Changes in predictor influence with time and with vegetation type identity in a post-abandonment situation

Imelda Somodi\textsuperscript{a,b,*}, Klára Virágh\textsuperscript{a}, Balázs Székely\textsuperscript{c,d}, Niklaus E. Zimmermann\textsuperscript{e}

\textsuperscript{a}Institute of Ecology and Botany of the Hungarian Academy of Sciences, 2-4. Alkotmány út, 2163 Vácrátót, Hungary
\textsuperscript{b}Department of Plant Taxonomy and Ecology, Institute of Biology, Eötvös Loránd University, I/C Pázmány Péter sétány, 1117 Budapest, Hungary
\textsuperscript{c}Institute of Photogrammetry and Remote Sensing of the Vienna University of Technology, Gusshausstrasse 27-29, A-1040 Wien, Austria
\textsuperscript{d}Space Research Group, Department of Geophysics and Space Science, Institute of Geography and Earth Sciences, I/A Pázmány Péter sétány, 1117 Budapest, Hungary
\textsuperscript{e}Swiss Federal Research Institute WSL, Zürcherstrasse 111, CH-8903 Birmensdorf, Switzerland

Received 9 January 2009; accepted 6 November 2009

Abstract

Variation in the explanatory potential of separate variable groups describing past vegetation patterns and/or abiotic environment were investigated using data from two points in time for three vegetation types, which are variants of the steppe grassland component of the forest steppe biome. First, we compared the predictive power of environmental variables assessed by the performance of generalised linear models (GLMs) explaining the distribution of three vegetation types in 1988 and 2002. Second, we fitted models based on conceptual hypotheses to explain vegetation distribution in 2002. Specifically, we wanted to examine, whether (i) current abiotic topo-environment, (ii) past neighbourhood configuration, or (iii) historical vegetation patterns or a combination of these determine best the current distribution of vegetation types. We developed basic predictor sets for each hypothesis, and using GLMs we tested to what degree these predictor sets were capable of explaining the currently observed patterns of individual vegetation types. We compared model accuracy by AUC and TSS values. Predictive performance of models changed both with time and with vegetation type. The analyses of changes over time showed that two of the three vegetation types had come closer to an equilibrium with abiotic conditions, while the third had moved farther away from equilibrium. Knowledge of past conditions was sufficient to predict the distribution of one of the three investigated vegetation types alone, thus no topo-environmental predictors were needed to successfully predict this type. The other two vegetation types were best explained by current topo-environmental predictors. We conclude that historical conditions clearly improve predictive models, though there may be variation in their
contribution to models of different vegetation types and this may depend on how far vegetation types are from an equilibrium state.

© 2010 Published by Elsevier GmbH on behalf of Gesellschaft für Ökologie.

**Zusammenfassung**


© 2010 Published by Elsevier GmbH on behalf of Gesellschaft für Ökologie.

**Keywords:** GLM; Equilibrium; Past neighbourhoods; Site history; Vegetation dynamics; Vegetation maps

---

**Introduction**

Predictive vegetation modelling typically assists vegetation ecology in two major ways: by providing a basis for inter- and extrapolation of known vegetation entities to remote or unknown areas (Cairns, 2001; Fischer, 1990; Miller & Franklin, 2002), and by exploring what factors best explain the spatial occurrence patterns (Guisan & Zimmerman, 2000; Miller, Franklin, & Aspinall, 2007; Pueyo & Alados, 2007). Whichever target is chosen, the most commonly used explanatory variables are those describing the actual abiotic environment, which are also assumed to exhibit time-invariant influence on the predicted entities (Fischer, 1990; Franklin, 1995; Vogiatzakis & Griffiths, 2006; Zimmermann & Kienast, 1999). It has long been recognized, however, that the history of land use and vegetation development can be of considerable importance as well (Cairns, 2001; de Blois, Domon, & Bouchard, 2001; Didier, Chabriere, Dutoit, Roche, & Langlois, 2005; Muster, Elsenbeer, & Conedera, 2007) and abiotic factors have predominant effect in equilibrium situations rather (Guisan & Zimmerman, 2000).

