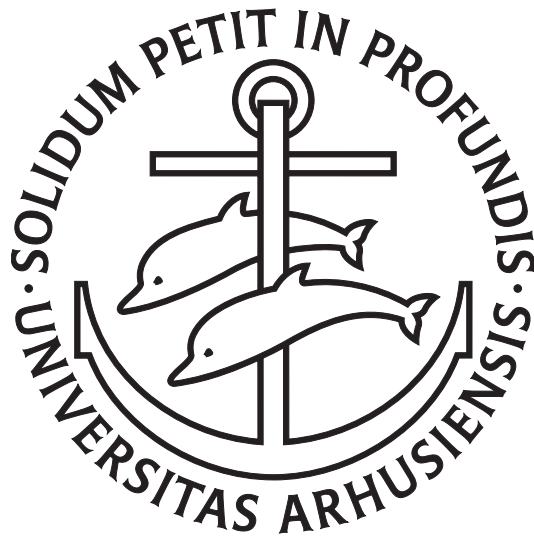# STATISTICAL INFERENCE FOR MICROSCOPY DATA

## – EFFICIENT SAMPLING AND MODELLING BY SPATIAL POINT PROCESSES



PhD Dissertation

## INA TROLLE ANDERSEN

Supervised by Eva B. Vedel Jensen, Ute Hahn & Jens R. Nyengaard

## Supervisors

Professor Eva B. Vedel Jensen
Department of Mathematics
Aarhus University
Denmark


Associate professor Ute Hahn
Department of Mathematics
Aarhus University
Denmark


Professor Jens R. Nyengaard
Department of Clinical Medicine – Stereological Research Laboratory
Aarhus University
Denmark

# Preface

This PhD thesis has been submitted to the Faculty of Science and Technology at Aarhus University. The work was conducted at the Department of Mathematics under the supervision of Associate Professor Ute Hahn in collaboration with Professor Eva B. Vedel Jensen and Professor Jens R. Nyengaard. The work was partly financed by the Centre for Stochastic Geometry and Advanced Bioimaging (CSGB), funded by the Villum Foundation.

The thesis comprises an introduction followed by four self-contained papers published in international journals or as research reports, and two supplementary chapters:

**Paper A**  Ina Trolle Andersen, Ute Hahn and Eva B. Vedel Jensen (2015). Optimal PPS sampling with vanishing auxiliary variables – with applications in microscopy. *Scandinavian Journal of Statistics 42(4)*, 1136–1148.

> **Supplement A**  *Supplementary material which is also available in the online version of Paper A at the publisher's web site.*

**Paper B**  Kresten K. Keller, Ina Trolle Andersen et al. (2013). Improving efficiency in stereology: a study applying the proportionator and the autodisector on virtual slides. *Journal of Microscopy 251(1)*, 68–76.

**Paper C**  Ina Trolle Andersen and Johanna F. Ziegel (2014). 2D non-uniform systematic sampling. *CSGB Research Report 2014–14, Department of Mathematics, Aarhus University.*

**Paper D**  Ina Trolle Andersen and Ute Hahn (2016). Matérn thinned Cox processes. *Spatial Statistics 15*, 1–21.

> **Supplement D**  Limit results and parameter estimation in Matérn thinned Matérn cluster processes. *Unpublished.*

The introduction consists of two parts concerning sampling (Paper A, B and C) and point processes (Paper D). These parts of the thesis present the theoretical background and research question of interest and summarize the main results. Additionally, Supplement D elaborates on Paper D with theoretical results and a simulation study.

# Acknowledgement

First and foremost, I would like to express my deep appreciation and gratitude to my supervisors, Ute Hahn, Eva B. Vedel Jensen and Jens R. Nyengaard, to whom I am indebted for their time, advise and training as a researcher. Ute Hahn, your devotion to research and scientific curiosity inspires me. I am grateful for our many interesting scientific discussions and your encouragement when the project troubled me. Your support during all phases of my research has been invaluable. I would also like to thank you for your personal dedication. You often have fun anecdotes, create irresistible cakes and I learned that you take a heartfelt interest in people around you. Eva B. Vedel Jensen, I appreciate the countless advices you have given me. You have a unique ability to find the possible path among all the impossible ones. Your attention to detail combined with a broad overview and your optimistic mind helped me complete my research. Jens R. Nyengaard, you have broadened my horizon within microscopy and stereology and enabled me contribute to the interdisciplinary research of statistics in microscopy. I thank you for sharing your vast knowledge of both theoretical and practical aspects of research within microscopy. I am truly fortunate to have had the opportunity to work with such inspiring researchers, as part of CSGB.

Furthermore, I would like to thank Johanna F. Ziegel, Kresten K. Keller and Ann Brinch Madelung. This thesis is the fruit of inspiring collaborations with these bright people. In a similar vein, I would like to thank colleagues at B.3 at the Department of Mathematics, in particular the girls in our "famous" Statistical Knitting Club, for providing a pleasant and inspiring working atmosphere. A special thanks goes to my office mate, Britta Anker Bak, for giving ear to my frustrations and to distract me with exciting informal conversations in the need of a break.

A humble thanks go to my friends, family and family-in-law, for being there when needed and creating a balance and perspective to my life. Thanks for all the "hygge" and great moments. Last, but certainly not least, I would like to acknowledge the sacrifices and support by my love, Søren, while I pursued this degree. You had faith in me and gave me the best gift in life, our son Holger, who's bright blue eyes and contagious laugh helped me through the last and most difficult time. I look forward to our future adventures.

# Summary

This PhD thesis presents advances in non-uniform sampling in microscopy. Furthermore, the thesis takes up concrete modelling aspects for microscopy data. The developed model class for point processes is an original contribution to the theory of spatial point processes.

Recent improvements in computerized image analysis of microscopy sections have opened up the possibility for obtaining precise quantitative information about inhomogeneous spatial structures. The idea is to use image analysis for the identification of the regions with high contents of the spatial structure of interest. Subsequently, small fields of view (FOV) of the section are considered at high magnification, thereby allowing for precise measurements of the structure by an expert user. The FOV are chosen according to a sampling design, where the sample inclusion probabilities depend on the auxiliary information from the image analysis. Thereby, efficient estimators can be obtained for the structure on the whole section.

In microscopy, the most commonly used non-uniform sampling design, the proportionator, is a systematic sampling design with sample inclusion probabilities proportional to the size of an auxiliary variable (PPS sampling). However, a common phenomenon in microscopy is vanishing auxiliary variables, i.e. zero-valued auxiliary variables with corresponding positive variables of interest. In PPS sampling this phenomenon makes part of the sample population inaccessible, hence it introduces a bias. In Paper A, we address this problem and propose a modified sampling design, which is optimal in a model-assisted sampling set-up.

In a practical application (Paper B), we compare the efficiency of the proportionator and the efficiency of standard systematic uniform random sampling (SURS). We find the proportionator 45 percent to 90 percent more time efficient than SURS. Furthermore, the applicability of our modified sampling design is illustrated for these data (Paper A).

Designs like the proportionator do not use the spatial information often available prior to sampling in microscopy. In order to preserve and utilize such information, we introduce in Paper C a generalization of non-uniform systematic sampling to the plane. The design has zero estimator variance under optimal auxiliary information. Simulation results show that this design is more efficient than standard sampling designs in a number of cases of area estimation, however, in general the design is not substantially better than simpler alternatives.

In bone marrow pathology, the degree of clustering of cells is used as a diag-

nostic criterion. However, pathologists categorize the degree of clustering by eye using vaguely defined rules. Therefore, we set out to derive statistical inference for clusters of cells. We aimed at describing such cell clusters by simple parametric models. This has the advantage in practice that the parameters often have simple biological interpretations and hypothesis regarding the parameters can be tested. Proper statistical inference can be obtained viewing the centres of the cells as a realization from a point process. For non-overlapping spherical cells, the diameter of the cells determine a lower bound for the distance between the centres. In Paper D, we achieve a combination of clustering and a hard core behaviour by applying dependent Matérn thinning to Cox processes. Formulae for first- and second-order characteristics are established using Palm theory. Simple approximations for these characteristics are suggested, for which the quality is supported by limit results (Supplement D) and simulations (Paper D and Supplement D). This allow for a quantitative characterization through summary statistics. We study parametric inference in the model (Supplement D) and illustrates its applicability for microscopy data (Paper D).

# Dansk Resumé

Denne afhandling præsenterer fremskridt inden for ikke-uniform sampling i mikroskopi. Endvidere beskæftiger afhandlingen sig med konkrete modellerings aspekter vedrørende data fra mikroskopi. Den nye modelklasse af punktprocesser er et originalt bidrag til teorien om rumlige punktprocesser.

Forbedret billedanalyse af snit fra mikroskopi har gjort det muligt at opnå præcise kvantitative informationer om inhomogene rumlige strukturer. Ideen er at benytte billedanalysen til at identificere de dele af snittet, der indeholder meget af den struktur man er interesseret i. Efterfølgende betragtes små observationsvinduer af snittet i høj forstørrelse, hvorpå en ekspert kan lave præcise målinger af strukturen. Observationsvinduerne er valgt ud fra et samplings design, hvor sample inklusionssandsynlighederne er baseret på den ekstra information der er tilgængelig fra billedanalysen. Derved kan der opnås præcise estimater for strukturen på hele snittet.

I mikroskopi er det oftest anvendte ikke-uniforme sampling design, proportionatoren, et systematisk sampling design med inklusionssandsynligheder, der er proportionale med en hjælpevariabel (PPS sampling). Imidlertid er det dog et udbredt fænomen i mikroskopi at hjælpevariablen kan tage værdien nul selv hvis interessevariablen tager en positiv værdi. I PPS sampling resulterer dette fænomen i, at en del af sample populationen er utilgængelig, hvorved der introduceres et bias. I Artikel A betragter vi dette problem og foreslår et modificeret sampling design, som er optimal under et model-assisteret set-up.

I en praktisk anvendelse (Artikel B), sammenlignes effektiviteten af proportionatoren med effektiviteten af standard systematisk uniform random sampling (SURS). Vi fandt proportionatoren 45 procent til 90 procent mere tidseffektiv end SURS. Endvidere blev anvendeligheden af vores modificerede sampling design illustreret for disse data (Artikel A).

Designs såsom proportionatoren anvender ikke den rumlige information, som ofte er tilgængelige forud for sampling i mikroskopi. For at bevare og udnytte denne information, introducerer vi i Artikel C en generalisering af ikke-uniform systematisk sampling til planen. Med dette design opnår man, at variansen på estimatet er nul, når optimale hjælpevariabler er tilgængelige. Simulationer viser, at dette design er mere effektivt end standard sampling designs i en række tilfælde af areal estimation, men generelt er designet ikke betydeligt bedre end simplere alternativer.

Inden for knoglemarvspatologi anvendes graden af klyngedannelse af celler som et diagnostisk kriterium. Patologerne kategoriserer graden af klyngedannelse på øjemål ved brug af nogle svagt definerede regler. Dette motiverede os til at udlede statistisk inferens for klyngedannelse af celler. Vi sigtede mod at kunne beskrive sådanne cellegrupperinger ved simple parametriske modeller. I praksis har det den fordel, at parametrene ofte har simple biologiske fortolkninger. Endvidere kan hypoteser vedrørende parametrene testes. Statistisk inferens kan opnås ved at betragte midtpunkterne af cellerne som en realisation af en punktproces. For ikke-overlappende sfæriske celler vil diameteren af cellerne bestemme en nedre grænse for afstanden mellem midtpunkterne. I Artikel D foreslår vi at anvende afhængig Matérn udtynding på Cox processer for at modellere klyngedannelse med en mindste positiv afstand mellem punkterne. Formler for første- og anden-ordensegenskaber bestemmes ud fra Palm teori. Simple approksimationer for disse egenskaber foreslås, for hvilke kvaliteten understøttes af grænseværdiresultater (Supplement D) og simuleringer (Artikel D og Supplement D). Dette giver mulighed for en kvantitativ karakterisering ved hjælp af summary statistics. Vi studerer parametrisk inferens i modellen (Supplement D) og illustrerer dets anvendelighed for data fra mikroskopi (Artikel D).

# Contents

# Part I

# Introduction

# Contributions to sampling in microscopy

## 1.1   Introduction

Analysis of spatial material by microscopy requires measurements on sections through the material. To reduce workload, the sections are partitioned into small *fields of view* (FOV) and a sample of these are considered at high magnification. Usually, the sample is chosen as a systematic uniform random set of FOV as illustrated in Figure 1.1. See also Baddeley and Jensen (2005, Section 12.4.2).
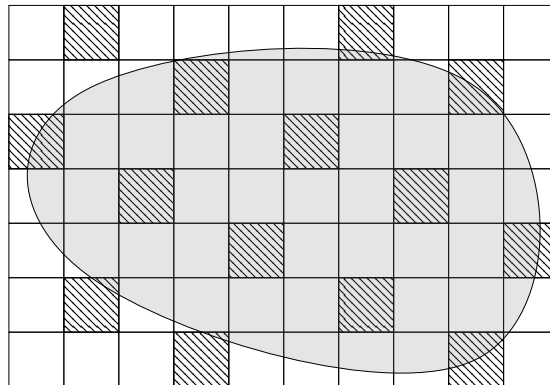


**Figure 1.1:** Illustration of a systematic sample of fields of view (FOV) from a section. Sampled FOV are shown hatched.

Biological material, however, is quite often inhomogeneous. For such material, *systematic uniform random samples* (SURS) of FOV consist frequently of many FOV with no or very small parts of the material of interest. The sampling scheme may be improved if the microscopy images allow for automatic identification of regions with high contents of the material of interest. An alternative uniform sampling scheme using such information, the so-called *smooth fractionator*, was proposed in Gundersen (2002). Later, Gardi et al. (2008a,b) developed this type of sampling into a systematic *non*-uniform random sampling scheme, called the *proportionator*, which is based on probability proportional to size (PPS) sampling. The proportionator has been shown to reduce the estimator variance considerably (Gardi et al., 2008a,b).

The non-uniform sampling scheme requires for each FOV the measurement

at low magnification of an auxiliary variable. In microscopy applications, this auxiliary variable may vanish for quite a number of FOV and these sampling units are thereby inaccessible by PPS sampling. In this thesis, we suggest a modification of the proportionator design for which an optimal solution can be found using a model-assisted approach. We also develop a genuine 2D non-uniform random sampling design that directly include the spatial information of the FOV in the sampling procedure.

In the following, we start by a description of design-based and model-assisted sampling (Section 1.2). We also give a short account of some of the currently used sampling designs in microscopy. Next, a description is given of the research questions studied, the results obtained and their relation to the state-of-the-art within sampling in microscopy for the three sampling publications of the thesis (Papers A, B and C), see Sections 1.3–1.5. Proofs and results from a robustness study of the optimal design in Paper A is given in Supplement A.

## 1.2   Design-based and model-assisted sampling

We consider a finite population of $N$ units and assume that a realization of a random variable $Y_i$ (the measurement of interest, e.g. number of cells in a FOV) is available for each unit $i$, $i = 1, \ldots, N$. Furthermore, we let $T = \sum_{i=1}^{N} Y_i$ denote the population total and assume that $T$ is the quantity of interest.

In the *design-based* framework, the aim is to estimate $T$ for a realization of $Y = \{Y_1, \ldots, Y_N\}$, based on a random sample $S \subseteq \{1, \ldots, N\}$, independent of $Y$. We shall assume that $S$ is a random sample without replacement of fixed size $n$. A popular estimator of the population total is the Horvitz-Thompson estimator (Särndal et al., 1992, p. 42)

$$\widehat{T}_{HT} = \sum_{i \in S} \frac{Y_i}{\pi_i}, \tag{1.1}$$

where $\pi_i = P(i \in S)$ is the probability that the $i$th unit is included in the sample. The estimator $\widehat{T}_{HT}$ is design-unbiased, i.e. $\mathbb{E}(\widehat{T}_{HT} \mid Y) = T$, if the inclusion probabilities $\pi_i$ are all positive. Design-based inference is often preferable in microscopy applications, as no model assumptions such as stationarity or isotropy are needed. Unfortunately, a general variance minimizing design and estimator (optimal strategy) does not exist in the design-based framework, which motivates some use of models.

In the *model-based* framework a model of $Y_i$ is assumed. The objective is to predict $T$ from a predictor $\widehat{T}$, based on a sample $S$. The predictor $\widehat{T}$ is not necessarily design-unbiased, but instead required to be model-unbiased, i.e. $\mathbb{E}(\widehat{T} - T \mid S) = 0$ for all possible samples $S$. Optimization in this framework is with respect to the model-variance $\text{Var}(\widehat{T} \mid S)$ (Isaki and Fuller, 1982).

Both design-based and model-based sampling for inference about particle populations (in particular, cell populations) are discussed in Baddeley and Jensen (2005, Chapter 10). In Hansen et al. (2011), model-based non-uniform sampling of point

processes was developed for microscopy data to obtain an optimal predictor under a proportional regression model.

A compromise between the design- and model-based framework is the *model-assisted* approach which has been advocated by Särndal et al. (1992). Here, both the design and the assumed model is important, as (approximately) design-unbiased estimation is required and one seek optimality results with respect to *the anticipated variance* of the estimation error $\text{Var}(\widehat{T} - T)$ (Isaki and Fuller, 1982). Under the assumptions that the $Y_i$'s are uncorrelated and $\widehat{T}$ is unbiased, $\mathbb{E}(\widehat{T} - T) = 0$, the anticipated variance simplifies and has following lower bound (Särndal et al., 1992, Remark 12.2.2, p. 453)

$$\text{Var}(\widehat{T} - T) = \mathbb{E}(\widehat{T} - T)^2 \geq \sum_{i=1}^{N} (\frac{1}{\pi_i} - 1)\text{Var}(Y_i). \tag{1.2}$$

The assumption regarding unbiasedness holds when $\widehat{T}$ is design-unbiased. In this case, the mean design variance, the anticipated variance and the mean square error of $\widehat{T}$ are identical,

$$\mathbb{E}\text{Var}(\widehat{T} \mid Y) = \text{Var}(\widehat{T} - T) = \mathbb{E}(\widehat{T} - T)^2. \tag{1.3}$$

Thus minimizing the anticipated variance is the same as minimizing the model expectation of the design variance. When $\mathbb{E}Y_i \propto \pi_i$ we obtain the Godambe-Joshi lower bound in (1.2) of the Horvitz-Thompson estimator.

In order to obtain an efficient estimator of $T$, auxiliary information is often used *directly* in the sampling design to choose the inclusion probabilities wisely, or *indirectly* to choose an efficient design with constant inclusion probabilities. A class of sampling schemes which is widely used in sampling due to its simplicity and efficiency is *(non-uniform) systematic sampling*, see e.g. Särndal et al. (1992). Here, auxiliary information may be used both directly and indirectly. A systematic sampling design with inclusion probabilities $\pi_i$, $i = 1, \dots, N$, and fixed sample size $n$, can be implemented as follows. Let $U \sim \text{Unif}([0, 1])$ and $U_j = U + j - 1$, $j = 1, \dots, n$. Then, the sample $S$ consists of those units $i$ for which a $U_j$ is contained in the interval

$$\Big[ \sum_{l=1}^{i-1} \pi_l, \sum_{l=1}^{i} \pi_l \Big].$$

Auxiliary variables $x_i$ associated with $Y_i$ may be used directly when $x_i$ is expected to be roughly proportional to $Y_i$. Then, the efficiency increases if we let the inclusion probability $\pi_i$ be proportional to $x_i$ (sampling proportional to size, PPS sampling), $\pi_i = n x_i / \sum_{i=1}^{N} x_i$. Auxiliary variables $x_i$ may also be used indirectly to choose an efficient design with constant inclusion probabilities. For instance, we may use the $x_i$s in an ordering of the units. If we simply order the units directly according to the size of the auxiliary variables, $x_1 \leq \dots \leq x_N$, then a linear (or monotonic) trend will often appear in the $y_i$s which means that the estimate $\widehat{T}$ will be monotonic with respect to the starting point of the samples. *Balanced systematic sampling*, described in Murthy et al. (1967, Section 5.9d) and illustrated in Figure 1.2, is a uniform

$x_1 < x_2 < \ldots$                                          $< x_N$

$x_{[1]} < x_{[2]} < \ldots$                                          $< x_{[N]}$
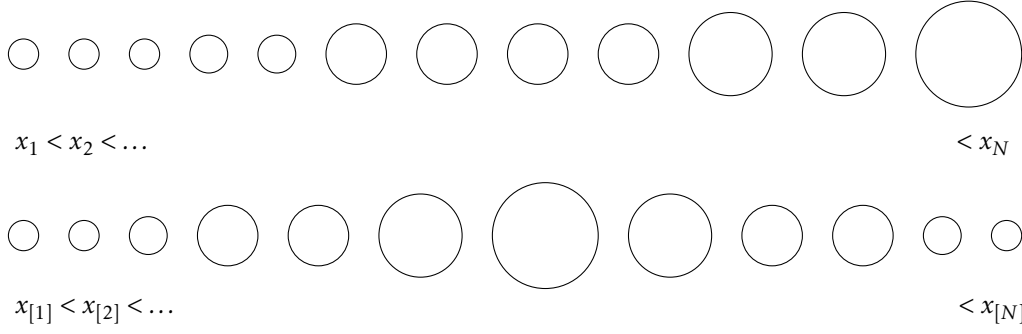
**Figure 1.2:** Ordered and balanced systematic sampling.

systematic sampling design with an ascending-descending ordering which is based on the idea of pairing the units together in such way that the trend cancels out. More precisely, if we let $1, \ldots, N$ denote the ordering such that $x_1 \leq x_2 \leq \ldots \leq x_N$, $M = \frac{N}{2}$ if $N$ is even and $M = \frac{N+1}{2}$ if $N$ is odd, the balanced ordering becomes $[1], \ldots, [N]$, where

$$[i] = \begin{cases} 2i - 1, & i \leq M, \\ 2(N + 1 - i), & i > M. \end{cases} \tag{1.4}$$

Balanced systematic sampling and other uniform systematic designs e.g. *modified systematic sampling* (Singh et al., 1968), have been considered in details under different models to find the optimal ordering. Balanced sampling with $x_i = i$ has been proven to be superior to uniform systematic sampling with an ordering according to increasing unit numbers when linear trend ($\mathbb{E}Y_i = a + bi, i = 1, \ldots N$) is present (Bellhouse and Rao, 1975).

When using the balanced uniform systematic sampling in stereology, the method is referred to as the smooth fractionator as proposed by Gundersen (2002) and considered for microscopy data in Gardi et al. (2006). Theoretical results with respect to optimal orderings for non-uniform systematic sampling do not exists, yet Gardi et al. (2008a,b) adapted the idea to their PPS sampling design, the proportionator, which now used in microscopy.

## 1.3   Optimal PPS Sampling (Paper A)

In applications in microscopy, a non-negligible part of the population under study may be inaccessible under PPS sampling due to vanishing auxiliary variables. In such cases, the Horvitz-Thompson estimator may be biased. Suggestions for resolving the situation have been given in Gardi et al. (2008a,b), but with the expense of an unnecessary large estimator variance. In Paper A, we propose a modification of the systematic PPS sampling design using stratified sampling for which an optimal solution can be found. The approach is model-assisted and optimality is in terms of minimal mean design variance.

### 1.3.1 Design and model set-up

As described in details in Section 1.2, we assume realizations of uncorrelated random variables $Y_i$ (the variable of interest), and associated auxiliary variables $x_i$, $i = 1\ldots,N$, are available. Let $Y = \{Y_1,\ldots,Y_N\}$ and $x = \{x_1,\ldots,x_N\}$. The aim is to estimate $T = \sum_i^N Y_i$ for a realization of $Y$ based on a sample $S$ using the Horvitz-Thompson estimator $\widehat{T} = \widehat{T}_{HT}$. As $x$ is completely known in the applications we have in mind, modelling and inference about $Y$ will be made in the conditional distribution given $x$. Thus, we may think of $x$ as non-random variables.

The common characteristics of the sampling designs we optimize in Paper A are fixed sample size $n$ and inclusion probabilities $\tilde{\pi}_i$, say, with

$$\tilde{\pi}_i = 0, \quad i = 1,\ldots,N_0, \tag{1.5}$$

$$\tilde{\pi}_i > 0, \quad i = N_0 + 1,\ldots,N. \tag{1.6}$$

(We use the notation $\tilde{\pi}_i$ here because $\pi_i$ is reserved for the modified design presented below.) Note that for $i = 1,\ldots,N_0$, the inclusion probabilities may in fact be zero or have been set to zero, because they are unknown. Consequently, part of the population is inaccessible. Hence, the purpose is to find an optimal modification of the design, such that the first $N_0$ units are assigned a constant *positive* inclusion probability and the remaining units have inclusion probabilities proportional to the original ones. The modified design should still have fixed sample size $n$.

In the microscopy example which motivated this set-up, we consider systematic PPS sampling with $\tilde{\pi}_i = nx_i/x_\bullet$, $x_\bullet = \sum_{i=1}^N x_i$ and $x_1 = \cdots = x_{N_0} = 0$. The modification proposed in Gardi et al. (2008a,b) is simply to let $x_i = \varepsilon > 0$, $i = 1,\ldots N_0$. This design is called *ε-corrected systematic PPS sampling* and is illustrated in Figure 1.3. Letting $\varepsilon$ be too small will lead to a large estimator variance. For instance, if there exists an $i$ with $Y_i > 0$ and $x_i = \varepsilon$, then

$$\text{Var}(\widehat{T} \mid Y) \geq \sum_{i=1}^N \frac{1}{\pi_i} Y_i^2 - T^2 \to \infty, \quad \text{as} \quad \varepsilon \to 0. \tag{1.7}$$

Therefore, we seek an optimal $\varepsilon$, which minimizes the variance.

The crucial step in the optimization is to realize that $\varepsilon$-corrected systematic PPS sampling has a built-in stratification mechanism, such that an almost fixed fraction of the sampled units belongs to the two strata, Stratum 0 ($1,\ldots,N_0$) and Stratum 1 ($N_0 + 1,\ldots,N$). The expected number of units sampled in Stratum 0 is given by

$$n_0 = n\frac{N_0\varepsilon}{x_\bullet + N_0\varepsilon}. \tag{1.8}$$

Hence, finding an optimal $\varepsilon > 0$ is equivalent to finding an optimal sample size $n_0 > 0$ in Stratum 0.

In Theorem 1 in Paper A, which holds under the mild assumption $\mathbb{E}Y_i = \mathbb{E}Y_j$, $i,j = 1,\ldots,N_0$, we verify to which extent the mean design variance of $\widehat{T}$ changes when it is modified into a genuinely stratified design, i.e. a design with independent
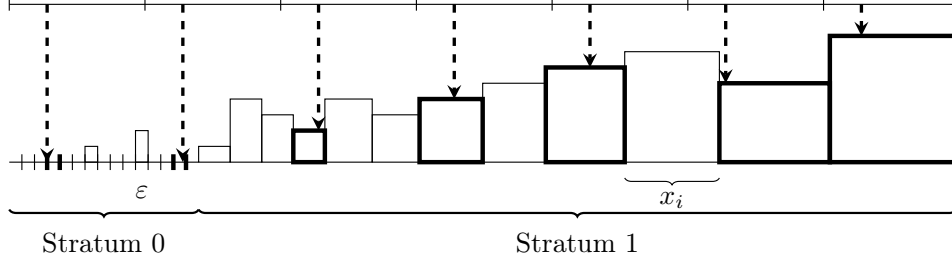
**Figure 1.3:** Illustration of $\varepsilon$-corrected systematic PPS sampling where a small constant $\varepsilon > 0$ has been added to each unit with $x_i = 0$ to ensure an unbiased estimator. Each box represents a sampling unit $i$ with height given by $Y_i$ and width proportional to the inclusion probabilities. The dashed arrows illustrate the sampled units in a systematic sample $S$. The sampling scheme has a built-in stratification mechanism such that an almost fixed fraction of the sampled units will be in each stratum. *(Paper A, Figure 2)*

sampling in each of the two strata. In particular, when $n_0$ is an integer, the two types of design have the same mean design variance.

In the general set-up described in (1.5)-(1.6), we similarly propose to sample independently on the two strata, Stratum 0 and Stratum 1, and consider modified designs with integer values of $n_0 > 0$, the sample size in Stratum 0, such that the inclusion probabilities of the modified design are given by

$$\pi_i = \begin{cases} \dfrac{n_0}{N_0}, & i = 1,\ldots,N_0, \\ \left(1 - \dfrac{n_0}{n}\right)\tilde{\pi}_i, & i = N_0 + 1,\ldots,N. \end{cases} \tag{1.9}$$

### 1.3.2  Optimal PPS sampling

In order to use a model-assisted approach to achieve optimal results, we will consider the following model

$$\mathbb{E}Y_i = \beta_0, \quad \mathrm{Var}\,Y_i = \sigma_0^2, \qquad i = 1,\ldots,N_0, \tag{1.10}$$

$$\mathbb{E}Y_i \propto \tilde{\pi}_i, \quad \mathrm{Var}\,Y_i = \sigma_i^2, \qquad i = N_0 + 1,\ldots,N, \tag{1.11}$$

where $\sigma_0^2 > 0$ and $\sigma_i^2 > 0, i = N_0 + 1,\ldots,N$.

In Theorem 2 in Paper A, we consider the Horvitz-Thompson estimator $\widehat{T} = \widehat{T}_0 + \widehat{T}_1$ where $\widehat{T}_0$ and $\widehat{T}_1$ are based on independent samples $S_0$ and $S_1$ in Stratum 0 and Stratum 1, respectively. We find the $n_0$ for which the mean design variance $\mathbb{E}\mathrm{Var}(\widehat{T}|Y)$ is minimized. In particular, we find a simple expression for the optimal $n_0$ under PPS sampling where $\tilde{\pi}_i = nx_i/x_.$. In the microscopy application, this $n_0$ may be converted to an expression for $\varepsilon$.

Further simplifications are possible if we assume a proportional regression model in Stratum 1 and, in addition, we assume that the mean-variance relationship is the same in the two strata, i.e.

$$\mathbb{E}Y_i = \beta_1 x_i, \ \mathrm{Var}\,Y_i = \sigma_1^2 x_i^g, \quad i = N_0 + 1,\ldots,N \quad \text{and} \quad \sigma_0^2/\beta_0^g = \sigma_1^2/\beta_1^g, \tag{1.12}$$

with $1 \leq g \leq 2$. Then, the optimal allocation becomes

$$n_0 = n \frac{\sqrt{N_0^{2-g}(qx.)^g}}{\sqrt{N_0^{2-g}(qx.)^g} + \sqrt{x.(x^{g-1}).}}, \quad (1.13)$$

where $(x^{g-1}). = \sum_{i=N_0+1}^{N} x_i^{g-1}$ and $q = \mathbb{E}T_0/\mathbb{E}T_1 = \frac{\beta_0 N_0}{\beta_1 x.}$. In the special cases with $g = 1$ and $g = 2$, we find

$$n_0 = \begin{cases} n \dfrac{\sqrt{qk}}{1 + \sqrt{qk}}, & g = 1, \\[2ex] n \dfrac{q}{1 + q}, & g = 2. \end{cases} \quad (1.14)$$

### 1.3.3 Robustness

We also investigated the robustness of the optimal allocation solution against parameter misspecifications and departures from the proportional regression model, described above. Numerical calculations as well as simulations from an inhomogeneous Poisson point process suggest that the optimum is robust against two types of misspecifications, namely; the guessed fraction of the variable of interest expected to be found in the part of the population with vanishing auxiliary variables, and departures from proportionality in a Poisson model with $Y_i \sim pois(x_i^\delta)$. Numerical calculation also showed that if the true variance parameter is $g = 1$ then choosing $g = 2$ may result in a substantial loss in efficiency, whereas the opposite case is less pronounced. These results may be found in Supplement A after Paper A in this thesis.

### 1.3.4 Analysis of microscopy data

To see how well the approach works in practice, microscopy data from Paper B were considered. The mean-variance relation in the extended regression model fitted a model with $g = 1$ quite well. The mean proportionality in Stratum 1 was not fulfilled. In fact, a model with $\delta = 2.4$ was more suitable. However, as also seen in the robustness study, lack of proportionality is not critical for the optimum when $g = 1$. The optimal allocation rule in (1.14) with $g = 1$ almost fitted the true optimum. Compared to using the optimum for $g = 2$ the variance was halved. This is illustrated in Figure 1.4.

### 1.3.5 Discussion and perspectives

Paper A shows both theoretically and in applications the consequence that adding a non-optimal correcting-constant $\varepsilon > 0$ has for the estimator variance. In the existing literature, either vanishing auxiliary variables are completely ignored (Hansen et al., 2011) or the importance of the value of $\varepsilon$ is underrated (Gardi et al., 2008a,b). We have shown that a model-assisted approach can be used to achieve optimal results
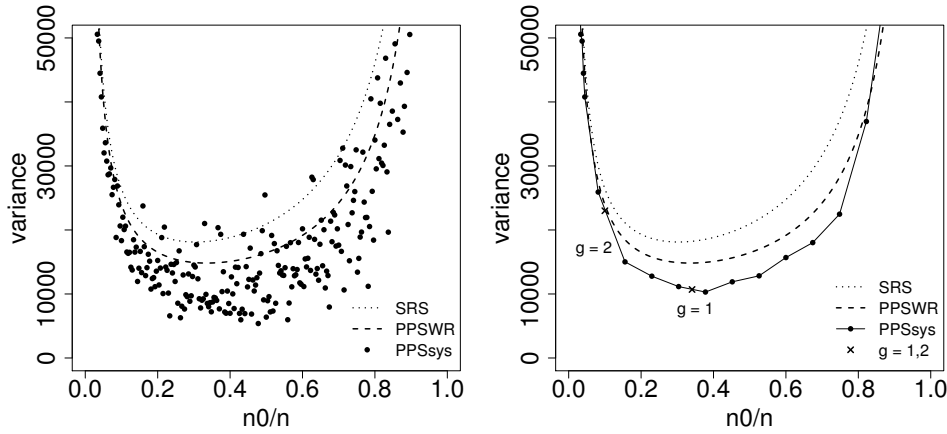
**Figure 1.4:** Left: The variance of simple random sampling (dotted line), PPS with replacement sampling (stippled line) and systematic PPS sampling with an ordering according to the size of the $x_i$s (dots), all under stratification as a function of $n_0/n$. Right: The same plot except that the variance for systematic PPS sampling is obtained by averaging bins of size 20. The sample allocations obtained by (1.14) with $g = 1$ and $g = 2$ are marked with crosses. (*Paper A*, *Figure 6*)

for the design-unbiased Horvitz-Thompson estimator. As the optimal solution only depends on a few parameters and has been shown to be robust, we expect that the optimal estimator will be used in the microscopy community, and also in a more general context where PPS sampling is used, and auxiliary variables can take the value zero.

For further research, it could be interesting to generalize the optimality result in Theorem 2 of Paper A. The result requires that the mean variance do not depend on the second-order inclusion probabilities. This holds under the assumption that $\mathbb{E}Y_i$ is proportional to the inclusion probabilities. However, some sampling designs have the same property for the mean variance using *arbitrary* inclusion probabilities, such as the design considered in Nedyalkova and Tillé (2008), where the samples are balanced on $\mathbb{E}Y_i$. This opens up to further optimization, using optimal inclusion probabilities. In a more general set-up with possible more than one type of auxiliary variable, the general regression estimator may also be considered (Särndal et al., 1992).

## 1.4   Implementation (Paper B)

In Paper B, published in *Journal of Microscopy*, the actual implementation of $\varepsilon$-corrected PPS sampling is discussed. One of the main objectives of the paper is to investigate the time efficiency of this new sampling design and compare it with the time efficiency of systematic uniform random sampling (SURS) in a practical application of cell counting in inhomogeneously distributed cell populations. Paper B is written for microscopy users. In the following, a mathematical and statistical description of the involved geometric sampling and variance estimation procedures
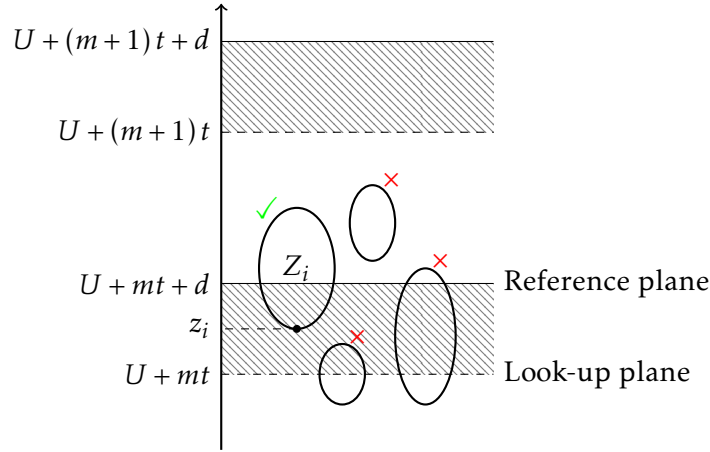
**Figure 1.5:** Illustration of the disector counting rule for particles.

is given.

### 1.4.1 Sampling

Let us consider a finite population of cells, $Z_1, \ldots, Z_N$. They can be arbitrary compact subsets of $\mathbb{R}^3$. Our aim is to estimate $N$ from observations in planar sections. We only consider planes of fixed orientation. Let $T_u$, $u \in \mathbb{R}$, be such a plane with position $u$, see Figure 1.5. A disector of height $d$ is a set of two parallel planes $(T_u, T_{u+d})$ a distance $d$ apart. Here, $T_{u+d}$ is called the reference plane and $T_u$ the look-up plane.

In Paper B, a systematic uniform random set of disectors with distance $t > d$ between neighbour disectors is used in the sampling. More specifically, the disectors are described as

$$(T_{U+mt}, T_{U+mt+d}), \qquad m \in \mathbb{Z}, \ U \sim \text{Unif}[0, t).$$

Let $z_i \in \mathbb{R}$ be the coordinate of the projection of the lower point of the cell $Z_i$ onto a line perpendicular to the planes, see Figure 1.5. The sample $S$ of cells consists of those cells with lower point between the planes of one of the disectors. More precisely,

$$S = \{i \mid \exists m : U + mt \leq z_i < U + mt + d\}. \tag{1.15}$$

The $i$th inclusion probability becomes

$$\pi_i = \sum_m P(U + mt \leq z_i < U + mt + d) \tag{1.16}$$

$$= \sum_m P(U \in [z_i - (mt + d), z_i - mt]) = d/t, \tag{1.17}$$

where we at the last equality sign use that $U$ is uniform and contained in exactly one of the intervals. Thus, all cells are equally likely to be included in the sample.

In Paper B, the number of sampled particles is denoted $Q^-$. We find

$$\mathbb{E}Q^- = \sum_{i=1}^{N} \pi_i = Nd/t, \tag{1.18}$$

so $\hat{N} = \frac{t}{d}Q^-$ is an unbiased estimator of $N$.

The sampling of sections described above requires that it is possible to identify whether the lower point of a cell lies in the space between the two planes of a disector. If this is not possible, the sampling may still be performed if the height of the cells is at least $d$. Then, no cell comes to lie entirely between two planes, and a cell is sampled in a disector if it hits the reference plane but not the look-up plane.

In applications, such as the one described in Paper B, we use sections, which are not exactly two-dimensinal planes but have a certain thickness. The object is cut completely into a random series of sections. The thickness corresponds to the disector height $d$ and $d/t$ will be the section sampling fraction (ssf). See further theoretical details in Baddeley and Jensen (2005) and e.g. Boyce et al. (2010); Evans et al. (2004) for a more applied point of view. Due to the low number of cells in the present study, the disector was applied in two directions, using both sections as reference and look-up section.

In Paper B, each section was subsampled, using either systematic uniform random sampling (SURS) or a version of $\varepsilon$-corrected PPS sampling, the proportionator, by the following procedure. For each sampled disector section, the area of interest was marked and subsequently divided into fields of view (FOVs) using a randomly placed grid and two measurements were determined for each FOV: the cell count ($Y_i = Q^-_{FOVi}$), obtained from manually counting the cells using the disector principle and a weight assignment ($w_i$) which were generated automatically from the image analysis program using a pre-chosen classification protocol based on the amount of colour from a staining. Subsequently, the weights were transformed ($x_i = \log(w_i + 1) + 1$) to avoid vanishing auxiliary variables and to obtain a better weight-count relation (i.e. $\varepsilon = 1$). To imitate a standard application procedure we simulated from the original exhaustive data two independent samples of half the desired sample size for each of the two designs; SURS and the proportionator. We used the Horvitz-Thompson estimator of the total number of cell in the sampled sections for each design, and used as the final estimate for each design, the mean of the two half samples

$$Q^- = \tfrac{1}{2}(Q^-_1 + Q^-_2). \tag{1.19}$$

The estimate $\hat{N}$ of the total number of cells was then simply obtained by scaling $Q^-$ with the inverse section sampling fraction and divided by two due to the double count in the disector.

### 1.4.2   Variance estimation

The precision of the estimator $\hat{N}$ is in Paper B measured by the coefficient of error, $CE(\hat{N}) = \sqrt{\mathrm{Var}(\hat{N})}/\mathbb{E}\hat{N}$. Two sources of randomness influence the precision of the estimator $\hat{N}$: sampling of the sections and sampling within the sections, often referred to as the counting noise. Conditioning on $U$, the random variable determining the sections, we may use the standard variance decomposition

$$CE^2_{\text{total}}\hat{N} = CE^2_{\text{sect}}\hat{N} + CE^2_{\text{noise}}\hat{N}, \tag{1.20}$$

where

$$CE^2_{\text{sect}}\hat{N} = \frac{\text{Var}\mathbb{E}(\hat{N} \mid U)}{\mathbb{E}^2\hat{N}} = \frac{\text{Var}\mathbb{E}(Q^- \mid U)}{\mathbb{E}^2 Q^-}, \tag{1.21}$$

$$CE^2_{\text{noise}}\hat{N} = \frac{\mathbb{E}\text{Var}(\hat{N} \mid U)}{\mathbb{E}^2\hat{N}} = \frac{\mathbb{E}\text{Var}(Q^- \mid U)}{\mathbb{E}^2 Q^-}. \tag{1.22}$$

The estimation of the variance of the counting noise may be performed under a model assumption. In the applied literature, a Poisson model for the cells is typically assumed. In Paper B, we used instead a direct method of estimating the counting noise as the empirical variance of two sub-samples. Ignoring the randomness of the grid of the FOVs, which is justified in the Appendix of Paper B, we have that $Q^-_1$ and $Q^-_2$ are conditionally independent and identically distributed. Thus, $\text{Var}(Q^- \mid U) = 1/2\text{Var}(Q^-_i \mid U)$, $i = 1, 2$, and therefore we may use the unbiased estimate

$$\widehat{\text{Var}}(Q^- \mid U) = 1/2\widehat{\text{Var}}(Q^-_i \mid U) = 1/2[(Q^-_1 - Q^-)^2 + (Q^-_2 - Q^-)^2] \tag{1.23}$$

$$= 1/4(Q^-_1 - Q^-_2)^2, \tag{1.24}$$

which is called $\text{Var}_{\text{noise}}$ in Paper B. Hence, an estimate of $CE^2_{\text{noise}}\hat{N}$ becomes

$$CE^2_{\text{noise}} = \frac{(Q^-_1 - Q^-_2)^2}{(Q^-_1 + Q^-_2)^2}. \tag{1.25}$$

To estimate the variance due to the randomness of the sections, we may write the estimate $Q^-$ as $Q^- = \sum_i Q^-_{\text{sect}i}$, where $Q^-_{\text{sect}i}$ denotes the number of cells sampled in section pair $i$. The variance between sections can be expressed as

$$\text{Var}\mathbb{E}(Q^- \mid U) = \text{Var}\left(\sum_i f(U + it)\right) \tag{1.26}$$

where $f(U + it) = \mathbb{E}(Q^-_{\text{sect}i} \mid U)$ is measured with error by $\hat{f}_i = Q^-_{\text{sect}i}$. The parameter of interest is

$$N = \frac{1}{d}\int f(x)\,\mathrm{d}x. \tag{1.27}$$

The variance of an estimator of such an integral of a measurement function $f$ is a well-studied problem in geometric sampling. In Paper B, we use the following estimate of the variance, including a correction-term ($\text{Var}_{\text{noise}}$) due to the measurement error,

$$CE^2_{\text{sect}} = \frac{3(A - \text{Var}_{\text{noise}}) - 4B + C)/240}{(Q^-)^2} \tag{1.28}$$

(Gundersen et al., 1999), where $A = \sum \hat{f}^2_i$, $B = \sum \hat{f}_i \hat{f}_{i+1}$ and $C = \sum \hat{f}_i \hat{f}_{i+2}$. In Baddeley and Jensen (2005, Chapter 13), approximations of such variances are discussed in detail where background references may also be found.

