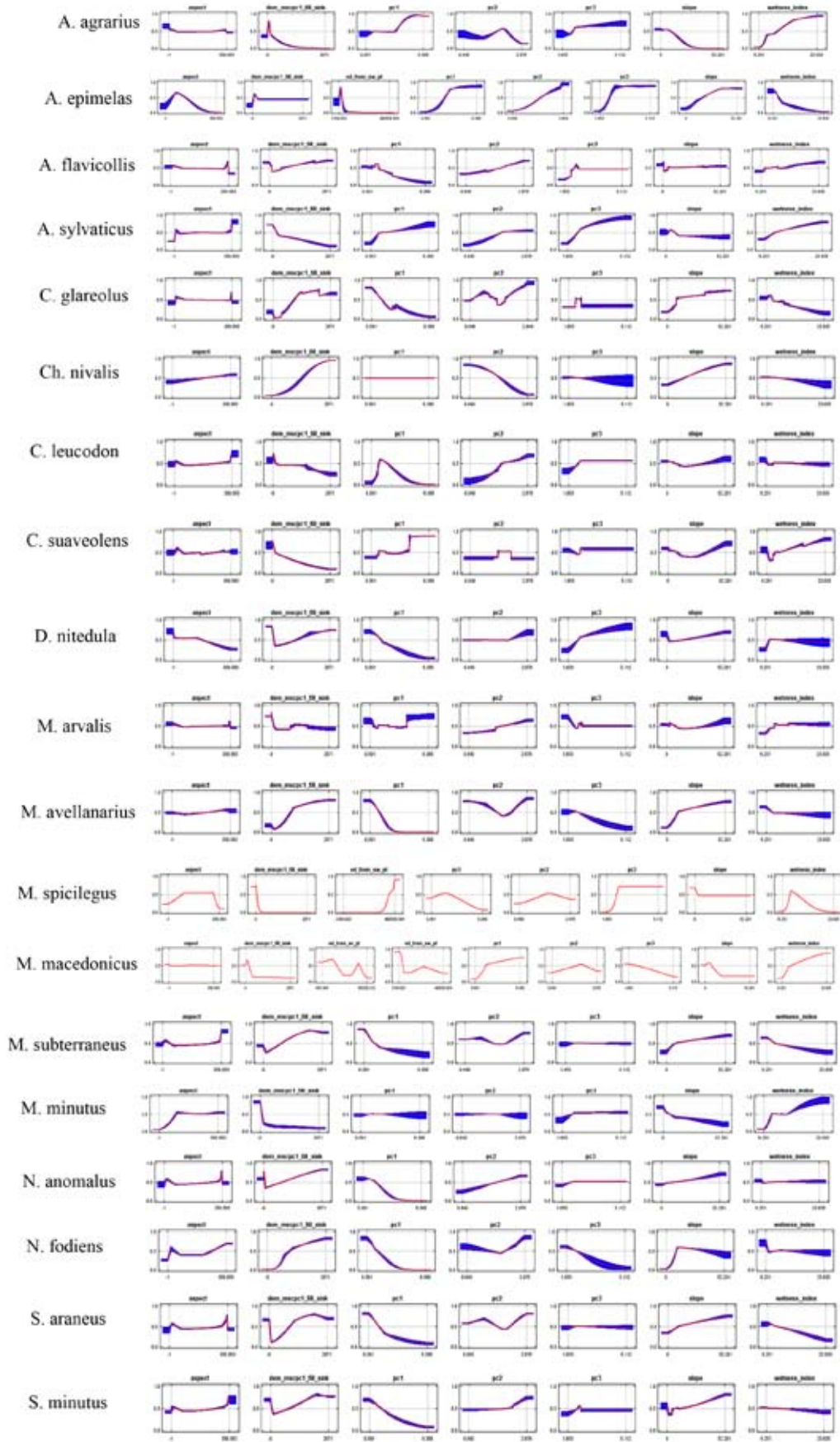


Fig. 2. Response curves of predictors for distribution, reflecting modelled range preferences of species in ecological space. Each curve represents a different model, created using only the corresponding variable

Discussion

Model performance

Comparisons between the three model types showed that MBD- and SF-models give generally lower values of AUC. In the setting of the analysis, this result could be regarded in favor of the interpretation that the output data of the ID-models suffer to some extent from sampling bias and autocorrelation. These comparisons show that probably, for some of the more common and widespread species, the models based on raw data contained a significant component of sampling bias and autocorrelation. Similar results were obtained from KRAMER-SCHADT *et al.* (2013). It was found that test AUCs were considerably reduced when using spatial filtering, while slightly reduced when background manipulation is used. The AUC of the range-restricted species did not decrease with spatial filtering and using bias files put more emphasis on the explanatory power of single variables, that is, the relative gain contribution increased (KRAMER-SCHADT *et al.* 2013). In the present study, the comparison between MBD- and SF-models shows that the latter are very variable across species. Most probably this is a result of a drastic reducing of the number of training points in some cases. The datasets, reduced by spatial filtering are less likely to represent the species' full niche and are often plagued by random vagaries of sampling (noise), both of which can hinder the calibration of realistic models (WISZ *et al.* 2008). Additionally, although this subsampling reduces the spatial autocorrelation it does not correct for the lack of data due to the low sampling effort in some areas. This method could also underestimate the contribution of suitable areas where the high density of occurrences reflects the true environmental suitability for the species (FOURCADE 2014). This is especially true for species with restricted ranges and thus with spatially clumped records. Spatial clumping in these cases is an ecological signal (DORMANN *et al.* 2007), and filtering the records can weaken the prediction. The presented data indicate that among the three model types MBD-models show greatest ecological relevancy. They were especially effective for species that used a wide range within the environmental predictors. Here, correcting for sampling bias prevents overemphasizing environmental variables that putatively correlate with the species' distribution simply due to occurrence records, based on uneven sampling. MBD-models outweigh the negative effects of sampling bias and substantially reduce omission errors (suitable areas not predicted by the models), particularly in areas where surveying was limited.

