



Jul 1st, 12:00 AM

# Statistical analysis of spatial plant patterns under the effect of forest use

I. López

T. Standovár

J. Garay

Z. Varga

M. Gámez

Follow this and additional works at: <https://scholarsarchive.byu.edu/iemssconference>

---

López, I.; Standovár, T.; Garay, J.; Varga, Z.; and Gámez, M., "Statistical analysis of spatial plant patterns under the effect of forest use" (2008). *International Congress on Environmental Modelling and Software*. 153.  
<https://scholarsarchive.byu.edu/iemssconference/2008/all/153>

This Event is brought to you for free and open access by the Civil and Environmental Engineering at BYU ScholarsArchive. It has been accepted for inclusion in International Congress on Environmental Modelling and Software by an authorized administrator of BYU ScholarsArchive. For more information, please contact [scholarsarchive@byu.edu](mailto:scholarsarchive@byu.edu), [ellen\\_amatangelo@byu.edu](mailto:ellen_amatangelo@byu.edu).

# Statistical analysis of spatial plant patterns under the effect of forest use

**I. López<sup>a</sup>, T. Standovár<sup>b</sup>, J. Garay<sup>b</sup>, Z. Varga<sup>c</sup> and M. Gámez<sup>a</sup>**

<sup>a</sup> Department of Statistics and Applied Mathematics, University of Almería. La Cañada de San Urbano 04120 Almería, Spain ([milopez@ual.es](mailto:milopez@ual.es), [mgamez@ual.es](mailto:mgamez@ual.es))

<sup>b</sup> Department of Plant Taxonomy and Ecology, L. Eotvos University, H-1117 Budapest, Pazmany Peter setany 1/c., Hungary  
([garayj@ludens.elte.hu](mailto:garayj@ludens.elte.hu), [standy@ludens.elte.hu](mailto:standy@ludens.elte.hu))

<sup>c</sup> Institute of Mathematics and Informatics, Szent István University, Péter K. u. 1., H-2103 Godollo, Hungary. ([Varga.Zoltan@gek.szie.hu](mailto:Varga.Zoltan@gek.szie.hu))

**Abstract:** The analysis of the consequences of land use (in particular forest use) may be considered as a partial step towards an integrated modelling of a land system. Let us consider a forest territory, where a gap-cut is made, and after a given time period the eventual change in the spatial distribution of undergrowth plants and tree seedlings is to be detected (see Mihók et al., 2005 and Gálhidy et al., 2006). If floristic data are collected along a line transect, we can try to detect the change in the plant distributions along the transect, the so-called *change-point*, and see whether this occurs at the geometric frontier of the human intervention.

The problem, on a theoretical level, can be addressed using the methodology of *change-point analysis* which is a technically involved branch of mathematical statistics (see e.g. Brodsky and Darkhovsky, 1993; Csörgö and Horváth, 1997), widely used to explore the possible temporal or spatial structure of local homogeneity from collected data. (The main application fields of change-point analysis include meteorology, hydrology, or environmental studies, economy, quality control in industry, biology and medicine.) In this paper we propose a practical, operative approach, using only technique of classical statistics.

In our case, given a plant species, along a line transect quadrats have been located and in each quadrat the individuals have been counted. We consider these data as samples of two distributions of the same type but with different parameters, separated by a change-point  $K$ . Based on the maximum likelihood approach, an algorithm is given to estimate  $K$ .

Since the distribution of the change-point estimate is not known, as a substitute of its confidence interval, the so-called *change-interval* will be calculated, using an adaptation of the *bootstrap* method. (For this widely applied simulation method see e.g. Efron and Tibshirani, 1993, a justification of the use of bootstrap in this case can be found in Ferger, 1993.) The implementation of the above algorithms was realized with the application of the statistical software "R". As an illustration, for a concrete plant species, the maximum likelihood estimation of the change-point and the calculation of the above mentioned change-interval will be presented.

**Keywords:** forest use, plant patches, change-point

## 1. INTRODUCTION

The presence of spatial patterns is commonly detected in vegetation studies. Differences in site conditions, outcome of interspecific interaction, growth characteristics of dominant species (e.g. clonality) can all play their roles in the development of these patterns depending on the spatial scale considered. Deliberate management activities can also change these vegetation patterns. There is a huge variety of methods that have been developed for pattern detection, ranging from analyzing spatial point patterns of individuals of a single species, to applying multivariate techniques like numerical classification (cluster

analysis) to distinguish compositionally more or less homogeneous vegetation patches (e.g. Grieg-Smith [1983], Jongman et al. [1987], Podani [2000]).

