Implementation and application of multiple potential natural vegetation models – a case study of Hungary

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

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 (re-scaling) 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 re-scaled probability maps of individual vegetation types for Hungary. Raw probabilities best suit within-type analyses, while re-scaled 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 viewing the potential vegetation composition of locations in far more detail than the PNV approach. Re-scaling 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 prioritization, landscape evaluation) suggests that the PNV concept with the extension towards vegetation distributions is useful both for research and application.

Keywords
Conservation prioritization; Landscape evaluation; Multilayer model; PNV; Predictive vegetation model; Probability distribution of vegetation types; Probability re-scaling; Restoration; Vegetation stochasticity

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.

Nomenclature
Bolöni et al. (2011) for vegetation types

Received 16 June 2016
Accepted 16 May 2017
Co-ordinating Editor: Ingolf Kühn

© 2017 International Association for Vegetation Science

1MTA Centre for Ecological Research, Institute for Ecology and Botany, 2-4 Alkotmány str., 2163 Vácstráit, Hungary;
2MTA Centre for Ecological Research, GINOP Sustainable Ecosystems Group, Klebesberg Kunó utca, 8237 Tihany, Hungary;
3European Topic Centre on Biological Diversity, Muséum national d'Histoire naturelle, 57 rue Cuvier, FR-75231 Paris Cedex 05, France;
4Department of Garden and Open Space Design, Faculty of Landscape Architecture and Urbanism, Szent István University, 29-43 Villányi út, Budapest, Hungary;
5National Adaptation Centre, Geological and Geophysical Institute of Hungary, 14 Stefánia út, 1143 Budapest, Hungary;
6Department of Environmental Informatics, Institute for Soil Sciences and Agricultural Chemistry, MTA Centre for Agricultural Research, 15 Herman Ottó út, 1022 Budapest, Hungary;
7Swiss Federal Research Institute WSL, 111 Zürcherstrasse, CH-8903 Birmensdorf, Switzerland;
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 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 pre-human vegetation is often treated as equivalent of PNV estimation (e.g. Hall & McGlone 2006; Carrión & Fernandez 2009). Such estimation may be misleading, however, because environmental conditions have clearly changed since pre-human 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 criticized (see e.g. 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 using formal methods (Hemsing & Bryn 2012; Fischer et al. 2013; Attorre et al. 2014; Reger et al. 2014) and maps showing 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 cannot 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 creates 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 (Hemsing & 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 (Hemsing & 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 type. We then developed a method to synthesize these maps into a comparable set of MPNV estimations.

Methods

Data

Vegetation data originate 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 organized with 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. 4, which 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, 2008; Bólöni 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 (Zőlyomi 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 seminatural vegetation in the MÉTA database from the training data set. The MÉTA database contains altogether 267 813 hexagons, out of which 87 830 were 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 19 bioclimatic indicators also advocated in WORLDCLIM (Hijmans et al. 2005), but from the locally optimized 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 on 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 in the R statistical environment (R Foundation for Statistical Computing, Vienna, AT). 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 criticized 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 data set 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 re-scaling, 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 with diverse vegetation types, serving as an example 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 nonhalophytic 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 data set (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 in 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 and H5a) are highly likely in both hexagons, with loess steppes (H5a) being somewhat more probable in hexagon “A”. Furthermore, when probabilities are re-scaled 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, re-scaling 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 dominate 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 which 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 northeastern 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 & Veitch 1993; Fischer et al. 2013), through rule-based methods (Hemsing & 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.

Fig. 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 bar plots. For abbreviations of vegetation types see Appendix S1.
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 re-scaling procedure. For example, MPNV after re-scaling revealed that loess steppes and reed beds 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 (Zőlyomi 1989; Molnár et al. 2009).
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. This within-unit heterogeneity also contributes to the multiplicity in PNV outcomes, and the MPNV representation of our models reflects these uncertainties better than a PNV estimation would. Unfortunately, our data do 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 & Godron 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 (Rondinini 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, and 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 & Crane 2008) can greatly differ among vegetation types. Therefore, a choice of 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 for 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 furthest 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 (Nyírsalovszki & Forian 2007; Lóczy & Sútó 2011). The most intensively used parts of the lowlands do not appear in this comparison due to the complete lack of natural vegetation.

The MPNV in its full complexity is probably difficult to use directly by practitioners, however applications are planned for facilitation in two ways:

1. Downloadable coarse-scale map overview of the potential distribution of individual PNV types available at 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 in 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 previously and provided a representation of model uncertainties and local variability. Our novel approach was
demonstrated using the potential distribution of forests vs grasslands as an example. Re-scaling 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), the Bolyai János research fellowship of the Hungarian Academy of Sciences (Bálint Czúcs), grant TAMOP 4.2.1-B/09-1/KMR-2010-0005 (Ákos Bede-Fazekas) and 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 Endrezs for corrections regarding grammar and language and further suggestions regarding the manuscript.

References


Multiple potential natural vegetation of Hungary

I. Somodi et al.


**Supporting Information**

Additional Supporting Information may be found in the online version of this article:

**Appendix S1.** Details of the vegetation types modelled.

**Appendix S2.** Explanatory variables used in the analysis.

**Appendix S3.** Parameters in the GBM models.

**Appendix S4.** The process of re-scaling 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.