The "Minimal Area" Problem in Ecology: A Spatial Poisson Process Approach

D. Pfeifer[1], H.-P. Bäumer[2], U. Schleier[3]

[1] FB Mathematik, Institut für Mathematische Stochastik,
Universität Hamburg, D-20146 Hamburg, FRG
[2] Hochschulrechenzentrum, [3] Fachbereich Mathematik,
Carl von Ossietzky Universität, D-26111 Oldenburg, FRG

Summary

Ecologists frequently deal with the problem of finding an appropriate size of the observation area such that all species of a certain community are represented in that area. A *minimal area* is – in a certain sense – the smallest observation area with this property. In this paper we show that modelling the spatial distribution of biological communitites by superpositions of homogeneous Poisson point processes not only allows for a simple explanation of "typical" forms of observed species–area curves, but gives also rise to explicit calculations of the distribution and moments (expectation, variance) of the size of the corresponding minimal area.

Keywords: Species–Area Curve, Sampling, Statistics of Extremes

1   Introduction

A very interesting empirically observed phenomenon which is treated in various papers in the ecological literature is the log–log–linear form of the so–called *species–area curve*; see e.g. Krebs (1985) or Begon, Harper, and Townsend (1990) for a survey. The empirical species–area curve is obtained
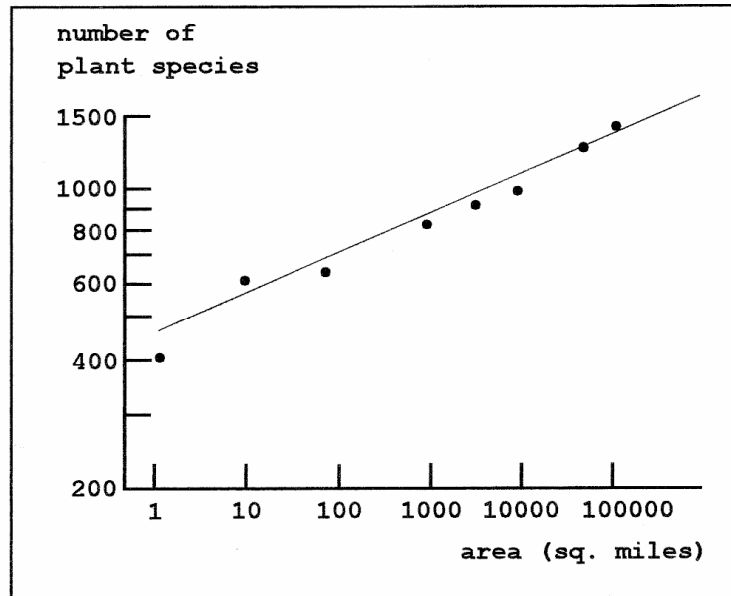
Fig. 1

when the number of observed species is plotted against the area size. If one assumes that the individuals of each species are dispersed in the plane according to some probability distribution and samples are taken from selected areas, each empirically observed species–area curve represents a realization of a certain stochastic process, which is discontinuous, but weakly increasing with area size and approaches in the limit the total number of species when area size grows to infinity. Taking expectations of these realizations, one obtains the [theoretical] species–area curve, which in general is continuous, while maintaining the other two analytical properties (monotonicity and asymptotics) of its empirical counterpart. In practical applications, the empirical species–area curve is usually fitted by an appropriate continuous function, representing the expected theoretical curve. Surprisingly, in many investigations it was found out that when the empirical species–area curve is plotted in a log–log scale, the fit of the empirical curve to a linear function seems most appropriate. The above figure shows the result of such an analysis for the species–area relationship for flowering plants in England, the smallest individual area size being measured in Surrey, the largest representing the whole of England (modified after Williams (1964); see also Krebs (1985), Fig. 24.8). There have been several attemps to explain from theory why such a striking relationship might occur. May (1975) has given a nice explanation by assuming a log–normal distribution for the species abundances. However, as was already pointed out above, the species–area curve must eventually approach the total number of species in the community with increasing area size. In this paper, we shall show that the assumption of homogeneous spatial Poisson processes for the dispersal of individuals of each species in the

plane does not not only give a likewise sufficient and simple explanation for an approximate local log–log–linear species–area curve but also allows for a mathematically rigorous definition of a "minimal area" for finding all or most of the abundant species in a given community. Recently Schleier and van Bernem (1995) have investigated this question using combinatorial arguments to determine the probability of finding a certain selection out of all abundant species, depending on sample size. They also consider similarity–area curves in the sense of Weinberg (1978) who defines a "minimal area" by the degree of similarity between biological characteristics of the sampling area and the total area supporting the whole community. In the present paper, we shall use methods from point process theory and stochastic geometry to show that a minimal area can quite naturally be defined as a certain random set in $\mathbb{R}^2$ with the property that it covers all species present, while any "smaller" subset does not. In particular, it is shown that formulas for the distribution of the size (=Lebesgue measure) of this "minimal area" as well as for its expectation and variance can easily be obtained in the Poisson model. Generalizations of this concept to "minimal areas" which cover not all but only a certain subselection of species are also possible using standard methods from statistics of extremes.

