

**Jose Raventós, Thorsten Wiegand, and Martin de Luis. Year. Evidence for the spatial segregation hypothesis: a test with nine-year survivorship data in a Mediterranean fire-prone shrubland. *Ecology* 91:2110-2120**

**Appendix A:** Details on methods of spatial pattern analysis.

The pair-correlation function  $g(r)$  is based on the distances between all pairs of points of a given point pattern. The estimator used for the pair correlation function is shown in equation A1. In essence, the univariate quantity  $\lambda g(r)$  gives the expected density of neighbours in a ring of radius  $r$  and ring-width of  $dr$  around an “average” point of the pattern, where  $\lambda$  is the intensity of the pattern (i.e., number of points of the pattern divided by the area of the study plot). The pair-correlation function isolates specific distance classes which makes it especially suitable for exploratory data analysis (Wiegand and Moloney 2004, Perry et al. 2006, Illian et al. 2008). The bivariate pair-correlation function for patterns comprising two types of points [i.e., dead (1) and surviving plants (2)] follows naturally; the bivariate quantity  $\lambda_2 g_{12}(r)$  gives the expected density of type 2 neighbours in a ring of radius  $r$  (and ring-width  $dr$ ) around an “average” type 1 individual, where  $\lambda_2$  is the intensity of the pattern of type 2 points.

#### *Test statistics*

For each year we know if a given plant was dead or alive. To identify the spatial structures in the process that assigned a label “mortality” to the plants the framework of “marked point patterns” is required (Goreaud and Pelissier 2003). The fundamental question is to find out if the process that distributed the labels was a random process. A suitable null model for this is “random labelling”. We used a Monte Carlo implementation of random labelling that involves random re-sampling of sets of  $n_1$  dead

individuals from the total of  $(n_1 + n_2)$  dead and surviving individuals, i.e., we randomly shuffle the label “dead” among the plants.

While the concept of random labelling is easy and intuitive, the test statistics needed to test departures from random labelling are slightly more complex because they need to be adapted to the specific question asked. Conventional random labelling analyzes spatial structure of the two patterns of dead and surviving plants conditionally on the locations of dead and surviving plants (Goreaud and Pelissier 2003, De la Cruz et al. 2008). In this case random labelling implies that  $g_{1+2,1+2}(r) = g_{11}(r) = g_{22}(r) = g_{12}(r) = g_{21}(r)$  where 1 symbolizes dead individuals, 2 surviving, and 1+2 symbolizes the joined pattern of dead and surviving individuals. This convenient property of random labelling allows for construction of different test statistics based on pair correlation functions. We use for this purpose the three specific test statistics  $g_{11}(r)$ ,  $g_{12}(r)$ ,  $g_{1,1+2} - g_{2,1+2}$ , and  $g_{1+2,1}$ . However, departures from random labelling may be influenced by a third pattern (e.g., De la Cruz et al. 2008; Biganzoli et al. 2009; Xu et al. 2009). Therefore we may also use test statistics that evaluate the probability of mortality in dependence on the distance from plants of a third pattern. Each of these test statistics we use evaluates a different biological effect of the spatial interactions that determine mortality (see below).

*Test of influence of heterospecific plants on mortality with  $p_{h,l}$ .*—Our first hypothesis requires quantification of the impact of interspecific plants on mortality of a given focal species. We ask if, in the probability of mortality of plants, the focal species depends on the distance from surviving heterospecific plants. The appropriate null model for this question is again random labelling, but the test statistic needs to consider additionally the impact of heterospecific plants. The software Programita (Wiegand and Moloney 2004) offers an appropriate test statistic for such a situation (Biganzoli et al.

