

A discrete mathematical method for the analysis of spatial pattern

P. Ittész¹, É. Jakó^{2,3}, Á. Kun^{1,3}, A. Kun⁴ and J. Podani^{3,5}

¹Collegium Budapest, Szentháromság u. 2, H-1014, Budapest, Hungary

²Research Group of Ecology and Theoretical Biology, Hungarian Academy of Sciences and Eötvös University, Pázmány P. sétány 1/c, H-1117 Budapest, Hungary

³Department of Plant Taxonomy and Ecology, Eötvös University, Pázmány P. sétány 1/c, H-1117 Budapest, Hungary

⁴Research Institute for Botany and Ecology, Hungarian Academy of Sciences, Alkotmány u. 2-4, H-2163 Vácraát, Hungary

⁵Corresponding author. E-mail: podani@judens.elte.hu

Keywords: Boolean functions, Cellular automata, Graphs, Species pattern, Vegetation.

Abstract: A discrete mathematical method, based on the Jakó Iterative Canonical Forms (ICF) of Boolean functions is proposed for the analysis of species combinations and the detection of characteristic areas in plant communities. Information on species combinations (or florulas) appearing in a sample is expressed in compact form to reveal fundamental properties of community pattern. The new method provides a complementary tool for the florula diversity approach: whereas florula diversity is indicative of the frequency distribution of species combinations regardless their interrelationships, the new procedure detects complexity in the abstract structure of species combinations. Graph-theoretical representations of the ICF promote understanding the new method and visualizing its results. A cellular automata model and field data provide illustrative examples.

Abbreviations: CA – Cellular Automata, CDNF – Canonical Disjunctive Normal Form, FD – Florula Diversity, ICF – Iterative Canonical Form, NSC – Number of species combinations.

Introduction

Vegetation studies are often centered around the analysis of interspecific associations. The traditional methods generally focus on spatial dependence relations between pairs of species (Greig-Smith 1983). However, in order to obtain a more faithful picture on community level properties and processes, such analyses should be extended to species combinations (florulas, Podani 1984, Bartha 1992). Based on observations in early successional plant assemblages, Bartha (1992) concluded that “the multi-species relationships projected into pairwise relations resulted in over-simplified, illogical structure of coalitions and many relationships remained unexplored”. However, studying multi-species combinations is a very complex task and the representation of the results is difficult since the number of species combinations exponentially increases with the number of species. Plexus graphs (McIntosh 1973, Dale 1977) offer a simple solution for visualization and, indeed, these have long been the only

available tools to demonstrate direction and strength of relationships between species coalitions.

Evaluating species composition is an important component of the analysis of temporal processes, such as succession or degradation (Bartha 1992, Wimberly and Spies 2001, Collins 1992). The relative proportion of species or functional groups changes in time and space, and their examination is difficult without a compact representation of species composition. In a previous communication (Jakó and Ittész 1998), we described a discrete mathematical method developed for this purpose. The essence of the approach is that a large number of sample quadrats are taken with increasing sizes and the species combinations found are encoded by Boolean functions which have an easily interpretable graph theoretical representation as well. The number of vertices in this graph, the cardinality, may be plotted against quadrat size, thus visualizing spatial complexity of plant pattern and depicting characteristic areas of the community. The new analytical tool is illustrated

here through a simple vegetation model that simulates plant patterns representing an intact successional series as well as a temporal process involving some degree of disturbance. A small set of field data was also analyzed for comparison. The method is compared with an information theory procedure (florula diversity, Juhász-Nagy and Podani 1983) to illustrate how the two approaches complement each other.

Materials

Simulated data

Structural analyses of open and dry grasslands (Bartha et al. 1998, Mucina and Bartha 1999, Kun et al. 2000) serve as the basis for a cellular automata (CA) model to generate simple community patterns that resemble actual communities. The starting point is the common observation that the spatial architecture of one or more dominant clonal plants determines the spatial pattern of other populations and the temporal processes. This effect is exceptionally strong in dry and open grasslands with diverse life history characteristics and a wide range of dispersal capabilities of the constituting species (Pärtel and Zobel 1995, van der Maarel 1996, Klimeš 1999). The mobility of the key species is one of the main forces affecting the development of spatial pattern (Herben et al. 1997). The aim of this study is not the detailed modelling of such open and dry grasslands, the CA model is used only to create community patterns for illustrating the utility of our method under controlled circumstances.

We assumed the existence of five species in the community which may be classified into three functional groups: (1) an annual plant with good dispersal capability, (2-4) three subdominant species with intermediate dispersal capability and (5) a dominant clonal plant that strongly affects the development of community structure. In the CA model, we recorded the age of the plants in each cell. Age determined seed production since the plant individual had to reach a minimal age before it could produce propagules (clonal offspring or seeds). Furthermore, an individual died after reaching its maximum life span, a species dependent parameter. For every species, propagules were placed from the parent plants at a distance randomly chosen from a Gaussian distribution with standard deviation 1 and mean 0. The fate of the propagules was determined by the occupancy of the newly reached cell: the species established itself only if the cell was empty.

The functional groups and the corresponding states of the CA model may be summarized as follows:

- The first species is a semelparous annual species, for example *Hornungia petraea* (Mucina and Bartha 1999, Klimeš 1999). This species has strong dispersal capability and a one „year” long lifespan, but it is unable to displace any other species. The only way to reproduce is via seed production.
- The three subdominant species are a semelparous perennial forb, e.g., *Seseli leucospermum* (Mucina and Bartha 1999, Klimeš 1999), a dwarf shrub, such as *Helianthemum canum* (Erschbamer et al. 1983, Klimeš 1999), and a narrow-leaved graminoid, e.g., *Festuca pallens* (Mucina and Bartha 1999). In our simple model, these three species have similar behaviour. Their lifespan is shorter than that of the dominant species, but they produce more seeds. They are only able to displace the annual group, but not each other, nor the dominant species.
- The fifth species is a dominant, clonal, dispersal-limited sedge (a typical representative is *Carex humilis*). It is able to displace any other species from a given site. However, since generative reproduction is rare its long-range dispersal is limited (Wikberg and Svensson 2003), and thus it is allowed to colonize only adjacent cells. To get the often-observed ring-like growth form, the dominant species has two life stages. The first active stage is able to colonize and spread, but after some time it turns into a second stage without reproduction. The lifespan of this stage is the longest of the five species.

The size of the cellular automata field was 1000 by 1000 cells, and the runs were initialized with a random pattern such that 1% of the field was filled with the dominant species, 5% with each of the subdominants and 10% with the annual. Note that each cell was allowed to contain only a single species or it was empty.

Two types of patterns were generated. The first type (*Intact*) was generated through a regular succession without disturbance. The second type (*Disturbed*) was obtained such that at each time step the contents in the randomly selected 3% of cells of the CA grid were deleted. The resulting pattern preserved most visual aspects of the original intact pattern, yet it served as a good example of a degradation process. For comparison, a random version of the Intact community was also created: in the map of the 200th generation the positions of individuals were swapped randomly, but the numbers of occurrences for every species were kept constant.

We took samples from the CA field after 50 and 200 simulation steps (“generations”) by applying a series of quadrat sizes. For each size, 1000 quadrats were taken.

The starting point in this spatial series was the elementary cell of the automata therefore each quadrat could contain only one species or nothing. In each subsequent step, the quadrat sides were increased with one cell up to the size of 100×100 cells (Fig. 1). Pattern simulation, sampling and calculations of Boolean functions were performed by programs written in C and Java languages by the first author.