Higher omission errors can have serious negative consequences for conservation because remote areas with scant surveying effort, which may be potentially important for a species, may be neglected (KRAMER-SCHADT *et al.* 2013). The results of this study confirms that an approach based on correcting for sampling bias is preferable, despite the putatively lower predictive performance according to the AUC_test values (PHILLIPS *et al.* 2009).

Ecological realism of the models

Although the choice of modelling methods and approach to model building and evaluation, is central to good model building (GUISAN and ZIMMERMAN 2000), the maintaining ecological realism throughout the model building process is also important (AUSTIN 2002). In this respect the evaluation of fitted functions throughout model building and interpretation of model outputs from an ecological point of view is fundamental to ensuring the ecological realism and general credibility of the final models (WINTLE *et al.* 2005).

Modelling of the species habitat suitability involves the use of environmental data, which is known to have a direct or indirect impact on a species. Direct factors are generally presented by such ecogeographical variables as temperature, precipitation, humidity, solar radiation, while indirect variables are those which are correlated with the first group and can be considered as their proxies (AUSTIN 2002). Usually climatic variables are a result of elevation-sensitive spatial interpolations of weather station data. Although from a mechanistic point of view, it is desirable to predict the distribution of biotic entities on the basis of direct ecological factors such as climatic variables, they often tend to be less precise than pure topographic characteristics, because of interpolation errors, lack of sufficient climate stations data, and the fact that standard weather stations do not reveal the biologically relevant microclimates (GUISAN and ZIMMERMAN 2000). On the other hand, available DEMs tend to be relatively accurate, even in mountainous terrain. Thus, directly derived topographic variables are generated without much loss of precision (GUISAN and ZIMMERMAN 2000). Four topographic variables were used in present work – altitude, as presented by DEM, aspect, steepness of the slope and topographic wetness index. DEM can be considered as a proxy of the main climatic features. In Bulgaria, temperature and annual temperature amplitude decrease with altitude, while rainfall mostly increased. Aspect represents the local features of the temperature and humidity associated with insolation. Usually southern slopes are drier and warmer while the northern ones are cool and moist. Steepness of

Table 5. Summary statistics of the first three Principal Component scores (PC1 - PC3) of reflectance data within CORINE land cover categories. For a clearer interpretation the data are sorted on the basis of the average values of PC1 and PC2

| Land cover | PC1 | | | | PC2 | | | | PC3 | | | |
|-------------------------------------|------|------|------|------|------|------|------|------|------|------|------|------|
| | Min | Max | X | SD | Min | Max | X | SD | Min | Max | X | SD |
| Coastal lagoons | 0.83 | 4.68 | 1.18 | 0.36 | 0.61 | 2.09 | 0.96 | 0.33 | 1.95 | 3.42 | 2.33 | 0.11 |
| Water bodies | 0.45 | 6.52 | 1.28 | 0.52 | 0.44 | 2.73 | 1.01 | 0.39 | 1.4 | 4.12 | 2.31 | 0.13 |
| Water courses | 0.39 | 6.86 | 1.3 | 0.61 | 0.49 | 2.61 | 1.01 | 0.4 | 1.76 | 3.58 | 2.33 | 0.14 |
| Coniferous forest | 0.39 | 7.12 | 1.3 | 0.37 | 0.51 | 3.07 | 1.54 | 0.22 | 1.18 | 5.3 | 2.19 | 0.08 |
| Mixed forest | 0.3 | 6.91 | 1.41 | 0.33 | 0.56 | 3.25 | 1.83 | 0.25 | 0.24 | 3.79 | 2.22 | 0.08 |
| Salt marshes | 1.02 | 2.71 | 1.42 | 0.32 | 0.68 | 1.93 | 1.1 | 0.39 | 2.1 | 2.66 | 2.45 | 0.11 |
| Broad-leaved forest | 0.45 | 7.19 | 1.46 | 0.26 | 0.31 | 3.4 | 1.98 | 0.19 | 0.91 | 4.39 | 2.25 | 0.08 |
| Salines | 0.79 | 4.67 | 1.56 | 0.46 | 0.35 | 2.03 | 1.01 | 0.25 | 1.92 | 4.48 | 2.58 | 0.33 |
| Moors and heatland | 0.41 | 7.8 | 1.63 | 0.59 | 0.39 | 3.07 | 1.65 | 0.26 | 1.24 | 6.44 | 2.2 | 0.14 |
| Peatbogs | 0.84 | 3.35 | 1.67 | 0.24 | 1.24 | 2.45 | 1.86 | 0.1 | 1.9 | 3.17 | 2.28 | 0.06 |
| Inland marshes | 0.64 | 6.25 | 1.68 | 0.4 | 0.58 | 2.48 | 1.72 | 0.26 | 1.59 | 3.25 | 2.33 | 0.11 |
| Transitional woodland-shrub | 0.44 | 7.37 | 1.88 | 0.5 | 0.54 | 3.25 | 1.87 | 0.18 | 0.09 | 7.09 | 2.2 | 0.11 |
| Green urban areas | 0.56 | 6.62 | 1.93 | 0.6 | 0.58 | 2.55 | 1.74 | 0.18 | 0.78 | 4.32 | 2.29 | 0.11 |
| Sport and leisure facilities | 0.48 | 7.26 | 1.99 | 0.67 | 0.42 | 2.86 | 1.76 | 0.22 | 1.06 | 4.58 | 2.27 | 0.14 |
| Agricultures and Natural vegetation | 0.46 | 8.09 | 2.13 | 0.56 | 0.25 | 3.2 | 1.88 | 0.18 | 0.3 | 6.56 | 2.23 | 0.12 |
| Fruit trees and berry plantations | 0.47 | 6.69 | 2.18 | 0.48 | 0.47 | 2.73 | 1.82 | 0.13 | 0.13 | 3.97 | 2.21 | 0.12 |
| Natural grassland | 0.34 | 7.87 | 2.21 | 0.56 | 0.39 | 3.14 | 1.89 | 0.16 | 0.12 | 7.02 | 2.19 | 0.13 |
| Complex cultivation patterns | 0.43 | 7.61 | 2.22 | 0.54 | 0.44 | 3.11 | 1.89 | 0.15 | 0.26 | 5.2 | 2.23 | 0.12 |
| Pastures | 0.56 | 7.56 | 2.29 | 0.45 | 0.41 | 2.98 | 1.86 | 0.13 | 0.91 | 4.2 | 2.21 | 0.11 |
| Road and railnetworks | 0.94 | 6.7 | 2.31 | 0.49 | 0.28 | 2.45 | 1.66 | 0.21 | 1.11 | 5.21 | 2.3 | 0.16 |
| Vineyards | 0.62 | 6.86 | 2.31 | 0.46 | 0.5 | 2.89 | 1.82 | 0.13 | 0.1 | 4.23 | 2.21 | 0.12 |
| Urban fabric | 1.09 | 6.49 | 2.34 | 0.54 | 0.74 | 2.36 | 1.44 | 0.17 | 1.66 | 3.8 | 2.29 | 0.15 |
| Non irrigated arable land | 0.43 | 7.55 | 2.37 | 0.49 | 0.3 | 3.39 | 1.75 | 0.16 | 0 | 5.22 | 2.25 | 0.17 |
| Rice fields | 0.77 | 5.96 | 2.46 | 0.67 | 0.54 | 2.49 | 1.62 | 0.26 | 1.44 | 3.66 | 2.34 | 0.14 |
| Industrial or commersila units | 0.46 | 8 | 2.56 | 0.63 | 0.22 | 3.44 | 1.7 | 0.22 | 0.58 | 5.67 | 2.32 | 0.18 |
| Sparcely vegetated areas | 0.4 | 8.01 | 2.64 | 0.68 | 0.24 | 2.8 | 1.78 | 0.19 | 1.11 | 6.95 | 2.26 | 0.32 |
| Mineral extraction sites | 0.47 | 7.8 | 2.79 | 0.9 | 0.25 | 2.84 | 1.7 | 0.24 | 1.13 | 5.01 | 2.36 | 0.3 |
| Dump sites | 0.58 | 6.87 | 2.82 | 1.29 | 0 | 2.61 | 1.44 | 0.4 | 0.97 | 5.88 | 2.69 | 0.72 |
| Beaches, dunes | 0.63 | 6.05 | 2.91 | 1.02 | 0.57 | 2.62 | 1.74 | 0.31 | 1.75 | 3.61 | 2.32 | 0.21 |
| Bare rock | 0.32 | 7.39 | 2.93 | 0.94 | 0.14 | 2.59 | 1.68 | 0.27 | 1.15 | 7.07 | 2.46 | 0.67 |