2   The Probabilistic Model

Let $\mathcal{B}$ denote the Borel σ–field over $\mathbb{R}^2$, $\mathcal{K}$ the subcollection of all compact sets, and $\mathfrak{m}$ the corresponding Lebesgue measure. Suppose further that in the observation areas $A \in \mathcal{B}$ with $\mathfrak{m}(A) > 0$ the individuals of s species are distributed according to independent homogeneous Poisson processes ξ_i, $i = 1, \ldots, s$ each, with intensities $\lambda_1, \ldots, \lambda_s > 0$. I.e., for all Borel sets $A \in \mathcal{B}$, the random variables $\xi_i(A)$ – counting the number of individuals of species i abundant in A – are Poisson–distributed with mean $E\xi_i(A) = \lambda_i \mathfrak{m}(A)$, and for pairwise disjoint subsets $A_1, \ldots, A_n \in \mathcal{B}$, the random variables $\xi_i(A_1), \ldots, \xi_i(A_n)$ are *independent*. Note that A might also be an *unbounded* set, e.g. $A = \mathbb{R}^2$. As a consequence of the model, the point process ξ with $\xi(A) = \sum_{i=1}^s \xi_i(A)$ counting the individuals of all species in Borel sets $A \in \mathcal{B}$ also is a Poisson process, with intensity $\lambda = \sum_{i=1}^s \lambda_i$ (for further details, see e.g. Diggle (1983)). The number $Z(A)$ of species present in a Borel set A can be represented as follows: Let

$$Z_i(A) = \begin{cases} 1 & \text{if } \xi_i(A) > 0 \quad \text{[i.e., species } i \text{ is present in } A] \\ 0 & \text{otherwise.} \end{cases} \qquad (1)$$

Then $Z(A) = \sum_{i=1}^s Z_i(A)$, which is a sum of independent 0–1 valued random variables with

$$\begin{aligned} E\big(Z_i(A)\big) &= P\big(Z_i(A) = 1\big) = P\big(\xi_i(A) > 0\big) \\ &= 1 - P\big(\xi_i(A) = 0\big) = 1 - e^{-\lambda_i \mathfrak{m}(A)} \end{aligned} \qquad (2)$$

4

and hence

$$E\big(Z(A)\big) = \sum_{i=1}^{s} E\big(Z_i(A)\big) = \sum_{i=1}^{s} \left(1 - e^{-\lambda_i \, \mathfrak{m}(A)}\right). \tag{3}$$

Clearly, $Z(A) \le s$, with equality being achieved only if all species are represented in A. A *minimal area* in this model now is a suitable set $\Xi \in \mathcal{B}$ such that $Z(\Xi) = s$, and $Z(A) < s$ for any other "smaller" set $A \subset \Xi$. To make such a definition mathematically rigorous, consider the collection $\mathfrak{U} = \{K \in \mathcal{K} \mid \mathfrak{m}(K) = 1\}$ of all compact sets with unit area which are *star–shaped* (i.e., the origin $\mathbf{0} \in K$ and for all points $\mathbf{x}$ in K, the line segment between $\mathbf{0}$ and $\mathbf{x}$ is contained in K). For such a set K, define

$$\rho_K = \inf\{r > 0 \mid Z(rK) = s\}, \quad \Xi_K = \rho_K K \tag{4}$$

where $rK = \{r\mathbf{x} \mid \mathbf{x} \in K\}$ is the set K scaled by the non–negative factor r. Call Ξ_K a *minimal area of type K*. In the language of stochastic geometry, Ξ_K is a so–called *random set* (see e.g. Stoyan, Kendall, and Mecke (1987)), and its size $W_K = \mathfrak{m}(\Xi_K)$ is a random variable whose distribution is independent of the particular shape of K:

$$\begin{aligned} P(W_K \le x) &= P\Big(\bigcap_{i=1}^{s}\{\xi_i(A_x) > 0\}\Big) = P\Big(\bigcap_{i=1}^{s}\{Z_i(A_x) = 1\}\Big) \\ &= \prod_{i=1}^{s}\left(1 - e^{-\lambda_i \mathfrak{m}(A_x)}\right) = \prod_{i=1}^{s}\left(1 - e^{-\lambda_i x}\right) \end{aligned} \tag{5}$$