Field data

A map representing the point pattern of the most abundant six species (Table 1) in a perennial sand steppe community at Csévharaszt (Kiskunság National Park, Hungary) was prepared by Szócs (1977). The area is fairly homogeneous and all species exhibit aggregated spatial pattern. The positions of all individuals were digitized in form of Cartesian coordinates within an area of $2.4 \text{ m} \times 10 \text{ m}$. Based on these coordinates, the sampling series

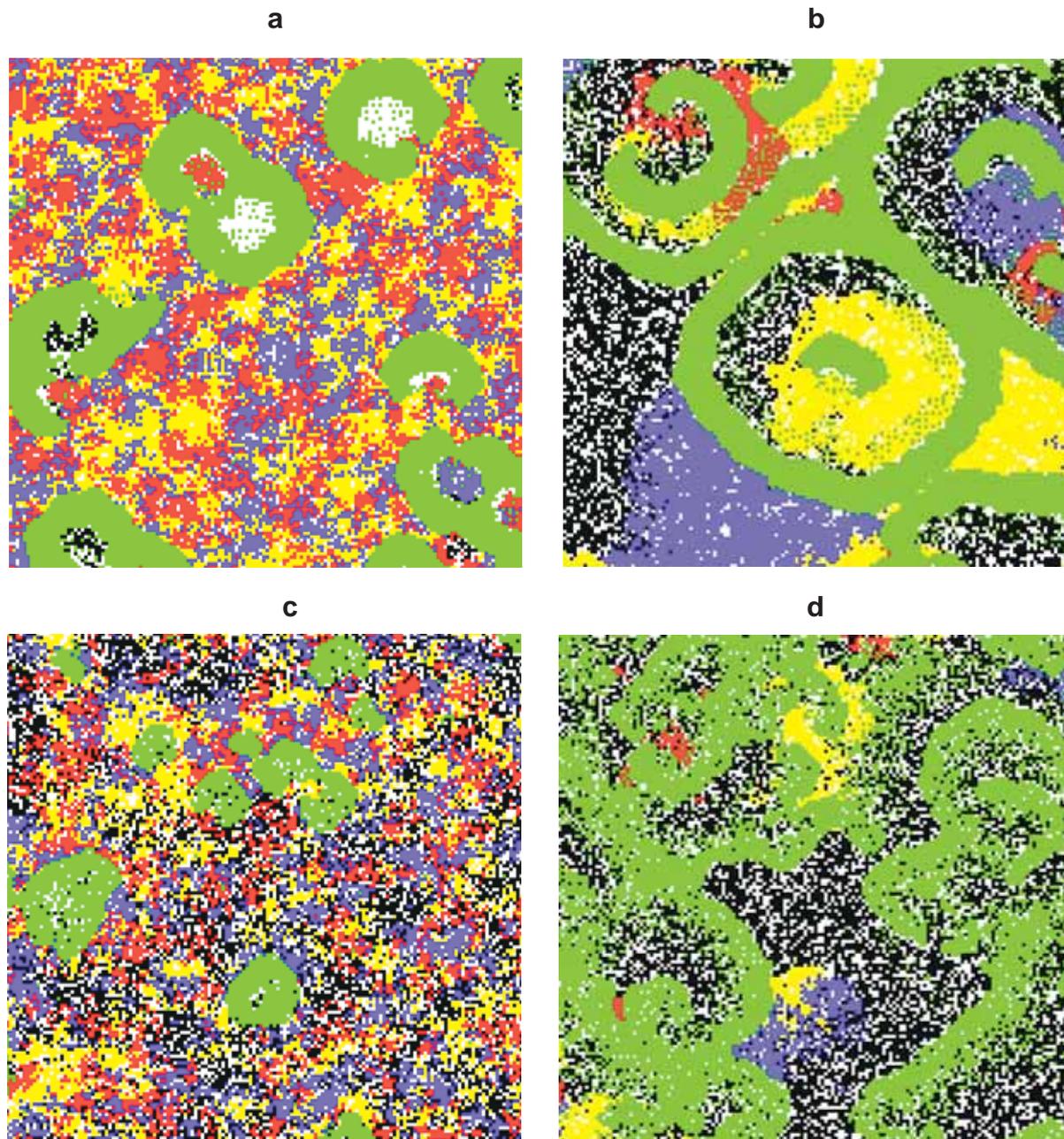


Figure 1. Patterns generated by the CA model. Segments of 300×300 cells from the original 1000×1000 field are shown. Intact: generation 50 (a), generation 200 (b), disturbed: generation 50 (c), generation 200 (d). Green indicates the dominant species. Red, blue and yellow colours reflect the three similar subdominant species. Black denotes the annual species and white identifies the empty cells.

Table 1. List of species occurring in the mapped region of the sand steppe community.

No.	Name	Number of individuals
1	<i>Minuartia verna</i>	50
2	<i>Tragus racemosus</i>	324
3	<i>Polygonum aviculare</i>	48
4	<i>Festuca vaginata</i>	287
5	<i>Stipa pennata</i>	147
6	<i>Koeleria glauca</i>	755

with increasing plot sizes was simulated using the simulator program ELSAM from the SYN-TAX program package (Podani 1994) to yield quadrat data for computing Boolean functions and florula diversity. The plot sizes were 78, 314, 707, 1256, 1963, 2827, 3838, 5026, 6361, and 7854 cm². A total of 300 plots were sampled for each plot size.

Methods

Developing a new Boolean technique based on the Iterative Canonical Form (ICF)

The Iterative Canonical Form (ICF) of Boolean functions (Jakó 1983) is used for deriving the so-called ICF-graphs which represent the species combinations. The new procedure advocated here has been described elsewhere (Jakó 1983, Frolov and Jakó 1991, Jakó and Ittész 1998, Jakó manuscript), while the general analytical expression and the algorithm for computing the ICF is given in the Appendix. Here, we describe the terminology and data structure in a simplified manner which requires minor mathematical background. We hope that this step-by-step presentation greatly facilitates understanding the fundamental concept.

a. A *Boolean algebra* is defined as a set and the following basic operations applied to its elements. Binary operations:

- OR (join, sum, disjunction, \vee , +),
- AND (meet, product, conjunction, \wedge , \cdot),

and a unary operation:

- NOT (negation, complementation, $'$, \neg).

b. A *Boolean variable* x_i can take two values: true (1) or false (0). In our case, variable x_i and its complement x'_i correspond respectively to the presence and absence of species i ($i = 1, \dots, n$) in a given sampling unit (quadrat).

c. A *Boolean function* is denoted by $f(x_1, \dots, x_i, \dots, x_n)$, in which each variable x_i takes the value of 0 or 1. The f function also returns the value of 0 or 1. In this paper, we use

Boolean functions to describe species combinations detected by m quadrats.

d. Let B^n denote an n -dimensional Boolean space or *binary n -cube*. The n -cube can be represented by a graph whose vertices correspond to binary vectors each of length n . The edges connect vertices differing in one element of the corresponding two vectors. Fig. 2.a illustrates a 3-cube. In ecological terms, a binary vector is a *species combination*. For example, (0,0,1) corresponds to a situation with species 1-2 absent and species 3 present.

In reality, usually a subset of possible combinations is realized in a sample. If a given species combination is detected by at least one quadrat, then the corresponding vertex becomes black, otherwise this vertex is white. Suppose that sampling by $m = 5$ quadrats yielded the following data matrix with three species:

	1	2	3	4	5
species 1	0	0	0	0	1
species 2	1	0	1	1	1
species 3	1	1	1	0	0

This matrix defines the colored 3-cube of Fig 2.b. The set of black vertices represents the species combinations realized in the sample, i.e., the number of black vertices equals the number of species combinations (*NSC*). Algebraically, the colored graph of Fig. 2.b is expressed by the following function:

$$f(x_1, x_2, x_3) = x'_1 x_2 x_3 + x'_1 x'_2 x_3 + x'_1 x_2 x'_3 + x_1 x_2 x'_3$$

Note that columns 1 and 3 are identical in the data matrix above, but these do not appear twice in the formula, known as the *Canonical Disjunctive Normal Form* (CDNF). The simplification of CDNF will be the cornerstone of the present approach, as we see below.

The number of possible combinations, that is, the number of vertices in the graph is $\omega = 2^n$. The number of possible colored graphs is therefore 2^ω , since each species combination may be present or absent (black or white).

e. It is seen in Fig. 2.a that the n -cube is organized to have several hierarchical levels. Each level corresponds to the number of species present in the combinations, ranging from $\binom{n}{0}$ to $\binom{n}{n}$. This number is called the *rank* of the combination. The graph also demonstrates that edges can only connect vertices between neighboring levels. That is, each edge connects a combination with another above it, which is generated by adding one more species, e.g., (001) \rightarrow (011).

f. In the CDNF, the binary vectors represented by black vertices in the n -cube are listed in an arbitrary ordering,

whereas the graph implies relationships via ordering by inclusion, expressed by the edges of the graph (Fig. 2.b). The principal question is therefore to find an analytical expression, which

- reduces redundancy of the CDNF,
- preserves coloring and partial ordering relationships implied by the edges of the n -cube.