the slope represents local features associated with the substrate and humidity. As a rule, steep slopes do not retain water, the soil layer is negligible, the vegetation is poorly developed. In contrast flat areas accumulate moisture, have a thick soil layer and as a rule – a good vegetation cover. The topographic wetness index represents water accumulation as a function of slope and catchment (PAROLO *et al.* 2008, FRANKLIN 2009). In general DEM can be considered as a proxy of regional climate while its derivatives describe the more local characteristics related to temperature, humidity, and condition of the substrate.

Another important local factor is land-cover, usually representing the vegetation cover. In contrast

to the above geomorphological factors, it reflects also the human impact on the landscape. It can be expected that land-cover exerts prevailing control of species' distribution at a finer spatial resolution than climate (PEARSON *et al.* 2004). In the present study land-cover is evaluated on the basis of three PC derived from Raw Landsat 7 ETM+ bands TM1/2/3/4/5/7. Although, remote sensing can be particularly useful when there is a need for complete spatial coverage over large areas, the use of raw reflectances in modelling species distributions poses some challenges. Information recorded by satellite sensors is typically restricted to energy returned in one or more wave bands and relative geographic position and is not

easily interpretable from an ecological point of view which may limit its relevance in revealing species–habitat associations during modelling process and evaluation. In this respect it is necessary to analyze the spectral information in a greater detail trying to interpret in terms of environmental gradients of ecological importance for a particular group of organisms. In order to elucidate the ecological significance of the first three PCs, an analysis of the relationship between the principal component scores and the land cover classes of the CORINE land cover was made. For this purpose, basic statistics (minimum, maximum, mean, standard deviation) of principal component scores were calculated for the third level categories of Corine Land Cover (Table 5).

It can be seen that PC1 reflects very well the nature of vegetation, as the mean score values form a clear gradient from forest vegetation and moist places (low mean scores of PC1) to xerophyllous bare surfaces (high mean scores of PC1). It is more difficult to evaluate the second principal component as a linear gradient of the vegetation types. From the data in Table 5, it is clear that it separates the coniferous forests (low mean scores) compared to other types of vegetation, mainly to deciduous forests (high mean scores). Open places and bushes occupy an intermediate position. In mountainous areas, it reflects very well the belts of forest vegetation. The PC3 scores show a little contrast (Table 5), but still separate the sites covered with vegetation (lower mean scores) from those without vegetation, including most places of anthropogenic origin (higher mean scores).

The review of the data in Table 5 shows another important feature of the reflectance data. Variability of the scores of the three PCs is very high within each of the CORINE land cover classes. Each class contains almost the entire range of possible values of the corresponding principal component scores. Analysis of detailed maps shows that in most part this is a result of the fact that the land cover classes do not reflect the fine heterogeneity of the landscape. CORINE has a minimum unit mapping size of 25 hectares vs 0.09 hectares of Landsat 7 ETM+.

