

THESIS FOR THE DEGREE OF LICENTIATE OF PHILOSOPHY

Tree influence on understory vegetation: an edge correction and a conditional model

SHARON KÜHLMANN-BERENZON

CHALMERS | GÖTEBORG UNIVERSITY



Department of Mathematical Statistics
Chalmers University of Technology and Göteborg University
SE-412 96 Göteborg, Sweden
March 2002

Tree influence on understory vegetation: an edge
correction and a conditional model
Sharon Kühlmann-Berenzon

© Sharon Kühlmann-Berenzon, 2002
ISSN-0347-2809
Technical report No. 2002:12
Department of Mathematical Statistics
Chalmers University of Technology
SE-412 96 Göteborg
Sweden
Telephone +46 (0)31-772 1000

Cover: Diagram of the concepts of
the edge correction for the
permanent sample plots of the
1985-86 National Forest Inventory
of Finland.

Layout produced in \LaTeX by Dan Mattsson.
School of Mathematical Sciences
Chalmers University of Technology and
Göteborg University
Göteborg, March 2002

Tree influence on understory vegetation: an edge correction and a conditional model

Sharon Kühlmann-Berenzon
Department of Mathematical Statistics
Chalmers University of Technology and Göteborg University

SUMMARY

The relation between the understory vegetation species and the surrounding trees was studied using data collected in the permanent sample plots of the 1985-86 National Forest Inventory of Finland. The influence potential index quantified the effect of the trees with an exponential function of the spatial location and size of the tree.

The observed values of the influence potential, however, were censored by the plot borders. An edge correction was therefore developed, which calculates the expected influence potential outside the plot borders; this was achieved with the Campbell theorem for a marked point process. The correction removed the bias in the influence potential of the large trees, but overcompensated in the case of the small trees.

A model for the presence of an understory vegetation species was then derived that avoided the large-scale factors implicit in the data. The result was a conditional logistic regression for the pattern of presence and absence in the quadrats of the plot. Observations of cowberry (*Vaccinium vitis-idaea*) were fitted using the edge-corrected influence potential of pine, spruce, and birch as explanatory variables; the results show that higher influence from pine and spruce reduce the odds of the observed pattern.

Keywords: edge correction, spatial point process, marked point process, conditional logistic regression, influence potential, forestry.

Acknowledgments

The interdisciplinary nature of this project, as well as other lucky circumstances, led Aila Särkkä and Urban Hjorth at the Department of Mathematical Statistics, and Juha Heikkinen at METLA to share the supervision of this project. Although this situation might be considered unorthodox by some, in reality it created a prolific and interesting research environment, while giving me the possibility to learn about different fields. I wish to thank Aila, Urban, and Juha for providing me with this opportunity, which together with their constant feedback and encouragement across three countries, made it a pleasure to work with them.

The continuation of my graduate studies has been possible in part by a scholarship from the Universidad de Costa Rica. The idea for the project came originally from METLA; there I would like to thank Prof. Erkki Tomppo and Raisa Mäkipää for their interest and help.

Nobody survives this far away from home and this long without a strong and wide network of caring people. I would not had made it without my friends in Costa Rica, Sweden, and elsewhere nor without the colleagues at the department. A special recognition to Ola Engvall, for being a patient and fun roommate; and to Dan Mattsson, for creating the layout of this document and providing his computer expertise.

My gratitude to Alexander Ploner, who did not let me forget that applied statistics is challenging, relevant, interesting, and more fun, and who has kept cheering me on along the way. Finally, my parents, my sister, and my brother have given me the love necessary to reach this point.

Table of contents

| | |
|---|----------|
| Introduction | 1 |
| 1 Part I | 2 |
| 2 Part II | 2 |
| Background | 5 |
| 1 Ecological Field Theory | 5 |
| 2 Data: Permanent sample plots (PSP) | 8 |
| 3 Ecological model: Influence Potential | 9 |

| | | |
|-----------|--|-----------|
| I | Edge correction for influence potential | 13 |
| 1 | Introduction | 13 |
| 2 | Spatial point processes | 14 |
| 3 | Edge-correction | 17 |
| 4 | Edge effects in the influence potential | 24 |
| 5 | Marked point processes | 25 |
| 6 | Derivation of edge correction | 27 |
| 7 | Application | 32 |
| 8 | Discussion and Conclusions | 46 |
| A | Appendix: Formulas by case | 49 |
| | | |
| II | Model for the local presence of an understory species | 61 |
| 1 | Introduction | 61 |
| 2 | Logistic regression | 62 |
| 3 | Logistic regression model for PSP data | 63 |
| 4 | Conditional logistic model for PSP data | 64 |
| 5 | Fitting of a conditional logistic model | 71 |
| 6 | Application | 72 |
| 7 | Conclusions | 77 |
| A | Appendix: Likelihood of conditional logistic model | 81 |
| | | |
| | References | 83 |

Introduction

The understory in a forest is composed of those plants that grow underneath the canopy of the taller trees, and include grasses, herbs, dwarf bushes, and mosses. It plays a vital role in giving shelter to animals, protecting and enriching the soil, and providing fodder and timber. Summary statistics at the level of tree stands (e.g. total crown coverage) have already been used to describe the understory vegetation. These measures, however, do not account for the diversity found within a stand, and it is possible that more accurate results may be obtained by including species, size, spatial patterns, and other characteristics of the trees. Information on the relationship between trees and understory vegetation can bring further understanding to the ecological processes that take place in the forest between trees and understory, which may have consequences for forestry management and biodiversity policies.

The data collected in the permanent sample plots of the National Forest Inventory of Finland provides a rich source of observations on trees and understory species. In collaboration with the Finnish Forest Research Institute (METLA), this project was set up to study the relationship between single trees and individual understory species using that data. The *influence potential* was chosen as the index to quantify the ecological effect of the trees; this measure has been previously applied to similar studies on small and homogeneous areas (Kuuluvainen and Pukkala 1989; Kuuluvainen et al. 1993; Økland et al. 1999; Saetre 1999).

The use of the permanent sample plots provided additional challenges to the analysis. First, since the study area covers all of Finland, it is very heterogeneous, including diverse environmental conditions and management practices; second, the measurements were carried out within small, delimited plots, which cause the observations on trees to be censored.

This thesis starts by presenting the main ideas on ecological field theory, which serves as background for the concept of influence potential. A description of the permanent sample plots and of the model of influence potential adopted in this study are then provided. In Part I we develop an edge correction method for the influence potential, and in Part II, a conditional model for the presence of an understory species. Additional exploratory results from this study are reported in Kühlmann, Heikkinen, Särkkä, and Hjorth (2001).

1 PART I

Before attempting any modeling, we first had to find an edge correction method for the influence potential measurements. This was necessary because the observations on the trees were censored by the borders of the plots; that meant that possible relevant information from trees standing outside the borders was missing. The method proposed is based on the Campbell theorem for a marked point process. The development of the edge correction and its application to the data from the permanent sample plots is presented in Part I of this thesis. Comparisons of the distribution of the influence potential with and without the correction show that the method works in a satisfactory way for the larger trees, but overcompensates for the smaller trees measured in the inner plot.

2 PART II

In Part II we derived a model for the presence of an individual understory species, where the influence potential of the three dominating tree species served as explanatory variables. The result was a conditional logistic model which avoids the large scale factors that were implicit in the data. This is the first attempt to find a relation between the characteristics

and spatial pattern of the trees, and the presence of an understory species. The model was fitted to observations of *Vaccinum vitis-idaea* and significant results were obtained that show a connection between the influence potential of pine and spruce, and the presence of this species.

Background

1 ECOLOGICAL FIELD THEORY

Ecological field theory (EFT) was originally introduced by Wu, Sharpe, Walker, and Penridge (1985) as a theoretical approach to the study of interactions among individual plants. Based on the field theory of Physics, EFT assumes that a field or domain exists around every plant where the plant influences the availability of resources according to its own characteristics and other environmental factors. As the plant adds or subtracts resources, it facilitates or suppresses the growth of other plants situated inside the influence field. This general framework allows different individuals from the same community to be compared because it considers specific characteristics of each plant, and it also incorporates the spatial configuration since it defines a physical domain. Furthermore, the way the plant influences its domain may be described mathematically.

The mathematical description of the influence on the domain is chosen empirically. Wu et al. (1985), for example, defined interference potential, a measure of the interference that a plant in a specific location is subject to and which must overcome in order to grow. The interference is a function of different system parameters that are determined by the surrounding plants. The parameters quantified were water, nutrient, and light availability, which themselves were functions of the crown, stem, and roots of the surrounding plants. Walker, Sharpe, Penridge, and Wu (1989) presented a simpler version of EFT where the interaction intensity between

two plants depended on the size of their influence domains, their individual response functions to environmental stress, and the overlap of the two domains.

Kuuluvainen and Pukkala (1989) applied EFT to study the relationship between Scots pine seed trees, and seedlings and understory vegetation in a Finnish forest. They introduced a single-tree index that quantified the influence of trees on a specific location, in this case, where the vegetation was observed. The calculation of this *influence potential* consisted of two steps: the first described the effect of an individual tree on the location; the second part combined the effect of the surrounding trees into the influence potential IP affecting the location.

An exponential decay based on the distance $t_z(q)$ between the tree z and the location q , and a parameter b described the effect Φ_z of a tree as

$$\Phi_z(q) = \Phi_z(0) \cdot \exp(-b_z \cdot t_z(q)^2).$$

The parameter b_z was determined by trial and error as $1/(0.4h_z)$, where h_z was the height of the tree; and the effect $\Phi_z(0)$ at the tree location was estimated as $D_z/35$, D_z standing for the diameter at breast height (DBH) of the tree. In this model the effect of a tree is strongest at the stem and decreases with distance from the tree.

The influence potential IP(q) at the location q was obtained as

$$\text{IP}(q) = 1 - \prod_z (1 - \Phi_z(q)). \quad (1)$$

This index ranges continuously between 0 and 1: low values indicate that the location q suffers of no interferences, i.e. it is at a greater distance from the trees; and high values mean that, because it is close to trees, the location is subject to high influence.

Økland, Rydgren, and Økland (1999) derived a formula for $\Phi(q)$ using more parameters and characteristics of the tree, such as height, crown radius, distance from the tree, and DBH. They concluded, however, that the best representation of the tree effect was a function of distance and

DBH, and then they modeled IP multiplicatively as in (Kuuluvainen and Pukkala, 1989). Their model of $\Phi(q)$, estimated with data collected in a Norwegian spruce forest, was

$$\Phi_z(q) = \left(\frac{D_z}{64}\right)^a \exp \left[-5.298 \left(\frac{0.621 \cdot t_z(q)}{c \cdot D_z^{0.689}} \right)^b \right];$$

in order to find the parameters a , b , and c , they recommended to optimize the previous function to the data using constrained ordination.

Another study, performed in a mixed spruce-birch stand by Saetre (1999), applied yet another formulation of the concept of influence potential. The main differences to the previous ones were the use of the linear distance from the tree and lack of additional parameters in $\Phi(q)$, and the additive function instead of a multiplicative one for IP; thus

$$\begin{aligned} \Phi_z(q) &= D_z \cdot \exp(-t_z(q)) \\ \text{IP}(q) &= \sum_z \Phi_z. \end{aligned} \quad (2)$$

The interpretation of Eq. 2 is the same as for Eq. 1, but there is no upper bound with this model, i.e. it increases as the number of trees increases.

Although the mathematical expressions for the tree effect $\Phi(q)$ and for the influence potential $\text{IP}(q)$ differed among the three studies, two key assumptions of EFT are present: there is an influence domain defined, and the individual characteristics of the trees are considered; the former is carried out by an exponential function, and the latter by the DBH and the height. Furthermore, implicit in the definitions of tree effect as an exponential model is the assumption that the effect is symmetrical around the stem.

Kuuluvainen and Pukkala (1989), Økland et al. (1999), and Saetre (1999) related their IP functions to the abundance of understory vegetation and seedlings, and to soil measurements, using statistical methods such as correlation, variogram analysis, linear regression, and constrained ordination. For the analysis of the understory vegetation, the species were

grouped in categories such as mosses, grasses and herbs, and dwarf shrubs (Kuuluvainen and Pukkala, 1989), or vascular plants and cryptograms (Økland et al., 1999); or described by principal components (Saetre, 1999). The data was obtained in all cases from boreal forests with one type of stand and relatively small study areas: pine in a 140 m² plot (Kuuluvainen and Pukkala, 1989), a spruce stand of 2 km² (Økland et al., 1999), and mixed spruce-birch forest in an area of 14 m x 22 m (Saetre, 1999); another similar study that applied the formulas of Kuuluvainen and Pukkala (1989), was carried out in a 1 ha Scots pine stand in Finland by Kuuluvainen, Hokkanen, Järvinen, and Pukkala (1993).

2 DATA: PERMANENT SAMPLE PLOTS (PSP)

The data for this study was collected in the permanent sample plots (PSP) established by the Finnish Forest Research Institute (METLA) as part of the 8th National Forest Inventory (1985-1986). PSP consists of 3009 plots located on a regular grid covering all of Finland (337 000 km²) (Fig. 1). In Southern Finland four plots were assigned to each cluster, with 400 m between the plots and 16 km between the clusters. In Northern Finland three plots at 600 m distance formed each cluster, and 24 km in N-S direction and 32 km in E-W direction separated the clusters.

The plots were located on forestry land and were circular with a radius of 9.77 m (area of 300 m²); figure 2 shows a schematic plot. Trees with DBH greater than 10.5 cm were measured in those circles, and their spatial location recorded, while trees with DBH between 4.5 cm and 10.5 cm were observed within a second radius of 5.64 m (area 100 m²) from the plot center. The stands within the plots were also identified. The dominating tree species were Scots pine (*Pinus sylvestris*), Norway spruce (*Picea abies*), and birch (*Betula pubescens* and *B. pendula*).

Within each plot six quadrats of 2 m² were systematically assigned: four at distances of 3 m and 8 m from the plot center on the N-S axis, and two more at 6 m from the plot center on the E-W axis. These were classified according to laying on mineral soils or peatlands. The understory vegetation was recorded by species and measured visually as percentage of the quadrat covered. Not all quadrats were consistently measured in every plot, but in 95% of the plots at least three were measured.

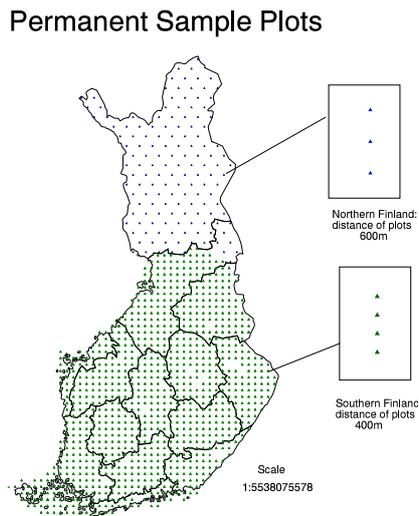


Figure 1: Permanent sample plots (PSP): Sampling grid.

This data has been used to monitor the health of the forest as well as to study the understory (e.g. Tonteri, Mikkola, and Lahti 1990; Korpela and Reinikainen 1996).

3 ECOLOGICAL MODEL: INFLUENCE POTENTIAL

The formulation for the influence potential IP used in this study is a combination of those from Kuuluvainen and Pukkala (1989), Økland et al. (1999), and Saetre (1999). We defined the effect of a tree of species T on a quadrat q as

$$\begin{aligned}\Phi_{z \in T}(q) &= D_z \cdot \exp\left(-\frac{\|z - q\|^2}{c_T}\right) \\ &= D_z \cdot \exp\left(-\frac{t_z(q)^2}{c_T}\right).\end{aligned}$$

The parameter c_T is a scale for the size of the domain of the influence field for the tree species T ; $t_z(q)$ represents the Euclidean distance between the

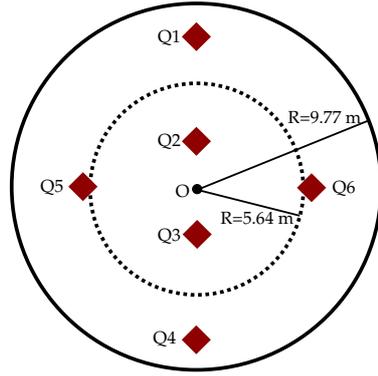


Figure 2: Permanent sample plot (PSP): Quadrats 1 and 4 (Q1 and Q4) are at 8 m, quadrats 2 and 3 (Q2 and Q3) at 3 m, and quadrats 5 and 6 (Q5 and Q6) at 6 m from the plot center (O); quadrats are 2 m^2 . Within the plot of $R = 9.77 \text{ m}$ (solid line), all trees with DBH greater than 10.5 cm were measured. Within the plot of $R = 5.64$ (dashed line), trees with DBH between 4.5 cm and 10.5 cm were also measured. Illustration at scale.

tree z and the quadrat q ; and D_z is the DBH of tree z . This expression of $\Phi(q)$ is symmetric around the stem; it assumes that the largest effect is found at the tree location and that the magnitude of the effect decreases with distance. Moreover, the domain for the species grows with c_T ; in other words, effects from the species are expected at larger distances when c_T is larger as in Fig. 3.

For the influence potential IP we adopted an additive model for all trees of species T in plot k , i.e.

$$\text{IP}_k(q; T, c) = \sum_{z \in T} \Phi(q; c).$$

For each quadrat in the plot, a measure of the influence potential of a specific tree species is calculated. That measure will depend on the distances between the trees of the species and the quadrat, the scale c of the tree species, and the size of those trees; e.g. large trees at short distances and with a high value of c will result in a high value of influence potential. Whether the calculated IP has a positive or negative influence on an un-

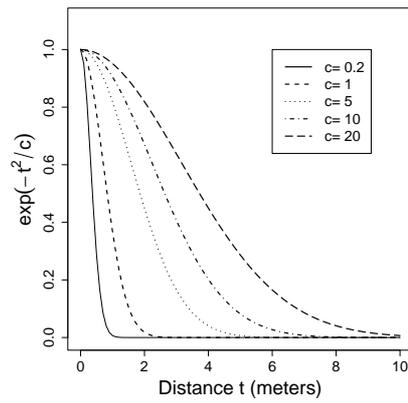


Figure 3: Function $\exp(-t_z(q)^2/c)$ for different values of c . The larger the value of c , the larger the influence domain is.

derstory species, will depend on the relationship between both species; i.e. while for some understory species it is detrimental to have trees close by, for others it might be beneficial. The same IP is related to all understory species observed in the quadrat, since IP is a function of only the trees.

Part I

Edge correction for influence potential

1 INTRODUCTION

The influence potential on a quadrat $IP(Q)$ is calculated as a function of the effects of surrounding trees. The sampling procedure used in the permanent sample plots (PSP) ignored the trees located outside the plot, and consequently the calculation of the influence potential IP may underestimate the true value. Such a situation is more severe when it is the nearby trees that are left out, as those are the ones with greater impact according to the definition of the tree effect function Φ . Therefore the closer a quadrat is to the edge of the plot, the fewer nearby and high-effect trees are taken into account, and the larger the bias.

The same type of problem is also encountered in the analysis of spatial point patterns, where that the information gathered on the events (i.e. the points of the point process) is censored by the boundaries of the study area; this problem is commonly known as edge-effects. The empty space function (F-function), the nearest-neighbor function (G-function), and the K-function (see e.g. Diggle, 1983) are some of the statistics often used to describe relationships among the events in a spatial point process. Several corrections that compensate for the edge effects have been developed for these functions. The trees in the PSP can also be considered as random events from a spatial point process, but the existing corrections are not suitable for this study: in our case we are not concerned with only

the trees, but also with the ecological relation between the trees and the quadrats as quantified by IP.