Here we apply a new approach for detecting spatial patterns of plant distribution.

## 2. STATISTICAL ANALYSIS OF HUMAN EFFECTS ON A FOREST

In forests, human management is aimed at only a few – though dominant – components (trees, wild game species) of the whole ecosystem. However, nowadays an increasing attention is paid to the loss of biodiversity. As a consequence there is a need to assess the effects of different land use activities – in our case forest management – on original biodiversity. Experience shows that not all plant species show a clear, easily detectable reaction to management activities, even if they create steep gradients (like the opening of a small canopy gap in an old forest). Often one can only find difference in the distribution of a species among patches of different quality. Consequently, new methods – capable of detecting such minor changes – could be used to detect minor, not readily detectable causes of human management.

There is a very wide range of applications of statistical methods of change-point analysis in ecology (see Becker et al. [2007], Schleip et al. [2006], Reed [2000] and their references). Our method not only provides an estimation for the location of a change-point and the different distributions laying in different patches, but also provides a so-called *change-interval* (C.I. for brevity) which localizes the distribution change with a high probability level. This interval can be considered as an estimate of a transient zone between patches. The latter has a particular importance in plant ecology since the change between patches usually is not point-like. In a transient zone there may be a mix of two patches, or a special plant composition. Our aim is to study case when the transient zone is small and contains only a mix of the patches.

## 3. THE FIELD EXPERIMENT

To test the applicability of our method, we used data that were collected in the framework of a study aimed at investigating the effects of canopy gap size on the resulting spatial distributions of key abiotic environmental variables (light and soil moisture) in gaps, and at studying how light and soil moisture affect the abundance and distribution of herb layer species. The study site is located in the Börzsöny Mountains, northern Hungary (47.9° N, 18.9° E). Mean annual temperature is 8

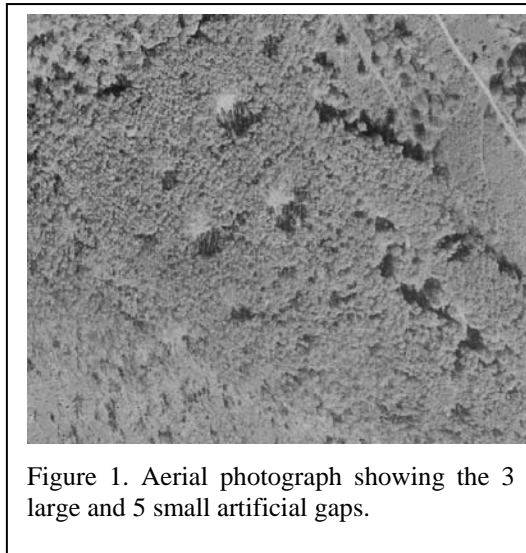


Figure 1. Aerial photograph showing the 3 large and 5 small artificial gaps.

°C, mean monthly temperature is -3.5 °C and 18 °C in January and in July, respectively. Annual precipitation is 700–800 mm. Bedrock is andesite, on which medium deep brown forest soil has developed. The study area is located at 540–610 m elevation, on a relatively steep east-northeast facing slope, that is covered by an almost pure stand of European beech (*Fagus sylvatica* L.). Average tree height is 25 m, mean diameter at breast height is 30 cm. Detailed site description is given in Gálhidy et al. (2006).

The selected stand was a good representative of even-aged, mature (86 year old), and dense forests – typical products of the common contemporary silvicultural system (uniform shelter-wood) [Matthews 1991].

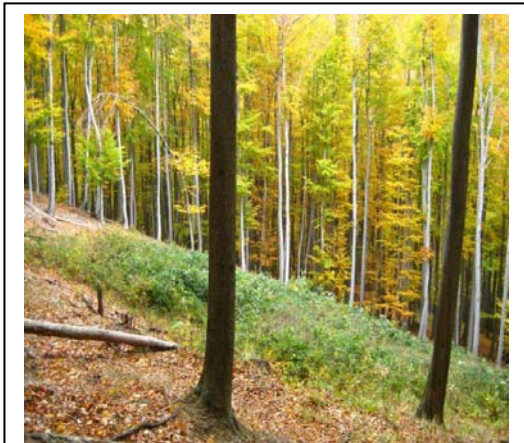


Figure 2. The photograph take in a large gap in late October 2007 (© László Gálhidy)