for all Borel sets A_x such that $\mathfrak{m}(A_x) = x$, $x > 0$. (For formal proofs, see Section 5.) Intuitively, this can be seen as follows: Let $x > 0$ and $A_x = \sqrt{x}K$ be of type K. Then $\mathfrak{m}(A_x) = x\mathfrak{m}(K) = x$ and for the corresponding events, $\{W_K \le x\} = \bigcap_{i=1}^{s}\{Z_i(A_x) = 1\} = \bigcap_{i=1}^{s}\{\xi_i(A_x) > 0\}$. The last equation reflects the idea that when any area A_x of size $\mathfrak{m}(A_x) = x$ already carries all abundant species, then the "minimal area" should be at most this size or smaller. Recall that for homogeneous Poisson processes, the distribution of $\xi_i(A_x)$ is the same for all Borel sets A_x having measure $\mathfrak{m}(A_x) = x$. Likewise, if $\Xi_K(i)$ denotes the minimal area of type K for the point process ξ_i representing the individuals of species i, with scaling factor $\rho_K(i)$, we have

$$\begin{aligned} \rho_K &= \max\{\rho_K(1),\ldots,\rho_K(s)\}, \quad \Xi_K = \bigcup_{i=1}^{s}\Xi_K(i), \\ W_K &= \max\{W_K(1),\ldots,W_K(s)\}. \end{aligned} \tag{6}$$

Note that for a meaningful definition of a minimal area, we cannot simply take "the" smallest Borel set containing all s species because this would with probability one be a random set consisting of exactly s distinct points in $\mathbb{R}^2$

(e.g. the collection of those points from each of the s Poisson processes being closest to the origin, or any other fixed location in the plane). Such a set would clearly always have measure zero and is hence of no worth for any practical application, not to speak of the fact that in order to determine such a set the locations of all points must be known, hence complete knowledge of the whole community would be necessary.

The definition of a minimal area as above also makes sense from an ecological point of view. For instance, if data are usually obtained from quadrat counts then K is a square, and if we assume that the location of the observer in the plane is defined to be the origin $\mathbf{0} \in \mathbb{R}^2$ – which we can do w.l.o.g. due to the fact that homogeneous Poisson processes are distributionally translation and rotation invariant – the minimal area Ξ_K can be considered to be the physically smallest square with center $\mathbf{0}$, say, which contains all abundant species. Note that Ξ_K varies with the location of the observer, which again reflects that actually Ξ_K is a random set. Therefore, characteristic parameters like mean and variance of the minimal area size $\mathfrak{m}(\Xi_K)$ are of specific interest.

3 Calculating Moments for the Minimal Area Size Distribution

The main mathematical facts from Section 2 (see also Section 5, Theorem 2) can be summarized as follows.

Proposition 1. Let $K \in \mathfrak{U}$ be a compact, star–shaped set of unit area and Ξ_K the minimal area of type K. Then the distribution of the minimal area size, $W_K = \mathfrak{m}(\Xi_K)$, is given by

$$P(W_K \leq x) = \prod_{i=1}^{s} \left(1 - e^{-\lambda_i x}\right), \quad x > 0, \tag{7}$$

and the species–area curve is given by

$$E\big(Z(A_x)\big) = \sum_{i=1}^{s} \left(1 - e^{-\lambda_i x}\right), \quad x > 0 \tag{8}$$

for any Borel set $A_x \in \mathcal{B}$ with size $\mathfrak{m}(A_x) = x$.

Using this formula, it is possible to give explicit expressions for the mean $E(W_K)$ and the variance $\sigma^2(W_K)$ in terms of $\lambda_1, \dots, \lambda_s$.

Proposition 2. Under the assumptions of Proposition 1, we have

$$E(W_K) = \sum_{k=1}^{s} (-1)^{k-1} \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq s} \cdots \sum \frac{1}{\lambda_{i_1} + \dots + \lambda_{i_k}} \tag{9}$$

$$E(W_K^2) \;=\; 2\sum_{k=1}^{s}(-1)^{k-1}\sum_{1\le i_1<i_2<\ldots<i_k\le s}\cdots\sum \frac{1}{(\lambda_{i_1}+\ldots+\lambda_{i_k})^2} \qquad (10)$$

with $\sigma^2(W_K) = E(W_K^2) - \left\{E(W_K)\right\}^2$, and, more generally, for the moment generating function,

$$E\left(e^{tW_K}\right) = \sum_{k=1}^{s}(-1)^{k-1}\sum_{1\le i_1<i_2<\ldots<i_k\le s}\cdots\sum \frac{\lambda_{i_1}+\ldots+\lambda_{i_k}}{\lambda_{i_1}+\ldots+\lambda_{i_k}-t} \qquad (11)$$

for $t < \min\{\lambda_1,\ldots,\lambda_s\}$, hence

$$E(W_K^n) \;=\; n!\sum_{k=1}^{s}(-1)^{k-1}\sum_{1\le i_1<i_2<\ldots<i_k\le s}\cdots\sum \frac{1}{(\lambda_{i_1}+\ldots+\lambda_{i_k})^n} \qquad (12)$$

for all $n \in \mathbb{N}$.

