

Pre-print of the final, accepted manuscript

Somodi I, Molnár Zs, Czúcz B, Bede-Fazekas Á, Bölöni J, Pásztor L, Laborczi A, Zimmermann NE (2017): Implementation and application of Multiple Potential Natural Vegetation models – a case study of Hungary. *Journal of Vegetation Science* 28(6): 1260–1269. DOI: 10.1111/jvs.12564

## **Implementation and application of Multiple Potential Natural Vegetation models – a case study of Hungary**

Imelda Somodi, Zsolt Molnár, Bálint Czúcz, Ákos Bede-Fazekas, János Bölöni, László Pásztor, Annamária Laborczi & Niklaus E. Zimmermann

**Somodi, I.** (corresponding author, somodi.imelda@okologia.mta.hu)<sup>1</sup>,

**Molnár, Zs.** (molnar.zsolt@okologia.mta.hu)<sup>1</sup>,

**Czúcz, B.** (czucz.balint@okologia.mta.hu)<sup>1,2</sup>,

**Bede-Fazekas, Á.** (bede-fazekas.akos@okologia.mta.hu)<sup>1,3,4</sup>,

**Bölöni, J.** (boloni.janos@okologia.mta.hu)<sup>1</sup>,

**Pásztor, L.** (pasztor@rissac.hu)<sup>5</sup>,

**Laborczi, L.** (laborczi@rissac.hu)<sup>5</sup>,

**Zimmermann, N. E.** (niklaus.zimmermann@wsl.ch)<sup>6,7</sup>

<sup>1</sup>MTA Centre for Ecological Research, 3 Klebelsberg Kunó utca, 8237 Tihany, Hungary;

<sup>2</sup>European Topic Centre on Biological Diversity, Muséum national d'Histoire naturelle, 57 rue Cuvier, FR-75231 Paris, Paris Cedex 05, France

<sup>3</sup>Department of Garden and Open Space Design, Faculty of Landscape Architecture and Urbanism, Szent István University, 29-43 Villányi út, Budapest, Hungary;

<sup>4</sup>National Adaptation Centre, Geological and Geophysical Institute of Hungary, 14 Stefánia út, 1143 Budapest, Hungary

<sup>5</sup>Department of Environmental Informatics, Institute for Soil Sciences and Agricultural Chemistry, MTA Centre for Agricultural Research, 15 Herman Ottó út 1022 Budapest, Hungary;

<sup>6</sup>Swiss Federal Research Institute WSL, 111 Zürcherstrasse, CH-8903 Birmensdorf, Switzerland

<sup>7</sup>Department of Environmental Systems Science, Swiss Federal Institute of Technology ETH, CH-8092 Zurich, Switzerland

Printed journal page estimate: 5902 words (7.4 pages), 0 table (0 pages), 4 figures (1.5 pages), total 8.9 pages.

## **Abstract**

### **Questions**

Multiple Potential Natural Vegetation (MPNV) is a framework for the probabilistic and multilayer representation of potential vegetation in an area. How can an MPNV model be implemented and synthesized for the full range of vegetation types across a large spatial domain such as a country? What additional ecological and practical information can be gained compared to traditional Potential Natural Vegetation (PNV) estimates?

### **Location**

Hungary

### **Methods**

MPNV was estimated by modelling the occurrence probabilities of individual vegetation types using gradient boosting models (GBM). Vegetation data from the Hungarian Actual Habitat Database (MÉTA) and information on the abiotic background (climatic data, soil characteristics, hydrology) were used as inputs to the models. To facilitate MPNV interpretation a new technique for model synthesis (rescaling) enabling comprehensive visual presentation (synthetic maps) was developed which allows for a comparative view of the potential distribution of individual vegetation types.

### **Results**

The main result of MPNV modelling is a series of raw and rescaled probability maps of individual vegetation types for Hungary. Raw probabilities best suit within-type analyses, while rescaled estimations can also be compared across vegetation types. The latter create a synthetic overview of a location's PNV as a ranked list of vegetation types, and make the comparison of actual and potential landscape composition possible. For example, a representation of forest vs grasslands in MPNV revealed a high level of overlap of the potential range of the two formations in Hungary.

### **Conclusion**

The MPNV approach allows for viewing the potential vegetation composition of locations in far more detail than the PNV approach. Rescaling the probabilities estimated by the models allows easy access to the results by making potential presence of vegetation types with different data structure comparable for queries and synthetic maps. The wide range of applications identified for MPNV (conservation and restoration prioritisation, landscape evaluation) suggests that the PNV concept with the extension towards vegetation distributions is useful both for research and applications.

## **Keywords**

conservation prioritization; landscape evaluation; multilayer model; PNV; predictive vegetation model; probability rescaling; restoration; probability distribution of vegetation types; vegetation stochasticity

**Nomenclature** Bölöni et al. (2011) for vegetation types

## **Abbreviations**

AUC – Area Under the Receiver Operating Characteristic (ROC) Curve

GBM – Gradient Boosting Model

MÉTA - Hungarian Actual Habitat Database

MPNV – Multiple Potential Natural Vegetation

PNV – Potential Natural Vegetation

Running head: Multiple potential natural vegetation of Hungary

## Introduction

Human disturbance has greatly transformed our environment and this has strongly affected the natural vegetation around the globe (Reynolds & Hessburg 2005; Jackson & Hobbs 2009; Kaplan et al. 2009). The high level of land transformation masks the vegetation potential of the landscape, and due to the low proportion and partial quasi-absence of natural and semi-natural remnants, it is now difficult to judge which vegetation type would cover the landscape in the absence of human land use. Therefore, knowledge of the potential natural vegetation (PNV) represents a crucial baseline for effective conservation and restoration actions, and for estimating the degree of habitat loss per landscape (Kowarik 1987; Prach et al. 2016).

An estimation of the prehuman vegetation is often treated as equivalent of PNV estimation (e.g. Hall & McGlone 2006; Carrión & Fernandez 2009). Such an estimation may be misleading, however, because environmental conditions have clearly changed since prehuman times across Europe (Kowarik 1987, Dotterweich 2008; Sillasoo et al. 2009). An alternative approach is to assess PNV from a comprehensive assessment of present day natural vegetation remnants, which offers more reliable estimates of what the natural vegetation would be under current environmental conditions (Tüxen 1956; Kowarik 1987; Somodi et al. 2012). Vegetation distribution models (similar to species distribution models; Zimmermann & Kienast 1999; Guisan & Zimmermann 2000; Elith et al. 2006) can be used to project PNV across space and time (e.g. Bittner et al. 2011).