The history of vegetation development can be assessed in various ways. The most straightforward and most commonly used approach is to take local site history into account, namely by including preceding land use, disturbance regime and/or vegetation type at a target location as potential explanatory variables (Baeza, Valdecantos, Alloza, & Vallejo, 2007; Benjamin, Domon, & Bouchard, 2005; de Blois et al., 2001). However, the spatial configuration and neighbourhood relationships of the past landscape surrounding a target location can also have a profound influence on vegetation development (Augustin, Cummins, & French, 2001). Considerable efforts have been made to incorporate current spatial dependencies into predictive modelling (Miller & Franklin, 2002; Dormann, McPherson, Araujo, Bivand, Bolliger et al., 2007; Miller et al., 2007). The spatial dependencies of past neighbourhoods – on the other hand – have primarily been included in dynamic spatial simulation studies only (Turner, 1987; Wu, Sklar, & Rutche, 1997; but see Augustin et al.,
Whichever variables are included into the potentially acting set of variables, there is a decision to be made whether the response variables (vegetation entities or species) are modelled using the same set of variables or are allowed to react to different explanatory variables. In fact, the distribution of different species or vegetation types may be best explained by different sets of predictors each (de Blois et al., 2001; Carmel et al., 2001; Pueyo & Alados, 2007; Zimmermann, Edwards, Moisen, Frescino, & Blackard, 2007), although predictive vegetation mapping often treats all vegetation types simultaneously without individual optimization per vegetation type (Cairns, 2001; Fischer, 1990; Pfeffer, Pebesma, & Burrough, 2003; Weaver & Perera, 2004).

To provide arguments in the above question, the main aim of our case study was to explore how the influence of different predictor sets varies with time and vegetation type with special emphasis on the effect of past vegetation conditions.

Our first question was how the dependence of various vegetation entities on abiotic conditions changes with ongoing regeneration. Our second question was how predictors expressing past conditions perform in models of different vegetation types describing their current patterns. A specific goal at this stage was also to evaluate the predictive potential of past neighbourhoods, in addition to the past vegetation type, which is the typical variable widely used to describe past vegetation conditions.

In order to provide a starting point for our modelling exercise, we formulated several hypotheses based on sets of differently acting predictors:

- Observed spatial vegetation patterns can be explained by
  - topo-environment alone (E),
  - past neighbourhood configuration alone (N),
  - past neighbourhood and topo-environment together (NE),
  - local site history (in our case the identity of the previous vegetation type) combined with topo-environment (PE) and
  - topo-environment, local site history and past neighbourhoods (PNE).

Local site history was not tested alone, because its influence cannot be separated from that of the topo-environment. This is due to a possible dependence of the past distribution of vegetation types on abiotic environment, which is not likely to have changed profoundly during two decades.

Methods

Study site and vegetation data

The study site is an abandoned pasture located at the foot of the Bükk Mountains, Hungary (47°53′40″ N; 20°34′40″ E). Sheep grazing prevailed until the late 1970s. Today, the vegetation consists of a mosaic of patches of different vegetation types including species-rich loess steppe types, natural dry grassland patches (dominated e.g. by Chrysopogon gryllus) and compositionally degraded habitat types. Vegetation types were named after the dominant species, their characteristic species compositions and abundance patterns as well as further site conditions are described in Somodi, Virágh, & Aszalós (2004). For this study, three abundant vegetation types with potentially contrasting expansion mechanisms were chosen based on an earlier study (Somodi et al., 2004): the Danthonia, the Chrysopogon and the Calamagrostis type. The first is the richest in species, and is a representative of the steppe vegetation with numerous rare and steppe-specialist species, and several potential co-dominants (Danthonia alpina, Festuca rupicola, Avenula praestans). Chrysopogon type is dominated by C. gryllus and is relatively species-rich, too, though less than the previous. The Calamagrostis type is dominated by Calamagrostis epigejos and includes only a low number of subordinate species, which are usually generalists. Species names follow Flora Europaea (Tutin, Heywood, Burges, Moore, Valentine et al., 1964–1993).

Dependent variables

We based our analysis on two vegetation maps, drawn at a scale of 1:1400, depicting the same ca. 300 × 300 m² area in 1988 and in 2002 (Fig. 1), thus spanning 14 years. The minimum patch size mapped was 5 m in diameter. The 1988 map was used both as a source of historical predictors and as a dependent variable when investigating the abiotic dependence in the past. Distribution patterns of vegetation types of the vegetation map of 2002 were used as the dependent variable for testing our conceptual hypothesis. We did not attempt, however, to predict the distribution of all the 9 vegetation types present in the maps, but only the 3 that are detailed above. A description of all transition patterns between vegetation types can be found in Somodi et al. (2004).