In this report we propose a correction for IP, which is based on the theory of marked point processes. First we summarize concepts and tools commonly used in the analysis of spatial point processes and include a survey of edge corrections found in the literature. We then present the basic theory of marked point processes and derive the new edge correction. Finally we compare the results of IP with and without the correction for the large and small trees of the PSP.

2 SPATIAL POINT PROCESSES

Spatial point processes deal with the study of events occurring on a n -dimensional space, with $n \geq 2$. The locations of those events are considered to be random, as may happen with trees, cells, stars, etc. Detailed discussions on the analysis of spatial point processes can be found, e.g. in Diggle (1983), Cressie (1991), Stoyan and Stoyan (1994), and Stoyan et al. (1995).

The processes we discuss here are restricted to those in \mathbb{R}^2 , as well as processes where every event within a specified window has been recorded. The point process Ψ is analyzed within that window A , but we assume that the process has been generated on a much larger area. The location of an event $\mathbf{x} \in \Psi$ is determined by its coordinates $\mathbf{x} = (x_1, x_2)$, usually expressed in the Cartesian system.

Two main characteristics can be associated with a process. A point process is said to be stationary or homogeneous if its properties do not vary according to the location, i.e. it is invariant to translation; and it may also be isotropic, if the properties remain the same independent of direction or rotation around the origin.

The first-order property of a spatial pattern is described by the intensity measure Λ :

$$\Lambda(B) = E[\Psi(B)] \quad (1)$$

$$= \int_B \lambda(\mathbf{x}) d\mathbf{x}, \quad (2)$$

where B is a Borel set.

The intensity function λ is defined as

$$\lambda(\mathbf{x}) = \lim_{|d\mathbf{x}| \rightarrow 0} \left\{ \frac{E[\Psi(d\mathbf{x})]}{|d\mathbf{x}|} \right\},$$

where $d\mathbf{x}$ is a disc of infinitesimal area centered at \mathbf{x} , and $|d\mathbf{x}|$ is the area of $d\mathbf{x}$; here we consider $\Psi(d\mathbf{x})$ as a measure that counts the number of events in $d\mathbf{x}$. The intensity measure of a stationary process can be described by $\lambda|A|$, where λ is usually estimated by the total number of events divided by the area of the observation window, $n/|A|$. In a pattern that is completely spatially random, the number of events has a Poisson distribution with parameter $\lambda|A|$, and the locations of the n events are distributed independently and uniformly in A .

The second-order intensity function is defined as

$$\gamma(\mathbf{x}, \mathbf{y}) = \lim_{|d\mathbf{x}|, |d\mathbf{y}| \rightarrow 0} \left\{ \frac{E[\Psi(d\mathbf{x})\Psi(d\mathbf{y})]}{|d\mathbf{x}||d\mathbf{y}|} \right\}.$$

For a stationary process, $\gamma(\mathbf{x}, \mathbf{y}) = \gamma(\mathbf{x} - \mathbf{y})$ only depends on the distance and direction between the events, and not on the absolute location of the events. In addition, if the process is isotropic, then the second order intensity will depend only on the distance and not on the direction.

There are three tools commonly used to characterize a pattern: the G- and F-functions explore first-order characteristics, and the K-function carries out a similar task at the second-order level.

The G-function, or nearest-neighbor function, and the F-function, or empty space function, are based on distances to the nearest neighbor. The first

looks at distances between events, and the second one, at distances between an event and a point, where a point is any location in the area.

The definition of the G- and F- functions are

$$\begin{aligned} G(r) &= P[\text{distance from an arbitrary event} \\ &\quad \text{to nearest other event} \leq r], \text{ and} \\ F(s) &= P[\text{distance from an arbitrary point} \\ &\quad \text{to the nearest event} \leq s]. \end{aligned}$$

These functions may be estimated from the observed pattern by

$$\begin{aligned} \hat{G}(r) &= \frac{\sum_i^n I\{r_i \leq r\}}{n}, \\ \hat{F}(s) &= \frac{\sum_j^m I\{s_j \leq s\}}{m}, \end{aligned}$$

with r_i as the distances from a point \mathbf{x}_i to its nearest event, and s_j as the distance from a sampled point \mathbf{y}_j to its nearest event; n as the total number of events in A ; m as the total number of sampled points in A ; and I as the indicator function counting the number of events or points satisfying the inequalities. Both $G(r)$ and $F(s)$ are estimated for a range of distances, often from 0 to the maximum possible distance within A .

In a stationary Poisson process

$$\begin{aligned} G(r) &= 1 - \exp(-\lambda\pi r^2), \quad r \geq 0; \\ F(s) &= 1 - \exp(-\lambda\pi s^2), \quad s \geq 0. \end{aligned} \tag{5}$$

The K-function was proposed by Ripley (1976) as a tool to summarize second-order properties. It requires a stationary and isotropic process. The function counts the number of other events within a determined radius around an arbitrary event:

$\lambda K(t) = E[\# \text{ of events at distance } \leq t \text{ from an arbitrary event } \mathbf{x}]$.

A naive estimator for $K(t)$ is

$$\hat{K}(t) = \frac{1}{\hat{\lambda}^2 |A|} \sum_{\mathbf{x}} \sum_{\mathbf{y} \neq \mathbf{x}} I\{\|\mathbf{x} - \mathbf{y}\| \leq t\}, \quad (7)$$

where the sum counts the number of ordered pairs of events (\mathbf{x}, \mathbf{y}) , such that \mathbf{y} is at distance t or closer from \mathbf{x} . As before λ can be approximated by $\hat{\lambda} = n/|A|$.

3 EDGE-CORRECTION

When a process is observed through a delimited window, some of the information is omitted. This problem, called edge effects, biases most analyzes of spatial point processes. In the G-function, for example, nearest neighbors that might be situated outside the borders are ignored; the result is that the probabilities are underestimated.

The guard, toroidal and border methods presented next are general strategies for correcting edge effects, and which can be applied in any analysis. The other strategies show how edge effects can be corrected in the estimation of the F-, G-, or K-functions. The emphasis of the presentation is on the reasoning behind each methodology; the formal theoretical derivations can be found in the appropriate references. Ripley (1982, 1988), Cressie (1991), and Stoyan et al. (1995) have more complete surveys of edge corrections.

3.1 GUARD AREA

This approach defines a guard area around the sampling window A . The analysis is carried out on those events that are located inside A . Moreover if information from outside the window is required, e.g. the nearest neighbor, then the events in the guard area are considered. This assures that the information is never censored.

Depending on the size of the sampling window and the guard area, the extra work to measure the location of and gather other information on the events in the guard area may be considered infeasible or unrealistic in terms of resources, especially since the additional information will not be fully utilized during the analysis. A study that used this correction is described by Rathbun and Cressie (1994), where the growth of a longleaf pine forest in space and time was modeled.

3.2 TOROIDAL CORRECTION

For certain shapes of sampling windows, such as rectangles, it is possible to use the principle of a torus. A circular torus is obtained by rotating a circle around a tangent, creating a three-dimensional object similar to a doughnut. Such an object has the particular characteristic of being connected on every side.

If the window is a rectangle, a torus can be created by copying the window around the original (and central) window. In this way the upper edge is connected to the bottom, and the left edge, to the right. The analysis is carried out only for those events in the central window, but any information necessary from outside the edges is taken from the neighboring window.

The approach is difficult to achieve with circles and irregular shapes, but convenient for windows with straight edges. For this reason it is often used in simulation studies.

3.3 ISOTROPIC CORRECTION

This correction was developed by Ripley for the K-function in Eq. 7 (Ripley 1976, 1977). For an event close to the border, some of the other events within distance t may have fallen outside A and not been observed. Ripley's strategy is to give weights to those pairs of events that have been observed, and in this way, estimate how many pairs of events were not recorded. This is achieved by counting every pair of events (\mathbf{x}, \mathbf{y}) $1/k(\mathbf{x}, \mathbf{y})$ times. Considering a circle with center at \mathbf{x} and passing through \mathbf{y} , the weight $k(\mathbf{x}, \mathbf{y})$ is the proportion of length of the circumference that is in-

side A ; i.e.

$$k(\mathbf{x}, \mathbf{y}) = \frac{\text{length}\{A \cap \delta b(\mathbf{x}, \|\mathbf{x} - \mathbf{y}\|)\}}{2\pi\|\mathbf{x} - \mathbf{y}\|}. \quad (8)$$

In Eq. 8, $b(\mathbf{x}, \|\mathbf{x} - \mathbf{y}\|)$ is the circle centered at \mathbf{x} and with radius $\|\mathbf{x} - \mathbf{y}\|$; and $\delta b(\mathbf{x}, \|\mathbf{x} - \mathbf{y}\|)$ represents its border. In general, $k(\mathbf{x}, \mathbf{y})$ is proportional to the probability of observing another event at distance $\|\mathbf{x} - \mathbf{y}\|$ from \mathbf{x} . The weight will be 1 if the complete circle falls inside A , in other words, if the distance between \mathbf{x} and its closest edge is larger than $\|\mathbf{x} - \mathbf{y}\|$.

The corrected estimate of $K(t)$ is

$$\hat{K}^c(t) = \frac{|A|}{n^2} \sum_{\mathbf{x}} \sum_{\mathbf{x} \neq \mathbf{y}} \frac{I\{\|\mathbf{x} - \mathbf{y}\| \leq t\}}{k(\mathbf{x}, \mathbf{y})}. \quad (9)$$

This method assumes that the process is stationary and isotropic, and uses $n/|A|$ as an estimate for λ . Cressie and Brant Collins (2001) applied this type of correction to the global estimate of the so-called product density (derivative of the K-function) and to the product density of local indicators of spatial associations (LISA functions).

3.4 TRANSLATION

Ohser and Stoyan calculated the number of possible translations to correct the K-function (Ripley, 1988; Stoyan and Stoyan, 1994). As in the isotropic correction the observed pairs of events are given weights $g(\mathbf{x}, \mathbf{y})$, which are calculated as the proportion of translations of \mathbf{x} inside the window A , such that \mathbf{y} remains inside A . Alternatively the weights can be measured in terms of rigid motions. In the first way, the weight is calculated as the intersecting area from translating A by \mathbf{x} and \mathbf{y} :

$$g(\mathbf{x}, \mathbf{y}) = \frac{|A_{\mathbf{x}} \cap A_{\mathbf{y}}|}{|A|},$$

where $A_{\mathbf{x}} = \{a + \mathbf{x} : a \in A\}$ is the translation of A by \mathbf{x} , such that \mathbf{x} lies at the origin of A ; and analogously for \mathbf{y} . To apply this method, stationarity is necessary but not isotropy. The estimate of the K-function is calculated as in Eq. 9, but $k(\mathbf{x}, \mathbf{y})$ is substituted by $g(\mathbf{x}, \mathbf{y})$.

The same idea was applied by Fiksel (1988) to correct kernel-estimates of the product density function and to the nearest-neighbor distance distribution; Gavrikov and Stoyan (1995) applied it to the product density of a marked point process; and Capobianco and Renshaw (1998) employed it also in a marked point process case, but to correct the correlation function for marks. In both (Gavrikov and Stoyan, 1995), and (Capobianco and Renshaw, 1998) the correction is carried out for a process in a rectangular window, where the area of the intersection is easy to calculate. No application of this method has been found for other shapes in two dimensions, although Stein et al. (2000) used and modified it for a process on \mathbb{R}^1 .

3.5 BORDER METHOD

This correction approach has been mainly applied to the G- and F-functions, but it is general and intuitive enough to be applied in any analysis as long as enough events exist in the window. Ripley (1988), for example, refers to it in connection with the K-function. This procedure was first mentioned in (Ripley, 1977) and then explicitly presented by Diggle (1979).

The idea behind this approach is to create a variable guard area inside the window. When it is applied to $G(r)$, for example, only the events that are at least at distance r from the edge are used in the estimate. If $d_{\mathbf{x}}$ is the distance from \mathbf{x} to its nearest neighbor, and $h_{\mathbf{x}}$ is the distance from \mathbf{x} to the nearest edge of the window, then the corrected version of the G-function is computed as

$$\hat{G}_1^c(r) = \frac{\sum_{\mathbf{x}} I\{d_{\mathbf{x}} \leq r, r < h_{\mathbf{x}}\}}{\sum_{\mathbf{x}} I\{r < h_{\mathbf{x}}\}}. \quad (11)$$

This represents the proportion of events with nearest neighbors at distance less than or equal to r , from those located at a distance greater than r from the nearest edge. In analogous way the F-function can be corrected,

Table I.1: Ordering of r , $h_{\mathbf{x}}$, and $d_{\mathbf{x}}$: Use in $\hat{G}_1^c(r)$, $\hat{G}_2^c(r)$, and $\hat{G}_3^c(r)$ (1 if yes, 0 if no), and presence of nearest neighbor within distance r , $I\{d_{\mathbf{x}} \leq r\}$ (1 if yes, 0 if no, ? if unknown).

| Case | Description | $\hat{G}_1^c(r)$ | $\hat{G}_2^c(r)$ | $\hat{G}_3^c(r)$ | $I\{d_{\mathbf{x}} \leq r\}$ |
|------|--|------------------|------------------|------------------|------------------------------|
| 1 | $r \leq h_{\mathbf{x}} < d_{\mathbf{x}}$ | 1 | 0 | 1 | 0 |
| 2 | $r \leq d_{\mathbf{x}} < h_{\mathbf{x}}$ | 1 | 1 | 1 | 0 |
| 3 | $d_{\mathbf{x}} \leq r < h_{\mathbf{x}}$ | 1 | 1 | 1 | 1 |
| 4 | $d_{\mathbf{x}} \leq h_{\mathbf{x}} < r$ | 0 | 1 | 1 | 1 |
| 5 | $h_{\mathbf{x}} < d_{\mathbf{x}} \leq r$ | 0 | 0 | 1 | 1 |
| 6 | $h_{\mathbf{x}} < r < d_{\mathbf{x}}$ | 0 | 0 | 0 | ? |

From: Doguwa (1989).

except using distances from a sampled point.

Hanisch (1984) modified Eq. 11 to include events whose nearest neighbor is known, i.e. those events whose nearest neighbor is closer than the nearest edge; thus

$$\hat{G}_2^c(r) = \frac{\sum_{\mathbf{x}} I\{d_{\mathbf{x}} \leq r, d_{\mathbf{x}} < h_{\mathbf{x}}\}}{\sum_{\mathbf{x}} I\{d_{\mathbf{x}} < h_{\mathbf{x}}\}}.$$

Doguwa (1989) analyzed both methods and stated that these estimates actually count over different subsets of events. Each of them includes different situations: e.g. $\hat{G}_1^c(r)$ includes the case when $r \leq h_{\mathbf{x}} \leq d_{\mathbf{x}}$, but not $\hat{G}_2^c(r)$; and the latter uses $d_{\mathbf{x}} \leq h_{\mathbf{x}} < r$, but not the former.

The disadvantage of either formulation is that number of events decrease as r increases, thus making the estimates more variable for larger r .

3.6 AREA

The discrepancies between $\hat{G}_1^c(r)$ and $\hat{G}_2^c(r)$ in terms of the subsets taken into account, led Doguwa and Upton (1990) to propose a different type of edge correction. They included in their estimator all six possible situations that can occur between r , $d_{\mathbf{x}}$, and $h_{\mathbf{x}}$; see Table I.1.

In the first five cases, it is clear whether the nearest neighbor is within distance r or not. The only situation when the outcome is not obvious is case 6, since the nearest neighbor might be outside the window. The correction consists in estimating the probability of finding the nearest event outside the window. If a circle $b(\mathbf{x}, r)$ is defined for case 6, then a part of that circle will lay outside the window. Using notation based on Floresroux and Stein (1996), let $I(\mathbf{x}, r)$ represent the area of the circle inside the window, and $O(\mathbf{x}, r)$, the area of the circle outside. By applying Eq. 5, the probability that the nearest event is in $O(\mathbf{x}, r)$ for a stationary Poisson process is $\hat{H}_1(\mathbf{x}, r) = 1 - \exp(-\lambda|O(\mathbf{x}, r)|)$.

The corrected estimate for the G-function is then

$$\begin{aligned} \hat{G}_3^c(r) &= \frac{1}{n} \sum_{\mathbf{x}} I\{d_{\mathbf{x}} < r\} + \\ &+ \frac{1}{n} \sum_{\mathbf{x}} (1 - I\{d_{\mathbf{x}} < r\}) (1 - I\{r < h_{\mathbf{x}}\}) \hat{H}_1(\mathbf{x}, r). \end{aligned} \quad (13)$$

The first term considers those events whose nearest neighbor is known (cases 1–5 in Table I.1). The second term is the probability of finding the nearest neighbor outside A and is applied for case 6, when it is possible that the nearest neighbor is outside.

3.7 ANALOGOUS EVENTS

If the underlying process is not stationary Poisson, $\hat{G}_3^c(r)$ may be biased. For that reason Floresroux and Stein (1996) looked for another way to estimate $H(\mathbf{x}, r)$ that would not depend on that assumption. They solved the situation by looking for analogous points in the window. Let $I(\mathbf{y}|\mathbf{x}, r)$ be $I(\mathbf{x}, r)$ translated to \mathbf{y} . Then the event \mathbf{y} is analogous to \mathbf{x} (where \mathbf{x} falls into case 6), if $b(\mathbf{y}, r)$ is completely inside the window, and $I(\mathbf{y}|\mathbf{x}, r)$ has no other event except \mathbf{y} . The probability of \mathbf{x} finding its nearest neighbor in $O(\mathbf{x}, r)$ is calculated as the proportion of analogous events that have their corresponding nearest neighbor in $O(\mathbf{y}|\mathbf{x}, r)$. This probability is expressed as

$$\hat{H}_2(\mathbf{x}, r) = \frac{\sum_{\mathbf{y}} T(\mathbf{x}, \mathbf{y}) I\{d_{\mathbf{y}} < r\}}{\sum_{\mathbf{y}} T(\mathbf{x}, \mathbf{y})},$$

where $T(\mathbf{x}, \mathbf{y})$ is 1 if \mathbf{x} and \mathbf{y} are analogous, and 0 otherwise. This method can also be used for isotropic processes, by including rotations as another way of determining analogy. The estimate of $G(r)$ is carried out as in Eq. 13, but using $\hat{H}_2(\mathbf{x}, r)$ instead.

The corrected estimator of Floresroux and Stein for the G-function has the advantage that it does not need to assume any process in particular, and it uses the information more efficiently than the other corrections.

3.8 KAPLAN-MEIER ESTIMATE

Baddeley and Gill (1997) suggested a new approach to edge correction, by comparing this problem to that in survival analysis. The main idea, e.g. in the case of the F-function, is to think of distance to the nearest event as failure, and the distance to the edge as a censoring time.

The Kaplan-Meier estimator for the failure distribution $B(t)$ is calculated as

$$\hat{B}(t) = 1 - \prod_j \left(1 - \frac{d_j}{r_j}\right),$$

where d_j is the number of observed failures at $t = j$, and r_j is the number of individuals who could have failed at $t = j$, also referred to as those in hazard.

In a rather straightforward way, $F(s)$ can then be estimated by

$$\hat{F}_1^c(s) = 1 - \prod_s \left(1 - \frac{\sum_{\mathbf{x}} I\{s_{\mathbf{x}} = s, s_{\mathbf{x}} \leq h_{\mathbf{x}}\}}{\sum_{\mathbf{x}} I(\min\{s_{\mathbf{x}}, h_{\mathbf{x}}\} \geq s)}\right).$$

The observed failures is the set of points whose nearest event is within

distance s and is closer than the border. The set in hazard are those with distance to the border or to the nearest event larger or equal to s .