The PNV concept has undergone changes and has frequently been criticised (see eg Carrión & Fernandez 2009; Chiarucci et al. 2010). This critique was partly due to misunderstandings, and partly due to a lack of mathematical formulation of the concept (Loidi et al. 2010; Somodi et al. 2012). Recently, several papers have presented PNV estimations by formal methods (Hemsing & Bryn 2012; Fischer et al. 2013; Attorre et al. 2014; Reger et al. 2014) and for mapping the most likely PNV type at any given location. However, there are two reasons why this approach may lead to oversimplification. (1) There are varying degrees of similarity among vegetation units and consequently they can not be viewed as categorical but rather as a fuzzy set (e.g. Roberts 2015). (2) Multiple stable states have been described from undisturbed environmental settings (e.g. Baker & Walford 1995, Petraitis 2013). It can therefore be argued that mapping solely the most likely vegetation type bears a loss of information regarding the local vegetation potential. Another solution is to group vegetation complexity into categories containing more than one type (Zólyomi 1989, Suck et al. 2014). Although this reflects fuzziness, it still represents a loss of information by limiting our view to potentially co-occurring complexes. The Multiple Potential Natural Vegetation (MPNV) concept was introduced as an alternative to avoid information loss. MPNV gives an estimate of likelihood of each PNV occurring at any given location. Thus PNV is assessed as a probability distribution of natural vegetation types that are possibly present under current environmental conditions. Besides avoiding information loss, MPNV includes the possibility of accounting for stochasticity in vegetation realizations. MPNV also accounts for estimation uncertainty by assigning a probability value to each vegetation type rather than declaring the most likely PNV.

Advantages of the MPNV approach have not been fully exploited so far. One possible reason is that the synthesis of predictions of individual vegetation types is a major challenge. So far, PNV models used a variety of methods to achieve a single outcome per location: including methods that

generate inherently categorical outcomes (Fischer 1990; Fischer et al. 2013), the determination of thresholds to consider a type as “present” (Bittner et al. 2011), or rules for choosing one type as “present” from probabilistic predictions (Hemsling & Bryn 2012; Attorre et al. 2014; Reger et al. 2014). In some studies, the predicted vegetation type with the highest probability per location was selected (Breziecki et al. 1993; Tichy 1999; Liu et al. 2009). However, this solution has one major drawback: the selection of the “most relevant” type can be biased and precludes the use of more subtle information, as often the dominant types are preferentially selected. Therefore, the use of initial model probabilities is not fully justified and has been avoided in more recent studies (Hemsling & Bryn 2012; Attorre et al. 2014; Reger et al. 2014). In summary, if more information is to be retained in an unbiased way, a transformation of probabilities to a common scale needs to be developed.

Hungary has exceptional data for MPNV estimation, because a country-wide assessment of the actual state of (semi-)natural vegetation is available (Molnár et al. 2007; Horváth et al. 2008). The Hungarian Habitat Database contains information on both natural and artificial habitats. Natural habitats are exclusively determined based on vegetation characteristics, so that they correspond to vegetation types broader than associations, but finer than alliances. We also refer to these units as vegetation types further on. Using this data source, we identified four major aims:

- 1) to formalize a model-based MPNV concept to estimate the potential natural vegetation of Hungary;
- 2) to develop a framework for synthesizing outcome of individual vegetation models into distributions of multiple vegetation types (MPNV) for each given location;
- 3) exploring additional information provided by MPNV for ecological applications.

To do so, we calibrated statistical models between vegetation observations and abiotic conditions in Hungary (Fig 1) and applied them to the entire country resulting in a series of probability maps per vegetation types. We then developed a method to synthesise these maps into a comparable set of MPNV estimations.

## Methods

### Data

Vegetation data originated from the Hungarian Actual Habitat Database (also referred to as “Landscape Ecological Vegetation Database & Map of Hungary”, MÉTA; Molnár et al. 2007; Horváth et al. 2008). This database contains field-based cover estimations for 86 main vegetation types per 35 ha hexagonal grid cells (ca. 700 m diameter) covering the entire country (Fig 1). Field mapping was organised by groups of hexagons contained by larger squares. Mapping was unsuccessful in 133 of the 2834 groups of hexagons and thus did not contain data. This is reflected later in Fig. 5 that shows masked predictions. The MÉTA database contains information on natural habitats based on vegetation characteristics, the thematic resolution corresponds to a level coarser than phytosociological plant associations, but finer than formations (Molnár et al. 2007; Molnár et al. 2008; Bölöni et al. 2011). Since our current goal was to estimate PNV, we chose the 38 vegetation types that are considered late-successional (and thus stable) based on reconstructions and studies in unmanaged landscapes (Zólyomi 1989). The influence of invasive species has been excluded from the analyses and so were vegetation types belonging to potential replacement vegetation (PRV, Chytrý 1998) if they were only sustainable with human management within the whole of Hungary (further details and justification is available in Appendix S1). Presence-absence of vegetation types was used as the dependent variable in the models. As MÉTA is a comprehensive national vegetation database based on extensive fieldwork, absence information can be considered reliable. However, there can still be two potential reasons for an absence: an unsuitable environment (“primary” absences) and human removal (“secondary” absences). Since secondary absences can

mislead model fitting, we excluded hexagons without any natural or semi-natural vegetation in the MÉTA database from the training dataset. The MÉTA database contains altogether 267 813 hexagons, out of which 87 830 have been retained after this screening.

All models for the different vegetation types were fitted with the same set of environmental variables. For climate variables, we calculated the nineteen bioclimatic indicators also advocated in WORLDCLIM (Hijmans et al. 2005), but from the locally optimised CarpatClim-Hu database (Appendix S2). The derivation of soil descriptors, indicators of water availability and topographic variation, as well as the resampling/interpolation of all predictor variables to the hexagon spatial scale are detailed in Appendix S2. The final set of 25 variables was developed from this starting set based upon inspection of individual variable effects and the correlation structure (Appendix S2).

## Analysis

Presence-absence of each considered vegetation type was related to the explanatory variables using Gradient Boosting Models (GBMs) as implemented in the “dismo” package (Hijmans, et al. 2016) in the R statistical environment (R Core Team 2016). GBM was chosen due to its flexibility when estimating response curves, and due to its explanatory variable selection approach, which is based on cross-validations rather than the criticised Akaike Information Criterion (AIC), and thus proved to be reliable in ecological modelling (Elith et al. 2006; Bühlmann & Hothorn 2007). For GBM, we followed the optimization procedure described in Elith et al. (2008) with a few exceptions (for details please consult Appendix S3). Model performance was assessed on the evaluation dataset by the well-established Area Under the Receiver Operating Curve method (AUC; Hanley & McNeil 1982).