On the base of these analyzes it can be stated that the spectral values and their derivatives – the principal components – contain much environmental information. The presented interpretations will help for an easier evaluation of the ecological relevancy of the modelling results. Final models were inspected for their ecological realism on the basis of relative contribution of ecogeographical variables and the shape of response curves within the context of the known ecology of species (POPOV 2007). In this light, the modelling results show that the general climatic

conditions, as presented by DEM, are the main factors determining the distribution of small mammals in the country. This is confirmed by the seemingly strange response curves of certain species to this factor. For example, the curves of *Sorex araneus*, *S. minutus*, *N. anomalus*, *M. subterraneus*, *D. nitedula*, *C. glareolus*, *A. flavicollis* show a high suitability at a high altitude, but also at the sea level. This can be explained by the fact that the coastal areas in eastern Bulgaria show some climatic characteristics similar to the mountainous regions – they have a higher humidity and a narrow annual temperature amplitude, due to the relatively low summer temperatures. Under these conditions, in the river valleys and near the beach, conditions similar to those in the middle mountain belt arise – relatively high soil moisture, well developed forest vegetation and as a result cooler and more humid microclimates. Therefore, some cold-loving mesophilous forest dwellers that have an optimal habitat in the mountainous parts of the country occur also near the seashore.

Response curves of many species on the PC1 of reflectance data, the second most important factor, show monotonous responses. This corresponds to the relatively linear gradient that this factor presents – from damp forested habitats to dry and open habitats. The reaction of the species highly dependent on this factor is consistent with this interpretation. Forest species show a high preference for low values of this gradient – wet and wooded habitats. In contrast, species associated with open areas have a high degree of probability of occurrence in higher values of this gradient.

Response curves on the PC2 of reflectance data represent the species preference for three vegetation types – coniferous forests, open habitats and shrubs and deciduous forests. Species preferring forests have curves with high values on both the left and right side of the scale (*C. glareolus*, *M. avellanarius*, *M. subterraneus*), those inhabiting predominantly deciduous forests have high values on the right side of the scale (*C. leucodon*, *D. nitedula*), those with peaks in the middle part of the gradient are associated with open areas and shrubs (*C. suaveolens*, *A. agrarius*, *M. macedonicus*, *M. spicilegus*). Response curves for the aspect of the slope also show important features of the species ecology. For many species they have two peaks near 0° and 360° indicating preference for slopes with northern exhibitions. In most cases, these are species that prefer moist wooded areas at a higher altitude – *S. minutus*, *S. araneus*, *N. fodiens*, *M. subterraneus*, *C. leucodon*, *C. glareolus*, *A. flavicollis*. In the lower parts of the country, these species occur in sites with wetter and cooler