In order to understand the derivation of the expression sought, consider the subgraph in Figure 2.c, derived from the graph of Figure 2.b. This subgraph contains all the information necessary to reproduce the one in Fig. 2.b if a set of certain rules are followed. For the reconstruction, consider the different role of white and black vertices in the graph of Fig. 2.c. A black vertex allows all those vertices above it, which are accessible from it on the n -cube. A white vertex prohibits all the accessible vertices above

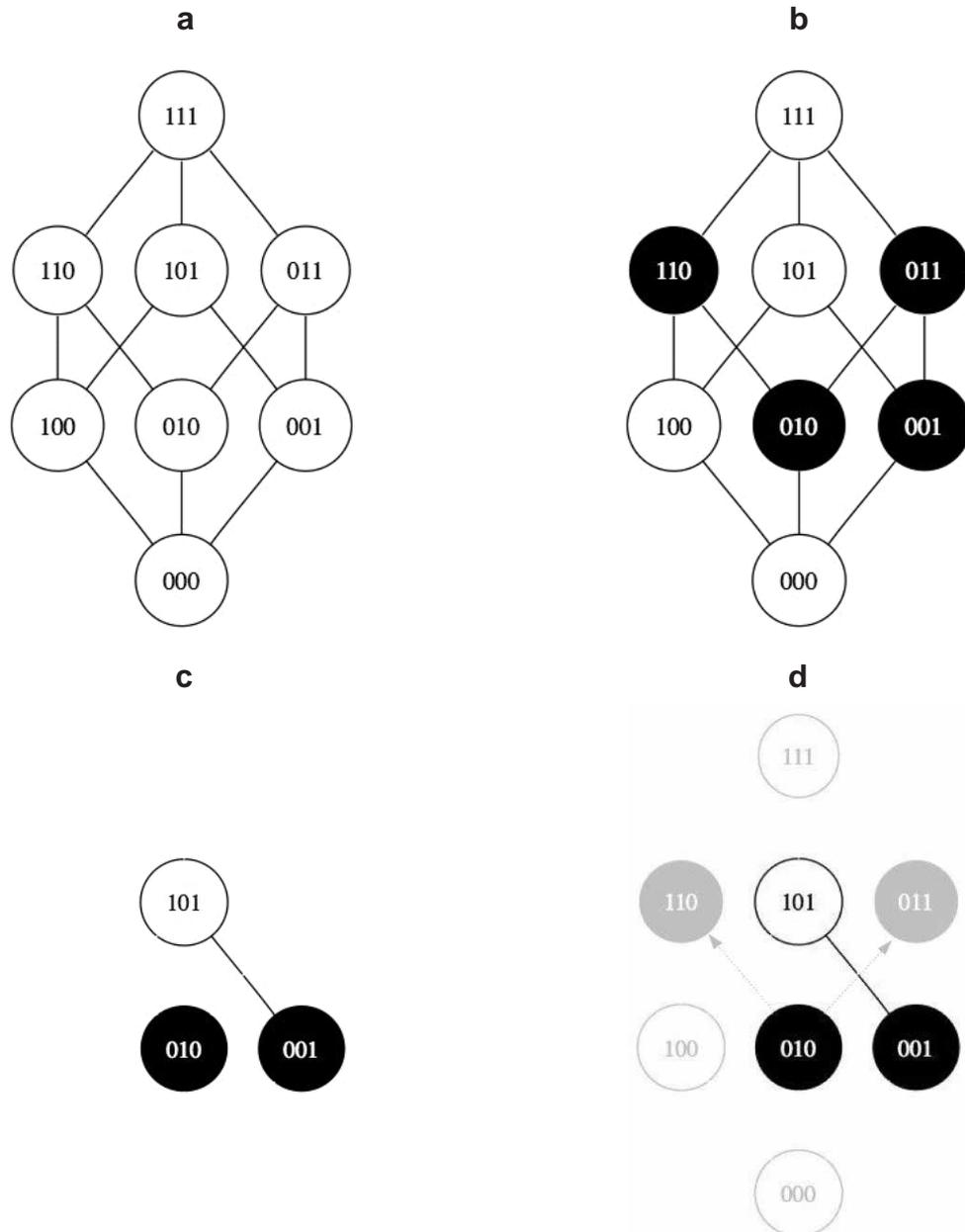


Figure 2. Derivation of the ICF for an artificial three-species case ($n = 3$). **a:** The n -cube shows all the possible species combinations (vertices) with links connecting species combinations that differ only in a single species. **b:** Graphical illustration of the CDNF showing realized species combinations in black, missing combinations in white. **c:** Graphical illustration of the ICF with black vertices representing realized species combinations that generate the other realized combinations of **b**, and a white vertex corresponding to the prohibited combination above it. **d:** Illustration of the generation process, vertices in full gray were generated from black vertices, as shown by arrows. Empty vertices in gray do not play any role in the ICF.

it on the n -cube. White vertices have priority; their presence overrides the effect of black vertices. White vertices are called the “prohibitors”. In the example, (110) and (011) are generated from (010), therefore they are not present in graph 2.c. The combination (111), which is not present in the data, does not appear because it is prohibited by (101) from below (see also Fig. 2.d). The edges play no direct role in the reconstruction, and only those remain that connect vertices differing in one element of the corresponding two vectors. Note the role of the lowest vertex representing the empty florula: regardless its color the vertices of rank 1 do not change. This explains why (100) is not regenerated. Its ecological rationale is that neither the presence nor the absence of empty florulas in the sample should influence the examination of species combinations.

The small example does not show a further possibility. It may happen that a given vertex is realized in the original data set, but a prohibitor appears below it. In this case, the realized has to be explicitly allowed and appears in the graph as a black vertex. An example will be seen later in Fig. 5.a.

The logical function corresponding to Figure 2.c is given by

$$f(x_1, x_2, x_3) = (x_2 \text{ OR } x_3) \text{ AND NOT } (x_1 \text{ AND } x_3) = (x_2 + x_3) \cdot (x_1 \cdot x_3)'$$

and is termed the *Iterative Canonical Form* (Jakó 1983, Jakó and Ittész 1998). It can be shown that the ICF is a unique representation of the CDNF (Jakó 1983, Jakó manuscript). As its name suggests, the determination of this function (and the corresponding ICF graph) requires an iterative algorithm. The details of the algorithm are described in the Appendix.

g. In a sense, the ICF can be considered as a parsimonious representation of the CDNF and therefore useful as an indicator of the complexity of the CDNF and, in turn, of the data from which the CDNF is derived. In this paper, we shall use the number of vertices of the ICF graph as the simplest possible measure of its complexity. This number is termed the *cardinality* of the ICF-graph and is abbreviated as C .

h. The ICF is scale-dependent for the simple reason that the underlying data matrix is also scale-dependent, i.e., it is influenced by the size of quadrats by which the community was sampled. Therefore, the ICF is determined for a series of quadrat sizes and its cardinality is plotted in the function of quadrat size as an illustration of the overall complexity of plant pattern. Particularly important is the

quadrat size at which C reaches its maximum, i.e., the characteristic area at which the community develops in its most *complex* form. Note that the relationship between C and NSC is not trivial, contrary to naive expectations, because C is determined by considering the rules by which the realized combinations can be generated from one another, and therefore C may be higher or lower than NSC .

Information theory methods

For the analysis of plant communities, Juhász-Nagy (1976, see also Juhász-Nagy and Podani 1983) developed a family of information theory functions that also rely on presence/absence data obtained by applying a series of quadrat sizes. The core of the approach may be described in terms of one formula selected from the model family, as follows. Let us identify the realized species combinations and determine the number of occurrences of each in order to obtain a frequency distribution. Then, similarly to the standard diversity approach in which Shannon's entropy is applied to the frequency distribution of species in a collection of plant individuals, apply the entropy function to the distribution of combinations:

$$m\hat{H} = m \log m - \sum_{k=1}^{\omega} f_k \log f_k$$

where m is the number of quadrats, f_k is the frequency of the k^{th} species combination and $\omega = 2^n$ as above. This formula is called the *florula diversity* and its value is also plotted against plot size to find the characteristic area at which the community develops in its most *diverse* form. (Note that log implies \log_2 in the sequel.) Florula diversity and several related functions have been extensively used for analyzing community pattern and for detecting underlying processes in a phase-space representation (Bartha et al. 1998).

Results and discussion

Artificial communities

Number of species combinations (NSC) vs. quadrat size.

The curves of NSC and ICF in the function of quadrat size differ with the number of simulation steps (generations) and disturbance. In the intact community, the number of species combinations reached its maximum (11) for quadrat size 25 after 50 steps. This maximum value increased to 14 by the 200th generation and the corresponding quadrat size became smaller (16 cells, Figs 3.a-b). An explanation of small characteristic areas is that the pattern-simulating model corresponds to an environment which generates a well-determined patchy structure (Fig. 1.a-b). The ring-form growing type of the dominant species pro-

duces empty patches that are easily invaded by the annual species and later the subdominants can settle down inside the rings. In the late stages of the succession, the species form separate patches determined by the growing and disappearing rings of the dominant species.