To enable the comparison of predicted probabilities between differing PNV types at any given location, the predicted probability values were split into a five-grade ordinal scale based on the distribution of probability values given presence and absence observations in the data (see Appendix S4 for details). We provide both graphical and mathematical descriptions as well as an R script (Appendix S6). After rescaling, the MPNV predictions represent a complex set of distributions of probability ranks. We explored the result of MPNV estimation by inspecting selected cases from the whole distribution and by a synthetic evaluation.

- 1) We report probabilities and rank distributions for a landscape, serving as an example with diverse vegetation types in contrast to deterministic PNV estimations designating a single vegetation type as PNV per location.
- 2) As the full MPNV estimation is very complex for the country, we only present examples. We chose to show the presence of forests vs. grasslands vegetation in the MPNV for the whole country in detail, because increase in forest cover has been in the spotlight recently (O'Leary & Elands 2002 and references therein). Additionally, we also produced a similar map for inland halophytic and non-halophytic grasslands.
- 3) To provide a comparison of actual and potential landscape composition based on standardized probability values, we calculated Kendall tau B between the binary observed values for all vegetation types (presence/absence) and the potential distribution represented by ranks. Kendall tau B is a form of rank correlation, corrected for ties between the vectors compared. Comparisons were calculated only for hexagons where natural vegetation had been recorded.

## Results

All models generated high AUC values on the independent evaluation dataset (Appendix S1). Figure 2 illustrates the difference between single and multiple PNV mapping. When the most likely PNV type is given per hexagon (see map on Fig. 2), the resulting pattern cannot reflect the potential diversity of the landscape. We note the emergence of seemingly illogical features when identifying the most probable type only. For example, in the hexagon marked “B” the vegetation type with the

highest probability is Eu- and mesotrophic reed and *Typha* beds (B1a), surrounded by hexagons covered by loess steppe (H5a) as PNV. Extending our inspection to the full MPNV range within that and a neighbouring hexagon (“A” on the map), it becomes clear that both vegetation types (B1a & H5a) are highly likely in both hexagons, with loess steppes (H5a) being somewhat more probable in hexagon “A”. Furthermore, when probabilities are rescaled to ranks, the potential vegetation composition of both hexagons appears highly similar. A related advantage of the MPNV approach can be seen in hexagon “C” lying near the edge of the flood-plain. At this location, rescaling reveals that closed sandy steppes (H5b) and hardwood gallery forests (J6) are equally possible. A few additional types such as open sand grasslands (G1), willow shrubs (J1a) also emerge. This range of vegetation types conveys the true character of the hexagon: an intermediate position between riverine wetlands and a sandy grassland vegetation complex.

The MPNV estimations are complex and thus there is a wide range of options for presenting the resulting MPNV map. For a broad overview at the country scale, we present the distribution of grasslands and forests within the MPNV for Hungary in detail (Fig. 3) and show a similar compilation for halophytic and non-halophytic grasslands in Appendix S5. In western Hungary, forests predominate the MPNV, while in the lowlands to the East the most likely PNV type is often grassland. In between, however, a wide belt of MPNV emerges in where both forest and grassland is possible. The MPNV composition is markedly different in western Hungary and in the mountain regions (i.e. above ca. 400 m a.s.l, see also Fig 1.). In both cases, forests are the more likely vegetation within MPNV. However, while there are no grassland components in the West, the potential for grasslands (mainly rocky and extra-zonal steppe grasslands) is in fact present in the mountain regions.

Finally, MPNV mapping allows for assessing the degree of naturalness at the landscape-level by comparing differences between actual and potential landscape compositions (Fig 4). Landscapes that reveal the highest correlations with the MPNV can be found primarily in the core of the mountain regions and in those parts of the lowlands where grasslands predominate in MPNV (in the southern and north-eastern Great Plain). The foothills and central lowlands, however, often reveal low correlations.

## Discussion

### Interpretation of MPNV estimation

Our models proved to be reliable tools for assessing the distribution of both single and complex PNV types since individual PNV models achieved high absolute AUCs (cf. Swets 1988; Liu et al. 2009; Pearman et al. 2011).

In earlier PNV estimations based on predictive vegetation models usually a single vegetation type was selected per location. Approaches ranged from inherently categorical methods (Fischer 1990; Aspinall & Veith 1993; Fischer et al. 2013), through rule-based methods (Hemling & Bryn 2012) to a selection of vegetation types that received the highest probability at each location (Breziecki et al. 1993; Tichy 1999; Liu et al. 2009). In general, these methods reduce the information regarding the site potential and they imply a fully deterministic view of vegetation distribution. Retaining estimations for all analysed vegetation types appears to be a more straightforward solution, albeit not without challenges. The distribution of probability values for individual PNV types strongly depends on the degree of human land transformation in the training data and data characteristics (e.g. Hernandez et al. 2006; Elith & Graham 2009). Therefore, the modelled probability values of individual PNV types are not directly comparable, which can be resolved with the proposed rescaling procedure. For example, MPNV after rescaling revealed that loess steppes and reedbeds are equally likely to occur in our example landscape west of the Danube river. Our interpretation is that these two types co-exist in a landscape characterized by its small-scale heterogeneous pattern.

Hungary lies at the border of two biomes: the forest biome, typical of Western and Central Europe and the forest-steppe biome having connections towards Eastern Europe and Central Asia (Zólyomi 1989; Molnár et al. 2012). Some authors even argue that the steppe biome is also present as a third biome (e.g. Fekete et al. 2010). This ecotone position combined with relatively low topographic variability makes the Hungarian vegetation subject to stochastic development and thus especially difficult to model. Therefore MPNV is highly suited to represent both the transient nature and the model uncertainties. Although PNV estimations in harsh environments, such as the nordic countries might be almost deterministic (Bryn et al. 2013), even such studies found similar local uncertainties regarding the position of the upper limits of some forest types despite relatively strong topographic gradients (Hemsing & Bryn 2012).

Our approach has some limitations, however. The spatial units that form the basis for our models are relatively large, thus they themselves include a level of environmental and vegetation heterogeneity. In fact, this within-unit heterogeneity contributes to the multiplicity in PNV outcome, and on the other hand the MPNV representation of our models reflects these uncertainties better than a PNV estimation would. Unfortunately, our data does not make it possible to separate the effects of vegetation stochasticity and background heterogeneity. Finer-scale modelling could thus improve the local model projections of individual PNV types. However, there is no universal resolution available, because typical vegetation patch size may vary between landscapes (Pickett & Thompson 1978, Forman & Gordon 1981).