Because of the dense tree canopy, understorey vegetation was extremely sparse before opening the experimental gaps. Three large gaps (the proportion of tree height of surrounding stand (H) to gap diameter (D) was 1:1.5) and five small gaps (H:D was 1:0.5) were created in February 2001 (Figure 1). We used a systematic sampling design with 5-meter grid resolution and 1x1 m quadrats. As Figure 2 shows, each large gap contained 123 quadrats, whereas small gaps contained 64 quadrats each. Vegetation data were recorded on seven occasions (in September 2000 –

before gap creation – May/ September 2001, and May/September 2002, August 2004 and 2006). On each occasion we determined the cover of each herbaceous species using visual estimation in each quadrat. Among other environmental variable, light conditions were studied in each quadrat, so we could reliably decide if a quadrat was in a gap or non-gap environment.

For the present study we used the data of one species, bramble (*Rubus fruticosus* L.) collected in one of the large gaps in 2006, containing 25 quadrats, i.e. in the seventh growing season after the artificial gaps had been opened, represented in Figures 1 and 2.

#### 4. MODEL DESCRIPTION, ALGORITHMS AND RESULTS

##### 4.1 Model Description

We fix  $0 << K << N$  and suppose that the number of plants in quadrats 1, 2, 3, ..., K are independent random variables with the same discrete probability distribution

$$\xi: \begin{cases} 0 & 1 & 2 & \dots & k \\ p_0 & p_1 & p_2 & \dots & p_k \end{cases}$$

whereas the number of plants in quadrats K+1, K+2, K+3, ..., N are independent random variables with the same discrete probability distribution

$$\eta: \begin{cases} 0 & 1 & 2 & \dots & l \\ q_0 & q_1 & q_2 & \dots & q_l \end{cases}$$

1	2	...	K-1	K	K+1	K+2	...	N
$\xi$	$\xi$	...	$\xi$	$\xi$	$\eta$	$\eta$	...	$\eta$

First, from a given sample vector  $X:=(x_1, x_2, \dots, x_N)$ , for each possible K, we estimate distributions of  $\xi$  and  $\eta$ , and the probability of “realization” of the given sample. Then, from the possible values of K we obtain the required estimate for K, applying the maximum likelihood approach.

#### 4.2 Estimation of distributions $\xi$ and $\eta$

For given  $1 \leq K \leq N-1$ , a possibility to estimate  $\xi$  in terms of relative frequencies may be the following: Let

$$k = \max_{j=1, \dots, K} x_j,$$

and for each  $i = 0, 1, \dots, k$ , we define the probability that the variable  $\xi$  takes each of its possible values:

$$\widehat{p}_{K,i} = P(\xi = i) = \frac{\text{number of indices } j = 1, 2, \dots, K \text{ with } x_j = i}{K},$$

providing an estimate for the distribution of  $\xi$ .

In analogous way we estimate the probability distribution of  $\eta$ : let

$$l = \max_{r=K+1, K+2, \dots, N-1} x_r,$$

and for each  $i = 0, 1, \dots, l$ , we define

$$\widehat{q}_{K,i} = P(\eta = i) = \frac{\text{number of indices } r = K+1, \dots, N \text{ with } x_r = i}{N - K}.$$

Let  $P_K$  be the probability of “realization” of the sample  $X := (x_1, x_2, \dots, x_N)$ , calculated with the above estimated probabilities:

$$P_K := \left( \prod_{i=1}^K \widehat{p}_{K,x_i} \right) \left( \prod_{s=K+1}^N \widehat{q}_{K,x_s} \right),$$

considered the “goodness” of  $K$ . Based on the given sample  $X$ , our purpose is to find a  $K$  which maximizes  $P_K$ , providing the “best” (i.e. the “most probable”) value of  $K$ . We shall deal with this in the next section.