A similar approach can be used for estimating the G- and K-functions. An application to a three-dimensional case is presented by Reed and Howard (1997).

4 EDGE EFFECTS IN THE INFLUENCE POTENTIAL

In the PSP the calculations of IP are biased since not all affecting trees are included, resulting this in edge effects. As the trees located outside the boundaries of the plot are ignored, the real influence potential is underestimated. Furthermore the tree effect Φ function weighs nearby trees more heavily than those further away. Therefore the edge effect problem is more pronounced the closer a quadrat is to the border of the plot, as in these cases more of the trees at shorter distances are missing.

During the measuring campaign of the PSP, no guard area was considered. Additionally the plots are too small and do not have enough trees to be able to use the border method, and the toroidal principal is difficult to implement for circular plots. The other edge corrections for the G- and K-functions concentrate on event to event relationships, while this study is interested in the relationship between an event (tree) and an arbitrary point (quadrat). This suggests that a correction applicable to the F-function, such as the Kaplan-Meier estimates, could be appropriate.

Nevertheless, since we needed to focus on the IP function, we decided to develop a correction specifically for this problem which would adjust the observed IP. The basic idea behind the proposed correction is to assume a stationary process in and around the plot and to calculate the expected value of IP in the area outside the plot. This expectation is then added to the observed IP to obtain the corrected and final IP. The theory of marked point processes provides the tools to compute the expectation, by taking each tree as an event and the diameter at breast height as the mark. Before describing the proposed edge correction, however, we review some basic results for marked point processes.

5 MARKED POINT PROCESSES

Theory on marked point processes can be found e.g. in Stoyan and Stoyan (1994) and Stoyan, Kendall, and Mecke (1995). Penttinen, Stoyan, and Henttonen (1992), Gavrikov and Stoyan (1995), and Stoyan and Penttinen (2000) review the use of spatial point processes in forestry, including marked point processes.

A marked point process can be considered as a spatial point process with an additional dimension that contains information regarding each event. The information may be continuous or discrete, e.g. diameters of trees, species of trees, volume of a particle. The marked process Ψ_M can be expressed as $\Psi_M = \{[\mathbf{x}_n; m_n]\}$, where the \mathbf{x}_n s are the locations of the events and m_n the marks of those events.

Stationarity and isotropy are defined as for the unmarked case. The intensity $\Lambda_M(B \times C)$ must take into account the mark distribution \mathbf{M} , which reflects the probabilities that a mark from a typical event lies in the mark set C . For the stationary case the intensity is defined as

$$\begin{aligned} \Lambda_M(B \times C) &= E[\Psi_M(B \times C)] \\ &= \lambda |B| \mathbf{M}(C). \end{aligned} \quad (17)$$

Other statistics related to the marks are the mean mark \bar{m} , defined by the distribution function \mathbf{M} as

$$\bar{m} = \int_{-\infty}^{\infty} m \, d\mathbf{M}(m), \quad (18)$$

and the mark sum measure S_m ,

$$S_m(B) = \sum_{[\mathbf{x}; m] \in \Psi_M} m I\{\mathbf{x} \in B\}, \quad (19)$$

which represents the sum of the marks for all events in B . The expected

value for S_m in the stationary case is $\lambda\bar{m}|B|$.

The estimate of the mean mark for a stationary marked point process observed in the window A is computed from

$$\hat{m} = \frac{S_m(A)}{n}, \quad (20)$$

and n is the number of events in A . In general the estimated mean mark \hat{m} is not always unbiased, but $\lambda\hat{m}$ is unbiased for $\lambda\bar{m}$ (Stoyan and Stoyan, 1994, pg. 278). The mark distribution $\mathbf{M}(u)$ for a continuous mark can be estimated by

$$\hat{\mathbf{M}}(u) = \frac{I\{m \leq u : [\mathbf{x}; m] \in A\}}{n},$$

that is, the proportion of events in A that have a mark less than or equal to u .

Another useful tool is the Campbell theorem. This theorem defines the expected value of the sum of a non-negative measurable function, and applies to a point process in any dimension; see Stoyan and Stoyan (1994) and Stoyan et al. (1995) for the application to a spatial point processes, and Kingman (1993) for the one-dimensional case and detailed proofs. In general, the theorem for a spatial point process can be seen as

$$E \left[\sum_{\mathbf{x} \in \Psi} f(\mathbf{x}) \right] = \int f(\mathbf{x}) \Lambda(d\mathbf{x}).$$

From Eq. 2, the expectation for a stationary process can be further simplified to

$$E \left[\sum_{\mathbf{x} \in \Psi} f(\mathbf{x}) \right] = \lambda \int f(\mathbf{x}) d\mathbf{x}.$$

With a parallel formulation, the Campbell theorem can also be obtained in

a marked point process for a function f of the events and marks (Stoyan and Stoyan, 1994; Stoyan et al., 1995). In the stationary case, and making use of Eq. 17:

$$\begin{aligned}
 E \left[\sum_{[\mathbf{x}; m] \in \Psi_M} f(\mathbf{x}; m) \right] &= \\
 &= \int f(\mathbf{x}, m) \Lambda_M d(\mathbf{x}, m) \\
 &= \lambda \int \int f(\mathbf{x}, m) d\mathbf{M}(m) d\mathbf{x} \quad (23)
 \end{aligned}$$

6 DERIVATION OF EDGE CORRECTION

The influence potential IP on quadrat Q is defined as a function of the tree effect Φ_z , where z are the trees and D is the diameter at breast height:

$$\begin{aligned}
 \text{IP}(Q, c) &= \sum_z \Phi_z(Q) \quad (24) \\
 &= \sum_z D_z \exp\left(-\frac{\|z - Q\|^2}{c}\right) \\
 &= \sum_z D_z \exp\left(-\frac{t_z(Q)^2}{c}\right);
 \end{aligned}$$

here we continue to use the simpler notation from the last equation where $\|z - Q\|^2 = t_z(Q)^2$.

The correction of IP for edge effects is carried out by finding the expected influence potential of the unobserved trees, and then adding this estimate to the observed IP; i.e.

$$\text{IP}^c = \text{IP}\{\text{observed}\} + \text{IP}\{\text{unobserved}\}. \quad (25)$$

In theory IP can be calculated for an infinite area around the quadrat. For the purpose of the correction, however, we define a circle $b(Q, s(c))$ of significant influence around quadrat Q , centered at the quadrat and with radius $s(c)$. The most important contributions to the IP of the quadrat come from trees located in this circle. The radius $s(c)$ is obtained by defining first a minimum significant effect Φ/D , say 0.01. This means that, independently of their size, only those trees with effects larger than 0.01 are considered to be relevant. Assuming that the parameter c has been previously determined, then $s(c)$ is obtained from

$$\begin{aligned} 0.01 &= \exp\left(-\frac{s(c)^2}{c}\right) \\ \Rightarrow s(c) &= \sqrt{-c \ln(0.01)}. \end{aligned}$$

Depending both on the position of the quadrat with respect to the border of the plot, and on the radius of minimum significant effect, $b(Q, s(c))$ might be completely contained inside the plot or not. If $b(Q, s(c))$ is inside the plot, then we have all the information necessary for determining IP for that quadrat. If $b(Q, s(c))$ is partly inside and partly outside the plot, we further call those two areas $I(Q, c)$ and $O(Q, c)$. More formally,

$$\begin{aligned} I(Q, c) &= b(Q, s(c)) \cap b(O, R) \\ O(Q, c) &= b(Q, s(c)) \setminus b(O, R), \end{aligned}$$

where $b(O, R)$ represents the plot centered at the origin and with radius R . Figure 1 illustrates all the necessary concepts.

Our objective is to determine the expected value of IP in $O(Q, c)$, and this is possible by applying the Campbell theorem for marked point processes. The possibility of using this theorem depends heavily on the fact that IP is defined as an additive model (Eq. 24). Furthermore, to be able to use the theorem, we consider the trees z as events in a spatial point process, and the diameter D as the mark of the event. If the process is stationary as well, then the Campbell theorem as defined in Eq. 23 serves as a way to calculate the expected value of IP in $O(Q, c)$ by

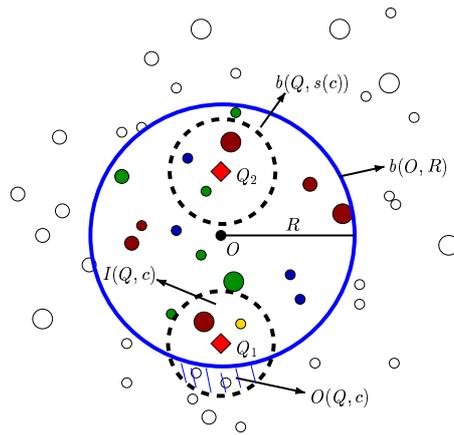


Figure 1: Edge correction concepts: The small circles represent trees: the filled ones have been observed and the empty ones not. The diamonds are the quadrats. The circle of influence $b(Q, s(c))$ (dashed line) is of the same size for both quadrats. Since Q_1 is closer to the plot border, its circle of influence extends beyond the borders and therefore a correction for IP is needed. The correction will estimate the expected value of IP in $O(Q, c)$ (dashed area). The upper quadrat Q_2 has a circle of influence completely inside the plot, thus all the information necessary for calculating IP has been measured and no correction is required.

$$\begin{aligned}
E[\text{IP}(O(Q, c))] &= \\
&= E \left[\sum_{[\mathbf{x}; m] \in \Psi} f(\mathbf{x}; m) \right] \\
&= E \left[\sum_{[z; \mathbf{D}] \in O(Q, c)} \Phi_z(Q) \right] \\
&= \lambda \int_{z \in O(Q, c)} \int_{\mathbf{D}} \Phi_z(Q) d\mathbf{M}(\mathbf{D}) dz \\
&= \lambda \int_{z \in O(Q, c)} \int_{\mathbf{D}} \mathbf{D} \exp\left(-\frac{t_z(Q)^2}{c}\right) d\mathbf{M}(\mathbf{D}) dz.
\end{aligned}$$

All the terms in the previous equation are only functions of the events or of the marks, and therefore it is possible to separate the integrals as

$$\begin{aligned}
E[\text{IP}(O(Q, c))] &= \\
&= \lambda \int_{\mathbf{D}} \mathbf{D} d\mathbf{M}(\mathbf{D}) \int_{z \in O(Q, c)} \exp\left(-\frac{t_z(Q)^2}{c}\right) dz.
\end{aligned}$$

The first integral represents the mean mark of \mathbf{D} as expressed in Eq. 18; thus

$$E[\text{IP}(O(Q, c))] = \lambda \bar{\mathbf{D}} \int_{z \in O(Q, c)} \exp\left(-\frac{t_z(Q)^2}{c}\right) dz.$$

The parameter λ can be approximated in the usual way with $\hat{\lambda} = n/|A|$. For $\bar{\mathbf{D}}$, we take the estimator in Eq. 20, $S_{\mathbf{D}}(A)/n$. Putting these two estimators together, we have that

$$\hat{\lambda} \hat{\bar{\mathbf{D}}} = \frac{\sum_{z \in A} \mathbf{D}_z}{|A|}.$$

The estimate of the expected value of IP in $O(Q, c)$ is then

$$\begin{aligned} \hat{E}[\text{IP}(O(Q, c))] &= \hat{\lambda} \hat{\bar{D}} \int_{z \in O(Q, c)} \exp\left(-\frac{t_z(Q)^2}{c}\right) dz \\ &= \frac{\sum_{z \in A} D_z}{|A|} \int_{z \in O(Q, c)} \exp\left(-\frac{t_z(Q)^2}{c}\right) dz. \end{aligned} \quad (27)$$

We can express the previous equation also in polar coordinates to facilitate the calculations of the boundaries of $O(Q, c)$:

$$\begin{aligned} \hat{E}[\text{IP}(O(Q, c))] &= \\ &= \frac{\sum_{z \in A} D_z}{|A|} \int_{\theta(O(Q, c))} \int_{r(O(Q, c))} r \exp\left(-\frac{r^2}{c}\right) dr d\theta \end{aligned} \quad (28)$$

Several advantages can be pointed out for this correction. It only requires stationarity of the process in and around the plot. Moreover, if information is available on the type of process that generated the trees in the plot, then a better estimate for λ can be utilized. In similar way, information on the distribution of the diameter may be available and a better estimate of $\hat{\bar{D}}$ can be included in the correction. Furthermore, IP may be interesting for a specific group of events, e.g. tree species, and this method allows for the correction to be calculated for each group separately, using their individual $\hat{\lambda}$ and $\hat{\bar{D}}$; this may be especially significant if the intensity or mean diameter differ greatly among the groups.

In this study the tree species was considered a relevant characteristic, so we wished to calculate IP and its correction for each species. The large number of plots, however, prevented us from making a detailed study of the process and distribution of the diameter in each plot. Thus, to estimate $\lambda \bar{D}$ for each species and plot, its corresponding $\sum_z D_z / |A|$ was calculated. Both the sum of the diameters and the area of the A were based on the entire plot instead of only $I(Q, c)$, since in this way we could take into account all the trees included in the uncorrected IP.

For the tree species T , then we define the influence potential on the quadrat Q , coming from the trees $k(T)$ in plot k and of species T as

$$\begin{aligned} \text{IP}_{k(T)}(Q) &= \text{IP}\{\text{observed trees of species } T \text{ in plot } k\} + \\ &+ \text{IP}\{\text{correction for species } T \text{ in plot } k\} \\ &= \sum_{z \in k(T)} D_z \exp\left(-\frac{t_z(Q)^2}{c}\right) + \\ &+ \frac{\sum_{z \in k(T)} D_z}{|A|} \int_{z \in O(Q,c)} \exp\left(-\frac{t_z(Q)^2}{c}\right) dz. \end{aligned}$$

If we assume independence among species, we can calculate a general IP for a quadrat based on the sum of the different IP_T 's, that is, as the sum of the influence potentials from all the species.

7 APPLICATION

The edge correction for IP was applied to the data set of big and small trees separately. The big trees were observed in an circle of radius 9.77 m, and the small trees in a plot with a radius of 5.64 m. The quadrats were therefore located at different positions with relation to the borders of these two plots, which represented different challenges when implementing of the edge correction.

7.1 BIG TREES

The big trees, those with diameter greater than 10.5 cm, were recorded in a circular plot of radius 9.77 m. The six quadrats were located at 3, 6, and 8 m away from the plot center, thus inside the plot. Three possible situations can occur in terms of $O(Q, c)$ and $b(O, R)$; these depend on the value of c that defines $s(c)$:

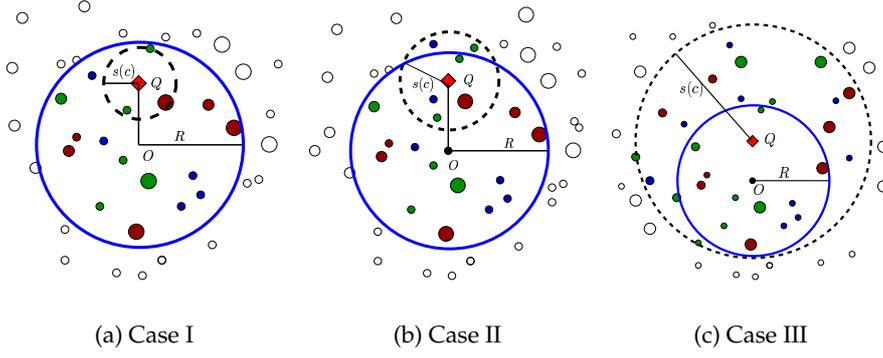


Figure 2: Cases I-III: The plot is represented by the large circle with solid line (center O and radius R), and the circle of influence by a dashed line (center Q and radius $s(c)$). The small circles are hypothetical trees: the filled ones are relevant trees, i.e. they are inside the circle of influence or the plot. Case II and III require an edge correction, since not all the relevant trees in the circle of influence have been observed.

- I. $s(c) < (R - \overline{OQ})$;
- II. $(R - \overline{OQ}) < s(c) < (R + \overline{OQ})$;
- III. $(R + \overline{OQ}) < s(c)$.

Figure 2 illustrate the three cases. Case I does not require any correction, because all the trees in $b(Q, s(c))$ needed for calculating IP have been observed. In case II $O(Q, c)$ has the shape of a crescent, and in case III that of an annulus of variable width. We present here explicitly the calculations of the boundaries of $O(Q, c)$ for case II; the approach for solving case III is explained further on, and the formulas for both are included in the Appendix.

One convenient way of evaluating the integral is by using polar coordinates as in Eq. 28. The limits of the integrals for case II are shown in detailed manner in Figure 3. The angle θ_{\max} represents the angle of intersection between the circle of influence and the plot, and $r(\theta)$ is the distance between the quadrat and the edge of the plot at angle θ . It follows that the integral term of the expectation can be calculated from

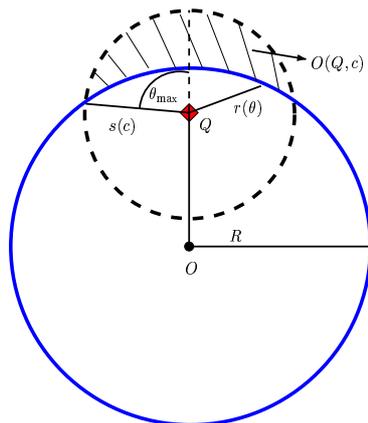


Figure 3: Geometry of the edge correction for case II. The angle θ_{\max} measures the intersection between the plot (centered at O and with radius R) and the circle of influence (centered at Q and with radius $s(c)$); $r(\theta)$ is the distance between the quadrat and the border of the plot at angle θ .

$$\begin{aligned}
& \int_{z \in O(Q,c)} \exp\left(-\frac{t_z(Q)}{c}\right) dz = \\
&= \int_{\theta(O(Q,c))} \int_{r(O(Q,c))} r \exp\left(-\frac{r^2}{c}\right) dr d\theta \\
&= \int_{-\theta_{\max}}^{\theta_{\max}} \int_{r(\theta)}^{s(c)} r \exp\left(-\frac{r^2}{c}\right) dr d\theta \\
&= \int_{-\theta_{\max}}^{+\theta_{\max}} \left[\frac{c}{2} \exp\left(-\frac{r(\theta)^2}{c}\right) - \frac{c}{2} \exp\left(-\frac{s(c)^2}{c}\right) \right] d\theta \\
&= \frac{c}{2} \int_{-\theta_{\max}}^{+\theta_{\max}} \exp\left(-\frac{r(\theta)^2}{c}\right) d\theta - c \theta_{\max} \exp\left(-\frac{s(c)^2}{c}\right); \quad (29)
\end{aligned}$$

and the expectations of $O(Q, c)$ for quadrat Q in plot k and tree species T is obtained from

$$\hat{E}[\text{IP}_{k(T)}(O(Q, c))] = \frac{\sum_{z \in k(T)} D_z}{|b(O, R)|} \cdot \left\{ \frac{c}{2} \int_{-\theta_{\max}}^{+\theta_{\max}} \exp\left(-\frac{r(\theta)^2}{c}\right) d\theta - c \theta_{\max} \exp\left(-\frac{s(c)^2}{c}\right) \right\}.$$

In this equation the sum of the diameters is computed over the $n_{k(T)}$ trees of species T in the plot, and $|b(O, R)|$ is the area of the plot and equal to 300 m². The remaining integral must be solved numerically. By applying the law of cosines, $r(\theta)$ and θ_{\max} can be calculated from

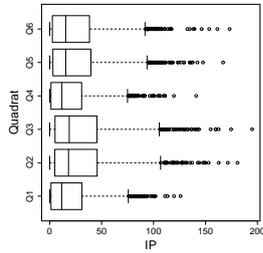
$$\begin{aligned} R^2 &= r(\theta)^2 + \overline{OQ}^2 - 2r(\theta)\overline{OQ}\cos(\pi - \theta) \\ \Rightarrow r(\theta) &= \sqrt{R^2 - \overline{OQ}^2 \sin^2 \theta - \overline{OQ}\cos(\theta)}; \end{aligned} \quad (30)$$

and

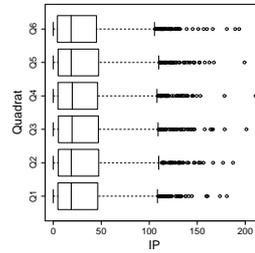
$$\theta_{\max} = \arccos\left(\frac{R^2 - \overline{OQ}^2 - s(c)^2}{2 \cdot \overline{OQ} \cdot s(c)}\right). \quad (31)$$

Equations 30 and 31 allow the estimate of the expectation to be described now in terms of known parameters: $|b(O, R)|$, c , R , and $s(c)$ are the same for all quadrats, plots, and species; \overline{OQ} depends on the quadrat; and $\sum_z D_z$ is specific for each species and plot.