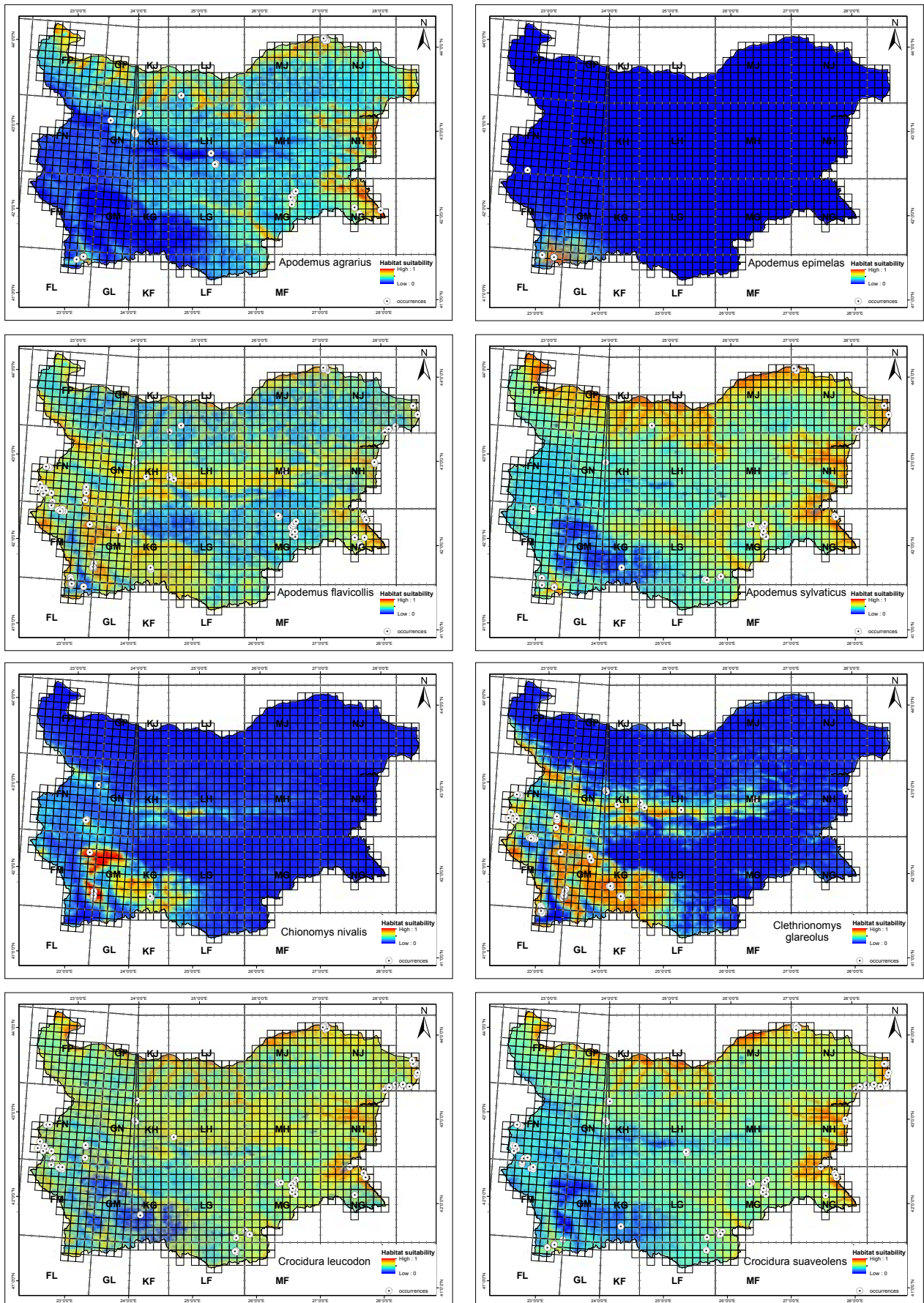


Fig. 3. Continued

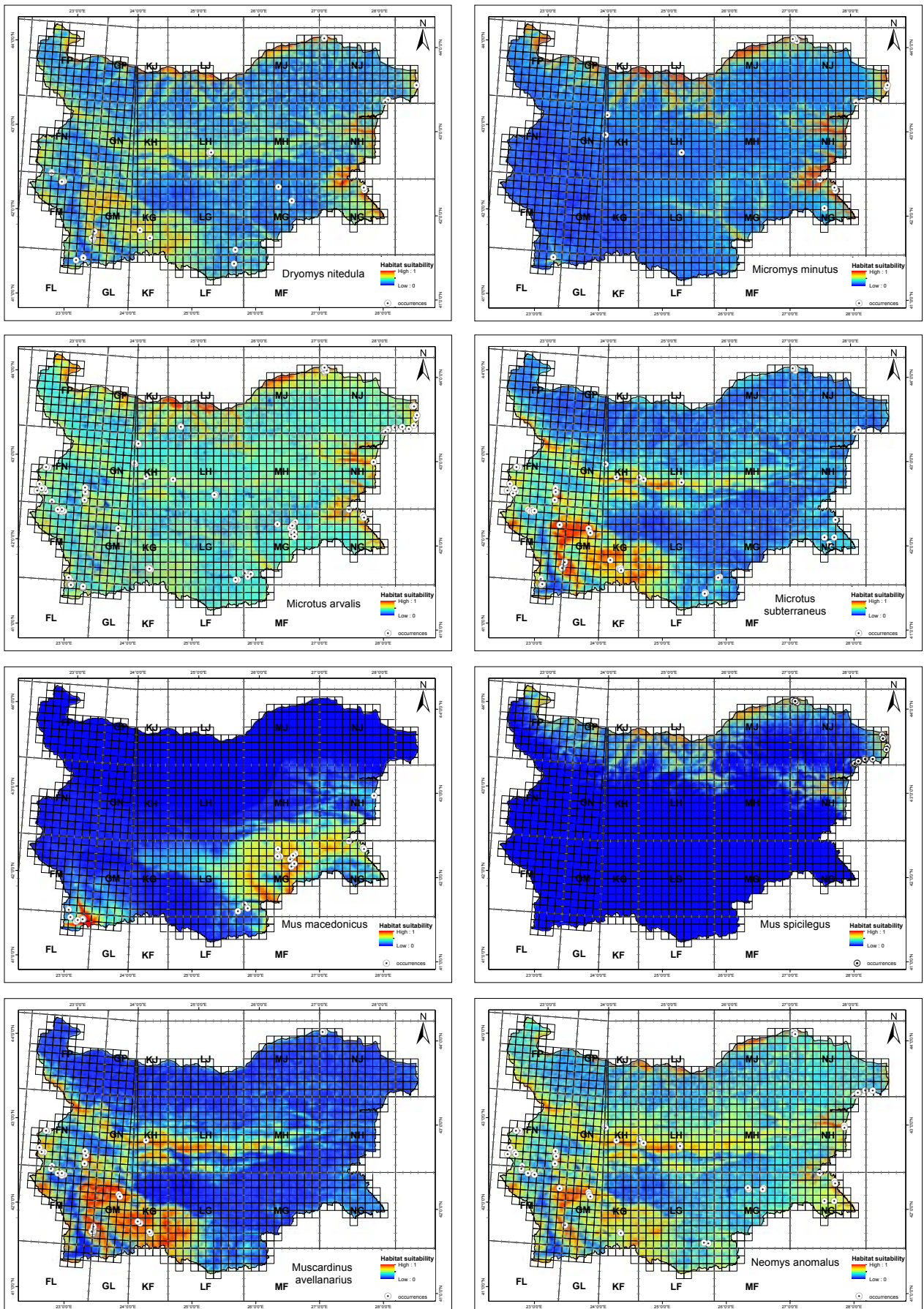


Fig. 3. Continued

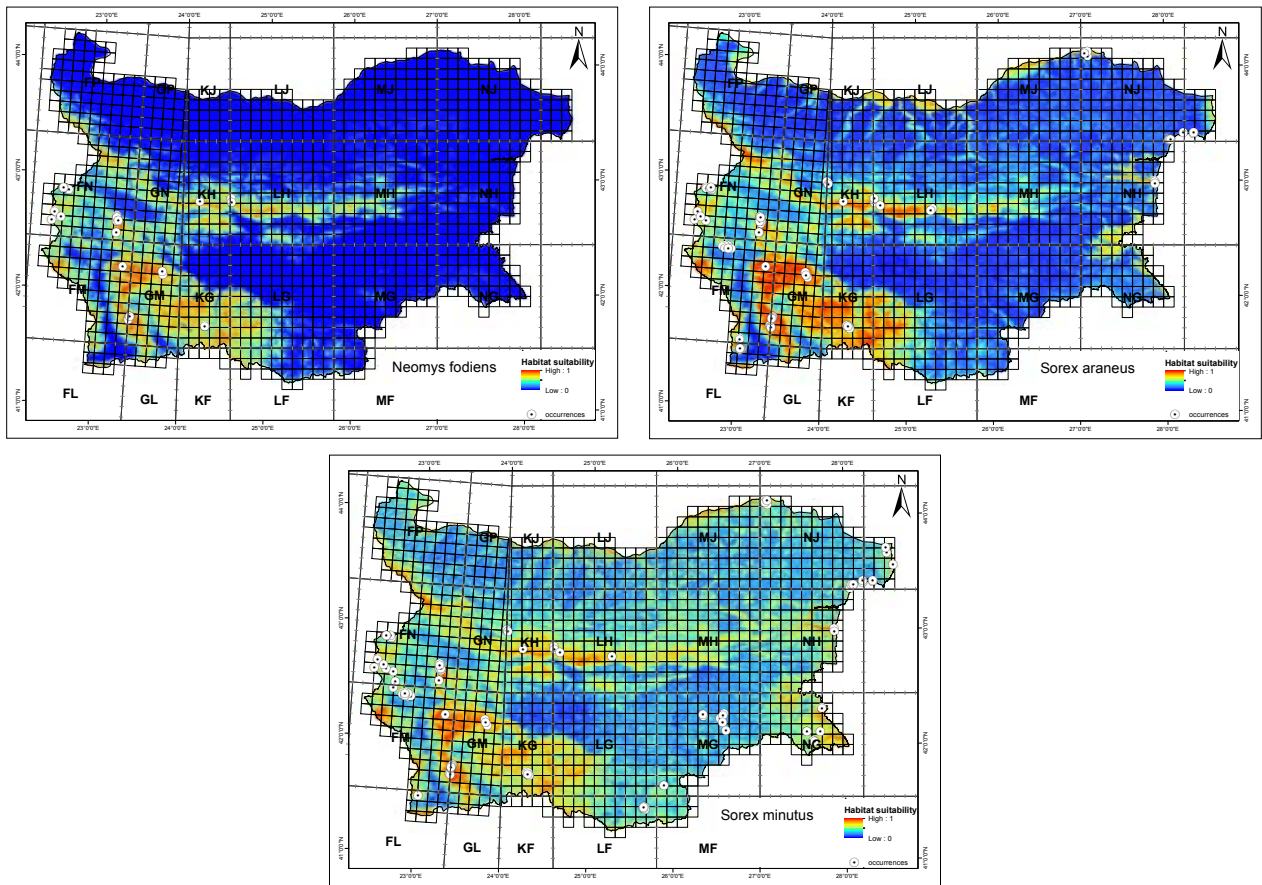


Fig. 3. Occurrences and modelled distribution (MBD-models) of 19 common small mammal species in Bulgaria (grid represents 10 x 10 km UTM squares)

microclimate. Usually these are places in north facing slopes. In contrast, some species that are known to be thermophilic and preferring habitats at a low altitude show an optimum at values around $90^{\circ} - 180^{\circ}$ and $90^{\circ} - 270^{\circ}$ – *M. macedonicus*, *A. epimelas*, *M. minutus*. The slope has a limited influence on species distribution. Most species occurring primarily in the mountains have a higher suitability at the higher values of the gradient unlike the species distributed in the lower parts of the country. An exception is *A. epimelas* that although only found in the lower parts of the country shows a preference for steep slopes. This is consistent with its ecology and preference for rocky outcrops and slopes with poor soil.

The above analyzes show that the species-environmental relationships registered by the models correspond very well to the known ecology of the studied species. This is particularly important with respect to the satellite variables. It is evident that they reflect significant environmental gradients mostly related to the nature of the vegetation and humidity. The response of the small mammals along these gradients confirms this interpretation. When comparing the satellite variables with land cover classes a significant variability