### 1.4.3   Discussion and perspectives

From the point of view of a statistician, this study yields a unique opportunity to work with data containing all information of the cell population within the sampled sections, not only a few sampled FOV. Hereby, we could detect problems with the weight assignment which e.g. results in a biased estimator due to vanishing weights (see Paper A). Furthermore, we could make a proper comparison between the designs based on exactly the same FOV and approximately the same cell count in the sample. Estimating the variance of systematic sampling is non-trivial, as can be seen from the wealth of literature dealing with this problem, see Baddeley and Jensen (2005, Chapter 13). The variance of systematic sampling of sections can often be estimated from formulae like (1.28), whereas no generally unbiased estimator for the within section variance which is based on only one sample is known. Often a very simple Poisson assumption is made, or, as in Paper B, two independent samples are used. This estimator, called the direct CE in Gardi et al. (2008a,b), is unbiased but can have a very skewed distribution which lead to underestimation in many cases. Consequently, better alternatives for estimation of the within section variance is an important topic for future research.

## 1.5   2D non-uniform sampling (Paper C)

Standard systematic uniform random sampling (SURS) in microscopy is, as described earlier, a sampling design, where each field of view (FOV) is equally likely to be sampled from a finite set of FOVs, and where the sampled FOVs are spatially spread across the section. The recently proposed proportionator design in microscopy (Gardi et al., 2008a,b) is non-uniform and therefore suitable for sampling of inhomogeneous material, but all spatial information is lost, prior to the sampling. Therefore, an idea to improve upon this non-uniform design, is to include spatial information in the sampling procedure.

In Paper C, we introduce a new 2D non-uniform systematic sampling design that balances samples spatially and study its efficiency. This design is a natural generalization of the 1D non-uniform systematic sampling introduced in the microscopy setting by Dorph-Petersen et al. (2000).

### 1.5.1   The generalized 2D design

The objective is to estimate the integral

$$Q = \int_{[0,1]^2} f(x,y)\,\mathrm{d}x\,\mathrm{d}y, \tag{1.29}$$

using values of the measurement function $f$ at a set of randomly sampled points in $[0,1]^2$. The aim is to spread the points spatially such that the resulting estimator is efficient. In microscopy applications, the value of $f$ at a point $(x,y)$ is typically a quantity observed in a small observation window (FOV) located at this point, e.g. the number of cells in the FOV.

To obtain a systematic non-uniform random set of points, we propose to use a systematic uniform random set of points subject to a transformation given by a diffeomorphism $G : [0,1]^2 \rightarrow [0,1]^2$. The aim is to choose $G$ such that the estimator of $Q$ has a low variance.

An unbiased estimator can be found using a set of $n \cdot m$ systematic random points, $G^{-1}((U_1 + i)\frac{1}{n}, (U_2 + j)\frac{1}{m})$, $i = 0, \ldots, n-1$, $j = 0, \ldots, m-1$, where $U_1, U_2 \sim \text{Unif}[0,1)$ are independent. The estimate is given by

$$\hat{Q}_{nm} = \frac{1}{nm} \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} \frac{f(G^{-1}((U_1 + i)\frac{1}{n}, (U_2 + j)\frac{1}{m}))}{|\det(G'(G^{-1}((U_1 + i)\frac{1}{n}, (U_2 + j)\frac{1}{m})))|} \tag{1.30}$$

where $G'$ denotes the Jacobi matrix of $G$. In the Appendix of Paper C, proof is given that $\hat{Q}_{nm}$ is in fact an unbiased estimator of $Q$. Assuming $f(x,y) > 0$ a.s., the variance of $\hat{Q}_{nm}$ is zero if we choose any $G$ with

$$|\det(G'(u,v))| = cf(u,v), \tag{1.31}$$

where $c > 0$. For instance, (1.31) is fulfilled if we let $G = (G_1, G_2)$, where

$$G_1(x,y) = \int_0^x g(u)\,\mathrm{d}u, \quad G_2(x,y) = \frac{1}{g(x)\Delta} \int_0^y f(x,v)\,\mathrm{d}v, \tag{1.32}$$

and

$$g(x) = \frac{1}{\Delta} \int_0^1 f(x,v)\,\mathrm{d}v, \quad \Delta = \int_{[0,1]^2} f(u,v)\,\mathrm{d}u\,\mathrm{d}v. \tag{1.33}$$

The unknown function $f$ in $G$ must be replaced by an auxiliary function $f_0$, ideally proportional to $f$, prior to sampling. In the microscopy applications, $f_0$ could be based on colour values obtained by automatic image analysis of the material of interest. This choice of $G$ is closely related to proportional-to-size sampling in the discrete set-up, as shown in the Appendix of Paper C. When replacing $f$ by $f_0$ in the expression of $G$, the estimator becomes

$$\hat{Q}_{nm} = \int_{[0,1]^2} f_0(u,v)\,\mathrm{d}u\,\mathrm{d}v \frac{1}{nm} \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} \frac{f(G^{-1}((U_1 + i)\frac{1}{n}, (U_2 + j)\frac{1}{m}))}{f_0(G^{-1}((U_1 + i)\frac{1}{n}, (U_2 + j)\frac{1}{m}))}. \tag{1.34}$$

A 2D non-uniform systematic set of sampling points is illustrated in Figure 1.6, where the uniform systematic set of points has been transformed according to the colour values by the formulae of $G$ in (1.32).

Using the proposed formulae, the transformed points are aligned in one direction. If one wants to avoid this, it is possible to introduce a composition $G = \Psi \circ \Phi$ of two transformations - see further details in Paper C. This results in a completely deformed grid of points, but may be difficult to work with in practice as inversion of $G$ may be difficult.
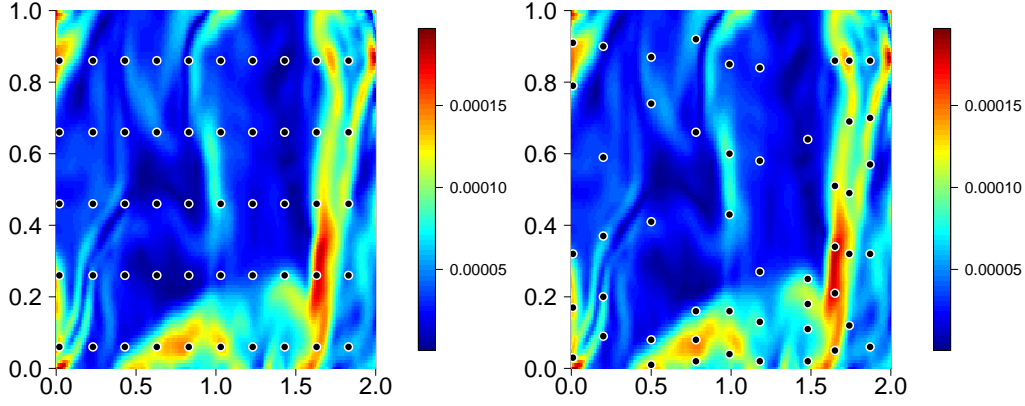
**Figure 1.6:** Left: 2D uniform systematic sampling. Right: 2D non-uniform systematic sampling where the points have been transformed according to the colour values. (*Paper C, Figure 1*)

### 1.5.2  Simulation study

A simulation study imitating sampling situations in microscopy was carried out, in order to compare the efficiency of the new design with that of existing designs. The simulations included number estimation, area estimation and general integral estimations. Below, we shall concentrate mainly on number estimation to explain how the simulations were carried out. Several sampling designs are compared: Non-uniform and uniform designs, systematic and non-systematic designs, continuous and discrete designs. For details, see Paper C. We compared

- 2D (non)-uniform systematic sampling (continuous, proposed in Paper C),

- 2D (non)-uniform systematic sampling (discrete, which corresponds to standard SURS in the uniform case and is proposed in the Appendix of Paper C for the non-uniform case),

- proportional-to-size sampling with replacement (PPS WR, discrete),

- simple random sampling without replacement (SRS WOR, discrete).

We model a cell population by a realization of a point process $Z = \{z_1, z_2, \ldots\}$ in a bounded region, for which we want to estimate the total number of cells $N_Z$. The associated measurement function $f(x, y)$ is the total number of points from the realized point process counted in the quadratic FOV $(x, y) + [0, w]^2$, normalized by the size of the FOV,

$$f(x, y) = \sum_{i=1}^{N_Z} \frac{1_{(x,y)+[0,w]^2}(z_i)}{w^2}.$$

(1.35)

With this definition of $f$, (1.29) is equal to $N_Z$. A realization of a point process and the corresponding $f$ is illustrated to the left in Figure 1.7 (the colours indicate values
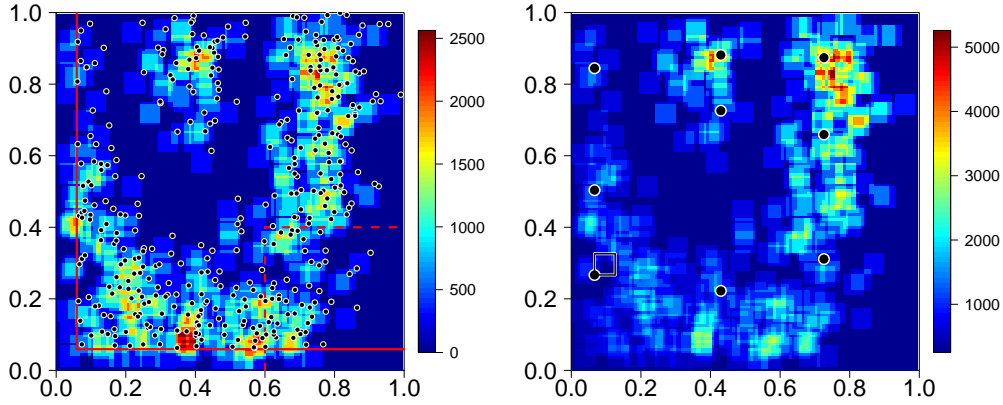
**Figure 1.7:** Left: The measurement function $f$ on $[0,1]^2$, for the realization of an point process illustrated by points (black dots). The full-drawn red lines indicates the lower and left boundary of $[w,1]^2$, where the point process has been generated, and the red dashed lines indicates the 'invisible' part of for cell population, that is points which are not expressed in $f_0$. Right: One of the simulated auxiliary measurement functions $f_0$ for number estimation ($f$ with some noise), and one realization from the corresponding 2D non-uniform systematic sample (black dots) and a FOV (black square). *(Paper C, Figure 2,8)*

of $f$). The "shadow" of side-length $w$, lower and left of each cell, shows where the cells contribute to the value of $f$.

Prior to sampling, several auxiliary functions $f_0$ similar to $f$ are generated:

$$f_0(x,y) = \sum_{z_i \in [0,1]^2 \setminus B} \alpha_i \frac{1_{(x,y)+[0,w]}(z_i)}{w^2} + c, \tag{1.36}$$

with different choices of spatial error $\alpha_i$ (may depend on the $(x,y)$ coordinate), $B$ (sub-set of $[0,1]^2$, where the cells are "invisible", i.e. do not contribute to $f_0$) and a constant $c$. An example of such function $f_0$ is seen to the right in Figure 1.7, where the black points are the set of sampling points and the small black square illustrates the corresponding FOV in one point.

### 1.5.3 Results

In the simulations, resembling the microscopy set-up, we found that the 2D non-uniform systematic designs (continuous and discrete) were more efficient than the various uniform designs and the more simple PPS WR design in a number of cases for area estimation. For number estimation, the new 2D non-uniform designs only had similar efficiency as PPS WR.

### 1.5.4 Discussion and perspectives

In Paper C, we have derived an optimal 2D non-uniform sampling design which reduces the variance substantially compared to the uniform case when the auxiliary function used in the design has a close connection to the measurement function.

Actual comparisons of efficiency have in Paper C been based on simulated data. Future studies should include analysis of real microscopy data. Also it seems worthwhile to compare our approach with the one presented in Grafström and Tillé (2013). Here, the distance between the sampling units are used in their sample algorithm to obtain a spatially balanced sample. Another idea is simply to divide the area into strata and sample independently in each stratum.

From a theoretical point of view, it could be interesting to investigate whether it is possible to construct a class of sampling situations for which 2D non-uniform systematic sampling is more efficient than independent 2D non-uniform sampling. Some considerations concerning this question may be found in Section 5 of Paper C.

## Bibliography

Baddeley, A. and E. B. V. Jensen (2005). *Stereology for statisticians*. Chapman and Hall/CRC.

Bellhouse, D. R. and J. Rao (1975). Systematic sampling in the presence of a trend. *Biometrika 62*(3), 694–697.

Boyce, R. W., K.-A. Dorph-Petersen, L. Lyck, and H. J. G. Gundersen (2010). Design-based stereology introduction to basic concepts and practical approaches for estimation of cell number. *Toxicologic pathology 38*(7), 1011–1025.

Dorph-Petersen, K.-A., H. J. G. Gundersen, and E. B. V. Jensen (2000). Non-uniform systematic sampling in stereology. *Journal of Microscopy 200*(2), 148–157.

Evans, S. M., A. M. Janson, and J. R. Nyengaard (2004). *Quantitative methods in neuroscience: A neuroanatomical approach*. Oxford University Press, USA.

Gardi, J. E., J. R. Nyengaard, and H. J. G. Gundersen (2006). Using biased image analysis for improving unbiased stereological number estimation–a pilot simulation study of the smooth fractionator. *Journal of Microscopy 222*(3), 242–250.

Gardi, J. E., J. R. Nyengaard, and H. J. G. Gundersen (2008a). Automatic sampling for unbiased and efficient stereological estimation using the proportionator in biological studies. *Journal of Microscopy 230*(1), 108–120.

Gardi, J. E., J. R. Nyengaard, and H. J. G. Gundersen (2008b). The proportionator: unbiased stereological estimation using biased automatic image analysis and non-uniform probability proportional to size sampling. *Computers in Biology and Medicine 38*(3), 313–328.

Grafström, A. and Y. Tillé (2013). Doubly balanced spatial sampling with spreading and restitution of auxiliary totals. *Environmetrics 24*(2), 120–131.

Gundersen, H. J. G. (2002). The smooth fractionator. *Journal of Microscopy 207*(3), 191–210.

Gundersen, H. J. G., E. B. V. Jensen, K. Kiêu, and J. Nielsen (1999). The efficiency of systematic sampling in stereology-reconsidered. *Journal of Microscopy 193*(3), 199–211.

Hansen, L. V., M. Kiderlen, and E. B. V. Jensen (2011). Image-based empirical importance sampling: An efficient way of estimating intensities. *Scandinavian Journal of Statistics 38*(3), 393–408.

Isaki, C. T. and W. A. Fuller (1982). Survey design under the regression superpopulation model. *Journal of the American Statistical Association 77*(377), 89–96.

Murthy, M. N. et al. (1967). *Sampling Theory and Methods.* Calcutta-35: Statistical Publishing Society, 204/1, Barrackpore Trunk Road, India.

Nedyalkova, D. and Y. Tillé (2008). Optimal sampling and estimation strategies under the linear model. *Biometrika 95*(3), 521–537.

Särndal, C. E., B. Swensson, and J. H. Wretman (1992). *Model assisted survey sampling.* Springer Verlag.

Singh, D., K. Jindal, and J. Garg (1968). On modified systematic sampling. *Biometrika 55*(3), 541–546.

# Contributions to point processes in microscopy

## 2.1   Introduction

The fully developed phase of several myeloproliferative disorders (i.e. hematopoietic stemcell disorders), described in Paper D, shows a wide and well-differentiated spectrum of clinical findings. By contrast, similar and sometimes overlapping clinical pictures can be observed in the early phases of these disorders. In these cases, bone marrow histopathology can be necessary for discriminating different disorders. The diagnostic criteria for these diseases by the World Health Organization (WHO) classification (Vardiman et al., 2009) includes the morphology of megakaryocytes, however the rationale and reproducibility of these diagnostic guidelines have been questioned. In Madelung et al. (2013) low consensus among pathologists concerning megakaryocyte morphological parameters such as the degree of clustering was found.

Figure 2.1 illustrates a bone marrow section at different magnification, where some megakaryocytes have been marked. Pathologists determine the level of clustering (i.e. no, "loose" and "dense" clustering) by a visual judgement following often vaguely described rules. In fact, a review of the literature revealed that the definitions were not unified and in several ways insufficient, as e.g. the cell intensities influenced several of the definitions, see e.g. Florena et al. (2004); Wilkins et al. (2008); Koopmans et al. (2011); Madelung et al. (2013); Vytrva et al. (2014). In order to use the degree of clustering of megakaryocytes as a classification marker, more precise and reproducible measures must therefore be found.

Proper statistical inference can be achieved viewing the centres of the cells as a realization of a point process. Parametric point process models often allow for simple biological interpretations of the parameters. In such cases, scientific hypotheses regarding the parameters can be used to e.g. test for significant differences in the degree of clustering between groups.

We propose to use *Matérn thinned Cox processes* as a natural, simple, and attractable class of point process models for centres of non-overlapping spherical cells in clusters. Dependent Matérn thinning is applied to Cox processes to achieve a hard core (minimum interpoint) distance between points in clustered point patterns. Thereby, the point pattern at short distances exhibits a repulsive behaviour, while at
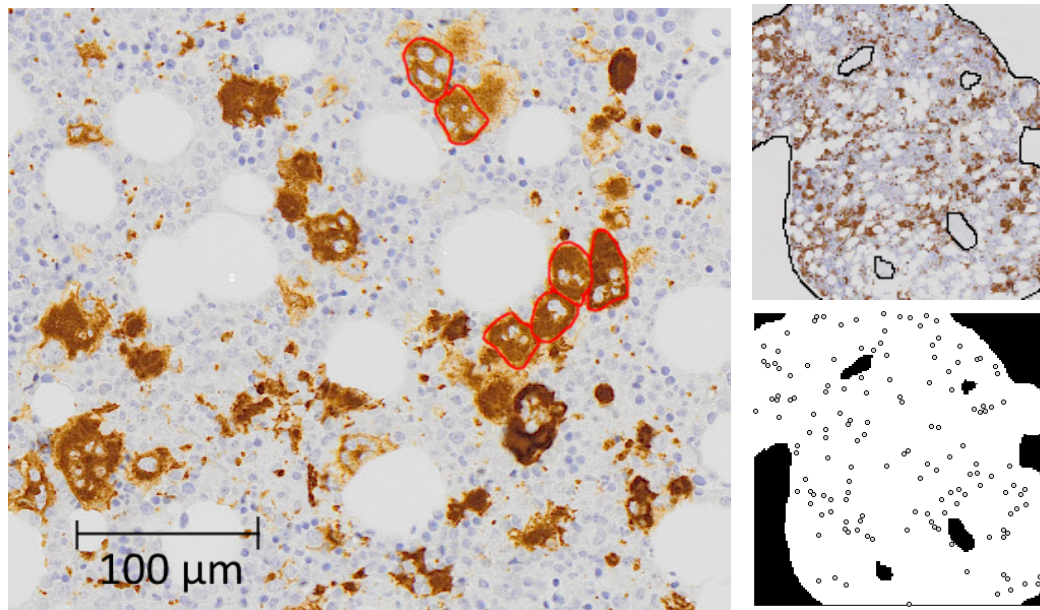
**Figure 2.1:** Left: A small part of a bone marrow biopsy section at high magnification, where a few megakaryocytes have been outlined with red. Right: A biopsy sub-section and the corresponding binary image, where the boarder of the tissue and midpoints of the megakaryocytes have been marked. (*Paper D*, *Figure 4 and 5*)

the same time at middle to large range distances exhibits a clustering behaviour. In Paper D, this new model class was proposed and theoretical properties were derived. The applicability in microscopy was illustrated by the megakaryocyte example. Furthermore, a simulation study was carried out to study parametric inference in the model. The work contributes not only to the particular study of megakaryocytes important to pathologists, but also to the point process literature in general.

Below, we start with a short background description of the relevant point process theory in Section 2.2, followed by a summary of the results obtained in Paper D, see Section 2.3. Additionally, limit results and further results from the simulation study are given in Supplement D.

## 2.2 Spatial point processes

Spatial point processes are used to model point patterns typically formed by positions of objects in 2D or 3D. The applications are numerous. The literature on spatial point processes spans from highly theoretical papers to genuine applications. Classical references include Diggle (2003); Møller and Waagepetersen (2004); Chiu et al. (2013). Modern technology increases the amount of applications and details for spatial data, including informations on marks, such as covariates and shape, size and type of the objects. See also Illian et al. (2008) with emphasis on statistical methods for applications. An overview of modern statistics for spatial point processes may be found in Møller and Waagepetersen (2007).

We consider mainly spatial point processes $X$ on $\mathbb{R}^d$, i.e. random countable subsets with realizations in $N_{\mathrm{lf}}$, the space of locally finite subsets of $\mathbb{R}^d$.

Let $X$ be a spatial point process on $\mathbb{R}^d$ with well-defined *intensity function* $\rho(\cdot)$ and *second-order product density* $\rho^{(2)}(\cdot, \cdot)$. The intensity measure $\alpha$ and the second factorial moment measure $\alpha^{(2)}$ are then given by

$$\alpha(B) = \mathbb{E}\left[\sum_{\xi \in X} \mathbb{1}(\xi \in B)\right] = \int_B \rho(\xi)\, \mathrm{d}\xi \qquad (2.1)$$

and

$$\alpha^{(2)}(B_1 \times B_2) = \mathbb{E}\left[\sum_{\xi, \eta \in X}^{\neq} \mathbb{1}(\xi \in B_1, \eta \in B_2)\right] = \int_{B_1} \int_{B_2} \rho^{(2)}(\xi, \eta)\, \mathrm{d}\xi\, \mathrm{d}\eta, \qquad (2.2)$$

for all $B$, $B_1$ and $B_2$ in the Borel $\sigma$-algebra on $\mathbb{R}^d$. Here $\sum^{\neq}$ denotes summation over distinct pairs. These measures describe first- and second-order properties of $X$. Heuristically, $\rho(\xi)\,\mathrm{d}\xi$ is the probability of the occurrence of a point from $X$ in an infinitesimal region around $\xi$. Correspondingly, for $\xi \neq \eta$, $\rho^{(2)}(\xi, \eta)\,\mathrm{d}\xi\,\mathrm{d}\eta$ is the probability of the occurrence of points in infinitesimal regions around $\xi$ and $\eta$. The value of a constant intensity function is called the *intensity* of the point process.

The interaction between pairs of points can be described by second-order summary statistics of which we will discuss (*Ripley's*) *K-function* and the *pair correlation function* (pcf)

$$g(\xi, \eta) = \rho^{(2)}(\xi, \eta)/(\rho(\xi)\rho(\eta)) \qquad (2.3)$$

(Møller and Waagepetersen, 2004, Chapter 4). We will mainly consider *stationary* and *isotropic* point processes, but some of the definitions can be generalized to so-called *second-order intensity reweighted stationary* processes which are explained and considered in Møller and Waagepetersen (2004, Section 4.1).

For the *reduced Palm distribution* in a point $\xi$, we will use the notation $\mathbb{P}^!_\xi$. This probability distribution can be interpreted as the conditional distribution of $X \setminus \xi$ given $\xi \in X$. The expectation with respect to $\mathbb{P}^!_\xi$ will be denoted $\mathbb{E}^!_\xi$. In the stationary case, $\mathbb{P}^!_o = \mathbb{P}^!_\xi$ for all $\xi$, where $o$ denotes origo, and this point can therefore be considered as a "typical" point in the process, see Møller and Waagepetersen (2004, Appendix C.2) and Illian et al. (2008, Section 4.1). Second- and higher-order Palm distributions can be defined analogously, and we let $\mathbb{P}^!_{\xi, \eta}$ denote the two-point reduced Palm distribution for $X$ in $(\xi, \eta)$ (Hanisch, 1982, p. 172).

In the stationary and isotropic case, the *K*-function is defined by

$$K(r) = \mathbb{E}^!_o\left[\sum_{\xi \in X} \mathbb{1}(\xi \in b(o, r))\right]\Big/\rho, \qquad (2.4)$$

which is the expected number of further points within distance $r$ from a "typical" point in $X$ divided by the intensity $\rho$. In this case, the pcf takes the form

$$g(r) = \rho^{(2)}(r)/\rho^2. \qquad (2.5)$$

Due to the relation

$$K(r) = \int_{b(o,r)} g(t)\,\mathrm{d}t = d|b(o,1)| \int_0^r t^{d-1} g(t)\,\mathrm{d}t, \tag{2.6}$$

$g(r)$ can be interpreted as the expected number of points at distance $r$ from a "typical" point in $X$, relative to the expected number for a Poisson process (the squared intensity).

The $K$-function and the pcf $g$ contain the same statistical information. Since the $K$-function is a cumulative function, it does not have as simple an interpretation as $g$. However, it has obtained its popularity due to the fact that the function is much easier to estimate. It is recommended e.g. in Illian et al. (2008) to use $g$ as the most informative second-order summary statistic. For the class of models considered in Paper D, an approximative expression for $g$ has been derived and this expression is used for parameter estimation. As shown in Supplement D, it is of great importance to find good estimators for $g$.

The existing point process literature provides models for a variety of interactions between points. In applications, models that allow for simple statistical inference are in favour. Often, the models describe patterns that exhibit one of the following three characteristics for the interaction between the points: Complete spatial randomness/no-interaction (*Poisson processes*), clustering/aggregation (e.g. *Cox processes* (Cox, 1955)) or repulsiveness/regularity (e.g. *Gibbs point processes* (Chiu et al., 2013), *Matérns hard core processes* (Matérn, 1986) or other hard core processes (Illian et al., 2008)).

A Cox process $X$ is a spatial point process with a non-negative driving (random) field $\Lambda$, such that conditionally on $\Lambda$, the process is a Poisson process with intensity function $\Lambda$. Simple expressions of e.g. intensity, second-order product density and void probabilities, follow directly from the conditional Poisson behaviour. We shall focus on *shot noise Cox processes* (SNCPs) (Møller, 2003), which can be regarded as Poisson cluster processes. Conditional on cluster locations and mean number of points in the clusters, the process has the same distribution as the superposition of independent Poisson processes. Each Poisson process generates a cluster with intensity functions proportional to a chosen kernel $k$, which may depend on the cluster location. *Neyman-Scott processes* (NSPs) (Neyman and Scott, 1958) may be regarded as a particular case of SNCPs with a constant mean number of points per cluster. More specifically, the driving field of a NSP is of the form

$$\Lambda(\xi) = \sum_{c \in C} \mu k(\xi - c), \tag{2.7}$$

where the process of cluster centres $C$ is a stationary Poisson point process with intensity $\kappa > 0$ (the cluster intensity) and $\mu$ is the mean number of points in a cluster. Finally, $k$ is a kernel function integrating to 1.

Two simple and popular stationary and isotropic NSPs are the *Matérn cluster process* (MCP) (Matérn, 1960, 1986) and the *Thomas process* (TP) (Thomas, 1949),

defined by kernels

$$k(\xi) = \mathbb{1}(\|\xi\| \leq R)/|b(o, R)| \quad \text{and} \quad k(\xi) = \exp(-\|\xi\|^2/2\sigma^2)/(2\pi\sigma^2)^{d/2}, \qquad (2.8)$$

for $R > 0$ and $\sigma > 0$, respectively, (the cluster "radius").

In practical applications, the above classification (i.e. no-interaction, clustering or repulsiveness) may be too simplistic. Patterns of centres of cells in biological material often show a repulsive behaviour at short distances due to cell size, while at the same time at larger range distances exhibit a clustering behaviour. A popular class of models for such behaviour is Gibbs point processes with an appropriate interaction function, as e.g. in Mattfeldt et al. (2006, 2007). Statistical inference is unfortunately based on simulations that require elaborate Markov chain Monte Carlo methods (Møller and Waagepetersen, 2004) which is less appealing in some applications. The interest for finding alternative models to Gibbs point processes, to achieve such dual behaviour is confirmed by recent work by Lavancier and Møller (2015) who extend previous work on interrupted point processes by Stoyan (1979). The idea is to apply thinning according to a random field on a regular point process. In Paper D, discussed in the next section, we take a reverse strategy and obtain the desired interaction by Matérn thinning of Cox processes.

## 2.3 Matérn thinned Cox processes (Paper D)

A class of mathematically tractable point process models that combine a clustering and a hard core behaviour, can be obtained by applying dependent Matérn type II thinning to a clustered Cox process. In recent years, generalizations of Matérn's hard core models have appeared in the literature, e.g. in Månsson and Rudemo (2002); Kiderlen and Hörig (2013); Teichmann et al. (2013), however, non of these models are appropriate for modelling considerable clustering behaviour as they are confined to thinning of a homogeneous Poisson point process. Very little seems to be known about thinning for other models.

In Paper D, we consider general processes subject to the Matérn type II thinning rule and introduce a general framework of Palm retention probabilities for calculating first- and second-order densities. In particular, we consider *Matérn thinned Cox processes* which we define as a point process with parameters $(\Lambda, h)$ given by

$$\text{MTCP}(\Lambda, h) := \{\xi \in X \mid \forall (\eta, m_\eta) \in X_M, \eta \neq \xi : \|\xi - \eta\| > h \vee m_\eta > m_\xi\}, \qquad (2.9)$$

where $X$ is a Cox process with driving field $\Lambda$ and $X_M = \{(\xi, m_\xi) \mid \xi \in X, m_\xi \sim \text{Unif}[0, 1]\}$ is the corresponding marked process with i.i.d. uniform marks. Sample realizations of Matérn thinned Cox processes with kernels as defined in (2.8), i.e. Matérn thinned MCPs (MTMCPs) and Matérn thinned TPs (MTTPs), are shown in Figure 2.2. These two types of models are considered in detail in Paper D.
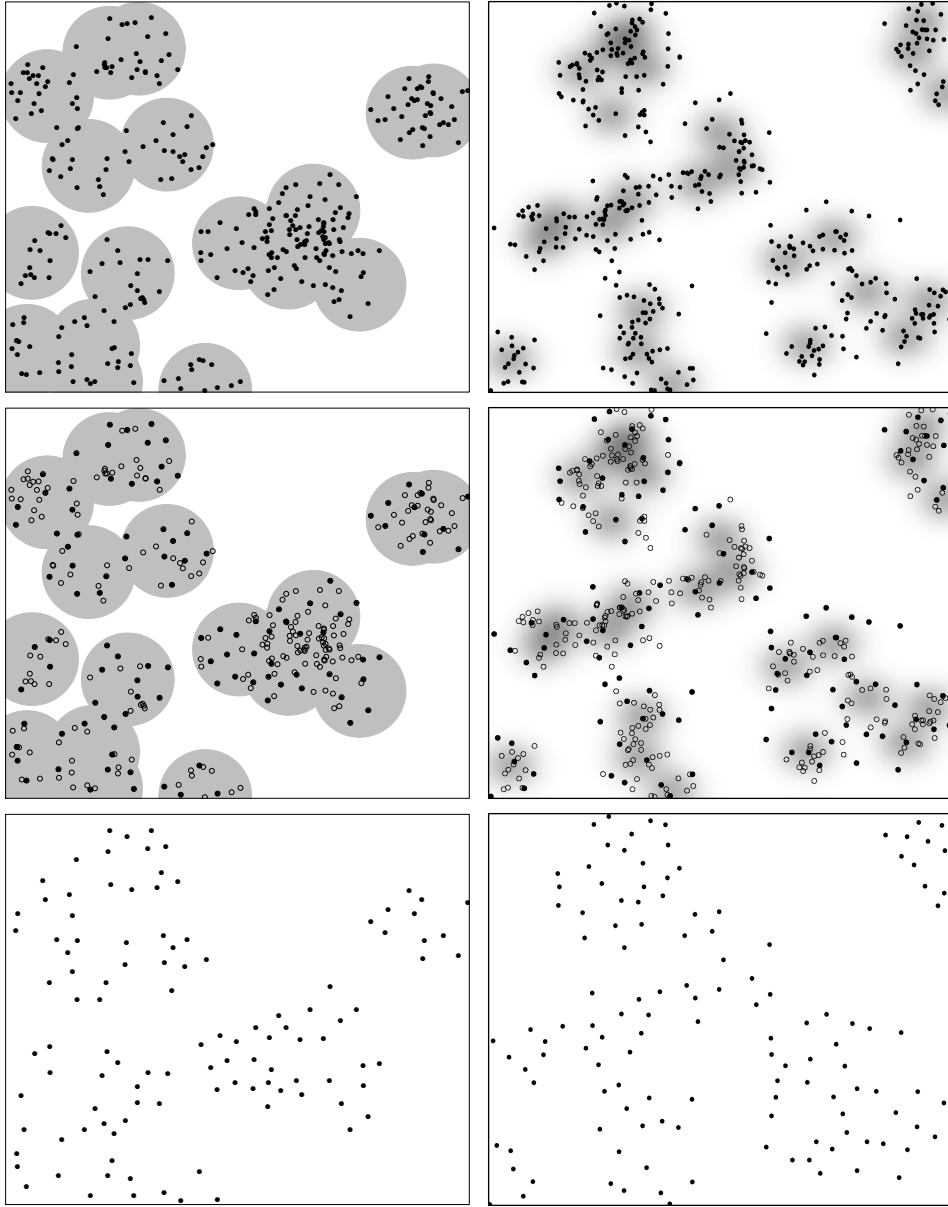
**Figure 2.2:** Realizations of MTMCPs (first column) and MTTPs (second column) with $\kappa = 25$, $R = 0.10$ or $\sigma = 0.04$, $h = R/3$ and $\mu = 20$. First row: An (un-thinned) MCP or TP $X$ is generated in the following way: generate a realization from a stationary planar Poisson point process $C$ of cluster locations with intensity $\kappa > 0$. Subsequently, for each location point $c \in C$, generate cluster points $X_c$ from an (inhomogeneous) Poisson point process with intensity function $\mu k(\cdot - c)$ (illustrated by grey scaled values), where $\mu > 0$ and $k(\cdot)$ is the kernel of the MCP or TP. All the clusters are generated from independent distributions and independently of the location process $C$. Letting $X$ be the superposition of all the cluster points results then in a realization of a MCP or TP. Second and third row: The thinned process (black points) is now obtained be generating i.i.d. marks $\{m_\xi : \xi \in X\}$ uniformly on $[0, 1)$, and letting pairs of points from $X$ with distance smaller than the hard core distance $h > 0$ compete to survive according to their marks. If a point does not survives one or more pair-competitions, it is deleted (open circles). The resulting thinned process is illustrated in the third row.

### 2.3.1 First- and second-order properties using Palm retention probabilities

The intensity function and second-order product density of the thinned process $\mathrm{MTCP}(\Lambda, h)$ can be obtained by multiplying the corresponding functions for the original process with the so-called retention probabilities,

$$\rho_{\mathrm{th}}(\xi) = p_{\mathrm{ret}}(\xi)\rho(\xi) \quad \text{and} \quad \rho_{\mathrm{th}}^{(2)}(\xi, \eta) = p_{\mathrm{ret}}^{(2)}(\xi, \eta)\rho^{(2)}(\xi, \eta). \tag{2.10}$$

Properties like stationarity and isotropy are inherited from the original process such that

$$\rho_{\mathrm{th}} = p_{\mathrm{ret}}\rho \quad \text{and} \quad \rho_{\mathrm{th}}^{(2)}(r) = p_{\mathrm{ret}}^{(2)}(r)\rho^{(2)}(r). \tag{2.11}$$

The ratio $p_{\mathrm{ret}}(\xi)$ between the intensity functions $\rho_{\mathrm{th}}(\xi)$ and $\rho(\xi)$ is also called the (first-order) *Palm retention probability* at the point $\xi$, since it can be expressed in terms of the reduced Palm distribution $\mathbb{P}^!_{(\xi, m_\xi)}$ for the marked process $X_M$. In fact,

$$p_{\mathrm{ret}}(\xi) = \int_0^1 \mathbb{P}^!_{(\xi, m_\xi)}(F(\xi, m_\xi; h))\, \mathrm{d}m_\xi, \tag{2.12}$$

where $F(\xi, m_\xi; h)$ is the set of marked point patterns $x_M$, for which the point $\xi$ with mark $m_\xi$ is retained in the thinned process. The second-order Palm retention probabilities can be found analogously using the two-point reduced Palm distribution $\mathbb{P}^!_{(\xi, m_\xi)(\eta, m_\eta)}$ for $X_M$.

One of the main results in Paper D (Theorem 2) is an exact formula for the first-order Palm retention probabilities of a Matérn thinned shot noise Cox process. In the special case of MTMCPs and MTTPs, the formula reduces nicely and can be determined by simple numerical calculations. For general driving fields $\Lambda$, the intensity function and second-order product density of $\mathrm{MTCP}(\Lambda, h)$ can alternatively be expressed using conditional retention probabilities *given* the driving field $\Lambda$. In Theorem 4 in Paper D we find under mild assumptions for $\Lambda$,

$$\rho_{\mathrm{th}}(\xi) = \mathbb{E}\big[p_{\mathrm{ret}|\Lambda}(\xi)\Lambda(\xi)\big] \quad \text{and} \quad \rho_{\mathrm{th}}^{(2)}(\xi, \eta) = \mathbb{E}\big[p_{\mathrm{ret}|\Lambda}^{(2)}(\xi, \eta)\Lambda(\xi)\Lambda(\eta)\big] \tag{2.13}$$

for $\|\xi - \eta\| > h$, otherwise 0, where for $\Lambda(\xi)$ and $\Lambda(\eta) > 0$,

$$p_{\mathrm{ret}|\Lambda}(\xi) = \frac{1 - \exp(-\Omega_\xi)}{\Omega_\xi} \tag{2.14}$$

and

$$p_{\mathrm{ret}|\Lambda}^{(2)}(\xi, \eta) = \frac{1 - \exp(-\Omega_\xi)}{\Omega_\xi \Omega_{\eta \setminus \xi}} + \frac{1 - \exp(-\Omega_\eta)}{\Omega_\eta \Omega_{\xi \setminus \eta}} - \frac{1 - \exp(-\Omega_{\xi \cup \eta})}{\Omega_{\xi \cup \eta}}\left(\frac{1}{\Omega_{\xi \setminus \eta}} + \frac{1}{\Omega_{\eta \setminus \xi}}\right), \tag{2.15}$$

with $\Omega_* = \int_{b_*} \Lambda(\tau)\, \mathrm{d}\tau$, $b_\xi = b(\xi, h)$, $b_{\xi \setminus \eta} = b(\xi, h) \setminus b(\eta, h)$, and $b_{\xi \cup \eta} = b(\xi, h) \cup b(\eta, h)$.

### 2.3.2 Approximations

The expressions for $\rho_{th}(\xi)$ and $\rho_{th}^{(2)}(\xi, \eta)$ are in general computational infeasible, and therefore we propose approximations which makes inference possible at least for Matérn thinned short noise Cox processes. Using

$$\Omega_\xi = \int_{b(\xi,h)} \Lambda(\vartheta) d\vartheta \approx \Lambda(\xi)|b(\xi,h)| \tag{2.16}$$

and similar approximations, simple expressions of $\rho_{th}(\xi)$ and $\rho_{th}^{(2)}(\xi, \eta)$ can be obtained, see Theorem 5 in Paper D. For stationary and isotropic Matérn thinned Neyman-Scott processes we find approximative expressions for the intensity and pair correlation function (pcf) given by

$$\rho_a = \frac{1 - \exp(-a)}{\tau_h}, \tag{2.17}$$

and

$$g_a(r) = \frac{\rho_a^{(2)}(r)}{\rho_a^2} = \frac{2\Gamma_h(r)\tau_h(1 - \exp(-a)) - 2\tau_h^2(1 - \exp(-b(r)))}{\Gamma_h(r)(\Gamma_h(r) - \tau_h)(1 - \exp(-a))^2}, \quad r > h, \tag{2.18}$$

where $\tau_h = |b(o,h)|$, $\Gamma_h(r) = |b(o,h) \cup b(\xi_r,h)|$, $\xi_r \in \mathbb{R}^d$, with $\|\xi_r\| = r$, and where $a$ and $b(r)$ depend on the kernel and parameters in the model through integration. For MTTPs and in particular MTMCPs, $a$ and $b(r)$ take simple forms. For MTMCPs, the formulas do not involve numerical integration.

### 2.3.3 Discussion and perspectives

One way of examining the quality of the approximations suggested in Section 2.3.2, is to compare the approximated functions with the theoretical functions in limiting cases of a scaling parameter of the intensity. Due to space limitations, such limit results were not included in Paper D, but are presented in Supplement D. A different approach is by evaluating the approximations by simulations. This was done in Paper D for MTMCPs and MTTPs, where it was found that the approximations capture most of the true behaviour of the models. Hence, the approximations are likely to be sufficiently accurate to enable simple model fitting in applications, using the minimum contrast method. Supplement D elaborates on the method used for parameter estimation, where we have found that parameter estimation for MTTPs often is as simple and efficient as for the simpler standard MCPs and TPs. An exception is the estimation of $\mu$ (the mean number of points per cluster before thinning) in highly regular MTTPs which is also discussed in Supplement D.

In the application in Paper D, point patterns of megakaryocytes in a control group and three disease groups were compared. MTMCPs and MTTPs were successfully fitted to the data. The fitted models seem to capture most of the behaviour of the average group pcfs from the observed point patterns. Thus, the new models can be used in a valid cluster detection procedure for megakaryocytes which otherwise has been missing in the literature within the field of pathology. The MTTPs

result in a slightly better fit. The fitted MTMCPs and MTTPs have slightly higher degree of clustering for short and large range distances and slightly lower degree of clustering for medium range distances compared to the observed point patterns. A future larger study must be performed to establish more detailed conclusions for the groups under considerations.

# Bibliography

Chiu, S. N., D. Stoyan, W. S. Kendall, and J. Mecke (2013). *Stochastic geometry and its applications*. John Wiley & Sons.

Cox, D. R. (1955). Some statistical methods connected with series of events. *Journal of the Royal Statistical Society. B*(17), 129–164.

Diggle, P. J. (2003). *Statistical Analysis of Spatial Point Patterns. Second edition*. Oxford University Press, New York.

Florena, A. M., C. Tripodo, E. Iannitto, R. Porcasi, S. Ingrao, and V. Franco (2004). Value of bone marrow biopsy in the diagnosis of essential thrombocythemia. *Haematologica 89*(8), 911–919.

Hanisch, K.-H. (1982). On inversion formulae for *n*-fold Palm distributions of point processes in LCS-spaces. *Mathematische Nachrichten 106*(1), 171–179.

Illian, J., A. Penttinen, H. Stoyan, and D. Stoyan (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*, Volume 70. Wiley, Chichester.

Kiderlen, M. and M. Hörig (2013). Matérn's hard core models of types I and II with arbitrary compact grains. *CSGB Research Report*. Submitted for journal publication.

Koopmans, S. M., F. J. Bot, K. H. Lam, A. M. Van Marion, H. De Raeve, and K. M. Hebeda (2011). Reproducibility of histologic classification in nonfibrotic myeloproliferative neoplasia. *American journal of clinical pathology 136*(4), 618–624.

Lavancier, F. and J. Møller (2015). Modelling aggregation on the large scale and regularity on the small scale in spatial point pattern datasets. *Available at arXiv:1505.07215*. Submitted for journal publication.

Madelung, A. B., H. Bondo, I. Stamp, P. Loevgreen, S. L. Nielsen, A. Falensteen, H. Knudsen, M. Ehinger, R. Dahl-Sørensen, N. B. Mortensen, et al. (2013). World Health Organization-defined classification of myeloproliferative neoplasms: Morphological reproducibility and clinical correlations — the Danish experience. *American journal of hematology 88*(12), 1012–1016.

Månsson, M. and M. Rudemo (2002). Random patterns of nonoverlapping convex grains. *Advances in Applied Probability 34*(4), 718–738.

Matérn, B. (1960). Spatial variation. Stochastic models and their application to some problems in forest surveys and other sampling investigations. *Medd. Statens Skogsforskningsinstitut 49*(5), 1–144.

Matérn, B. (1986). *Spatial Variation.* Lecture Notes in Statistics 36, Springer-Verlag, Berlin.

Mattfeldt, T., S. Eckel, F. Fleischer, and V. Schmidt (2006). Statistical analysis of reduced pair correlation functions of capillaries in the prostate gland. *Journal of microscopy 223*(2), 107–119.

Mattfeldt, T., S. Eckel, F. Fleischer, and V. Schmidt (2007). Statistical modelling of the geometry of planar sections of prostatic capillaries on the basis of stationary Strauss hard-core processes. *Journal of microscopy 228*(3), 272–281.

Møller, J. (2003). Shot noise Cox processes. *Advances in Applied Probability 35*, 614–640.

Møller, J. and R. P. Waagepetersen (2004). *Statistical Inference and Simulation for Spatial Point Processes.* Chapman and Hall/CRC, Boca Raton, FL.

Møller, J. and R. P. Waagepetersen (2007). Modern statistics for spatial point processes. *Scandinavian Journal of Statistics 34*(4), 643–684.

Neyman, J. and E. L. Scott (1958). Statistical approach to problems of cosmology. *Journal of the Royal Statistical Society. Series B 20*, 1–43.

Stoyan, D. (1979). Interrupted point processes. *Biometrical Journal 21*(7), 607–610.

Teichmann, J., F. Ballani, and K. G. van den Boogaart (2013). Generalizations of Matérn's hard-core point processes. *Spatial Statistics 3*, 33–53.

Thomas, M. (1949). A generalization of Poisson's binomial limit for use in ecology. *Biometrika 36*, 18–25.

Vardiman, J. W., J. Thiele, D. A. Arber, R. D. Brunning, M. J. Borowitz, A. Porwit, N. L. Harris, M. M. Le Beau, E. Hellström-Lindberg, A. Tefferi, et al. (2009). The 2008 revision of the world health organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood 114*(5), 937–951.