Explanatory variables

Predictor variables were grouped to provide building stones for the conceptual hypotheses (Table 1). The first variable group (P) contained one factorial variable only,
the first-order vegetation history (the vegetation type found at the same location in the previous vegetation map in 1988). All 9 vegetation types present at our sites were included. The second group \((N)\) described the past neighbourhood and spatial configuration in two different ways: (a) distance to the nearest patch of the target vegetation type and (b) proportion of cells occupied by the target vegetation type in circular neighbourhoods. As it is difficult to determine a single optimal neighbourhood radius to study any organism or community and optimal neighbourhoods may vary between entities (Addicott, Aho, Antolin, Padilla, Richardson et al., 1987; Weaver & Perera, 2004) we calculated this measure for five different neighbourhood sizes with radii of 20–100 m, with intervals of 20 m. Since the distance variable was heavily skewed, we applied a square-root transformation prior to using it in our statistical analyses.

The third group of variables \((E)\) described abiotic environment and topography-related variation, which also characterise soil conditions and microclimate indirectly (sensu Franklin 1995). They were derived from a 2.5-m resolution digital elevation model (DEM). The following variables were included: aspect value, slope, relative topographic exposure calculated across a range of scales, an integrated index of topographic position (unsmoothed and smoothed in a 3 \(\times\) 3 pixel neighbourhood), and three maps of potential direct solar radiation (Table 1). The variable aspect value was transformed from the circular aspect variable into a linearized form ranging from 0 to 1 by using the following transformation:

\[
\text{asp.val} = ((\cos(\text{aspect} - 22.5) + 1)/2)
\]

It represents a NNE-SSW gradient, which is believed to be the most discriminating gradient in aspects for plants on the temperate Northern Hemisphere. Relative topographic exposure was obtained for different spatial scales by applying circular moving windows to the DEM.

### Table 1. Sets of explanatory variables used in the analysis. The variables were grouped to form the building stones of the conceptual hypotheses.

<table>
<thead>
<tr>
<th>Variable group</th>
<th>Described phenomenon</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(P)</td>
<td>Past vegetation in the same cell in 1988.</td>
<td>past.veg</td>
</tr>
<tr>
<td>(N)</td>
<td>Distance to nearest patch in 1988.</td>
<td>dist</td>
</tr>
<tr>
<td></td>
<td>Proportion of neighbouring cells occupied by the vegetation type (radius of 20, 40, 60, 80 and 100 m) in 1988.</td>
<td>neigh20, neigh40, etc.</td>
</tr>
<tr>
<td>(E)</td>
<td>Aspect value</td>
<td>asp.val</td>
</tr>
<tr>
<td></td>
<td>Slope angle</td>
<td>slope</td>
</tr>
<tr>
<td></td>
<td>Relative topographic exposures at 6 different moving window sizes : 10, 30, 50, 70, 90 and 110 m</td>
<td>top.exp10, top.exp30, top.exp50, etc.</td>
</tr>
<tr>
<td></td>
<td>Topographic position index as a composite of the relative topographic exposures.</td>
<td>top.pos</td>
</tr>
<tr>
<td></td>
<td>Smoothed topographic position index (smoothing with a 3 (\times) 3 cell window)</td>
<td>top.poss</td>
</tr>
<tr>
<td></td>
<td>Potential direct radiation of summer (June) and winter (December) solstice, and of the spring equinox (March).</td>
<td>rad.june, rad.dec, rad.mar</td>
</tr>
</tbody>
</table>
Statistical analysis

Generalised linear models (GLMs) of the binomial family (McCullagh & Nelder, 1989) were fitted using the R software package (R Development Core Team 2008) to explain vegetation patterns depending on abiotic environment in 1988 and on the variable groups corresponding to our conceptual hypotheses in 2002. In each model, the response variable was represented by the presence/absence of one of the three target vegetation types in 1988 or 2002 (Table 2.). The modelling procedure consisted of four steps for each vegetation type independently: first, a correlation matrix among all continuous-scale explanatory variables was tested. No two variables with a Pearson correlation coefficient higher than 0.8 were allowed to enter any of the P, N or E models. Among those variables with higher correlations, we chose the one that showed the highest predictive power when entered into a GLM alone. Second, GLMs describing the abiotic dependence in the past as well as corresponding each conceptual hypothesis were built based on all remaining variables in the respective variable group. Linear and quadratic terms were included. Third, both backward and forward stepwise variable selection was performed based on the Bayesian information criterion (BIC; Schwarz, 1978). Models containing variables from simpler models (NE, PE, PNE) were built using all variables that remained in the final models of those, i.e. N and E, respectively (P was always used as one factorial variable; Appendix A and B). Composite models were always subject to stepwise selection, too. Finally, model quality was evaluated in a ten-fold cross-validation using both the widely reported area under the ROC curve (AUC; Swets, 1988) and the maximum of the “True Skill Statistics” (TSS; Allouche, Tsoar, & Kadmon 2006) over possible cutoffs. The latter has been introduced as an alternative of Cohen’s Kappa (Landis & Koch 1977), free of the prevalence bias typical to kappa. We used both accuracy measures because they are sensitive to different aspects of model performance.