Proof. Recall that for any real numbers $a_1,\ldots,a_s$ we have:

$$1-\prod_{i=1}^{s}(1-a_i) = \sum_{k=1}^{s}(-1)^{k-1}\sum_{1\le i_1<i_2<\ldots<i_k\le s}\cdots\sum \prod_{j=1}^{k} a_{i_j}. \qquad (13)$$

It suffices to prove relation (11). If we denote by $F(x) = P(W_K \le x)$, $x > 0$, the distribution function of W_K and by f the corresponding density, then, by partial integration, for $t < \min\{\lambda_1,\ldots,\lambda_s\}$,

$$\begin{aligned}
E\left(e^{tW_K}\right) &= \int_0^\infty e^{tx}f(x)\,dx = 1 + t\int_0^\infty e^{tx}\left[1-F(x)\right]dx \\
&= 1 + t\int_0^\infty \sum_{k=1}^{s}(-1)^{k-1}\sum_{1\le i_1<i_2<\ldots<i_k\le s}\cdots\sum e^{-\left\{\sum_{j=1}^{k}\lambda_{i_j}-t\right\}x}\,dx \\
&= 1 + \sum_{k=1}^{s}(-1)^{k-1}\sum_{1\le i_1<i_2<\ldots<i_k\le s}\cdots\sum t\int_0^\infty e^{-\left\{\sum_{j=1}^{k}\lambda_{i_j}-t\right\}x}\,dx \\
&= 1 + \sum_{k=1}^{s}(-1)^{k-1}\sum_{1\le i_1<i_2<\ldots<i_k\le s}\cdots\sum \frac{t}{\lambda_{i_1}+\ldots+\lambda_{i_k}-t} \qquad (14)
\end{aligned}$$

from which relation (11) follows readily since by (13), for $a_i = 1$,

$$\begin{aligned}
E\left(e^{tW_K}\right) &= 1 + \sum_{k=1}^{s}(-1)^{k-1}\sum_{1\le i_1<i_2<\ldots<i_k\le s}\cdots\sum \frac{t}{\lambda_{i_1}+\ldots+\lambda_{i_k}-t} \\
&= \sum_{k=1}^{s}(-1)^{k-1}\sum_{1\le i_1<i_2<\ldots<i_k\le s}\cdots\sum \left[1+\frac{t}{\lambda_{i_1}+\ldots+\lambda_{i_k}-t}\right] \\
&= \sum_{k=1}^{s}(-1)^{k-1}\sum_{1\le i_1<i_2<\ldots<i_k\le s}\cdots\sum \frac{\lambda_{i_1}+\ldots+\lambda_{i_k}}{\lambda_{i_1}+\ldots+\lambda_{i_k}-t}. \qquad (15)
\end{aligned}$$

The remaining formulas are proved by differentiation of relation (11).

In most practical situations, when typically s is large (more than 10, say), it might be very cumbersome to compute the mean and variance of the minimal area size by formulas (9) and (10) unless all λ_i are equal, a case which will be treated separately in the sequel. However, a suitable application of symbolic Computer Algebra programs like MAPLE© could be recommended for such cases. The following figure shows a possible worksheet for MAPLE V Release 3 under WINDOWS© for calculating mean and variance of the size of the minimal area and for plotting the theoretical species–area curve and the density f of the size distribution of the minimal area. Figures 3 and 4 were created just this way, using as input the following 10 intensities for the corresponding Poisson processes: $\lambda_1 = \ldots = \lambda_7 = 0.1$, $\lambda_8 = 0.15$, $\lambda_9 = 0.3$, $\lambda_{10} = 2.7$.

```
Maple V - MINAREA.MS
File  Edit  Format  View  Options  Help

> with(linalg):
> with(plots):
> readlib(readdata):
> lambda:=convert(readdata(`minarea.dat`),rational);
```

$$\lambda := \left[\frac{1}{10}, \frac{1}{10}, \frac{1}{10}, \frac{1}{10}, \frac{1}{10}, \frac{1}{10}, \frac{1}{10}, \frac{3}{20}, \frac{3}{10}, \frac{27}{10} \right]$$

```
> s:=vectdim(lambda):
> S:=sum(1-E^(-lambda[i]*x),i=1..s):
> F:=product(1-E^(-lambda[i]*x),i=1..s):
> G:=expand(1-F):
> H:=int(G,x):
> mu:=-evalf(subs(x=0,H));
```

$$\mu := 26.32648819$$

```
> expected:=evalf(subs(x=mu,S));
```

$$expected := 9.477139467$$

```
> J:=int(H,x):
> variance:=evalf(2*subs(x=0,J)-mu^2);
```

$$variance := 147.270896$$

```
> sigma:=sqrt(variance);
```

$$\sigma := 12.13552207$$

```
> f:=diff(F,x):
> lower:=1/max(seq(lambda[j],j=1..s)):
> upper:=(s+1)/min(seq(lambda[j],j=1..s)):
> loglogplot(S,x=lower..upper,labels=[area,number]);
> plot(f,x=0..upper,labels=[area,density]);
```