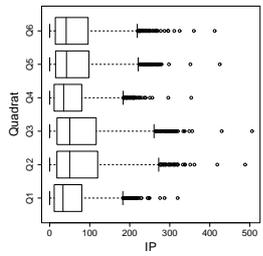
The results of cases II and III can be seen in Figure 4, where boxplots of IP using c at 20 and 75 are shown with and without the correction. The computations were carried out on 1240 plots with one stand and with quadrats located on mineral soils. For each tree species studied, i.e. pine, spruce, and birch, IP and the correction were calculated and then added together to obtain one corrected IP per quadrat. For each plot, the calculations were carried out for all the six quadrats, even when less quadrats had been observed in the field. In this way, every quadrat had the same number of IP measurements; this was possible since the formulas for IP



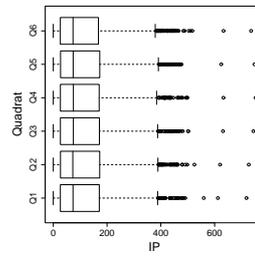
(a) $c = 20$ ($s(c) = 9.5$ m): No correction



(b) $c = 20$ ($s(c) = 9.5$ m): With correction



(c) $c = 75$ ($s(c) = 18.6$ m): No correction



(d) $c = 75$ ($s(c) = 18.6$ m): With correction

Figure 4: Big trees ($D \geq 10.5$ cm) : IP by quadrat with c at 20 and 75, corresponding to cases II (a and b) and III (c and d). A total of 1240 plots with one stand and with quadrats in mineral soils were analyzed assuming that six quadrats were measured in each plot. The calculation were conducted separately for spruce, pine, and birch; then the three corrected IP were added up to obtain one IP for each quadrat and plot.

and the correction do not require any information from the quadrat beside \overline{OQ} .

In Figs. 4(a) and 4(c), the distributions of quadrats 1 and 4 concentrate on a smaller scale compared to the other quadrats. The reason is that those quadrats are positioned closest to the edge of the plot, so fewer trees are taken into account when calculating the influence potential. This bias is more subtle for quadrats 5 and 6 situated further inside the plot. Quadrats 2 and 3, close to the center, also suffer of edge effects with these values of c , because the $b(Q, s(c))$ extends beyond the plot borders. For these quadrats, however, $O(Q, c)$ is smaller than for the other quadrats, so a smaller adjustment is required.

Figures 4(b) and 4(d) show the distributions of all quadrats when the IP has been corrected by adding the expectation over $O(Q, c)$. Although the correction cannot reproduce observations on the tail, it adds an important proportion of tree effects. The boxes are now almost perfectly aligned, as was anticipated since IP for all six quadrats should be in the same range. These results confirm the presence of edge effects and proves the effectiveness of the proposed correction to reduce bias.

The weight of the correction was also evaluated as the percentage it represented in the corrected IP. At $c = 20$, the mean percentages were 51% for quadrats 1 and 4; 30% and 32% for quadrats 5 and 6, and 9% for both quadrats 2 and 3. The mean percentages increased for c at 75: 57% at quadrats 1 and 4, 46% at quadrats 5 and 6, and 33% at quadrats 2 and 3. Again we observe that the correction needed is larger the closer the quadrat is to the border. Furthermore the correction is also larger as c , and therefore $O(Q, c)$, increases.

Some plots required a correction of 100% of the adjusted IP. One of these extreme cases is shown in Figure 5, where only one birch tree was observed. This tree was situated in the exact opposite direction to quadrat 1, and thus the influence of birch on quadrat 1 was very small, $1.7 \cdot 10^{-6}$ with $c = 20$. After computing the correction with $\hat{\lambda} = 1/300$ and $\sum D = 11$ cm, the new adjusted IP was 0.8. The correction was consequently 100% for birch, while for Norway spruce it yielded 73%. Quadrat 4, which was more closely surrounded by trees of both species, required a correction of only 8% for birch and 24% for spruce. The correction for both quadrat 1 and 4 is the same; the difference lies in the observed IP.

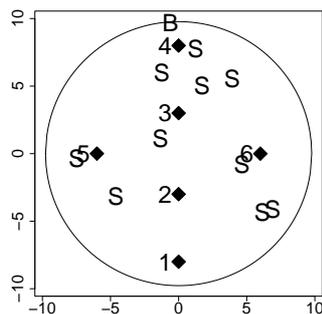


Figure 5: Plot 815904: "S"=Norway spruce, "B"=Birch, \diamond =quadrat; not at scale.

7.2 SMALL TREES

The application of the IP correction to the small trees ($4.5 \text{ cm} \leq D < 10.5 \text{ cm}$) of the PSP represented a greater challenge than the big trees. The plot in this case has a radius of 5.64 m, which means that quadrats 1, 4, 5, and 6 are located outside. We identified seven possible cases; the first three cases apply only to quadrats 2 and 3 and are equivalent to those for the big trees, since these quadrats lay inside the plot.

For quadrats 2 and 3, cases are

- I. $s(c) < (\overline{OQ} - R)$;
- II. $(\overline{OQ} - R) < s(c) < (\overline{OQ} + R)$;
- III. $(R + \overline{OQ}) < s(c)$.

Four additional cases for quadrats 1, 4, 5, and 6 were identified (see Fig. 6):

- IV. $s(c) < (\overline{OQ} - R)$;
- V. $(\overline{OQ} - R) < s(c) < \overline{OQ}$;
- VI. $\overline{OQ} < s(c) < (\overline{OQ} + R)$;
- VII. $(\overline{OQ} + R) < s(c)$.

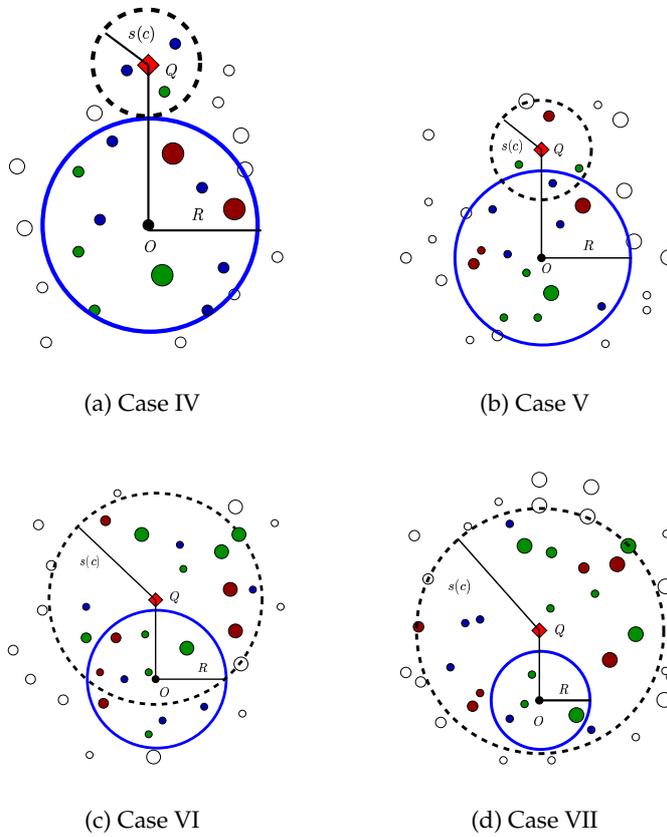


Figure 6: Cases IV-VII. The plot is represented by a circle with solid line (center O and radius R), and the circle of influence by a dashed line (center Q and radius $s(c)$). The small circles are hypothetical trees: the filled ones are relevant trees, i.e. they are inside the circle of influence or the plot.

Case IV is particular for the quadrats 1 and 4: it happens when the $s(c)$ is small enough that the circle of significant influence does not intersect with the plot at all. That means that $b(Q, s(c)) = O(Q, c)$, and the correction integrates over the complete area of $b(Q, s(c))$. This does not happen to quadrats 5 and 6, because they are too close to the border of the small plot.

The integral for cases V, VI, and VII are more easily implemented as the difference of the integrals over $b(Q, s(c))$ and over $b(Q, s(c)) \cap b(O, R)$:

$$\begin{aligned} & \int_{z \in O(Q, c)} \exp\left(-\frac{t_z(Q)^2}{c}\right) dz = \\ &= \int_{z \in b(Q, s(c))} \exp\left(-\frac{t_z(Q)^2}{c}\right) dz - \\ & - \int_{z \in \{b(Q, s(c)) \cap b(O, R)\}} \exp\left(-\frac{t_z(Q)^2}{c}\right) dz \end{aligned}$$

The formulas and limits for all cases are included in the Appendix, but here we will show in detail the procedure for case VI. Figure 7 presents the geometrical details for this case. The integral for $b(Q, s(c))$ is straightforward: it is a circle centered at the quadrat and with constant radius, such that

$$\begin{aligned} & \int_{z \in b(Q, s(c))} \exp\left(-\frac{t_z(Q)^2}{c}\right) dz = \\ &= \int_{-\pi}^{+\pi} \int_0^{s(c)} r \exp\left(-\frac{r^2}{c}\right) dr d\theta \\ &= \pi c \left[1 - \exp\left(-\frac{s(c)^2}{c}\right)\right]. \end{aligned} \tag{32}$$

The area of $b(Q, s(c)) \cap b(O, R)$ is separated into two by the intersection of $b(O, R)$ and $b(Q, s(c))$. The lateral regions require two radii, one from

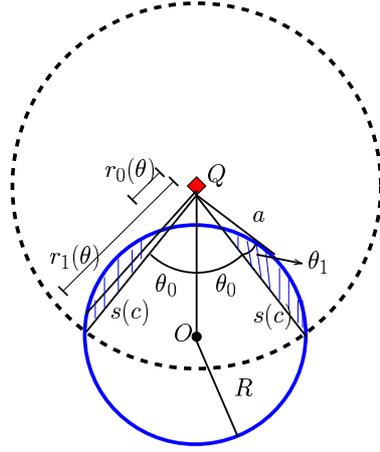


Figure 7: Geometry of the edge correction for case VI. The angle θ_0 measures the intersection between the plot (centered at O and with radius R) and the circle of influence (centered at Q and with radius $s(c)$); $r_0(\theta)$ is the distance between the quadrat and the closest border of the plot at angle θ ; $r_1(\theta)$ is the distance between the quadrat and the furthest border of the plot at angle θ ; a is the tangent to the plot.

the quadrat to the closest border of $b(O, R)$, and another to the furthest border; the central region is integrated between $-\theta_0$ to $+\theta_0$ and from $r_0(\theta)$ to $s(c)$. Thus the area of intersection between $b(O, R)$ and $b(Q, s(c))$ can be obtained from

$$\begin{aligned}
 & \int_{z \in \{b(Q, s(c)) \cap b(O, R)\}} \exp\left(-\frac{t_z(Q)^2}{c}\right) dz = \\
 & = 2 \int_{\theta_0}^{\theta_0 + \theta_1} \int_{r_0(\theta)}^{r_1(\theta)} r \exp\left(-\frac{r^2}{c}\right) dr d\theta \\
 & + \int_{-\theta_0}^{+\theta_0} \int_{r_0(\theta)}^{s(c)} r \exp\left(-\frac{r^2}{c}\right) dr d\theta \quad (33)
 \end{aligned}$$

The first integral, i.e. the laterals, results in

$$\begin{aligned}
& \int_{\theta_0}^{\theta_0+\theta_1} \int_{r_0(\theta)}^{r_1(\theta)} r \exp\left(-\frac{r^2}{c}\right) dr d\theta = \\
& = \frac{c}{2} \int_{\theta_0}^{\theta_0+\theta_1} \exp\left(-\frac{r_0(\theta)^2}{c}\right) d\theta - \frac{c}{2} \int_{\theta_0}^{\theta_0+\theta_1} \exp\left(-\frac{r_1(\theta)^2}{c}\right) d\theta;
\end{aligned}$$

and the second in

$$\begin{aligned}
& \int_{-\theta_0}^{+\theta_0} \int_{r(\theta)}^{s(c)} r \exp\left(-\frac{r^2}{c}\right) dr d\theta = \\
& = \frac{c}{2} \int_{-\theta_0}^{+\theta_0} \exp\left(-\frac{r(\theta)^2}{c}\right) d\theta - c \theta_0 \exp\left(-\frac{s(c)^2}{c}\right).
\end{aligned}$$

To obtain formulations of θ_0 , θ_1 , $r_0(\theta)$, and $r_1(\theta)$ in terms of known constants, we first need to find a ; a is the tangent to $b(O, R)$ that runs through Q , in other words, it is the outermost edge of the area. Furthermore, we define θ_{\max} as $\theta_0 + \theta_1$. Then, by using the law of cosines, the unknown terms are solved this way:

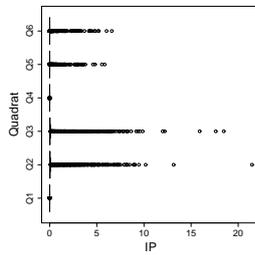
$$\begin{aligned}
a & : & a^2 & = & \overline{OQ}^2 - R^2 \\
\theta_{\max} & : & R^2 & = & \overline{OQ}^2 + a^2 - 2\overline{OQ}a \cos(\theta_{\max}) \\
& & \theta_{\max} & = & \arccos\left(\frac{\overline{OQ}^2 + a^2 - R^2}{2\overline{OQ}a}\right) \\
& & & = & \theta_0 + \theta_1 \\
\theta_0 & : & R^2 & = & \overline{OQ}^2 + s(c)^2 - 2\overline{OQ}s(c) \cos(\theta_0) \\
& & \theta_0 & = & \arccos\left(\frac{\overline{OQ}^2 + s(c)^2 - R^2}{2\overline{OQ}s(c)}\right) \\
\theta_1 & : & \theta_1 & = & \theta_{\max} - \theta_0 \\
r_{0/1}(\theta) & : & R^2 & = & r(\theta)^2 + \overline{OQ}^2 - 2r(\theta)\overline{OQ} \cos(\theta) \\
& & r_{0/1}(\theta) & = & \overline{OQ} \cos(\theta) \pm \sqrt{R^2 - \overline{OQ}^2 \sin(\theta)^2}
\end{aligned}$$

As can be observed from Fig 7, $r_0(\theta)$ and $r_1(\theta)$ are solved using the same formulation. The shorter radius results from subtracting the two terms, and the longer, from adding them.

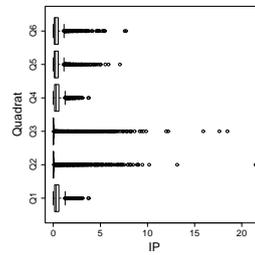
For each quadrat, the result of Eq. 33 is subtracted from that of Eq. 32. These values correspond to the integral term of the expectation of IP (Eq. 28). The correction is then obtained putting together these values, the sum of the diameters of the species in the plot, and the area of the plot (100 m²).

To illustrate the efficacy of the edge correction on small trees, we chose c at 0.5, 1, 2, and 10, corresponding to cases IV–VII in quadrats 1, 4, 5, and 6, as well as cases II and III (for both $c = 1$ and $c = 2$, case II applies) in quadrats 2 and 3. The censoring effect of the plot boundaries are clear in Figs. 8(a), 8(c), 9(a), and 9(c). The boxplots of the corrected IP distributions (Figs. 8(b), 8(d), 9(b), and 9(d)), however, are not aligned as with the big trees. In fact, there seems to be an overestimation of IP after the correction, related to the distance of the quadrat to the edge of the plot: the closer the quadrat, the larger the overestimation.

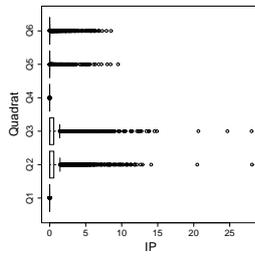
The reason for this problem may perhaps fall on the estimation of $\hat{\lambda}\hat{D}$, the intensity of the trees times the mean diameter. If the true pattern and



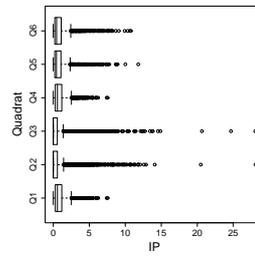
(a) $c = 0.5$ ($s(c) = 9.5$ m): No correction



(b) $c = 0.5$ ($s(c) = 9.5$ m): With correction

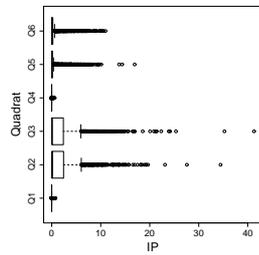


(c) $c = 1$ ($s(c) = 18.6$ m): No correction

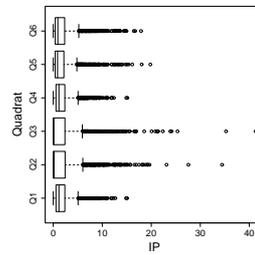


(d) $c = 1$ ($s(c) = 18.6$ m): With correction

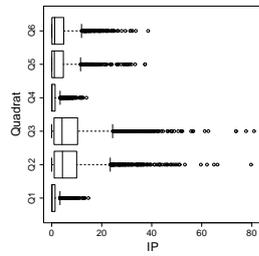
Figure 8: Small trees ($4.5 \text{ cm} \leq D \leq 10.5 \text{ cm}$): IP by quadrat with c at 0.5 and 1, corresponding to cases I, II, IV, and V. A total of 1158 plots with one stand and quadrats in mineral soils were analyzed assuming that six quadrats were measured in each plot. The calculation were conducted separately spruce, pine, and birch; then the three corrected IP were added up to obtain one IP for each quadrat and plot.



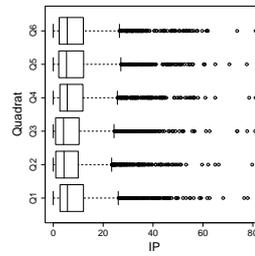
(a) $c = 2$ ($s(c) = 9.5$ m): No correction



(b) $c = 2$ ($s(c) = 9.5$ m): With correction



(c) $c = 10$ ($s(c) = 18.6$ m): No correction



(d) $c = 10$ ($s(c) = 18.6$ m): With correction

Figure 9: Small trees ($4.5 \text{ cm} \leq D \leq 10.5 \text{ cm}$): IP by quadrat with c at 2 and 10, corresponding to cases II, III, VI and VII. A total of 1158 plots with one stand and quadrats in mineral soils were analyzed assuming that six quadrats were measured in each plot. The calculation were conducted separately spruce, pine, and birch; then the three corrected IP were added up to obtain one IP for each quadrat and plot.

mark distribution for this subset of trees deviates greatly from the stationarity assumption, then the estimate of the expected IP in $O(Q, c)$ will be over- or underestimated. The more they deviate from the assumptions, the more biased the expectation will be.