Vytrva, N., E. Stacher, P. Regitnig, W. Zinke-Cerwenka, S. Hojas, E. Hubmann, A. Porwit, M. Bjorkholm, G. Hoefler, and C. Beham-Schmid (2014). Megakaryocytic morphology and clinical parameters in essential thrombocythemia, polycythemia vera, and primary myelofibrosis with and without jak2 v617f. *Archives of pathology & laboratory medicine 138*(9), 1203–1209.

Wilkins, B. S., W. N. Erber, D. Bareford, G. Buck, K. Wheatley, C. L. East, B. Paul, C. N. Harrison, A. R. Green, and P. J. Campbell (2008). Bone marrow pathology in essential thrombocythemia: interobserver reliability and utility for identifying disease subtypes. *Blood 111*(1), 60–70.

# Part II

# Papers

# Paper A

# Optimal PPS sampling with vanishing auxiliary variables – with applications in microscopy

By I. T. Andersen, U. Hahn & E. B. V. Jensen

# Optimal PPS Sampling with Vanishing Auxiliary Variables – with Applications in Microscopy

INA TROLLE ANDERSEN

*Department of Mathematics, Aarhus University*
*Stereological Research Laboratory, Aarhus University*

UTE HAHN AND EVA B. VEDEL JENSEN

*Department of Mathematics, Aarhus University*

ABSTRACT. Recently, non-uniform sampling has been suggested in microscopy to increase efficiency. More precisely, proportional to size (PPS) sampling has been introduced, where the probability of sampling a unit in the population is proportional to the value of an auxiliary variable. In the microscopy application, the sampling units are fields of view, and the auxiliary variables are easily observed approximations to the variables of interest. Unfortunately, often some auxiliary variables vanish, that is, are zero-valued. Consequently, part of the population is inaccessible in PPS sampling. We propose a modification of the design based on a stratification idea, for which an optimal solution can be found, using a model-assisted approach. The new optimal design also applies to the case where 'vanish' refers to missing auxiliary variables and has independent interest in sampling theory. We verify robustness of the new approach by numerical results, and we use real data to illustrate the applicability.

*Key words:* Horvitz–Thompson estimator, microscopy, model-assisted sampling, optimal allocation, proportional regression models, systematic PPS sampling, vanishing auxiliary variables

## 1. Introduction

Non-uniform sampling has considerable practical interest in microscopy, as the structures under study often show pronounced inhomogeneity. In microscopy, one aims at estimating the total number of cells (or some other feature of interest) in a section, by evaluating only a few fields of view (FOV) at high magnification. The FOV are thus the sampling units. When using uniform sampling on an inhomogeneous microscopical section, most of the sampled FOV will contain no or only little information of the feature of interest, and, as a consequence, the sampling becomes highly inefficient. The idea is therefore to use a 'cheap' auxiliary variable that is a crude approximation to the 'expensive' variable of interest to determine sampling weights. In microscopy, the auxiliary variables can be obtained by automatic computerized analysis of low magnification images, therewith detecting areas where a high content of the feature of interest is expected. Combining this information with non-uniform sampling may then lead to a considerable reduction in estimator variance compared with the traditional systematic uniform sampling (Gardi *et al.*, 2008a, 2008b).

This idea of empirical importance sampling has been given a stochastic formulation in Hansen *et al.* (2011), using point process theory. In the paper by Hansen *et al.* (2011), statistical tools are developed for assessing the efficiency and constructing optimal model-based estimators of intensities in the class of generalized proportional regression models. These estimators can be used in practice, but several problems arise, which motivates further research.

One of the problems is that if the proportionality assumption is not met, the model-based estimator may be biased, which is unacceptable for the majority of researchers working

in microscopy. Therefore, it may be preferable to keep the original design-based Horvitz–Thompson estimator, which preserves unbiasedness regardless of proportionality or not, and focus on modifying the sampling design to improve efficiency of the estimator.

Another important problem which is not addressed in Hansen *et al.* (2011) is vanishing auxiliary variables. We use the term vanishing auxiliary variables to cover *either* the case where the auxiliary variables are zero-valued but may have positive value of the corresponding variables of interest, which often occurs in the microscopy application, *or* the case where the auxiliary variables are simply missing. In a study of Keller *et al.* (2013) where the variables of interest were cell counts and the auxiliary variables were the amount of colour associated with a staining of the cells, 10% of the cells were in fact found in FOV with vanishing (zero-valued) auxiliary variables. Unbiasedness of the Horvitz–Thompson estimator however requires positive inclusion probabilities for all FOV with a positive cell count. Sampling proportional to size (PPS), where 'size' is measured by the auxiliary variables, needs therefore to be modified if vanishing auxiliary variables can occur.

The workaround in microscopy, as suggested in Gardi *et al.* (2008a, 2008b), is to add a small constant $\varepsilon > 0$ to all auxiliary variables before sampling. Current software (Keller *et al.*, 2013) uses by default an unrealistically small constant. In cases such as the one described in Keller *et al.* (2013), the unbiasedness would therefore be paid for by an extremely high variance if the default had been used. This is caused by rare cases, where the sample includes the problematic FOV mentioned previously. On the other hand, large values of $\varepsilon$ may decrease the efficiency one hopes to gain from PPS sampling compared with uniform sampling. Therefore, optimal ways of choosing such $\varepsilon$ are important in practical application. This problem is addressed in the present paper. We consider a sampling design for a finite population of units, numbered $\{1, \dots, N\}$. The sample is a random subset $S \subseteq \{1, \dots, N\}$ of the population with $n$ elements, say. Some of the sampling units $i$ have zero inclusion probabilities $\pi_i = P(i \in S)$, for instance,

$$\pi_1 = \cdots = \pi_{N_0} = 0,$$

where $N_0 < N$. We modify the design, such that the resulting sample still has size $n$, and such that it retains a constant *positive* inclusion probability $\pi_0$ for the units $1, \dots, N_0$ and inclusion probabilities proportional to the original inclusion probabilities for the remaining units. Under mild regularity conditions, we find the optimal design of this type. This result, which is of independent interest in sampling theory, can be used to determine an optimal value of $\varepsilon$ in the original problem described previously.

The composition of the paper is as follows. The sampling set-up is presented in Section 2, while the optimal design is derived in Section 3, where it is also shown, that under a proportional regression model, the optimality result simplifies. This framework, where both design and model play a role, is often referred to as a model-assisted approach (Särndal *et al.*, 1992). Furthermore, Section 3 gives a summary of the robustness study of the optimal design against parameter misspecification and departures from proportionality. An analysis of data from microscopy, using the developed methods, is presented in Section 4. Conclusions and further perspectives are found in Section 5. The Supporting Information contains proofs of theorems and a detailed investigation of the robustness of the optimal design.

## 2. Set-up

We consider a finite population of $N$ units and assume that a realization of a random variable $Y_i$ (the variable of interest) is available for each unit $i, i = 1, \dots, N$. Additionally, we assume that $Y_1, \dots, Y_N$ are uncorrelated. The aim is to estimate the population total, $T = \sum_{i=1}^{N} Y_i$, for a realization of $Y = \{Y_1, \dots, Y_N\}$.

## 2.1. The Horvitz–Thompson estimator

Let $S \subseteq \{1, \ldots, N\}$ be a random sample without replacement of fixed size $n$ and independent of $Y$. An estimator of the population total, well known from survey sampling theory (Särndal *et al.*, 1992, p. 42), is the Horvitz–Thompson estimator

$$\hat{T} = \sum_{i \in S} \frac{Y_i}{\pi_i}, \tag{1}$$

where $\pi_i = P(i \in S)$ is the probability that the $i$th unit is included in the sample. The sampling design is called non-uniform if the inclusion probabilities $\pi_i$ are non-constant. The estimator $\hat{T}$ is design-unbiased, that is,

$$\mathbb{E}(\hat{T}|Y) = T,$$

if the inclusion probabilities $\pi_i$ are all positive. Under the same assumption, the design variance takes the form

$$\text{Var}(\hat{T}|Y) = \sum_{i=1}^{N} \sum_{j=i+1}^{N} (\pi_i \pi_j - \pi_{ij}) \left( \frac{Y_i}{\pi_i} - \frac{Y_j}{\pi_j} \right)^2, \tag{2}$$

where $\pi_{ij}$ is the $(i, j)$th joint inclusion probability, $\pi_{ij} = P(i \in S, j \in S)$.

A sampling design is called optimal under a model for $Y$, if it minimizes the mean design variance $\mathbb{E}\text{Var}(\hat{T}|Y)$. Because $\hat{T}$ is design-unbiased, the mean design variance is in fact equal to the mean square error of $\hat{T}$

$$\mathbb{E}\text{Var}(\hat{T}|Y) = \text{Var}\hat{T} - \text{Var}T = \mathbb{E}(\hat{T} - T)^2.$$

If the inclusion probabilities are proportional to the mean values of the $Y_i$s, that is,

$$\pi_i \propto \mathbb{E}Y_i,$$

then we have the following result for the mean design variance

$$\mathbb{E}\text{Var}(\hat{T}|Y) = \sum_{i=1}^{N} \left( \frac{1}{\pi_i} - 1 \right) \text{Var}Y_i, \tag{3}$$

which follows directly from Särndal *et al.* (1992, Result 2.6.2). Note that the mean variance only depends on the inclusion probabilities $\pi_i$ and the variances of the $Y_i$s. More detailed properties of the sampling design such as the joint inclusion probabilities do not appear in the formula.

## 2.2. Systematic proportional to size sampling

In order to obtain an efficient estimator of $T$, the information from an auxiliary variable $x_i$ associated with $Y_i$, $i = 1, \ldots, N$, may be used. We let $x = \{x_1, \ldots, x_N\}$. In the applications we have in mind, both $x$ and $Y$ are random, but $x$ is completely known. Henceforth, modelling and inference about $Y$ will be made in the conditional distribution given $x$. Compared with the previous subsection, the variables of interest $Y_1, \ldots, Y_N$ are now assumed to be uncorrelated in the conditional distribution given $x$, and the sample $S$ and $Y$ are independent given $x$.

Often, the inclusion probability $\pi_i$ is chosen proportional to $x_i$ (sampling proportional to size and PPS sampling), as one expects $x_i$ to be roughly proportional to $Y_i$. It is possible to

construct such a PPS sample $S$ of size $n$ if $nx_i/x_\bullet \leq 1$ for all $i$, where $x_\bullet = \sum_{i=1}^{N} x_i$. Under this common constraint on the sample size $n$ (Särndal *et al.*, 1992, Remark 3.6.2), we have

$$\pi_i = n \frac{x_i}{x_\bullet}.$$

A PPS sampling scheme that is widely used in sampling because of its simplicity and efficiency is the systematic PPS sampling (Särndal *et al.*, 1992, Section 3.6). It can be implemented as follows. Let $1, \ldots, N$ refer to an ordering of the units. Sampling is performed on cumulative weights with a random starting point in $[0, \frac{x_\bullet}{n}]$, followed by equidistant selections of the units. More precisely, let $W_i = \sum_{j=1}^{i} x_j, i = 1, \ldots, N$, denote the cumulated weights with $W_0 = 0$. Let $V_1 \sim \text{unif}([0, \frac{x_\bullet}{n}])$, independent of $Y$, and let $V_j = V_1 + (j-1)\frac{x_\bullet}{n}, j = 2, \ldots, n$. Then, the sample $S$ consists of those units $i$ for which $[W_{i-1}, W_i]$ contains at least one $V_j$. Because of the constraint that $x_i \leq x_\bullet/n$, each interval $[W_{i-1}, W_i]$ can contain at most one $V_j$, and therefore,

$$\sum_{j=1}^{n} \mathbf{1}\{V_j \in [W_{i-1}, W_i]\} \leq 1$$

for all $i$, and the inclusion probabilities take the intended form

$$\pi_i = P\left( \sum_{j=1}^{n} \mathbf{1}\{V_j \in [W_{i-1}, W_i]\} = 1 \right)$$
$$= \mathbb{E}\left( \sum_{j=1}^{n} \mathbf{1}\{V_j \in [W_{i-1}, W_i]\} \right)$$
$$= n \frac{x_i}{x_\bullet}.$$

Figure 1 illustrates systematic PPS sampling with the natural ordering of the units, according to the size of the auxiliary variable $x$.

A specific variant of systematic PPS sampling, called the proportionator, was suggested in Gardi *et al.* (2008a, 2008b) for analysis of microscopy images. The design uses the principles of the so-called smooth fractionator (Gundersen, 2002) to order the sampling units, which corresponds to the balanced systematic sampling described in Murthy (1967, Section 5.9d). If we let $1, \ldots, N$ denote the ordering such that $x_1 \leq x_2 \leq \ldots \leq x_N, M = \frac{N}{2}$ if $N$ is even and $M = \frac{N+1}{2}$ if $N$ is odd, the smooth ordering becomes $[1], \ldots, [N]$, where

$$[i] = \begin{cases} 2i - 1, & i \leq M, \\ 2(N + 1 - i), & i > M. \end{cases} \tag{4}$$
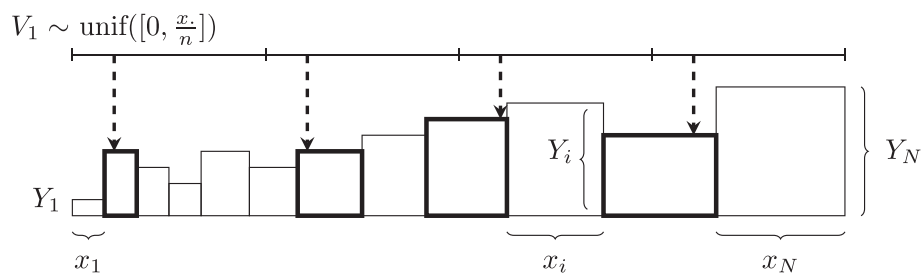


*Fig. 1.* Illustration of systematic proportional to size sampling. Each box represents a sampling unit $i$ with height given by the variable of interest $Y_i$ and width given by the auxiliary variable $x_i$. The sampling units have been ordered according to the sizes of the auxiliary variables, and sampling is then performed on the cumulative weights from a random starting point in $[0, \frac{x_\bullet}{n}]$, followed by equidistant selections of the units.

This alternative ordering with $x_i = i$ has been proven to be superior to the ordinary (according to increasing unit number) ordering in uniform systematic sampling, for example, when linear trend ($\mathbb{E}Y_i = a + bi, i = 1, \dots N$) is present (Bellhouse & Rao, 1975), and the efficiency is illustrated in Gundersen (2002). Similar theoretical results for non-uniform sampling do not exists, to the best of our knowledge, yet Gardi *et al.* (2008a, 2008b) adapted the idea, and this ordering is used in microscopy.

In the sampling literature, there also exists a with-replacement version of PPS sampling. Here, the estimator takes the form (Särndal *et al.*, 1992, p. 97)

$$\hat{T}^{WR} = \frac{1}{n} \sum_{i=1}^{N} \#\{i \in S\} \frac{Y_i}{p_i},$$

where $\#\{i \in S\}$ denotes the number of times unit $i$ is sampled, and the draw-by-draw inclusion probability of unit $i$ is given by $p_i = x_i / x_{\bullet\bullet}$.

### 2.3. ε-corrected proportional to size sampling

In the present paper, we address the problem of vanishing auxiliary variables, that is, *either* $x_i = 0$ for some $i$ *or* $x_i$ is missing for some $i$. In the latter case, we set $x_i = 0$ too. As a consequence, in both cases, there may exist units $i$ with $Y_i > 0$ and $x_i = 0$. In PPS sampling, these units are sampled with probability $\pi_i = 0$, and thus, the Horvitz–Thompson estimator (1) will be biased. To adjust for this, one can add a small constant $\varepsilon > 0$ to the auxiliary variables which are zero. The resulting PPS sampling design will be called $\varepsilon$-corrected.

Let $N_0 = \#\{i | x_i = 0\}$, and suppose that the units are ordered such that

$$x_1 = \cdots = x_{N_0} = 0.$$

Then, the inclusion probabilities of the $\varepsilon$-corrected PPS sampling design with sample size $n$ become

$$\pi_i = \begin{cases} n \frac{\varepsilon}{x_{\bullet\bullet} + N_0 \varepsilon}, & i = 1, \dots, N_0, \\ n \frac{x_i}{x_{\bullet\bullet} + N_0 \varepsilon}, & i = N_0 + 1, \dots, N. \end{cases} \tag{5}$$

It is important that $\varepsilon$ is not chosen too small. When $\varepsilon$ is chosen unrealistically small, like it was done in microscopy until recently, the result is an extremely large variance. In fact, with inclusion probabilities as specified in (5), $\text{Var}(\hat{T}|Y) \to \infty$, when $\varepsilon \to 0$, if $Y_i > 0$ for just one $i \in \{1, \dots, N_0\}$. To see this, note that

$$\text{Var}(\hat{T}|Y) \geq \sum_{i=1}^{N} \frac{1}{\pi_i} Y_i^2 - T^2 \to \infty, \text{ as } \varepsilon \to 0. \tag{6}$$

On the other hand, $\varepsilon$ should not be chosen too large, because then the sampling is directed towards the first $N_0$ units, and units in Stratum 1 with proportionality between $x_i$ and $Y_i$ are not sampled so often.

The $\varepsilon$-corrected systematic PPS sampling can be considered as a kind of stratification, based on the auxiliary variables (Fig. 2). This observation opens up for the possibility of finding an optimal $\varepsilon$, using optimal allocation in stratified sampling (Särndal *et al.*, 1992, Section 3.7).

Stratification is a standard variance reduction technique in sampling, where the population is divided into strata, and independent samples are taken from each stratum. In $\varepsilon$-corrected systematic PPS sampling, we can regard the population as divided into two strata, Stratum 0
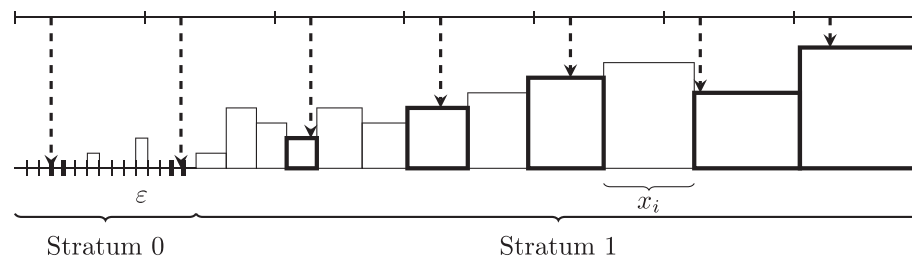
*Fig. 2.* Illustration of $\varepsilon$-corrected systematic proportional to size sampling, where a small constant $\varepsilon$ has been added to each unit with $x_i = 0$ to ensure an unbiased estimator. Due to the systematic sampling, the sampling scheme has a built-in stratification mechanism, such that an almost fixed fraction of the sampled units will be in each stratum.

consisting of the sampling units with $x_i = 0$ and Stratum 1 consisting of the sampling units with $x_i > 0$. If we let $U_0 = \{1, \dots, N_0\}$ and $U_1 = \{N_0 + 1, \dots, N\}$ be the notation used for the two strata, we have for the Horvitz–Thompson estimator based on $\varepsilon$-corrected PPS sampling

$$\hat{T} = \sum_{i \in S} \frac{Y_i}{\pi_i} = \sum_{i \in S \cap U_0} \frac{Y_i}{\pi_i} + \sum_{i \in S \cap U_1} \frac{Y_i}{\pi_i} = \hat{T}_0 + \hat{T}_1, \tag{7}$$

say. Because we have no detailed information about Stratum 0, where all auxiliary variables vanish, it is natural to make the model assumption that the $Y_i$s in Stratum 0 have the same mean. Under this model, systematic PPS sampling leads to zero mean covariance of the two estimator parts $\hat{T}_0$ and $\hat{T}_1$ if the expected number of units sampled from Stratum 0,

$$n_0 = n \frac{N_0 \varepsilon}{x_\cdot + N_0 \varepsilon} \tag{8}$$

is an integer, as stated in theorem 1. This is essentially due to the number of sampled units in each stratum being fixed, if $n_0$ is integer.

**Theorem 1.** *Assume that $Y_i, Y_j, i \neq j$ are uncorrelated, and that $\mathbb{E}Y_i = \mathbb{E}Y_j$ if both $x_i = 0$ and $x_j = 0$. Under $\varepsilon$-corrected systematic PPS sampling with ordering according to the variables of size,*

$$\mathbb{E}\,\mathrm{Cov}\left(\hat{T}_0, \hat{T}_1 \mid Y\right) = -\nu(1 - \nu)\tau_0\tau_1,$$

*where $\lfloor n_0 \rfloor$ is the integer part and $\nu = n_0 - \lfloor n_0 \rfloor$ is the fractional part of $n_0$, $\tau_0 = \mathbb{E}T_0/n_0$ and*

$$\tau_1 = \begin{cases} \mathbb{E}T_1/(n - n_0) \text{ (or some arbitrary constant)}, & \nu = 0 \\ \frac{1}{\nu}\mathbb{E}\left(\mathbb{E}\left(\hat{T}_1 - T_1 \mid S\right) \,\middle|\, \#(S \cap U_0) = \lfloor n_0 \rfloor\right), & \nu > 0. \end{cases}$$

*If we furthermore assume proportionality for units $i$ in Stratum 1, that is $\mathbb{E}Y_i \propto x_i$, then $\tau_1$ simplifies to $\tau_1 = \mathbb{E}T_1/(n - n_0)$ for all values of $\nu$.*

For the proof, refer to the Supporting Information. This built-in stratification effect of systematic sampling motivates to consider alternatively in the following sections, a genuinely stratified design, with fixed sample size $n_0$ in Stratum 0, and where sampling is carried out independently on the two strata.

## 3. An optimal stratified design

### 3.1. The optimality results

Consider a sampling design with fixed sample size $n$ and inclusion probabilities $\tilde{\pi}_i$ such that $\sum_{i=1}^{N} \tilde{\pi}_i = n$. Let us suppose that

$$\tilde{\pi}_i = 0, \quad i = 1, \ldots, N_0,$$

$$\tilde{\pi}_i > 0, \quad i = N_0 + 1, \ldots, N.$$

The model example is the systematic PPS sampling in Section 2.2 with $\tilde{\pi}_i = n x_i / x_{\bullet}$ and $x_1 = \cdots = x_{N_0} = 0$. We want to modify the design such that the first $N_0$ units are assigned a constant positive inclusion probability and the remaining units have inclusion probabilities proportional to the original ones. We will only consider modified designs for which the sample size among the first $N_0$ units is a fixed integer $n_0$. The modified sampling design will have the following inclusion probabilities

$$\pi_i = \begin{cases} \frac{n_0}{N_0}, & i = 1, \ldots, N_0, \\ \left(1 - \frac{n_0}{n}\right) \tilde{\pi}_i, & i = N_0 + 1, \ldots, N. \end{cases} \tag{9}$$

Motivated by the result of theorem 1, we propose to sample independently on the two strata, Stratum 0 $(1, \ldots, N_0)$ and Stratum 1 $(N_0 + 1, \ldots, N)$. Otherwise, the sampling design remains unchanged, up to the rescaled inclusion probabilities. The theorem below gives the optimal stratified design of this type. The proof may be found in the Supporting Information. The result holds under the following model assumptions

$$\mathbb{E}Y_i = \beta_0, \quad \text{Var}Y_i = \sigma_0^2, \quad i = 1, \ldots, N_0, \tag{10}$$

$$\mathbb{E}Y_i \propto \tilde{\pi}_i, \quad \text{Var}Y_i = \sigma_i^2, \quad i = N_0 + 1, \ldots, N, \tag{11}$$

where $\sigma_0^2 > 0$ and $\sigma_i^2 > 0, i = N_0 + 1, \ldots, N$.

**Theorem 2.** *Consider a sampling design with positive inclusion probabilities of the form (9) and suppose that (10) and (11) hold. Then, under stratified sampling, the mean variance $\mathbb{E}Var(\hat{T}|Y)$ of the Horvitz–Thompson estimator $\hat{T} = \hat{T}_0 + \hat{T}_1$, where $\hat{T}_0$ and $\hat{T}_1$ are based on independent samples $S_0$ and $S_1$ in Stratum 0 and Stratum 1, respectively, is minimized if the sample size $n_0$ in Stratum 0 is chosen as one of the integers closest to*

$$\min \left( n \frac{N_0}{N_0 + \sqrt{n \sum_{i=N_0+1}^{N} \sigma_i^2 / (\sigma_0^2 \tilde{\pi}_i)}}, N_0 \right). \tag{12}$$

We will from now on, for ease of exposition, focus on the case, where the first term inside the minimum sign in (12) is an integer $n_0$ smaller than or equal to $N_0$, such that this term becomes the optimal allocation for $n_0$.

In the case of PPS sampling with probabilities according to an auxiliary variable, as described in the previous section, we have $\tilde{\pi}_i = n x_i / x_{\bullet}$ with $x_1 = \cdots = x_{N_0} = 0$. It follows from theorem 2 that under PPS sampling with (10) and (11) fulfilled, the optimal allocation becomes

$$n_0 = n \frac{N_0}{N_0 + \sqrt{x_{\bullet} \sum_{i=N_0+1}^{N} \sigma_i^2 / (\sigma_0^2 x_i)}}. \tag{13}$$

Using the expression (8) for $n_0$ in terms of $\varepsilon$, we find $\varepsilon = n_0 x_\bullet / [(n - n_0)N_0]$ and the $\varepsilon$ minimizing $\mathbb{E}\text{Var}(\hat{T}|Y)$ then becomes

$$\varepsilon = \frac{\sqrt{\sigma_0^2 x_\bullet}}{\sqrt{\sum_{i=N_0+1}^{N} \sigma_i^2 / x_i}}. \tag{14}$$

Further simplifications are possible if we assume that the $Y_i$s in Stratum 1 fulfil a proportional regression model with $1 \leq g \leq 2$,

$$\mathbb{E}Y_i = \beta_1 x_i, \quad \text{Var}Y_i = \sigma_1^2 x_i^g, \quad i = N_0 + 1, \ldots, N \tag{15}$$

and, in addition, the mean–variance relationship is the same in the two strata, that is,

$$\frac{\sigma_0^2}{\beta_0^g} = \frac{\sigma_1^2}{\beta_1^g}. \tag{16}$$

Then, the optimal allocation becomes

$$n_0 = n \frac{(\beta_0/\beta_1)^{g/2} N_0}{(\beta_0/\beta_1)^{g/2} N_0 + \sqrt{x_\bullet (x^{g-1})_\bullet}}, \tag{17}$$

where $(x^{g-1})_\bullet = \sum_{i=N_0+1}^{N} x_i^{g-1}$. The optimal allocation in (17) can alternatively be expressed, using the natural parameter $q = \mathbb{E}T_0/\mathbb{E}T_1$, where $T_0 = \sum_{i=1}^{N_0} Y_i$ and $T_1 = \sum_{i=N_0+1}^{N} Y_i$. We find

$$n_0 = n \frac{\sqrt{N_0^{2-g} (q x_\bullet)^g}}{\sqrt{N_0^{2-g} (q x_\bullet)^g} + \sqrt{x_\bullet (x^{g-1})_\bullet}}. \tag{18}$$

In the special cases with $g = 1$ and $g = 2$, we find

$$n_0 = \begin{cases} n \frac{\sqrt{qk}}{1 + \sqrt{qk}}, & g = 1, \\ n \frac{q}{1+q}, & g = 2, \end{cases} \tag{19}$$

where $k = N_0/(N - N_0)$. Under the model specified in (10), (15) and (16), $q = \frac{\beta_0 N_0}{\beta_1 x_\bullet}$.

Notice that under the assumptions of theorem 2, the optimal choice of $n_0$ (or $\varepsilon$) does not depend on joint inclusion probabilities within the strata.

### 3.2. Robustness

We have investigated robustness of optimal allocation against departures from proportionality in the regression model and against misspecification of the parameters $g$ and $q$. A detailed description of the results may be found in the Supporting Information under the heading Robustness. Throughout the study, the extended proportional regression model given by (10), (15) and (16) was assumed, and stratified sampling was considered. The model we used in the investigations for departure from proportionality is inspired by the applications we have in mind, namely, cell counting (Hansen *et al.*, 2011). We assume that

$$\begin{aligned} Y_i &\sim \text{pois}(\beta_0), & i &= 1, \ldots, N_0, \\ Y_i | X_i = x_i &\sim \text{pois}\left(x_i^\delta\right), & i &= N_0 + 1, \ldots, N. \end{aligned} \tag{20}$$

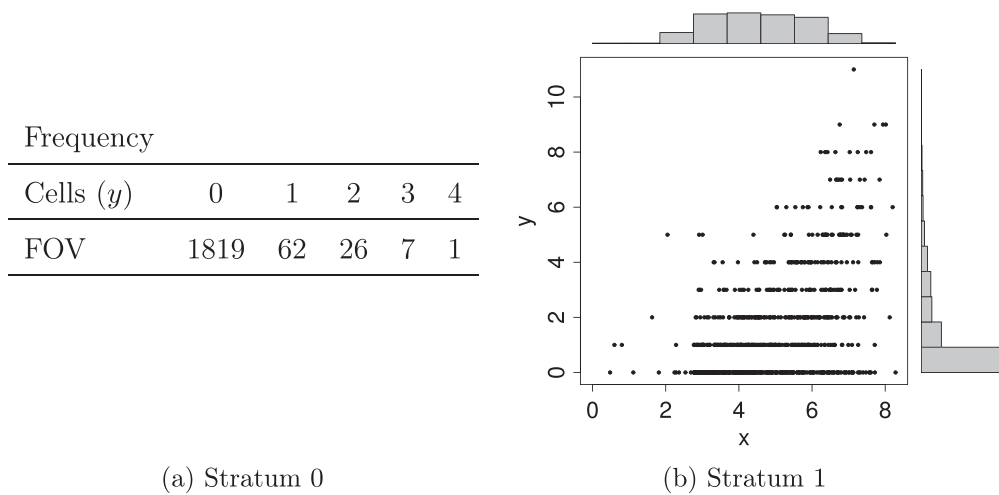| Frequency | | | | | |
|---|---|---|---|---|---|
| Cells ($y$) | 0 | 1 | 2 | 3 | 4 |
| FOV | 1819 | 62 | 26 | 7 | 1 |

(a) Stratum 0         (b) Stratum 1

*Fig. 3.* The empirical distribution of the number of cells $y$ in Stratum 0 ($N_0 = 1915$ FOV with $x = 0$) and a scatterplot of the auxiliary variable $x$ and the number of cells $y$ in Stratum 1 ($N - N_0 = 788$ fields of view with $x > 0$).

To summarize, investigations based on numerical calculations and simulations suggest that the optimum is robust against departures of proportionality in the regression model and misspecification of the parameter $q$, determining the part of the population total stemming from sampling units with vanishing auxiliary variables. Numerical calculations also showed that the parameter $g$, controlling the variance of the variables of interest, given the auxiliary variable, must be chosen with care. Under the model considered with $g = 1$, choosing $g = 2$ caused a substantial loss in efficiency, whereas the opposite case was less pronounced.

## 4. Analysis of real microscopy data

In this section, we use the developed methods in the analysis of a data set from microscopy (Keller *et al.*, 2013), collected with the purpose of estimating osteoclast cell numbers in paws from mice with experimental arthritis. The tissue sections analyzed were divided by a grid into $N = 2703$ small FOV or observation windows. The realized random variable $Y_i$ is the number of cells in FOV $i$, while $x_i$ indicates the amount of a pre-chosen colour in FOV $i$ associated with the staining of the cells. The $x_i$ values are easily determined by automatic image analysis at a low magnification. This is in contrast to the cell counts $Y_i$ which are time-consuming to determine, as they have to be carried out at high magnification by an expert user.

The data set from Keller *et al.* (2013) is unique in the sense that it is exhaustive comprising 100% of FOV and covering the total section areas. Figure 3(a) shows the empirical distribution of cells in Stratum 0 ($N_0 = 1915$ FOV with $x = 0$), and Fig. 3(b) shows a scatterplot of the auxiliary variable $x$ and the number of cells $y$ in Stratum 1 ($N - N_0 = 788$ FOV with $x > 0$).

As the population analyzed in Keller *et al.* (2013) is completely known, containing 10% cells in FOV with $x = 0$ ($q = 1/9$), these data are suitable for testing how far the allocation provided by the new approach using models in combination with optimal allocation is from the actual optimum. We first check the proportionality assumption and choose model parameters, and then study the variance as a function of allocation.

### 4.1. Proportionality

In Fig. 4, it is investigated whether a linear relationship between $x$ and $y$ is a satisfactory description of the data in Stratum 1, consisting of $N_1 = 788$ FOV. The data were partitioned into bins of size 35 ($\pm 1$), resulting in a total number of 22 bins.
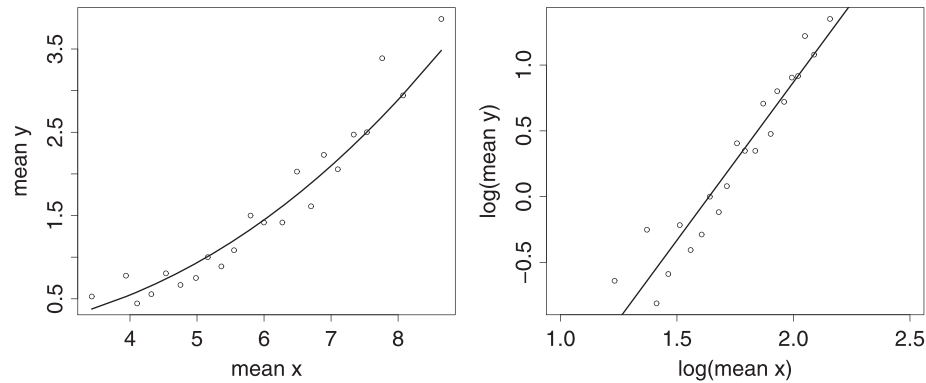
*Fig. 4.* Means (left) and log means (right) of x and y-values in 22 bins in Stratum 1. The curve in the left panel was obtained by transforming the regression line from the right panel.

The mean proportionality in Stratum 1 is not fulfilled (see the left panel in Fig. 4). A linear regression on the log-transformed and binned $x$ and $y$ gave a much more satisfactory description (see Fig. 4 right panel). The estimated relation is $\log y = -3.93 + 2.40 \log x$, which corresponds to a model with $\delta = 2.4$ in (20).

### 4.2. Relation between mean and variance

Under the extended proportional regression model, it is assumed that $\text{Var} Y_i \propto (\mathbb{E} Y_i)^g$ holds in Stratum 1 (cf. (15)). To choose an appropriate value for $g$ for use in (18), we compare the empirical means $\bar{y}$ and variances $s^2$ estimated in the bins in Stratum 1 as specified in Section 4.1. Linear regression of the log-transformed $s^2$ and $\bar{y}$ gives a relation $\log s^2 = 0.44 + 1.22 \log \bar{y}$, which is shown in the left panel of Fig. 5 (full drawn line). Although the slope 1.22 is significantly different from 1 ($p = 0.021$), we will use $g = 1$ in further investigations. For completeness, a line with slope 1 is also shown in the left panel (dotted line). Figure 5, right panel, shows the same estimated relations in a mean–variance plot by transformation of the line in the left panel (full drawn curve), together with a fitted line through the origin (stippled line).

An additional assumption of the model is given in (16), which means that both strata have the same mean–variance relation, that is, the ratio $\text{Var} Y_i / (\mathbb{E} Y_i)^g$ must be the same. Here, we assume $g = 1$. While $s^2 / \bar{y} = 1.69$ in Stratum 0, the ratio was found to be 2.03 in Stratum 1, thus (16) with $g = 1$ appears only approximately fulfilled.



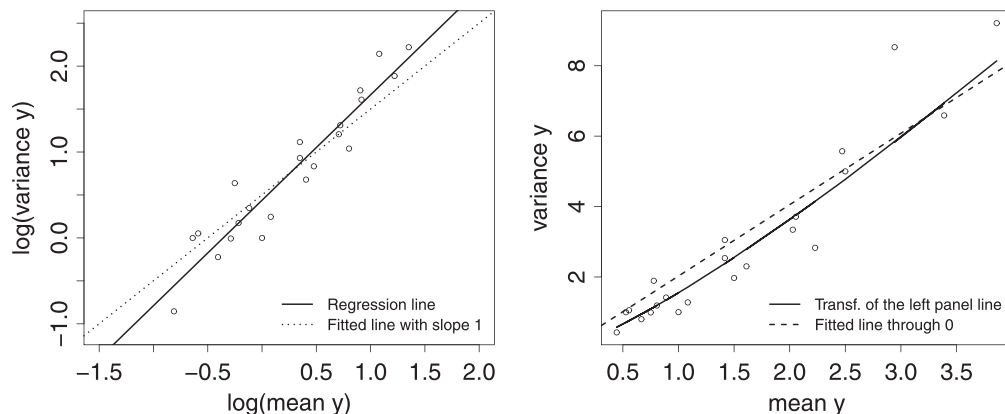*Fig. 5.* The left panel shows log-variances plotted against log-means of y in each of the 22 bins from Stratum 1, together with a regression line (full drawn) and a fitted line with slope 1 (dotted). The right panel shows variances plotted against means of y in each bin, together with a transformation of the left panel regression line (full drawn) and a fitted line through 0 (stippled).
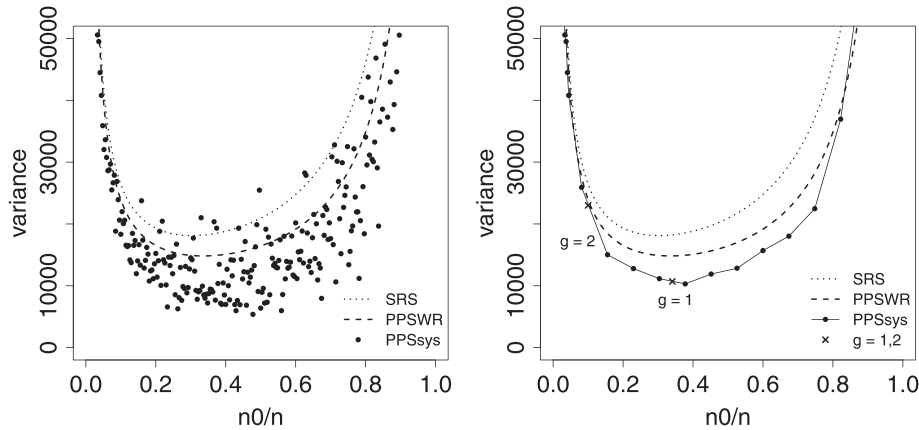
*Fig. 6.* Left: The variance of stratified simple random sampling (dotted line), proportional to size with replacement sampling (stippled line) and systematic proportional to size sampling with an ordering according to the size of the $x_i$s (dots), all under stratification, as a function of $n_0/n$. Right: The same plot, except that the variance for systematic proportional to size sampling is obtained from a partition of the original ones into bins of size 20, where data begin to fluctuate. The sample allocations obtained by (19) with $g = 1$ and $g = 2$ are marked with crosses.

### 4.3. Optimal allocation

Following the results of Section 4.1 and 4.2, we will describe the data by a proportional regression model with $g = 1, \sigma^2/\beta = 2, q = 1/9$ and $\delta = 2.4$. The simulations presented in the Supporting Information indicated that moderate departures from proportionality ($\delta \neq 1$) are not critical for optimal allocation. It thus seems reasonable to use the optimal allocation given in (19) for $g = 1$. We investigate how well this fits the actual optimum and compare with the allocation given by (19) for $g = 2$.

Figure 6 shows for completeness the mean variance of stratified simple random sampling (SRS), stratified PPS with replacement (WR) sampling and stratified systematic PPS sampling as functions of the proportion $n_0/n$ of the sample of size $n = 0.10N$ that is allocated to Stratum 0. The mean variances of PPS WR sampling and SRS are smooth functions, whereas the mean variance of systematic PPS sampling shows a more complicated behaviour, because of the systematic sampling. It is however clear that the mean variance of systematic PPS sampling becomes very large if $n_0$ is chosen too small, and in most cases, the mean variance is smaller than the one for PPS WR sampling and SRS with the same allocation. An overall impression of the mean variance of systematic PPS sampling is obtained by binning of size 20 except for the small (and large) values of $n_0$, which removes the huge fluctuations (see Fig. 6, right panel). Using the binned data, optimal allocation based on (18) for $g = 1$, corresponding to $n_0 = 0.34n$, very well fits the true optimum. If $g = 2$ is used instead, corresponding to $n_0 = 0.10n$, the mean variance becomes a factor 2 larger. From (8) we obtain that $n_0 = 0.34n$ corresponds to $\varepsilon = 1.05$ and $n_0 = 0.10n$ corresponds to $\varepsilon = 0.23$, hence $\varepsilon = 1$ used in Keller *et al.* (2013) was in fact remarkably close to the optimal choice.

## 5. Conclusion and further perspectives

In applications of PPS sampling in microscopy, vanishing, that is zero-valued, auxiliary variables are dealt with by adding a small positive constant $\varepsilon$. In the present paper, it has been shown, both theoretically and in applications, that it is of great importance to choose this constant wisely, in order to obtain an efficient estimator. To solve the problem of choosing the constant in an optimal manner, a model-assisted approach has been suggested, where the mean variance is minimized. The optimization depends on the choice of just a few parameters in the

model given by (10), (15) and (16). If not yet any or only little data are available, we recommend to use $g = 1$ as a robust guess and give a rough guess on the remaining parameter $q$ in the formula for the optimal $\varepsilon$.

The optimal $\varepsilon$ may be so large that it makes the Stratum 0 sampling probabilities larger than the Stratum 1 sampling probabilities for units with the smallest positive $x$-values. For our data, this happened only for 3 out of 788 units in Stratum 1. If this phenomenon is more predominant and it is expected that the vanishing variables are zero-valued, we suggest to enlarge Stratum 0 with units having small positive $x$-values until the sampling probability for units in Stratum 0 is smaller than any of the sampling probabilities for units in Stratum 1.

Throughout the paper, we have mainly used the term vanishing auxiliary variables for zero-valued size variables, and both the robustness results and analysis of microscopy data are examples of this case. However, the optimality result also applies to the case, where Stratum 0 solely consists of units where the auxiliary variable has not been determined.

In the present paper, we have focused on the Horvitz–Thompson estimator. Note, however, that under PPS sampling with one auxiliary variable, the Horvitz–Thompson estimator actually coincides with the ($\pi$-weighted) ratio estimator and the general regression estimator (Särndal *et al.*, 1992, p. 180 and 225). An interesting future research question is to try to generalize theorem 2 such that the assumption $\mathbb{E}Y_i \propto \tilde{\pi}_i$ in (11) is relaxed. Theorem 2 depends crucially on formula (3) for the mean variance, which does not depend on the second-order inclusion probabilities. We have already seen that (3) holds when $\mathbb{E}Y_i$ is proportional to the sampling probability of unit $i$, but it can in fact also be shown, that for specific sampling designs with samples balanced on $\mathbb{E}Y_i$, (3) holds for arbitrary inclusion probabilities – see Nedyalkova & Tillé (2008) for details on balanced sampling. This observation opens up for establishing a theorem 2 in a more general setting, allowing for other inclusion probabilities, for example, the optimal inclusion probabilities found in Nedyalkova & Tillé (2008), which are proportional to the standard deviations of the $Y_i$s. Note also that under balanced sampling on $\mathbb{E}Y_i$, in the sense of Nedyalkova & Tillé (2008), the Horvitz–Thompson estimator and the regression estimator coincide.

## Acknowledgements

## References

Bellhouse, D. R. & Rao, J. (1975). Systematic sampling in the presence of a trend. *Biometrika* **62**, (3), 694–697.

Gardi, J. E., Nyengaard, J. R. & Gundersen, H. J. G. (2008a). Automatic sampling for unbiased and efficient stereological estimation using the proportionator in biological studies. *J. Microsc.* **230**, (1), 108–120.

Gardi, J. E., Nyengaard, J. R. & Gundersen, H. J. G. (2008b). The proportionator: unbiased stereological estimation using biased automatic image analysis and non-uniform probability proportional to size sampling. *Comput. Biol. Chem.* **38**, (3), 313–328.

Gundersen, H. J. G. (2002). The smooth fractionator. *J. Microsc.* **207**, (3), 191–210.

Hansen, L. V., Kiderlen, M. & Jensen, E. B. V. (2011). Image-based empirical importance sampling: an efficient way of estimating intensities. *Scand. J. Stat.* **38**, (3), 393–408.

Keller, K. K., Andersen, I. T., Andersen, J. B., Hahn, U., Steengaard-Pedersen, K., Hauge, E.-M. & Nyengaard, J. R. (2013). Improving efficiency in stereology: a study applying the proportionator and the autodisector on virtual slides. *J. Microsc.* **251**, (1), 68–76.

Murthy, M. N. (1967). *Sampling theory and methods*, Statistical Publishing Society, Calcutta.

Nedyalkova, D. & Tillé, Y. (2008). Optimal sampling and estimation strategies under the linear model. *Biometrika* **95**, (3), 521–537.

Särndal, C. E., Swensson, B. & Wretman, J. H. (1992). *Model assisted survey sampling*, Springer Verlag, New York.