`Bytes Used:  1050K   Time Used:   48 sec   Free Memory:   11219K`
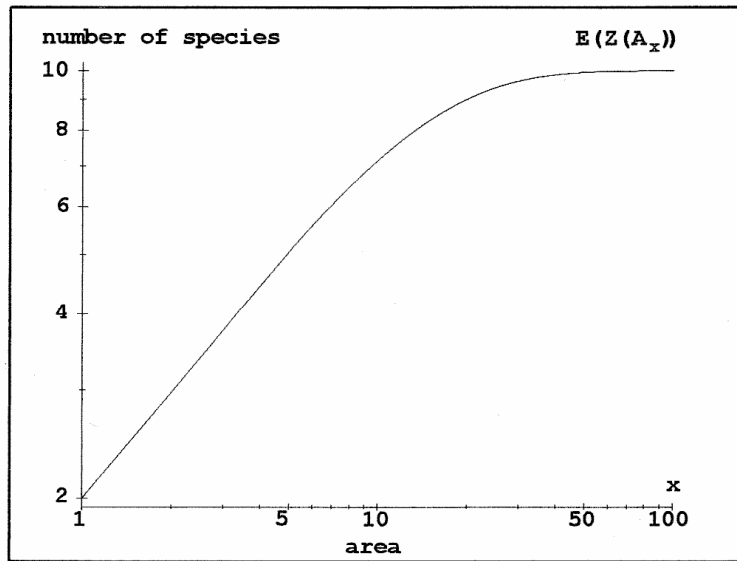
Fig. 2

number of species $E(Z(A_x))$

Fig. 3

During the execution of the worksheet, data are first read from the file MINAREA.DAT containing the intensities as non-negative numbers in ASCII format, collected in the vector **lambda**. The letter S denotes the species–area curve, **F** the cumulative distribution function of W_K, and **f** its density. μ and σ represent the mean $E(W_K)$ and the standard deviation $\sigma(W_K)$. Additionally, the expected number of species detected in an area of size μ, i.e. $E\big(Z(A_\mu)\big)$, is computed. For the example under consideration, $\mu = 26.32$, $\sigma = 12.13$. The expected number of detected species is 9.47 which is close to the number of all species, 10. Note that the present worksheet has been optimized w.r.t. the computation speed and memory capacity, so that good results are obtained already on PC's with 8 MB RAM or less.

The graph in figure 3 shows quite clearly that a local approximate log–log linear form of the species–area curve can be obtained by a suitable choice of intensities. Typically, in such a community we have a few dominant species characterized by "large" values of the intensities (here: $\lambda_{10} = 2.7$) and a larger number of more rare species characterized by "small" values of the intensities (here: $\lambda_1 = \ldots = \lambda_7 = 0.1$). A Poisson model of the above form therefore explains quite naturally why the log–log linear shape of the species–area curve is frequently observed. For practical applications, the intensities have of course to be estimated, e.g. in a "pilot study", before the minimal area characteristics can be calculated explicitly. Statistical methods for this type of problem are well–developed, see e.g. Diggle (1983), sections 3.2.2 and 3.3.3. In order to see the influence of "rare" species on the minimal area size, one could add a few fictional species to the observed assemblage with "low" values for the intensities, and compare the results.
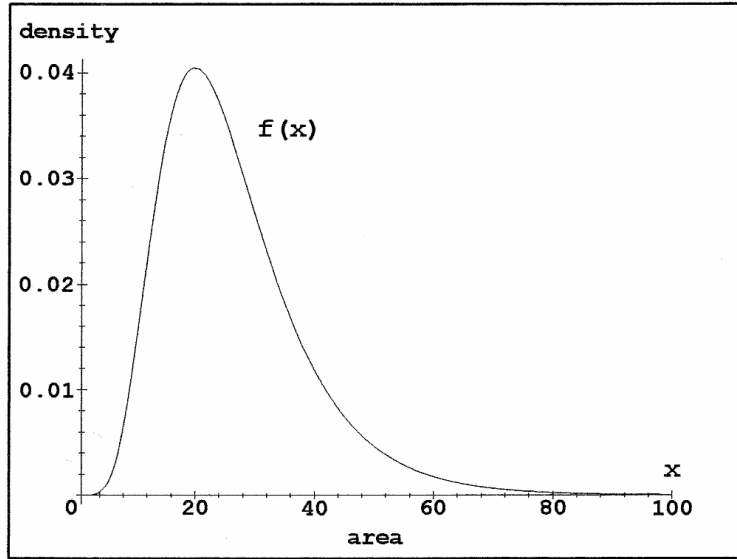
Fig. 4