8 DISCUSSION AND CONCLUSIONS

The initial measurement of the influence potential IP from the PSP are biased due to edge-effects. Edge corrections for spatial point patterns are available. The measurements of IP, however, requires a method that takes into account the relationship between trees and quadrats as defined by a function. For this reason a new correction was proposed, which consisted of estimating the expected IP outside the plot, and adding it to the observed IP. The correction is based on the Campbell theorem for a marked point processes.

The method integrates the formula for tree effect over $O(Q, c)$, the area of the circle of influence that lies outside the plot. The limits of the integral are defined by polar coordinates and simple geometrical properties such as the law of cosines. The quadrat as reference point for the integral limits is important, since the tree effect is measured according to the distance from the quadrat. The integrals are solved numerically, but this is only done once for each quadrat, as they depend solely on the distance between the plot center and the quadrat, and therefore are the same for all plots and species. Once the integral is obtained, the sum of the diameters for each species and plot, and the area of the plot remain to be inserted in the formula. For other plot shapes, similar procedures should also be possible.

We applied the correction to observations of big and the small trees from the PSP. In each case we assumed that all six quadrats in the plot had been measured, so that each distribution was based on the same number of observations. The results for big trees were satisfactory, and the correction eliminated most of the bias that was present in the original IP calculations. For small trees, the corrections over-adjusted, adding more influence than what was anticipated.

The procedure we have introduced is based on stationarity of the process

in and around the plot, which might not necessarily be true. The PSP are very heterogeneous in terms of types of forests that were measured. Some of the plots are located in natural forests, others in harvested forests; within the latter different treatments might have been employed (e.g. thinning strategies). Together with environmental factors such as latitude, this leads to different species composition and age structures between stands, which in turn may affect the intensity and the diameter distribution of the trees. As a result, the assumption of stationarity in all plots seems unrealistic.

Because of their size and age, the big trees might be expected to have a more uniform diameter distribution. In the small trees, however, more variation due to treatments and dependence between trees, might allow for inhomogeneous processes and different distributions of the diameters. The following quote summarizes the difficulties found in characterizing the distribution of the diameters:

"[In] even-aged, one species crops (...) the distribution and frequency of different diameters in a crop may vary enormously with the crop's species, age and history. Even-aged crops that have been regularly thinned may have a small variation in diameter around the mean; in contrast, plantations that have not been thinned may have a much larger range in diameter. Usually even-aged crops tend towards a simple distribution, normal or slightly skewed, although in some instances thinning may produce bimodal distributions. (Philip, 1994, p. 111).

The difficulties pointed out are enhanced by the large number of plots in a large study area analyzed in this study. It remains to be seen if the problems encountered with the small trees can be overcome with homogeneous plots, e.g. one stand composition, even-aged, equally treated. Simulated non-stationary processes with different diameter distributions could also be used to determine if the stationary assumption causes the overestimation. It would also be interesting to see if the satisfactory results with the big trees are obtained with datasets collected in other types of forests.

NOTATION

| Notation | Definition |
|-------------------------------|---|
| A | Window where spatial point process is observed. |
| $ A $ | Area of A . |
| $\ \mathbf{x} - \mathbf{y}\ $ | Euclidean distance between event \mathbf{x} and event \mathbf{y} . |
| $b(\mathbf{x}, r)$ | Circle centered at \mathbf{x} and with radius r . |
| $b(O, R)$ | PSP defined as a circle centered at origin and with radius R . |
| $b(Q, s(c))$ | Circle of influence, centered at Q , and with radius $s(c)$. |
| $\delta b(\mathbf{x}, r)$ | Border of the circle centered at \mathbf{x} with radius r . |
| D | Diameter at breast height (DBH) of a tree. |
| $I\{\cdot\}$ | Indicator function that takes the value 1, if \cdot is true, and the value 0 otherwise. |
| IP | Influence potential. |
| $I(Q, c)$ | Inside part of a circle of influence $b(Q, s(c))$. |
| λ | Intensity of the point process. |
| n | Number of points in A . |
| $O(Q, c)$ | Outside part of the circle of influence $b(Q, s(c))$. |
| O | Plot origin, i.e. center of plot. |
| \overline{OQ} | Euclidean distance between plot origin O and quadrat Q . |
| Ψ | Spatial point process. |
| Ψ_M | Marked point process. |
| Φ_z | Tree effect of tree z . |
| Q | Location of quadrat. |
| R | Radius of the plot. |
| $s(c)$ | Radius of minimum significant tree effect. |
| θ_{\max} | Maximum angle for integration. |
| $t_z(Q)$ | Euclidean distance between tree z and quadrat Q . |
| \mathbf{x} | Random event in a spatial point process. |
| z | Location of a tree. |

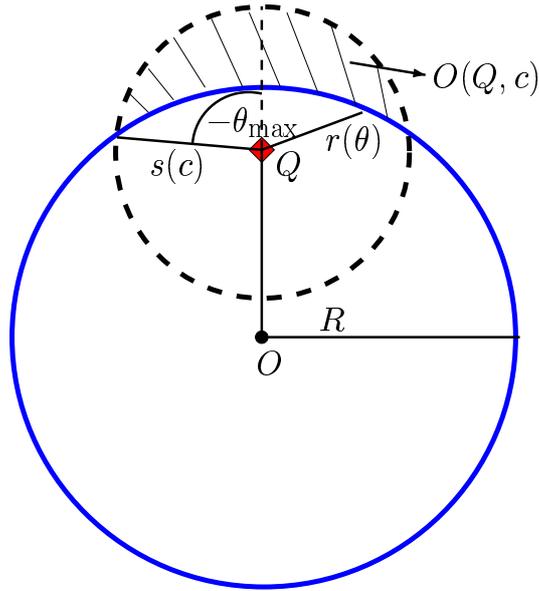


Figure 10: Case II: Geometry of the edge correction

A APPENDIX: FORMULAS BY CASE

A.1 CASE II FOR QUADRATS 2/3, 5/6 AND 1/4, BIG TREES ($R= 9.77$ M);
AND QUADRATS 2/3, SMALL TREES ($R= 5.64$ M)

$$\begin{aligned}
 \int_{z \in O(Q,c)} \exp\left(-\frac{t_z(Q)^2}{c}\right) dz &= \\
 &= \int_{-\theta_{\max}}^{+\theta_{\max}} \int_{r(\theta)}^{s(c)} r \exp\left(-\frac{r^2}{c}\right) dr d\theta \\
 &= \int_{-\theta_{\max}}^{+\theta_{\max}} \left[\frac{c}{2} \exp\left(-\frac{r(\theta)^2}{c}\right) - \frac{c}{2} \exp\left(-\frac{s(c)^2}{c}\right) \right] d\theta \\
 &= \frac{c}{2} \int_{-\theta_{\max}}^{+\theta_{\max}} \exp\left(-\frac{r(\theta)^2}{c}\right) d\theta - c \theta_{\max} \exp\left(-\frac{s(c)^2}{c}\right)
 \end{aligned}$$

θ_{\max} :

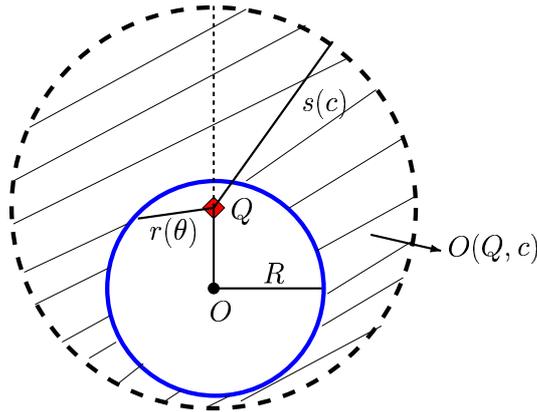


Figure 11: Case III: Geometry of the edge correction

$$\begin{aligned}
 R^2 &= \overline{OQ}^2 + s(c)^2 - 2 \overline{OQ} s(c) \cos(\pi - \theta_{\max}) \\
 &= \overline{OQ}^2 + s(c)^2 + 2 \overline{OQ} s(c) \cos(\theta_{\max}) \\
 \theta_{\max} &= \arccos \left(\frac{R^2 - \overline{OQ}^2 - s(c)^2}{2 \overline{OQ} s(c)} \right)
 \end{aligned}$$

$r(\theta)$: distance from Q to plot edge at angle θ

$$\begin{aligned}
 R^2 &= r(\theta)^2 + \overline{OQ}^2 - 2 r(\theta) \overline{OQ} \cos(\pi - \theta) \\
 0 &= r(\theta)^2 - 2 \overline{OQ} \cos(\pi - \theta) r(\theta) + (\overline{OQ}^2 - R^2) \\
 &= r(\theta)^2 + 2 \overline{OQ} \cos(\theta) r(\theta) + (\overline{OQ}^2 - R^2) \\
 r(\theta) &= -\overline{OQ} \cos(\theta) \pm \sqrt{\overline{OQ}^2 \cos^2(\theta) - (\overline{OQ}^2 - R^2)} \\
 &= -\overline{OQ} \cos(\theta) \pm \sqrt{R^2 - \overline{OQ}^2 (1 - \cos^2(\theta))} \\
 r(\theta) &= \sqrt{R^2 - \overline{OQ}^2 \sin^2(\theta)} - \overline{OQ} \cos(\theta)
 \end{aligned}$$

A.2 CASE III FOR QUADRATS 2/3, 5/6, AND 1/4, BIG TREES ($R=9.77$ M); AND QUADRATS 2/3, SMALL TREES ($R=5.64$ M)

$$\begin{aligned} & \int_{z \in O(Q,c)} \exp\left(-\frac{t_z(Q)^2}{c}\right) dz = \\ &= \int_{z \in b(Q,s(c))} \exp\left(-\frac{t_z(Q)^2}{c}\right) dz - \\ & - \int_{z \in \{b(Q,s(c)) \cap b(O,R)\}} \exp\left(-\frac{t_z(Q)^2}{c}\right) dz \end{aligned}$$

$$\begin{aligned} \int_{z \in b(Q,s(c))} \exp\left(-\frac{t_z(Q)^2}{c}\right) dz &= \int_{-\pi}^{+\pi} \int_0^{s(c)} r \exp\left(-\frac{r^2}{c}\right) dr d\theta \\ &= \pi c \left[1 - \exp\left(-\frac{s(c)^2}{c}\right)\right] \end{aligned}$$

$$\begin{aligned} & \int_{z \in \{b(Q,s(c)) \cap b(O,R)\}} \exp\left(-\frac{t_z(Q)^2}{c}\right) dz = \\ &= \int_0^{2\pi} \int_0^{r(\theta)} r \exp\left(-\frac{r^2}{c}\right) dr d\theta \\ &= \int_0^{2\pi} \left[-\frac{c}{2} \exp\left(-\frac{r(\theta)^2}{c}\right) + \frac{c}{2}\right] d\theta \\ &= \int_0^{2\pi} \frac{c}{2} d\theta - \frac{c}{2} \int_0^{2\pi} \exp\left(-\frac{r(\theta)^2}{c}\right) d\theta \\ &= \pi c - \frac{c}{2} \int_0^{2\pi} \exp\left(-\frac{r(\theta)^2}{c}\right) d\theta \end{aligned}$$

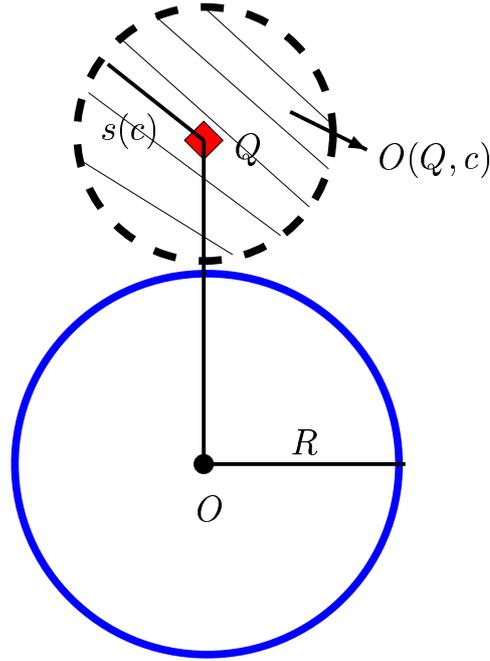


Figure 12: Case IV: Geometry of the edge correction

$r(\theta)$: distance from Q to plot edge at angle θ

$$\begin{aligned}
 R^2 &= r(\theta)^2 + \overline{OQ}^2 - 2r(\theta)\overline{OQ}\cos(\pi - \theta) \\
 0 &= r(\theta)^2 - 2\overline{OQ}\cos(\pi - \theta)r(\theta) + (\overline{OQ}^2 - R^2) \\
 &= r(\theta)^2 + 2\overline{OQ}\cos(\theta)r(\theta) + (\overline{OQ}^2 - R^2) \\
 r(\theta) &= -\overline{OQ}\cos(\theta) \pm \sqrt{\overline{OQ}^2\cos(\theta)^2 - (\overline{OQ}^2 - R^2)} \\
 &= -\overline{OQ}\cos(\theta) \pm \sqrt{R^2 - \overline{OQ}^2(1 - \cos(\theta)^2)} \\
 r(\theta) &= \sqrt{R^2 - \overline{OQ}^2\sin(\theta)^2} - \overline{OQ}\cos(\theta)
 \end{aligned}$$

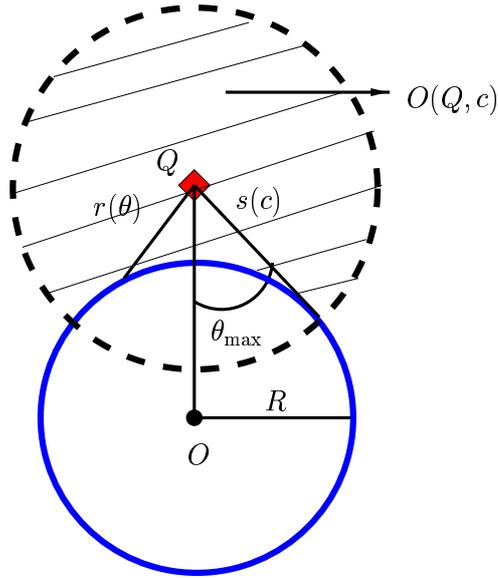


Figure 13: Case V: Geometry of the edge correction

A.3 CASE IV FOR QUADRANTS 1/4 AND 5/6; SMALL TREES ($R= 5.64$ M)

$$\begin{aligned}
 \int_{z \in O(Q,c)} \exp\left(-\frac{t_z(Q)^2}{c}\right) dz &= \int_{z \in b(Q,s(c))} \exp\left(-\frac{t_z(Q)^2}{c}\right) dz \\
 &= \int_{-\pi}^{+\pi} \int_0^{s(c)} r \exp\left(-\frac{r^2}{c}\right) dr d\theta \\
 &= \pi c \left[1 - \exp\left(-\frac{s(c)^2}{c}\right)\right]
 \end{aligned}$$

A.4 CASE V FOR QUADRANTS 1/4 AND 5/6, SMALL TREES ($R= 5.64$ M):

$$\begin{aligned}
 \int_{z \in O(Q,c)} \exp\left(-\frac{t_z(Q)^2}{c}\right) dz &= \\
 &= \int_{z \in b(Q,s(c))} \exp\left(-\frac{t_z(Q)^2}{c}\right) dz - \\
 &- \int_{z \in \{b(Q,s(c)) \cap b(O,R)\}} \exp\left(-\frac{t_z(Q)^2}{c}\right) dz
 \end{aligned}$$

$$\begin{aligned} \int_{z \in b(Q, s(c))} \exp\left(-\frac{t_z(Q)^2}{c}\right) dz &= \int_{-\pi}^{+\pi} \int_0^{s(c)} r \exp\left(-\frac{r^2}{c}\right) dr d\theta \\ &= \pi c \left[1 - \exp\left(-\frac{s(c)^2}{c}\right)\right] \end{aligned}$$

$$\begin{aligned} \int_{z \in \{b(Q, s(c)) \cap b(O, R)\}} \exp\left(-\frac{t_z(Q)^2}{c}\right) dz &= \\ &= \int_{-\theta_{\max}}^{+\theta_{\max}} \int_{r(\theta)}^{s(c)} r \exp\left(-\frac{r^2}{c}\right) dr d\theta \\ &= \int_{-\theta_{\max}}^{+\theta_{\max}} \left[\frac{c}{2} \exp\left(-\frac{r(\theta)^2}{c}\right) - \frac{c}{2} \exp\left(-\frac{s(c)^2}{c}\right) \right] d\theta \\ &= \frac{c}{2} \int_{-\theta_{\max}}^{+\theta_{\max}} \exp\left(-\frac{r(\theta)^2}{c}\right) d\theta - c \theta_{\max} \exp\left(-\frac{s(c)^2}{c}\right) \end{aligned}$$

θ_{\max} :

$$\begin{aligned} R^2 &= \overline{OQ}^2 + s(c)^2 - 2 \overline{OQ} s(c) \cos(\theta_{\max}) \\ \theta_{\max} &= \arccos\left(\frac{\overline{OQ}^2 + s(c)^2 - R^2}{2 \overline{OQ} s(c)}\right) \end{aligned}$$

$r(\theta)$: distance from Q to plot edge at angle θ

$$\begin{aligned} R^2 &= r(\theta)^2 + \overline{OQ}^2 - 2 r(\theta) \overline{OQ} \cos(\theta) \\ 0 &= r(\theta)^2 - 2 \overline{OQ} \cos(\theta) r(\theta) + (\overline{OQ}^2 - R^2) \\ r(\theta) &= \overline{OQ} \cos(\theta) \pm \sqrt{\overline{OQ}^2 \cos^2(\theta) - (\overline{OQ}^2 - R^2)} \\ &= \overline{OQ} \cos(\theta) \pm \sqrt{\overline{OQ}^2 (\cos^2(\theta) - 1) + R^2} \\ &= \overline{OQ} \cos(\theta) \pm \sqrt{R^2 - \overline{OQ}^2 (1 - \cos^2(\theta))} \\ r(\theta) &= \overline{OQ} \cos(\theta) - \sqrt{R^2 - \overline{OQ}^2 \sin^2(\theta)} \end{aligned}$$