*Received December 2013, in final form February 2015*

Ute Hahn, Department of Mathematics, Aarhus University.
E-mail: ute@math.au.dk

**Supporting information**

Additional supporting information may be found in the online version of this article at the publisher's web site.

# Supplement A

This chapter contains proofs of the theorems in Paper A and a detailed investigation of the robustness of the optimal design. This information may also be found in the online version of Paper A at the publisher's web site. References in this chapter are with respect to Paper A.

## AA.1 Proofs

*Proof of Theorem 1.* To find $\mathbb{E}\mathrm{Cov}(\widehat{T_0}, \widehat{T_1} \mid Y)$, write

$$\mathbb{E}\mathrm{Cov}(\widehat{T_0}, \widehat{T_1} \mid Y) = \mathrm{Cov}(\widehat{T_0}, \widehat{T_1}) - \mathrm{Cov}\Big(\mathbb{E}(\widehat{T_0} \mid Y), \mathbb{E}(\widehat{T_1} \mid Y)\Big). \tag{AA.1}$$

Obviously, the parts $\widehat{T_h} := \sum_{i \in S \cap U_h} Y_i/\pi_i$, $h = 0, 1$, are design-unbiased estimators for the stratum totals $T_h = \sum_{i \in U_h} Y_i$. Together with the uncorrelatedness of $Y_i, Y_j, i \neq j$, which results in uncorrelatedness of $T_0$ and $T_1$, we thus have

$$\mathrm{Cov}\Big(\mathbb{E}(\widehat{T_0} \mid Y), \mathbb{E}(\widehat{T_1} \mid Y)\Big) = \mathrm{Cov}(T_0, T_1) = 0. \tag{AA.2}$$

Furthermore, $\widehat{T_0}$ and $\widehat{T_1}$ are also uncorrelated, given the sample $S$. Thus,

$$\begin{aligned}
\mathrm{Cov}(\widehat{T_0}, \widehat{T_1}) &= \mathrm{Cov}\Big(\mathbb{E}(\widehat{T_0} \mid S), \mathbb{E}(\widehat{T_1} \mid S)\Big) + \mathbb{E}\mathrm{Cov}(\widehat{T_0}, \widehat{T_1} \mid S) \\
&= \mathrm{Cov}\Big(\mathbb{E}(\widehat{T_0} \mid S), \mathbb{E}(\widehat{T_1} \mid S)\Big) \\
&= \mathbb{E}\Big(\mathbb{E}(\widehat{T_0} \mid S) - \mathbb{E}\widehat{T_0}\Big)\Big(\mathbb{E}(\widehat{T_1} \mid S) - \mathbb{E}\widehat{T_1}\Big).
\end{aligned} \tag{AA.3}$$

Since the $Y_i$s in Stratum 0 all have the same mean $\mathbb{E}Y_i = \mathbb{E}T_0/N_0$, and the same inclusion probability $\pi_i = n_0/N_0$,

$$\mathbb{E}(\widehat{T_0} \mid S) = \sum_{i \in S \cap U_0} \frac{\mathbb{E}Y_i}{\pi_i} = M_0 \frac{\mathbb{E}T_0}{n_0} \tag{AA.4}$$

depends only on the number $M_0$ of sampled units from $U_0$. Due to the systematic sampling procedure, this number only takes the values $\lfloor n_0 \rfloor$ or $\lfloor n_0 \rfloor + 1$, where $\lfloor n_0 \rfloor$ is the integer part of $n_0 = \mathbb{E}M_0$. With probability $v = n_0 - \lfloor n_0 \rfloor$, we have that $M_0 = \lfloor n_0 \rfloor + 1$, and with probability $1 - v$, it is $M_0 = \lfloor n_0 \rfloor$. Writing

$$A := \{S : M_0 = \lfloor n_0 \rfloor + 1\}, \tag{AA.5}$$

and $\tau_0 = \mathbb{E}T_0/n_0$, we get

$$\mathbb{E}(\widehat{T_0} \mid S) - \mathbb{E}\widehat{T_0} = M_0\tau_0 - n_0\tau_0 = \begin{cases} (1-\nu)\tau_0, & S \in A, \\ -\nu\tau_0, & S \notin A. \end{cases} \tag{AA.6}$$

Thus,

$$\Big(\mathbb{E}(\widehat{T_0} \mid S) - \mathbb{E}\widehat{T_0}\Big)\Big(\mathbb{E}(\widehat{T_1} \mid S) - \mathbb{E}\widehat{T_1}\Big) = \begin{cases} (1-\nu)\tau_0\Big(\mathbb{E}(\widehat{T_1} \mid S) - \mathbb{E}T_1\Big), & S \in A, \\ -\nu\tau_0\Big(\mathbb{E}(\widehat{T_1} \mid S) - \mathbb{E}T_1\Big), & S \notin A. \end{cases} \tag{AA.7}$$

Let $f_1(S) = \mathbb{E}(\widehat{T_1} \mid S) - \mathbb{E}T_1$. Taking the expectation of (AA.7), we obtain

$$\begin{aligned} \mathrm{Cov}(\widehat{T_0}, \widehat{T_1}) &= (1-\nu)\tau_0\mathbb{E}(f_1(S) \mid S \in A)P(S \in A) - \nu\tau_0\mathbb{E}(f_1(S) \mid S \notin A)P(S \notin A) \\ &= (1-\nu)\tau_0\mathbb{E}(f_1(S) \mid S \in A)\nu - \nu\tau_0\mathbb{E}(f_1(S) \mid S \notin A)(1-\nu), \end{aligned} \tag{AA.8}$$

so $\widehat{T_0}$ and $\widehat{T_1}$ are uncorrelated if $\nu = 0$. For $\nu > 0$, use the notation $\tau_1 = \mathbb{E}(f_1(S) \mid S \notin A)/\nu$; since $\mathbb{E}f_1(S) = 0$, and $P(S \in A) = \nu$, we get

$$\mathbb{E}(f_1(S) \mid S \in A)P(S \in A) = -\mathbb{E}(f_1(S) \mid S \notin A)P(S \notin A) = -\tau_1\nu(1-\nu), \tag{AA.9}$$

hence

$$\mathrm{Cov}(\widehat{T_0}, \widehat{T_1}) = -(1-\nu)^2\nu\tau_0\tau_1 - \nu^2(1-\nu)\tau_0\tau_1 = -\nu(1-\nu)\tau_0\tau_1. \tag{AA.10}$$

For a model with proportionality in Stratum 1 between $\mathbb{E}Y_j$ and $\pi_j$ such that $\mathbb{E}Y_j/\pi_j = \mathbb{E}T_1/n_1$ for all $N_0 + 1 \leq j \leq N$, we get

$$\mathbb{E}(\widehat{T_1} - T_1 \mid S) = \sum_{j \in S \cap U_1} \frac{\mathbb{E}Y_j}{\pi_j} - \mathbb{E}T_1 = (\#(S \cap U_1) - n_1)\frac{\mathbb{E}T_1}{n_1}. \tag{AA.11}$$

Thus, the definition of $\tau_1$ reduces to $\tau_1 = \mathbb{E}T_1/n_1$, where $n_1 = n - n_0$ is the expected number of sampled units in Stratum 1.                                                                                                    $\square$

*Proof of Theorem 2.* The expected variance of $\widehat{T}$ is

$$\mathbb{E}\mathrm{Var}(\widehat{T}|Y) = \mathbb{E}\mathrm{Var}(\widehat{T_0}|Y) + \mathbb{E}\mathrm{Var}(\widehat{T_1}|Y), \tag{AA.12}$$

since, conditionally on $Y$, $\widehat{T_0}$ and $\widehat{T_1}$ are independent. Within each stratum the mean values of $Y_i$ is proportional to the inclusion probabilities $\pi_i$ given in (9) and, using (3) on each stratum separately, we find for $n_0 \leq N_0$ and $n_0 < n$

$$\mathbb{E}\mathrm{Var}(\widehat{T}|Y) = N_0\left(\frac{N_0}{n_0} - 1\right)\sigma_0^2 + \sum_{i=N_0+1}^{N}\left(\frac{1}{\left(1 - \frac{n_0}{n}\right)\tilde{\pi}_i} - 1\right)\sigma_i^2 = f(n_0), \tag{AA.13}$$

say. We will now, as is usual in optimal allocation problems, replace the fundamentally discrete problem, the choice of the number $n_0$ of units to be sampled in Stratum

0, by a continuous one, so we replace $n_0$ in $f(n_0)$ by a positive real parameter $\eta_0$, say. We find

$$f'(\eta_0) = -\frac{N_0^2 \sigma_0^2}{\eta_0^2} + \frac{V}{n\left(1 - \frac{\eta_0}{n}\right)^2}, \tag{AA.14}$$

where

$$V = \sum_{i=N_0+1}^{N} \frac{\sigma_i^2}{\tilde{\pi}_i}. \tag{AA.15}$$

The equation $f'(\eta_0) = 0$ can be written as

$$\left(\frac{N_0^2}{n}\sigma_0^2 - V\right)\eta_0^2 - 2\sigma_0^2 N_0^2 \eta_0 + n\sigma_0^2 N_0^2 = 0, \tag{AA.16}$$

which has the following solution for $\eta_0 < n$

$$\eta_0 = \frac{nN_0}{N_0 + \sqrt{nV/\sigma_0^2}}. \tag{AA.17}$$

This solution is a minimum of $f(\eta_0)$. Restricting to integer values $n_0$ with $n_0 \leq N_0$, we get that the optimal $n_0$ is one of the integers closest to (12). $\qquad\square$

## AA.2 Robustness

In this section, we investigate the robustness of the optimal allocation under the extended proportional regression model against departures from this model and parameter misspecification. We study the relative inflation in mean variance

$$R = \frac{\mathbb{E}\mathrm{Var}(\widehat{T}(n_0')|Y)}{\mathbb{E}\mathrm{Var}(\widehat{T}(n_0)|Y)}, \tag{AA.18}$$

where $n_0$ and $n_0'$ are calculated according to (18), $n_0$ with the true values of the parameters and the correct model, and $n_0'$ with alternative parameter values or the wrong model, and $\widehat{T}(.)$ is the resulting estimator.

### AA.2.1 Robustness against parameter misspecification

Consider the case that the extended proportional regression model given by (10), (15) and (16) holds, but that the parameter $g$ is misspecified. This parameter controls the mean-variance relation, viz.

$$\mathrm{Var}\,Y_i \propto (\mathbb{E}Y_i)^g. \tag{AA.19}$$

In Table AA.1, $R$ is shown for the case $g = 1$. For $n_0'$, the true value of $q$ is used, but $g$ is wrongly assumed to be 2. Using (19), we find

$$R = \frac{(1+q)(1+k)(1 - \frac{n}{N})}{(1 + \sqrt{qk})^2 - (1+q)(1+k)\frac{n}{N}}. \tag{AA.20}$$

Table AA.1 shows the value of $R$ for different combinations of $k$ and $q$, with $n/N = 0.1$. The range of $k$ and $q$ has been chosen such that it represents what is expected in the application we have in mind, see Section 4. It is seen, that the mean variance is increased quite markedly if we wrongly assume that $g = 2$. The same conclusion holds for other choices of $n/N$. We also considered the 'opposite' case, where the

**Table AA.1:** The relative inflation $R$ in mean variance, given by (AA.20), by wrongly using $g = 2$, when the true value is $g = 1$. The results are for varying values of $q$ and $k$, when $n/N = 0.1$.

| $k$ \ $q$ | 0.025 | 0.05 | 0.10 | 0.15 |
|---|---|---|---|---|
| 0.25 | 1.113 | 1.069 | 1.028 | 1.010 |
| 0.50 | 1.278 | 1.197 | 1.115 | 1.071 |
| 1.00 | 1.624 | 1.468 | 1.309 | 1.222 |
| 2.00 | 2.326 | 2.000 | 1.683 | 1.514 |
| 4.00 | 3.781 | 3.011 | 2.341 | 2.010 |

true value of $g$ is 2, but $g$ is wrongly assumed to be 1. In this case, $R$ also depends on the realized values of the auxiliary variable. With $C_x = (N - N_0)(x^2)_{\cdot}/(x_{\cdot})^2 = N(x^2)_{\cdot}/((1 + k)(x_{\cdot})^2)$, we get

$$R = \frac{(1 + \sqrt{qk})(1 + \frac{q^2}{\sqrt{qk}}) - \frac{n}{N}(1 + k)(\frac{q^2}{k} + C_x)}{(1 + q)^2 - \frac{n}{N}(1 + k)(\frac{q^2}{k} + C_x)}. \tag{AA.21}$$

If the auxiliary variables $x_{N_0+1}, \ldots, x_N$ are i.i.d. realizations of a random variable $X$, we have $C_x \approx \mathbb{E}(X^2)/\mathbb{E}(X)^2$. Since $\pi_i = (n - n_0)x_i/x_{\cdot} \leq 1$, very skew distributions are excluded and thus large values of $C_x$ are prevented. In the case illustrated in Table AA.2, where we let $C_x = 1.3$, $R$ was much closer to one than in the previous case where $g = 1$, but $g = 2$ is assumed. The same conclusion holds for other choices of $n/N$.

**Table AA.2:** The relative inflation $R$ in the mean variance, given by (AA.21), by wrongly using $g = 1$, when the true value is $g = 2$. The results are for varying values of $q$ and $k$, when $n/N = 0.1$ and $C_x = 1.3$.

| $k$ \ $q$ | 0.025 | 0.05 | 0.10 | 0.15 |
|---|---|---|---|---|
| 0.25 | 1.042 | 1.036 | 1.020 | 1.009 |
| 0.50 | 1.079 | 1.082 | 1.068 | 1.050 |
| 1.00 | 1.142 | 1.160 | 1.156 | 1.137 |
| 2.00 | 1.267 | 1.315 | 1.329 | 1.311 |
| 4.00 | 1.670 | 1.780 | 1.802 | 1.752 |

Table AA.3 shows values of $R$, when $n'_0$ is calculated, using the true values of $g = 1$ and $k = 2$, and varying values of a guessed value $\hat{q}$ and a true value of $q$. The formula is here

$$R = \frac{(1 + \sqrt{\hat{q}k})(1 + \sqrt{\frac{q^2}{\hat{q}}k}) - (1 + q)(1 + k)\frac{n}{N}}{(1 + \sqrt{qk})^2 - (1 + q)(1 + k)\frac{n}{N}}. \tag{AA.22}$$

These results indicate that $R$ is robust against misspecification of $q$.

**Table AA.3:** The relative inflation $R$ in mean variance, given by (AA.22), for varying values of the guessed value $\hat{q}$ and the true value $q$, when $n/N = 0.1$, $g = 1$ and $k = 2$.

| $\hat{q}$ \ $q$ | 0.025 | 0.05 | 0.10 | 0.15 |
|---|---|---|---|---|
| 0.025 | 1.000 | 1.030 | 1.133 | 1.185 |
| 0.05 | 1.023 | 1.000 | 1.031 | 1.055 |
| 0.10 | 1.096 | 1.028 | 1.000 | 1.002 |
| 0.15 | 1.155 | 1.065 | 1.009 | 1.000 |

## AA.2.2 Robustness against departures from proportionality

Let us now investigate the robustness of the optimal allocation based on the proportional regression model against departures from proportionality between $x_i$ and $\mathbb{E}Y_i$. The model we used in the investigations for departure from proportionality is inspired by the applications we have in mind, namely cell counting, see Hansen et al. (2011). We assume that

$$
\begin{aligned}
Y_i &\sim \text{pois}(\beta_0), & i &= 1,\dots,N_0, \\
Y_i|X_i = x_i &\sim \text{pois}(x_i^\delta), & i &= N_0+1,\dots,N.
\end{aligned}
\tag{AA.23}
$$

Note that in Stratum 1, a model of this type is obtained if $Y_i$ is the number of cells observed in field of view $i$ and the cell centres follow a Cox point process with cumulated random intensity of the form $x_i^\delta$. For all choices of $\delta$, the model (AA.23) represents a mean-variance relation with $g = 1$, i.e., $\text{Var}\,Y_i \propto (\mathbb{E}Y_i)^1$. If $\delta = 1$, the extended proportional regression model holds with $\sigma_0^2 = \beta_0$ and $\sigma_1^2 = \beta_1$. If instead $\delta \neq 1$, the proportional regression relationship between $x_i$ and $Y_i$ does not hold for $i > N_0$.

We study the consequences of using $n_0$ as given in (19) with $g = 1$, even though the underlying assumptions are not fulfilled. We use PPS sampling in Stratum 1 with inclusion probabilities proportional to the $x_i$s. In contrast to the case $\delta = 1$, the mean variance for $\delta \neq 1$ is influenced by the specific design. We focus on systematic PPS sampling, with an ordering according to the size of the $x_i$s.

Various distributions of the auxiliary variable have been tested. Here, we present the results for the case where $x_i$ is a realization of

$$
X_i \sim \text{beta}(\gamma_1,\gamma_2)\rho + \tau,
\tag{AA.24}
$$

$i = N_0+1,\dots,N$. The relative inflation $R$ in mean variance due to allocation following (19) with $g = 1$ was calculated for various values of $q$ and $\delta$ in the model given by (AA.23) and (AA.24). The correct value of $q$ was used in the allocation. For each pair of parameters $q$ and $\delta$, one realization of the $X_i$s was considered, and the true optimal value of $n_0$, which is needed for calculation of the denominator of $R$, was determined. In all the cases considered, $0.025 \leq q \leq 0.15$ and $0.5 \leq \delta \leq 2$, optimal allocation assuming proportionality showed robustness against departures from proportionality ($R \in [1;1.03]$).

Table AA.4 shows the empirical distribution of $Y$ in Stratum 0 and Figure AA.1 shows scatterplots of $(X, Y)$ for realizations of the model with $\delta$=0.5, 1 and 2, respectively. The parameter $\beta_0$ is chosen such that the parameter $q = \mathbb{E}T_0/\mathbb{E}T_1 = 0.05$ is the same in all three cases.

**Table AA.4:** The empirical distribution of $Y$ in Stratum 0 for a realization of the model specified in (AA.23) and Figure AA.1.

| Frequency | | | |
|---|---|---|---|
| $y$ | 0 | 1 | 2 |
| Units | 618 | 46 | 2 |
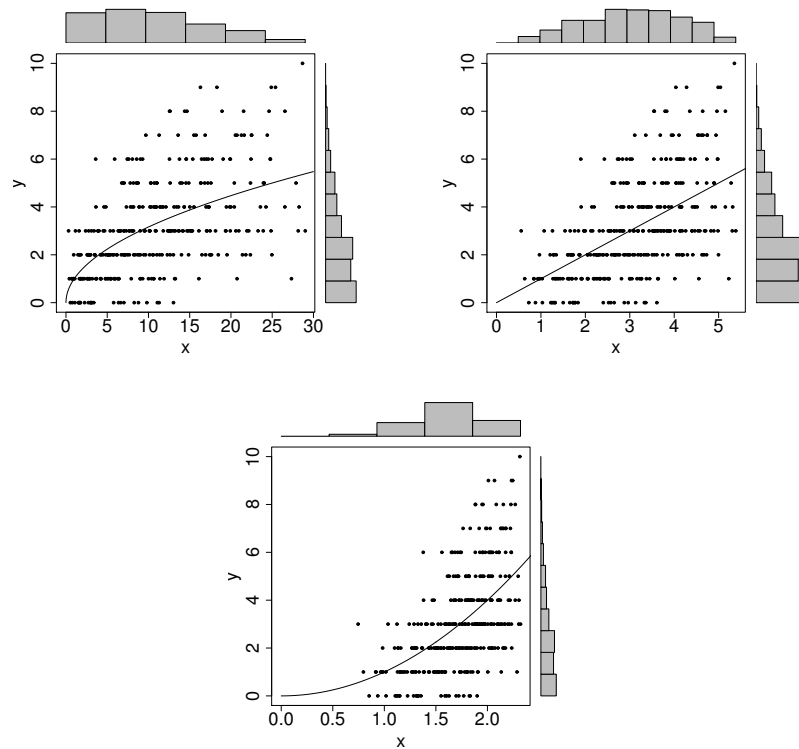


**Figure AA.1:** Scatterplots of $(X, Y)$ in Stratum 1 for realizations of the model given by (AA.23) and (AA.24) with (left to right) $\delta$=0.5, 1 and 2, respectively, together with the mean relation $Y = X^\delta$. The parameters in the distribution (AA.24) are $\gamma_1 = \gamma_2 = 2$, $\rho = 5$, $\tau = 0.5$ while the parameter $q$ is equal to 0.05 in all three cases. The values of $N$ and $N_0$ are $N = 1000$ and $N_0 = 2/3N$. The empirical marginal distributions of $X$ and $Y$ are shown on the upper and right side of the graphs.

In Figure AA.2, the mean variance is shown for $n$ fixed ($n = 100$) as a function of the sample proportion $n_0/n$ in Stratum 0. The variances are shown for simple random sampling (SRS), PPS with replacement (WR) sampling, and systematic PPS sampling, all under stratification. Note that the variances are only shown for a range of the values of $n_0$, as the variance becomes very large for extreme choices of $n_0$. This emphasizes the importance of good choices of $n_0$. Although the variances differ,
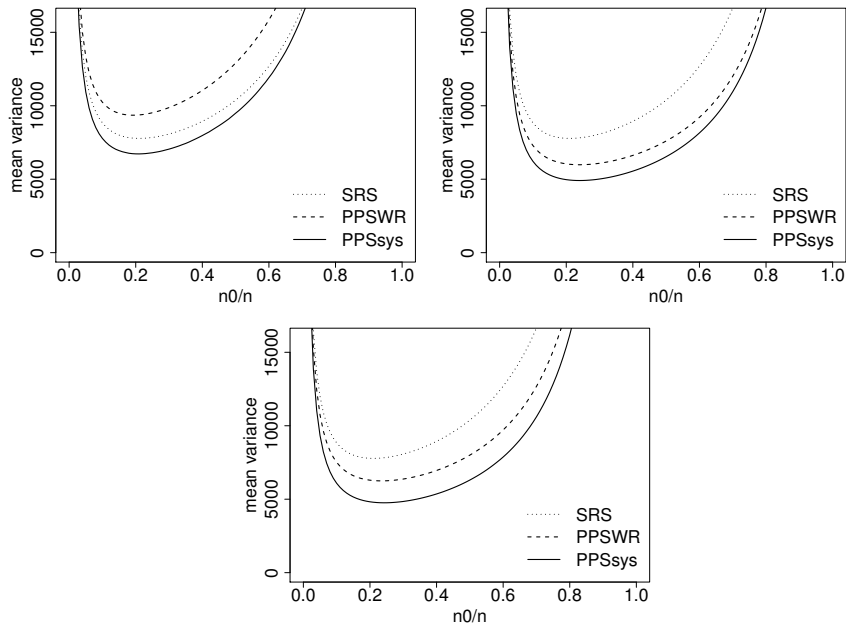
**Figure AA.2:** The mean variance of the model given by (AA.23) and (AA.24) under SRS, PPS WR sampling and systematic PPS sampling, all under stratification, is shown for $q = 0.05$ and (from left to right) $\delta$=0.5, 1 and 2, respectively, as a function of the sample proportion in Stratum 0. The remaining parameter values are specified in Figure AA.1. For more details, see the text.

the optimal allocations are almost identical for PPS WR sampling and systematic PPS sampling. In the case $\delta = 0.5$, shown to the left of Figure AA.2, we gain much from using systematic sampling, as PPS WR and systematic PPS sampling, differ the most in this case. Here, SRS actually performs better than PPS WR sampling. This can be mainly ascribed to stratification, as without stratification, the variance of SRS is approximately 35.000 in the case of $\delta = 0.5$.

# Bibliography

Hansen, L. V., M. Kiderlen, and E. B. V. Jensen (2011). Image-based empirical importance sampling: An efficient way of estimating intensities. *Scandinavian Journal of Statistics 38*(3), 393–408.

**Paper B**

# Improving efficiency in stereology: a study applying the proportionator and the autodisector on virtual slides

By K. K. Keller, I. T. Andersen, J. B. Andersen, U. Hahn, K. Stengaard-Pedersen, E.–M. Hauge & J. R. Nyengaard

# Improving efficiency in stereology: a study applying the proportionator and the autodisector on virtual slides

K.K. KELLER*, I.T. ANDERSEN†, J.B. ANDERSEN‡,§, U. HAHN†, K. STENGAARD-PEDERSEN*, E.-M. HAUGE* & J.R. NYENGAARD‡

*Department of Rheumatology, Aarhus University Hospital, Aarhus, Denmark*

†*Department of Mathematics, Center for Stochastic Geometry and Advanced Bioimaging, University of Aarhus, Aarhus, Denmark*

‡*Stereology and Electron Microscopy Laboratory, Center for Stochastic Geometry and Advanced Bioimaging, Aarhus University Hospital, Aarhus, Denmark*

§*Visiopharm A/S, Hørsholm, Denmark*

## Summary

Cell counting in stereology is time-consuming. The proportionator is a new stereological sampling method combining automatic image analysis and non-uniform sampling. The autodisector on virtual slides combines automatic generation of disector pairs with the use of digital images. The aim of the study was to investigate the time efficiency of the proportionator and the autodisector on virtual slides compared with traditional methods in a practical application, namely the estimation of osteoclast numbers in paws from mice with experimental arthritis and control mice. Tissue slides were scanned in a digital slide scanner and the autodisector was applied on the obtained virtual tissue slides. Every slide was partitioned into fields of view, and cells were counted in all of them. Based on the original exhaustive data set comprising 100% of fields of view and covering the total section area, a proportionator sampling and a systematic, uniform random sampling were simulated. We found that the proportionator was 50% to 90% more time efficient than systematic, uniform random sampling. The time efficiency of the autodisector on virtual slides was 60% to 100% better than the disector on tissue slides. We conclude that both the proportionator and the autodisector on virtual slides may improve efficiency of cell counting in stereology.

## Introduction

Systematic, uniform random sampling (SURS) has been the gold standard in the field of stereology for decades (Gundersen & Jensen, 1987). This principle is commonly applied when determining the total number of cells in an organ from parallel sections; the sections to be evaluated are selected by SURS. In the next step, SURS is used in a computer-aided system to select a known and homogeneously distributed fraction of the fields of view (FOVs) that are subsequently evaluated using stereological estimators. These methods may be time-consuming, if the cell population is very inhomogeneous, as a large number of FOVs must be examined to obtain a reasonable coefficient of error (CE = square root of standard error, divided by the mean). Thus, new methods to increase sampling efficiency are needed.

Recently, the proportionator was introduced by Gardi *et al.* (2008a, b), combining automatic image analysis and non-uniform sampling. This method aims at avoiding evaluation of FOVs with low cell number, and is applicable if regions with high cell number can be visually distinguished from regions with low cell number, as is the case when the cells are well stained. First, the experimenter determines, e.g. a colour of interest (colour of the stained cells). This colour is used by the image analysis system to assign a weight to every FOV, based on number of pixels with similar colour. Subsequently, the FOVs are sampled with probability proportional to their weights, using an advanced systematic sampling scheme that combines various variance reduction ideas known from traditional statistics, and from modern stereology. Prior to sampling, the FOVs are arranged according to the principles of the smooth fractionator (Gundersen, 2002), which corresponds to so-called balanced systematic sampling described in Murthy (1967, Section 5.9d). Next, systematic probability proportional to size (pps) sampling (Madow, 1949) is applied to select the FOVs. Finally, the total cell number is estimated design-unbiased using the Horvitz–Thompson estimator, which is

Correspondence to: K.K. Keller, Department of Rheumatology, Aarhus University Hospital, Nørrebrogade 44, DK-8000 Aarhus C, Denmark. Tel: +45 7846 4411; fax: +45 7846 4412; e-mail: kresten@ana.au.dk

a weighted sum of the sampled cells (Horvitz & Thompson, 1952). The CE of this estimate depends on the relation between the weights and the counts. Ideally, the weights will be directly proportional to the counts. However, even with a poor or no relation, the proportionator will still yield a design-unbiased estimate, but the CE may be higher than for SURS.

The proportionator has been successfully implemented in a computer simulation as well as in three different tissues but only using one or two individuals per study setup (Gardi *et al.*, 2008a, b). In order to learn more about the proportionator, it is necessary to investigate it in a setting with several individuals.

Osteoclasts are the main cells responsible for bone degradation in arthritis (Gravallese *et al.*, 1998). Therefore, estimation of the number of osteoclasts on bone surfaces is important in arthritis research. Recently, we estimated the total number of osteoclasts in the SKG model of rheumatoid arthritis using the fractionator (Gundersen, 1986; Keller *et al.*, 2011). The SKG mouse model shares many similarities with rheumatoid arthritis and is characterised by polyarthritis, elevated cytokines as well as extra-articular manifestations of arthritis (Sakaguchi *et al.*, 2003; Kobayashi *et al.*, 2006). Since osteoclasts are very inhomogeneously distributed in arthritic mouse paws, this cell type seems appropriate for studying the proportionator.

The routine use of digital images in the field of pathology has considerable potential, but has until recently not been used. However, this is changing in the current years due to improvement of the technique and distribution of slide scanners (Jara-Lazaro *et al.*, 2010; Taylor, 2011). To date, no stereological study has applied these techniques, and therefore it is unknown whether the use of digital images can improve time efficiency in counting. In addition, no study has combined the use of digital images with the automated aligning of images for the physical disector (autodisector).

The primary objective was to compare the proportionator with the traditional SURS. The secondary objective was to investigate time efficiency of the autodisector on virtual slides compared to the traditional disector on tissue slides.

## Methods and materials

### SKG mice

Paws from 7 female SKG mice with and 5 without arthritis from a larger study were evaluated (Keller *et al.*, 2012). Six weeks after arthritis induction with 20 mg mannan intraperitoneally (Hashimoto *et al.*, 2010), the mice were anesthetised with isoflurane (Baxter, Deerfield, Illinois, USA) and euthanised by cervical dislocation. The study was approved by the Danish Animal Experiments Inspectorate.

### Histological preparation of mouse paws

The right hind paw was cut 0.5 cm above the ankle joint and immersed in 70% alcohol for fixation, followed by dehydration, and finally the paws were embedded undecalcified in methyl methacrylate (Fig. 1a).

The width of the paw was measured and we aimed at obtaining approximately 10 section levels using SURS (Gundersen & Jensen, 1987) (Fig. 1b). Section pairs, 7-$\mu$m-thick, were sampled at each section level using a microtome (R. Jung GmbH, Heidelberg, Germany) (Fig. 1c). The number of section levels was only approximate, because it relies upon the measured thickness of the paw, which will never be precise. The enzymatic TRAP stain was used to detect osteoclasts.

### Virtual tissue slides

Tissue slides were scanned in a high-resolution digital slide scanner (NanoZoomer 2.0 series, Hamamatsu, Japan) using the software provided with the scanner (vers. 2.2.60). Virtual slides were generated at a total magnification of $\times 589$.

### Data collection

The newCAST software vers. 3.6.5.0 and the add-on module autodisector (Visiopharm, Hoersholm, Denmark) were used for automated aligning of images for the physical disector (Sterio, 1984). In short, overview images were automatically created from the virtual slides and linked. Subsequently, the software automatically aligned each pair of virtual slides, and after delineating the region of interest with a mask, the autodisector automatically sampled and stored corresponding disector images. We sampled an area fraction of 100% using a counting frame of 107912 $\mu$m$^2$ at a total magnification of $\times 589$. Finally, the samples were loaded, and cell counting was performed.

### Image analysis and weights

The protocol for calculating the weights was made using the visiomorph software (Visiopharm). A linear Bayesian pixel classifier was applied in RGB space running a training process of marking the background, tissue and osteoclasts on a representative image ending up with three possible classes (Gonzales & Woods, 2008). It was checked that the classification protocol was robust across images. A post process step was included in the protocol with an output calculation defined as *weight* ($w_{raw}$) which equals the area of classified osteoclasts (in pixels) as described by Gardi *et al.* (2008a). The weights were generated on all FOVs (counting frames) serving as reference sections. For sampling of FOVs, the weights were transformed to $w = \log(w_{raw} + 1) + 1$, since the original weights had a very skewed distribution, and many of the original weights were equal to zero. The logarithm transformation ensures that the distribution of weights resembles the distribution of cell numbers, and the constant was added to avoid bias due to vanishing weights.
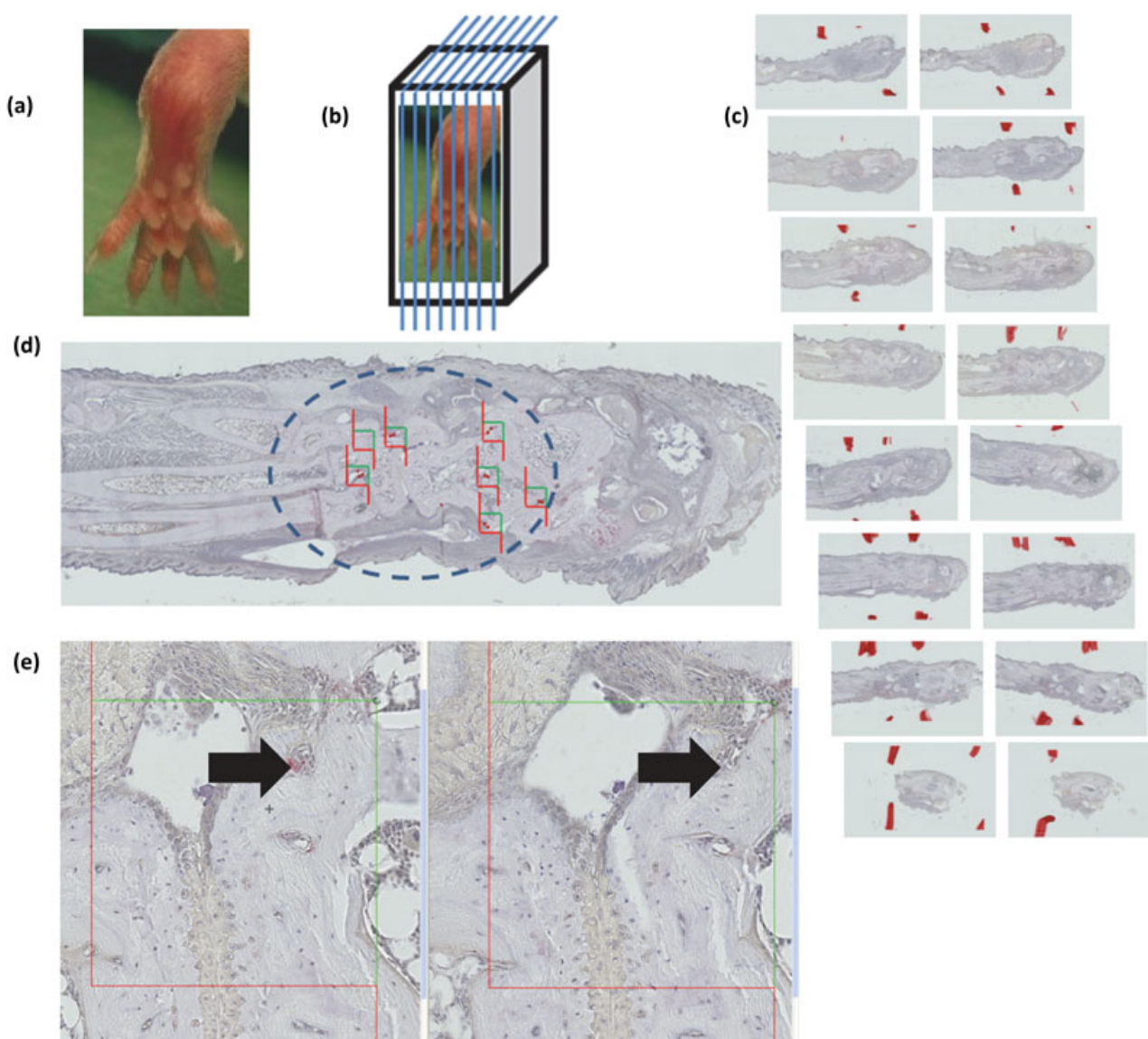
Fig. 1. Stereological design. Right hind paw was cut 0.5 cm above the ankle joint (a). The paws were embedded in methyl methacrylate and cut according to the principles of SURS with a random start and a fixed distance between section levels (b). Two sections 7 $\mu$m apart were generated at each section level (c). The region of interest was drawn with a dotted line according to the tarsus, and subsequently the FOVs were sampled using non-uniform sampling with a probability correlated to a weight determined by image analysis – the proportionator (d). The amount of the osteoclast colour indicates the weight of the field of view. Here, area with detected osteoclast colour is marked with red dots. Finally, osteoclasts were counted on the sampled FOVs using the physical disector (e). Black arrows indicate an osteoclast on the left picture and the corresponding spot on the right picture.

*Simulated SURS and proportionator sampling*

For each individual, a proportionator sampling and SURS were simulated using the original exhaustive data set comprising 100% of FOVs and covering the total section area. For the proportionator sampling, the transformed weights from all FOVs were arranged by the smooth fractionator principle (Gundersen, 2002), followed by systematic sampling, as described by Gardi *et al.* (2008a, b) (Fig. 1d and e). The SURS mimicked the meander sampling procedure used by the newCAST software, as described by Gardi *et al.* (2006).

The SURS sample covered 50% of the FOVs, which may seem quite large. However, the entire data set contained about 200 cells per animal. Thus, counting with an area fraction of 50% amounts to an average count of about 100 cells per specimen, which is commonly acknowledged as the number necessary to achieve a reasonable precision. The proportionator sample comprised 20% of the FOVs and was chosen as to obtain

approximately the same cell count as in SURS. Both samples consisted of two independent parts of half the sample size, with the purpose of estimating the CE (see below).

*Estimation of the number of cells*

Osteoclasts were counted in each of the $n$ sampled FOVs using a physical disector. We applied the disector in two directions, using the two sections both as reference and as lookup section. In order to estimate the total number of cells in the sampled section pairs from one individual ($Q^-$), the counted number of cells ($Q^-_{\mathrm{FOV}i}$) in a FOV was summed up and multiplied by the inverse area sampling fraction (*asf*) for SURS:

$$Q^- = \frac{1}{asf} \sum Q^-_{\mathrm{FOV}i}. \qquad (1)$$

In the proportionator sampling, the estimate is the weighted sum of the counted cells:

$$Q^- = \frac{W}{n} \cdot \sum \frac{Q^-_{\mathrm{FOV}i}}{w_i}, \qquad (2)$$

where $W$ is the sum of the weights of all FOVs, $w_i$ and $Q^-_{\mathrm{FOV}i}$ are the weight and cell count in a given FOV and $n$ is the number of FOVs. $\frac{W}{n}$ can be regarded as a systematic sampling period. We made two samples, and thus obtained two estimates $Q^-_1$ and $Q^-_2$; hence, the final estimate ($Q^-$) was obtained as the mean

$$Q^- = \frac{1}{2}(Q^-_1 + Q^-_2). \qquad (3)$$

In order to estimate the total number of osteoclasts ($N$) in the paw, we used the physical fractionator principle (Gundersen, 1986):

$$\hat{N} = \frac{1}{2} \cdot \frac{1}{ssf} \cdot Q^-, \qquad (4)$$

where *ssf* is the section sampling fraction. The result was divided by two because we counted both ways in the physical fractionator.

*Time estimation for SURS and the proportionator*

We estimated the time spent in the following four situations for both arthritic and control mice:

(1) *SURS using the autodisector on virtual slides.* The time was estimated as the sum of the time for preparation of the autodisector, the time for automatic sampling by the autodisector and the time for counting cells. Total counting time was estimated as the number of FOVs with a given cell count multiplied with the counting time for a FOV with this cell count. The results were summed for all given cell counts in a FOV from 0 to 10.

(2) *Proportionator using the autodisector on virtual slides.* Time was estimated as in (1).

(3) *SURS using the disector on sections.* Time was estimated as the sum of the time for preparation of the autodisector and counting cells. Time for counting was estimated as the number of FOVs with a given cell count multiplied with the sum of the counting time for a FOV with this cell count and the time for microscope stage movement. The results were summed for all given cell counts in a FOV from 0 to 10.

(4) *Proportionator using the disector on sections.* Time was estimated as in (3).

The preparation time for the autodisector was recorded during the 100% SURS. In this study, we used the same preparation time for the disector as for the autodisector, although the autodisector may be faster due to the automatic alignment of the virtual slides.

The time for microscope stage movement from position to position was obtained as the average time recorded approximately 40 times in total, on four different sections. Time was recorded from when a picture of the reference section was taken, until pictures of the lookup section were obtained and a virtual image was generated. In this study, the image acquired from the lookup section was expanded by 50% to allow for any small dis-alignment.

In approximately 200 FOVs located on 12 different sections, we recorded the time for evaluating a FOV and the number of cells counted in a given FOV. The average time for counting a given number of cells was calculated. At least five FOVs with a given number of cells were recorded before an average was estimated. Therefore, FOVs with six or more cells were extrapolated from the results of 0 to 5 cells in a FOV.

In order to investigate the time efficiency of the different evaluation and counting methods, the relative time efficiency ($R_{\mathrm{te}}$) was estimated:

$$R_{\mathrm{te}} = \frac{t_1}{t_2}, \qquad (5)$$

where $t_1$ and $t_2$ are the total time for evaluation using two different evaluation or counting methods. The $R_{\mathrm{te}}$ was calculated when comparing SURS and the proportionator as well as the autodisector and the disector. For each comparison (e.g. evaluation method), the other two parameters (e.g. group and counting method) were fixed in four different combinations.

*Coefficient of error*

Two sources of randomness influence the precision of the estimator $\hat{N}$: sampling of the sections and sampling of the FOVs within the sections; the latter is also known as 'cell counting noise'. Accordingly, the coefficient of error is composed of two parts, $CE_{\mathrm{sect}}$ and $CE_{\mathrm{noise}}$. Here, we describe the practical estimation of the total CE ($CE_{\mathrm{total}} = \sqrt{\mathrm{Var}\hat{N}}/N$). For the mathematical background, see the Appendix.

**Table 1.** The average of the number of FOVs, counted cells, total number of osteoclasts ($\hat{N}$) and CEs for the proportionator and SURS. The total CE ($CE_{total}$) is composed of $CE_{noise}$ and $CE_{sect}$.

| Group | Evaluation method | FOV | Count (cells) | N^ (cells) | $CE_{noise}$ (%) | $CE_{sect}$ (%) | $CE_{tot}$ (%) |
|---|---|---|---|---|---|---|---|
| Arthritis | SURS | 194 | 97 | 4823 | 15.6 | 2.5 | 15.8 |
| | Proportionator | 77 | 92 | 4707 | 13.4 | 2.3 | 13.6 |
| Control | SURS | 154 | 14 | 862 | 18.3 | 7.2 | 19.7 |
| | Proportionator | 62 | 19 | 657 | 20.8 | 6.3 | 21.7 |

**Table 2.** The average time consumption for the different groups, counting and evaluation methods. Time consumption is divided into preparation, automatic sampling and counting.

| Group | Counting method | Evaluation method | Preparation (min) | Automatic sampling (min) | Counting (min) | Total time (min) |
|---|---|---|---|---|---|---|
| Arthritis | Autodisector | SURS | 15:47 | 16:23 | 27:26 | 59:36 |
| | | Proportionator | 15:47 | 6:33 | 17:54 | 40:15 |
| | Disector | SURS | 15:47 | 00:00 | 99:19 | 115:06 |
| | | Proportionator | 15:47 | 00:00 | 46:48 | 62:35 |
| Control | Autodisector | SURS | 14:31 | 13:32 | 13:25 | 41:27 |
| | | Proportionator | 14:31 | 5:25 | 7:08 | 27:03 |
| | Disector | SURS | 14:31 | 00:00 | 70:12 | 84:43 |
| | | Proportionator | 14:31 | 00:00 | 29:58 | 44:29 |

To obtain $CE_{noise}$, we applied the so-called *direct CE* method described by Gardi *et al.* (2008a), which is simply the empirical CE of two counts $Q_1^-$ and $Q_2^-$ from two independent samples of half the wanted size, i.e.

$$CE_{noise} = \frac{\sqrt{Var_{noise}}}{Q^-}, \qquad (6)$$

where

$$Var_{noise} = \frac{1}{2} \cdot est.Var Q \qquad (7)$$

is half the empirical variance of the two values $Q_1^-$ and $Q_2^-$, which can be calculated as est. $Var Q = \frac{1}{2}(Q_1^- - Q_2^-)^2$. The other part of the total CE ($CE_{sect}$) explains how the tissue varies from section to section (Gundersen *et al.*, 1999). It is estimated as

$$CE_{sect} = \frac{\sqrt{(3(A - Var_{noise}) - 4B + C)/240}}{Q^-}, \qquad (8)$$

where $A = \sum Q_{secti}^- \cdot Q_{secti}^-$, $B = \sum Q_{secti}^- \cdot Q_{secti+1}^-$, $C = \sum Q_{secti}^- \cdot Q_{secti+2}^-$ and $Q_{secti}^-$ denotes the estimated number of cells in section pair *i*. The total CE was calculated as

$$CE_{total} = \sqrt{CE_{sect}^2 + CE_{noise}^2}. \qquad (9)$$

Finally, the average CE for all *m* individuals was calculated as follows:

$$\overline{CE} = \sqrt{\frac{CE_1^2 + CE_2^2 \cdots + CE_m^2}{m}}. \qquad (10)$$

**Results**

*Characterization of data*

The study was designed to obtain almost identical cell counts for the simulated SURS and proportionator sampling (Table 1).