2009; Xu et al. 2009) which was also independently proposed by De la Cruz et al. (2008). We have heterospecific plants (subscript h), dead plants of the focal species (subscript 1) and surviving plants of the focal species (subscript 2). The test statistic which estimates the probability of mortality of plants of the focal species in dependence on the distance  $r$  from heterospecific plants (subscript h) is given by

$$p_{h,1}(r) = \frac{\lambda_1}{(\lambda_1 + \lambda_2)} \frac{g_{h,1}(r)}{g_{h,1+2}(r)} \quad (\text{A.1})$$

where  $\lambda_1 + \lambda_2$  and  $\lambda_1$  are the intensities of plants of the focal species and of the dead plants of the focal species, respectively, and  $g_{h,1+2}(r)$  and  $g_{h,1}(r)$  are the bivariate pair correlation functions measuring the intensity normalized neighborhood density of surviving and dead focal plants (1+2) and dead focal plants (1), respectively, around heterospecific plants (h).

The expectation of this test statistic under random labelling is the overall probability of mortality, i.e., number of dead plants divided by number of plants. In the case of negative interactions exerted by plants of other species at distance  $r$  (i.e., competition) we expect a higher probability of mortality, i.e.,  $p_{h,1}(r) > \lambda_1/(\lambda_1 + \lambda_2)$ , whereas positive interactions would be indicated by a lower probability of mortality in the proximity of heterospecific plants, i.e.,  $p_{h,1}(r) < \lambda_1/(\lambda_1 + \lambda_2)$ .

Note that this test statistic is formally analogous to mark-connection functions (Getzin et al. 2008, Illian et al. 2008) because we normalize with the expectation under random labelling, thereby removing the effect of the spatial structure of the pre-mortality pattern from the test statistic. The test statistic  $p_{h,i}$  therefore shows the effects we are interested in (i.e., influence of heterospecific plants) much more clearly than by

simply using the pair correlation function  $g_{h,1}(r)$  or the L-function as done in De la Cruz et al. (2008).

Testing of hypothesis H1 requires not only to show that proximity of heterospecific plants does not influence the probability of mortality but also to show that proximity of conspecific plants does influence the probability of mortality. Therefore we used the conspecific counterpart of the a test statistic  $p_{h,1}(r)$ , which is  $p_{1+2,1}(r)$  where 1+2 symbolizes surviving and dead conspecifics, to test for effects of the proximity of heterospecific plants on the probability of mortality. We estimated the  $p_{1+2,1}(r)$  analogously to Eq. A.1 as :

$$p_{1+2,1}(r) = \frac{\lambda_1}{(\lambda_1 + \lambda_2)} \frac{g_{1+2,1}(r)}{g_{1+2,1+2}(r)} \quad (\text{A.2})$$

*Test of aggregation of dead plants with  $g_{11}$ .*—We used the test statistic  $g_{11}(r)$  (i.e., univariate random labelling; Wiegand and Moloney 2004) primarily to test hypothesis H2, but also used it for hypotheses H1 and H4. Under scramble competition (H2), we expect that clumps of interacting seedlings should die together, thus dead seedlings should be spatially aggregated (e.g., Kenkel 1998). The test statistic  $g_{11}(r)$  is especially tailored to detect clustering of dead individuals: if  $g_{11}(r)$  shows a positive departure from the random labelling null model, dead plants are significantly clustered at scale  $r$ , conditionally on the joined pattern of dead and surviving individuals.

*Test of attraction vs. segregation of dead and surviving plants.*—We used the test statistic  $g_{12}(r)$  (Goreaud and Pelissier 2003) primarily to test hypothesis H2, but used it also for hypotheses H1 and H4. Under two-sided scramble competition (H2), negative interactions would be most pronounced in areas of higher seedling density,

leading eventually to clumps of dead seedlings which are spatially segregated from surviving seedlings. Conversely, one-sided contest competition leads to a situation where suppressed “losers” are located near its superior rivals. Thus, in the language of point pattern analysis, dead plants will be “attracted” by surviving plants (Kenkel 1988). The test statistic  $g_{12}(r)$  is tailored to detect correlation between dead and surviving trees (Goreaud and Pelissier 2003). Attraction (or segregation) between dead and surviving plants occurs if  $g_{12}(r)$  shows positive (or negative) departures from the random labelling null model.