Another limit of our approach is that there is a subtle difference between the resolution of the environmental and vegetation data. This is, however, a general limit to almost any predictive study, since resolution often varies with different predictor variables (Rondini et al. 2006, Dormann 2011). We aimed to overcome this problem by using interpolation and resampling to adjust original resolution of environmental data to that of vegetation data.

## **Applications**

Three main areas emerge as most promising for MPNV applications: (1) landscape conservation, (2) restoration planning, and (3) evaluation of naturalness of landscapes. Cost effectiveness has been increasingly considered when planning new conservation actions (Hoekstra et al. 2005; Reynolds & Hessburg 2005; Wilson et al. 2007). Focusing new efforts on areas, where the actual vegetation is close to the PNV will likely increase conservation effectiveness (Humphries et al. 2008). Management in such a situation may be less demanding and success is more likely. MPNV allows for a range of vegetation types to be considered locally, thus offers more options to conservationists than a simple categorical map.

Besides conservation planning, ecological restoration activities can also benefit from potential vegetation models that quantify the site requirements of different vegetation types (Rodwell 2005, Shafroth et al. 2008, Loidi et al. 2010, ). Restoring the potential vegetation is more sustainable especially if continuous management is not planned. The feasibility of restoration actions (Bakker & Berendse 1999; Ehrenfeld 2000) and its acceptance by the local communities (Pfadenhauer 2001; Buckley & Crone 2008) can greatly differ among vegetation types. Therefore, choice options can help enhance restoration activities, compared to restoring single PNVs that may not always be preferred locally.

Finally, PNV estimations can support the analysis of the degree of naturalness at the landscape scale. MPNV can serve as a reference for comparison (Ricotta et al. 2002) when the goal is to quantify the departure of the actual landscape from its potential. Our comparison of actual and potential landscape composition identified that the core of the mountain regions was the most similar to its potential. This is likely due to the presence of protected zonal forests, which are utilized only by low-intensity forestry exploitation, and thus are close to natural vegetation types. Similar reasons may explain the high degree of naturalness in some lowland areas. These are regions with natural grassland types, which can only be used economically as pastures. Therefore, landscape composition remains closer to natural (Molnár et al. 2012). The foothills of the mountain

regions were found to be the farthest from natural in this comparison, which can be explained by the long history of increased human pressure, especially agricultural use, while not completely destroying the vegetation (Nyizsalovszki & Forian 2007; Lóczy & Sütő 2011). The most intensively used parts of the lowlands do not appear in this comparison due to the complete lack of natural vegetation.

MPNV in its full complexity is probably difficult to use directly by practitioners, however applications is planned to be facilitated in two ways:

1) downloadable coarse-scale map overview of the potential distribution of individual PNV types available at [www.novenyzetiterkep.hu/node/1411](http://www.novenyzetiterkep.hu/node/1411)

2) As an additional data layer to the hexagon-wise MÉTA database, which can be accessed for different PNV types the same way as for the actual vegetation layers.

## Conclusions

The MPNV estimation approach allowed us to view the potential composition of locations in far more detail than before and provided a representation of model uncertainties and local variability. Our novel approach was demonstrated with the potential distribution of forests vs. grasslands as an example. Rescaling of predicted probabilities ensured that vegetation types with different prevalence and specifically those with high rates of human conversion to artificial vegetation received appropriate weights in the MPNV distribution analyses. PNV estimations were developed into a useful basis for vegetation and landscape conservation and restoration planning.

## Acknowledgements

The authors thank to the Hungarian Actual Habitat Database (MÉTA) Curatorium for allowing the use of the database, as well as to the field mappers for their contribution to the database. The study was supported by the Hungarian Scientific Research Fund (OTKA) grant no. PD-83522 (Imelda Somodi), by the Bolyai János research fellowship of the Hungarian Academy of Sciences (Bálint Czúcz), by the grant TÁMOP 4.2.1/B-09/1/KMR-2010-0005 (Ákos Bede-Fazekas) and by the GINOP-2.3.2-15-2016-00019 grant. Niklaus E. Zimmermann acknowledges additional support from the Swiss SNF (grant: 40FA40\_158395). Financial support was also received from Iceland, Liechtenstein and Norway through the EEA Grants and the REC. The authors are responsible for the content. The authors are grateful to Douglas Evans and Gábor Endresz for corrections regarding grammar and language correctness and further suggestions regarding the manuscript.

## References

- Aspinall, R. & Veitch, N. 1993. Habitat mapping from satellite imagery and wildlife survey data using a Bayesian modeling procedure in a GIS. *Photogrammetric Engineering and Remote Sensing* 59: 537–543.
- Attorre, F., Francesconi, F., De Sanctis, M., Alfò, M., Martella, F., Valenti, R. & Vitale, M. 2014. Classifying and mapping potential distribution of forest types using a finite mixture model. *Folia Geobotanica* 49: 313–335.
- Baker, W.L., Walford, G.M. 1995. Multiple Stable States and Models of Riparian Vegetation Succession on the Animas River, Colorado. *Annals of the Association of American Geographers* 85: 320–338.
- Bakker, J.P. & Berendse, F. 1999. Constraints in the restoration of ecological diversity in grassland and heathland communities. *Trends in Ecology & Evolution* 14: 63–68.
- Bittner, T., Jaeschke, A., Reineking, B. & Beierkuhnlein, C. 2011. Comparing modelling approaches at two levels of biological organisation—Climate change impacts on selected Natura 2000 habitats. *Journal of Vegetation Science* 22: 699–710.