The situation is analytically more simple if we assume that all $\lambda_i = \lambda > 0$ are equal, which corresponds to the case of maximal diversity in the community (see e.g. Magurran (1988)). In this case, the minimal area size W_K is simply the maximum of s independent, identically exponentially distributed random variables with parameter λ, which is – after suitable normalization – asymptotically doubly exponentially distributed when $s \to \infty$. The calculation of moments reduces to

$$\mu = \quad E(W_K) \quad = \frac{1}{\lambda} \sum_{i=1}^{s} \frac{1}{i} \approx \frac{\ln s + C}{\lambda} \tag{16}$$

$$\sigma^2 = \quad \sigma^2(W_K) \quad = \frac{1}{\lambda^2} \sum_{i=1}^{s} \frac{1}{i^2} \approx \frac{\pi^2}{6\lambda^2} \tag{17}$$

for large s, where $C = 0.577216\ldots$ denotes Euler's constant (see e.g. Pfeifer (1987), Lemma 0.3 for a simple direct proof). The expected number of species detected in an area $A_\mu \in \mathcal{B}$ of size μ here is approximately

$$E\big(Z(A_\mu)\big) = s\left(1 - e^{-\lambda\mu}\right) \approx s\left(1 - \frac{e^{-C}}{s}\right) = s - e^{-C} = s - 0.5614\ldots \tag{18}$$

for large s, i.e. the loss in the number of detected species is less than one (the r.h.s. of (16) also provides a lower bound of $E(W_K)$ if formally C is chosen to be zero). Note that the size of the minimal area here increases only as $\ln s$ when s goes to infinity.

4 Possible Generalizations and Conclusions

At first glance, two mainstreams of generalizations seem to be of particular interest here:

    1. consideration of non–homogeneous Poisson processes

    2. "minimal areas" which cover t out of s species with $t < s$.

In the first case, it makes sense to maintain the original definition of Ξ_K, assuming again that the origin $\mathbf{0}$ corresponds to the location of the observer, i.e. the shaping set K is scaled around the observer by the random factor ρ_K until all s species are covered. For the minimal area size distribution, we have, similarly as above,

$$
\begin{aligned}
P(W_K \leq r) &= P\Big(\bigcap_{i=1}^{s}\{\xi_i(\sqrt{r}K) > 0\}\Big) = \prod_{i=1}^{s} P\big(\{\xi_i(\sqrt{r}K) > 0\}\big) \\
&= \prod_{i=1}^{s}\Big(1 - P\big(\xi_i(\sqrt{r}K) = 0\big)\Big) \hspace{3cm} (19) \\
&= \prod_{i=1}^{s}\Big(1 - e^{-E\xi_i(\sqrt{r}K)}\Big), \quad r > 0, \hspace{2cm} (20)
\end{aligned}
$$

which can easily be evaluated if the intensity measures $E\xi_i$ are explicitly known. Note, however, that unlike in the homogeneous case, the distribution of W_K now also depends on the shape of K, in general, as well as on the "physical" location of the origin $\mathbf{0}$ in the observation plane.

In the second case, the definition of the individual $\rho_K(i)$ can likewise be maintained; an appropriate definition of a t–out–of–s minimal area could then be

$$
\rho_K = \rho_K[t:s], \quad \Xi_K = \rho_K K, \hspace{3cm} (21)
$$

where $\rho_K[1:s] \leq \rho_K[2:s] \leq \ldots \leq \rho_K[s:s]$ denote the corresponding *order statistics*. The distribution of W_K can generally be evaluated by combinatorial arguments, similarly as above; e.g., one obtains in the homogeneous Poisson case, for $t < s$:

$$
\begin{aligned}
P(W_K \leq r) &= P\Big(\bigcup_{1\leq j_1<\ldots<j_t\leq s}\bigcap_{k=1}^{t}\{\xi_{j_k}(\sqrt{r}K) > 0\} \cap \bigcap_{i\in\{1,\ldots,s\}\setminus\{j_1,\ldots,j_t\}}\{\xi_i(\sqrt{r}K) = 0\}\Big) \\
&= \sum_{1\leq j_1<\ldots<j_t\leq s}\prod_{k=1}^{t}P(\xi_{j_k}(\sqrt{r}K) > 0)\prod_{i\in\{1,\ldots,s\}\setminus\{j_1,\ldots,j_t\}}P\big(\{\xi_i(\sqrt{r}K) = 0\}\big) \\
&= \sum_{1\leq j_1<\ldots<j_t\leq s}\exp\Big\{-\sum_{i\notin\{j_1,\ldots,j_t\}}\lambda_i r\Big\}\prod_{k=1}^{t}\Big(1 - e^{-\lambda_{j_k}r}\Big), \quad r > 0. \hspace{1cm} (22)
\end{aligned}
$$

Of course, both aspects could also be combined in one model. The practical limitation of such approaches, however, lies in the fact that detailed spatial

information about the dispersal of individuals is necessary, which will often not be available.