*Test of density dependent mortality.*— We developed the third test statistic  $g_{1,1+2} - g_{2,1+2}$  specifically to provide a direct test of hypothesis H3, i.e., to detect density dependent effects in mortality (Yu et al. 2009). This test statistic compares the density of dead and surviving plants (i.e., 1+2) around dead plants (i.e., pattern 1) with the density of dead and surviving plants (i.e., 1+2) around surviving plants (i.e., pattern 2). The expected value of this test statistic is zero under random labelling, but under density dependent mortality, dead plants would occur preferably in areas with high pre-mortality densities, i.e.,  $g_{1,1+2} > g_{2,1+2}$ . Conversely, under positive density dependent mortality (i.e., facilitation), we expect surviving plants to have more neighbors (i.e.,  $g_{1,1+2} < g_{2,1+2}$ ). Note that the test statistic  $g_{1,1+2} - g_{2,1+2}$  is in structure somewhat similar to the test statistic  $p_{1+2,1}(r)$  because both evaluate effects of all seedlings on mortality. However, we could not derive a direct analytical relationship between the test statistics and used therefore both.

### *Simulation envelopes*

We performed for each plot and year 999 Monte Carlo simulations of the random labelling null model to generate approximately 95% simulation envelopes. For each test

statistic we used the 25th-lowest and 25th-highest value at a given scale  $r$  as simulation envelopes. However, the simulation envelopes cannot be interpreted as confidence intervals. Due to simultaneous inference (i.e., we test at several spatial scales,  $r$ , simultaneously), Type I error may occur if the value of the test statistic is close to a simulation envelope (i.e., the null model may be rejected even if it is true; Loosmore and Ford 2006).

To assess the overall fit of the random labelling null model for a given test statistic we used a Goodness-of-Fit (GoF) test with a test statistic  $u_i$  which represents the total squared deviation between the observed pattern and the theoretical result across the distances of interest (in our case the scales 0-25cm). The  $u_i$  were calculated for the observed data ( $i = 0$ ) and for the data created by the  $i = 1, \dots, 999$  simulations of the null model, and if the rank of  $u_0$  among all  $u_i$  was larger than 950 (990), the data showed a significant departure from the null model with an error rate  $\alpha = 0.05$  ( $\alpha = 0.01$ ) (Diggle 2003, Loosmore and Ford 2006).

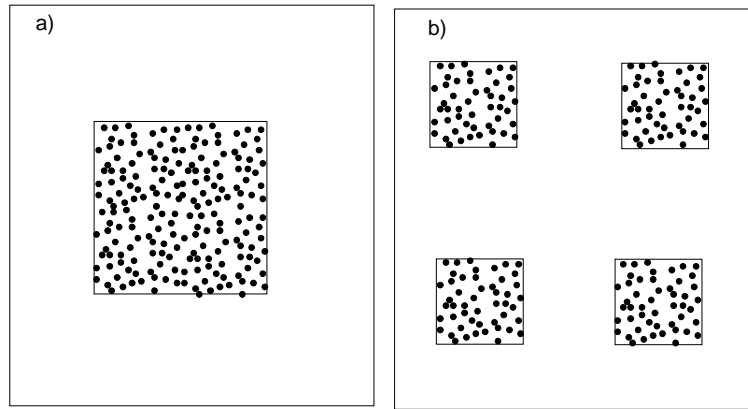


FIG. A1. One large plot (a) vs. several small plots (b).

*Combining the results of several point pattern analyses/subplots into one average test statistic*

Mapping a plant community for the purpose of point pattern analysis involves a trade-off between one large plot and several smaller plots. One large plot has the advantage that edge effects are lower (i.e., there are less plants at the border of the plot which suffer from the influence of unknown plants outside the plot), but one larger plot may not sufficiently represent the typical conditions occurring at the study site.

Point pattern analysis can deal with replicate analysis of several replicate plots. In this case, the resulting test statistics of the individual replicate plots can be combined into average test statistics (Diggle 2003: page 123, Illian et al. 2008: page 263). Average test statistics are also an effective way of summarizing the results of several replicate plots. This is of particular interest if the number of points in each replicate plot is relatively low. In this case, the simulation envelopes of individual analyses would become wide, but combining the data of several replicate plots into average test statistics increases the sample size and thus narrows the simulation envelopes. Note that combining the results may even yield significant effects for the combined plots if the results for the individual plots are not significant. This is a consequence of the increased sample size.