In the disturbed situation, *NSC* peaked (16) at quadrat size 4 after 50 steps, and then the maximum decreased to 8 in step 200. The quadrat size pertaining to this maximum became extremely large, ca. 100 times larger than 150 steps before (Fig. 3.c-d). That is, the number of the realized species combinations is higher in the earlier stages, but in the later stages the subdominant species cannot spread as effectively as in the intact situation therefore the structure becomes poorer.

ICF cardinality vs. quadrat size. The ICF cardinality curves of generations 50 and 200 from the intact community differ more substantially than the *NSC* curves (Figs 3.a-b). The peak of the 50th generation is at quadrat size 16 while the peak of the generation 200 is at the quadrat size of 121 cells. In the 200th generation, the ICF cardi-

nality curve has a local minimum at quadrat size 16 where the *NSC* curve is peaked. It is interesting that there are abrupt changes around the ICF cardinality peak in the intact community after the 200th generation.

The ICF cardinality curve in the 50th generation of the disturbed community has a local minimum at quadrat size 9 and its peak is at quadrat size 36. Note that the maximum point of the *NSC* curve is at quadrat size 4 (Figs 3.c-d). The maximum points are at 16 and 25 in the case of the 200th generation. The minimum point or minimum range of the 200th step is the same as the maximum range of the *NSC* curve. The characteristic area of the ICF cardinality is greater in the case of the 50th, disturbed generation than in the same generation of the intact scenario, but in the 200th disturbed generation the ICF characteristic area is much smaller than in the intact community. Random disturbance explains the observations for both situations. In the early phase of the succession, patch structure is disturbed, but the abundances of the species are not changed dramatically compared to the starting propor-

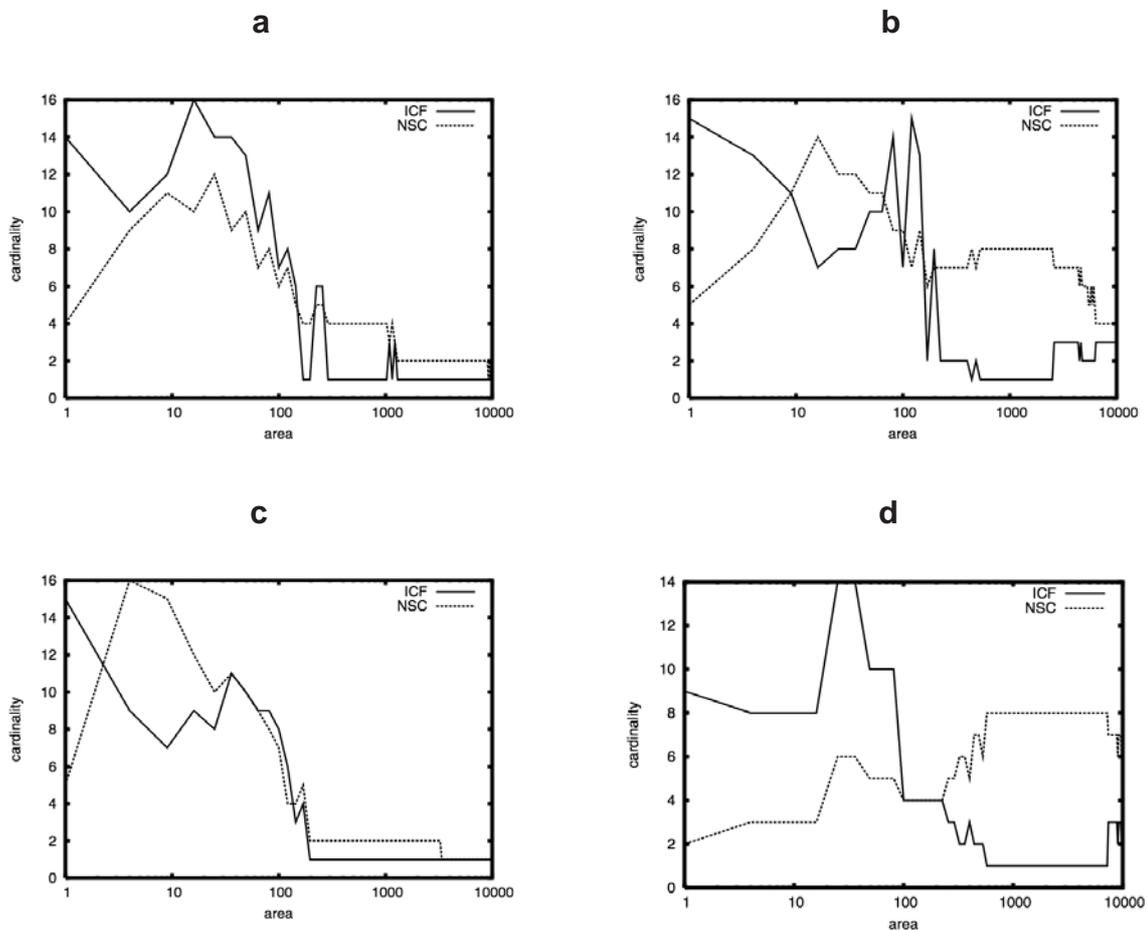


Figure 3. The cardinality of ICF and the number of species combinations (*NSC*) in the different stages of the simulated succession. Intact: generation 50 (a), generation 200 (b), Disturbed: generation 50 (c), generation 200 (d).

tions. In the later phase (200th generation), however, the subdominants are less abundant and the patches are less defined as well, therefore smaller area is enough to find the maximum ICF area. The most interesting point of this result is that there are cases when *NSC* is high, but the cardinality of the ICF is low. Since the ICF algorithm performs a logical minimization on the input data, a monotonous input structure results in low ICF cardinality. If the distributions of the species are closer to random, then the realized combinations in a data set tend to be monotonous. Monotonous means that each possible combination is realized from the 1-species combinations to the n -species combinations.

Florula diversity vs. quadrat size. The entropy curves of spatial series show that the maximum points for patterns from later stages of the simulations are shifted to the larger quadrat sizes (Fig. 4). The patchy growing type of the dominant and the subdominant species caused this phenomenon. As patches extend, greater sample area is needed to get a larger number of species combinations

and hence higher *FD*. Disturbing the structure causes smaller and fewer subdominant patches to occur, therefore the *NSC* curve has a very low local minimum at generation 200, which is emphasized by the entropy curve as well (Fig. 4.d). The peak of the florula diversity in the case of the disturbed community in the 200th generation overlaps the minimum range of the ICF cardinality. This shows that the structure is weak, and it is closer to a random situation.

ICF graphs. The ICF-graphs for a selected quadrat size (100) in the 50th and 200th generations are shown in Figure 5. In the case of the intact community, the ranks range from 1 to 5. Species 5 (dominant) is the only representative of the rank 1 combinations and species 1 (annual) and 5 can form a two-species combination of rank 2. Note that this combination is hidden in the graph (Fig. 5.a) because the rank 1 node generates it. We can conclude that the few scattered generating and prohibiting nodes of the graph refer to a strong structure in the simulated pattern.