The total coefficient of error ($CE_{total}$) in this study was composed mainly of the $CE_{noise}$ (Table 1). The average CE for SURS (16%) was approximately the same as for the proportionator (14%) in arthritic mice. Similar results were found for the control group (Table 1). However, in both groups, the results varied among individuals, and in some cases, the $CE_{total}$ was larger for the proportionator than for SURS.

*Time efficiency*

We investigated the theoretical time consumption for the different evaluation and counting methods in arthritic and control mice. The results are shown in Table 2.

The proportionator was 48% to 90% more time efficient than SURS depending on group and counting method (Table 3). SURS versus the proportionator using the disector was superior to other comparisons (Table 3). The average time difference between the proportionator and SURS was explained by a time gain in both the automatic sampling and the following cell counting for the autodisector, and a time gain in the cell counting for the disector (Table 2).

Knowing that the proportionator improved time efficiency, we compared the autodisector with the disector in order to evaluate the difference in time consumption. We found that

**Table 3.** The relative time efficiency ($R_{te}$) for comparisons of the different counting and evaluation methods. During the comparisons, the two other variables were fixed in four different combinations for each comparison. $R_{te}$ is the total time for evaluating one parameter (e.g. SURS) divided with the other (e.g. proportionator).

| Fixed variables | Comparison | $R_{te}$ |
|---|---|---|
| Arthritis, Autodisector | SURS vs. Proportionator | 1.48 |
| Arthritis, Disector | SURS vs. Proportionator | 1.84 |
| Control, Autodisector | SURS vs. Proportionator | 1.53 |
| Control, Disector | SURS vs. Proportionator | 1.90 |
| Arthritis, SURS | Disector vs. Autodisector | 1.93 |
| Arthritis, Proportionator | Disector vs. Autodisector | 1.55 |
| Control, SURS | Disector vs. Autodisector | 2.04 |
| Control, Proportionator | Disector vs. Autodisector | 1.64 |

the autodisector was 55% to 104% faster than the disector (Table 3).

*Relation between counts per FOV and number of FOVs*

In this study, the number of evaluated FOVs was much larger in SURS than in the proportionator sampling for both arthritis and control mice. The difference was almost entirely explained by a high number of FOVs with 0 counts (Fig. 2a and c). In order to directly compare the two sampling methods, we adjusted the number of frames using 50% SURS to the 20% proportionator sampling (multiplied with 0.4). We found that there were still more frames with 0 counts using SURS. However, for all other counts, there was a larger number of FOVs using the proportionator (Fig. 2b and d).

*Relation between counts per FOV and weight*

The association between the counts per FOV and the weight was not ideal in this study. In many FOVs with high weight, no cells were counted. On the other hand, Figure 3 demonstrates that some of the FOVs with the lowest weight 1 contained one
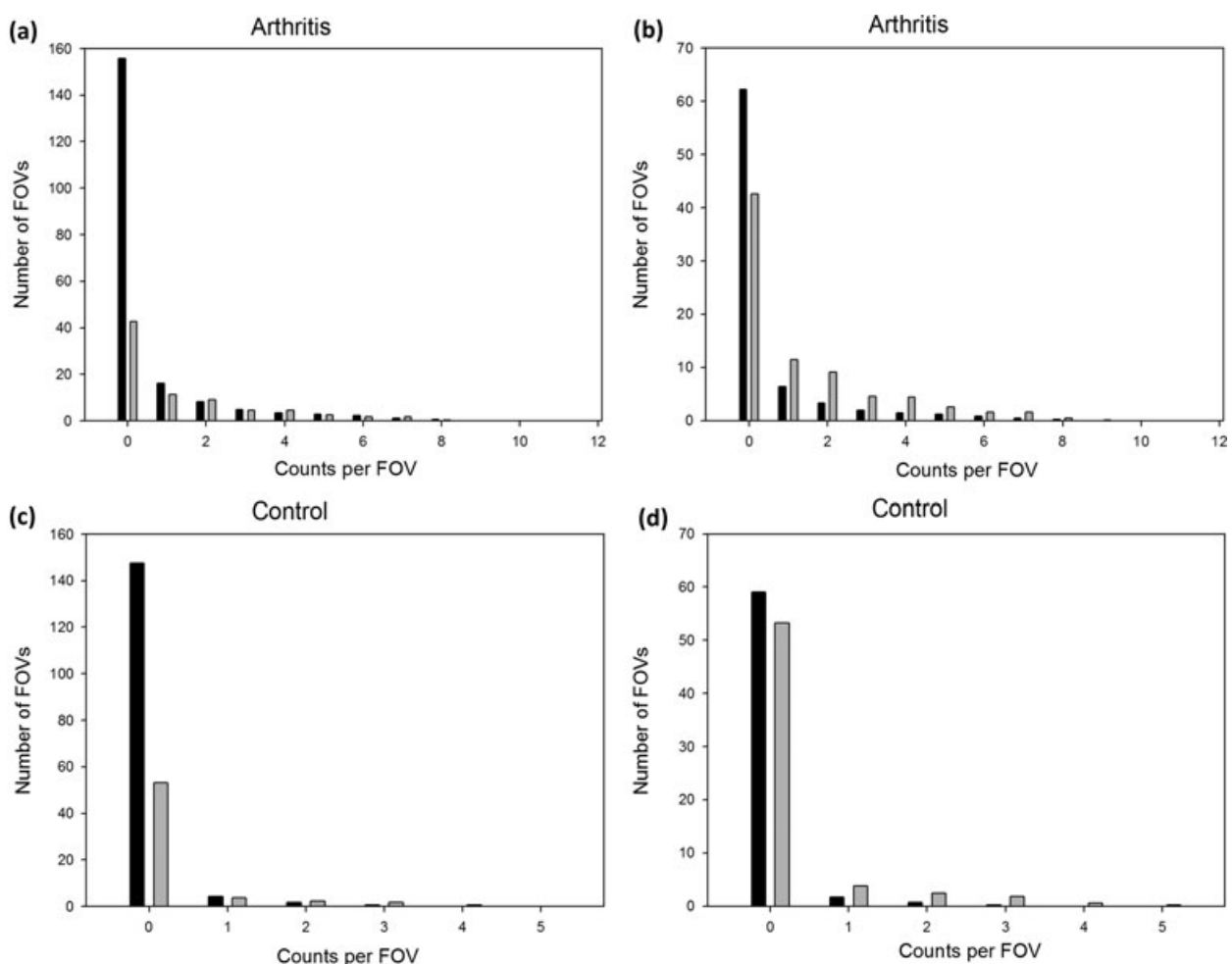


**Fig. 2.** Association between the counts per FOV and the number of FOVs for SURS and the proportionator in arthritic and control mice. The counts per FOV were compared to the number of FOVs (a) and (c). The number of FOVs using the 50% SURS was adjusted to the number of FOVs using the 20% proportionator (multiplied with 0.4) (b) and (d). Black boxes illustrate SURS and grey boxes illustrate the proportionator.
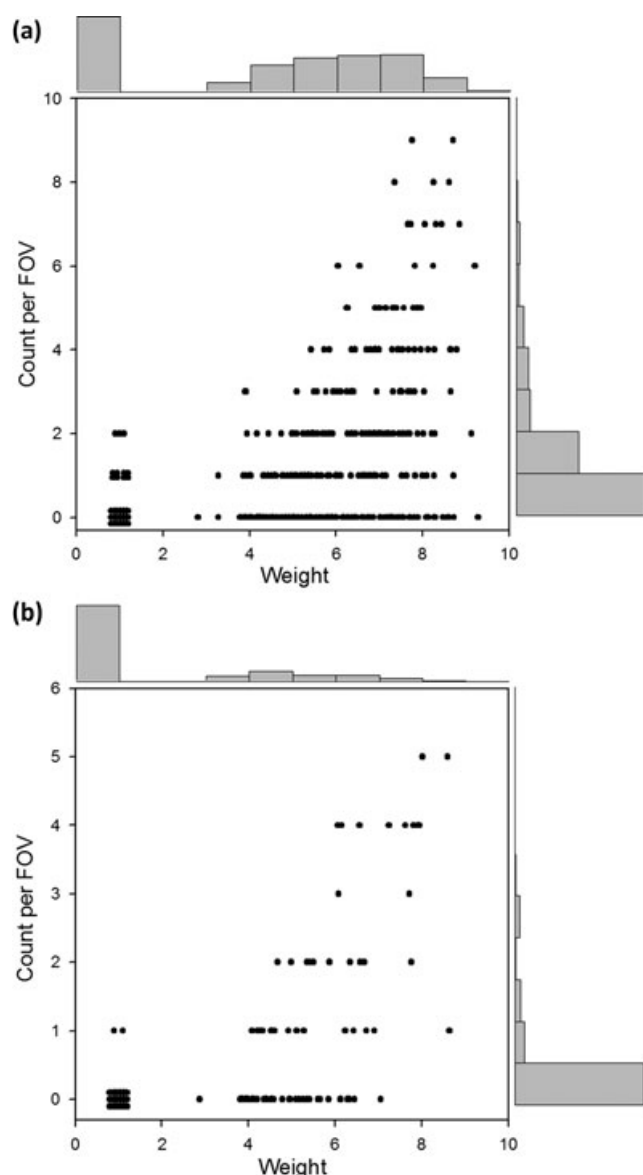
**Fig. 3.** Association between the weight and the counts per FOV. Some FOVs with a high weight achieved no counts and some FOVs with a weight of 1 achieved one or two counts in arthritic mice (eight and three FOVs) and two FOVs in normal mice achieved one count. The results for a weight of 1 are jittered. Marginal relative distributions for each axis are shown due to many overlapping dots.

or two cells in arthritic mice (eight and three FOVs, respectively) and one cell in control mice (two FOVs). As the original weights have been transformed, a weight of 1 corresponds to a raw weight of 0. This means that the corresponding cells have not been detected by the image analysis procedure.

**Discussion**

In this study, we compared the proportionator with SURS using number estimation of osteoclasts in arthritic joints. In our

setup, we found that the proportionator was 48% to 90% more time efficient than SURS. Hence, the proportionator was efficient in both arthritic and control mice. Previously, Gardi *et al.* (2008a) demonstrated that the proportionator was approximately 100% more time efficient in some tissues (pancreas tissue, pancreas $\beta$ cells and cerebellar granule cells), but only approximately 50% as time efficient in another tissue (GFP orexin neurons).

Although the proportionator was more time efficient, we also encountered challenges. The association between the weight and the count was poor. Firstly, we found that some FOVs with a weight of 1 yielded one or two counts. A weight of 1 as used in sampling corresponds to a weight of 0 of the original image analysis. Hence, some cells were counted in FOVs where no colour was detected. In this study setup, some of the cell counts in FOVs with a weight of 1 can probably be explained by the fact that we counted both ways, but the weight was only recorded in the first section. However, the major reason for these cell counts is probably that the image protocol was not sensitive enough. This does not lead to a biased estimate, but may result in a high CE (or, equivalently, high variance), a problem common in sampling with unequal probability and well known to statistics, as drastically illustrated by Basu (1971) with his famous circus elephant example. Therefore, the pilot study to calibrate the weights should ensure that the protocol is 'over sensitive' thereby making sure that all countable areas are classified. Consequently, efficiency will be reduced as more empty areas will be presented to the user, but the CE will be much lower. Secondly, we also observed that a lot of FOVs with a high weight yielded none or few counts. In some cases, a large weight and no count is explained by large cells that are less likely to be counted. Bridges, defined as cells that appear as one cell profile on the reference section but as two cell profiles on the lookup section (Gundersen *et al.*, 1993), are also a challenge for the proportionator, because they contribute to the weight. We did not account for bridges although they amount to 4–5% of the total number of osteoclasts (Keller *et al.*, 2011). How bridges can be accounted for in proportionator counting needs further investigation.

CE estimation for the proportionator is difficult, due to the systematic sampling that is part of the principles of the smooth fractionator. We applied the direct CE of two individual samplings (Gardi *et al.*, 2008b) on the proportionator sample as well as the SURS sample in order to compare the two methods under the same conditions. This estimate is design-unbiased, but can have a skewed distribution, which will lead to underestimation in most cases. The skewness may be reduced by more independent samples, but the true CE of the estimator will be increased. Consequently, the development of reliable estimators for the CE remains a challenging problem for future research.

Time efficiency of the autodisector on virtual slides was also compared with the disector on tissue slides. We found that the autodisector on virtual slides was 55% to 104% faster than the

disector on tissue slides. Consequently, the autodisector was superior to the disector in both arthritic and control mice. The time spent on scanning slides in the digital slide scanner was not accounted for in this study, and therefore the real time gain will be somewhat smaller. However, the scanning process is not very laborious because more than 200 tissue slides can be automatically scanned in one session. One could also argue that the automatic sampling should not be accounted for, because this process is entirely automatic, and thereby the total time for the autodisector would actually be decreased. In conclusion, the net time gain is substantial and we believe that the autodisector on virtual slides will be able to save time during stereological cell counting in the future.

In this study, we applied the autodisector on virtual slides, but the autodisector can also be applied directly on tissue slides on a standard microscope equipped for stereology. However, efficiency will be lower in this process primarily because the image generation is slower. In addition, time must be spent on taking super images of the slides and changing slides in the slide loader.

The use of virtual slides could also be implemented for stereology on volume, length and surfaces. However, the time gained would probably be less because the stage movement time would be almost identical in the two setups. In our study, time was primarily gained by the fast and automatic generation of the disector pairs. When using tissue slides for the estimation of volume, length and surfaces, time must be spent on taking super images of the slides and changing slides in the slide loader.

When using virtual slides, a concern is whether the quality of the virtual slides is as good as the original tissue slides, and especially whether the plane of focus is acceptable. In general, we found the quality of the virtual slides to be quite acceptable for evaluation in a stereological setup.

There are limitations in the design of this study. Although the time estimation for the proportionator and SURS was based on a real 100% SURS sample, it was, however, simulated.

In summary, we demonstrated that the proportionator was more time efficient than the traditional SURS. Secondarily, we showed that cell counting on virtual slides using the autodisector is time efficient compared with traditional cell counting on tissue slides using the disector. Hence, the autodisector on virtual slides and the proportionator may be important for stereological cell counting in the future.

## Acknowledgements

## References

Basu, D. (1971) An essay on the logical foundations of survey sampling, part 1 (with discussion). *Foundations of Statistical Inference* (ed. by V.P. Godambe & D.A. Sprott), pp. 203–242. Holt, Reinhart and Winston, Toronto.

Gardi, J.E., Nyengaard, J.R. & Gundersen, H.J. (2006) Using biased image analysis for improving unbiased stereological number estimation – a pilot simulation study of the smooth fractionator. *J. Microsc.* **222**, 242–250.

Gardi, J.E., Nyengaard, J.R. & Gundersen, H.J. (2008a) Automatic sampling for unbiased and efficient stereological estimation using the proportionator in biological studies. *J. Microsc.* **230**, 108–120.

Gardi, J.E., Nyengaard, J.R. & Gundersen, H.J. (2008b) The proportionator: unbiased stereological estimation using biased automatic image analysis and non-uniform probability proportional to size sampling. *Comput. Biol. Med.* **38**, 313–328.

Gonzales, R.C. & Woods, R.E. (2008) *Digital Image Processing.* Prentice Hall, New Jersey, USA.

Gravallese, E.M., Harada, Y., Wang, J.T., Gorn, A.H., Thornhill, T.S. & Goldring, S.R. (1998) Identification of cell types responsible for bone resorption in rheumatoid arthritis and juvenile rheumatoid arthritis. *Am. J. Pathol.* **152**, 943–951.

Gundersen, H.J. (1986) Stereology of arbitrary particles. A review of unbiased number and size estimators and the presentation of some new ones, in memory of William R. Thompson. *J. Microsc.* **143**, 3–45.

Gundersen, H.J. (2002) The smooth fractionator. *J. Microsc.* **207**, 191–210.

Gundersen, H.J., Boyce, R.W., Nyengaard, J.R. & Odgaard, A. (1993) The Conneulor: unbiased estimation of connectivity using physical disectors under projection. *Bone* **14**, 217–222.

Gundersen, H.J. & Jensen, E.B. (1987) The efficiency of systematic sampling in stereology and its prediction. *J. Microsc.* **147**, 229–263.

Gundersen, H.J., Jensen, E.B., Kieu, K. & Nielsen, J. (1999) The efficiency of systematic sampling in stereology–reconsidered. *J. Microsc.* **193**, 199–211.

Hashimoto, M., Hirota, K., Yoshitomi, H. et al. (2010) Complement drives Th17 cell differentiation and triggers autoimmune arthritis. *J. Exp. Med.* **207**, 1135–1143.

Horvitz, D.G. & Thompson, D.J. (1952) A generalization of sampling without replacement from a finite universe. *J. Am. Statist. Assoc.* **47**, 663–685.

Jara-Lazaro, A.R., Thamboo, T.P., Teh, M. & Tan, P.H. (2010) Digital pathology: exploring its applications in diagnostic surgical pathology practice. *Pathology* **42**, 512–518.

Keller, K.K., Stengaard-Pedersen, K., Dagnaes-Hansen, F., Nyengaard, J.R., Sakaguchi, S. & Hauge, E.M. (2011) Histological changes in

chronic autoimmune SKG-arthritis evaluated by quantitative three-dimensional stereological estimators. *Clin. Exp. Rheumatol.* **29**, 536–543.

Keller, K.K., Thomsen, J.S., Stengaard-Pedersen, K., Dagnaes-Hansen, F., Nyengaard, J.R. & Hauge, E.M. (2012) Bone formation and resorption are both increased in experimental autoimmune arthritis. *PLoS One* **7**, e53034.

Kobayashi, K., Suda, T., Nan-Ya, K., Sakaguchi, N., Sakaguchi, S. & Miki, I. (2006) Cytokine production profile of splenocytes derived from zymosan A-treated SKG mice developing arthritis. *Inflamm. Res.* **55**, 335–341.

Madow, W. (1949) On the theory of systematic sampling, II. *Ann. Math. Stat.* **20**, 333–354.

Murthy, M.N. (1967) *Sampling: Theory and Methods.* Statistical Publishing Society, Calcutta, India.

Sakaguchi, N., Takahashi, T., Hata, H. et al. (2003) Altered thymic T-cell selection due to a mutation of the ZAP-70 gene causes autoimmune arthritis in mice. *Nature* **426**, 454–460.

Sterio, D.C. (1984) The unbiased estimation of number and sizes of arbitrary particles using the disector. *J. Microsc.* **134**, 127–136.

Taylor, C.R. (2011) From microscopy to whole slide digital images: a century and a half of image analysis. *Appl. Immunohistochem. Mol. Morphol.* **19**, 491–493.

## Appendix

In the Material and Methods section, we showed a formula for the CE of $\hat{N}$, the estimator of the total number $N$ of cells. This appendix exposes the mathematical background.

Three randomization steps contribute to the variance of $\hat{N}$, namely:

(1) the randomization of the start section $U$ in systematic sampling of the sections,

(2) the uniform selection of random origins $O$ of the grid of the FOVs on the sections and

(3) the sample $S$ of FOVs given the grid.

For the last step, we chose either the traditional spatially systematic uniform meander design (see Gardi *et al.*, 2006), or the proportionator.

Following the standard variance decomposition when conditioning on $U$, the squared coefficient of error of $\hat{N} = Q^-/(2ssf)$ can be written as $CE^2_{\text{total}}\hat{N} = CE^2_{\text{sect}}\hat{N} + CE^2_{\text{noise}}\hat{N}$, where

$$CE^2_{\text{sect}}\hat{N} = \frac{\text{Var}\,E(\hat{N}|U)}{E^2\hat{N}} = \frac{\text{Var}\,E(Q^-|U)}{E^2Q^-} \qquad (A1)$$

can be estimated by $CE^2_{\text{sect}}$ given in Eq. (8), see Gundersen *et al.* (1999). The $CE^2$ due to randomization and sampling of the FOVs,

$$CE^2_{\text{noise}}\hat{N} = \frac{E\,\text{Var}(\hat{N}|U)}{E^2\hat{N}} = \frac{E\,\text{Var}(Q^-|U)}{E^2Q^-}, \qquad (A2)$$

is estimated using empirical variance and mean of two realizations of $Q^-$ using the same set of sections, by

$$CE^2_{\text{noise}} = \frac{\frac{1}{2}\text{est.Var}\,Q}{\left(\frac{1}{2}(Q_1^- + Q_2^-)\right)^2} = \frac{(Q_1^- - Q_2^-)^2}{(Q_1^- + Q_2^-)^2}, \qquad (A3)$$

see Equations (6) and (7). The empirical variance est.Var $Q$ is a design-unbiased (though not very precise) estimator for $\text{Var}(Q^-|U)$, even if the same set of FOV grid origins $O$ on the sections is used for both $Q_1^-$ and $Q_2^-$: Since the sections are entirely partitioned into FOVs, the conditional mean $E(Q^-|U, O)$ is independent of the grid origins $O$; thus, $\text{Var}_O\,E_S(Q^-|U, O) = 0$ and therefore $\text{Var}(Q^-|U) = E_O\,\text{Var}_S(Q^-|U, O)$.

# Paper C

# 2D non-uniform systematic sampling

By I. T. Andersen & J. F. Ziegel

# 2D non-uniform systematic sampling

Ina Trolle Andersen[1,2] and Johanna F. Ziegel[3]

[1]Department of Mathematics, Aarhus University, ita@math.au.dk
[2]Stereological Research Laboratory, Department of Clinical Medicine, Aarhus University
[3]Institute of Mathematical Statistics and Actuarial Science, University of Bern,
johanna.ziegel@stat.unibe.ch

### Abstract

In this paper, we introduce a natural generalization of non-uniform systematic sampling to two dimensions. Under optimal auxiliary information about the function of interest, this design yields an estimator with zero variance. In a simulation study imitating sampling situations in microscopy, the 2D non-uniform systematic designs show similar efficiency as proportional-to-size sampling with replacement. An exception is area estimation where the 2D non-uniform systematic designs are superiour in a number of cases considered.

*Keywords:* Cavalieri estimator, efficiency, Horvitz-Thompson estimator, microscopy, systematic sampling, 2D sampling

## 1 Introduction

In this paper, we introduce a new 2D non-uniform systematic sampling design that respects the spatial information available and study its efficiency. In designs like the proportionator (Gardi et al., 2008a,b), which is used in microscopy, all spatial information is lost, prior to sampling. Therefore, a natural idea to improve upon such designs, is to include spatial information in the sampling procedure. We propose to use transformations of 2D uniform systematic sampling into non-uniform sampling, while still maintaining some of the spatially balanced arrangement. Unlike Grafström and Tillé (2012) and Stevens Jr and Olsen (2004), which also try to balance non-uniform samples spatially, our proposal is a genuine 2D sampling procedure.

The suggested 2D non-uniform systematic sampling design is a generalization of Dorph-Petersen et al. (2000), where non-uniform systematic sampling was introduced in 1D. In that paper it was concluded from simulations, that non-uniform sampling was more efficient than traditional uniform sampling, known as the classical 2D Cavalieri estimator, in an example of area estimation from lengths of linear intercepts. In the present paper we propose a 2D non-uniform systematic sampling design, which under optimal auxiliary information about the function of interest, yields zero variance of the estimator. Furthermore, a simulation study resembling sampling in microscopy, has been performed to investigate the applicability of the method.

The paper is organised as follows. First we recall results from Dorph-Petersen et al. (2000) for the 1D sampling procedure. Then, we derive the generalization to higher dimensions and propose various transformations of 2D uniform systematic sampling. Subsequently, methods and results of the simulation study are presented, followed by a discussion. Technical proofs are deferred to two appendices.

## 2   Generalized systematic sampling in 1D

### 2.1   Theoretical considerations

Let $f$ be a bounded non-negative function with bounded support, assumed to be the interval $[0,1]$ without loss of generality. The objective is to estimate the integral

$$Q = \int_0^1 f(x)\,\mathrm{d}x, \tag{2.1}$$

using values of $f$ at a set of random sampling points. In uniform systematic sampling (Cruz-Orive, 1993; Gundersen et al., 1999), a random systematic set of $n$ points is selected, $Y_i = (U+i)/n$, where $U \sim \mathrm{unif}[0,1)$ and $i = 0,1\ldots,n-1$, from which $f$ is estimated by a simple step function. Thus, $Q$ can be estimated unbiasedly by

$$\hat{Q}_n = \frac{1}{n}\sum_{i=0}^{n-1} f(Y_i). \tag{2.2}$$

Instead of equidistant sampling points, it may be more efficient to choose points closer to each other in some areas, while keeping a spatially spread sample. This can be obtained by considering an increasing bijective function $G : [0,1] \to [0,1]$, which transforms the sampling points into new points $X_i = G^{-1}(Y_i), i = 0,\ldots,n-1$. Using that

$$Q = \int_0^1 f(x)\,\mathrm{d}x = \int_0^1 f(G^{-1}(x))\frac{1}{G'(G^{-1}(x))}\,\mathrm{d}x, \tag{2.3}$$

we obtain a new unbiased estimator

$$\hat{Q}_n = \frac{1}{n}\sum_{i=0}^{n-1} \frac{f(X_i)}{G'(X_i)} = \frac{1}{n}\sum_{i=0}^{n-1} \frac{f(G^{-1}(Y_i))}{G'(G^{-1}(Y_i))},$$

which corresponds to the uniform systematic sampling estimator of the function

$$\tilde{f}(y) = f(G^{-1}(y))/G'(G^{-1}(y)).$$

### 2.2   Choice of sampling points

Finding efficient sampling points, hence $G$ and thereby $\tilde{f}$, is considered in Dorph-Petersen et al. (2000), where three choices are investigated. An obvious possibility is to choose $G$ such that $\tilde{f}$ is constant. Then $\hat{Q}_n$ is constant and always yields the true value $Q$. Under the assumption that $f > 0$, the function $G$ can be defined by

$$G(x) = \frac{\int_0^x f(t)\,\mathrm{d}t}{\int_0^1 f(t)\,\mathrm{d}t}. \tag{2.4}$$

In Dorph-Petersen et al. (2000), the relation between this choice and sampling proportional-to-size in the discrete setting, is established. Another suggestion uses that the transformation corresponds to uniform systematic sampling for $\tilde{f}$, hence asymptotic results in the uniform case can be used. Transitive methods yield that the asymptotic variance depends on the smoothness of $\tilde{f}$, hence it is preferable to have $\tilde{f}$ smoother than $f$. The third suggestion is based on the idea of sampling the most, where $f$ varies the most.

The three possibilities considered in the paper each depend on the function $f$, thus prior knowledge of $f$ or properties of $f$ are needed. In practice $f$ is unknown, therefore a function $f_0$ similar to $f$ is used instead. For instance using this in (2.4), the estimator $\hat{Q}_n$ becomes

$$\hat{Q}_n = \int_0^1 f_0(t)\,\mathrm{d}t \cdot \frac{1}{n}\sum_{i=0}^{n-1} \frac{f(X_i)}{f_0(X_i)}, \quad X_i = G^{-1}((U+i)\tfrac{1}{n}). \tag{2.5}$$

## 3 Generalized systematic sampling in 2D

### 3.1 Theoretical considerations

The idea is now to generalize this to higher dimensions. For simplicity, we restrict our considerations to two dimensions as the theoretical considerations are directly transferable to any dimension. Let $f$ be a bounded non-negative function with bounded support $[0,1]^2$. The objective is to estimate the integral

$$Q = \int_{[0,1]^2} f(x,y)\,\mathrm{d}x\,\mathrm{d}y, \tag{3.1}$$

using values of $f$ at a set of random sampling points in $[0,1]^2$. In uniform systematic sampling, a random systematic set of $n \times m$ points is selected,

$$(Y_{1i}, Y_{2j}) = ((U_1+i)\tfrac{1}{n}, (U_2+j)\tfrac{1}{m}),$$

where $U_1$ and $U_2$ are independent and $U_1, U_2 \sim \mathrm{unif}[0,1)$, $i = 0, 1\ldots, n-1$ and $j = 0, 1\ldots, m-1$. From these points $f$ is estimated by a simple step function, hence $Q$ can be estimated unbiasedly by

$$\hat{Q}_{nm} = \frac{1}{nm}\sum_{i=0}^{n-1}\sum_{j=0}^{m-1} f(Y_{1i}, Y_{2j}).$$

Similar to the 1D case, it might be better to choose points closer to each other in some areas. This can be obtained by considering a diffeomorphism $G : [0,1]^2 \rightarrow [0,1]^2$, which transforms the sampling points into new points $(X_{1i}, X_{2j}) = G^{-1}(Y_{1i}, Y_{2j})$, $i = 0, \ldots, n-1$, $j = 0, \ldots, m-1$. Using that

$$Q = \int_{[0,1]^2} f(x,y)\,\mathrm{d}x\,\mathrm{d}y$$
$$= \int_{[0,1]^2} f(G^{-1}(x,y))\frac{1}{|\det(G'(G^{-1}(x,y)))|}\,\mathrm{d}x\,\mathrm{d}y,$$

where $G'$ denotes the Jacobi matrix of $G$, we obtain a new unbiased estimator

$$\hat{Q}_{nm} = \frac{1}{nm} \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} \frac{f(X_{1i}, X_{2j})}{|\det(G'(X_{1i}, X_{2j}))|}$$

$$= \frac{1}{nm} \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} \frac{f(G^{-1}(Y_{1i}, Y_{2j}))}{|\det(G'(G^{-1}(Y_{1i}, Y_{2j})))|}$$

which, again, corresponds to the uniform systematic sampling estimator of the function

$$\tilde{f}(x,y) = f(G^{-1}(x,y))/|\det(G'(G^{-1}(x,y)))|.$$

In the Appendix, we proof that $\hat{Q}_{nm}$ is in fact unbiased.

## 3.2   Choice of sampling points

Analogously to 1D we would like to choose the sampling points in an efficient manner. In particular, we can determine $G$ such that $\tilde{f}$ is constant, which implies that the variance of $\hat{Q}_{nm}$ becomes zero. Assuming $f(x,y) > 0$, $\tilde{f}$ is constant if we choose any $G$ with

$$|\det(G'(u,v))| = cf(u,v), \tag{3.2}$$

for some constant $c > 0$. There are many possible choices of diffeomorphisms with property (3.2), one example is $G = (G_1, G_2)$, where

$$G_1(x,y) = \int_0^x g(u) \, du, \quad G_2(x,y) = \frac{1}{g(x)\Delta} \int_0^y f(x,v) \, dv, \tag{3.3}$$

and

$$g(x) = \frac{1}{\Delta} \int_0^1 f(x,v) \, dv, \quad \Delta = \int_{[0,1]^2} f(u,v) \, du \, dv. \tag{3.4}$$

The Jacobi matrix is given by

$$\begin{pmatrix} \frac{\partial G_1}{\partial x} & \frac{\partial G_1}{\partial y} \\ \frac{\partial G_2}{\partial x} & \frac{\partial G_2}{\partial y} \end{pmatrix} = \begin{pmatrix} g(x) & 0 \\ \dots & \frac{f(x,y)}{g(x)\Delta} \end{pmatrix},$$

hence $|\det(G'(u,v))| = f(u,v)/\Delta$. Replacing $f$ with a known function $f_0$, which is similar to $f$, will be used in practice. This could be colour values obtained by image analysis. Like in 1D this choice of $G$ is related to sampling proportional-to-size in the discrete set-up. In the Appendix, this relation is established. Replacing $f$ with $f_0$ in (3.3) and (3.4), the estimator becomes

$$\hat{Q}_{nm} = \int_{[0,1]^2} f_0(u,v) \, du \, dv \frac{1}{nm} \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} \frac{f(X_{1i}, X_{2j})}{f_0(X_{1i}, X_{2j})}, \tag{3.5}$$

$$(X_{1i}, X_{2j}) = G^{-1}((U_1 + i)\tfrac{1}{n}, (U_2 + j)\tfrac{1}{m}).$$

Figure 1 illustrates a 2D uniform systematic sample, which is transformed into a non-uniform systematic sample, proportional to the colour values. Note that the non-uniform sample is still spatially spread. However, the statistical behaviour of the estimator $\hat{Q}_{nm}$
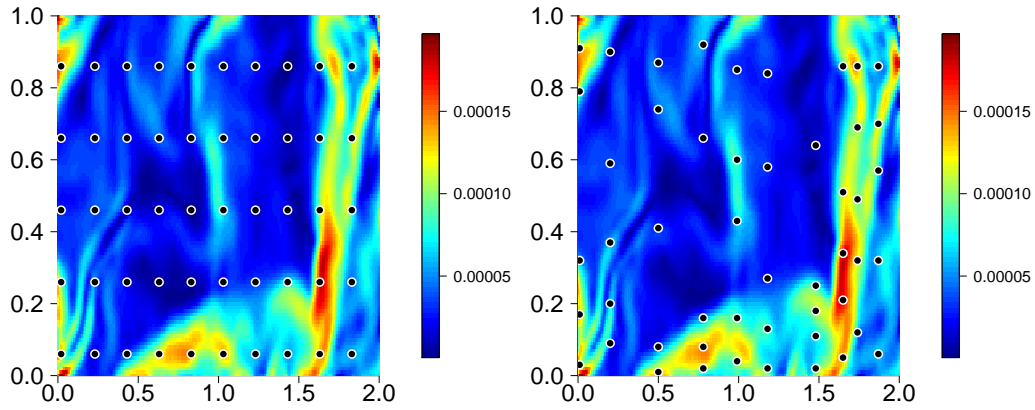
**Figure 1:** Left: 2D uniform systematic sampling. Right: 2D non-uniform systematic sampling, where the points have been transformed proportional to the colour values using (3.3) and (3.4). The colours are obtained from the *bei*-data in the R-package Spatstat, see Baddeley and Turner (2005).

depends on which coordinate in (3.3) and (3.4) is chosen first. If we interpret $\Phi$ as the result of a gravitational force acting on the systematic grid of points, $G$ has to be given as $G(x, y) = \nabla \Phi(x, y)$ for some scalar potential $\Phi$. Hence to fulfil property (3.2), we seek a solution of the Monge-Amper equation

$$\frac{\partial^2 \Phi}{\partial x^2} \frac{\partial^2 \Phi}{\partial y^2} - \frac{\partial^2 \Phi}{\partial x \partial y} = c f(x, y)$$

see e.g. Kołodziej (1998), subject to some suitable boundary and convexity conditions. Unfortunately this equation is known to be very difficult to solve even numerically.

Instead, keeping the ideas behind $G$ in (3.3) and (3.4), it is possible to decrease the strong dependence on the order of the coordinates by working with the composition $G = \Psi \circ \Phi$ of two transformations, where $\Phi : [0, 1]^2 \to [0, 1]^2$ is given by

$$\Phi_1(x, y) = \int_0^x g(u) \, \mathrm{d}u,$$

$$\Phi_2(x, y) = \frac{1}{g(x) \Delta_g} \int_0^y f(x, v)^\alpha \, \mathrm{d}v,$$

where $\alpha \in (0, 1)$,

$$g(x) = \frac{1}{\Delta_g} \int_0^1 f(x, v)^\alpha \, \mathrm{d}v, \quad \Delta_g = \int_{[0,1]^2} f(u, v)^\alpha \, \mathrm{d}u \, \mathrm{d}v.$$

and $\Psi : [0, 1]^2 \to [0, 1]^2$ is given by

$$\Psi_1(x, y) = \frac{1}{h(y) \Delta_h} \int_0^x (f \circ \Phi^{-1})(u, y)^{1-\alpha} \, \mathrm{d}u,$$

$$\Psi_2(x, y) = \int_0^y h(v) \, \mathrm{d}v,$$

where

$$h(y) = \frac{1}{\Delta_h} \int_0^1 (f \circ \Phi^{-1})(u,y)^{1-\alpha} \, \mathrm{d}u,$$

$$\Delta_h = \int_{[0,1]^2} (f \circ \Phi^{-1})(u,v)^{1-\alpha} \, \mathrm{d}u \, \mathrm{d}v.$$

It is straight forward to show, that this choice of $G$ fulfils (3.2), and we expect that the resulting estimator is more efficient. Unlike $G$ in (3.3) and (3.4), where the transformed points are aligned in one direction, this composition of $G$ results in a completely deformed grid of points, which is not effected much by the order of coordinates. In practice, it may be difficult to work with a transformation of the form $G = \Psi \circ \Phi$ as inverting $G$ can be very hard even numerically.

## 4   Simulation study

### 4.1   Motivation

To support our theoretical findings and to investigate the efficiency of the new design compared to existing designs, a simulation study was carried out. As the project was motivated by sampling in microscopy, it is natural to construct set-ups, which resemble data obtained from this field of research. Here, non-uniform sampling is mainly used for counting purposes, for instance for determining the total number of cells in a cell population, but also area estimation is considered, see e.g. Gardi et al. (2008a,b). Both number estimation and area estimation will be studied in the present paper, as well as estimation of the integral of a function. The latter case will be called integral estimation and involves investigation of more smooth functions than the measurement functions in the first two cases, which are based on indicator functions. In microscopy, the value of the measurement function in a point corresponds to complete information from a small observation window, located at this point.

We use a selection of stochastic point processes to generate the centres of the imaginary cells for number estimation, and thereby a selection of measurement functions $f$. Various choices of the function $f_0$, which controls the transformation $G$ of the sampling points, is investigated in all three cases of estimation. The function $f_0$ is constructed from $f$ with different levels of noise including spatial errors. In this set-up, $f_0$ corresponds to known auxiliary information, obtainable from automatic image analysis of a tissue section, for instance the amount of a predetermined color identifying a staining of the cells.

### 4.2   Sampling approximations

The inverse of some non-standard integrals is needed to obtain the samples, but as this is quite time consuming, the integrals are approximated by sums, and only the non-decomposed choice of $G$ given in (3.3) and (3.4) is considered. Furthermore, $f_0$ is only determined in a finite number of points, corresponding to a pixelation of the image. Let $f_0$ be determined in a finite number of points in $[0,1]^2$, say in $N \times M$ equidistant points, $(x_k, y_l)$, $k = 1, \ldots, N$, $l = 1, \cdots, M$, where $x_k = (k-1)/N$ and $y_l = (l-1)/M$. These values correspond to an approximation of $G$ using the lower left corners of the 'pixels'.

This choice allows for a particularly simple approximation of $G^{-1}$. We assume that $N$ and $M$ are large compared to the size of the observation window, thus we have a set-up, which resembles a continuous set-up. Replacing $f$ with $f_0$ in (3.3) and (3.4), we obtain for $x \in [x_k, x_{k+1}) = [\frac{k-1}{N}, \frac{k}{N})$, $y \in [y_l, y_{l+1}) = [\frac{l-1}{M}, \frac{l}{M})$, $k = 1, \ldots, N$, $l = 1, \ldots, M$, discrete approximations given by

$$\Delta = \int_{[0,1]^2} f_0(u, v) \, \mathrm{d}u \, \mathrm{d}v \approx \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} f_0(x_i, y_j),$$

$$g(x) = \frac{1}{\Delta} \int_0^1 f_0(x, v) \, \mathrm{d}v \approx \frac{1}{\Delta M} \sum_{j=1}^{M} f_0(x_k, y_j),$$

$$G_1(x) = \int_0^x g(u) \, \mathrm{d}u = \sum_{i=1}^{k-1} \int_{x_i}^{x_{i+1}} g(u) \, \mathrm{d}u + \int_{x_k}^{x} g(u) \, \mathrm{d}u$$

$$\approx \frac{1}{N} \sum_{i=1}^{k-1} g(x_i) + (x - \tfrac{k-1}{N}) g(x_k)$$

$$\approx \frac{1}{\Delta NM} \sum_{j=1}^{M} \Big( \sum_{i=1}^{k-1} f_0(x_i, y_j) + (Nx - k + 1) f_0(x_k, y_j) \Big),$$

$$G_2(x, y) = \frac{1}{\Delta g(x)} \int_0^y f_0(x, v) \, \mathrm{d}v$$

$$\approx \frac{1}{\Delta g(x_k)} \Big( \sum_{j=1}^{l-1} \int_{y_j}^{y_{j+1}} f_0(x_k, v) \, \mathrm{d}v + \int_{y_l}^{y} f_0(x_k, v) \, \mathrm{d}v \Big)$$

$$\approx \frac{1}{\Delta g(x_k) M} \Big( \sum_{j=1}^{l-1} f_0(x_k, y_j) + (My - l + 1) f_0(x_k, y_l) \Big).$$

In particular,

$$G_1(x_k) \approx \frac{1}{\Delta NM} \sum_{j=1}^{M} \sum_{i=1}^{k-1} f_0(x_i, y_j),$$

$$G_2(x_k, y_l) \approx \frac{1}{\Delta g(x_k) M} \sum_{j=1}^{l-1} f_0(x_k, y_j),$$

for $k = 1, \ldots, N$, $l = 1, \ldots, M$. To approximate the inverse we take for $u, v \in [0, 1)$, $G^{-1}(u, v) \approx (x_k, y_l)$, where $k = \max \{i \mid G_1(x_i) \leq u\}$, $l = \max \{j \mid G_2(x_k, y_j) \leq v\}$. For a given choice of $f_0$, using the approximations above, it is straight forward to simulate non-uniform samples from uniformly generated random variables, and calculate the estimate from (3.5).

## 4.3   Sampling designs

In order to investigate the efficiency of the sampling design, the variance or $CE^2$ (squared coefficient of error) of the design is compared to the ones for standard sampling designs. The designs which are considered are

- 2D continuous: 2D non-uniform systematic sampling, described above and in Section 3.2 (continuous sampling).

- 2D discrete: 2D non-uniform systematic sampling, described in Appendix B, corresponding to proportional-to-size sampling (PPS sampling) with spatial information. This design is a new design proposed in the present paper. It is thought as a compromise between the continuous design above and standard PPS sampling with no spatial information (discrete sampling).

- PPS WR: Proportional-to-size sampling with replacement (discrete sampling).

- SRS WOR: Simple random sampling without replacement (discrete sampling).

When discrete sampling is considered, the section is divided by an equally spaced grid into fields of size $w \times w$ (determined in the section below), thus sampling is performed on a finite number of elements.

## 4.4   The measurement function

Without loss of generality we assume, that the measurement function is defined on the unit square $[0,1]^2$. We imitate observations in microscopy by letting the measurement function $f(x,y)$ be proportional to the measurement of interest on a whole observation window $(x,y) + [0,w]^2$, with side-length $0 \leq w \leq 1$. As we shall see, the parameters of interest are expressible as an integral (3.1) of $f$ over $[0,1]^2$ if the spatial structure is contained in $[w,1]^2$.

### 4.4.1   The measurement function for number estimation

Let $Z = \{z_1, z_2, \ldots\}$ be points in $[w,1]^2$, generated by a pre-chosen point process, indicating centres of a cell population, and assume we want to estimate the number of points in $Z$, $N_Z$. Then, $W_i = z_i - [0,w]^2$ is the set of $(x,y)$ in $[0,1]^2$ for which $z_i$ is counted in the observation window $(x,y) + [0,w]^2$. If we let $f(x,y)$ be the total number of points, counted in $(x,y) + [0,w]^2$, normalized with $1/w^2$,

$$f(x,y) = \sum_{i=1}^{N_Z} \frac{1_{W_i}(x,y)}{w^2}, \tag{4.1}$$

we obtain, as desired, the total number of points $N_Z$ by

$$Q = \int_{[0,1]^2} f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_{[0,1]^2} \sum_{i=1}^{N_Z} \frac{1_{W_i}(x,y)}{w^2}\,\mathrm{d}x\,\mathrm{d}y$$

$$= \sum_{i=1}^{N_Z} \frac{|W_i|}{w^2} = N_Z.$$

### 4.4.2   The measurement function for area or integral estimation

Let $\lambda$ be a non-negative function on $\mathbb{R}^2$ that is identically 0 outside $[w,1]^2$, and assume we want to estimate

$$\int_{[w,1]^2} \lambda(u,v)\,\mathrm{d}u\,\mathrm{d}v,$$

which for $\lambda(x,y) = 1((x,y) \in A)$ corresponds to the area of the set $A \subseteq [w,1]^2$. If we let $f(x,y)$ be the integral of $\lambda$ over $(x,y) + [0,w]^2$, normalized with $1/w^2$,

$$f(x,y) = \frac{1}{w^2} \int_{(x,y)+[0,w]^2} \lambda(u,v)\,\mathrm{d}u\,\mathrm{d}v, \tag{4.2}$$

we obtain the complete integral by

$$\begin{aligned}
Q &= \int_{[0,1]^2} f(x,y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \frac{1}{w^2} \int_{[0,1]^2} \left( \int_{(x,y)+[0,w]^2} \lambda(u,v)\,\mathrm{d}u\,\mathrm{d}v \right) \mathrm{d}x\,\mathrm{d}y \\
&= \frac{1}{w^2} \int_{[0,1]^2} \left( \int_{[0,w]^2} \lambda(u+x,v+y)\,\mathrm{d}u\,\mathrm{d}v \right) \mathrm{d}x\,\mathrm{d}y \\
&= \frac{1}{w^2} \int_{[0,w]^2} \left( \int_{[0,1]^2} \lambda(u+x,v+y)\,\mathrm{d}x\,\mathrm{d}y \right) \mathrm{d}u\,\mathrm{d}v \\
&= \frac{1}{w^2} \int_{[0,w]^2} \left( \int_{(u,v)+[0,1]^2} \lambda(x,y)\,\mathrm{d}x\,\mathrm{d}y \right) \mathrm{d}u\,\mathrm{d}v \\
&= \frac{1}{w^2} \int_{[0,w]^2} \left( \int_{[w,1]^2} \lambda(x,y)\,\mathrm{d}x\,\mathrm{d}y \right) \mathrm{d}u\,\mathrm{d}v \\
&= \int_{[w,1]^2} \lambda(u,v)\,\mathrm{d}u\,\mathrm{d}v.
\end{aligned}$$

## 4.5 Computational details

The simulations have been performed under the following specifications:

- The constructed images consist of $128 \times 128$ pixels covering $[w,1]^2$.
- We let the observation windows consist of $8 \times 8$ pixels, thus $|W_i| = w^2$, $w = 8/(128 + 8 - 1)$.
- We have $N = M = 128 + 8 - 1 = 135$.
- The set $[w,1]^2$ can be covered by $16 \times 16$ observation windows, which is considered in the cases of discrete sampling.
- For each set-up we used sample sizes $n \times m$, where $n = m = 3$.
- For each set-up $10\,000$ samples were simulated to approximate the theoretical variance of the estimator.