- Bölöni, J., Molnár, Zs. & Kun, A. (eds.) 2011. Magyarország élőhelyei. A hazai vegetációtípusok leírása és határozója. ÁNÉR 2011 [Habitats in Hungary. Description and identification guide of the Hungarian vegetation]. MTA ÖBKI, Vácrátót, Hungary.
- Breziecki, B., Kienast, F. & Wildi, O. 1993. A simulated map of the potential natural forest vegetation of Switzerland. *Journal of Vegetation Science* 4: 499–508.
- Bryn, A., Dourojeanni, P., Hemsing, L. Ø. & O'Donnell, S. 2013. A high-resolution GIS null model of potential forest expansion following land use changes in Norway. *Scandinavian Journal of Forest Research* 28: 81–98.
- Buckley, M.C. & Crone, E.E. 2008. Negative off-site impacts of ecological restoration: understanding and addressing the conflict. *Conservation Biology* 22: 1118–1124.
- Bühlmann, P. & Hothorn, T. 2007. Boosting algorithms: regularization, prediction and model fitting. *Statistical Science* 22: 477–505.
- Carrión, J.S. & Fernandez, S. 2009. The survival of the 'natural potential vegetation' concept (or the power of tradition). *Journal of Biogeography* 36: 2202–2203.
- Chiarucci, A., Araújó, M.B., Decocq, G., Beierkuhnlein, C. & Fernández-Palacios, J.M. 2010. The concept of potential natural vegetation: an epitaph? *Journal of Vegetation Science* 21: 1172–1178.
- Chytrý, M., 1998. Potential replacement vegetation: an approach to vegetation mapping of cultural landscapes. *Applied Vegetation Science*, 1(2), pp.177-188.
- Dormann, C.F., 2011. Modelling species' distributions. In: Jopp, F., Reuter, H., Brackling, B. (eds): *Modelling Complex Ecological Dynamics* (pp. 179-196). Springer Berlin.
- Dotterweich, M. 2008. The history of soil erosion and fluvial deposits in small catchments of central Europe: deciphering the long-term interaction between humans and the environment—a review. *Geomorphology* 101: 192–208.
- Ehrenfeld, J.G. 2000. Defining the limits of restoration: the need for realistic goals. *Restoration Ecology* 8(1): 2–9.
- Elith, J. & Graham, C.H. 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography* 32: 66–77.
- Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K.S., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S. & Zimmermann, N.E. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29: 129–151.
- Elith, J., Leathwick, J.R. & Hastie, T. 2008. A working guide to boosted regression trees. *Journal of Animal Ecology* 77: 802–813.
- Fekete G., Somodi I. & Molnár Zs. 2010. Is chorological symmetry observable within the forest steppe biome in Hungary? – A demonstrative analysis of floristic data. *Community Ecology* 11: 140–147.
- Fischer, H.S. 1990. Simulating the distribution of plant communities in an alpine landscape. *Coenoses* 5: 37–43.
- Fischer, H.S., Winter, S., Lohberger, E., Jehl, H. & Fischer, A. 2013. Improving transboundary maps of potential natural vegetation using statistical modeling based on environmental predictors. *Folia Geobotanica* 48: 115–135.

- Forman, R. T., & Godron, M. 1981. Patches and structural components for a landscape ecology. *BioScience*, 31, 733-740.
- Guisan, A. & Zimmermann, N.E. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135: 147-186.
- Hall, G.M.J. & McGlone, M.S. 2006. Potential forest cover of New Zealand as determined by an ecosystem process model. *New Zealand Journal of Botany* 44: 211–232
- Hanley, J.A. & McNeil, B.J. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143: 29–36.
- Hemning, L.Ø. & Bryn, A. 2012. Three methods for modelling potential natural vegetation (PNV) compared: A methodological case study from south-central Norway. *Norsk Geografisk Tidsskrift – Norwegian Journal of Geography* 66: 11–29.
- Hernandez, P.A., Graham, C.H., Master, L.L. & Albert, D.L. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* 29: 773–785.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25: 1965–1978.
- Hijmans, R.J., Phillips, S., Leathwick, J. & Elith, J. 2016. dismo: Species Distribution Modeling. R package version 1.1-1. <https://CRAN.R-project.org/package=dismo>.
- Hoekstra, J.M., Boucher, T.M., Ricketts, T.H. & Roberts, C. 2005. Confronting a biome crisis: Global disparities of habitat loss and protection. *Ecology Letters* 8: 23–29.
- Horváth, F., Molnár, Zs., Bölöni, J., Pataki, Zs., Polgár, L. Révész, A., Oláh, K., Krasser, D. & Illyés, E. 2008. Fact sheet of the MÉTA Database 1.2. *Acta Botanica Hungarica* 50: 11–34.
- Humphries, H.C., Bourgeron, P.S. & Reynolds, K.M. 2008. Suitability for conservation as a criterion in regional conservation network selection. *Biodiversity and Conservation* 17: 467–492.
- Jackson, S.T. & Hobbs, R.J. 2009. Ecological restoration in the light of ecological history. *Science* 325: 567–569.
- Kaplan, J.O., Krummhardt, K.M. & Zimmermann, N.E. 2009. The prehistoric and preindustrial deforestation of Europe. *Quaternary Science Reviews* 28: 3016–3034.
- Kowarik, I., 1987. Kritische Anmerkungen zum theoretischen Konzept der potentiellen natürlichen Vegetation mit Anregungen zu einer zeitgemäßen Modifikation. *Tuexenia : Mitteilungen der Floristisch-Soziologischen Arbeitsgemeinschaft* 7: 53-67.
- Liu, H., Wang, L., Yang, J., Nakagoshi, N., Liang, C., Wang, W. & Lv, Y. 2009. Predictive modeling of the potential natural vegetation pattern in northeast China. *Ecological Research* 24: 1313–1321.
- Loidi, J., Del Arco, M., Pérez de Paz, P. L., Asensi, A., Díez Garretas, B., Costa, M., Díaz González, T., Fernández-González, F., Izco, J., Penas, Á., Rivas-Martínez, S. & Sánchez-Mata, D. (2010). Understanding properly the 'potential natural vegetation' concept. *Journal of Biogeography*, 37: 2209-2211.
- Lóczy, D. & Sütő, L. 2011. Human activity and geomorphology. In: Gregory, K. & Goudie, A.S. (eds.): *SAGE Handbook of Geomorphology*, pp. 267–285. Sage Publications, London, UK.
- Molnár, Zs., Bartha, S., Seregélyes, T., Illyés, E., Tímár, G., Horváth, F., Révész, A., Kun, A., Botta-Dukát, Z., Bölöni, J., Biró, M., Bodoncz, L., Deák, J.Á., Fogarasi, P., Horváth, A., Isépy, I., Karas, L., Kecskés, F., Molnár, Cs., Ortmann-né Ajkai, A. & Rév, Sz. 2007. A grid-based,

satellite-image supported, multi-attributed vegetation mapping method (MÉTA). *Folia Geobotanica* 42: 225–247.