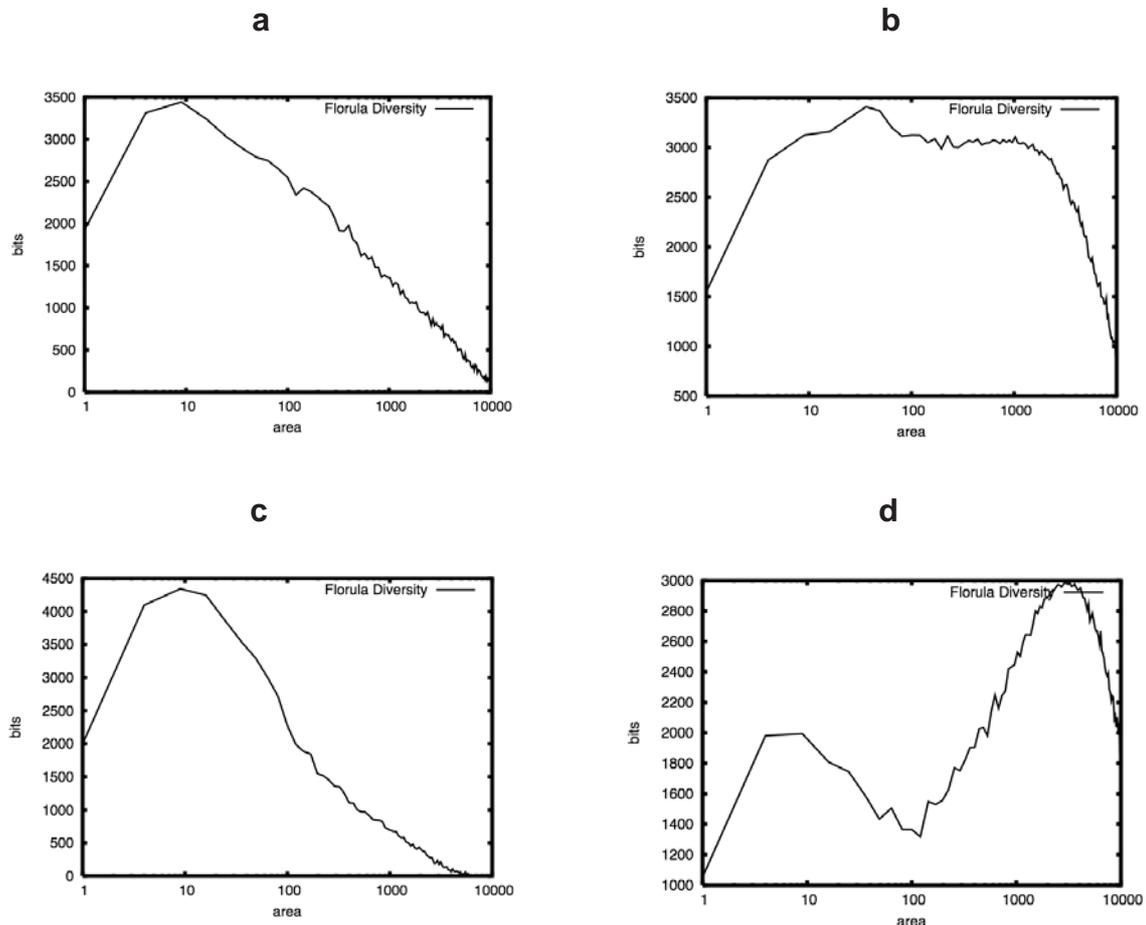


Figure 4. Florula diversity from the simulated communities. Intact: generation 50 (a) generation 200 (b), Disturbed: generation 50 (c), generation 200 (d).

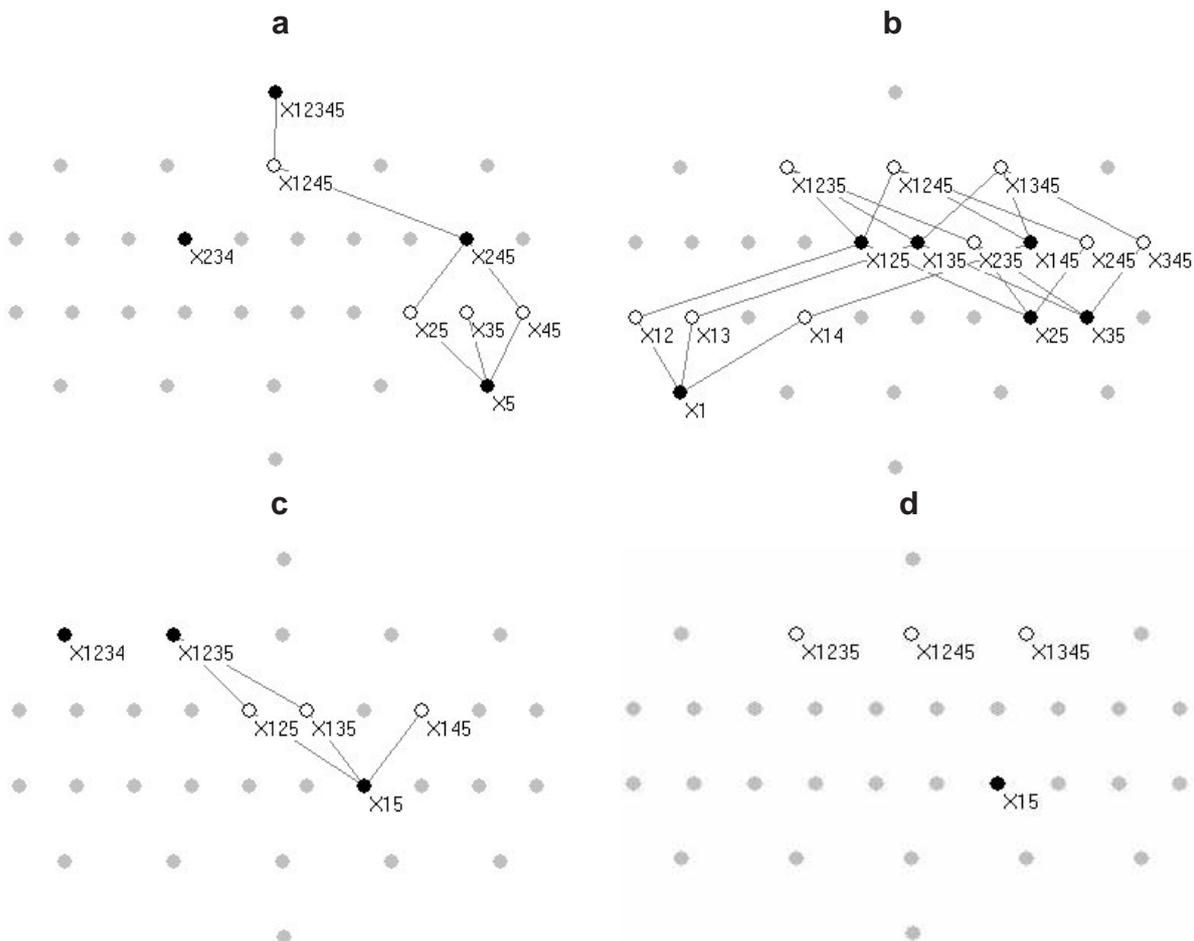


Figure 5. ICF-graphs obtained for quadrat size 10×10 . Intact: generation 50 (a), generation 200 (b), Disturbed: generation 50 (c), generation 200 (d). Note that this figure uses different symbols from those in Fig. 2 due to the limited space. For example, X1234 represents the species combination (11110), whereas X15 = (10001).

In the 200th generation of the intact community the number of the nodes in the ICF graph is higher than in the 50th generation, but at the same time the range of realized ranks is dropped to 1-3 (Fig. 5.b). The 4th level of the graph contains only prohibiting elements and there is no generating element on the 3rd level, which would be able to generate any hidden element from the 4th level.

The growth type of the dominant species causes most of the differences between generations. In the 50th generation, the patch size of the dominant species is smaller and the distribution of the subdominants is more random. The annual species is very rare in this stage (Fig. 1.a). In the 200th generation, the patch size of the dominant is greater and the ring shape of the patches is more distinctive. This dynamical structure creates empty habitats for the annual species, which is therefore more frequent than in the 50th generation and the subdominants are considerably segregated (Fig. 1.b). This process explains the shift of the ICF cardinality peak towards greater quadrat sizes in the case of the 200th generation. The quadrat size 100 is around the

maximum point of the ICF cardinality curve in the 200th generation, but this size is in the descending part of the curve in the 50th generation (Fig. 3.a-b).

In the 50th generation of the disturbed community (Figure 5.c) the range of the ranks is narrow (2-4) compared to the intact community. There is only one two-species combination (1,5), the combination of the annual and the dominant species. There is no three-species combination and there are two elements at the 4th rank.