Results

Both the abiotic models for 1988 and the models calibrated according to the conceptual hypotheses are given in Appendix A, their coefficients are given in Appendix B. Typically, one variable describing topographic exposure was dropped by stepwise selection from model PNE compared to the original NE models in the case of each vegetation type. Nevertheless, at least one variable describing topographic exposure was always retained. Performance ranking was nearly identical for both accuracy measures, although one minor difference in the significance of model differences occurred (Appendix C).

We found significant differences in the performance of abiotic models aimed at explaining the distribution of the same vegetation types in 1988 and 2002 (Table 2.). For the Calamagrostis type the abiotic dependence was

<table>
<thead>
<tr>
<th>Measures</th>
<th>AUC Medians</th>
<th>Maximised TSS Medians</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vegetation type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calamagrostis</td>
<td>0.8344</td>
<td>0.7678</td>
</tr>
<tr>
<td>Danthonia</td>
<td>0.7529</td>
<td>0.8496</td>
</tr>
<tr>
<td>Chrysopogon</td>
<td>0.8894</td>
<td>0.9371</td>
</tr>
</tbody>
</table>
significantly lower in 2002 than in 1988, while for the other two types abiotic factors were significantly better in explaining the 2002 pattern than the one from 1988 (Fig. 2, Table 2).

There were not only differences regarding the two time points, however. The relative performance of the various models targeted at explaining the vegetation patterns in 2002 greatly differed among the three vegetation types (Fig. 2). The model with the highest median performance was the *PNE* model (containing all the variable groups), the performance of which, however, was never significantly better than the next best models for any of the three vegetation types (Appendix C). We consider those models the most parsimonious – and thus the ones with the best theoretical explanation – that had the least number of predictor sets included, while being only insignificantly worse than the best performing and more complete model. For the *Calamagrostis* type, the model based on the variable group of distance and past neighbourhoods (N) performed the best, while the identity of the vegetation type previously recorded at today’s locations of the *Calamagrostis* type did not add significant information. The other two types, however, were the most parsimoniously predicted by current topo-environmental predictors alone (Fig. 3).

**Discussion**

The median performance of the best model for each vegetation type ranged between AUC values of 0.85 and 0.95, which is reasonably high when compared to earlier studies (see Zimmermann et al., 2007 for a discussion of prediction results). Regarding AUC, many authors agree that a performance between 0.7 and 0.9 corresponds to high model performance and values above 0.9 express excellent performance (Swets, 1988; McKenzie, Peterson, & Thornton, 2003; Guisan, Zimmermann, Elith, Graham, & Thuiller, 2007).

![Fig. 2](image-url) Performance of the models for the three vegetation types according to maximised TSS and area under the ROC curve (AUC). Medians and quartiles of 10-fold crossvalidation runs are shown. Legend: C – *Calamagrostis* type, D – *Danthonia* type, H – *Chrysopogon* type. x-Axis labels refer to the environmental model for 1988 distribution of the vegetation type in question (E88), and to the conceptual hypotheses: the distribution of the vegetation type in 2002 depends on environmental constrains (E), on past neighbourhoods (N), on the combination of the two (NE), on environmental constraints and local site history (PE), and on the combination of all three groups of variables (PNE).
We thus believe that the sets of stepwise optimised variables in our most parsimonious models contained the most important explanatory variables responsible for vegetation development after abandonment of grazing. TSS can be regarded as a successor of the kappa measure, but there is much less experience in interpreting its absolute values; therefore, we only interpret them in comparisons.