of satellite data within individual classes is found. This is obviously a result of the larger resolution of satellite data. Land cover classification removes within-class heterogeneity by the omission of fine-scale local features (BELLIS *et al.* 2008, ST-LOUIS *et al.* 2009). The homogeneity of the land cover classes (“thematic resolution”) leads to the circumstance that many occurrence points from different small habitats fall into one category, and in contrast, occurrence records from similar small habitats fall into different land cover classes. Furthermore, the boundaries of CORINE classes are too inaccurate, and therefore a lot of occurrence records may fall into the neighboring class often with different ecological interpretation. It was observed during the fieldwork and in comparison with Landsat imagery, that CORINE data were at times of poor spatial accuracy (deviations of up to 1 km) at known species occurrence sites, especially near polygon borders. In this sense satellite data can be regarded as more suitable for modelling small and less mobile (small home ranges) animals.

These analyzes show that in addition to the good statistical performance, the models have a high degree of ecological realism. They reflect both

the general type of distribution of each species in the country, and peculiarities associated with local conditions. In this sense, the resulting maps of the habitat suitability for 19 small mammal species covering the whole country in a resolution of 900 m² would be useful for many practical aspects such

as epidemiology, conservation, management of protected areas, landscape planning, etc.

Acknowledgement: The author is grateful to three Anonymous Referees for their valuable comments that improved the manuscript.