## 4.6 Number estimation

### 4.6.1 Choices of $f$ and $f_0$

Several set-ups have been tested, including different sample sizes ($n$ and $m$) and population sizes ($N_Z$). Only some of these will be discussed in detail here. We consider three cases of the measurement function $f$, which are shown in Figure 2. In all cases, $f$ is given by (4.1), generated by point patterns obtained as realizations of different inhomogeneous point processes, resulting in approximately 400 points. We define the auxiliary information $f_0$
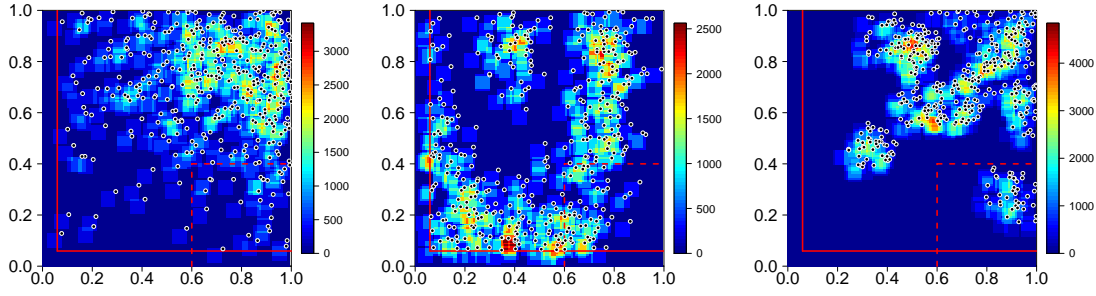
**Figure 2:** The three cases of the measurement function $f$, cf. (4.1), together with the points from realizations of the point process $Z$. The full-drawn red lines indicates the lower and left boundary of $[w, 1]^2$, and the red dashed lines indicates the 'invisible' part of $Z$, for some choices of $f_0$, see (4.3) and Table 1.

by the same sum of indicator functions as the measurement function $f$ in (4.1), but with additional spatial error and a constant added, to avoid division by zero. More precisely,

$$f_0(x,y) = \sum_{z_i \in [0,1]^2 \setminus B} \alpha_i \frac{1_{W_i}(x,y)}{w^2} + c, \qquad (4.3)$$

with different choices of $\alpha_i$, $B$ and $c$. The 9 choices are shown in Table 1 and illustrated in Figure 8 in Appendix A.3. Both $\alpha_i$ and $B$ are used to introduce a spatial error, in particular $B$ is used to make some of the points 'invisible', that is, the points are not expressed in $f_0$.

**Table 1:** Parameter choices for $f_0$ in formula (4.3).

| $f_0$ | $\alpha_i$ | $B$ | $c$ |
|---|---|---|---|
| 1 | $x_i + y_i$ | $\emptyset$ | 1000 |
| 2 | 1 | $\emptyset$ | 1000 |
| 3 | $x_i + y_i$ | $\emptyset$ | 250 |
| 4 | 1 | $\emptyset$ | 250 |
| 5 | $x_i + y_i$ | $[0.6; 1] \times [0; 0.4]$ | 1000 |
| 6 | 1 | $[0.6; 1] \times [0; 0.4]$ | 1000 |
| 7 | $x_i + y_i$ | $[0.6; 1] \times [0; 0.4]$ | 250 |
| 8 | 1 | $[0.6; 1] \times [0; 0.4]$ | 250 |
| 9 | 0 | 0 | 1 |

### 4.6.2 Results

Figure 3 shows the estimated $CE^2$ (Row 1) and the estimated relative $CE^2$ (Row 2) obtained by 10 000 sample simulations, for different choices of the auxiliary information $f_0$ ($x$-axis), as detailed in Table 1, with $n = m = 3$.

The values of the three measurements functions $f$ lies in the intervals $[0; 3417]$, $[0; 2563]$ and $[0; 4841]$, thus when constructing $f_0$ from $f$ with error, the non-uniformity is considerable and large 'smoothing' parameters are needed, e.g. large values of the constant $c$ in
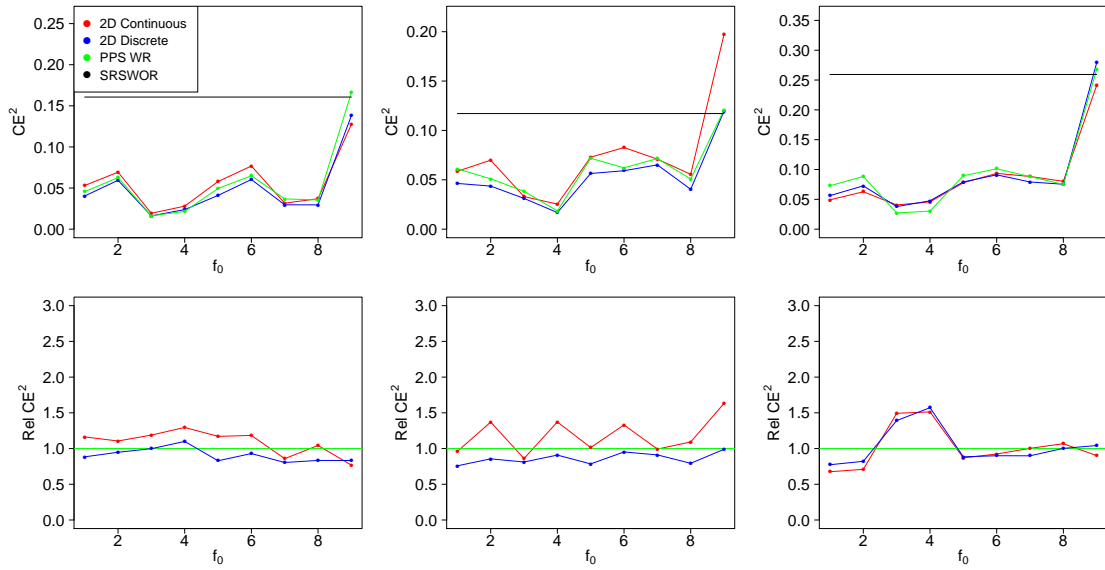
**Figure 3:** Results for number estimation with $n = m = 3$. Row 1: The estimated $CE^2$ for the four sampling designs, in the three cases of $f$. For each choice of $f$, the nine choices of $f_0$ are displayed. Row 2: The estimated relative $CE^2$ for the two 2D non-uniform systematic sampling designs, relative to PPS WR, in the three cases of $f$. For each choice of $f$, the nine choices of $f_0$ are displayed.

Table 1, to prevent extreme estimates. Moreover, extreme non-uniformity results in overlaps between the sample windows in the 2D continuous systematic sampling design and non-fixed effective sample-sizes in the 2D discrete systematic sampling design, even for relative small sample sizes. Due to the extreme non-uniformity, $n = m = 3$ is the upper bound on the sample size in the three cases considered here, and this sample size was therefore chosen.

Figure 3 shows that the main effect on efficiency is from the choice of $f_0$, and there is almost no difference in efficiency between the three designs, 2D continuous, 2D discrete and PPS WR, which uses non-uniform sampling. The effect from the two more complicated designs, which use spatial information, compared to the much more simple PPS WR design, is therefore negligible. The variances naturally become larger in the non-uniform cases, when many of the cells are placed in the invisible part of the section, but due to the large value of $c$, this effect is not clearly expressed in the results. In the uniform case, choice 9 of $f_0$, the effect solely from systematic sampling can be seen. Systematic uniform sampling surprisingly yields in its continuous implementation a higher $CE^2$ than SRS WOR for the second choice of $f$, and only little effect is seen for the remaining two choices of $f$, which may explain why the non-uniform systematic sampling designs do not have the expected effect. In general, it seems that when spatial error is introduced (choices 1,3,5 and 7 of $f_0$), 2D continuous systematic sampling is either almost as efficient or slightly more efficient than PPS WR, and overall the 2D discrete systematic sampling design performs better than the continuous version and often even slightly better than PPS WR.

## 4.7 Area estimation

### 4.7.1 Choices of $f$ and $f_0$
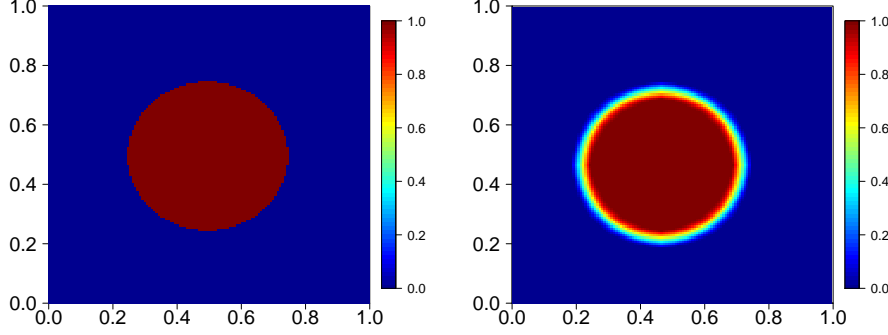


**Figure 4:** Left: The function $\lambda$ equal to the indicator function of the set of interest. Right: The corresponding measurement function $f$, see (4.2).

The integration function $\lambda$ is in this case an indicator function, indicating where the tissue of interest is located. A simple example is investigated here, see Figure 4, where we consider the area of a circular disc with centre $(0.5, 0.5)$ and radius 0.25, denoted $B((0.5, 0.5), 0.25)$. More precisely, using (4.2), the measurement function $f$ can be written as

$$f(x, y) = |B((0.5, 0.5), 0.25) \cap ((x, y) + [0, w]^2)|/w^2. \tag{4.4}$$

We define the auxiliary information $f_0$ from the measurement function $f$ by adding spatial error in the following way

$$\begin{aligned} f_0(x, y) &= c_1 f(x, y) + c_2 f(x, y) \\ &\quad \times \Gamma(x, y)(\Phi_R(x, y) - \Phi_r(x, y)) + c_3, \end{aligned} \tag{4.5}$$

where $\Phi_r(x, y) = 1((x - 0.5)^2 + (y - 0.5)^2 > r^2)$ and $\Phi_R(x, y) = 1((x - 0.5)^2 + (y - 0.5)^2 \leq R^2)$, creating a smoother band around the boundary of the circular disk, and with different choices of $\Gamma$, $c_1$, $c_2$ and $c_3$. The choices are shown in Table 2 and illustrated in Figure 9 in Appendix A.3. In all cases $r = 0.15$ and $R = 0.35$.

### 4.7.2 Results

Figure 5 shows the estimated $CE^2$ and the estimated relative $CE^2$ obtained by 10 000 sample simulations, for different choices of the auxiliary information $f_0$, as detailed in Table 2, with $n = m = 3$.

For most choices of $f_0$, the non-uniformity is only moderate, hence sample sizes up to $n \times m$, with $n = m = 5$, without non-fixed effective sample sizes are possible in 2D discrete sampling. Nevertheless, in order to minimize the probability of repeated sampling of the same window in PPS WR, we consider only results for $n = m = 3$. Difference between PPS WR and the 2D systematic sampling designs will therefore not solely be due to the re-sampling probability.

**Table 2:** Parameter choices for $f_0$ in formula (4.5).

| $f_0$ | $\Gamma$ | $c_1$ | $c_2$ | $c_3$ |
|-------|----------|-------|-------|-------|
| 1 | $x + y$ | 1 | 0.4 | 0.2 |
| 2 | 1 | 1 | 0.4 | 0.2 |
| 3 | $x + y$ | 0 | 0.4 | 0.2 |
| 4 | 1 | 0 | 0.4 | 0.2 |
| 5 | $x + y$ | 1 | 1 | 0.2 |
| 6 | 1 | 1 | 1 | 0.2 |
| 7 | $x + y$ | 0 | 1 | 0.2 |
| 8 | 1 | 0 | 1 | 0.2 |
| 9 | $x + y$ | 1 | 0.4 | 0.4 |
| 10 | 1 | 1 | 0.4 | 0.4 |
| 11 | $x + y$ | 0 | 0.4 | 0.4 |
| 12 | 1 | 0 | 0.4 | 0.4 |
| 13 | $x + y$ | 1 | 1 | 0.4 |
| 14 | 1 | 1 | 1 | 0.4 |
| 15 | $x + y$ | 0 | 1 | 0.4 |
| 16 | 1 | 0 | 1 | 0.4 |
| 17 | 0 | 0 | 0 | 1 |

The effect from just introducing systematic sampling in the uniform case, choice 17 of $f_0$, is clearly substantial, and might explain why, in contrast to the case of number estimation, 2D non-uniform systematic sampling (2D continuous and 2D discrete in Figure 5) overall performs better than PPS WR. Although uniform systematic sampling is an efficient design here, the efficiency may still be increased by combining the systematic sampling with non-uniform sampling. The efficiency is most pronounced for moderate non-uniformity, as large values of $c_3$ reduces the variance of the 2D sampling designs relative to both SRS WOR and PPS WR. There seems to be almost no difference in the results from the discrete and the continuous 2D systematic sample designs.

Clearly the parameter $c_1$ affects the efficiency relative to SRS WOR, as $c_1 = 1$ results in high agreement between $f$ and $f_0$, thus non-uniform sampling is close to optimal. The influence of $c_1$ for the non-uniform designs is less clear. The parameter $c_2$ controls the magnitude of spatial error, and clearly larger values of $c_2$ results in a higher variance, but in contrast to what is expected, it also results in less efficiency compared to PPS WR. There seems to be no or negligible effect from the spatial error introduced by $\Gamma$.

## 4.8 Integral estimation

### 4.8.1 Choices of $f$ and $f_0$

Here we consider three cases of functions $\lambda$, which together with the corresponding measurement functions $f$, see (4.2), are shown in Figure 6. The integrals have been scaled, such that they integrate to one, which unifies the set-up and choices of parameters. We define the auxiliary information $f_0$ from the measurement function $f$ by adding spatial error in
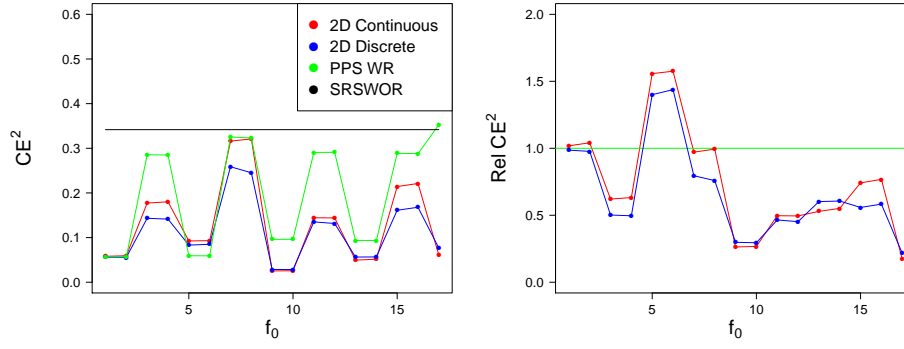
**Figure 5:** Results for area estimation with $n = m = 3$. Left: The estimated $CE^2$ for the four sampling designs for the nine choices of $f_0$ are displayed. Right: The estimated relative $CE^2$ for the two 2D non-uniform systematic sampling designs, relative to PPS WR.
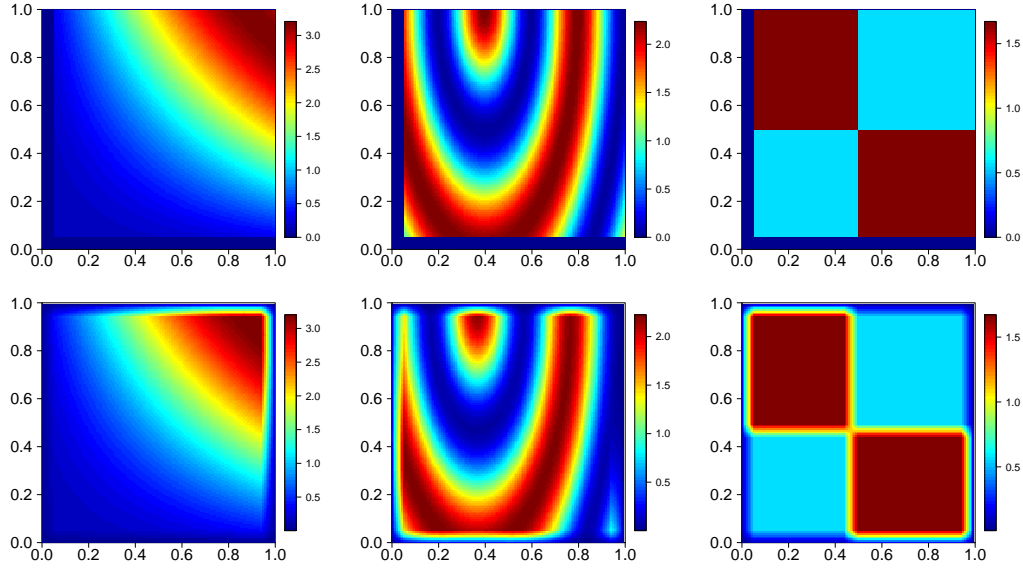


**Figure 6:** Row 1: The three functions $\lambda$. Row 2: The three corresponding measurement functions $f$.

the following way

$$f_0(x, y) = c_1 f(x, y) + c_2 f(x, y)\Gamma(x, y) + c_3, \tag{4.6}$$

with different choices of $\Gamma$, $c_1$, $c_2$ and $c_3$. The choices are shown in Table 3 and illustrated in Figure 10 in Appendix A.3.

**Table 3:** Parameter choices for $f_0$ in formula (4.6), with $A = [0,1]^2 \backslash ([0.6,1] \times [0,0.4])$.

| Sim | $\Gamma$ | $c_1$ | $c_2$ | $c_3$ |
|---|---|---|---|---|
| 1 | $x + y$ | 1 | 0.4 | 0.4 |
| 2 | 1 | 1 | 0.4 | 0.4 |
| 3 | $x + y$ | 0 | 0.4 | 0.4 |
| 4 | 1 | 0 | 0.4 | 0.4 |
| 5 | $x + y$ | 1 | 1 | 0.4 |
| 6 | 1 | 1 | 1 | 0.4 |
| 7 | $x + y$ | 0 | 1 | 0.4 |
| 8 | 1 | 0 | 1 | 0.4 |
| 9 | $(x + y + 0.1)1((x,y) \in A)$ | 1 | 0.4 | 0.4 |
| 10 | $1((x,y) \in A)$ | 1 | 0.4 | 0.4 |
| 11 | $(x + y + 0.1)1((x,y) \in A)$ | 0 | 0.4 | 0.4 |
| 12 | $1((x,y) \in A)$ | 0 | 0.4 | 0.4 |
| 13 | $(x + y + 0.1)1((x,y) \in A)$ | 1 | 1 | 0.4 |
| 14 | $1((x,y) \in A)$ | 1 | 1 | 0.4 |
| 15 | $(x + y + 0.1)1((x,y) \in A)$ | 0 | 1 | 0.4 |
| 16 | $1((x,y) \in A)$ | 0 | 1 | 0.4 |
| 17 | 0 | 0 | 0 | 1 |

### 4.8.2 Results

Figure 7 shows the estimated $CE^2$ and the estimated relative $CE^2$, obtained by $10\,000$ sample simulations, for different choices of the auxiliary information $f_0$, detailed in Table 3, with $n = m = 3$.

In all cases of $f_0$, the non-uniformity is only moderate, hence allow sample sizes up to $n \times m$, with $n = m = 5$, without non-fixed effective sample sizes in 2D discrete sampling. Nevertheless, with arguments stated in the previous section, we consider only results for $n = m = 3$.

The main effect is again from the choice of $f_0$, and there is almost no difference between the three designs, which use non-uniform sampling. The relative differences seem on the other hand more significant. In particular the 2D continuous design does not perform well, but the 2D discrete design perform similar or better than PPS WR.

Introducing systematic sampling in the uniform case, choice 17 of $f_0$, increases the efficiency in the first and third case, where the non-uniformity is most systematic, whereas in the second case the efficiency decreases. Interchanging the sampling order (first on the $y$-axis, then on the $x$-axis instead of the opposite order) did not change the results much.
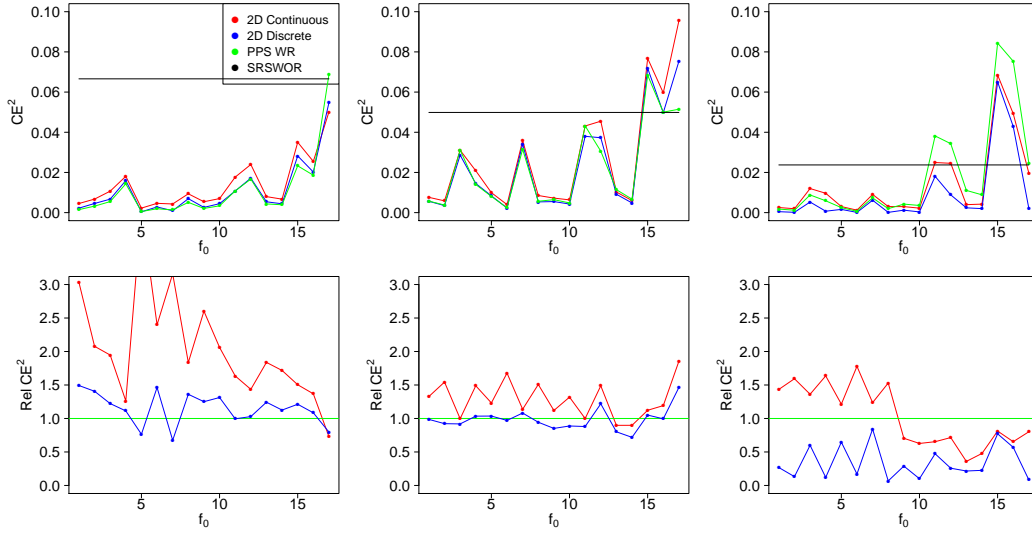
**Figure 7:** Results for area estimation with $n = m = 3$. Row 1: The estimated $CE^2$ for the four sampling designs, in the three cases of $f$. Row 2: The estimated relative $CE^2$ for the two 2D non-uniform systematic sampling designs, relative to PPS WR, in the three cases of $f$. For each choice of $f$, the seventeen choices of $f_0$ are displayed.

# 5   Discussion and other perspectives

We have shown the existence of an optimal choice of 2D non-uniform systematic sampling. The design reduces the variance substantially compared to uniform sampling, when the auxiliary information used to construct the sampling inclusion probabilities have a close connection to the measurement function under consideration. More precisely, when a function proportional to the measurement function is known, the design yields zero variance of the estimator.

To support our theoretical findings and investigate the efficiency of this new design compared to other designs, a simulation study was carried out. As the project was motivated by sampling in microscopy, it was natural to construct a set-up, which resembles data obtained from this field of research. Number estimation, area estimation and general integral estimation was simulated for several choices of the measurement function and the auxiliary information, with different levels of spatial error added.

In most cases considered, the 2D non-uniform systematic designs had similar efficiency as PPS WR. An exception was area estimation where the non-uniform systematic designs were superior in a number of cases considered. Within the non-uniform systematic designs, the new discrete design is more efficient than the continuous design in a number of cases. The discrete design is easy to simulate and with a moderate constant added a robust and efficient choice, if one suspects that the auxiliary information $f_0$ is not optimal.

In the simulation study, other sample sizes ($n$ and $m$) were considered such that the range 3 % to 25 % of the total number of observation windows was covered. The conclusion concerning the relative efficiency of the designs was the same as for the sample sizes considered in the present paper.

In the case of number estimation in a Poisson point process, it is expected that 2D non-uniform systematic sampling does not have higher efficiency than the proportionator,

simply because the numbers counted in disjoint observation windows are independent. The simulation study shows that in a wider range of sampling situations in microscopy the gain in efficiency, if any, is modest.

One might wonder whether it is possible to construct a theoretical class of sampling situations for which the parameter of interest is more efficiently estimated, using 2D non-uniform systematic sampling compared to independent 2D non-uniform sampling. For this purpose, let $f_{\text{hom}}$ be a bounded non-negative function with bounded support $[0,1]^2$. Suppose that

$$\text{Var}\Big(\frac{1}{nm}\sum_{i,j} f_{\text{hom}}(Y_{1i}, Y_{2j})\Big) < \text{Var}\Big(\frac{1}{nm}\sum_{i,j} f_{\text{hom}}(U_{1i}, U_{2j})\Big),$$

where $(Y_{1i}, Y_{2j}) = ((U_1 + i)\frac{1}{n}, (U_2 + j)\frac{1}{m})$ and $U_1, U_2 \sim \text{unif}[0,1)$ independent, while the $U_{1i}$s and $U_{2j}$s are all independent and unif[0,1). Furthermore, let $G : [0,1]^2 \to [0,1]^2$ be any diffeomorphism. Then,

$$f_{\text{inhom}}(x,y) = f_{\text{hom}}(G(x,y)) \cdot |\det G'(x,y)|$$

is more efficiently estimated using 2D non-uniform systematic sampling than independent 2D non-uniform sampling, induced by $G$. More precisely, the variance of

$$\frac{1}{nm}\sum_{i,j} \frac{f_{\text{inhom}}(G^{-1}(Y_{1i}, Y_{2j}))}{|\det G'(G^{-1}(Y_{1i}, Y_{2j}))|}$$

is smaller than the variance of

$$\frac{1}{nm}\sum_{i,j} \frac{f_{\text{inhom}}(V_{ij})}{|\det G'(V_{ij})|},$$

where $V_{ij} = (V_{1ij}, V_{2ij})$ are independent and with common density $|\det G'(v)|$. The practical consequences of this finding are subject of future research.

# 6 Acknowledgement

# References

Baddeley, A. and R. Turner (2005). Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software 12*(6), 1–42.

Cruz-Orive, L. (1993). Systematic sampling in stereology. *Proceedings 49th Session Int. Statist. Inst. 2*, 451–468.

Dorph-Petersen, K.-A., H. J. G. Gundersen, and E. B. V. Jensen (2000). Non-uniform systematic sampling in stereology. *Journal of Microscopy 200*(2), 148–157.

Gardi, J. E., J. R. Nyengaard, and H. J. G. Gundersen (2008a). Automatic sampling for unbiased and efficient stereological estimation using the proportionator in biological studies. *Journal of Microscopy 230*(1), 108–120.

Gardi, J. E., J. R. Nyengaard, and H. J. G. Gundersen (2008b). The proportionator: unbiased stereological estimation using biased automatic image analysis and non-uniform probability proportional to size sampling. *Computers in Biology and Medicine 38*(3), 313–328.

Grafström, A. and Y. Tillé (2013). Doubly balanced spatial sampling with spreading and restitution of auxiliary totals. *Environmetrics 24*(2), 120–131.

Gundersen, H. J. G., E. B. V. Jensen, K. Kiêu, and J. Nielsen (1999). The efficiency of systematic sampling in stereology – reconsidered. *Journal of Microscopy 193*(3), 199–211.

Kołodziej, S. (1998). The complex Monge-Ampere equation. *Acta Mathematica 180*(1), 69–117.

Stevens Jr, D. L. and A. R. Olsen (2004). Spatially balanced sampling of natural resources. *Journal of the American Statistical Association 99*(465), 262–278.

# Appendix

## A.1  Unbiasedness of $\hat{Q}_{nm}$

The estimator $\hat{Q}_{nm}$ is unbiased, which follows directly by using the distribution of $U_1$ and $U_2$. We let $A_n = [0, \frac{1}{n}] \times [0, \frac{1}{m}]$ and obtain

$$
\begin{aligned}
\mathbb{E}(\hat{Q}_{nm}) &= \mathbb{E}\Big( \frac{1}{nm} \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} \frac{f(G^{-1}(Y_{1i}, Y_{2j}))}{|\det(G'(G^{-1}(Y_{1i}, Y_{2j})))|} \Big) \\
&= \frac{1}{nm} \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} \mathbb{E}\Big( \frac{f(G^{-1}(\frac{U_1+i}{n}, \frac{U_2+j}{m}))}{|\det(G'(G^{-1}(\frac{U_1+i}{n}, \frac{U_2+j}{m})))|} \Big) \\
&= \frac{1}{nm} \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} \int_{A_n} \frac{nmf(G^{-1}(u+\frac{i}{n}, v+\frac{j}{m}))}{|\det(G'(G^{-1}(u+\frac{i}{n}, v+\frac{j}{m})))|} \, \mathrm{d}u \, \mathrm{d}v \\
&= \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} \int_{[\frac{i}{n}, \frac{i+1}{n}] \times [\frac{j}{m}, \frac{j+1}{m}]} \frac{f(G^{-1}(u,v))}{|\det(G'(G^{-1}(u,v)))|} \, \mathrm{d}u \, \mathrm{d}v \\
&= \int_{[0,1]^2} f(G^{-1}(u,v)) \frac{1}{|\det(G'(G^{-1}(u,v)))|} \, \mathrm{d}u \, \mathrm{d}v \\
&= \int_{[0,1]^2} f(x,y) \, \mathrm{d}x \, \mathrm{d}y \\
&= Q.
\end{aligned}
$$

## A.2  Relation to proportional-to-size sampling

2D systematic sampling proportional-to-size can be described as follows. Let us consider a population of $N \times M$ units $\mathscr{P} = \{(i,j), i = 1, \ldots, N, j = 1, \ldots M\}$, where the numbers refer to the spatial arrangement. For each unit we have an unknown variable of interest $x_{ij}$, and a known auxiliary variable $z_{ij}$, which is positively correlated with $x_{ij}$. The object is to estimate

$$
Q = \sum_{(i,j) \in \mathscr{P}} x_{ij},
$$

from a systematic sample $S \subseteq \mathscr{P}$ of fixed size $n \times m$. The sampling procedure performed below generates samples, where the probability of including unit $(i,j)$ in $S$, is proportional to $z_{ij}$. The procedure uses the (marginal) cumulative values of $z_{ij}$ in two steps, where in each step, systematic sampling is performed analogous to the procedure in 1D. First, the units are divided into $N$ groups according to the values of the first coordinate, from which $n$ groups are sampled, followed by sampling $m$ units, for each of the $n$ groups. Let $z_g = \sum_{j=1}^{M} z_{gj}, g = 1, \ldots, N$, denote the marginal auxiliary variables and $\Delta_0 = 0$, $\Delta_g = \sum_{i=1}^{g} z_i, g = 1, \ldots, N$, the cumulative ones. Then, letting $U_1 \sim \text{unif}[0,1)$, the group $g$ is chosen if there exists an $i = 0, \ldots, n-1$, such that

$$
(U_1 + i) \tfrac{1}{n} \Delta_N \in [\Delta_{g-1}, \Delta_g).
$$

Assume $z_g < \Delta_N / n, g = 1, \ldots, N$, such that no group is sampled more than once. Next, for each of the $n$ sampled groups $g$, let $\Delta_{g0} = 0$, $\Delta_{gk} = (\Delta_{gM}/z_g) \sum_{j=1}^{k} z_{gj}, k = 1, \ldots, M$,

denote the (scaled) cumulative auxiliary variables within the group $g$. Then, letting $U_2 \sim$ unif$[0,1)$, unit $k$ is chosen if there exists an $j = 0, \ldots, m-1$, such that

$$(U_2 + j)\tfrac{1}{m}\Delta_{gM} \in [\Delta_{g(k-1)}, \Delta_{gk}).$$

Assume again that $z_{gk} < \Delta_{gM}/m, k = 1, \ldots, M$. It can be shown (see calculations below), that $\pi_{ij} = \mathbb{P}((i,j) \in S) = nmz_{ij}/\Delta_N$ from which we get the Horvitz-Thompson estimator

$$\hat{Q}_{nm} = \sum_{(i,j)\in S} \frac{x_{ij}}{\pi_{ij}} = \frac{1}{nm}\Delta_N \sum_{(i,j)\in S} \frac{x_{ij}}{z_{ij}}.$$

The following calculation verifies that the scaling of the auxiliary variables results in the correct inclusion probabilities, using the distribution of $U_1, U_2$ and the assumptions $z_g < \Delta_N/n$ and $z_{gk} < \Delta_{gM}/m$. We have with $A_g = [\Delta_{g-1}, \Delta_g)$, $A_{gk} = [\Delta_{g(k-1)}, \Delta_{gk})$ and $B_g(i,j) = [i, i+1)\tfrac{1}{n}\Delta_N \times [j, j+1)\tfrac{1}{m}\Delta_{gM}$

$$
\begin{aligned}
\pi_{gk} &= \mathbb{P}((g,k) \in S) \\
&= \sum_{i=0}^{n-1}\sum_{j=0}^{m-1} \mathbb{P}\left((U_1 + i)\tfrac{1}{n}\Delta_N \in A_g, (U_2 + j)\tfrac{1}{m}\Delta_{gM} \in A_{gk}\right) \\
&= \sum_{i=0}^{n-1}\sum_{j=0}^{m-1} \int_{A_g \times A_{gk} \cap B_g(i,j)} \frac{nm}{\Delta_N \Delta_{gM}}\, \mathrm{d}u\, \mathrm{d}v \\
&= \frac{nm}{\Delta_N \Delta_{gM}} |A_g \times A_{gk}| \\
&= \frac{nm}{\Delta_N \Delta_{gM}} z_g \frac{\Delta_{gM} z_{gk}}{z_g} \\
&= nm\frac{z_{gk}}{\Delta_N}.
\end{aligned}
$$

The relation to the set-up with integrals of a measurement function follows when we let

$$
\begin{aligned}
f(x,y) &= NMx_{ij}, \\
f_0(x,y) &= NMz_{ij}, \qquad \text{when} \\
(x,y) &\in [i-1, i)\tfrac{1}{N} \times [j-1, j)\tfrac{1}{M}
\end{aligned}
$$

and

$$
G_1(x,y) = \int_0^x g(u)\, \mathrm{d}u = \frac{1}{\Delta}\left(\sum_{k=1}^{i-1}\sum_{j=1}^{M} z_{kj} + (Nx - i + 1)\sum_{j=1}^{M} z_{ij}\right)
$$

$$
G_2(x,y) = \frac{1}{g(x)\Delta}\int_0^y f_0(x,v)\, \mathrm{d}v = \frac{1}{\sum_{j=1}^{M} z_{ij}}\left(\sum_{k=1}^{j-1} z_{ik} + (My - j + 1)z_{ij}\right),
$$

where

$$
g(x) = \frac{\int_0^1 f_0(x,y)\, \mathrm{d}y}{\Delta} = \frac{N\sum_{j=1}^{M} z_{ij}}{\Delta},
$$

$$
\Delta = \int_{[0,1]^2} f_0(x,y)\, \mathrm{d}x\, \mathrm{d}y = \sum_{(i,j)\in\mathscr{P}} z_{ij},
$$

when $(x,y) \in [i-1, i)/N \times [j-1, j)/M$, for $i = 1 \ldots, N$ and $j = 1, \ldots, M$.

## A.3 Figures
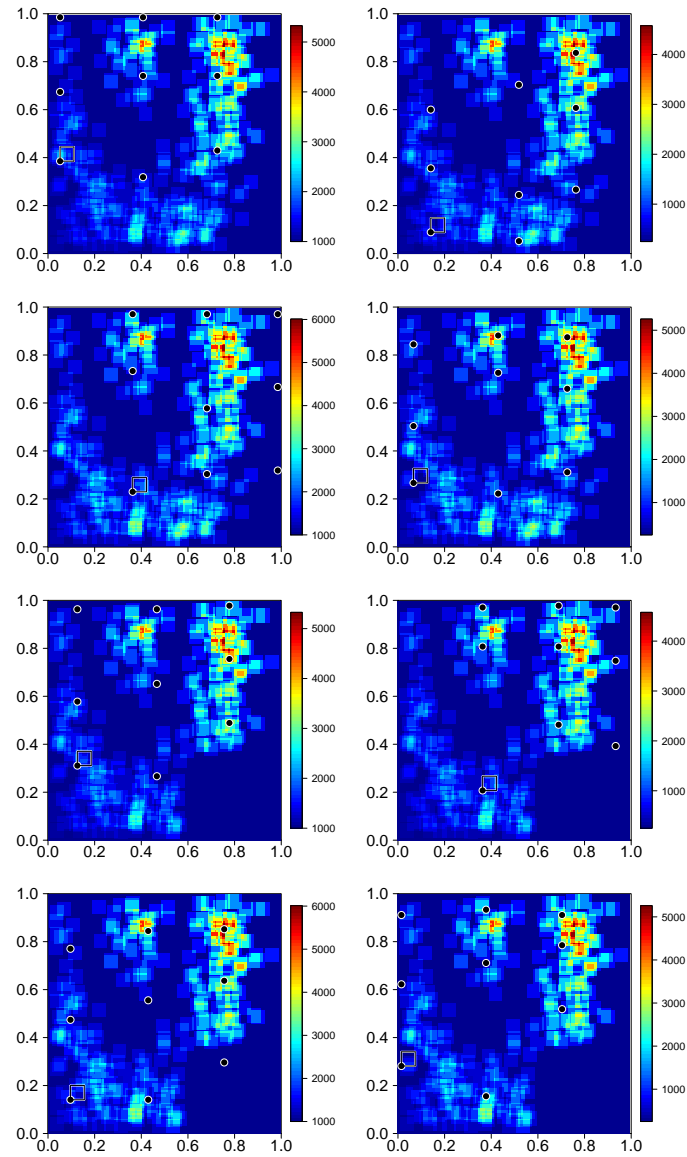


**Figure 8:** The eight cases of non-uniform auxiliary information $f_0$ for number estimation, for the second measurement function, together with one example of sample points.
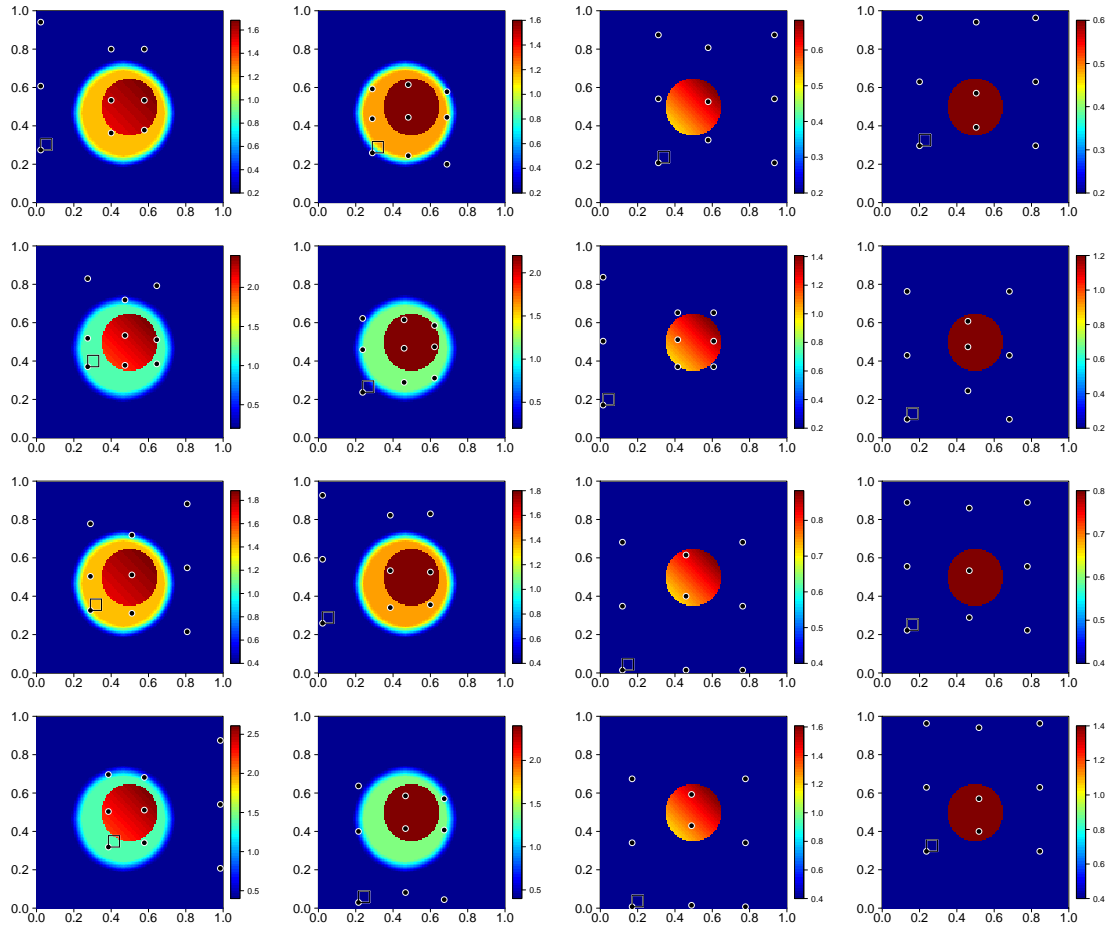
**Figure 9:** The sixteen cases of non-uniform auxiliary information $f_0$ for area estimation, together with one example of sample points.
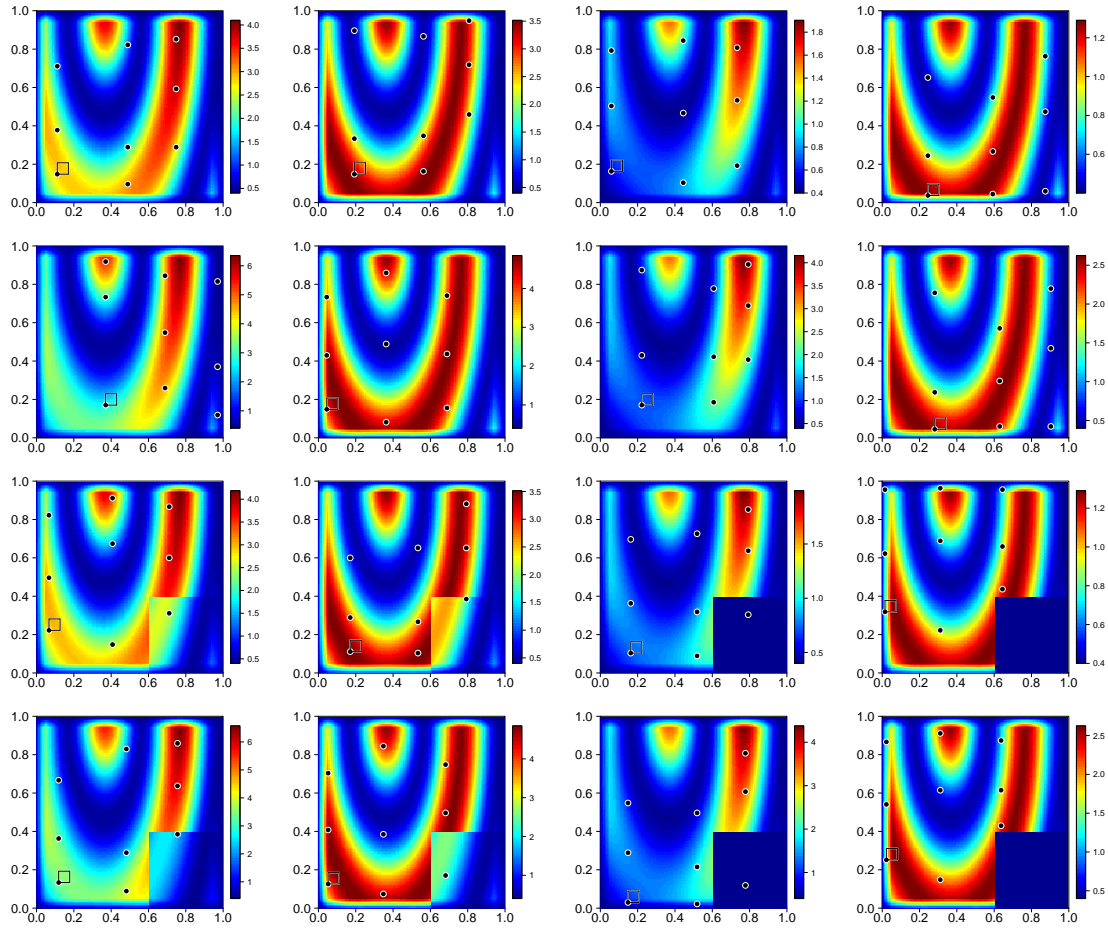
**Figure 10:** The auxiliary information in the sixteen cases of $f_0$ for integral estimation, for the second measurement function, together with one example of sample points.