- Molnár, Zs., Biró, M. & Bölöni, J. (eds.) 2008. APPENDIX (English names of the Á-NÉR habitat types; Natura 2000 habitats and their Á-NÉR habitats equivalents). *Acta Botanica Hungarica* 50: 249–255.
- Molnár, Zs., Biró, M., Bartha, S. & Fekete, G. 2012. Past Trends, Present State and Future Prospects of Hungarian Forest-Steppes. In: Werger, MJA & van Staalduinen, MA (eds.): *Eurasian Steppes. Ecological Problems and Livelihoods in a Changing World*, pp. 209–252. Springer, Dordrecht, Heidelberg, New York, London.
- Nyizsalovszki, R. & Forian, T. 2007. Human impact on the landscape in the Tokaj Foothill Region, Hungary. *Geografia Fisica e Dinamica Quaternaria* 30: 219–224.
- O'Leary, T.N. & Elands, B.H.M. 2002. Anyone for more forests? Current perspectives and future expectations on afforestation and forest functions across Europe. In: Wiersum, K.F. & Elands, B.H.M (eds.) *The changing role of forestry in Europe: perspectives for rural development*, pp. 51–73. Wageningen University, Wageningen, Netherlands.
- Pearman, P.B., Guisan, A. & Zimmermann, N.E. 2011. Impacts of climate change on Swiss biodiversity: an indicator taxa approach. *Biological Conservation* 144: 866–875.
- Petraitis, P., 2013. *Multiple stable states in natural ecosystems*. OUP Oxford.
- Pickett, S. T., & Thompson, J. N. (1978). Patch dynamics and the design of nature reserves. *Biological Conservation*, 13(1), 27–37.
- Pfadenhauer, J. 2001. Some remarks on the socio-cultural background of restoration ecology. *Restoration Ecology* 9: 220–229.
- Prach, K., Tichý, L., Lencová, K., Adámek, M., Koutecký, T., Sádlo, J. Bartošová, A., Novák, J., Kovář, P. & Šmilauer, P. 2016. Does succession run towards potential natural vegetation? An analysis across seres. *Journal of Vegetation Science* 27: 515–523.
- R Core Team 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- Reger, B., Häring, T. & Ewald, J. 2014. The TRM-model of potential natural vegetation in mountain forests. *Folia Geobotanica* 49: 337–359.
- Reynolds, K.M. & Hessburg, P.F. 2005. Decision support for integrated landscape evaluation and restoration planning. *Forest Ecology and Management* 207: 263–278.
- Ricotta, C., Carranza, M.L., Avena, G. & Blasi, C. 2002. Are potential natural vegetation maps a meaningful alternative to neutral landscape models? *Applied Vegetation Science* 5: 271–275.
- Roberts D.W. 2015. Potential natural vegetation and environment: a critique of Kusbach, Shaw & Long. *Applied Vegetation Science* 18: 733–738.
- Rodwell, J. 2005. Restoring Cultural Landscapes Using the Vegetation Map of Europe. In: Bohn, U., Hettwer, C., & Gollub, G. Anwendung und Auswertung der Karte der natürlichen Vegetation Europas. (Application and Analysis of the Map of the Natural Vegetation of Europe.) BfN Skripten 156. Bundesamt für Naturschutz, Bonn, Germany.
- Rondinini, C., Wilson, K.A., Boitani, L., Grantham, H. & Possingham, H.P., 2006. Tradeoffs of different types of species occurrence data for use in systematic conservation planning. *Ecology letters*, 9(10), pp.1136–1145.
- Shafroth, P.B., Beauchamp, V.B., Briggs, M.K., Lair, K., Scott, M.L. & Sher, A.A. 2008. Planning Riparian Restoration in the Context of Tamarix Control in Western North America. *Restoration Ecology* 16: 97–112.

- Sillasoo, Ü., Poska, A., Seppä, H., Blaauw, M. & Chambers, F.M. 2009. Linking past cultural developments to palaeoenvironmental changes in Estonia. *Vegetation History and Archaeobotany* 18: 315–327.
- Somodi, I., Molnár, Zs. & Ewald, J. 2012. Towards a more transparent use of the potential natural vegetation concept - an answer to Chiarucci et al. *Journal of Vegetation Science* 23: 590–595.
- Suck, R.; Bushart, M., Hofmann, G. & Schröder, L. (2014): Karte der Potentiellen Natürlichen Vegetation Deutschlands Band I Grundeinheiten. Bonn-Bad Godesberg: Bundesamt für Naturschutz (BfN-Skripten), 348.
- Swets, J.A. 1988. Measuring the accuracy of diagnostic systems. *Science* 240: 1285–1293.
- Tichy, L. 1999. Predictive modeling of the potential natural vegetation in the Podyjí National Park, Czech Republic. *Folia Geobotanica* 34: 243–252.
- Tüxen, R. 1956. Die heutige potentielle natürliche Vegetation als gegenstand der Vegetationskartierung. *Angewandte Pflanzensociologie (Stolzenau)* 13: 5–42.
- Wilson, K.A., Underwood, E.C., Morrison, S.A., Klausmeyer, K.R., Murdoch, W.W., Reyers, B., Wardell-Johnson, G., Marquet, P.A., Rundel, P.W., McBride, M.F., Pressey, R.L., Bode, M., Hoekstra, J.M., Andelman, S., Looker, M., Rondinini, C., Kareiva, P., Shaw, M.R. & Possingham, H.P. 2007. Conserving Biodiversity Efficiently: What to Do, Where, and When. *PLoS Biology* 5: e223.
- Zimmermann, N.E. & Kienast, F. 1999. Predictive mapping of alpine grasslands in Switzerland: species versus community approach. *Journal of Vegetation Science* 10: 469–482.
- Zólyomi B 1989. Magyarország természetes növénytakarója. [Map of the natural vegetation of Hungary.] In: Pécsi M (ed) *Nemzeti Atlasz. [National Atlas.]*, p. 89. Kartográfia Vállalat, Budapest, Hungary.