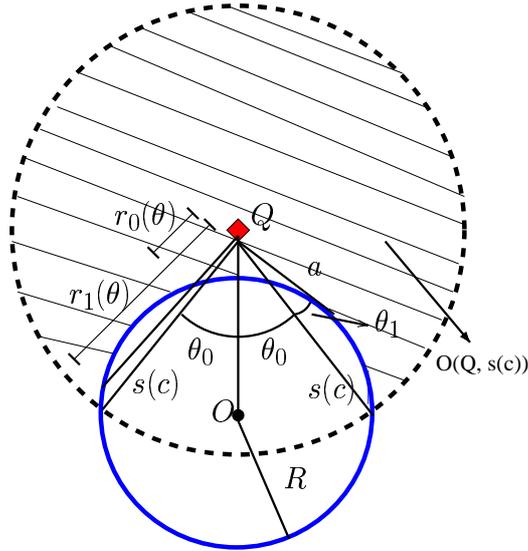


Figure 14: Case VI: Geometry of the edge correction

A.5 CASE VI FOR QUADRANTS 1/4 AND 5/6, SMALL TREES ($R = 5.64$ M):

$$\begin{aligned}
 & \int_{z \in O(Q,c)} \exp\left(-\frac{t_z(Q)^2}{c}\right) dz = \\
 & = \int_{z \in b(Q,s(c))} \exp\left(-\frac{t_z(Q)^2}{c}\right) dz - \\
 & - \int_{z \in \{b(Q,s(c)) \cap b(O,R)\}} \exp\left(-\frac{t_z(Q)^2}{c}\right) dz
 \end{aligned}$$

$$\begin{aligned}
 \int_{z \in b(Q,s(c))} \exp\left(-\frac{t_z(Q)^2}{c}\right) dz & = \int_{-\pi}^{+\pi} \int_0^{s(c)} r \exp\left(-\frac{r^2}{c}\right) dr d\theta \\
 & = \pi c \left[1 - \exp\left(-\frac{s(c)^2}{c}\right) \right]
 \end{aligned}$$

$$\begin{aligned}
& \int_{z \in \{b(Q, s(c)) \cap b(O, R)\}} \exp\left(-\frac{t_z(Q)^2}{c}\right) dz = \\
& = 2 \int_{\theta_0}^{\theta_0 + \theta_1} \int_{r_0(\theta)}^{r_1(\theta)} r \exp\left(-\frac{r^2}{c}\right) dr d\theta \\
& + \int_{-\theta_0}^{+\theta_0} \int_{r_0(\theta)}^{s(c)} r \exp\left(-\frac{r^2}{c}\right) dr d\theta
\end{aligned}$$

$$\begin{aligned}
& \int_{\theta_0}^{\theta_0 + \theta_1} \int_{r_0(\theta)}^{r_1(\theta)} r \exp\left(-\frac{r^2}{c}\right) dr d\theta = \\
& = \frac{c}{2} \int_{\theta_0}^{\theta_0 + \theta_1} \exp\left(-\frac{r_0(\theta)^2}{c}\right) d\theta - \frac{c}{2} \int_{\theta_0}^{\theta_0 + \theta_1} \exp\left(-\frac{r_1(\theta)^2}{c}\right) d\theta
\end{aligned}$$

$$\begin{aligned}
& \int_{-\theta_0}^{+\theta_0} \int_{r(\theta)}^{s(c)} r \exp\left(-\frac{r^2}{c}\right) dr d\theta = \\
& = \int_{-\theta_0}^{+\theta_0} \left[\frac{c}{2} \exp\left(-\frac{r(\theta)^2}{c}\right) - \frac{c}{2} \exp\left(-\frac{s(c)^2}{c}\right) \right] d\theta \\
& = \frac{c}{2} \int_{-\theta_0}^{+\theta_0} \exp\left(-\frac{r(\theta)^2}{c}\right) d\theta - c \theta_0 \exp\left(-\frac{s(c)^2}{c}\right)
\end{aligned}$$

a = tangent to plot and running through Q.

$$a^2 = \overline{OQ}^2 - R^2$$

$$\theta_{\max} = \theta_0 + \theta_1:$$

$$\begin{aligned}
R^2 &= \overline{OQ}^2 + a^2 - 2\overline{OQ}a \cos(\theta_{\max}) \\
\theta_{\max} &= \arccos\left(\frac{\overline{OQ}^2 + a^2 - R^2}{2\overline{OQ}a}\right) \\
&= \theta_0 + \theta_1
\end{aligned}$$

θ_0 :

$$\begin{aligned}
R^2 &= \overline{OQ}^2 + s(c)^2 - 2\overline{OQ}s(c) \cos(\theta_0) \\
\theta_0 &= \arccos\left(\frac{\overline{OQ}^2 + s(c)^2 - R^2}{2\overline{OQ}s(c)}\right)
\end{aligned}$$

θ_1 :

$$\theta_1 = \theta_{\max} - \theta_0$$

$r(\theta)$: distance from Q to plot edge at angle θ

$$\begin{aligned}
R^2 &= r(\theta)^2 + \overline{OQ}^2 - 2r(\theta)\overline{OQ} \cos(\theta) \\
0 &= r(\theta)^2 - 2\overline{OQ} \cos(\theta) r(\theta) + (\overline{OQ}^2 - R^2) \\
r(\theta) &= \overline{OQ} \cos(\theta) \pm \sqrt{\overline{OQ}^2 \cos^2(\theta) - (\overline{OQ}^2 - R^2)} \\
&= \overline{OQ} \cos(\theta) \pm \sqrt{\overline{OQ}^2 (\cos^2(\theta) - 1) + R^2} \\
&= \overline{OQ} \cos(\theta) \pm \sqrt{R^2 - \overline{OQ}^2 (1 - \cos^2(\theta))} \\
r_{0/1}(\theta) &= \overline{OQ} \cos(\theta) \pm \sqrt{R^2 - \overline{OQ}^2 \sin^2(\theta)}
\end{aligned}$$

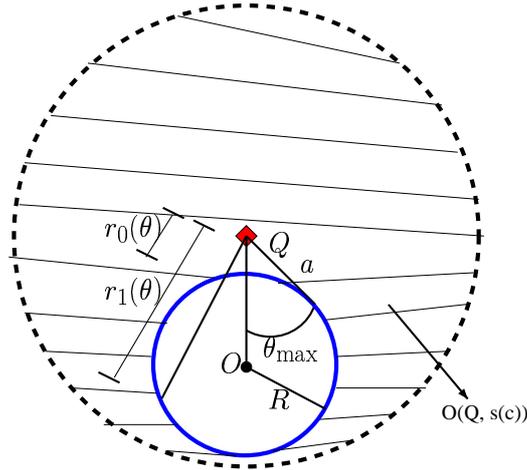


Figure 15: Case VII: Geometry of the edge correction

A.6 CASE VII FOR QUADRATS 1/4 AND 5/6, SMALL TREES ($R = 5.64 M$):

$$\begin{aligned}
 & \int_{z \in O(Q,c)} \exp\left(-\frac{t_z(Q)^2}{c}\right) dz = \\
 & = \int_{z \in b(Q,s(c))} \exp\left(-\frac{t_z(Q)^2}{c}\right) dz - \\
 & - \int_{z \in \{b(Q,s(c)) \cap b(O,R)\}} \exp\left(-\frac{t_z(Q)^2}{c}\right) dz
 \end{aligned}$$

$$\begin{aligned}
 \int_{z \in b(Q,s(c))} \exp\left(-\frac{t_z(Q)^2}{c}\right) dz & = \int_{-\pi}^{+\pi} \int_0^{s(c)} r \exp\left(-\frac{r^2}{c}\right) dr d\theta \\
 & = \pi c \left[1 - \exp\left(-\frac{s(c)^2}{c}\right) \right]
 \end{aligned}$$

$$\begin{aligned}
& \int_{z \in \{b(Q, s(c)) \cap b(O, R)\}} \exp\left(-\frac{t_z(Q)^2}{c}\right) dz = \\
& = \int_{-\theta_{\max}}^{+\theta_{\max}} \int_{r_0(\theta)}^{r_1(\theta)} r \exp\left(-\frac{r^2}{c}\right) dr d\theta \\
& = \frac{c}{2} \int_{-\theta_{\max}}^{+\theta_{\max}} \exp\left(-\frac{r_0(\theta)^2}{c}\right) d\theta - \frac{c}{2} \int_{-\theta_{\max}}^{+\theta_{\max}} \exp\left(-\frac{r_1(\theta)^2}{c}\right) d\theta
\end{aligned}$$

a = longest segment from Q to plot edge, i.e. tangent to plot and running through Q .

$$\begin{aligned}
a^2 & = (\overline{OQ} - R)(\overline{OQ} + R) \\
& = \overline{OQ}^2 - R^2
\end{aligned}$$

θ_{\max} :

$$\begin{aligned}
R^2 & = \overline{OQ}^2 + a^2 - 2\overline{OQ}a \cos(\theta_{\max}) \\
\theta_{\max} & = \arccos\left(\frac{\overline{OQ}^2 + a^2 - R^2}{2\overline{OQ}a}\right)
\end{aligned}$$

$r(\theta)$: distance from Q to plot edge at angle θ

$$\begin{aligned}
R^2 & = r(\theta)^2 + \overline{OQ}^2 - 2r(\theta)\overline{OQ} \cos(\theta) \\
0 & = r(\theta)^2 - 2\overline{OQ} \cos(\theta) r(\theta) + (\overline{OQ}^2 - R^2) \\
r(\theta) & = \overline{OQ} \cos(\theta) \pm \sqrt{\overline{OQ}^2 \cos^2(\theta) - (\overline{OQ}^2 - R^2)} \\
& = \overline{OQ} \cos(\theta) \pm \sqrt{\overline{OQ}^2 (\cos^2(\theta) - 1) + R^2} \\
& = \overline{OQ} \cos(\theta) \pm \sqrt{R^2 - \overline{OQ}^2 (1 - \cos^2(\theta))} \\
r_{0/1}(\theta) & = \overline{OQ} \cos(\theta) \pm \sqrt{R^2 - \overline{OQ}^2 \sin^2(\theta)}
\end{aligned}$$

Part II

Model for the local presence of an understory species

1 INTRODUCTION

The influence potential index IP quantifies the effect of surrounding trees on a quadrat. In order to model the relationship between IP and the presence of an understory species in a quadrat, we must take into consideration that definition of IP: as IP is modeled as an exponential decay, it follows that the effect of the trees works at small ranges. The measurements from the permanent sample plots (PSP) from the National Forest Inventory, however, implicitly reflect large scale factors such as latitude, climate, and soil types, which interfere with the local scale signal measured by IP.

In order to eliminate the large scale factors in the data, a conditional model is necessary. This report begins by presenting the basic ideas of logistic regression, and then follows with the derivation of a conditional logistic model for this study. The model is then applied to PSP data for *Vaccinium vitis-idaea*, and then the results are discussed.

2 LOGISTIC REGRESSION

Binary data are observations of an event Y with two possible outcomes, for example success and failure. The presence and absence of an understory species is an example of such data, where the presence may be represented by a 1 and the absence by a 0. A model for this type of data estimates the probability of the outcomes, i.e. $P(Y_k = 1) = p_k$ and $P(Y_k = 0) = 1 - p_k$, using information from independent variables or covariates \mathbf{X} , where \mathbf{X} represents the design matrix (see e.g. McCullagh and Nelder 1989; Hosmer and Lemeshow 1989; and Collett 1991). In this particular problem, we are interested in using IP from each of the tree species as covariates for modeling the probability of the presence of an understory species.

Normal linear models such as

$$v_k = \mathbf{X}\boldsymbol{\beta} + \epsilon_k, \quad (1)$$

with $\boldsymbol{\beta}$ as coefficients and ϵ as the normal random error with mean zero and constant variance, put no restriction on the possible values that the response variable v_k may take. For binary data the probability of success p_k must be limited to the interval $[0, 1]$. The solution is to transform the probabilities to a new variable η_k that is allowed to vary continuously in the interval $(-\infty, +\infty)$; an inverse transformation of η_k should give the fitted probabilities between 0 and 1.

One such transformation is the logistic transformation

$$\eta_k = \log \left(\frac{p_k}{1 - p_k} \right).$$

In the theory of generalized linear models, the transformation is known as the link function. Other links for binary data are the probit and complementary log-log transformations (McCullagh and Nelder, 1989). A linear model similar to Eq. 1 may be applied using η_k as

$$\begin{aligned}
\eta_k &= \log\left(\frac{p_k}{1-p_k}\right) \\
&= \sum_i \beta_i X_{ik} \\
&= \boldsymbol{\beta} \mathbf{x}_k,
\end{aligned}$$

where \mathbf{x}_k is the k -th row vector of the design matrix.

The inverse transformation results in

$$p_k = \frac{\exp(\mathbf{x}_k \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_k \boldsymbol{\beta})}.$$

The assumption behind the logistic model is that the conditional distribution of $Y_k \mid \mathbf{x}_k$ is Binomial with expected value p_k .

The fitting of such models may be carried out using the maximum likelihood method. The likelihood function for N binary observations is given by

$$L(\boldsymbol{\beta}) = \prod_{k=1}^N p_k^{y_k} (1-p_k)^{1-y_k}.$$

3 LOGISTIC REGRESSION MODEL FOR PSP DATA

We wish to find a model that can explain the probability of presence of an understory species in a specific quadrat by using IP from pine, spruce, and birch. A logistic model for it could be set up as

$$P(W_{kq} = 1) = \frac{\exp(\beta_0 + \beta_1 \text{IP}_{\text{pine},kq} + \beta_2 \text{IP}_{\text{spruce},kq} + \beta_3 \text{IP}_{\text{birch},kq})}{1 + \exp(\beta_0 + \beta_1 \text{IP}_{\text{pine},kq} + \beta_2 \text{IP}_{\text{spruce},kq} + \beta_3 \text{IP}_{\text{birch},kq})}, \quad (2)$$

where W_{kq} represents the presence or absence of a species in quadrat q of plot k .

The ideas behind the ecological field theory indicate that the trees exert their greatest influence at short ranges. Therefore we wish to model the relationship between understory and trees at the local scale, i.e. within short domains. The data from the PSP, however, consists of measurements carried out throughout a very large study area. It is then expected that the observations on the presence and absence of understory reflect some large-scale factors such as latitude and climate: e.g. certain species prefer warmer conditions and are present more often in plots situated in the southern part of Finland (see e.g. Reinikainen et al. (2000)). In similar way, IP measurements also include implicitly large-scale factors that affect the presence and size of the trees; Scots pines, for example, are less frequent and smaller towards the North (Anon., 2000). Therefore information from both large- and local-scales are contained in the data, and since our purpose is to concentrate on the local scale, the model in Eq. 2 is not appropriate.

4 CONDITIONAL LOGISTIC MODEL FOR PSP DATA

One way of avoiding the large-scale factors is to condition the probability by another event that also includes those same factors. By conditioning it is possible to cancel the large-scale factors and to leave the local-scale characteristics for further modeling. To find this other large-scale event, we take into account that all quadrats in the same plot are affected by identical large-scale factors, but not by the same local-scale ones: IP is different for each quadrat because the distance from a tree to a quadrat is different in each case, but variables such as latitude are the same for all quadrats. We therefore require an event that is also related to the set of all quadrats in a plot and which is affected by large-scale factors. One such event is the number of quadrats in the plot where the species was found. Conditioning on this event means that we shift the focus from the presence in a quadrat to the pattern of presence of the species in all quadrats of the plot.

In order to emphasize the local scale effects of IP on the presence of an understory species, we require a new model. The model at quadrat level

that includes both large- and local-scale factors in Eq. 2 can be written as

$$P(W_{kq} = 1) = \frac{\exp(a + \mathbf{x}_{kq}\boldsymbol{\beta})}{1 + \exp(a + \mathbf{x}_{kq}\boldsymbol{\beta})}.$$

Here W_{kq} represents the presence or absence of the understory species in quadrat q of plot k ; a collects the large scale information for the entire study area; and \mathbf{x}_{kq} is the row vector of IP covariates for the specific quadrat.

In similar way, the probability of absence in quadrat q can be modeled from

$$P(W_{kq} = 0) = \frac{1}{1 + \exp(a + \mathbf{x}_{kq}\boldsymbol{\beta})},$$

and in general

$$P(W_{kq} = w) = \frac{\exp[w(a + \mathbf{x}_{kq}\boldsymbol{\beta})]}{1 + \exp(a + \mathbf{x}_{kq}\boldsymbol{\beta})}, \quad (3)$$

where $w \in \{0, 1\}$.

The set of quadrats in a plot, however, is the new basis for calculating probabilities. We denote such a set of quadrats in a plot k by the vector

$$\mathbf{w} = (w_1, \dots, w_{n_k})^t;$$

each element w_q represents one of the n_k quadrats in the plot and may take the value of 1 or 0 depending on whether the species was found in that quadrat or not. Furthermore the conditioning event, i.e. the number of quadrats in the plot occupied by the species, is denoted by z_k .

Given the number n_k of quadrats in the plot and the total number z_k of quadrats where the species was present, different patterns of 0's and 1's

may arise in a plot. For example, if three quadrats were measured in a plot, and the understory species was seen in two of them, then any of the following patterns could have been observed: $(1, 1, 0)$, $(1, 0, 1)$, or $(0, 1, 1)$. In other words, the species might have been found in quadrats 1 and 2, 1 and 3, or 2 and 3. The number of patterns that are possible in a plot is

$$M_k + 1 = \binom{n_k}{z_k},$$

i.e. all possible combinations of z_k occurrences in n_k quadrats. The pattern that was actually observed in the field is denoted by $\mathbf{w}_k^{(0)}$, and all other possible, unobserved patterns by $\mathbf{w}_k^{(j)}$, where $j = 1, \dots, M_k$.

The conditional probability of the observed pattern $\mathbf{w}_k^{(0)}$ given the z_k occurrences of the species in the plot is expressed as

$$\begin{aligned} P(\mathbf{w}_k^{(0)} | z_k) &= \frac{P(\mathbf{w}_k^{(0)} \cap z_k)}{P(z_k)} \\ &= \frac{P(\mathbf{w}_k^{(0)})}{P(z_k)}. \end{aligned}$$

If we consider that the quadrats are sufficiently separated in terms of local-scale effects, then independence among the quadrats may be assumed. It follows that the probability of any pattern will depend on the probabilities of presence in each of the quadrats as in Eq. 3; thus

$$\begin{aligned} P(\mathbf{w}_k^{(j)}) &= \prod_{q=1}^{n_k} P(w_{kq}^{(j)}) \\ &= \prod_{q=1}^{n_k} \frac{\exp[w_{kq}^{(j)}(a + \mathbf{x}_{kq}\boldsymbol{\beta})]}{1 + \exp(a + \mathbf{x}_{kq}\boldsymbol{\beta})}. \end{aligned} \quad (4)$$

Furthermore the probability of z_k occurrences is the sum of the probabilities of those patterns with z_k 1's in the n_k quadrats. The set of such pat-

terns is denoted by $A_{z_k} = \left\{ j : \sum_{q=1}^{n_k} w_{kq}^{(j)} = z_k, j = 0, \dots, M_k \right\}$, and their probabilities can be obtained from Eq. 4. Therefore,

$$\begin{aligned}
 P(Z_k = z_k) &= \sum_{j \in A_{z_k}} P(\mathbf{w}_k^{(j)}) \\
 &= \sum_{j \in A_{z_k}} \prod_{q=1}^{n_k} P(w_{kq}^{(j)}) \\
 &= \sum_{j \in A_{z_k}} \prod_{q=1}^{n_k} \frac{\exp[w_{kq}^{(j)}(a + \mathbf{x}_{kq}\boldsymbol{\beta})]}{1 + \exp(a + \mathbf{x}_{kq}\boldsymbol{\beta})}. \tag{5}
 \end{aligned}$$

Combining the equations for the probability of the observed pattern (Eq. 4) and of the number of quadrats where the species was found (Eq. 5), the conditional probability is

$$P(\mathbf{w}_k^{(0)} | z_k) = \frac{\prod_{q=1}^{n_k} \frac{\exp[w_{kq}^{(0)}(a + \mathbf{x}_{kq}\boldsymbol{\beta})]}{1 + \exp(a + \mathbf{x}_{kq}\boldsymbol{\beta})}}{\sum_{j \in A_{z_k}} \prod_{q=1}^{n_k} \frac{\exp[w_{kq}^{(j)}(a + \mathbf{x}_{kq}\boldsymbol{\beta})]}{1 + \exp(a + \mathbf{x}_{kq}\boldsymbol{\beta})}}.$$

This represents the probability of the observed pattern in relation to the probability of all possible patterns with the same number of quadrats and occurrences of the understory species. If the species is present in all or absent in all the quadrats of a plot, then only one pattern is possible, i.e. the observed one, and in that case no valuable information is obtained from that plot.