For generation 200, the difference between the disturbed and the intact communities is even more emphasized (Fig. 5.d). The range of the ranks is narrower, containing only two and three species combinations. The fourth level contains prohibiting elements, similarly to the intact case (Fig. 5.b). Another important feature of this graph is that the third level contains all the combinations which include species 1 and 5, but they are hidden and therefore not shown in the graph. These differences reflect the weaker structure of the disturbed case. The ran-

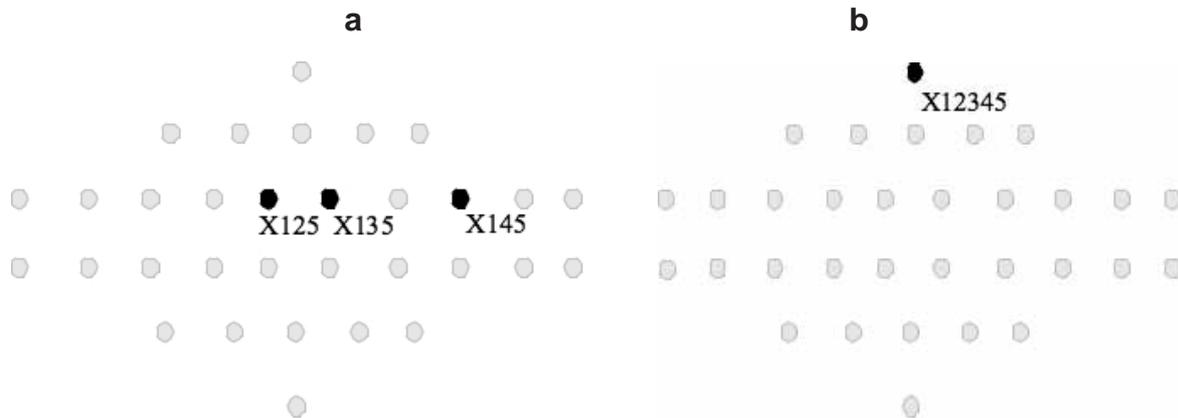


Figure 6. ICF-graphs for combinations detected by 5×5 (a) and 10×10 (b) quadrats from the randomized community of the intact CA model at generation 200.

dom disturbance of the community makes the patch structure weaker and the frequencies of the subdominant species are decreased while the frequency of the annual is increased (Fig. 1.c-d).

ICF graphs for the randomized pattern. The ICF graphs for the randomized pattern are shown in Fig. 6 for two quadrat sizes. At size 5×5 , three combinations with 3 species and all the combinations with higher ranks appeared in the sample. The only restriction is that the dominant species and the annual species are represented in every quadrat because these species are more abundant than the three subdominant species. The ICF-graph for quadrat size 10×10 consists only of one node at the highest level of the n-cube, meaning that all the species were present in every 10×10 cells size quadrat. Needless to say that for quadrat size 10×10 plant pattern was less structured than in the previous example with patchy growing types (cf. Figure 5), and the quadrats containing 25 cells indicate much less structure as well.

Sand steppe community

Figure 7 shows the change of *NSC*, *C* and *FD* in the function of quadrat size. The maxima appear at the same quadrat size (1256 cm^2), i.e., the three functions indicate the same characteristic area. Nevertheless, the shapes of the curves radically differ. Whereas the *NSC* curve fluctuates in the narrow range of 5 to 12, the other two functions suggest more drastic changes. Florula diversity has one maximum, which is often reported from similar studies. The ICF curve, however, has a more complex shape. It has four characteristic points: at quadrat sizes 707 cm^2 , 1256 cm^2 , 6361 cm^2 and 7854 cm^2 , representing a local minimum, the maximum point, the absolute minimum and a final increase of the ICF cardinality, respectively.

In order to capture the details of the spatial series producing these extreme points, let us consider the ICF graphs in Fig. 8. For the smallest quadrat size (78 cm^2) each sampling unit contains only a single species or nothing, leading to the ICF graph of Fig. 8.a. It has a relatively high number of vertices because vertices of the rank 1 are black and all rank 2 vertices have to be white to prohibit all higher level combinations. Thus, the high cardinality for these small quadrats is a misleading indication of high

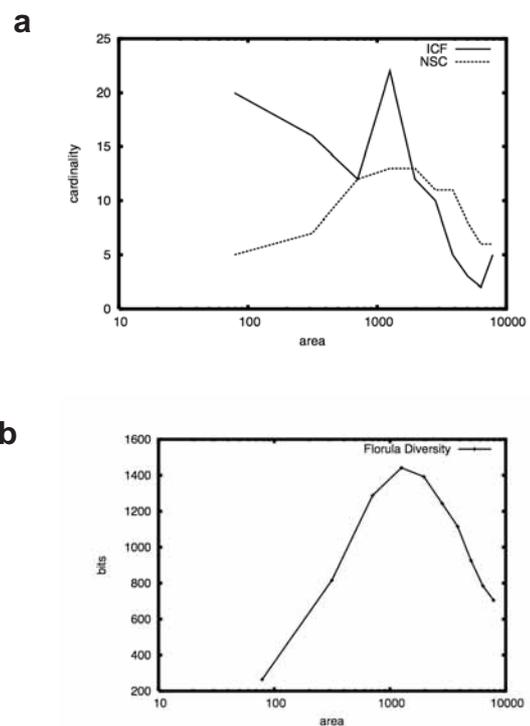


Figure 7. Results for field data. Cardinality of ICF and the number of species combinations (a) and florula diversity (b).

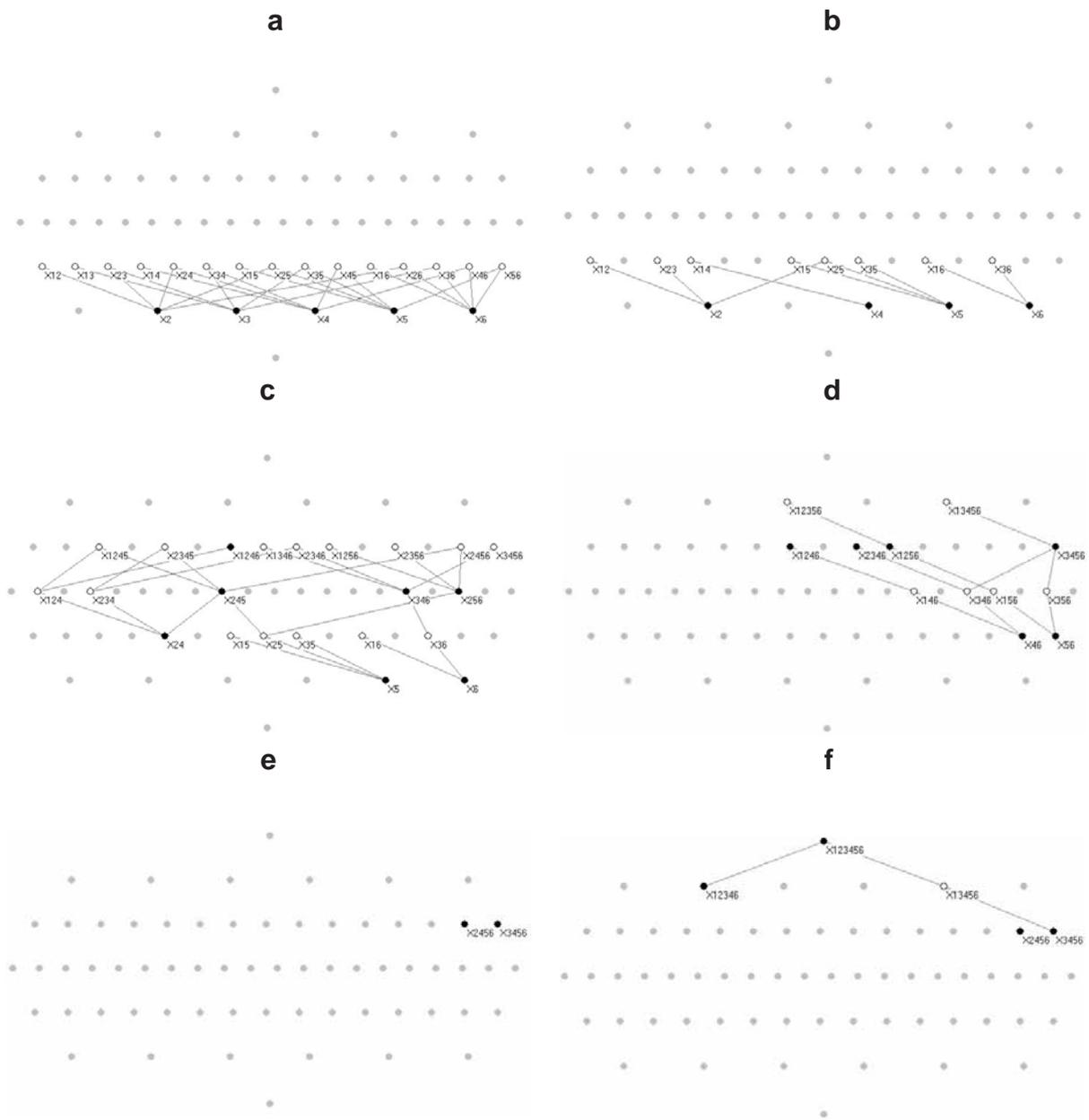


Figure 8. The ICF-graphs for the characteristic points of the ICF set cardinality curve from the field data at quadrat sizes 78 cm^2 (a), 707 cm^2 (b), 1256 cm^2 (c), 1963 cm^2 (d), 6361 cm^2 (e), and 7854 cm^2 (f).

pattern complexity because each quadrat can in fact include only one species.