Corresponding modifications in the MAPLE worksheet given above can easily be made by replacing the original expression for F by the appropriate one for the cumulative distribution function of W_K.

5   Proofs & Facts from Stochastic Geometry and Point Processes

Our notation closely follows Stoyan, Kendall, and Mecke (1987). Let $\mathcal{F}$ denote the collection of all closed subsets of $\mathbb{R}^2$, and for a compact set $K \in \mathcal{K}$, $\mathcal{F}_K = \{F \in \mathcal{F} \mid F \cap K \neq \emptyset\}$. Let further $\mathfrak{F} = \sigma\{\mathcal{F}_K \mid K \in \mathcal{K}\}$ denote the correspondingly generated sigma–field. A two–dimensional *random* closed set Ξ is a $(\mathcal{F}, \mathfrak{F})$–measurable map defined on a suitable probability space $(\Omega, \mathcal{A}, P)$ where $\mathcal{A}$ is a σ–field over the non–empty set Ω.

Lemma 1. Let $K \in \mathcal{K}$ be a compact set and ρ a non–negative and finite random variable. Then $\Xi = \rho K$ is a random compact set.

Proof. Let $L \in \mathcal{K}$ be an arbitrary compact set. For every real number $r > 0$, the set rK is also compact, and either $rK \cap L \neq \emptyset$, or rK and L have a positive distance of each other. Hence the set $T = \{r > 0 \mid rK \cap L = \emptyset\}$ is a countable union of open intervals in $\mathbb{R}^1$ and thus a Borel set. Since $\{r > 0 \mid rK \cap L \neq \emptyset\} = (0, \infty) \setminus T$, the map $r \mapsto rK$ is $(\mathcal{F}, \mathfrak{F})$–measurable, and so is ρK.

Lemma 2. Let $H : \mathcal{F} \times \mathbb{R}^2 \to \{0, 1\} : (F, \mathbf{x}) \mapsto \mathbf{1}_F(\mathbf{x})$ where $\mathbf{1}_F$ denotes the indicator random variable for the set F. Then H is measurable w.r.t. the product σ–field $\mathfrak{F} \otimes \mathcal{B}$.

Proof. By the closedness of sets in $\mathcal{F}$, we have

$$H^{-1}(\{1\}) = \{(F, \mathbf{x}) \in \mathcal{F} \times \mathbb{R}^2 \mid \mathbf{x} \in F\} = \bigcap_{n=1}^{\infty} \bigcup_{i=1}^{\infty} \mathcal{F}_{K_{in}} \times K_{in} \in \mathfrak{F} \otimes \mathcal{B}$$

where $K_{in} \in \mathcal{K}$ can be chosen as compact sets with diameters $\Phi(K_{in}) = \varepsilon_n \downarrow 0$, $n \to \infty$ for all i and $\bigcup_{i=1}^{\infty} K_{in} = \mathbb{R}^2$ for all n.

Theorem 1. Let μ be a σ–finite measure on $\mathcal{B}$ and Ξ a random closed set. Then $\mu(\Xi)$ is a non–negative random variable.

Proof. By Lemma 2, the map $(\Xi, \mathbf{x}) \mapsto \mathbf{1}_\Xi(\mathbf{x})$ for $\mathbf{x} \in \mathbb{R}^2$ is a product–measurable random variable, hence by the σ–finiteness of μ,

$$\mu(\Xi) = \int \mathbf{1}_\Xi(\mathbf{x}) \, \mu(d\mathbf{x})$$

is also measurable and thus a random variable.