The particular formula to combine results from several replicate plots depends on the estimators of the test statistics used. The grid based software *Programita* (Wiegand and Moloney 2004) allows analysis of study areas of irregular shape and is thus able to analyse several plots as shown in Fig. A1 directly, if they are placed sufficiently apart (i.e., distances among plots are larger than the largest scale  $r$  analyzed). Following the notation in Wiegand and Moloney (2004) (their equation 11),

the numerical estimator of the O-ring statistic  $O_{12}(r) = \lambda_2 g_{12}(r)$  implemented in Programita is calculated as:

$$\hat{O}_{12}^w(r) = \frac{\frac{1}{n_1} \sum_{i=1}^{n_1} \mathbf{Points}_2[R_{1,i}^w(r)]}{\frac{1}{n_1} \sum_{i=1}^{n_1} \mathbf{Area}[R_{1,i}^w(r)]} \quad (\text{A.3})$$

where  $n_1$  is the number of points of pattern 1,  $R_{1,i}^w(r)$  is the ring with radius  $r$  and width  $w$  centred in the  $i$ th point of pattern 1,  $\mathbf{Points}_2[X]$  counts the points of pattern 2 in a region  $X$ , and the operator  $\mathbf{Area}[X]$  determines the area of the region  $X$ .

Equation A.3 applies for both, the one large plot (Fig. A1a) and the several small plots (Fig. A1b). If several subplots are sufficiently separated in space as shown in Fig. A1b, no point pair appears in the estimator where points are from different subplots. In this case, the estimator Eq. A.3 can be rewritten, by simply grouping the contributions from the  $M$  different subplots:

$$\hat{O}_{12}^w(r) = \frac{\sum_{i^1=1}^{n_1^1} \mathbf{Points}_2[R_{1,i^1}^w(r)] + \dots + \sum_{i^M=1}^{n_1^M} \mathbf{Points}_2[R_{1,i^M}^w(r)]}{\sum_{i^1=1}^{n_1^1} \mathbf{Area}[R_{1,i^1}^w(r)] + \dots + \sum_{i^M=1}^{n_1^M} \mathbf{Area}[R_{1,i^M}^w(r)]} \quad (\text{A.4})$$

which is equivalent to

$$\hat{O}_{12}^w(r) = \frac{\left( \frac{n_1^1}{N} \left( \frac{1}{n_1^1} \sum_{i^1=1}^{n_1^1} \mathbf{Points}_2[R_{1,i^1}^w(r)] \right) + \dots + \frac{n_1^M}{N} \left( \frac{1}{n_1^M} \sum_{i^M=1}^{n_1^M} \mathbf{Points}_2[R_{1,i^M}^w(r)] \right) \right)}{\left( \frac{n_1^1}{N} \left( \frac{1}{n_1^1} \sum_{i^1=1}^{n_1^1} \mathbf{Area}[R_{1,i^1}^w(r)] \right) + \dots + \frac{n_1^M}{N} \left( \frac{1}{n_1^M} \sum_{i^M=1}^{n_1^M} \mathbf{Area}[R_{1,i^M}^w(r)] \right) \right)} \quad (\text{A.5})$$



where  $i^j$  is the  $i$ th point of pattern 1 and replicate  $j$ ,  $n_1^j$  is the number of points of pattern 1 and analysis/replicate  $j$ , and  $N = \sum_j n_1^j$  is the total number of points of pattern 1 in all analyses. Note that the estimator Eq. A.5 sums up for each subplot the average intensity-weighted number of points of pattern 2 in rings with radius  $r$  and width  $w$  centred at point of pattern 1 and the intensity weighted area of these rings:

$$\hat{O}_{12}^w(r) = \frac{\left( \frac{n_1^1}{N} (\overline{\mathbf{Points}_2[R_{1,i^1}^w(r)]}) + \dots + \frac{n_1^M}{N} (\overline{\mathbf{Points}_2[R_{1,i^M}^w(r)]}) \right)}{\left( \frac{n_1^1}{N} (\overline{\mathbf{Area}[R_{1,i^1}^w(r)]}) + \dots + \frac{n_1^M}{N} (\overline{\mathbf{Area}[R_{1,i^M}^w(r)]}) \right)} \quad (\text{A.6})$$