# Paper D

# Matérn thinned Cox processes

By I. T. Andersen & U. Hahn

# Matérn thinned Cox processes

Ina Trolle Andersen [a,b], Ute Hahn [a,*]

[a] *Department of Mathematics, Aarhus University, Ny Munkegade 118, 8000 Aarhus C, Denmark*

[b] *Stereological Research Laboratory, Department of Clinical Medicine, Aarhus University, Nørrebrogade 44, 8000 Aarhus C, Denmark*

## ABSTRACT

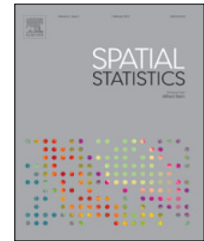A class of spatial point process models that combine short range repulsion with medium range clustering is introduced. The model is motivated by patterns of centres of non-overlapping spherical cells in biological tissue which tend to have a clustering behaviour. Such a combination of clustering and hard core behaviour can be achieved by applying a dependent Matérn thinning to a Cox process. An exact formula for the intensity of a Matérn thinned shot noise Cox process is derived from the Palm distribution. For the more general class of Matérn thinned Cox processes, formulae for the intensity and second-order characteristics are established using the conditional Poisson assumption. These formulae include more complicated integrals for which approximations are suggested to simplify calculations. An example from pathology illustrates the applicability of the models.

© 2015 Published by Elsevier B.V.

## 1. Introduction

The existing point process literature provides models for a variety of interactions between points, of which the models that allow for simple statistical inference often are those used in applications. Most models yield point patterns with either clustering *or* hard core behaviour; however in practical applications, one may observe both types of interaction on different scales simultaneously. One such example is the pattern formed by centres of cells in cell clusters—the centres cannot come closer than the diameter of the cells, but nevertheless they show clustering on a mid range of spatial distance. This

example has motivated the present study. In the literature, such cases are often modelled by Gibbs point processes with an appropriate interaction function, as e.g. in Mattfeldt et al. (2006, 2007). While they have an intuitive physical interpretation through the interaction function, theoretical properties and summary statistics of Gibbs models are accessible only by simulation. Another modelling option is to start with a stationary hard core process and obtain a clustered behaviour by independent thinning with probability according to a random field. Second-order summary statistics of these so-called interrupted point processes (Stoyan, 1979; Lavancier and Møller, in press) can be obtained straightforwardly from the properties of the hard core process and the random field. In particular if the random field can take zero values, such as a Boolean model, this approach yields a clustered appearance. However, the spatial arrangement of points inside the clusters is influenced by "invisible" points outside the clusters that have been thinned from the original homogeneous pattern, which, depending on the application, may seem less natural from a physical point of view.

In the present paper, we introduce and investigate a class of mathematically tractable point process models that combine clustering and hard core property, namely by applying dependent Matérn type II thinning (Matérn, 1960, 1986) to a clustered Cox process (Cox, 1955; Møller and Waagepetersen, 2004). In a nutshell, Matérn's thinning algorithms remove points from an existing pattern that have neighbours that are closer than a given hard core distance $h$. The thinning condition can be interpreted as the condition that balls with diameter $h$ attached to the points may not overlap. In recent years, generalizations of Matérn hard core models have appeared in the literature. These papers modify the thinning condition, by replacing the non-overlapping balls with more general (random) convex sets (Månsson and Rudemo, 2002; Kiderlen and Hörig, 2013), or by thinning according to more general functions of the distance between points (Teichmann et al., 2013), but they still are confined to thinning a homogeneous Poisson point process, as in Matérn's original work. Very little is known about the point processes resulting from application of Matérn thinning rules to other models. In his Ph.D. thesis (2005), Tscheschel gives a formula for the intensity of Matérn thinned Matérn cluster point processes as a model for microstructure of rubber. With the present paper, we supply a general framework of Palm retention probabilities for calculating both first- and second-order densities of point processes with known Palm distribution in Section 3. For the applications, we focus on Cox point processes as a very flexible class for clustered patterns.

The paper is organized as follows. In Section 2 we give a short theoretical overview of the applied standard point process models, and we recall some of their properties. In Section 3 we derive general expressions for Palm retention probabilities obtained after a Matérn type II thinning procedure. These probabilities are important in the analysis of Matérn thinned Cox processes, defined in Section 4. Sample realizations and theoretical results with respect to Palm retention probabilities, first- and second-order characteristics of the Matérn thinned Cox processes are also presented in Section 4. Simple approximations are suggested in Section 5, for which the quality is supported by simulations of two examples of Matérn thinned Cox processes, the Matérn thinned Matérn cluster process (MCP) and the Matérn thinned Thomas process (TP). The applicability of the proposed class of point processes is illustrated in Section 6 by means of an example from pathology of patterns of megakaryocytes in bone marrow. Finally, a short discussion is found in Section 7 and all the included proofs are found in the Appendix.

## 2. Preliminaries

This section introduces the notation and basic properties of the point process models considered in this paper. References for detailed description of the theory of point processes include Stoyan et al. (1995), Illian et al. (2008) and Møller and Waagepetersen (2004).

### 2.1. Spatial point processes

Let $N_{\mathrm{lf}}$ denote the set of locally finite subsets of the $d$-dimensional Euclidean space $\mathrm{R}^d$, equipped with an appropriate $\sigma$-algebra $\mathcal{N}_{\mathrm{lf}}$. Then, *a spatial point process X* is a random variable taking values in

$N_{lf}$. Assume $X$ has well-defined *intensity function* $\rho(\cdot)$ and *second-order product density* $\rho^{(2)}(\cdot, \cdot)$, such that the intensity measure $\alpha$ and the second factorial moment measure $\alpha^{(2)}$ are given by

$$\alpha(B) = \text{E}\left[\sum_{\xi \in X} \mathbb{1}(\xi \in B)\right] = \int_B \rho(\xi)\,d\xi,$$

$$\alpha^{(2)}(B_1 \times B_2) = \text{E}\left[\sum_{\xi, \eta \in X}^{\neq} \mathbb{1}(\xi \in B_1, \eta \in B_2)\right] = \int_{B_1} \int_{B_2} \rho^{(2)}(\xi, \eta)\,d\xi\,d\eta,$$

for $B$, $B_1$ and $B_2$ in the Borel $\sigma$-algebra $\mathcal{B}$ on $\text{R}^d$. Here $\sum^{\neq}$ denotes summation over distinct pairs. The interaction between pairs of points can be described by the *pair correlation function*

$$g(\xi, \eta) = \rho^{(2)}(\xi, \eta)/(\rho(\xi)\rho(\eta)).$$

The process is said to be *stationary* if its distribution is translation invariant. Stationarity implies that the intensity function is constant and that $\rho^{(2)}(\xi, \eta) = \rho^{(2)}(\xi - \eta)$. The process is said to be *isotropic* if its distribution is invariant with respect to rotations around the origin. For stationary and isotropic point processes, we have $\rho^{(2)}(\xi, \eta) = \rho^{(2)}(\|\xi - \eta\|)$. In this case, the pair correlation function is effectively a function on R, $g(r) = \rho^{(2)}(r)/\rho^2$, $r \in \text{R}$, and the function can be interpreted as the mean number of points at distance $r$ from a "typical" point in $X$, relative to the mean number for a Poisson process. Here "typical" point is to be understood by means of Palm distribution theory. For the reduced Palm distribution in a point $\xi$, we will use the notation $\mathbb{P}^!_\xi$, which can be interpreted as the conditional distribution of $X \backslash \xi$ given $\xi \in X$ (Møller and Waagepetersen, 2004, p. 249).

## 2.2. Cox processes

Let $\{\Lambda(\xi), \xi \in \text{R}^d\}$ be a non-negative random field, which is almost surely locally integrable with respect to the Lebesgue measure. A point process $X$ is by definition a *Cox process with driving field* $\Lambda$, if conditionally on $\Lambda$, $X$ is a Poisson process with intensity function $\Lambda$ (Cox, 1955; Møller and Waagepetersen, 2004). It follows from the conditional behaviour of $X$, that the intensity function and the second-order product density are given by

$$\rho(\xi) = \text{E}[\Lambda(\xi)] \quad \text{and} \quad \rho^{(2)}(\xi, \eta) = \text{E}[\Lambda(\xi)\Lambda(\eta)]. \tag{1}$$

Furthermore, the *void probability* for bounded $B \in \mathcal{B}$ is given by

$$\mathbb{P}(X \cap B = \emptyset) = \text{E}\left[\exp\left(-\int_B \Lambda(\xi)\,d\xi\right)\right], \tag{2}$$

and the *generating functional* $G_X(f) = \text{E}[\prod_{\xi \in X} f(\xi)]$ for a function $f : \text{R}^d \to [0, 1]$ is

$$G_X(f) = \text{E}\left[\exp\left(-\int_{\text{R}^d} (1 - f(\xi))\Lambda(\xi)\,d\xi\right)\right], \tag{3}$$

cf. Møller and Waagepetersen (2004, p. 60). A Cox process is stationary and isotropic if the driving field $\Lambda$ is stationary and isotropic.

*Shot noise Cox processes* (SNCPs) are characterized by random intensity functions of the form

$$\Lambda(\xi) = \sum_{(c, \gamma) \in \Phi} \gamma k(c, \xi), \tag{4}$$

where $\Phi$ is a Poisson process of cluster centres and cluster intensities on $\text{R}^d \times (0, \infty)$ with a locally finite diffuse intensity measure denoted by $\zeta$, and $k(c, \cdot)$ is a kernel, i.e. a probability density function on $\text{R}^d$. See more details in Møller (2003). SNCPs can be regarded as *Poisson cluster processes*, as $X|\Phi$ is distributed as the superposition of independent Poisson processes $X_{(c,\gamma)}$, $(c, \gamma) \in \Phi$, with intensity functions $\gamma k(c, \cdot)$.

*Neyman–Scott processes* (NSPs) (Neyman and Scott, 1958), which are also Cox processes, can be regarded as a particular case of SNCPs with constant $\gamma$ in (4). In the stationary case, $\Lambda$ takes the form

$$\Lambda(\xi) = \sum_{c \in C} \mu k(\xi - c), \tag{5}$$

where $C$ is a stationary Poisson process on $\mathrm{R}^d$ with intensity $\kappa > 0$, say. The intensity for a stationary NSP is $\rho = \mu\kappa$. Two examples of well-known stationary NSPs are given below. These processes will be used throughout the paper to show how the general formulae simplify in concrete examples.

**Example 1** (*Matérn Cluster Process*)**.** A simple and popular NSP is the *Matérn cluster process* (MCP) (Matérn, 1960, 1986), where the kernel

$$k(\xi) = \mathbb{1}(\|\xi\| \leq R)/|B(o, R)|$$

is the uniform density on a ball $B(o, R)$ centred at the origin $o$ and of radius $R > 0$. Here and in the following, $|B|$ denotes the volume of a set $B$. Note that a MCP is distributed as $\cup_{c \in C} X_c$, where $X_c|C$ is a stationary Poisson point process on $B(c, R)$ with mean number of points $\mu > 0$, and where $X_c, c \in C$, are mutually independent and independent of $C$.

**Example 2** (*Thomas Process*)**.** The class of NSPs also includes the Thomas process (TP) (Thomas, 1949), where the kernel

$$k(\xi) = \exp(-\|\xi\|^2/2\sigma^2)/(2\pi\sigma^2)^{d/2}$$

is the density for $d$ independent normally distributed variables with mean 0 and variance $\sigma^2 > 0$. As for the MCP, the TP can be constructed as a Poisson cluster process.

The reduced Palm distribution $\mathbb{P}^!_\xi$ of a SNCP takes a particularly simple form, as it is just the distribution of $X$ superposed with an independent cluster containing $\xi$. This turns out to be useful in the following sections. The result is restated below. The proof may be found in Møller (2003, Proposition 2).

**Proposition 1** (*Reduced Palm Distribution of SNCPs*)**.** *Let $X$ be a SNCP with random intensity of the form* (4)*. For $\rho(\xi) > 0$, let*

$$\Lambda_\xi(\eta) = \gamma_\xi k(c_\xi, \eta), \quad \eta \in \mathrm{R}^d,$$

*where $(c_\xi, \gamma_\xi)$ is a random variable with distribution*

$$\mathbb{P}((c_\xi, \gamma_\xi) \in D) = \frac{\int_D \gamma k(c, \xi) \, \mathrm{d}\zeta(c, \gamma)}{\rho(\xi)}, \quad \text{for Borel sets } D \subseteq \mathrm{R}^d \times (0, \infty).$$

*Let $X_\xi|(c_\xi, \gamma_\xi)$ be a Poisson process with intensity function $\Lambda_\xi$ and let $(c_\xi, \gamma_\xi, X_\xi)$ be independent of $(\Phi, X)$. Then, for Lebesgue almost all $\xi$ with $\rho(\xi) > 0$,*

$$\mathbb{P}^!_\xi(F) = \mathbb{P}(X \cup X_\xi \in F), \quad F \in \mathcal{N}_{\mathrm{lf}}.$$

### 2.3. Matérn's hard core models

Matérn (1960, 1986) introduced three hard core models, obtained by dependent thinning of a stationary Poisson process, according to a hard core distance $h > 0$. These models are known as Matérn's processes of types I, II and III. The most simple model, based on Matérn type I thinning, solely depends on the point configuration of the original process. All points with a neighbour closer than the distance $h$ are deleted, and this thinning rule results in the most sparse process. In contrast, the most popular thinning algorithm, the Matérn type II thinning, depends also on independent "arrival time" marks, which are attached to the points of the original Poisson process. The hard core property is then achieved by removing points if there are other points within a distance $h$ with a smaller mark.

Naturally, the intensity of the type II model becomes higher than that for the type I model. To increase the intensity further, Matérn proposed an iterative procedure, where a point is only influenced by other points that have not already been deleted (type III model). Due to the resulting long range dependence, closed forms for the summary statistics of the type III model do not exist, and therefore we only consider Matérn type II thinning in this paper.

## 3. Palm retention probabilities for Matérn type II thinned processes

An important tool in analysing the Matérn thinned Cox processes is calculation of Palm retention probabilities. They have the intuitive interpretation as the probability that a given point (or a pair of points) that is present in the original pattern "survives" the thinning. In this section, we derive expressions for these probabilities, valid for an arbitrary point process $X$.

Let the mark attached to a point $\xi \in X$ of the point process be denoted as $m_\xi$. The marks are assumed independent, continuously and identically distributed, and independent of $X$. Without loss of generality, the marks will be assumed to be uniformly distributed on $[0, 1)$ (Stoyan and Stoyan, 1985). Let $X_M = \{(\xi, m_\xi) | \xi \in X, m_\xi \sim \text{unif}[0, 1)\}$ denote the marked process. The Matérn type II thinned process is then given by

$$\text{MatII}(X_M; h) := \{\xi \in X \mid \forall (\eta, m_\eta) \in X_M, \eta \neq \xi : \|\xi - \eta\| > h \vee m_\eta > m_\xi\}. \tag{6}$$

The ratio between the intensity function $\rho_{\text{th}}$ of the thinned process $X_{\text{th}} = \text{MatII}(X_M; h)$ and the intensity function $\rho$ of the original process $X$ can be interpreted as a retention probability

$$p_{\text{ret}}(\xi) := \rho_{\text{th}}(\xi)/\rho(\xi). \tag{7}$$

The probability will be called the *Palm retention probability*, since it can be expressed in terms of the reduced Palm distribution $\mathbb{P}^!_{(\xi,m_\xi)}$ for the marked process $X_M$. In fact,

$$p_{\text{ret}}(\xi) = \int_0^1 p_{\text{ret}}(\xi, m_\xi) \, dm_\xi, \tag{8}$$

where

$$p_{\text{ret}}(\xi, m_\xi) = \mathbb{P}^!_{(\xi,m_\xi)}(F(\xi, m_\xi; h)), \tag{9}$$

and $F(\xi, m_\xi; h)$ is the set of marked point patterns $x_M$, for which the point $\xi$ with mark $m_\xi$ is retained in the thinned process,

$$F(\xi, m_\xi; h) := \{x_M \mid \forall (\eta, m_\eta) \in x_M : \|\xi - \eta\| > h \vee m_\eta > m_\xi\}.$$

A proof of (8) can be found in the Appendix.

Second- and higher-order Palm retention probabilities can be treated analogously. Let $\rho_{\text{th}}^{(2)}$ and $\rho^{(2)}$ denote the second-order product density of $X_{\text{th}}$ and $X$, respectively. Then, the second-order retention probability is defined by

$$p_{\text{ret}}^{(2)}(\xi, \eta) := \rho_{th}^{(2)}(\xi, \eta)/\rho^{(2)}(\xi, \eta). \tag{10}$$

As in (8) and (9), we have

$$p_{\text{ret}}^{(2)}(\xi, \eta) = \int_0^1 \int_0^1 p_{\text{ret}}^{(2)}((\xi, m_\xi)(\eta, m_\eta)) \, dm_\xi \, dm_\eta \tag{11}$$

with

$$p_{\text{ret}}^{(2)}((\xi, m_\xi), (\eta, m_\eta)) = \mathbb{1}(\|\xi - \eta\| > h)\mathbb{P}^!_{(\xi,m_\xi),(\eta,m_\eta)}(F(\xi, m_\xi; h) \cap F(\eta, m_\eta; h)). \tag{12}$$

Here, $\mathbb{P}^!_{(\xi,m_\xi)(\eta,m_\eta)}$ is the two-point reduced Palm distribution for $X_M$ (Hanisch, 1982, p. 172). A proof of (11) can be found in the Appendix.

In cases where it is possible to calculate the retention probabilities, the intensity and second-order product density of the thinned process can be obtained by multiplying the ones for the original process with the retention probabilities. If $X_{th}$ is obtained by thinning a stationary and isotropic point process $X$, $X_{th}$ will likewise be stationary and isotropic, and the intensity and second-order product density are given by

$$\rho_{th} = p_{ret}\rho \quad \text{and} \quad \rho_{th}^{(2)}(r) = p_{ret}^{(2)}(r)\rho^{(2)}(r). \tag{13}$$

## 4. Matérn thinned Cox processes

A Matérn (type II) thinned Cox process is defined as follows.

**Definition 1.** A *Matérn thinned Cox process* with parameters $(\Lambda, h)$ is a point process given by

$$\text{MTCP}(\Lambda, h) := \{\xi \in X \mid \forall (\eta, m_\eta) \in X_M, \eta \neq \xi : \|\xi - \eta\| > h \vee m_\eta > m_\xi\}, \tag{14}$$

where $X$ is a Cox process with driving field $\Lambda$ and $X_M = \{(\xi, m_\xi) \mid \xi \in X, m_\xi \sim \text{unif}[0, 1)\}$ is the marked process with i.i.d. uniform marks.

Sample realizations of Matérn thinned MCPs and TPs in $\mathbb{R}^2$ are shown in Fig. 1. For Matérn thinned Cox processes, there exist (at least) two possible methods to derive the intensity function and the second-order product density for the thinned process. Either one can try to calculate the retention probabilities in (8) and (11) directly, or one can use the standard approach for Cox processes, and first condition on the random intensity function, and then calculate the retention probabilities given the random intensity function. The first approach is in general more difficult, but as we shall see in Theorem 2 it is possible to find explicit expressions for the first-order retention probability for Matérn thinned SNCPs. A proof of Theorem 2 can be found in the Appendix.

**Theorem 2** (*Retention Probabilities of Matérn Thinned SNCPs*)**.** *The first-order Palm retention probability* (8) *of* $\text{MTCP}(\Lambda, h)$ *with shot noise driving field* $\Lambda$ *specified by* (4) *is given by*

$$p_{ret}(\xi) = \int_0^1 \exp\left(-\int_{\mathbb{R}^d \times (0,\infty)} p_{\xi,m}(c, \gamma)\, \mathrm{d}\zeta(c, \gamma)\right)$$

$$\times \left(\int_{\mathbb{R}^d \times (0,\infty)} \gamma(1 - p_{\xi,m}(c, \gamma))\frac{k(c, \xi)}{\rho(\xi)}\, \mathrm{d}\zeta(c, \gamma)\right) \mathrm{d}m,$$

*where*

$$p_{\xi,m}(c, \gamma) = 1 - \exp\left(-m\int_{B(\xi,h)} \gamma k(c, \eta)\, \mathrm{d}\eta\right).$$

From Theorem 2 we can easily derive the retention probabilities for Matérn thinned NSPs.

**Corollary 3** (*Retention Probabilities of Matérn Thinned NSPs*)**.** *The first-order Palm retention probability* (8) *of* $\text{MTCP}(\Lambda, h)$ *with Neyman–Scott driving field* $\Lambda$ *specified by* (5) *with* $\mu, \kappa > 0$ *is given by*

$$p_{ret} = \int_0^1 \exp\left(-\kappa\int_{\mathbb{R}^d} p_m(c)\, \mathrm{d}c\right)\left(\int_{\mathbb{R}^d} (1 - p_m(c))k(-c)\, \mathrm{d}c\right) \mathrm{d}m,$$

*where*

$$p_m(c) = 1 - \exp\left(-\mu m\int_{B(o,h)} k(\eta - c)\, \mathrm{d}\eta\right).$$

**a**     Matérn thinned MCPs         Matérn thinned TPs

$R = 0.05, h = R/3, \mu = 10$         $\sigma = 0.02, h = 0.05/3, \mu = 14$

**b**

$R = 0.05, h = R/5, \mu = 7$         $\sigma = 0.02, h = 0.05/5, \mu = 8$

**c**

$R = 0.10, h = R/3, \mu = 18$         $\sigma = 0.04, h = 0.10/3, \mu = 35$

**d**

$R = 0.10, h = R/5, \mu = 8$         $\sigma = 0.04, h = 0.10/5, \mu = 8$
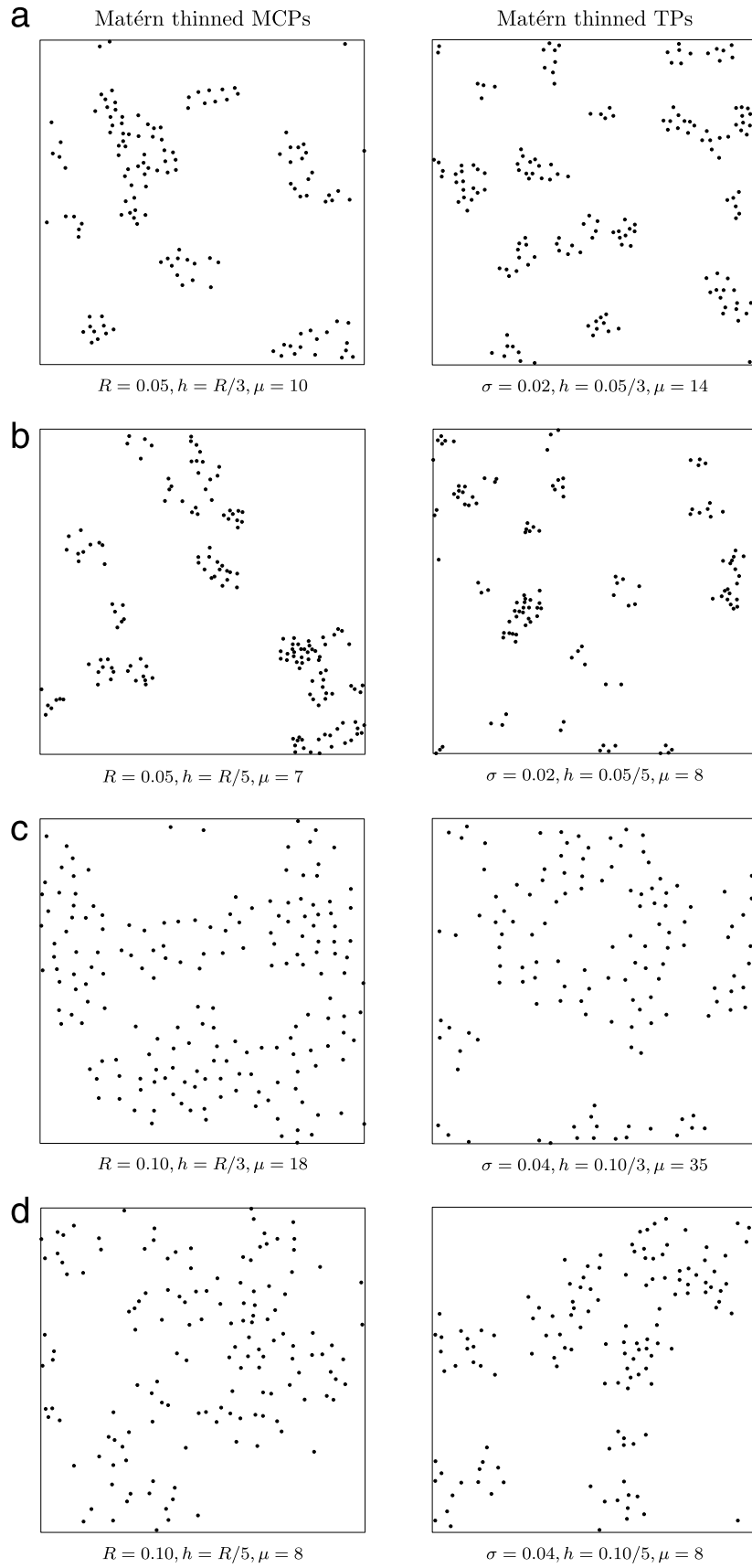
**Fig. 1.** Sample realizations of Matérn thinned MCPs (left) and Matérn thinned TPs (right) in observation window $W = [0, 1]^2$, with four choices of parameters, all with intensities $\rho_{\text{th}} \approx 150$ and $\kappa = 25$.

Further simplifications are possible in the case of isotropy, which is illustrated in Examples 3 and 4 for the Matérn thinned MCPs and Matérn thinned TPs, respectively.

**Example 3** (Example 1, *Continued*). For $d = 2$, the first-order Palm retention probability of a Matérn thinned MCP reduces to

$$p_{\text{ret}} = \int_0^1 \exp\left(-2\pi\kappa \int_0^{R+h} p_m(r) r \, dr\right)\left(2/R^2 \int_0^R (1 - p_m(r)) r \, dr\right) dm, \tag{15}$$

where

$$p_m(r) = p_m(\|\xi_r\|) = 1 - \exp(-\mu\, m|B(o, h) \cap B(\xi_r, R)|/(\pi R^2)).$$

**Example 4** (Example 2, *Continued*). For $d = 2$, the first-order Palm retention probability of a Matérn thinned TP reduces to

$$p_{\text{ret}} = \int_0^1 \exp\left(-2\pi\kappa \int_0^\infty p_m(r) r \, dr\right)\left(1/\sigma^2 \int_0^\infty (1 - p_m(r)) \exp(-r^2/(2\sigma^2)) r \, dr\right) dm, \tag{16}$$

where

$$p_m(r) = 1 - \exp\left(-\mu\, m/(2\pi\sigma^2) \int_{-h}^h \int_{-\sqrt{h^2-\eta_2^2}}^{\sqrt{h^2-\eta_2^2}} \exp(-(\eta_1^2 + (\eta_2 - r)^2)/(2\sigma^2)) \, d\eta_1 \, d\eta_2\right).$$

For general Matérn thinned Cox processes the intensity function and the second-order product density can be determined by conditioning on the random intensity function.

**Theorem 4** (*Intensity Function and Second-Order Product Density of Matérn Thinned Cox Processes*)**.** *Assume $\Lambda$ is a driving field of a Cox process that almost surely fulfils*

$$\Lambda(\xi) > 0 \Rightarrow \int_{B(\xi,r)} \Lambda(\eta) \, d\eta > 0, \quad \text{for all } r > 0, \tag{17}$$

*which can be regarded as the assumption, that the random intensity function almost surely has no singularities.*

*Then for* $\text{MTCP}(\Lambda, h)$*, the intensity function and second-order product density are given by*

$$\rho_{\text{th}}(\xi) = \text{E}\left[p_{\text{ret}|\Lambda}(\xi)\Lambda(\xi)\right] \tag{18}$$

*and*

$$\rho_{\text{th}}^{(2)}(\xi, \eta) = \text{E}\left[p_{\text{ret}|\Lambda}^{(2)}(\xi, \eta)\Lambda(\xi)\Lambda(\eta)\right] \tag{19}$$

*for* $\|\xi - \eta\| > h$*, otherwise* $0$*, where for* $\Lambda(\xi)$ *and* $\Lambda(\eta) > 0$*,*

$$p_{\text{ret}|\Lambda}(\xi) = \frac{1 - \exp(-\Omega_\xi)}{\Omega_\xi} \tag{20}$$

*and*

$$p_{\text{ret}|\Lambda}^{(2)}(\xi, \eta) = \frac{1 - \exp(-\Omega_\xi)}{\Omega_\xi \Omega_{\eta\setminus\xi}} + \frac{1 - \exp(-\Omega_\eta)}{\Omega_\eta \Omega_{\xi\setminus\eta}} - \frac{1 - \exp(-\Omega_{\xi\cup\eta})}{\Omega_{\xi\cup\eta}}\left(\frac{1}{\Omega_{\xi\setminus\eta}} + \frac{1}{\Omega_{\eta\setminus\xi}}\right), \tag{21}$$

*with* $\Omega_* = \int_{B_*} \Lambda(\tau) \, d\tau$*, and* $B_\xi = B(\xi, h)$*,* $B_{\xi\setminus\eta} = B_\xi\setminus B_\eta$*, and* $B_{\xi\cup\eta} = B_\xi \cup B_\eta$*.*

A proof of Theorem 4 may be found in the Appendix. If the original process is an inhomogeneous Poisson process (corresponding to a Cox process case, where $\Lambda(\xi) = \rho(\xi)$ is deterministic), the retention probabilities reduce to

$$p_{\text{ret}}(\xi) = \frac{1 - \exp(-\omega_\xi)}{\omega_\xi} \tag{22}$$

and

$$p_{\text{ret}}^{(2)}(\xi, \eta) = \frac{1 - \exp(-\omega_\xi)}{\omega_\xi \omega_{\eta \setminus \xi}} + \frac{1 - \exp(-\omega_\eta)}{\omega_\eta \omega_{\xi \setminus \eta}} - \left( \frac{1 - \exp(-\omega_{\xi \cup \eta})}{\omega_{\xi \cup \eta}} \right) \left( \frac{1}{\omega_{\xi \setminus \eta}} + \frac{1}{\omega_{\eta \setminus \xi}} \right) \tag{23}$$

for $\|\xi - \eta\| > h$, otherwise 0, where $\omega_* = \int_{b_*} \rho(\vartheta) \, d\vartheta$ and $b_\xi$ is defined in Theorem 4.

In the stationary and isotropic case (corresponding to the Matérn type II model), the intensity is given by

$$\rho_{\text{th}} = \frac{1 - \exp(-\rho \tau_h)}{\tau_h}, \tag{24}$$

and the second-order product density is given by

$$\rho_{\text{th}}^{(2)}(r) = \begin{cases} 0, & r \leq h, \\ \dfrac{2\Gamma_h(r)(1 - \exp(-\rho \tau_h)) - 2\tau_h(1 - \exp(-\mu \Gamma_h(r)))}{\tau_h \Gamma_h(r)(\Gamma_h(r) - \tau_h)}, & r > h, \end{cases} \tag{25}$$

where $\rho$ is the intensity of the process before thinning, $\tau_h = |B(o, h)|$, $\Gamma_h(r) = |B(o, h) \cup B(\xi_r, h)|$, $\xi_r \in \mathbb{R}^d$ with $\|\xi_r\| = r$. These formulae may also be found in e.g. Matérn (1960) and Stoyan et al. (1995).

## 5. Approximated first- and second-order densities

### 5.1. Derivation of approximations

The first- and second-order densities $\rho_{\text{th}}(\xi)$ and $\rho_{\text{th}}^{(2)}(\xi, \eta)$, given by Eqs. (18) and (19) in Theorem 4, are in general hard to evaluate numerically or express explicitly, since they require calculation of mean values as complex as

$$\mathbb{E}\left[ \frac{1 - \exp(-\Omega_*)}{\Omega_*} \Lambda(\xi) \right] \quad \text{and} \quad \mathbb{E}\left[ \frac{1 - \exp(-\Omega_*)}{\Omega_* \Omega_\diamond} \Lambda(\xi) \Lambda(\eta) \right], \qquad \Omega_{\cdot} = \int_{B_{\cdot}} \Lambda(\vartheta) \, d\vartheta,$$

where $B_*$ or $B_\diamond$ stand for sets of the form

$$B_\xi = B(\xi, h), \quad B_{\xi \cup \eta} = B_\xi \cup B_\eta \quad \text{and} \quad B_{\xi \setminus \eta} = B_\xi \setminus B_\eta.$$

When both clustering and hard core behaviour are visible in a point pattern, the hard core distance will always appear small compared to the diameter of the clusters. Thus, in an appropriate model, the driving field $\Lambda$ will be often appear constant or slowly varying on a scale of the hard core distance $h$. It is therefore sensible to approximate $\Lambda(\vartheta)$ for $\vartheta \in B(\xi, h)$ by $\Lambda(\xi)$, which yields the following expressions for the integrals $\Omega_{\cdot}$:

$$\Omega_\xi = \int_{B_\xi} \Lambda(\vartheta) d\vartheta \approx \Lambda(\xi) \tau_h, \tag{26}$$

$$\Omega_{\xi \cup \eta} = \int_{B_{\xi \cup \eta}} \Lambda(\vartheta) d\vartheta \approx (\Lambda(\xi) + \Lambda(\eta)) \Gamma_h(r)/2, \tag{27}$$

$$\Omega_{\xi \setminus \eta} = \int_{B_{\xi \setminus \eta}} \Lambda(\vartheta) d\vartheta \approx \Lambda(\xi)(\Gamma_h(r) - \tau_h), \tag{28}$$

where $r = \|\xi - \eta\|$, $\tau_h = |B(o, h)|$, $\Gamma_h(r) = |B(o, h) \cup B(\xi_r, h)|$, $\xi_r \in \mathbb{R}^d$, with $\|\xi_r\| = r$.

We will thus approximate the intensity function and second-order product density of a Matérn thinned Cox process by

$$\rho_{\text{a}}(\xi) := \mathbb{E}[p_{\text{a}|\Lambda}(\xi) \Lambda(\xi)], \tag{29}$$

and

$$\rho_{\mathrm{a}}^{(2)}(\xi, \eta) := \mathrm{E}[p_{\mathrm{a}|\Lambda}^{(2)}(\xi, \eta)\Lambda(\xi)\Lambda(\eta)], \tag{30}$$

for $\|\xi - \eta\| > h$, otherwise 0, where $p_{\mathrm{a}|\Lambda}(\xi)$ and $p_{\mathrm{a}|\Lambda}^{(2)}(\xi, \eta)$ are obtained by replacing $\Omega_*$ in the retention probabilities (20) and (21) with the approximations (26)–(28). Simple formulae for the approximated intensity function and second-order product density can then be found.

**Theorem 5** (*Approximated First- and Second-Order Densities for Matérn Thinned Cox Processes*)**.** *The approximated intensity function and second-order product density of a* $\mathrm{MTCP}(\Lambda, h)$*, defined by* (29) *and* (30), *are given by*

$$\rho_{\mathrm{a}}(\xi) = \frac{1 - \mathrm{E}[\exp(-\Lambda(\xi)\tau_h)]}{\tau_h} \tag{31}$$

*and*

$$\rho_{\mathrm{a}}^{(2)}(\xi, \eta) = \frac{\Gamma_h(r)(2 - \mathrm{E}[\exp(-\Lambda(\xi)\tau_h)] - \mathrm{E}[\exp(-\Lambda(\eta)\tau_h)]) - 2\tau_h(1 - \mathrm{E}[\exp(-(\Lambda(\xi) + \Lambda(\eta))\Gamma_h(r)/2)])}{\tau_h \Gamma_h(r)(\Gamma_h(r) - \tau_h)} \tag{32}$$

*for* $\|\xi - \eta\| = r > h$, *otherwise* 0.

The proof of Theorem 5 may be found in the Appendix. From Theorem 5 we can derive the approximated intensity function and second-order product density for Matérn thinned SNCPs.

**Corollary 6** (*Approximated First- and Second-Order Densities for Matérn thinned SNCPs*)**.** *For SNCPs with driving field* $\Lambda$ *given by* (4), *the expected values in Theorem 5 are for* $\|\xi - \eta\| = r$ *given by*

$$\mathrm{E}[\exp(-\Lambda(\xi)\tau_h)] = \exp(-a(\xi)), \tag{33}$$
$$\mathrm{E}[\exp(-(\Lambda(\xi) + \Lambda(\eta))\Gamma_h(r)/2)] = \exp(-b(\xi, \eta)), \tag{34}$$

*where*

$$a(\xi) = \int_{\mathrm{R}^d \times (0,\infty)} (1 - \exp(-\gamma k(c, \xi)\tau_h)) \, \mathrm{d}\zeta(c, \gamma),$$

$$b(\xi, \eta) = \int_{\mathrm{R}^d \times (0,\infty)} (1 - \exp(-\gamma \{k(c, \xi) + k(c, \eta)\}\Gamma_h(r)/2)) \, \mathrm{d}\zeta(c, \gamma).$$

*For stationary Matérn thinned NSPs with driving field* $\Lambda$ *given by* (5), $a(\xi)$ *and* $b(\xi, \eta)$ *reduce to*

$$a = \kappa \int_{\mathrm{R}^d} (1 - \exp(-\mu k(\vartheta)\tau_h)) \, \mathrm{d}\vartheta,$$

$$b(\xi - \eta) = \kappa \int_{\mathrm{R}^d} (1 - \exp(-\mu \{k(\xi - \eta + \vartheta) + k(\vartheta)\}\Gamma_h(r)/2)) \, \mathrm{d}\vartheta.$$

The formulae become particularly simple for stationary and isotropic Matérn thinned NSPs, which is illustrated in Examples 5 and 6 for Matérn thinned MCPs and Matérn thinned TPs in the plane.

**Example 5** (Example 1, *Continued*)**.** For $d = 2$, $k(\xi) = \mathbb{1}(\|\xi\| \leq R)/(\pi R^2)$ and $b(\xi - \eta) = b(\|\xi - \eta\|)$, say, $a$ and $b$ in Corollary 6 for Matérn thinned MCPs reduce to

$$a = \kappa \pi R^2 (1 - \exp(-\mu h^2/R^2)) \quad \text{and}$$

$$b(r) = 2\kappa(1 - \exp(-\mu \Gamma_h(r)/(2\pi R^2)))(\Gamma_R(r) - \pi R^2)$$
$$\quad + \kappa(1 - \exp(-\mu \Gamma_h(r)/(\pi R^2)))(2\pi R^2 - \Gamma_R(r)).$$

**Example 6** ([Example 2](), *Continued*). For $d = 2$, $k(\xi) = \exp(-\|\xi\|^2/(2\sigma^2))/(2\tau_\sigma)$ and $b(\xi - \eta) = b(\|\xi - \eta\|)$, say, $a$ and $b$ in [Corollary 6]() for Matérn thinned TPs reduce to

$$a = 2\pi\kappa \int_0^\infty (1 - \exp[-\mu\, h^2/(2\sigma^2)\exp(-r^2/(2\sigma^2))])r\, \mathrm{d}r,$$

$$b(r) = \kappa \int_{\mathrm{R}} \int_{\mathrm{R}} (1 - \exp(-\mu\{\exp(-(\vartheta_1^2 + \vartheta_2^2)/(2\sigma^2))$$
$$+ \exp(-(\vartheta_1^2 + (\vartheta_2 - r)^2)/(2\sigma^2))\}\Gamma_h(r)/(4\pi\sigma^2)))\, \mathrm{d}\vartheta_1\, \mathrm{d}\vartheta_2.$$

Combining [Theorem 5]() and [Corollary 6](), the approximated intensity function and the second-order product density for stationary and isotropic Matérn thinned NSPs simplify to

$$\rho_{\mathrm{a}} = \frac{1 - \exp(-a)}{\tau_h}, \tag{35}$$

and for $r > h$,

$$\rho_{\mathrm{a}}^{(2)}(r) = \frac{2\Gamma_h(r)(1 - \exp(-a)) - 2\tau_h(1 - \exp(-b(r)))}{\tau_h \Gamma_h(r)(\Gamma_h(r) - \tau_h)}. \tag{36}$$

The approximated pair correlation function is defined by $g_{\mathrm{a}}(\xi, \eta) := \rho_{\mathrm{a}}^{(2)}(\xi, \eta)/(\rho_{\mathrm{a}}(\xi)\rho_{\mathrm{a}}(\eta))$. For stationary and isotropic Matérn thinned NSPs, we find

$$g_{\mathrm{a}}(r) = \frac{\rho_{\mathrm{a}}^{(2)}(r)}{\rho_{\mathrm{a}}^2} = \frac{2\Gamma_h(r)\tau_h(1 - \exp(-a)) - 2\tau_h^2(1 - \exp(-b(r)))}{\Gamma_h(r)(\Gamma_h(r) - \tau_h)(1 - \exp(-a))^2}. \tag{37}$$

### 5.2. Properties of the approximations in Matérn thinned MCPs and Matérn thinned TPs

In order to examine the quality of the approximations $\rho_{\mathrm{a}}$ and $g_a$ in [(35)]() and [(37)](), Matérn thinned MCPs and Matérn thinned TPs have been considered in the plane for different combinations of the parameters.

The approximated intensity $\rho_{\mathrm{a}}$ is compared to the theoretical intensity $\rho_{\mathrm{th}} = p_{\mathrm{ret}}\rho$, where $p_{\mathrm{ret}}$ is found from [(15)]() in [Example 3]() for Matérn thinned MCPs and [(16)]() in [Example 4]() for Matérn thinned TPs (obtained by numerical integration). The results are summarized in [Fig. 2](), which shows the relative bias as a function of the hard core distance $h$, relative to the cluster radius $R$, for the nine combinations of parameters $(\kappa, \mu)$ indicated in the figure. For Matérn thinned MCPs, $R$ is simply the radius for which the kernel is positive, while for the Matérn thinned TPs, $R$ is determined such that

$$\int_{\mathrm{R}^2} k(\|\xi\|)\mathbb{1}(\|\xi\| \le R)\, \mathrm{d}\xi \approx 95\%.$$

For Matérn thinned MCPs, $\rho_{\mathrm{a}}$ yields a negative, but relatively small bias, when $h$ is reasonably small compared to $R$. For Matérn thinned TPs, $\rho_{\mathrm{a}}$ approximates $\rho_{\mathrm{th}}$ very well. To assess the quality of the approximation for the pair correlation function (pcf), $g_a$ is compared to an estimated theoretical pcf obtained by averaging 500 empirical estimates from sample realizations. We consider four combinations of parameter values for each of the two types of processes, Matérn thinned MCPs and Matérn thinned TPs. See [Fig. 1]() for parameter details and sample realizations.

For the simulated point patterns from each of the chosen combinations of parameters, non-parametric estimates $\hat{g}(r)$ of the pcfs were found using the `pcf.ppp` function from the `spatstat` package in R ([Baddeley and Turner, 2005]()), with default values of $r$, the Epanechnikov kernel, fixed kernel bandwidth parameter `bw` $= h/\sqrt{5}$, `translate` for the choice of edge correction and `divisor = d`. Since the estimator $\hat{g}(r)$ is based on kernel smoothing, it is biased. In particular, it will assign values $\hat{g}(r) > 0$ for $r < h$. One can show that $\hat{g}(r)$ with the above choice of parameters
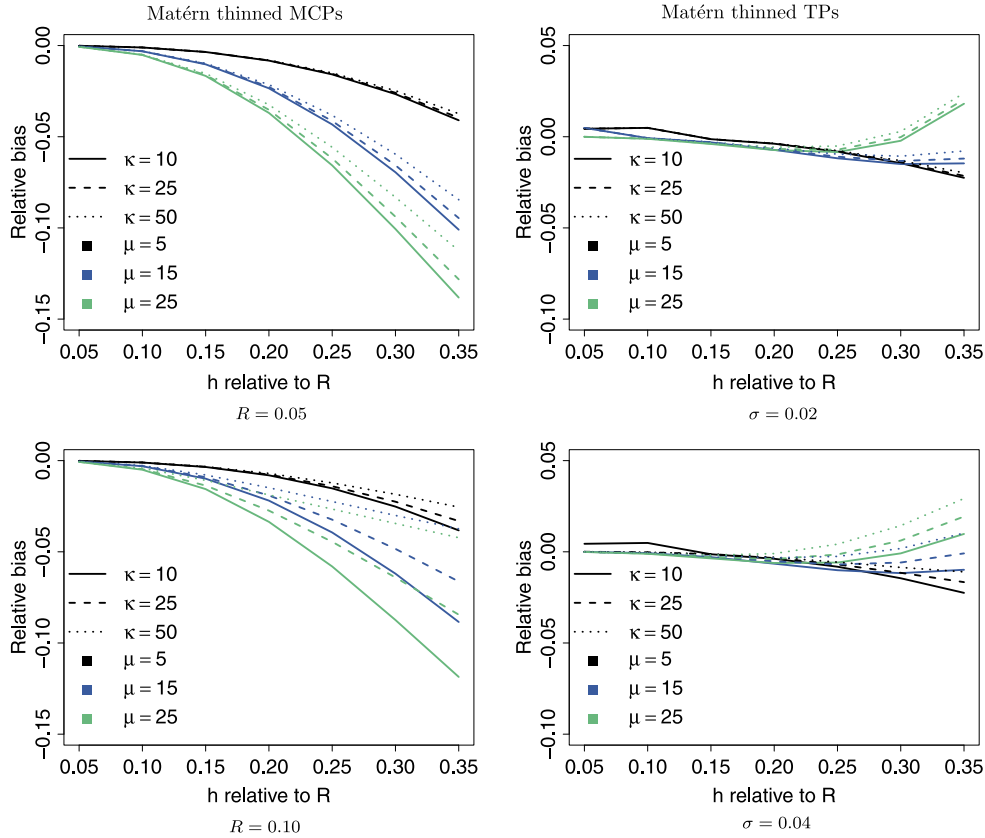
**Fig. 2.** The relative bias of the approximated intensities of Matérn thinned MCPs (left) and Matérn thinned TPs (right) from the theoretical intensities obtained by numerical integration. The bias is shown as a function of the hard core distance $h$, relative to the cluster radius $R$. For Matérn thinned MCPs, $R$ is simply the radius for which the kernel is positive, while for Matérn thinned TPs, a comparable $R$ is used. For details, see text. In each sub-figure, nine different models are considered, corresponding to all the combinations of three cluster intensities, $\kappa = 10$, 25 and 50 (unbroken, dashed and dotted lines, respectively), and three mean number of points in a cluster before thinning, $\mu = 5$, 15 and 25 (black, blue and green lines, respectively).

is a ratio unbiased estimator for the convolution of the true pair correlation function with the kernel, see Fiksel (1988). We therefore consider the kernel smoothed version

$$\widetilde{g_a}(r) := \int g_a(s)k(r-s)ds \tag{38}$$

when judging the quality of approximation by simulation.