References

- AUSTIN M. P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modeling. – *Ecological Modeling*, **157**: 101-118.
- BELLIS L.M., A.M. PIDGEON, V. C. RADELOFF, V.ST-LOUIS, J. L. NAVARRO AND M.B. MARTELLA 2008. Modeling habitat suitability for greater rheas based on satellite image texture. – *Ecological Applications*, **18**: 1956-1966.
- BERON P., V. BESHKOV, V.POPOV, M. VASSILEV, R. PANDURSKA, T. IVANOVA 2000a. Biodiversity of small vertebrates (Pisces, Amphibia, Reptilia, Mammalia-Insectivora, Chiroptera, Lagomorpha and Rodentia) in Central Balkan National Park. Pp. 363-391 in: Biological Diversity of the Central Balkan National Park. PENSOFT, Sofia.
- BERON P., V. BESHKOV, V.POPOV, M. VASSILEV, R. PANDURSKA, T. IVANOVA 2000b. Biodiversity of small vertebrates (Pisces, Amphibia, Reptilia, Mammalia-Insectivora, Chiroptera, Lagomorpha and Rodentia) in the Rila National Park. Pp. 333-360 in: Biological Diversity of the Rila National Park. PENSOFT, Sofia.
- BRADLEY B. A., E. FLEISHMAN 2008. Can remote sensing of land cover improve species distribution modelling.- *Journal of Biogeography*, **35**: 1158-1159.
- CORD A. F., D. KLEIN, F. MORA, S. DECH 2014. Comparing the suitability of classified land cover data and remotesensing variables for modeling distribution patterns of plants. – *Ecological Modelling*, **272**: 129-140.
- DORMANN C. F., J. M. MCPHERSON, M. B. ARAU'JO, R. BIVAND, J. BOLLIGER, G. CARL, R. G. DAVIES, A. HIRZEL, W. JETZ, W. DANIEL KISSLING, I. KUHN, R. OHLEMLER, P. R. PERES-NETO, B. REINEKING, B. SCHRÖDER, F. M. SCHURR and R. WILSON 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. – *Ecography*, **30**: 609-628.
- DUDIK M., R. E. SCHAPIRE and S.J. PHILLIPS 2005. Correcting sample selection bias in maximum entropy density estimation. *Advances in Neural Information Processing Systems*, 18 (eds Y. Weiss, B. Schölkopf & J. Platt), pp. 323-330. MIT Press, Cambridge, Massachusetts, USA.
- EAA. 2000. CORINE Land Cover Database. EU DG XI/European Environmental Agency/ JRC & ETC/LC, Brussels.
- FOURCADE Y., J. O. ENGLER, D. RÖDDER, J. SECONDI. 2014. Mapping species distributions with MAXENT using a geographically biased sample of presence data: a performance assessment of methods for correcting sampling bias. – *PLoS one*, **9** (5): e97122.
- FRANKLIN J. 2009. Mapping species distributions spatial inference and prediction. Cambridge University Press, Cambridge, UK, 320 pp.
- GOSLEE S. C. 2011. Analyzing remote sensing data in R: The landsat package. – *Journal of Statistical Software*, **43** (4): 1-25.
- GRASS Development Team, 2008. Geographic resources analysis support system (grass gis) software. <http://grass.osgeo.org>.
- GUISAN A., W. THULLER 2005. Predicting species distribution: Offering more than simple habitat models. – *Ecology Letters*, **8**: 993-1009.
- GUISAN A., N. E. ZIMMERMANN 2000. Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**: 147-186.
- KERR J.T., M. OSTROVSKY 2003. From space to species: ecological applications for remote sensing. – *Trends in Ecology and Evolution*, **18**: 299-305.
- KRAMER-SCHADT S., J. NIEDBALL, J. D. PILGRIM, B. SCHRÖDER, J. LINDENBORN, V. REINFELDER, M. STILLFRIED, I. HECKMANN, A. K. SCHARF, D. M. AUGERI, S. M. CHEYNE, A. J. HEARN, J. ROSS, D. W. MACDONALD, J. MATHAI, J. EATON, A. J. MARSHALL, G. SEMIADI, R. RUSTAM, H. BERNARD, R. ALFRED, H. SAMEJIMA, J. W. DUCKWORTH, C. BREITENMOSER-WUERSTEN, J. L. BELANT, H.T. HOFER and A. WILTING 2013. The importance of correcting for sampling bias in MaxEnt species distribution models. – *Diversity and Distributions*, **1**: 1-14.
- KÜHN, I. 2007. Incorporating spatial autocorrelation may invert observed patterns. – *Diversity and Distributions*, **13**: 66-69.
- LEYEQUIEN, E., J. VERRELST, M. SLOT, G. SCHAEPMAN-STRUB, I. M. A. HEITKONIG, A. SKIDMORE 2007. Capturing the fugitive: applying remote sensing to terrestrial animal distribution and diversity. – *International Journal of Applied Earth Observation and Geoinformation*, **9**: 1-20.
- LOBO J. M., A. JIMÉNEZ-VALVERDE and R. REAL 2008. AUC: a misleading measure of the performance of predictive distribution models. – *Global Ecology and Biogeography*, **17**: 145-151.
- LUOTO M., R. VIRKKALA and R. K. HEIKKINEN 2007. The role of land cover in bioclimatic models depends on spatial resolution. – *Global Ecology and Biogeography*, **16**: 34-42.
- MINKOVA T. V., V. V. POPOV 2002. Spatial patterns of terrestrial small mammal communities in Central Western Bulgaria (Mammalia: Insectivora, Rodentia). – *Acta zoologica bulgarica*, **54** (3): 55-74.
- MÜCHERA C. A., S. M. HENNEKENS, R. G.H. BUNCEA, J. H.J. SCHAMINÉEA, M. E. SCHAEPMANC 2009. Modelling the spatial distribution of Natura 2000 habitats across Europe. – *Landscape and Urban Planning*, **92**: 148-159.
- NASA Land Processes Distributed Active Archive Center (LP DAAC). 2014. ASTER L1B. USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota.
- PAROLO G., G. ROSSI, and A. FERRARINI 2008. Toward improved species niche modelling: *Arnica montana* in the Alps as a case study. – *Journal of Applied Ecology*, **45**: 1410-1418.
- PEARSON R. G., T. P. DAWSON and C. LIU 2004. Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. – *Ecography*, **27**: 285-298.
- PEARSON R. G, C. J. RAXWORTHY, M. NAKAMURA and A. T. PETERSON 2007. Predicting species distributions from small numbers