#### 4.3 Algorithms

##### Algorithm 1 (Estimation of the change-point $K$ ):

1. Introduce sample  $X$ .  $N := \text{Size}(X)$ .
2. FOR  $K=1$  until  $N-1$ :
  - a) Calculate:  $\widehat{p}_{K,i}$  and  $\widehat{q}_{K,i}$ , for each  $i$ .
  - b) Calculate:  $\text{Log } P_K = \sum_{i=1}^K \text{Log } \widehat{p}_{K,x_i} + \sum_{s=K+1}^N \text{Log } \widehat{q}_{K,x_s}$ .  
(Logarithm is introduced to avoid too small probability values.)
  3.  $\text{LogProbSample} := (\text{Log}P_1, \dots, \text{Log}P_{N-1})$ .
4.  $\text{Estimate}K :=$  Position  $K$  with maximum value among the coordinates of  $\text{LogProbSample}$ .
5. Return  $\text{Estimate}K$ .

To find a change-interval for  $K$  we elaborate a resampling method based on the known Bootstrap, but with certain modifications in the choice of the elements of each sample of the simulations, in order to fit the method to our problem. The original sample is divided in two homogenous parts, such that the order of the elements of the new samples is important, since, by the linear arrangement, we must not mix *all elements* in a random way. The

generated samples must keep the particularity of having two homogenous parts. The process to follow is explained below:

**Algorithm 2 (Calculation of a 90% level Change-Interval):**

1. Introduce the sample  $X:=(x_1,x_2,\dots,x_N)$ .  $N:= \text{Size}(X)$ .
2. FOR  $K=1$  until  $N-1$ 
  - a) Calculate a weight for each  $K$ :

$$W_K = \left( \prod_{i=1}^K 10 \widehat{p}_{K,x_i} \right) \left( \prod_{s=K+1}^N 10 \widehat{q}_{K,x_s} \right)$$

- b) Normalize the weights (and denote them by  $WN_K$ ).

- c) FOR  $L=1$  until  $m$  (we generate  $m$  samples for each  $K$ ):

c1) Generate  $K$  random numbers  $\{u_1, \dots, u_K\}$  of a discrete uniform distribution  $U[1, K]$ , and  $N-K$  random numbers  $\{u_{K+1}, \dots, u_N\}$  of distribution  $U[K+1, N]$ .

c2) We generate each sample with two homogenous zones, selecting the elements of the original sample according to the random positions obtained in c1) for both zones:

$$X_L := (x_{u_1}, \dots, x_{u_K}, x_{u_{K+1}}, \dots, x_{u_N}).$$

c3) We apply Algorithm 1 to the sample  $X_L$ , to obtain an estimate  $K_L$  for the change-point.

d) From the obtained values  $K_1, \dots, K_m$ , ( $m$  large enough) we calculate a distribution  $d_K$  of the change-point, for each fixed  $K$ .

3. We combine all these new distributions to obtain a unique distribution  $\sum_K WN_K d_K$  for  $K$ , for which we calculate the 90%-level change-interval, with percentile 5 and percentile 95 as extremes.

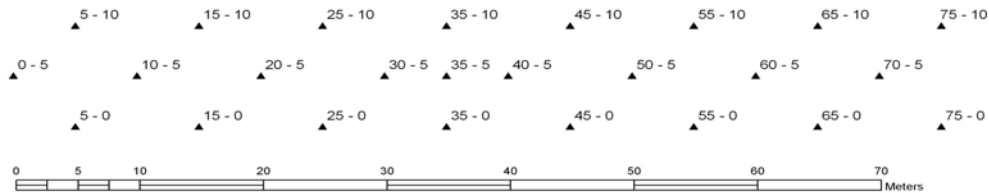
If we have a small amount of data, we can increase the number of data in the following way: we uniformly divide each quadrat of the linear transect into 100 small quadrats along a straight line. In each small quadrat the species will be present (value 1) or not (value 0). Let  $w$  be the number of small quadrats where the species is present. Then, using the statistical software “R” we generate randomly  $w$  values of a discrete uniform distribution between 1 and 100. These  $w$  values will indicate the positions of the small quadrats with one plant, and the remaining  $100-w$  will be the small quadrats with no plant. Therefore, in these 100 small quadrats we represent a  $w\%$  presence of the species. We denote this new data vector by  $S$ . Now, however, we may have too many data for a reasonable run time for the calculations. To reduce them, we sum the values of each 10 consecutive quadrats, and denote this new data vector by  $Z$ . We carry out this in the following:

**Algorithm 3:**

1. Introduce the original sample  $X = (x_1, x_2, \dots, x_N)$ .  $N := \text{Size}(X)$ .
2. We increase its size to  $100N$ , then  $X$  changes to a vector  $S = (s_1, s_2, \dots, s_{100 \cdot N})$  of 1s and 0s, and the frequency of 1s, uniformly placed in a random way between the positions  $(i - 1) \cdot 100 + 1, i \cdot 100$  in  $S$ , will be  $x_i$ .
3. We sum every 10 values of  $S$  obtaining a sample vector  $Z = (z_1, z_2, \dots, z_{10N})$ , to which we apply Algorithm 1 and Algorithm 2 (with  $m = 100$ ), obtaining the estimate for  $K$  and the 90%-level change-interval.