It is striking that for the third quadrat size (707 cm^2) certain two-species combinations appear, yet the cardinality of the ICF is much lower (Figure 8.b). The presence of four rank 1 vertices shows that four species can form fairly large homogeneous patches.

The 4th step (1256 cm^2 , Figure 8.c) corresponds to the peak of the curve and the graph is undoubtedly the most complex of all with the number of prohibiting elements

still higher than the number of positive elements. Two species (5, 6) can still form pure patches at this quadrat size and there are higher level combinations as well. The species combination with the highest rank is shown explicitly, i.e., {X1246} is of rank 4. We have to emphasize that in spite of the lack of prohibiting elements above this generating element, the great number of prohibiting elements on the same level make redundant the prohibiting elements of rank 5. In addition, two pairs of species (5, 6) and (2, 4) can form patches together at this size, although only the second combination has to be shown explicitly.

Species combinations including these two elements together can also be found on the higher levels of the ICF-graph. It can be concluded, therefore, that this combination plays an important role in shaping the structure of the vegetation.

To evaluate the radical changes in the ICF graph, let us see the graph for the 5th step (1963 cm²) in Figure 8.d. The cardinality of the ICF is dropped from 22 to 12 from step 4 to step 5, meanwhile *NSC* is slightly increased. The lowest rank is moved to 2 and the highest rank is moved to 5, but the 5th rank combinations are not represented explicitly in the graph. The background of the cardinality change is that the number of prohibiting elements is lower, which suggests that the species are less segregated at this quadrat size.

Step 9 (6361 cm²) is the absolute minimum of the ICF cardinality curve. Only two vertices from rank 4 are represented in the ICF graphs and there are no prohibiting vertices at all. It means that the two combinations at rank 4 generate all their extended combinations at rank 5 and rank 6.

Step 10 (7854 cm²) produces a slight increase in ICF cardinality (with no increase in *NSC* and florula diversity). From the ICF-graphs (Fig. 8.e-f) it is clear that at step 9 the realized combinations are extensions of two combinations from rank 4. Thus, the elements from higher ranks are redundant. At step 10, there is a prohibiting element of rank 5 (1,3,4,5,6) and a new positive element (1,2,3,4,6) appears at rank 5 as well. The prohibiting element (1,3,4,5,6) at step 10 (Figure 8.f), an extension of element (3,4,5,6), is implicitly allowed at step 9 (Figure 8.e), therefore we can suspect some segregation of species at plot size 10. The examination of the graphs shows that species 2 and 6 are rarely found together in the same quadrat and at this level the spatial series reached a characteristic size again where we can find a prohibited composition of species 6 and 2 and they can be found together only in the trivial case when all of the species appeared together.

Concluding remarks

The application of discrete mathematical methods opens new perspectives in the analysis of plant pattern based on presence/absence information. The Iterative Canonical Form and its graphical variant express how the realized species combinations can be generated from one another most effectively. In a sense, therefore, the ICF is a most parsimonious representation of interrelationships between species combinations. If we display the cardinality of the ICF graph against quadrat size, characteristic ar-

reas can be identified. The maximum area which corresponds to a scale at which the species combinations develop in the most *complex* way is of particular interest. More precisely the cardinality of the ICF shows the minimum number of elements required to characterize the pattern described by the sampling units. Our simulation experiments and analyses of field data demonstrated that, in comparison with the florula diversity approach, the ICF-based analysis also provides meaningful and interpretable results. Since *FD* depicts underlying structure in terms of diversity, the two methods are complementary and cannot be expected to coincide under all circumstances. However, ICF is much more sensitive to changes in the composition of the set of realized combinations, especially concerning changes in the ranks of the combinations. As a consequence, random data structure is manifested as an ICF graph with much less vertices than in an ICF graph derived from a structured community.

The new approach detects differences between intact and slightly disturbed successional processes efficiently. Also, the differences between a simulated pattern and its random variant are obvious when the two ICF graphs are compared. Joint examination of the cardinality curves and the ICF graphs shows, however, that whereas the cardinality emphasizes crucial points in the spatial series, it is only a preliminary measure of complexity. Very simple structures (e.g., one species per each quadrat at the given scale) may appear similar to much more complex structures in terms of cardinality. A more elaborated graph theoretical measure, which considers the number of generating and prohibiting elements and their relationships expressed by present and hidden edges awaits development. This is especially important in communities with more species than used in this paper, because the ICF graph cannot be displayed efficiently for more than 8 species. Future extension of ICF includes its application to multivariate analysis of communities based on their inherent complexity. Distances can be calculated for pairs of ICF graphs, and then the distance matrix subjected to clustering or ordination. Another promising extension of the method is to involve the frequencies of the realized species combinations in the representation of the results. This frequency-weighted analysis would be more important in the case of classification of communities and it would emphasize the importance of given multi-species combinations or combination groups.

Acknowledgements. Financial support from the Hungarian National Scientific Grant (OTKA) is greatly acknowledged. Á. Kun is a postdoctoral fellow supported by OTKA (D048406), J. Podani was supported financially by Grant no. SU43732. We acknowledge comments from two referees, M. Dale and S. Bartha.

References

- Bartha, S. 1992. Preliminary scaling for multi-species coalitions in primary succession. *Abstracta Botanica* 16:31-41.
- Bartha, S., T. Czárán and J. Podani. 1998. Exploring plant community dynamics in abstract coenostate spaces. *Abstracta Botanica* 22:49-66.
- Bartha, S., T. Rédei, Gy. Szollát, J. Bódis and L. Mucina. 1998. Compositional diversity and fine-scale spatial patterns of dolomite grasslands on contrasting slopes. In: Csontos, P. (ed.), *Sziklagyeppek szünbotanikai kutatása*. Zólyomi Bálint professzor emlékének. Scientia Kiadó, Budapest. pp. 159-182.
- Collins, S.L. 1992. Fire frequency and community heterogeneity in tallgrass prairie vegetation. *Ecology* 73: 2001-2006.
- Dale, M.R.T. 1977. Graph theoretical analysis of the phytosociological structure of plant communities: the theoretical basis. *Vegetatio* 34: 137-154.
- Erschbamer, B., G. Grabherr and H. Reisigl. 1983. Spatial pattern in dry grassland communities of the Central Alps and its ecophysiological significance. *Plant Ecology* 54: 143-151.
- Frolov, A. and É. Jakó. 1991. Algorithms for recognition of partially ordered objects and their application. *Scripta Technica* 29: 109-118.
- Greig-Smith, P. 1983. *Quantitative Plant Ecology*. 2nd ed. Blackwell, Oxford.
- Herben, T., F. Krahulec, V. Hadincová and S. Pecháčková. 1997. Is a grassland community composed of coexisting species with low and high spatial mobility? *Folia Geobot. Phytotax.* 29: 459-468.
- Jakó, É. 1983. *Iterative Canonical Decomposition of Boolean Functions and its Application to Logical Design*. PhD Thesis, Technical University, Moscow.
- Jakó, É. (manuscript). Recognition and classification of discrete objects by Iterative Canonical Forms (ICF) of Boolean functions. *Journal of Mathematical Chemistry*.
- Jakó É. and P. Itzész. 1998. A discrete mathematical approach to the analysis of spatial pattern. *Abstracta Botanica* 22:121-142.
- Juhász-Nagy, P. 1976. Spatial dependence of plant populations. I. Equivalence analysis (an outline for a new model). *Acta Bot. Hung.* 22: 61-78.
- Juhász-Nagy, P. and J. Podani. 1983. Information theory methods for the study of spatial processes and succession. *Vegetatio* 51:129-140.
- Klimeš, L. 1999. Small-scale plant mobility in a species-rich grassland. *Journal of Vegetation Science* 10: 209-218.
- Kun, A., P. Itzész and D. Krasser. 2000. Coenological gradients in the Hungarian rocky vegetation I. Landscape-level studies. V. Magyar Ökológus Kongresszus előadásainak és poszttereinek kivonatai. Debrecen, 2000. p. 93.
- McIntosh, R. 1973. Matrix and plexus techniques. In: R. H. Whittaker (ed.), *Ordination and Classification of Plant Communities*. Junk, The Hague. pp. 157-191.
- Mucina, L. and S. Bartha. 1999. Variance in species richness and guild proportionality in two contrasting dry grassland communities. *Biologia (Bratislava)* 54: 67-75.
- Pärtel, M. and M. Zobel. 1995. Small-scale dynamics and species richness in successional alvar plant communities. *Ecography* 18: 83-90.
- Podani, J. 1984. Analysis of mapped and simulated vegetation patterns by means of computerized sampling techniques. *Acta Bot. Hung.* 30: 403-425.
- Podani, J. 1994. *Multivariate analysis in ecology and systematics*. SPB Publishing, The Hague, The Netherlands.
- Podani, J., T. Czárán and S. Bartha. 1993. Pattern, area and diversity: the importance of spatial scale in species assemblages. *Abstracta Botanica* 17:37-52.
- Szöcs, Z. 1977. New computer-oriented methods for the study of natural and simulated vegetation structure. In: L. Orlóci, C. R. Rao and W. M. Stiteler (eds.), *Multivariate Methods in Ecological Work*. International Co-operative Publishing House, Burtonsville, MD, USA. pp. 301-308
- van der Maarel, E. 1996. Pattern and process in the plant community: fifty years after A.S. Watt. *Journal of Vegetation Science* 7: 19-28.
- Wikberg, S. and B.M. Svensson. 2003. Ramet demography in a ring-forming clonal sedge. *Journal of Ecology* 91: 847-854.
- Wimberly, M.C. and T.A. Spies. 2001. Influences of environment and disturbance on forest patterns in coastal Oregon watersheds. *Ecology* 82: 1443-1459.