The univariate estimators of  $O(r)$  are calculated in a manner analogous to the bivariate function by setting pattern 1 equal to pattern 2. In this case, however, the focal points of the circles are not counted:

$$N = \sum_{j=1}^M (n^j - 1). \quad (\text{A.7})$$

In practice, we analyzed the data for each subplot separately and performed the simulations of the random labelling null model separately within each subplot and combined the results of the subplots using the estimator Eq. A.5. Note that we applied Eq. A.5 for the observed data and the simulations of the null model. Alternatively, we could use the feature of *Programita* to analyze irregularly shaped study regions and analyze all subplots at once, thereby applying the random labelling null model over all subplots (e.g., Fig. A1b). However, we decided to treat each plot separately in the null

model to better account for smaller differences between the subplots. This treatment is analogous to homogeneous vs. heterogeneous Poisson processes.

The estimator Eq. A.5 is not only an effective way of combining the results of several replicate plots, it can also be used to test in plant communities selectively for the impact of intraspecific interactions on mortality. Random labelling can be applied to the data on the survival/mortality of individual species or to plants of all species together. However, random labelling using the data of all four species implies that the estimator of the pair correlation function is composed of both intraspecific and interspecific point pairs. To study intraspecific effects selectively, the estimator needs to remove all interspecific point pairs. This results in the same estimator as using Eq. A.5 to combine the results of the individual analysis of the four species. In this case, the averages in Eq. A.6 would be interpreted as the average for individual species.

#### LITERATURE CITED

- Biganzoli, F., T. Wiegand and W.B. Batista. 2009. Fire-mediated interactions between shrubs in a South American temperate savannah. *Oikos*, *in press*. DOI: 10.1111/j.1600-0706.2009.17349.x
- De la Cruz, M., R. L. Romao, A. Escudero, and F. T. Maestre. 2008. Where do seedlings go? A spatio-temporal analysis of seedling mortality in a semi-arid gypsophyte. *Ecography* 31:1–11.
- Diggle, P. J. 2003. *Statistical Analysis of Spatial Point Patterns*. Second Edition. Arnold, London, UK.

- Getzin, S., T. Wiegand, K. Wiegand, and F. L. He. 2008. Heterogeneity influences spatial patterns and demographics in forest stands. *Journal of Ecology* 96:807–820.
- Goreaud, F., and R. Pelissier. 2003. Avoiding misinterpretation of biotic interactions with the intertype  $K_{12}$ -function: population independence vs. random labelling hypothesis. *Journal of Vegetation Science* 14:681–692.
- Illian, J., A. Penttinen, H. Stoyan, and D. Stoyan. 2008. Statistical analysis and modelling of spatial point patterns. John Wiley and Sons. Chichester.
- Kenkel, N.C. 1988. Pattern of self-thinning in jack pine: testing the random mortality hypothesis. *Ecology* 69:1017–1024.
- Loosmore N. B., and E. D. Ford. 2006. Statistical inference using the G or K point pattern spatial statistics. *Ecology* 87:1925–1931
- Perry, G. L. W., B. P. Miller, and N. J. Enright. 2006. A comparison of methods for statistical analysis of spatial point patterns in plant ecology. *Plant Ecology* 187:59–82.
- Wiegand, T., and K. A. Moloney. 2004. Rings, circles, and null-models for point pattern analysis in ecology. *Oikos* 104:209–229.
- Xu, X., T. D. Harwood, M. Pautasso, and M. J. Jeger. 2009. Spatio-temporal analysis of an invasive plant pathogen (*Phytophthora ramorum*) in England and Wales. *Ecography* 32:504–516. DOI: 10.1111/j.1600-0587.2008.05597.x
- Yu, H. T. Wiegand, X. Yang, and L. Ci. 2009. The impact of fire on the spatial patterns and density-dependent mortality in a Mongolian pine forest chronosequence in the Hulunbeier sand region, Inner Mongolia, China. *Forest Ecology and Management* 257:2098–2107.