The change in the abiotic dependence of the three vegetation types can well be explained in relation to equilibrium. The *Danthonia* and *Chrysopogon* types are natural parts of the system and, therefore, it can be expected that during regeneration they get closer to a distribution mirroring their abiotic requirements, just as our results showed. On the contrary, the explanatory power of abiotic factors for the *Calamagrostis* type decreased from 1988 to 2002. As abandonment has occurred recently in our study site and *C. epigejos*, the dominant species of this type, is typically overtaking abandoned areas elsewhere as well, the *Calamagrostis* type is likely to have been in a phase of heavy spreading and this is the reason why its current distribution does not reflect abiotic constraints. This is also in good accordance with the results of the comparison of our models based on conceptual hypotheses for the 2002 distribution: models incorporating variables describing past conditions were most successful for this type. We conclude that historical patterns are specifically important where vegetation is not in pseudo-equilibrium with climate and land use, which was the case in our study, as grazing was abandoned in the late 1970’s only.

Yet, past conditions are most commonly assessed as the local site history exclusively, and past neighbourhood configuration is rarely considered. Predictive studies tend to use rather the current and not the historical neighbourhood information (e.g. in de Blois et al., 2001; Miller & Franklin, 2002; Benjamin et al., 2005; Miller et al., 2007). In these cases, neighbourhood relationships carry information about autocorrelation patterns, yet they lack the explicit mechanism and include dynamic vegetation processes only implicitly. A more mechanistic approach requires the investigation of past neighbourhoods or spatial configurations (Augustin et al., 2001; Carmel et al., 2001; Carmel & Flather, 2004; Muster et al., 2007). Past neighbourhoods, however, typically revealed marginal influence on vegetation dynamics in previous research, in sharp contrast to our results. In the case of *Calamagrostis* type, past neighbourhoods alone were capable of explaining the current distribution pattern. We conclude, that past neighbourhoods can add considerable information to the understanding of the development of vegetation patterns.

Although past conditions proved to be crucial predictors in our study, vegetation types greatly differed regarding which hypothesis provided the best model for their currently observed patterns. This is in accordance with results of recent studies, where different predictor sets were examined for different vegetation types separately (e.g. Augustin et al., 2001, de Blois et al., 2001, Carmel et al., 2001, Carmel & Flather, 2004; Pueyo & Alados, 2007). These studies, however, investigated differences between herbaceous and woody vegetation types only.
Our results prove that considerable differences may occur among herbaceous types as well. Such large differences may impede predictive vegetation modelling if the predictor set used for a complete site with various vegetation entities or species does not include all relevant variables (Fischer, 1990; Cairns, 2001; Pfeiffer et al., 2003; Weaver & Perera, 2004).

Having built models according to hypotheses (including e.g. historical aspects) rather than submitting a larger set of predictors to automatic variable selection, allows us to formulate conclusions about the possible mechanism how current vegetation patterns arose. The strong dependence of the current pattern of the Calamagrostis type on the distance and frequency of past Calamagrostis patches (N) indicates a vegetative propagation mechanism, which is in good accordance with the autecology of the dominant species (Rebele & Lehmann, 2001). As neither the identity of the previous vegetation type did influence its spread, nor abiotic constraints tightened for it in the last decade, the Calamagrostis type seems to be actively expanding without significant control by other types or by abiotic factors. Thus, it seems to be more competitive than all the other types at the site (possibly with the exception of shrubs). As a contrast, the distribution of the other two types is well-predicted by abiotic conditions alone. From this, we may infer that the expansion of these types was primarily directed by the availability of suitable sites after abandonment, which might also mean that they colonised vegetation types with similar abiotic requirements. We might even say that they can be considered to have come close to an equilibrium distribution since their distribution corresponds to that of the suitable sites and their dependence on these has increased lately. Thus, these types and their associated species can be expected to respond to changing environmental and land use situations most quickly. We are aware that our study is limited to a comparably small site. Yet, we believe that the demonstration of the role of the past, and especially the role of the past neighbourhood, as well as that of the differences in predictor sets for vegetation types does not depend on scale. We expect to find a similar importance in the predictive power of past conditions and similar differences between differing predictor sets at larger spatial scales as well.

Acknowledgements

Klára Virágh was supported by the Hungarian Research Fund (OTKA) Grant no. SUP062338. Imelda Somodi was also funded by the Hungarian Research Fund (OTKA), partly by the Grant no. SUP062338, partly the Grant no. NL68218. In the early phase of the project Balázs Székely contributed as a Békésy György Postdoctoral Fellow. Niklaus Zimmermann was supported by the EU FP6 project ECOCHANGE (GOCE 046866).

Appendix A–C. Supplementary material

The online version of this article contains additional supplementary data. Please visit doi:10.1016/j.baae.2009.11.006.

References