## Appendices

- Appendix S1. Abbreviation, name and characteristics of studied vegetation types, and AUC of the models
- Appendix S2. Explanatory variables used in the analysis
- Appendix S3. Parameters in the GBM models
- Appendix S4. The process of rescaling of probability distributions
- Appendix S5. A representation of MPNV estimations regarding halophytic vs. non-halophytic grasslands
- Appendix S6. R script of the proposed probability ranking procedure

## Figures

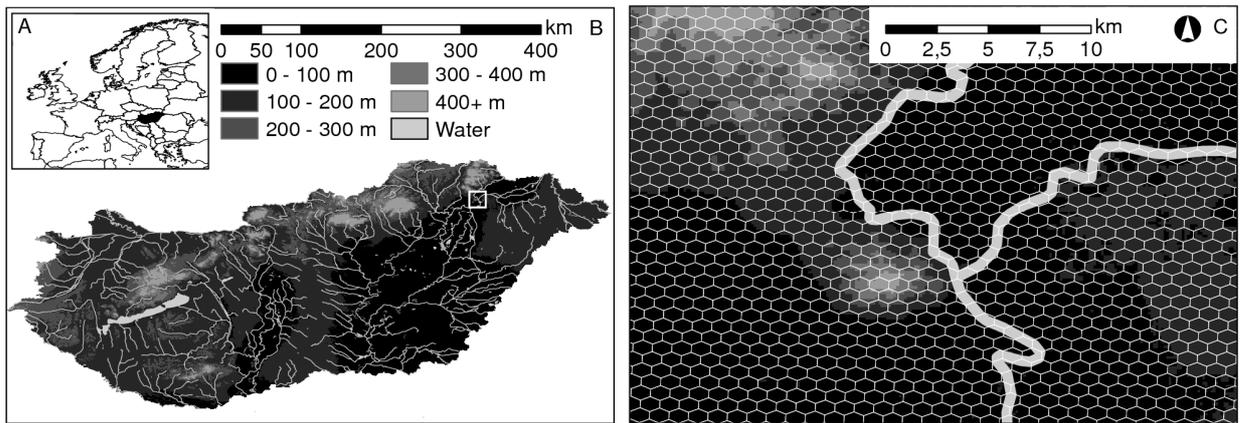


Figure 1. Location of Hungary and its coverage by the MÉTA hexagon lattice. A: location of Hungary within Europe; B: classified elevation map of Hungary; C: representation of the hexagon structure in an inset (see panel B) along the Bodrog watershed. White lines depict rivers and streams. Also lakes are shown in white.

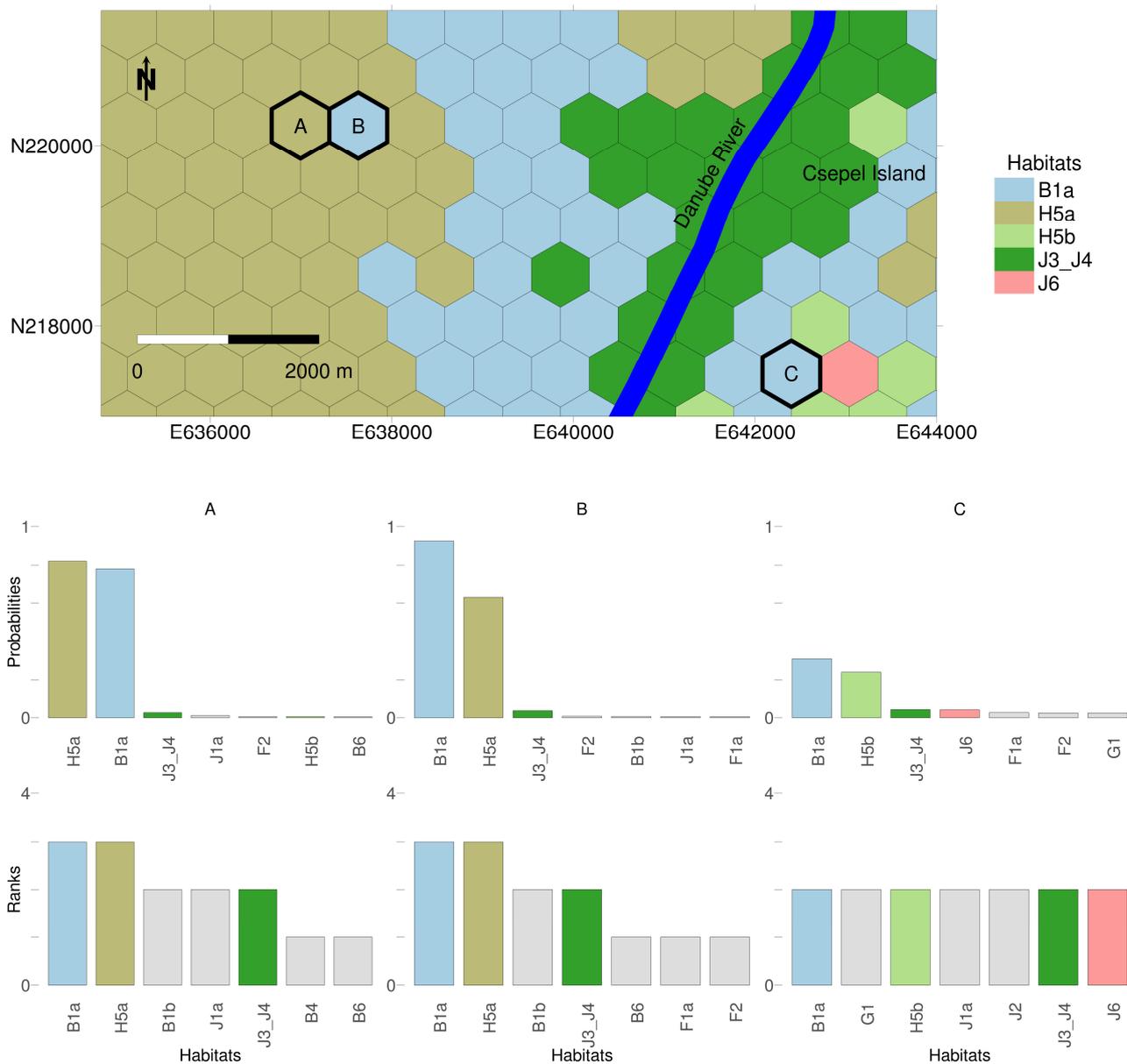


Figure 2: Distribution of potential probabilities and ranks of vegetation types in three selected hexagons in an area Southwest of Budapest. Only the seven highest ranked vegetation types are shown. The map is coloured according to the habitat with the highest probability per hexagon for contrast (equivalent to the traditional PNV estimation). Vegetation types not shown on the map inset are left grey in the barplots. For abbreviations of vegetation types see Appendix S1.