Lemma 3. Let η be a homogeneous Poisson point process over $\mathbb{R}^2$ with intensity $\nu > 0$, and $K \in \mathcal{K}$ a non–empty compact set with $\mathfrak{m}(K) > 0$. Then

$$\rho = \inf\{r > 0 \mid \eta(rK) > 0\}$$

is – up to possible modifications on sets of measure zero – a finite random variable with

$$P(\rho > r) = e^{-\nu \mathfrak{m}(K)r^2}, \quad r > 0,$$

i.e. ρ^2 has an exponential distribution with parameter $\nu \mathfrak{m}(K)$:

$$P(\rho^2 > r) = P(\rho > \sqrt{r}) = e^{-\nu \mathfrak{m}(K)r}, \quad r > 0.$$

Proof. We have $\{\rho > r\} = \{\eta(rK) = 0\}$, $r > 0$, hence

$$P(\rho > r) = P\big(\eta(rK) = 0\big) = e^{-\nu \mathfrak{m}(rK)} = e^{-\nu \mathfrak{m}(K)r^2}.$$

Note that it is always possible to modify ρ on sets where $\rho = \infty$ (which are of measure zero), so that w.l.o.g. one can always assume that ρ is finite everywhere. For the subsequent Theorem, it is tacitly assumed that such modifications are made where necessary.

Theorem 2. Let, with the notation of Section 2, denote

$$\rho_K = \inf\{r > 0 \mid Z(rK) = s\}, \quad \Xi_K = \rho_K K$$

for $K \in \mathfrak{U}$. Then ρ_K is a finite random variable, ρ_K^2 is distributed as the maximum of s independent exponentially distributed random variables with parameters $\lambda_1, \ldots, \lambda_s$ each, i.e.

$$P(\rho_K^2 \leq r) = \prod_{i=1}^{s} \left(1 - e^{-\lambda_i r}\right), \quad r > 0,$$

and Ξ_K is a random compact set. The size distribution of $W_K = \mathfrak{m}(\Xi_K)$ coincides with the distribution of ρ_K^2.

Proof. Let for $i = 1, \ldots, s$ denote $\rho_K(i) = \inf\{r > 0 \mid Z_i(rK) = 1\}$, as above. By Lemma 3, $\rho_K(i)$ is a finite random variable with $\rho_K^2(i)$ being exponentially distributed with parameter λ_i, i.e.

$$P\left(\rho_K^2(i) \leq r\right) = 1 - e^{-\lambda_i \mathfrak{m}(K)r} = 1 - e^{-\lambda_i r}, \quad r > 0.$$

Hence

$$\rho_K = \inf\left\{r > 0 \mid \bigcap_{i=1}^{s}\{\xi_i(rK) > 0\}\right\} = \max\{\rho_K(1), \ldots, \rho_K(s)\},$$

from which the first statement follows readily. For the second part, first note that Ξ_K is a random compact set by Lemma 1. Hence by Theorem 1, $\mathfrak{m}(\Xi_K)$ is a random variable which is finite by the compactness of K and the finiteness of ρ_K with

$$
\begin{aligned}
P\Big(\mathfrak{m}(\rho_K K) \leq r\Big) &= P\Big(\mathfrak{m}(\rho_K K) \leq \mathfrak{m}(\sqrt{r}K)\Big) = P\Big(\rho_K^2\,\mathfrak{m}(K) \leq r\,\mathfrak{m}(K)\Big) \\
&= P(\rho_k^2 \leq r) = P\Big(\max\{\rho_K^2(1),\ldots,\rho_K^2(s)\} \leq r\Big) \\
&= \prod_{i=1}^{s} P(\rho_K^2(i) \leq r) = \prod_{i=1}^{s}\Big(1 - e^{-\lambda_i r}\Big)
\end{aligned}
$$

for $r > 0$.

Acknowledgement.

6   References

M. Begon, J.L. Harper, and C.R. Townsend (1990): Ökologie. Individuen • Populationen • Lebensgemeinschaften. Birkhäuser, Basel.

P. Diggle (1983): Statistical Analysis of Spatial Point Patterns. Mathematics in Biology Series. Ac. Press, London.

C.J. Krebs (1985): Ecology. 3rd ed., Harper & Row, New York.

A.E. Magurran (1988): Ecological diversity and Its Measurement. Croom Helm, London.

R.M. May (1975): Patterns of species abundance and diversity. In: *Ecology and Evolution of Communities,* ed. M.L. Cody, and J.M. Diamond, 81 – 120. Harvard Univ. Press, Cambridge, Mass.

D. Pfeifer (1987): Einführung in die Extremwertstatistik. Teubner Skripten zur Mathematischen Stochastik. Teubner, Stuttgart.

U. Schleier, and K.H. van Bernem (1995): Statistical methods to determine sample size for the estimation of species richness and species' abundances in benthic communities. To appear in: Archive for Fish and Marine Research.

D. Stoyan, W.S. Kendall, and J. Mecke (1987): Stochastic Geometry and Its Applications. Akademie–Verlag, Berlin; reprinted by Wiley, New York.

S. Weinberg (1978): The minimal area problem in invertebrate communities of mediterranean rocky substrata. Marine Biology **49**, 33 – 40.

C.B. Williams (1964): Patterns in the Balance of Nature and Related Problems in Quantitative Ecology. Ac. Press, New York.