**Table A1.** Main hypothesis, test statistics, biological questions to be tested, and related figures. In all cases we use the random labelling null model but different test statistics to detect departures from random mortality.

Test statistics	Expectation		Biological questions to be tested	Fig.
$p_{h,1}(r) = \frac{\lambda_1}{(\lambda_1 + \lambda_2)} \frac{g_{h,1}(r)}{g_{h,1+2}(r)}$	$p_{h,1}(r) > \lambda_1/(\lambda_1 + \lambda_2)$	Test statistic above simulation envelopes	<b>negative interactions</b> will be exerted <b>from heterospecific plants</b> (= subscript h) or <b>conspecific plants</b> (subscript 1+2) at distance $r$ (i.e., competition) causing a higher probability of mortality (=subscript 1)  <b>positive interactions</b> would be indicated by a lower probability of mortality in the proximity of <b>heterospecific plants</b> (= subscript h) or <b>conspecific plants</b> (subscript 1+2) .	Fig. 1
$p_{1+2,1}(r) = \frac{\lambda_1}{(\lambda_1 + \lambda_2)} \frac{g_{1+2,1}(r)}{g_{1+2,1+2}(r)}$	$p_{h,1}(r) < \lambda_1/(\lambda_1 + \lambda_2)$	Test statistic below simulation envelopes		
$g_{11}(r)$	$g_{1+2,1+2}(r)$	Test statistic above simulation envelopes	<b>Dead individuals</b> (= pattern 1) are <b>aggregated</b> at scale $r$ , conditionally on the positions of dead and surviving individuals	Fig. 2
$g_{12}(r)$	$g_{1+2,1+2}(r)$	Test statistic below simulation envelopes	<b>Dead and surviving individuals are segregated</b> , conditionally on the positions of dead and surviving individuals	Fig. 2
$g_{1,1+2}(r) - g_{2,1+2}(r)$	0	Test statistic above simulation envelopes	Dead individuals are located in areas of higher initial seedlings density. <b>Negative density dependence</b>	Fig. 2
$g_{1,1+2}(r) - g_{2,1+2}(r)$	0	Test statistic below simulation envelopes	Surviving individuals are located in areas of higher initial seedlings density. <b>Positive density dependence</b>	Fig. 2

Hypotheses	Test	Figures
<b>H1:</b> intraspecific interactions are more important than interspecific interactions.	Use random labelling with test statistics $p_{h,1}(r)$ and $p_{1+2,1}(r)$ where <b>heterospecific plants</b> (= subscript h), <b>dead plants</b> of the focal species (= subscript 1) and <b>surviving plants</b> of the focal species (= subscript 2). The $p_{h,1}(r)$ and $p_{1+2,1}(r)$ estimate the probability of mortality of plants of the focal species in dependence on the distance $r$ from heterospecific and conspecific plants, respectively.	Fig 1 (A to H) vs. Fig 1 (I to P)
<b>H2:</b> Competition should show characteristics of the two-sided scramble type (Kenkel 1988)	Use random labelling with test statistics $g_{11}(r)$ and $g_{12}(r)$ : <ul style="list-style-type: none"> <li><math>g_{11}(r)</math>: dead individuals should be aggregated</li> <li><math>g_{12}(r)</math>: dead and surviving individuals should be segregated</li> </ul>	Fig 2 (A to P)
<b>H3:</b> Positive density dependent effects on mortality during the earlier stage of seedling establishment, but negative effects when plants grow to maturity.	Use random labelling with test statistic $g_{1,1+2}(r)$ - $g_{2,1+2}(r)$ <ul style="list-style-type: none"> <li>test for <math>g_{1,1+2}(r)</math>- <math>g_{2,1+2}(r) &lt; 0</math> in first years</li> <li>test for <math>g_{1,1+2}(r)</math>- <math>g_{2,1+2}(r) &gt; 0</math> in later years</li> </ul>	Fig 2 (Q to X)
<b>H4:</b> The simulated extreme rainfall event cause non-specific mortality of seedlings reducing intraspecific competition	Compare results of hypothesis 1 for the two treatments.	Figs 1,2 but Figs. F vs. FR