With further simplifications the conditional probability may be expressed as

$$\begin{aligned}
P(\mathbf{w}_k^{(0)} | z_k) &= \tag{6} \\
&= \frac{\prod_{q=1}^{n_k} \exp \left[w_{kq}^{(0)} (a + \mathbf{x}_{kq} \boldsymbol{\beta}) \right]}{\prod_{q=1}^{n_k} 1 + \exp(a + \mathbf{x}_{kq} \boldsymbol{\beta})} \cdot \frac{\prod_{q=1}^{n_k} 1 + \exp(a + \mathbf{x}_{kq} \boldsymbol{\beta})}{\sum_{j=0}^{M_k} \prod_{q=1}^{n_k} \exp \left[w_{kq}^{(j)} (a + \mathbf{x}_{kq} \boldsymbol{\beta}) \right]} \\
&= \frac{\prod_q \exp \left(w_{kq}^{(0)} a \right) \prod_q \exp \left(w_{kq}^{(0)} \mathbf{x}_{kq} \boldsymbol{\beta} \right)}{\sum_j \left[\prod_q \exp \left(w_{kq}^{(j)} a \right) \prod_q \exp \left(w_{kq}^{(j)} \mathbf{x}_{kq} \boldsymbol{\beta} \right) \right]} \\
&= \frac{\exp(a \sum_q w_{kq}^{(0)}) \prod_q \exp \left(w_{kq}^{(0)} \mathbf{x}_{kq} \boldsymbol{\beta} \right)}{\sum_j \left[\exp \left(a \sum_q w_{kq}^{(j)} \right) \prod_q \exp \left(w_{kq}^{(j)} \mathbf{x}_{kq} \boldsymbol{\beta} \right) \right]} \\
&= \frac{\exp(a z_k) \prod_q \exp \left(w_{kq}^{(0)} \mathbf{x}_{kq} \boldsymbol{\beta} \right)}{\sum_j \left[\exp(a z_k) \prod_q \exp \left(w_{kq}^{(j)} \mathbf{x}_{kq} \boldsymbol{\beta} \right) \right]} \\
&= \frac{\prod_{q=1}^{n_k} \exp \left(w_{kq}^{(0)} \mathbf{x}_{kq} \boldsymbol{\beta} \right)}{\sum_{j=0}^{M_k} \prod_{q=1}^{n_k} \exp \left(w_{kq}^{(j)} \mathbf{x}_{kq} \boldsymbol{\beta} \right)}. \tag{7}
\end{aligned}$$

The advantage of this approach is that the large-scale factor a is canceled out. This is possible because all patterns in a plot, both observed and unobserved, are subject to the same large-scale effects, and because they all have the same number of 0's and 1's, i.e. $\sum \mathbf{w}_k^{(j)} = z_k$. The resulting model in Eq. 7 is a *conditional logistic model*, also known in medical applications as a matched case-control model (Collett 1991; Woodward 1999). The set of patterns from a specific plot, or the corresponding cases and controls, are sometimes called a "matched set".

This model can be written more simply in matrix notation:

$$\begin{aligned}
P(\mathbf{w}_k^{(0)} | z_k) &= \frac{\exp(\sum_{q=1}^{n_k} w_{kq}^{(0)} \mathbf{x}_{kq} \boldsymbol{\beta})}{\sum_{j=0}^{M_k} \exp(\sum_{q=1}^{n_k} w_{kq}^{(j)} \mathbf{x}_{kq} \boldsymbol{\beta})} \\
&= \frac{\exp(\mathbf{w}_k^{(0)} \mathbf{X}_k \boldsymbol{\beta})}{\sum_{j=0}^{M_k} \exp(\mathbf{w}_k^{(j)} \mathbf{X}_k \boldsymbol{\beta})} \quad (8) \\
&= \left[\frac{\sum_{j=0}^{M_k} \exp(\mathbf{w}_k^{(j)} \mathbf{X}_k \boldsymbol{\beta})}{\exp(\mathbf{w}_k^{(0)} \mathbf{X}_k \boldsymbol{\beta})} \right]^{-1} \\
&= \left[1 + \frac{\sum_{j=1}^{M_k} \exp(\mathbf{w}_k^{(j)} \mathbf{X}_k \boldsymbol{\beta})}{\exp(\mathbf{w}_k^{(0)} \mathbf{X}_k \boldsymbol{\beta})} \right]^{-1} \\
&= \left\{ 1 + \sum_{j=1}^{M_k} \exp\left[\left(\mathbf{w}_k^{(j)} \mathbf{X}_k - \mathbf{w}_k^{(0)} \mathbf{X}_k\right) \boldsymbol{\beta}\right] \right\}^{-1}.
\end{aligned}$$

It is important to notice that the design matrix $\mathbf{X}_k = (\mathbf{x}_{k1}^t, \dots, \mathbf{x}_{kn_k}^t)^t$ for the plot k is the same for the observed and the unobserved patterns. The actual covariate matrix is in fact $\mathbf{w}_k^{(j)} \mathbf{X}_k$, which is in terms of the patterns; in other words the model compares patterns instead of quadrats. These pattern covariates we call the *influence potential on a pattern* or IPP:

$$\begin{aligned}
\text{IPP}_k^{(j)} &= \mathbf{w}_k^{(j)} \mathbf{X}_k \\
&= \sum_{q=1}^{n_k} w_{kq}^{(j)} \cdot \text{IP}_{kq}.
\end{aligned}$$

As an example, we can take a plot where quadrats 1, 2, 3, and 4 were measured; quadrats 1 and 3 were occupied by a species; and the following values of IP for pine and spruce were obtained:

| Quadrat | Presence | IP _{pine} | IP _{spruce} |
|---------|----------|--------------------|----------------------|
| 1 | 1 | 2 | 0 |
| 2 | 0 | 5 | 0 |
| 3 | 1 | 1 | 0 |
| 4 | 0 | 8 | 0 |

IP for spruce is zero in this example because no trees of that species were found in the plot. Six patterns are possible with this number of quadrats and occurrences. The matrix of original IP measurements is the same for all patterns, but after it is multiplied by the patterns $\mathbf{w}_k^{(j)}$, a different IPP is obtained for each pattern:

| j | \mathbf{w} | IP _{pine} | IP _{spruce} | IPP _{pine} = $\mathbf{w}IP_{pine}$ | IPP _{spruce} = $\mathbf{w}IP_{spruce}$ | Y |
|-----|--------------|--------------------|----------------------|--|--|-----|
| 0 | (1, 0, 1, 0) | | | 3 | 0 | 1 |
| 1 | (1, 1, 0, 0) | 2 | 0 | 7 | 0 | 0 |
| 2 | (1, 0, 0, 1) | | | 10 | 0 | 0 |
| 3 | (0, 1, 1, 0) | | | 6 | 0 | 0 |
| 4 | (0, 1, 0, 1) | | | 13 | 0 | 0 |
| 5 | (0, 0, 1, 1) | | | 9 | 0 | 0 |

The matrix with IPP_{pine} and IPP_{spruce} as columns is the new design matrix for the conditional logistic model. The response variable is no longer the presence or absence in a quadrat, but whether the pattern is the observed ($Y = 1$) or an unobserved ($Y = 0$) one, as in the last column before. For this same reason, the number the number of rows of the design matrix of IPP is now the total number M of possible patterns in all plots, i.e.

$$\begin{aligned}
 M &= \sum_k \binom{n_k}{z_k} \\
 &= \sum_k (M_k + 1).
 \end{aligned}$$

In logistic regression, an important measure is the odds $p/(1-p)$, i.e. the relation between the probability of success and the probability of failure.

The odds ratio $\Psi = [p_1/(1 - p_1)]/[p_2/(1 - p_2)]$ compares the odds of one event to that of another (Woodward, 1999). In this study Ψ represents the relative probability of observing pattern r with respect to pattern s , given that both patterns are possible in the same plot. If the respective conditional probabilities (e.g. as in Eq. 8) of finding patterns r and s in the forest are denoted by p_r and p_s , and $\mathbf{w}_k^{(r)} \mathbf{X}_k$ are the covariates measured for pattern r , and $\mathbf{w}_k^{(s)} \mathbf{X}_k$ those for pattern s , then the Ψ in a conditional logistic model is

$$\begin{aligned} \Psi &= \frac{p_r/(1 - p_r)}{p_s/(1 - p_s)} \\ &= \frac{\exp(\mathbf{w}_k^{(r)} \mathbf{X}_k \beta)}{\sum_j \exp(\mathbf{w}_k^{(j)} \mathbf{X}_k \beta)} \cdot \frac{\sum_j \exp(\mathbf{w}_k^{(j)} \mathbf{X}_k \beta)}{\exp(\mathbf{w}_k^{(s)} \mathbf{X}_k \beta)} \\ &= \exp \left[\left(\mathbf{w}_k^{(r)} \mathbf{X}_k - \mathbf{w}_k^{(s)} \mathbf{X}_k \right) \beta \right]. \end{aligned}$$

From this follows that $\exp(\beta_i)$ can be interpreted as the odds ratio when covariate i increases by a unit and all other covariates are kept constant. In other words, it indicates how much more likely is pattern r relative to pattern s when there is a unit increase in covariate i . This is often more interesting and useful than estimating the actual probability $P(\mathbf{w}_k^{(0)} | z_k)$, because it gives information on the effect of the covariate on the odds of the studied event.

5 FITTING OF A CONDITIONAL LOGISTIC MODEL

As with logistic regression, the parameters β in a conditional logistic model can be estimated using the maximum likelihood method. For completeness, in the appendix we include the likelihood, the log-likelihood, and the first and second derivatives of the log-likelihood for the conditional logistic model for the data from the PSP.

Since no analytical solutions exist, we implemented the Newton-Raphson algorithm to find the estimates and asymptotic variances of the parameters of the model, using the first and second derivatives of the log likeli-

hood. An alternative method for estimation is the Cox proportional hazard model when all possible patterns are considered as the risk set, and the observed pattern as the failure; this is the method employed by the package `survival` of the statistical software R.

The goodness of fit of a model is tested by comparing the log likelihood of two nested models, as in $D = -2(\log L_1 - \log L_2)$, where the deviance D has asymptotically a χ^2 distribution with as many degrees of freedom as the difference in the number of parameters between the two models. In conditional logistic regression the log likelihood of the null model, i.e. when $\beta = \mathbf{0}$, simplifies to a constant equal to $\sum_k \log(M_k + 1)$.

While an apparent consensus exists in normal linear models as to what tools should be used when checking the model, a large number of different methods have been suggested in the literature for conditional logistic regression; cf. Pregibon (1984), Hosmer and Lemeshow (1989), Breslow and Day (1980), and Collett (1991). A useful tool for this type of model is the graph of k and $\Delta_k \beta_i$, known as delta-beta (Pregibon, 1984). Delta-beta measures how the estimate of the parameter i changes if the matched set k is ignored, and is usually standardized to standard error units. Also useful for model checking are Pearson residuals and the general delta-beta for individual observations (Hosmer and Lemeshow 1989; Collett 1991).

6 APPLICATION

For the purpose of illustrating the application of the conditional logistic model to data from the PSP, we chose to fit the model to the presence of the dwarf shrub *Vaccinium vitis-idaea*, commonly known as cowberry in English ("lingon" in Swedish, "Preiselbeere" in German, "puolukka" in Finnish, and "arándano rojo" in Spanish). The analyzes were carried out in the statistical software R, version 1.2.1, where the package "survival" version 2.8-2 provided the functions for fitting the conditional logistic regression.

Although 3009 permanent sample plots were originally measured, a more homogeneous subset was analyzed that consisted of those plots with only one stand and with all quadrats situated on mineral soils. Additionally, the restrain that $z_k \neq n_k$ and $z_k \neq 0$, i.e. that the occurrences should not be

Table II.1: *Vaccinum vitis-idaea*: Mean (st.dev.) of IPP for pine, spruce, and birch at $c = 1$ according to pattern.

| Pattern | IPP | | | | | |
|------------|------|--------|--------|--------|-------|--------|
| | Pine | | Spruce | | Birch | |
| Observed | 1.48 | (3.80) | 0.72 | (2.19) | 0.88 | (3.19) |
| Unobserved | 2.87 | (5.95) | 1.51 | (3.79) | 0.90 | (3.09) |

equal to the number of quadrats or to zero, reduced further the available number of plots. For cowberry 342 plots matched the necessary criteria for this study, and additional 1400 other patterns were possible to construct, resulting in a design matrix with 1742 rows.

Before starting the analysis, the c parameter in the IP equation needed to be fixed. This parameter determines the range of influence of the tree species. For this study, two levels of c were chosen arbitrarily and applied to all tree species, namely $c = 1$ and $c = 10$. The range for the former is approximately 2 m, and for the latter it is less than 7 m. Additionally, as noted before, the actual covariates are $\mathbf{w}_k^{(j)} \mathbf{X}_k = \text{IPP}$, where the pattern is multiplied by the measured values of IP.

6.1 VACCINUM VITIS-IDAEA WITH $c = 1$

For cowberry, the model was first fitted with $c = 1$. Table II.1 compares the mean and standard deviations of IPP in the observed and unobserved patterns; the respective histograms are shown in Fig. 1. Although the standard deviations were relatively large compared to the means, there seemed to be differences among the values of IPP for spruce and pine among the groups, with higher IPP among the unobserved patterns than the observed. The distributions of IPP in both groups were highly skewed due to the large proportion of values equal to zero that result from the absence of the tree species in the plot. Furthermore, large differences were found between the maximum value of IPP of spruce in the observed patterns, 15.0, and the unobserved patterns, 29.70; such differences were not so obvious among pine and birch. Correlations among the tree species were very low; the largest was -0.09 between spruce and pine.

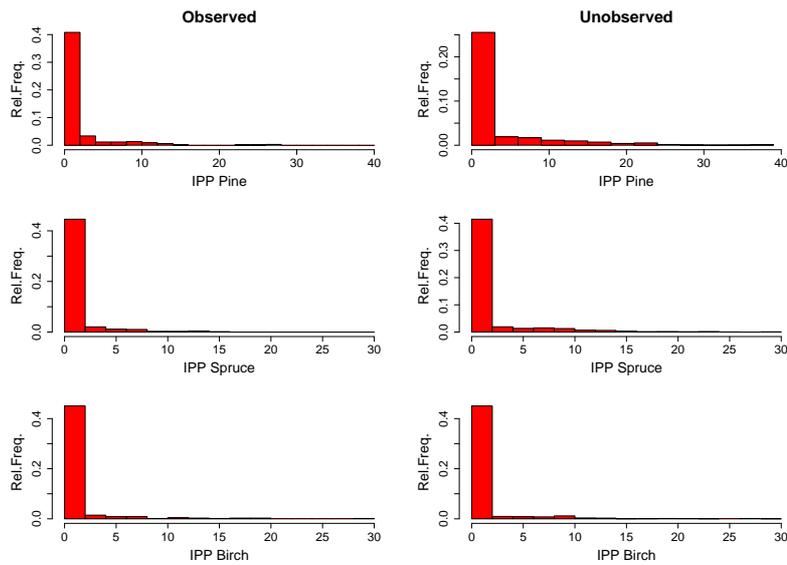


Figure 1: *Vaccinium vitis-idaea*: Histograms of IPP of pine, spruce, and birch at $c = 1$ for observed (left column) and unobserved (right column) patterns.

Table II.2 shows the results of the three models fitted during the analysis. A model with only main effects was first attempted (model 1), which was significant when compared to the null model (model 0). The influence from birch, however, was not significant according to the Wald's test. A new model with only pine and spruce was then considered (model 2). The difference in the log likelihoods with and without the influence of birch (-498.97 and -498.25) was not significant, meaning that incorporating birch did not provide any additional valuable information to the model.

The log likelihood of model 2 with two covariates was also significant compared with the null model. The interaction between pine and spruce was included (model 3), but it did not improve the explanatory power. The results suggest that all three models are basically equivalent in terms of the information contained in them; therefore, based on parsimony, the model with pine and spruce was deemed as the most appropriate one.

In the delta-beta graphs (Fig. 2) several highly influential matched sets were identified. The effect of these sets was large enough that deleting them resulted in a significant improvement in the fit, which might imply

Table II.2: *Vaccinium vitis-idaea*: Conditional logistic models that were fitted with $c = 1$.

| Model | No. coef. | Log lik | Odds ratio | | | |
|-------|-----------|---------|------------|--------|-------|-------------|
| | | | Pine | Spruce | Birch | Pine*Spruce |
| 0 | 0 | -528.20 | | | | |
| 1 | 3 | -498.25 | 0.91 | 0.84 | 1.03 | |
| 2 | 2 | -498.97 | 0.91 | 0.84 | | |
| 3 | 3 | -498.90 | 0.91 | 0.84 | | 1.00 |

Table II.3: *Vaccinium vitis-idaea*: Estimates of model 2 for pine and spruce at $c = 1$.

| | Coef. | s.e.(coef) | Odds ratio | 95% CI Odds ratio |
|--------|---------|------------|------------|-------------------|
| Pine | -0.0964 | 0.0221 | 0.908 | (0.870; 0.948) |
| Spruce | -0.1763 | 0.0352 | 0.838 | (0.782; 0.898) |

overdispersion (Collett, 1991). The reason for their large effect was the higher influence of pine and spruce in the observed patterns than in the unobserved, contrary to what the model dictates; in other words, these plots did not fit well with the model. These outlying plots, however, did not show any special characteristics that justified deleting them, and the results of model 2 with the 342 original plots were retained.

Detailed results of model 2 with all the plots are provided in Table II.3. As described earlier, the odds ratio is a more useful measure in these type of problems than $P(\mathbf{w}_k^{(0)} | z_k)$. The estimated odds ratio according to this model is

$$\begin{aligned} \hat{\Psi} &= \exp \left[-0.0964 \left(\mathbf{w}^{(j)} \mathbf{X}_{\text{pine}} - \mathbf{w}^{(0)} \mathbf{X}_{\text{pine}} \right) - \right. \\ &\quad \left. - 0.1763 \left(\mathbf{w}^{(j)} \mathbf{X}_{\text{spruce}} - \mathbf{w}^{(0)} \mathbf{X}_{\text{spruce}} \right) \right] \\ &= \exp \left[-0.0964 \left(\text{IPP}_{\text{pine}}^{(j)} - \text{IPP}_{\text{pine}}^{(0)} \right) - 0.1763 \left(\text{IPP}_{\text{spruce}}^{(j)} - \text{IPP}_{\text{spruce}}^{(0)} \right) \right]. \end{aligned}$$

This model gives information on how IPP from pine and spruce affect the odds of an observed pattern. It suggests that, when IPP from pine in-

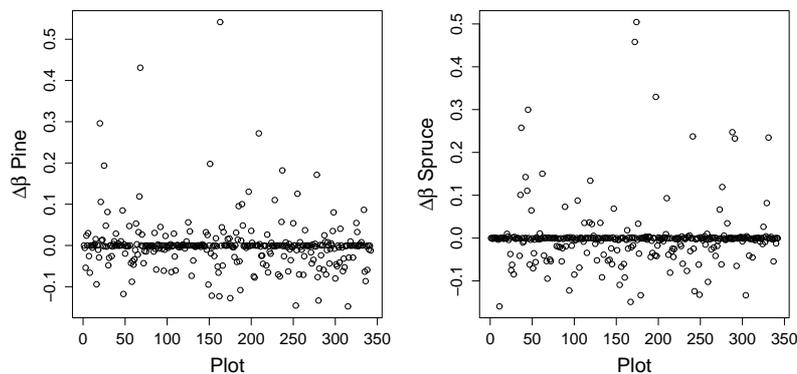


Figure 2: *Vaccinium vitis-idaea*: Standardized $\Delta\beta_i$ against plot number from model with main effects of pine and spruce at $c = 1$.

creases in one unit, the odds of finding the observed pattern decreases by a factor of $\exp(-0.0964) = 0.91$ compared with that of an unobserved pattern, given that the patterns belong to the same plot and IPP for spruce is kept constant. In similar way, the odds of the observed pattern are $\exp(-0.1763) = 0.84$ times that of an unobserved one when IPP from spruce increases by one. A graph with the estimated odds ratio as a function of the change in IPP for pine and spruce can be seen in Fig. 3.