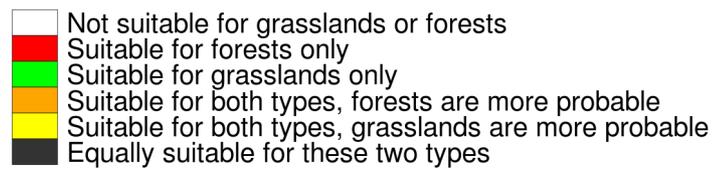
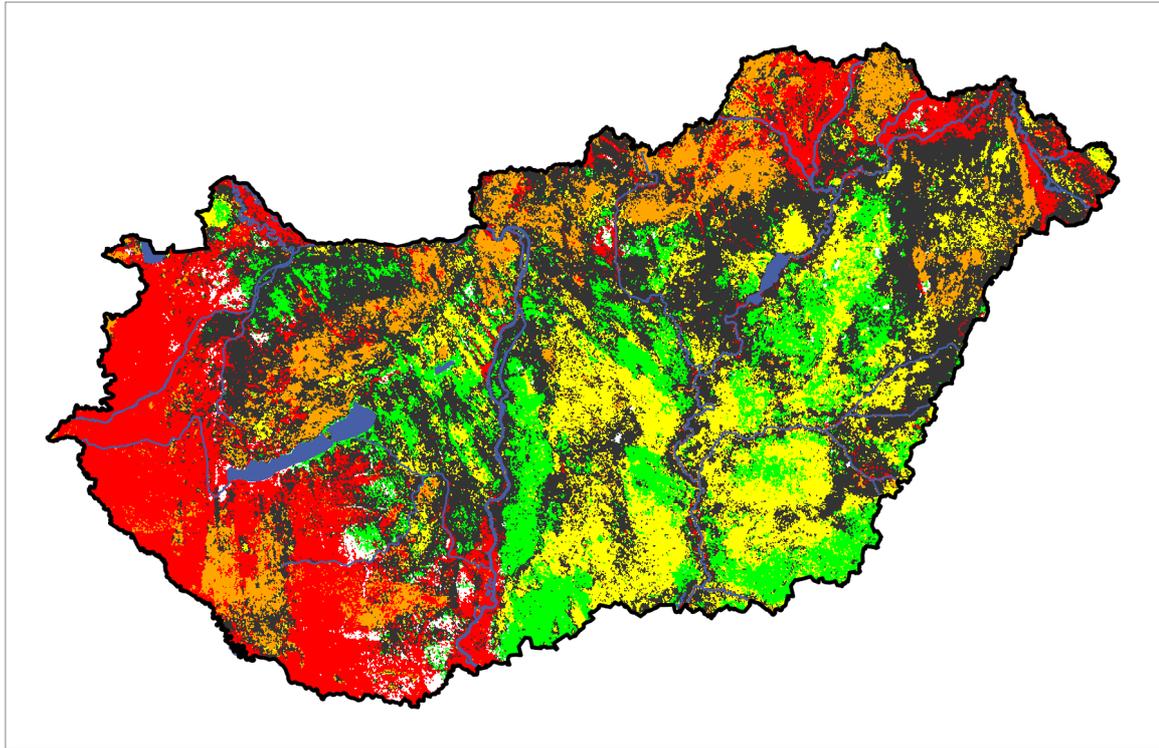


Figure 3. Representation of the modelled MPNV aggregated to forests and grasslands (see Appendix S1 for aggregation scheme). Forests include open and closed forests. Wetlands and shrubs are not included in this representation. Blue colour depicts water bodies.

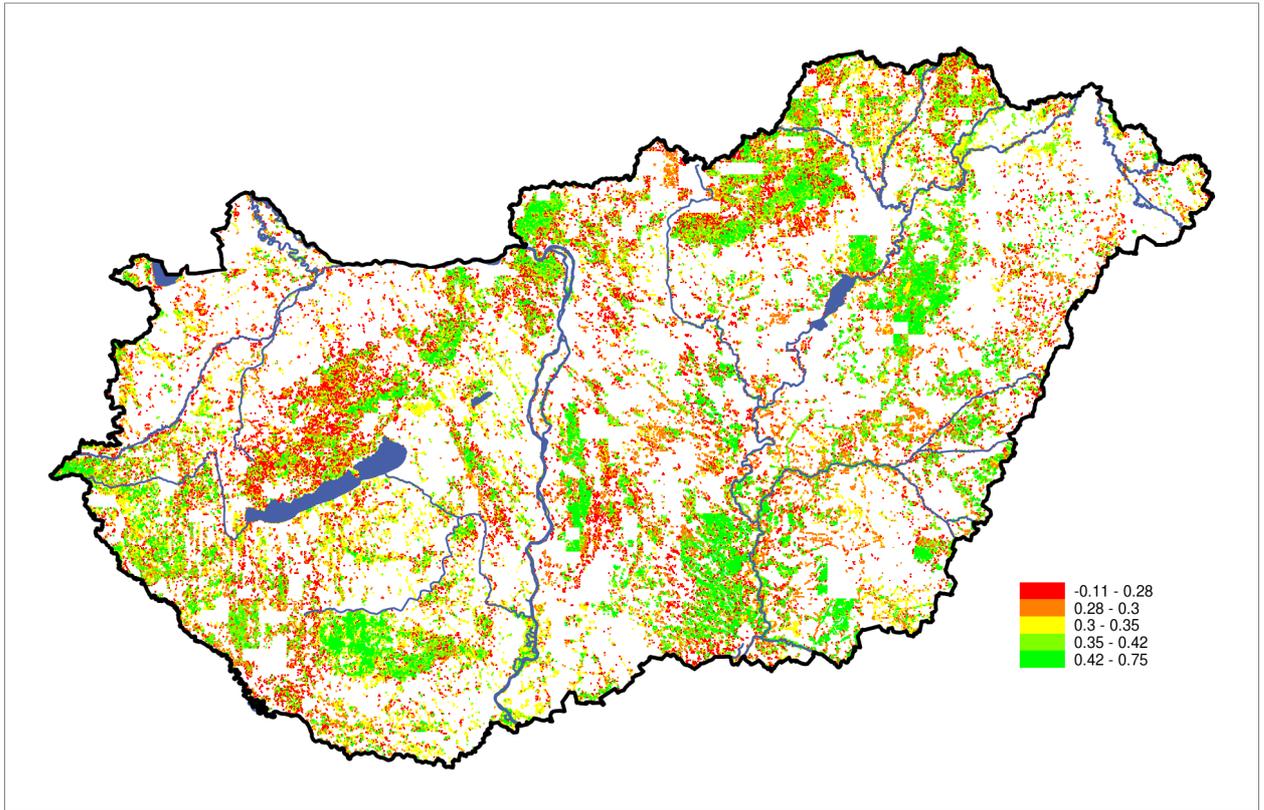


Figure 4: Rank correlation (Kendall tau B coefficient) between observed and potential vegetation type distribution for hexagons, in which at least one natural vegetation type was observed (i.e. natural vegetation has not been completely removed by humans). White represents areas where no natural vegetation was observed.