**4.4 The linea**

Consider the species *Rubus fruticosus* with data of 2006, taken from the following area, and described by Cartesian coordinates “X-Y”:



Below, in the first three columns we present the data for the species at each location with coordinates  $X$  and  $Y$ . We know that the distribution at the centre of this area is different from those observed at the extremes. The change of distribution is observed around  $X=25$  on the left, and around  $X=55$  on the right. In order to estimate this change-point we distribute the data along a straight line in the way shown in the 5<sup>th</sup> and 6<sup>th</sup> columns. The new data set is given in the last column:

X	Y	RUB_FRU_06	K (new order)	Data taken from		
				X	Y	RUB_FRU
0	5	0,00	1	0	5	0
5	0	0,00	2	75	0	0
5	10	0,00	3	75	10	0
10	5	0,00	4	5	10	0
15	0	0,00	5	5	0	0
15	10	0,00	6	70	5	0
20	5	0,00	7	10	5	0
25	0	5,00	8	65	10	0
25	10	25,00	9	65	0	0
30	5	15,00	10	15	0	0
35	0	40,00	11	15	10	0
35	5	35,00	12	60	5	0
35	10	0,00	13	20	5	0
40	5	45,00	14	55	10	0
45	0	5,00	15	55	0	0
45	10	25,00	16	25	0	5
50	5	25,00	17	25	10	25
55	0	0,00	18	50	5	25
55	10	0,00	19	30	5	15
60	5	0,00	20	45	0	5
65	0	0,00	21	45	10	25
65	10	0,00	22	35	10	0
70	5	0,00	23	35	0	40
75	0	0,00	24	35	5	35
75	10	0,00	25	40	5	45





- Brodsky, B.E. and Darkhovsky, B.S., *Nonparametric Methods in Change-point Problems*. Kluwer Academic Publishers, The Netherlands, 1993.
- Csörgö, M. and Horváth, L., *Limit Theorems in Change-point Analysis*. Wiley, Chichester, 1997.
- Efron, B. and Tibshirani, R. *An Introduction to the Bootstrap*. Chapman & Hall, New York, 184-188, 1993.
- Ferger, D. Asymptotic distribution theory of change-point estimators and confidence intervals based on bootstrap approximation. *Mathematical Methods of Statistics*, Vol. 3, No. 4, 362-378, 1993.
- Gálhidy, L., Mihók, B., Hagyó, A., Rajkai, K. and Standovár, T. 2006. Effects of gap size and associated changes in light and soil moisture on the understorey vegetation of a temperate deciduous forest. *Plant Ecology* 183:133-145.
- Greig-Smith P. 1983. *Quantitative Plant Ecology*, 3rd edition. Blackwell, Oxford.
- Jongman R. H. G., C. J. F. ter Braak, and O. F. R. van Tongeren. (eds.) 1987. *Data Analysis in Community and Landscape Ecology*. Pudoc, Wageningen
- Matthews, J.D. 1991 *Silvicultural Systems*. Calderon Press, Oxford.
- Mihók B., Gálhidy L., Kelemen K., and Standovár T. 2005. Study of Gap-phase Regeneration in a Managed Beech Forest: Relations between Tree Regeneration and Light, Substrate Features and Cover of Ground Vegetation. *Acta Silv. Lign. Hung.*, 1:25-38.
- Podani, J. 2000. *Introduction to the Exploration of Multivariate Biological Data*. Backhuys Publishers, Leiden.
- Reed, W.J reconstruction the history of forest fire frequency: Identifying hazard rate change points the Bayesian information criterion. *Can. Jour. Stat.* **28**, 353-365. 2000
- Schleip, C. Menzel, A., Estrella, N., Dose, V. The use of Bayesian analysis to detect recent change in phenological events throughout of years. *Agri. Forest Meteor.* **141**, 179-191 . 2006.