Appendix

The Iterative Canonical Form (ICF) of Boolean functions

The Iterative Canonical Form (ICF) of Boolean functions (Jakó 1983, Jakó submitted) is used in the present study to derive the ICF-graphs representing the species combinations. The formal expression of the ICF corresponds to an external analytical structure $\vee(D \& \neg D)$, where D (or $\neg D$) denote disjunctions (or a negation of disjunctions) of multivariable monotone conjunctions. These disjunctions (or negations of disjunctions) are defined on the basis of two disjoint, logically reduced subsets of elements X and Y , representing the *on*-set, and *off*-set values of the given Boolean function $f(x_1, \dots, x_n)$, respectively. The program of **J** ICF performs a logical partitioning and minimization of the initial data structure, represented as binary n -cube subsets $M_f = M_f^{(1)} \cup M_f^{(0)}$, where $M_f^{(1)} \cap M_f^{(0)} = \emptyset$. There exists a one-to one correspondence between the set of all binary vectors $(\mathbf{a}_1, \dots, \mathbf{a}_n)$ of the binary n -cube and the set of all monotone conjunctions of ranks 0, 1, ..., n . In other words, by logical partitioning of the given data structure according to the algorithm of the ICF, the elements of the sets $M_f^{(1)}$ and $M_f^{(0)}$ are transformed into some logically reduced disjoint subsets of elements $\{X_i, Y_i\}$, where $X_i \subseteq M_f^{(1)}$, and $Y_i \subseteq M_f^{(0)}$. Let $\phi_{i1}(x_1, \dots, x_n)$ and $\phi_{i0}(x_1, \dots, x_n)$ be two partially defined Boolean subfunctions that take the binary n -tuples from the set X_i to the value '1', and from the set Y_i to the value '0', respectively. Then, the algebraic expressions of the structural units $\{X_i, Y_i\}$ can be represented by the subfunctions $\phi_{i1}(x_1, \dots, x_n)$ and $\phi'_{i0}(x_1, \dots, x_n)$ as disjunctions or *negations* of disjunctions of monotone conjunctions, respectively. The general expression of a Boolean function $f(x_1, \dots, x_n)$ in Iterative Canonical Form is defined as:

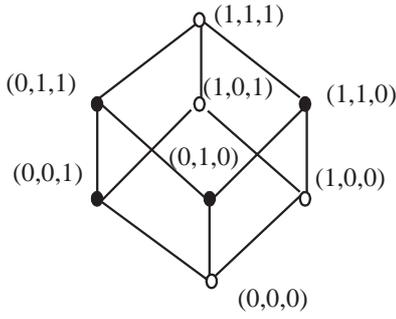


Figure A.1. The set $M = \{(0,0,1), (0,1,0), (0,1,1), (1,1,0)\} \subseteq B_3$

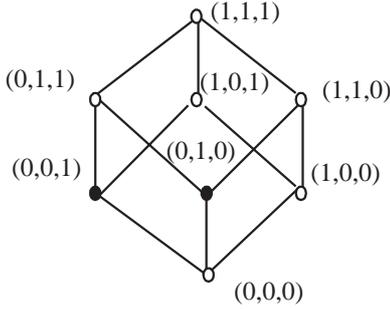


Figure A.2. The set $\beta(M) = \beta\{(0,0,1), (0,1,0), (0,1,1), (1,1,0)\} = \{(0,0,1), (0,1,0)\} \subseteq B_3$ as a result of β -contraction.

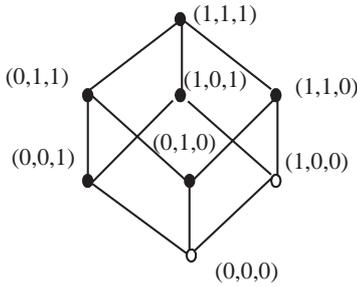


Figure A.3. The set $\alpha(\beta(M)) = \alpha(\beta\{(0,0,1), (0,1,0), (0,1,1), (1,1,0)\}) = \alpha(\{(0,0,1), (0,1,0)\}) = (x,x,1) \cup (x,1,x) \subseteq B_3$ as a result of α -expansion following the β -contraction.

$$f(x_1, \dots, x_n) = \bigvee_{i=1}^{i=k} \phi_{i1}(x_1, \dots, x_n) \cdot \phi'_{i0}(x_1, \dots, x_n),$$

where

$$i \leq \left\lceil \frac{n+1}{2} \right\rceil$$

is the number of iteration steps during the logical partitioning, and $j = \overline{1,0}$ denotes the parities of the subfunctions ϕ_{i1} and ϕ_{i0} , respectively.

For logical minimization of the binary n -cube subsets, two dual logical operations for *contraction* (β) and *expansion* (α) of terms were introduced, named as β -contraction, and α -expansion, respectively. The operation of β -contraction takes the given subset M to the set of its minimal elements, whereas the operation of α -expansion takes the subset M to the union of its subset elements, defined by the corresponding intervals on the binary n -cube. It is important that the α -, and β - operations are dual, and are not restricted to the distance 1 concept as in conventional methods of logical minimization (where two terms combine only if they are distance 1 apart).

Example. The initial data structure, and the logical operations of β -contraction and α -expansion are illustrated schematically in Figures A.1-A.3, respectively.

As seen from Figure A.2, the logical operation of β -contraction takes the given subset M to the set $\beta(M)$ of minimal elements of this subset. Thus, the term $(0,0,1)$ reduces by relation of inclusion the term $(0,1,1)$, whereas the term $(0,1,0)$ reduces the terms $(0,1,1)$ and $(1,1,0)$, respectively. In turn, as shown in Figure A.3, the operation of α -expansion takes the subset M to the union $\alpha(M) = \bigcup_{a \in M} \alpha(a)$ of its subset elements. In other words, the terms $(0,0,1)$ and $(0,1,0)$ generate a subset of terms $\{(0,1,1), (1,0,1), (1,1,0), (1,1,1)\}$, which corresponds to the set of vertices of the binary n -cube connected by the relation of inclusion (or partial order).

The main steps of the algorithm of the ICF for derivation of the general analytical expression (1) based on structural units $\{S_{i1}, S_{i0}\}$ are as follows:

Define the initial data sets: $M_f^{(1)}$ and $M_f^{(0)}$.

1. Open an empty list L of sets. Assign $M = M_f^{(1)}$, $i=1$.

2. Until $M \neq \emptyset$ do

 Calculate $S_{i1} = \beta(M)$ and insert the set S_{i1} into the list L of temporary results.

 Calculate $M = \alpha S_{i1} \cap M$.

 Calculate $S_{i0} = \beta(M)$ and insert the set S_{i0} into the list L of temporary results.

 Calculate $M = \alpha S_{i0} \cap M$. Assign $i=i+1$.

When the algorithm terminates, the list L will contain the sets of structural units S_{i1} and S_{i0} .

Remark. The process of logical partitioning is finished when any of the subsets S_{ij} is an empty subset. Obviously, $S_{11} = \emptyset$ is an unlikely occurrence, whereas if S_{i0} is zero, it will not be represented in the formulae, since $\neg S_{i0} = 1$.