6.2 VACCINUM VITIS-IDAEA WITH $c = 10$

The second example fitted the model with $c = 10$. The results did not differ greatly from the first example. The influence from birch was not significant here either: the log likelihood of the model with the three main effects was -496.19 , and with only pine and spruce -497.71 ; the latter was highly significant compared to the null model. Likewise, the interaction between pine and spruce did not provide further information ($\log L = -496.70$). The estimates of the coefficients of the final model, which included only the main effects from pine and spruce, are shown in

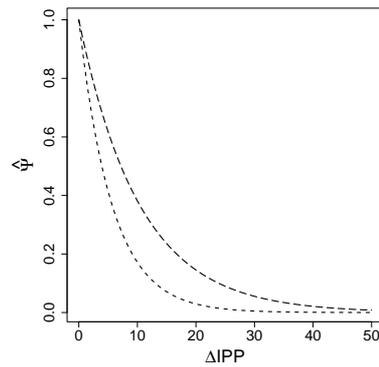


Figure 3: *Vaccinium vitis-idaea*: Estimated odds ratio of model 2 according to the change in IPP at $c = 1$, given that the other covariates are constant: pine (long dash) and spruce (short dash).

Table II.4: *Vaccinium vitis-idaea*: Estimates for the conditional logistic regression with main effects of pine and spruce at $c = 10$.

| | Coef. | s.e.(coef) | Odds ratio | 95% CI Odds ratio |
|--------|---------|------------|------------|-------------------|
| Pine | -0.0319 | 0.00676 | 0.969 | (0.956; 0.982) |
| Spruce | -0.0717 | 0.00866 | 0.950 | (0.934; 0.966) |

Table II.4. The odds ratio for pine was higher than the one from spruce, but both were below 1, indicating that the odds of an observed pattern decreases with larger IPP. The respective 95% confidence intervals were rather narrow: from 0.96 to 0.98 for pine, and from 0.93 to 0.97 for spruce. The delta-beta graphs were less compact, but no obvious outlying plots appeared (Fig 4).

7 CONCLUSIONS

Logistic regression is an appropriate method for modeling binary data such as the presence of an understory species. In this study we wished to model the presence as a function of the influence potential IP of tree species, which quantifies short range effects. Since the PSP data was col-

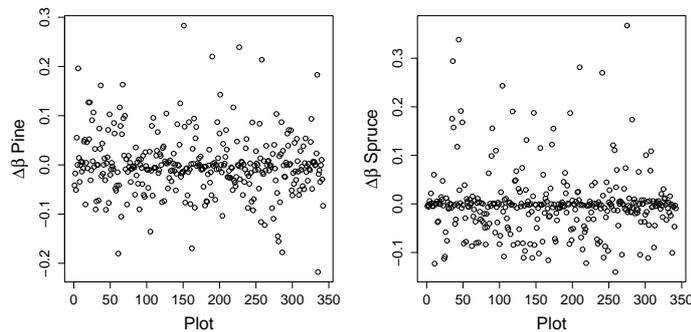


Figure 4: *Vaccinium vitis-idaea*: Standardized $\Delta\beta_i$ against plot number from model with main effects of pine and spruce at $c = 10$.

lected in a very large study area, the measurements on presence and IP implicitly reflect large-scale factors as well as the local-scale effects. To be able to isolate the local-scale effects, we derived a conditional logistic model.

The basis for this model is the pattern of presence and absence in the set of quadrats in the plot. In other words the event of interest is the pattern that has been observed in the field. This is compared to all other patterns that are possible given the number of quadrats in the plot and the number of quadrats occupied by the understory species. With this setting it is not possible to answer directly the initial question as to whether IP can explain the occurrence of a species in a quadrat. Nonetheless, this model gives an idea of how the influence potential of a tree species can modify the relative odds of the presence of an understory species. In more general terms, this model provides some light on the conditions in the plot, while using detailed measurements from the quadrats.

The application of the conditional logistic model to *Vaccinium vitis-idaea* was carried out with the c parameter of the IP function set at 1 and 10. The model that contained the main effects of pine and spruce was the most parsimonious one for both values of c ; the influence from birch as

well as the interaction of pine and spruce were not important. The results indicated that higher values of IPP from pine and from spruce decreased the odds of finding the observed pattern in the forest, but the estimates of the odds ratio for pine and spruce were different for each ,

A considerable amount of variability, however, could not be explained with either c value. Several causes could be pointed out in this regard. Some variation is expected due to measurement error: the data was collected over the summer periods of two years, and during that time a species could have appeared and disappeared in a given quadrat. Otherwise, the more pertinent reason for the high unexplained variation could be the need for further covariates, which is a reasonable argument considering the complexity of the system: dynamics in a forest are due to many other local-scale factors beside the surrounding trees, such as fertility levels (see e.g. Tonteri et al. 1990).

The fit also depended to a certain degree on the parameter c . Although the same covariates were deemed significant when c was 1 and 10, for the latter the fit was better (smaller log likelihood). Additionally there were no obvious outliers in the delta-beta graphs for $c = 10$, implying that c can reduce overdispersion. Moreover, IP of birch was close to being significant when $c = 10$ (p-value of 0.08), which suggests that it could perhaps be included in the model with an even larger parameter. On the other hand, the odds ratio for the smaller c were more interesting, since odds ratios close to 1 indicate that IP does not play a major role in the occurrence of the understory species.

Three main aspects have been identified for future research:

1. The way the goodness of fit and the odds ratio changed according to c point towards the importance of finding its "correct" value, and maybe even a different one for each tree species. Only two values were explored during this study, and further analyzes are required to determine the behavior of the fit as a function of c . To find optimal c 's statistical criteria can be used, but the "correct" values should also have ecological relevance and interpretation.
2. The conditional logistic model proposed for the PSP data depends on the assumption of independence between the quadrats of the same plot. A more general alternative is to include random effects

as a component for the correlation among the quadrats; this could also solve overdispersion problems (Collett, 1991). A recent paper by Chowdhury and McGilchrist (2001) explores such a possibility in this type of models using a generalized linear mixed model (GLMM) approach.

3. Further improvements could be possible by reconsidering the use of the covariate $IPP = \mathbf{w}IP$ to represent the influence potential on the pattern. The current method agrees with the definition of IP used in this study, in the sense that it describes the influence as a sum of the effects; IPP corresponds then to the sum of the IP values of those quadrats where the understory species was observed. It could be argued that this strategy does not represent the ecological dynamics correctly, and a multiplicative model for IP and for the patterns could be more appropriate.

NOTATION

| Symbol | Definition |
|----------------------|---|
| β | : vector of coefficients |
| IP | : influence potential |
| IPP | : influence potential on a pattern |
| j | : index for pattern, $j = 0, \dots, M_k$ |
| k | : index for plot |
| $M_k + 1$ | : number of possible patterns |
| M | : total number of patterns in all plots |
| n_k | : number of quadrats in plot k |
| Ψ | : odds ratio |
| q | : index for quadrat, $q = 1, \dots, n_k$ |
| W_{kq} | : presence/absence in quadrat q |
| $\mathbf{w}_k^{(0)}$ | : observed pattern of presence/absence |
| $\mathbf{w}_k^{(j)}$ | : pattern of presence/absence |
| \mathbf{x}_{kq} | : row vector of covariates for quadrat q |
| \mathbf{x}_{ki} | : column vector of covariate i |
| \mathbf{X}_k | : matrix of covariates |
| z_k | : number of presence observed in the plot k |

A APPENDIX: LIKELIHOOD OF CONDITIONAL LOGISTIC MODEL

A.1 LIKELIHOOD

$$\begin{aligned}
 L &= \prod_k P(\mathbf{w}_{0k} \mid z_k) \\
 &= \prod_k \left\{ 1 + \sum_{j=1}^{M_k} \exp \left[\left(\mathbf{w}_k^{(j)} \mathbf{X}_k - \mathbf{w}_k^{(0)} \mathbf{X}_k \right) \boldsymbol{\beta} \right] \right\}^{-1}
 \end{aligned}$$

$$\begin{aligned}
 l &= \log L \\
 &= - \left\{ \sum_k \ln \left\{ 1 + \sum_j \exp \left[\left(\mathbf{w}_k^{(j)} \mathbf{X}_k - \mathbf{w}_k^{(0)} \mathbf{X}_k \right) \boldsymbol{\beta} \right] \right\} \right\}
 \end{aligned}$$

A.2 FIRST DERIVATIVE

(\mathbf{X}_k represents the matrix of covariates and patterns; \mathbf{x}_{ki} the column vector of covariate i).

$$\frac{\delta l}{\delta \beta_i} = - \sum_k \frac{\sum_{j=1}^{M_k} \exp \left[\left(\mathbf{w}_k^{(j)} \mathbf{X}_k - \mathbf{w}_k^{(0)} \mathbf{X}_k \right) \boldsymbol{\beta} \right] \cdot \left(\mathbf{w}_k^{(j)} \mathbf{x}_{ki} - \mathbf{w}_k^{(0)} \mathbf{x}_{ki} \right)}{1 + \sum_{j=1}^{M_k} \exp \left[\left(\mathbf{w}_k^{(j)} \mathbf{X}_k - \mathbf{w}_k^{(0)} \mathbf{X}_k \right) \boldsymbol{\beta} \right]}$$

A.3 SECOND DERIVATIVE

$$\begin{aligned}
\frac{\delta^2 l}{\delta \beta_i \delta \beta_m} &= \\
&= \sum_k \left\{ \frac{1}{\left\{ 1 + \sum_{j=1}^{M_k} \exp \left[\left(\mathbf{w}_k^{(j)} \mathbf{X}_k - \mathbf{w}_k^{(0)} \mathbf{X}_k \right) \beta \right] \right\}^2} \right. \\
&\cdot \left[\sum_{j=1}^{M_k} \exp \left[\left(\mathbf{w}_k^{(j)} \mathbf{X}_k - \mathbf{w}_k^{(0)} \mathbf{X}_k \right) \beta \right] \left(\mathbf{w}_k^{(j)} \mathbf{x}_{ki} - \mathbf{w}_k^{(0)} \mathbf{x}_{ki} \right) \cdot \right. \\
&\cdot \left. \left. \sum_{j=1}^{M_k} \exp \left[\left(\mathbf{w}_k^{(j)} \mathbf{X}_k - \mathbf{w}_k^{(0)} \mathbf{X}_k \right) \beta \right] \left(\mathbf{w}_k^{(j)} \mathbf{x}_{km} - \mathbf{w}_k^{(0)} \mathbf{x}_{km} \right) \right] - \right. \\
&\left. - \sum_{j=1}^{M_k} \frac{\exp \left[\left(\mathbf{w}_k^{(j)} \mathbf{X}_k - \mathbf{w}_k^{(0)} \mathbf{X}_k \right) \beta \right] \left(\mathbf{w}_k^{(j)} \mathbf{x}_{ki} - \mathbf{w}_k^{(0)} \mathbf{x}_{ki} \right) \left(\mathbf{w}_k^{(j)} \mathbf{x}_{km} - \mathbf{w}_k^{(0)} \mathbf{x}_{km} \right)}{1 + \exp \left[\left(\mathbf{w}_k^{(j)} \mathbf{X}_k - \mathbf{w}_k^{(0)} \mathbf{X}_k \right) \beta \right]} \right\}
\end{aligned}$$

References

- Anon. (2000), *Finnish statistical yearbook of forestry*, Finnish Forest Research Institute.
- Baddeley, A. and Gill, R. D. (1997), "Kaplan-Meier Estimators of Distance Distributions for Spatial Point Processes," *The Annals of Statistics*, 25, 263–292.
- Breslow, N. E. and Day, N. E. (1980), *Statistical methods in cancer research 1: The analysis of case-control studies*, I.A.R.C.
- Capobianco, R. and Renshaw, E. (1998), "The autocovariance function for marked point processes: A comparison between two different approaches," *Biometrical Journal*, 40, 431–446.
- Chowdhury, S. R. and McGilchrist, C. A. (2001), "Matched Case Control Studies with Random Exposure Effects," *Biometrical Journal*, 43, 271–281.
- Collett, D. (1991), *Modelling Binary Data*, Chapman & Hall.
- Cressie, N. (1991), *Statistics for Spatial Data*, John Wiley & Sons.
- Cressie, N. and Brant Collins, L. (2001), "Analysis of spatial point patterns using bundles of product density LISA functions," *Journal of Agricultural, Biological, and Environmental Statistics*, 6, 118–135.

- Diggle, P. J. (1979), "On Parameter Estimation and Goodness-of-fit Testing for Spatial Point Patterns," *Biometrics*, 35, 87–101.
- (1983), *Statistical Analysis of Spatial Point Patterns*, Mathematics in Biology, Academic Press.
- Doguwa, S. (1989), "A comparative study of the edge-corrected kernel-based nearest neighbour density estimators for point processes," *Journal of Statistical Computation and Simulation*, 33, 83–100.
- Doguwa, S. I. and Upton, G. J. G. (1990), "On the Estimation of the Nearest Neighbour Distribution, $G(t)$, for Point Processes," *Biometrical Journal*, 32, 863–876.
- Fiksel, T. (1988), "Edge-Corrected Density Estimators for Point Processes," *Statistics*, 19, 67–75.
- Floresroux, E. M. and Stein, M. L. (1996), "A new method of edge correction for estimating the nearest neighbor distribution," *Journal of Statistical Planning and Inference*, 50, 353–371.
- Gavrikov, V. and Stoyan, D. (1995), "The use of marked point processes in ecological and environmental forest studies," *Environmental and Ecological Statistics*, 2, 331–344.
- Hanisch, K. H. (1984), "Some Remarks on Estimators of the Distribution Function of Nearest Neighbour Distance in Stationary Spatial Point Processes," *Mathematische Operationsforschung und Statistik*, 15, 409–412.
- Hosmer, D. W. and Lemeshow, S. (1989), *Applied logistic regression*, John Wiley & Sons.
- Kingman, J. F. C. (1993), *Poisson Processes*, Oxford University Press.
- Korpela, L. and Reinikainen, A. (1996), "A numerical analysis of mire margin forest vegetation in South and Central Finland," *Ann. Bot. Fennici*, 33, 183–197.
- Kühlmann, S., Heikkinen, J., Särkkä, A., and Hjorth, U. (2001), "Relating abundance of ground vegetation species and tree patterns using ecological field theory," in *Proceedings of IUFRO 4.11 Conference: Forest Biometry, Modelling and Information Science*, ed. Rennolls, K., University of Greenwich, Greenwich, UK: cms1.gre.ac.uk/conferences/iufro/proceedings/.

- Kuuluvainen, T., Hokkanen, T. J., Järvinen, E., and Pukkala, T. (1993), "Factors related to seedling growth in a boreal Scots pine stand: a spatial analysis of a vegetation-soil system," *Canadian Journal of Forest Research*, 23, 2101–2109.
- Kuuluvainen, T. and Pukkala, T. (1989), "Effect of Scots pine seed trees on the density of ground vegetation and tree seedlings," *Silva Fennica*, 23, 159–167.
- McCullagh, P. and Nelder, J. (1989), *Generalized Linear Models*, Monographs on Statistics and Applied Probability, Chapman and Hall, 2nd ed.
- Økland, R., Rydgren, K., and Økland, T. (1999), "Single-tree influence of understorey vegetation in a Norwegian boreal spruce forest," *OIKOS*, 87, 488–498.
- Penttinen, A., Stoyan, D., and Henttonen, H. M. (1992), "Marked Point Processes in Forest Statistics," *Forest Science*, 38, 806–824.
- Philip, M. S. (1994), *Measuring Trees and Forests*, CABI, 2nd ed.
- Pregibon, D. (1984), "Data analytic methods for match case-control studies," *Biometrics*, 40, 639–651.
- Rathbun, S. L. and Cressie, N. (1994), "A Space-Time Survival Point Processes for a Longleaf Pine Forest in Southern Georgia," *Journal of the American Statistical Association*, 89, 1164–1174.
- Reed, M. G. and Howard, C. V. (1997), "Edge-corrected estimators of the nearest-neighbor distance distribution function for three-dimensional point patterns," *Journal of Microscopy*, 186, 177–184.
- Reinikainen, A., Mäkipää, R., Vanha-Majamaa, I., and Hotanen, J. P. (eds.) (2000), *Kasvit Muuttuvassa Metsäluonnossa*, Kustannusosakeyhtiö Tammi.
- Ripley, B. D. (1976), "The second-order analysis of stationary point processes," *Journal of Applied Probability*, 13, 255–266.
- (1977), "Modelling Spatial Patterns (with discussions)," *Journal of the Royal Statistical Society-Series B*, 39, 172–192.

- (1982), "Edge effects in spatial stochastic processes," in *Statistics in Theory and Practice. Essays in Honour of Bertil Matérn*, ed. Ranney, B., Swedish University of Agricultural Sciences, Section of Forest Biometry, pp. 247–262.
- (1988), *Statistical Inference for Spatial Processes*, Cambridge University Press.
- Saetre, P. (1999), "Spatial patterns of ground vegetation, soil microbial biomass and activity in a mixed spruce-birch stand," *Ecography*, 22, 183–192.
- Stein, M. L., Quaschnock, J. M., and Loh, J. M. (2000), "Estimating the K function to a point process with an application to cosmology," *The Annals of Statistics*, 28, 1503–1532.
- Stoyan, D., Kendall, W. S., and Mecke, J. (1995), *Stochastic Geometry and its Applications*, John Wiley & Sons, 2nd ed.
- Stoyan, D. and Penttinen, A. (2000), "Recent Applications of Point Process Methods in Forestry Statistics," *Statistical Science*, 15, 61–78.
- Stoyan, D. and Stoyan, H. (1994), *Fractals, Random Shapes and Point Fields. Methods for Geometrical Statistics*, John Wiley & Sons.
- Tonteri, T., Mikkola, K., and Lahti, T. (1990), "Compositional gradients in the forest vegetation," *Journal of Vegetation Science*, 1, 691–698.
- Walker, J., Sharpe, P. J. H., Penridge, L. K., and Wu, H. (1989), "Ecological Field Theory: the concept and field tests," *Vegetatio*, 83, 81–95.
- Woodward, M. (1999), *Epidemiology: study design and data analysis*, Chapman & Hall.
- Wu, H., Sharpe, P. J. H., Walker, J., and Penridge, L. K. (1985), "Ecological Field Theory: A spatial analysis of resource interference among plants," *Ecological Modelling*, 29, 215–243.