- of occurrence records: a test case using cryptic geckos in Madagascar. – *Journal of Biogeography*, **24**: 102-117.
- PESHEV TS., D. PESHEV, and V. POPOV 2004. The Fauna of Bulgaria. Vol. 27. Mammalia. Acad. Publ. House “Acad. Marin Drinov”, Sofia. 632 p. (In Bulgarian).
- PETERSON A. T., M. PAPES and M. EATON 2007. Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. – *Ecography*, **30**: 550-560.
- PETTORELLI N., S. RYAN, T. MUELLER, N. BUNNEFELD, B. JEDRZEJEWSKA, M. LIMA, K. KAUS-RUD 2011. The Normalized Difference Vegetation Index (NDVI): unforeseen successes in animal ecology. – *Climate Research*, **46**: 15-27.
- PHILLIPS J., R. P. ANDERSON and R. E. SCHAPIRE 2006. Maximum entropy modeling of species geographic distributions. – *Ecological Modelling*, **190**: 231-259.
- PHILLIPS S.J., M. DUDÍK 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. – *Ecography*, **31**: 161-175.
- PHILLIPS, S.J., M. DUDÍK, J. ELITH, C. H. GRAHAM, A. LEHMANN, J. R. LEATHWICK AND S. FERRIER 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. – *Ecological Applications*, **19**: 181-197.
- POPOV V. 1999. Estimation of the faunistic diversity of terrestrial small mammals (Mammalia: Insectivora, Lagomorpha, Rodentia) in Vitosha national Park. In: “65 years of Park Vitosha”. Litera, Sofia, 12-13.
- POPOV V. 2000. Epigeobiont animal assemblages from two landscapes of the Bulgarian Black sea coast: relationship to environmental gradients, assemblage structure and biodiversity. III. Small mammals (Insectivora, Rodentia). – *Acta zoologica bulgarica*, **52** (3): 75 -87.
- POPOV V. 2007. Biogeographical and ecological spatial patterns of terrestrial mammals in Bulgaria. – In: Fet, V., Popov, A. (Eds.): Ecology and Biogeography of Bulgaria. Monographiae Biologicae, Vol. **82**: 9-38, Spinger.
- POPOV V., I. PANDURSKI, R. PANDURSKA-WHITCHER, V. BESHKOV 2006. Small mammals (Insectivora, Chiroptera, Lagomorpha, Rodentia) in the area of Strandzha Mountain, South-Eastern Bulgaria. In: Challenges of establishment and management of a trans-border biosphere reserve between Bulgaria and Turkey in Strandzha Mountain. UNESCO-BAS-MOEW. 87-104.
- POPOV V., N. SPASSOV, T. IVANOVA, B. MIHOVA, K. GEORGIEV 2007. Mammals important for conservation in Bulgaria. (in Bulgarian). Dutch Mammal Society VZZ, Arnhem, The Netherlands, 328 p.
- Quantum and Development Team, 2011. Quantum GIS Geographic Information System. Open Source Geospatial Foundation. <http://qgis.osgeo.org>.
- R Development Core Team 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- RENNER I. W., D. I. WARTON 2013. Equivalence of MAXENT and Poisson point process models for species distribution modeling in ecology. – *Biometrics*, **69** (1): 274-281.
- RONDININI C., K. A. WILSON, L. BOITANI, H. GRANTHAM and H.P. POSSINGHAM 2006. Tradeoffs of different types of species occurrence data for use in systematic conservation planning. – *Ecology Letters*, **9**: 1136-1145.
- SEGURADO P., M. B. ARAÚ 2004. An evaluation of methods for modelling species distributions. – *Journal of Biogeography*, **31**: 1555-1568.
- SHIRLEY S. M., Z. YANG, R. A. HUTCHINSON, J. D. ALEXANDER, K. MCGARIGAL and M. G. BETTS. 2013. Species distribution modelling for the people: unclassified landsat TM imagery predicts bird occurrence at fine resolutions. – *Diversity and Distributions*, **1**: 1-12.
- SICHANOV D., V. POPOV, V. BISERKOV, S. ZIDAROVA, N. CHIPEV 2006. Spatial pattern and diversity of small mammal assemblages in the area of Srebarna Managed Nature Reserve (NE Bulgaria). – *Acta zoologica bulgarica*, **58** (2): 209-222.
- ST-LOUIS V., A.M. PIDGEON, M. K. CLAYTON, B. A. LOCKE, D. BASH and V.C. RADELOFF 2009. Satellite image texture and a vegetation index predict avian biodiversity in the Chihuahuan Desert of New Mexico. – *Ecography*, **32**: 468-480.
- SWETS J. A. 1988. Measuring the accuracy of diagnostic systems. – *Science*, **240**: 1285-1.
- SYFERT M. M., M. J. SMITH, D. A. COOMES 2013. The Effects of Sampling Bias and Model Complexity on the Predictive Performance of MaxEnt Species Distribution Models. – *PLoS One*, **8** (2): e55158.
- TATTONI C., F. RIZZOLLI, P. PEDRINI 2012. Can LiDAR data improve bird habitat suitability models? – *Ecological Modelling*, **245**: 103-110.
- THOMAS A., T. KERVYN, D. MAES. 2008. Applying species distribution modelling for the conservation of the threatened saproxylic Stag Beetle (*Lucanus cervus*). – *Biological Conservation*, **141**: 1400-1410.
- TURNER W., S. SPECTOR, N. GARDINER, M. FLADELAND, E. STERLING and M. STEININGER 2003. Remote sensing for biodiversity science and conservation. – *Trends in Ecology and Evolution*, **18** (6): 306-314.
- VELOZ S. D. 2009. Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. – *Journal of Biogeography*, **36**: 2290-2299.
- WINTLE B. A., J. ELITH and J. M. POTTS 2005. Fauna habitat modelling and mapping: A review and case study in the Lower Hunter Central Coast region of NSW. – *Austral Ecology*, **30**: 719-738.
- WISZ M. S., R. J. HJIMANS, J. LI, A. T. PETERSON, C. H. GRAHAM, A. GUISAN, NCEAS Predicting Species Distributions Working Group. 2008. Effects of sample size on the performance of species distribution models. – *Diversity and Distribution*, **14**: 763-773.
- WOODCOCK C. E., R. ALLEN, M. ANDERSON, A. BELWARD, R. BINDSCHADLER, W. COHEN, F. GAO, S. N. GOWARD, D. HELDER, E. HELMER, R. NEMANI, L. OREOPOULOS, J. SCHOTT, P. S. THENKABAIL, E. F. VERMOTE, J. VOGELMANN, M. A. WULDER, R. WYNNE, Landsat Sci, T. 2008. Free access to Landsat imagery. – *Science*, **320**: 1011.
- ZIMMERMANN N. E., T. C. EDWARDS JR., G. G. MOISEN, T. S. FRESCINO and J. A. BLACKARD 2007. Remote sensing-based predictors improve distribution models of rare, early successional and broadleaf tree species in Utah. – *Journal of Applied Ecology*, **44**: 1057-1067.

Received: 21.11.2014

Accepted: 11.02.2015