In Fig. 3, $g_a$ and $\widetilde{g_a}$ are compared to estimated pcfs for Matérn thinned MCPs (left) and Matérn thinned TPs (right). The grey areas mark pointwise central 95%-regions. The black full drawn lines are the average estimated pcfs, and the red stippled and full drawn lines represent $g_a$ and $\widetilde{g_a}$, respectively. The smooth versions $\widetilde{g_a}$ of $g_a$ are very similar to the average estimated pcfs in all cases considered, in particular for the Matérn thinned MCPs.

## 6. Applications

In this section, we consider point patterns describing the position of megakaryocytes in bone marrow biopsy sections for a control group and three case groups with haematopoietic stem cell disorders, essential thrombocythemia (ET), polycythaemia vera (PV) and primary myelofibrosis (PMF). The diseases result in an increased number of megakaryocytes and different levels of clustering, which is used as diagnostic indicators in pathology. Pathologists determine the level of clustering (i.e. no, "loose" and "dense" clustering) by a visual judgement following different vaguely described rules, see e.g. Madelung et al. (2013) and Vytrva et al. (2014) for further details. To the best of our knowledge, no precise quantitative measure for the degree of clustering between megakaryocytes has been used in the existing literature within the field of pathology and this is therefore the first time, point process
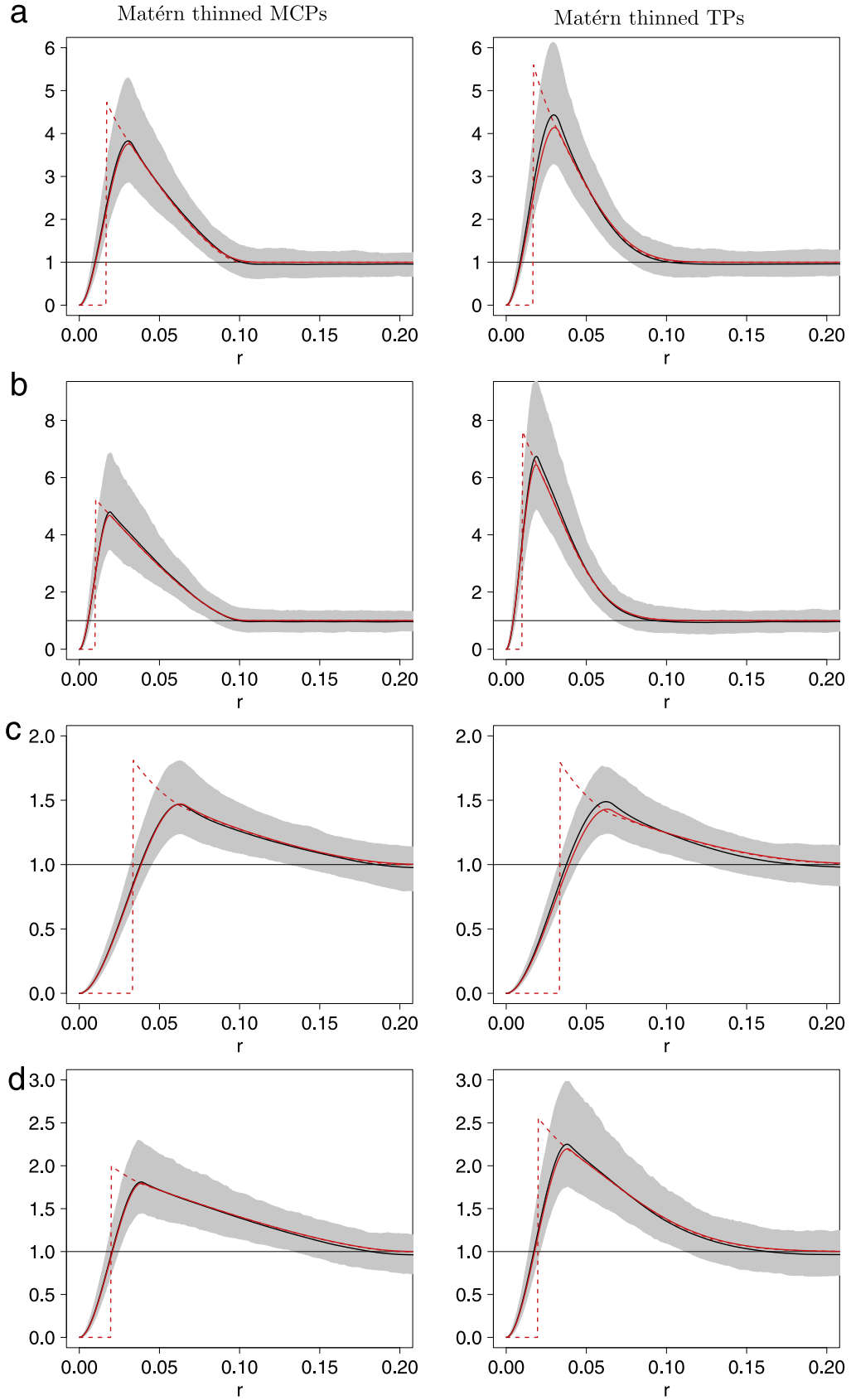
**Fig. 3.** For Matérn thinned MCPs (left) and Matérn thinned TPs (right) and each of the parameter combinations (a)–(d) shown in Fig. 1, estimates of the pcfs are shown together with the approximated pcfs $g_a$ (red stippled lines). The estimates of the pcfs are represented by the grey areas, which mark pointwise central 95%-regions, based on 500 simulations. The average pcfs, based on these simulations, are also shown (black full drawn lines) as well as the smooth versions $\widetilde{g}_a$ of $g_a$ (red full drawn lines).
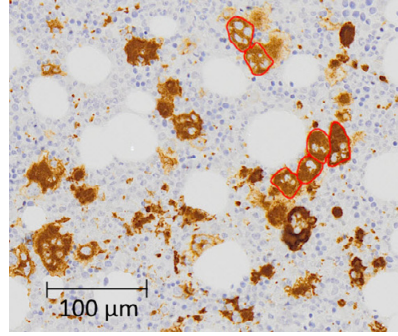
**Fig. 4.** A small part of a bone marrow biopsy section at high magnification. The bone marrow consists mostly of blood forming cells, among them megakaryocytes, a few of which have been outlined with red. On the section, huge white fat cells are also seen.

theory has been applied to measure the level of clustering in this application. One means to get to a quantitative characterization of the spatial arrangement of the cells are summary functions, such as the pair correlation function considered here. Another option is to fit a point process model that includes parameters directly linked to clustering. A similar approach has been taken by Mattfeldt et al. (2007) for the description of capillary profiles in pathological sections by a Strauss hard core model.

Megakaryocytes are large cells, which are easy to distinguish from other cells. See Fig. 4 for an illustrative example of megakaryocytes at high magnification. We considered 1, 5, 3 and 4 biopsies for the control, ET, PMF and PV groups, respectively. One section per biopsy was scanned and divided into sub-sections, resulting in 9, 13, 10 and 10 sub-sections, respectively. These were chosen as parts of the sections, where the tissue was fairly regular and unbroken. For the case groups, the sub-sections were of size 1540 $\mu$m $\times$ 1540 $\mu$m, but for the control group we chose parts of different sizes to include enough cells for the point process analysis. In Fig. 5, sub-section examples are shown for the four groups, where the tissue and megakaryocytes have been outlined. The distinct clustering behaviour and low density in the pattern formed by the centres of the megakaryocyte profiles suggest that Matérn thinned Cox processes are suitable models. In order to reveal differences between the groups, we fitted Matérn thinned MCPs and Matérn thinned TPs to the data. The hard core distance $h$ is considered as a nuisance parameter. Considering minimum point-to-point distances, we chose $h = 25$ $\mu$m. A preliminary investigation showed that the results were not affected substantially by choosing other reasonable values of $h$. To estimate the other model parameters, we applied the minimum contrast method in combination with the pcf (Diggle, 1983, Chapter 6). More concretely, the estimated parameter vector $\hat{\theta}$ is found as the minimizer of the discrepancy measure

$$D(\theta) = \int_{r_1}^{r_2} (\hat{g}(r)^q - g(r; \theta)^q)^2 \, dr$$

between the theoretical function $g(\cdot; \theta)$ and its empirical non-parametric counterpart $\hat{g}$, where $r_1$, $r_2$ and $q$ are tuning parameters specified below. We estimated the pair correlation functions for each group as average of pcfs estimated on the sub-sections, using `spatstat` as described in Section 5.2. To avoid bias due to kernel estimation, we replaced the theoretical pcf with the convoluted version $\widetilde{g}_a$ given in (38). To find the parameter vector $\hat{\theta}$ that minimizes $D(\hat{\theta})$, we used the `spatstat` function `mincontrast`, which is based on the Nelder–Mead (1965) algorithm. The set of candidate parameters for the search of $\hat{\theta}$ was restricted to parameters that yield an approximated intensity $\rho_a$ calculated from (35) to be equal to the observed one. The tuning parameters were set to $r_1 = 0$, $r_2 = 200$ $\mu$m, and $q = 1$. The value of $q$ can be used to add weight to the part of the functions most important for the analysis. The short range clustering seems to be of most importance in the pathological application, thus we chose $q = 1$.

In Table 1 an overview of the fitted parameters is given for the Matérn thinned MCP and Matérn thinned TP models. Besides the model parameters we also show the usual non-parametric estimate for the intensity $\hat{\rho}_{th} = \#\text{Cell}/|W|$, the estimate for the mean number of cells in a cluster after thinning $\hat{\mu}_{th} = \hat{\rho}_{th}/\hat{\kappa}$ and the discrepancy measure.
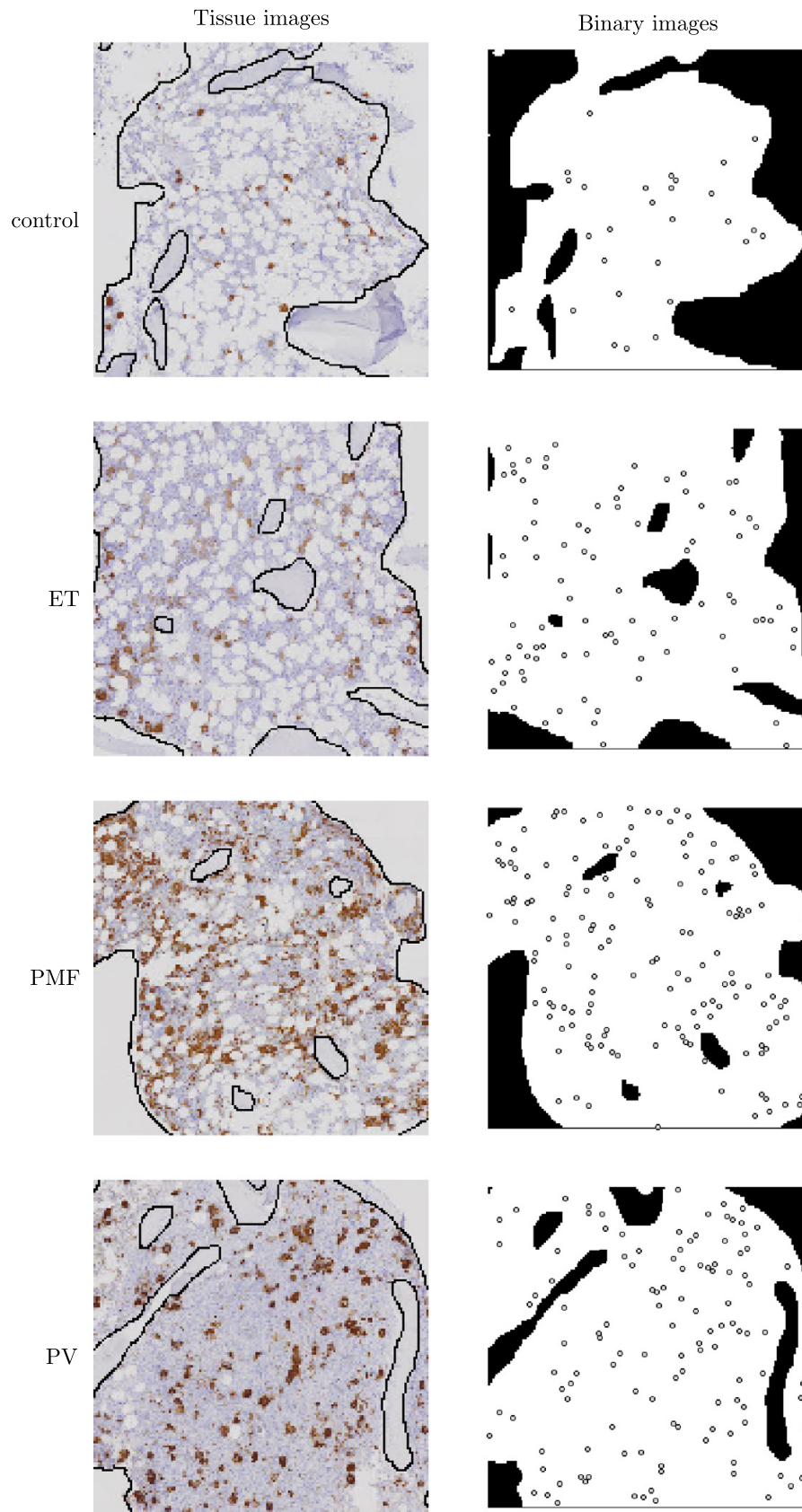
**Fig. 5.** Sub-section examples (1540 μm × 1540 μm) of the control, ET, PMF and PV groups. Tissue images and binary images are shown, where the tissue and cells have been marked. The cells have been marked manually and they are represented by balls of radius $r = h/2 = 12.5$ μm.

**Table 1**

Estimated parameters from fitting Matérn thinned MCPs and Matérn thinned TPs to the observed point patterns using the estimated pcf for each group. Besides the model parameters we also estimated the following quantities: the usual non-parametric estimate for the intensity $\hat{\rho}_{\text{th}} = \#\text{Cell}/|W|$, the estimate for the mean number of cells in a cluster after thinning $\hat{\mu}_{\text{th}} = \hat{\rho}_{\text{th}}/\hat{\kappa}$ and the discrepancy measure.

| Estimated parameters | Groups and Matérn thinned models | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Control | | ET | | PMF | | PV | |
| | MC | T | MC | T | MC | T | MC | T |
| $\hat{\kappa} \times 10^5$ (per $(\mu m)^2$) | 3.95 | 3.44 | 2.59 | 2.57 | 1.48 | 1.48 | 2.38 | 2.39 |
| $\hat{R}$ or $\hat{\sigma}$ ($\mu m$) | 56.29 | 34.83 | 86.78 | 45.67 | 76.40 | 38.61 | 113.16 | 59.28 |
| $\hat{\mu}$ | 0.43 | 0.48 | 1.54 | 1.54 | 4.16 | 4.11 | 2.47 | 2.50 |
| $\hat{\rho}_{\text{th}} \times 10^5$ (per $(\mu m)^2$) | 1.58 | 1.58 | 3.62 | 3.62 | 4.74 | 4.74 | 5.25 | 5.25 |
| $\hat{\mu}_{\text{th}}$ | 0.40 | 0.46 | 1.40 | 1.41 | 3.21 | 3.21 | 2.21 | 2.20 |
| Discrepancy measure | 2.57 | 2.16 | 0.50 | 0.37 | 2.82 | 1.98 | 0.40 | 0.41 |

In Fig. 6, we consider the four groups (top to bottom) and the fitted Matérn thinned MCPs (left) and Matérn thinned TPs (right). The estimated pcf for each group from the observed point patterns is shown as black full drawn lines. The approximated pcfs $g_a$ for the Matérn thinned MCPs and Matérn thinned TPs with fitted parameters are also shown, as well as the smooth versions $\widetilde{g}_a$ (red stippled and full drawn lines, respectively). The grey areas mark pointwise central 95%-regions based on 500 simulations from the fitted Matérn thinned MCPs and Matérn thinned TPs.

The fitted Matérn thinned MCPs and Matérn thinned TPs seem to capture most of the behaviour of the pcfs from the observed point pattern, although we see that the models slightly underestimate the level of clustering for short and large range distances and slightly overestimate the level of clustering for medium range distances. The Matérn thinned TPs result in slightly better fit, except for the PV group, where the models are almost equally good (see the discrepancy measure in Table 1).

All groups indicate a clustering behaviour, but the control group can easily be distinguished from the case groups due to the low intensity. The PMF group is the most clustered group and seems to have a much higher mean number of cells per cluster than the rest of the groups. The ET and PV groups show similar behaviour of the pcfs, with ET having slightly smaller clusters with fewer cells per cluster.

## 7. Discussion

In the present paper, we have defined the class of Matérn thinned Cox processes and derived formulae for first- and second-order characteristics. Furthermore, we have suggested approximations to simplify calculations. For Matérn thinned MCPs and Matérn thinned TPs, the accuracy of the approximations was examined by numerical integration (for the intensity) and by simulations (for the pair correlation function), which indicates that the approximations capture most of the true behaviour of the models. The approximations enable simple fitting of the models, using e.g. the minimum contrast method, without the need of simulations. Model fitting using both Matérn thinned MCPs and Matérn thinned TPs were performed in a study of the pattern of megakaryocytes under different diseases. The models fitted very well the observed behaviour of the estimated pcfs, resulting in a valid analysis procedure to detect clustering of these cells, which is missing in the literature within the field of pathology. As only a pilot study with few sections was available, a future larger study must be performed to establish more robust conclusions for the groups under consideration. The fitted Matérn thinned MCPs and Matérn thinned TPs are very similar, and as numerical integration is only necessary for the approximations of the more complicated Matérn thinned TPs, the Matérn thinned MCPs could very well be preferred.

There are still some open questions regarding the approximations. It may be possible to obtain results regarding upper bounds for the deviation between the theoretical and approximated quantities. Furthermore, a thorough, future simulation study of the Matérn thinned Cox processes is needed to investigate in detail for which choices of parameters efficient estimation of the parameters is possible.
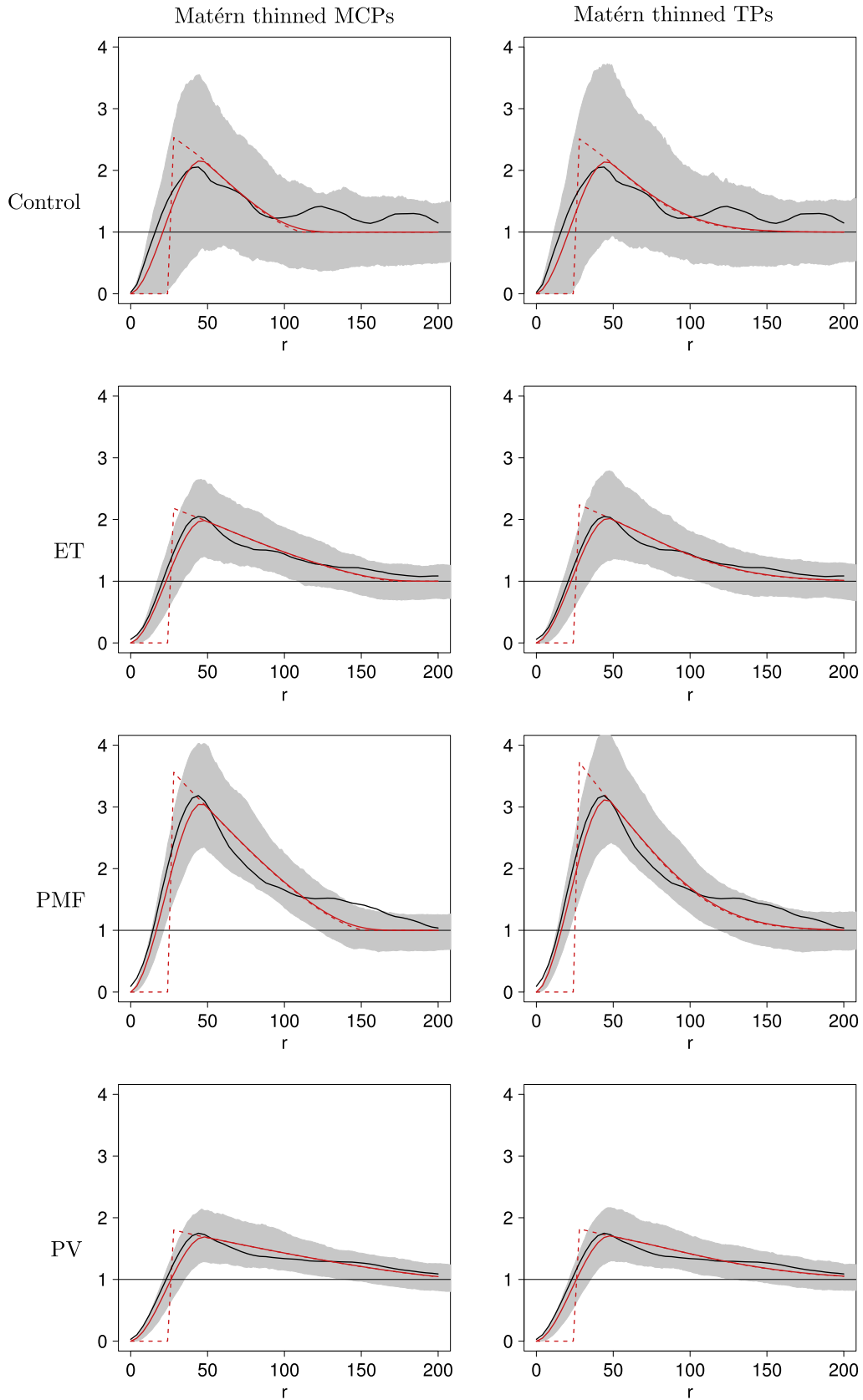
**Fig. 6.** Group estimated pcfs from the observed point patterns are shown (black full drawn lines) for the control, ET, PMF and PV groups. The approximated pcfs as well as the ones after smoothing (see details in the text) of the fitted models are shown (red stippled and full drawn lines, respectively) for Matérn thinned MCPs (left) and Matérn thinned TPs (right). The grey areas mark pointwise central 95%-regions based on 500 simulations from the fitted models.

**Acknowledgements**

**Appendix.  Proofs**

**Proof of (8).** To show (8), let $X_{M\text{th}}$ denote the thinned marked process. Then, the intensity measure of $X_{\text{th}}$ is for $B \in \mathcal{B}$ given by

$$\alpha_{X_{\text{th}}}(B) = \alpha_{X_{M\text{th}}}(B \times [0, 1)) = \mathrm{E} \sum_{(\xi, m_\xi) \in X_M} h((\xi, m_\xi), X_M \backslash (\xi, m_\xi)),$$

where

$$h((\xi, m_\xi), X_M \backslash (\xi, m_\xi)) = \mathbb{1}(\xi \in B)\mathbb{1}(X_M \backslash (\xi, m_\xi) \in F(\xi, m_\xi; h)).$$

Applying the Campbell–Mecke theorem (Møller and Waagepetersen, 2004, p. 249) on the marked process $X_M$ and the function $h((\xi, m_\xi), X_M \backslash (\xi, m_\xi))$, we obtain

$$\alpha_{\text{th}}(B) = \int_B \int_0^1 \mathbb{P}^!_{(\xi, m_\xi)}(F(\xi, m_\xi; h)) \, \mathrm{d}m_\xi \, \rho(\xi) \, \mathrm{d}\xi.$$

This equality holds for all $B \in \mathcal{B}$, and therefore we have $\rho_{\text{th}}(\xi) = p_{\text{ret}}(\xi)\rho(\xi)$ for Lebesgue almost all $\xi$, where $p_{\text{ret}}(\xi)$ is given by (8).    □

**Proof of (11).** To derive (11), we rewrite the second factorial moment measure of the thinned process for $B_1, B_2 \in \mathcal{B}$

$$\alpha_{\text{th}}^{(2)}(B_1 \times B_2) = \mathrm{E} \sum_{(\xi, m_\xi), (\eta, m_\eta) \in X_M}^{\neq} h((\xi, m_\xi), (\eta, m_\eta), X_M \backslash \{(\xi, m_\xi), (\eta, m_\eta)\}),$$

where

$$\begin{aligned} & h((\xi, m_\xi), (\eta, m_\eta), X_M \backslash \{(\xi, m_\xi), (\eta, m_\eta)\}) \\ & \quad = \mathbb{1}((\xi, \eta) \in B_1 \times B_2)\mathbb{1}(\|\xi - \eta\| > h)\mathbb{1}(X_M \backslash \{(\xi, m_\xi), (\eta, m_\eta)\} \\ & \quad \in F(\xi, m_\xi; h) \cap F(\eta, m_\eta; h)). \end{aligned}$$

Using Hanisch (1982, (2.4) and (2.6)) with $n = 2$, we find

$$\begin{aligned} \alpha_{\text{th}}^{(2)}(B_1 \times B_2) = \int_{\mathrm{R}^d} \int_0^1 \int_{\mathrm{R}^d} \int_0^1 & \mathrm{E}^!_{(\xi, m_\xi)(\eta, m_\eta)} h((\xi, m_\xi), (\eta, m_\eta), X_M) \\ & \times \alpha_M^{(2)}(\, \mathrm{d}(\xi, m_\xi), \, \mathrm{d}(\eta, m_\eta)). \end{aligned}$$

Since $\alpha_M^{(2)}(\, \mathrm{d}(\xi, m_\xi), \, \mathrm{d}(\eta, m_\eta)) = \rho^{(2)}(\xi, \eta) \, \mathrm{d}m_\xi \, \mathrm{d}m_\eta \, \mathrm{d}\xi \, \mathrm{d}\eta$, we get (11).    □

**Proof of Theorem 2.** Recall that $X = \cup_{(c,\gamma) \in \Phi} X_{(c,\gamma)}$, where $\Phi$ is a Poisson process and $X_{(c,\gamma)}|\Phi$ are independent Poisson processes on $\mathrm{R}^d$ with intensity functions $\gamma k(c, \cdot)$.

The marked process $X_M = \{(\xi, m_\xi)|\xi \in X, m_\xi \sim \text{unif}[0, 1)\}$ can then similarly be written as $X_M = \cup_{(c,\gamma) \in \Phi} X_{(c,\gamma)M}$, where $X_{(c,\gamma)M} = \{(\xi, m_\xi)|\xi \in X_{(c,\gamma)}, m_\xi \sim \text{unif}[0, 1)\}$ and $X_{(c,\gamma)M}|\Phi$ are independent Poisson processes on $\mathrm{R}^d \times [0, 1)$ with intensity functions $h_{(c,\gamma)}(\xi, m) = \gamma k(c, \xi)$ for all $(c, \gamma) \in \Phi$. Thus, $X_M$ is again a SNCP, and the retention probability can be obtained using Proposition 1 on the marked SNCP $X_M$.

Let $X_{M(\xi,m)}$ denote the random process specified in Proposition 1. Then by definition $X_{M(\xi,m)}|(c_{\xi,m}, \gamma_{\xi,m})$ is a Poisson process with intensity function $\Lambda_{\xi,m}(\eta, \tilde{m}) = \gamma_{\xi,m}k(c_{\xi,m}, \eta)$, the distribution of $(c_{\xi,m}, \gamma_{\xi,m})$ is given by

$$\mathbb{P}((c_{\xi,m}, \gamma_{\xi,m}) \in D) = \frac{\int_D \gamma k(c, \xi)\, d\zeta(c, \gamma)}{\rho(\xi)}, \quad \text{for Borel sets } D \subseteq \mathbb{R}^d \times (0, \infty),$$

and $(c_{\xi,m}, \gamma_{\xi,m}, X_{M(\xi,m)})$ is independent of $(\Phi, X_M)$. Next, we notice that

$$F(\xi, m; h) = \{x_M | x_M \cap (B(\xi, h) \times [0, m]) = \emptyset\}.$$

Using the simple description of the reduced Palm distribution in Proposition 1, we obtain

$$\begin{aligned}
p_{\text{ret}}(\xi, m) &= \mathbb{P}^!_{(\xi,m)}(F(\xi, m; h)) \\
&= \mathbb{P}(X_M \cap (B(\xi, h) \times [0, m]) = \emptyset)\mathbb{P}(X_{M(\xi,m)} \cap (B(\xi, h) \times [0, m]) = \emptyset),
\end{aligned}$$

where the void probabilities are

$$\mathbb{P}(X_M \cap (B(\xi, h) \times [0, m]) = \emptyset) = \exp\left(-\int_{\mathbb{R}^d \times (0,\infty)} p_{\xi,m}(c, \gamma)\, d\zeta(c, \gamma)\right), \tag{A.1}$$

$$\mathbb{P}(X_{M(\xi,m)} \cap (B(\xi, h) \times [0, m]) = \emptyset) = \int_{\mathbb{R}^d \times (0,\infty)} \gamma(1 - p_{\xi,m}(c, \gamma))\frac{k(c, \xi)}{\rho(\xi)}\, d\zeta(c, \gamma), \tag{A.2}$$

with

$$p_{\xi,m}(c, \gamma) = 1 - \exp\left(-m \int_{B(\xi,h)} \gamma k(c, \eta)\, d\eta\right). \tag{A.3}$$

The result in (A.1) follows by using that $X_M$ is a SNCP together with (2) and (3) in the Poisson case. The result in (A.2) is derived using Proposition 1 as follows

$$\begin{aligned}
&\mathbb{P}(X_{M(\xi,m)} \cap (B(\xi, h) \times [0, m]) = \emptyset) \\
&= \mathbb{E}\,\mathbb{P}(X_{M(\xi,m)} \cap (B(\xi, h) \times [0, m]) = \emptyset | (c_{\xi,m}, \gamma_{\xi,m})) \\
&= \mathbb{E}\left[\exp\left(-m \int_{B(\xi,h)} \gamma_{\xi,m}k(c_{\xi,m}, \eta)\, d\eta\right)\right] \\
&= \int_{\mathbb{R}^d \times (0,\infty)} \gamma(1 - p_{\xi,m}(c, \gamma))\frac{k(c, \xi)}{\rho(\xi)}\, d\zeta(c, \gamma). \quad \square
\end{aligned}$$

**Proof of Theorem 4.** The intensity function and the second-order product density of $X_{\text{th}}$ can be written as

$$\rho_{\text{th}}(\xi) = \mathbb{E}\left[\Lambda_{\text{th}}(\xi)\right],$$

$$\rho_{\text{th}}^{(2)}(\xi, \eta) = \mathbb{E}\left[\Lambda_{\text{th}}^{(2)}(\xi, \eta)\right],$$

where $\Lambda_{\text{th}}(\xi)$ and $\Lambda_{\text{th}}^{(2)}(\xi, \eta)$ are the intensity function and second-order product density function of $X_{\text{th}}$ given $\Lambda$. Using (7) and (10), we get that

$$\Lambda_{\text{th}}(\xi) = p_{\text{ret}|\Lambda}(\xi)\Lambda(\xi),$$

$$\Lambda_{\text{th}}^{(2)}(\xi, \eta) = p_{\text{ret}|\Lambda}^{(2)}(\xi, \eta)\Lambda(\xi)\Lambda(\eta).$$

To find $p_{\text{ret}|\Lambda}(\xi)$ and $p_{\text{ret}|\Lambda}^{(2)}(\xi, \eta)$ it remains to derive the retention probabilities for inhomogeneous Poisson processes.

Let $\mathbb{P}^!_{(\xi,m_\xi)|\Lambda}$ be the reduced Palm distribution for the marked point process $X_M$ given $\Lambda$. Using (8), we find

$$p_{\text{ret}|\Lambda}(\xi) = \int_0^1 \mathbb{P}^!_{(\xi,m_\xi)|\Lambda}(F(\xi, m_\xi; h)) \, \mathrm{d}m_\xi. \tag{A.4}$$

Since the conditional distribution of $X_M$ given $\Lambda$ is inhomogeneous Poisson with intensity function $\Lambda_M(\xi, m_\xi) = \Lambda(\xi)$, we can use the Slivnyak–Mecke theorem (Møller and Waagepetersen, 2004, (3.7)), and get

$$\mathbb{P}^!_{(\xi,m_\xi)|\Lambda}(F(\xi, m_\xi; h)) = \mathbb{P}(X_M \cap (B(\xi, h) \times [0, m_\xi]) = \emptyset|\Lambda) = \exp(-\Omega_\xi m_\xi).$$

Inserting in (A.4), we get (20).

Likewise, using (11), we find

$$p_{\text{ret}}^{(2)}(\xi, \eta) = \int_0^1 \int_0^1 \mathbb{1}(\|\xi - \eta\| > h) \mathbb{P}^!_{(\xi,m_\xi),(\eta,m_\eta)|\Lambda}(F(\xi, m_\xi; h) \cap F(\eta, m_\eta; h)) \, \mathrm{d}m_\xi \, \mathrm{d}m_\eta,$$

where

$$\mathbb{P}^!_{(\xi,m_\xi),(\eta,m_\eta)|\Lambda}(F(\xi, m_\xi; h) \cap F(\eta, m_\eta; h))$$
$$= \mathbb{P}(X_M \cap (B(\xi, h) \times [0, m_\xi]) = \emptyset, X_M \cap (B(\eta, h) \times [0, m_\eta]) = \emptyset|\Lambda)$$
$$= \exp(-(\Omega_{\xi \setminus \eta} m_\xi + \Omega_{\eta \setminus \xi} m_\eta + \Omega_{\xi \cap \eta}(m_\xi \vee m_\eta))),$$

for $\|\xi - \eta\| > h$, otherwise 0, for which (21) similarly appears from integration over the marks. □

**Proof of Theorem 5.** Let $\Lambda(\xi), \Lambda(\eta) > 0$, and $\|\xi - \eta\| = r$. Using the approximations (26)–(28), the approximated retention probabilities reduce to

$$p_{\mathrm{a}|\Lambda}(\xi) = \frac{1-\exp(-\Lambda(\xi)\tau_h)}{\Lambda(\xi)\tau_h}$$

and

$$p_{\mathrm{a}|\Lambda}^{(2)}(\xi, \eta) = \frac{1-\exp(-\Lambda(\xi)\tau_h)}{\Lambda(\xi)\tau_h \Lambda(\eta)(\Gamma_h(r)-\tau_h)} + \frac{1-\exp(-\Lambda(\eta)\tau_h)}{\Lambda(\eta)\tau_h \Lambda(\xi)(\Gamma_h(r)-\tau_h)}$$
$$- \frac{1-\exp(-(\Lambda(\xi)+\Lambda(\eta))\Gamma_h(r)/2)}{(\Lambda(\xi)+\Lambda(\eta))\Gamma_h(r)/2} \left( \frac{1}{\Lambda(\xi)(\Gamma_h(r)-\tau_h)} + \frac{1}{\Lambda(\eta)(\Gamma_h(r)-\tau_h)} \right),$$
$$= \frac{\Gamma_h(r)(2-\exp(-\Lambda(\xi)\tau_h)-\exp(-\Lambda(\eta)\tau_h))-2\tau_h(1-\exp(-(\Lambda(\xi)+\Lambda(\eta))\Gamma_h(r)/2))}{\tau_h \Gamma_h(r)(\Gamma_h(r)-\tau_h)\Lambda(\xi)\Lambda(\eta)},$$

and therefore, using the definitions (29) and (30), we obtain the results in (31) and (32). □

**Proof of Corollary 6.** We find

$$\mathrm{E}[\exp(-\Lambda(\xi)\tau_h)] = \mathrm{E}\left[ \prod_{(c,\gamma)\in\Phi} \exp(-\gamma k(c, \xi)\tau_h) \right] = \exp(-a(\xi)),$$

where we at the last equality sign have used the form of the generating functional for the Poisson process $\Phi$, see Møller and Waagepetersen (2004, Proposition 3.3). The formula (34) is proved in a similar fashion. The result for NSPs is obtained by noting that a NSP is a special case of a SNCP with $k(c, \xi) = k(\xi - c)$ and intensity measure of cluster centres and cluster intensities $\mathrm{d}\zeta(c, \gamma) = \mathrm{d}c \, \mathrm{d}\delta_\mu(\gamma)$, where $\delta_\mu$ is a measure concentrated in the point $\mu$ with $\delta_\mu(\{\mu\}) = \kappa$. □

# References

Baddeley, A., Turner, R., 2005. Spatstat: An R package for analyzing spatial point patterns. J. Stat. Softw. 12 (6), 1–42.

Cox, D.R., 1955. Some statistical methods connected with series of events. J. R. Stat. Soc. Ser. B B (17), 129–164.

Diggle, P.J., 1983. Statistical Analysis of Spatial Point Patterns. Academic Press, London.

Fiksel, T., 1988. Edge-corrected density estimators for point processes. Statistics 19 (1), 67–75.

Hanisch, K.-H., 1982. On inversion formulae for $n$-fold Palm distributions of point processes in LCS-spaces. Math. Nachr. 106 (1), 171–179.

Illian, J., Penttinen, A., Stoyan, H., Stoyan, D., 2008. Statistical Analysis and Modelling of Spatial Point Patterns. In: Statistics in Practice, John Wiley & Sons Ltd., Chichester.

Kiderlen, M., Hörig, M., 2013. Matérn's hard core models of types I and II with arbitrary compact grains. CSGB Res. Rep. 2013-05, Aarhus University.

Lavancier, F., Møller, J., 2015. Modelling aggregation on the large scale and regularity on the small scale in spatial point pattern datasets. Scand. J. Stat. http://dx.doi.org/10.1111/sjos.12193. in press.

Madelung, A.B., Bondo, H., Stamp, I., Loevgreen, P., Nielsen, S.L., Falensteen, A., Knudsen, H., Ehinger, M., Dahl-Sørensen, R., Mortensen, N.B., et al., 2013. World Health Organization-defined classification of myeloproliferative neoplasms: Morphological reproducibility and clinical correlations —the Danish experience. Am. J. Hematol. 88 (12), 1012–1016.

Månsson, M., Rudemo, M., 2002. Random patterns of nonoverlapping convex grains. Adv. Appl. Probab. 34 (4), 718–738.

Matérn, B., 1960. Spatial variation. Stochastic models and their application to some problems in forest surveys and other sampling investigations. Medd. Statens Skogsforskningsinst 49 (5), 1–144.

Matérn, B., 1986. Spatial Variation. In: Lecture Notes in Statistics, vol. 36. Springer, Berlin.

Mattfeldt, T., Eckel, S., Fleischer, F., Schmidt, V., 2006. Statistical analysis of reduced pair correlation functions of capillaries in the prostate gland. J. Microsc. 223 (2), 107–119.

Mattfeldt, T., Eckel, S., Fleischer, F., Schmidt, V., 2007. Statistical modelling of the geometry of planar sections of prostatic capillaries on the basis of stationary Strauss hard-core processes. J. Microsc. 228 (3), 272–281.

Møller, J., 2003. Shot noise Cox processes. Adv. Appl. Probab. 35, 614–640.

Møller, J., Waagepetersen, R.P., 2004. Statistical Inference and Simulation for Spatial Point Processes. Chapman and Hall/CRC, Boca Raton, FL.

Nelder, J.A., Mead, R., 1965. A simplex method for function minimization. Comput. J. 7 (4), 308–313.

Neyman, J., Scott, E.L., 1958. Statistical approach to problems of cosmology. J. R. Stat. Soc. B 20, 1–43.

Stoyan, D., 1979. Interrupted point processes. Biom. J. 21 (7), 607–610.

Stoyan, D., Kendall, W.S., Mecke, J., 1995. Stochastic Geometry and Its Applications, second ed. Wiley, Chichester.

Stoyan, D., Stoyan, H., 1985. On one of Matérn's hard-core point process models. Math. Nachr. 122 (1), 205–214.

Teichmann, J., Ballani, F., van den Boogaart, K.G., 2013. Generalizations of Matérn's hard-core point processes. Spat. Stat. 3, 33–53.

Thomas, M., 1949. A generalization of Poisson's binomial limit for use in ecology. Biometrika 36, 18–25.

Tscheschel, A., 2005. Räumliche Statistik zur Charakterisierung gefüllter Elastomere (Ph.D. thesis), TU Bergakademie Freiberg.

Vytrva, N., Stacher, E., Regitnig, P., Zinke-Cerwenka, W., Hojas, S., Hubmann, E., Porwit, A., Bjorkholm, M., Hoefler, G., Beham-Schmid, C., 2014. Megakaryocytic morphology and clinical parameters in essential thrombocythemia, polycythemia vera, and primary myelofibrosis with and without JAK2 V617F. Arch. Pathol. Lab. Med. 138 (9), 1203–1209.

# Supplement D

This chapter elaborates on the results found in Paper D. We examine the quality of the proposed approximations of first- and second-order characteristics in limiting cases of a scaling parameter. Furthermore, we study parametric inference based on the minimum contrast method in combination with the approximated characteristics.

## DD.1   Limit results

Obviously, the retention probabilities in Paper D are close to one if the intensity of the un-thinned process is low, and close to zero, if the original process has high intensity. In Theorem DD.1 and Theorem DD.2 below, we describe the limit behaviour of the pair correlation function $g(r)$ and proposed approximation $g_a(r)$ of a Matérn thinned Cox process, when the intensity of the original process is multiplied by a constant $\beta$, which goes either to zero or to infinity. This is particularly interesting for NSPs, as it explains how $g(r)$ depends on the parameter $\mu$, the mean number of points in a cluster before thinning. The proofs of Theorem DD.1 and Theorem DD.2 are deferred to the Appendix at the end of the thesis.

**Theorem DD.1** (Limit behaviour of the pair correlation function under scaled driving measures). *Consider Matérn thinned Cox processes with driving field $\beta\Lambda$, where $\beta > 0$. Assume that $\Lambda$ has no singularities, in particular that $\Omega_\xi > 0$ a.s. when $\Lambda(\xi) > 0$, where $\Omega_\xi = \int_{b(\xi,h)} \Lambda(\eta)d\eta$.*
*Then, the pair correlation functions $g_\beta$ of the thinned processes satisfy*

$$\lim_{\beta\to 0} g_\beta(\xi,\eta) = \bar{g}(\xi,\eta), \qquad \|\xi - \eta\| > h, \tag{DD.1}$$

$$\lim_{\beta\to\infty} g_\beta(\xi,\eta) = \breve{g}(\xi,\eta), \qquad \|\xi - \eta\| > 2h, \tag{DD.2}$$

*where $\bar{g}$ is the pair correlation function of a Cox process with driving field $\Lambda$ and $\breve{g}$ is the pair correlation function of a Cox process with driving field given by*

$$\breve{\Lambda}(\xi) = \frac{\Lambda(\xi)}{\Omega_\xi}\mathbb{1}(\Lambda(\xi) > 0), \tag{DD.3}$$

*Example* 1.  Let $X$ be a MTMCP with cluster intensity $\kappa$, cluster radius $R$ and mean number of points per cluster before thinning $\mu$. If $\mu$ is large, we can use the approximation $\Lambda(\xi)/\Omega_\xi \approx 1/\tau_h$ (a constant) to see that the pair correlation function $g(r)$

of $X$ can for $r > 2h$ be approximated by the pair correlation function of a Boolean model with spherical grains of radius $R$ and germ intensity $\kappa$.

**Theorem DD.2** (Limit behaviour of the approximation pair correlation under scaled driving measures). *Consider Matérn thinned Cox processes with driving field $\beta\Lambda$, where $\beta > 0$ and $\Lambda$ fulfils the assumptions in Theorem DD.1.*

*Then, the approximated pair correlation functions $g_{a\beta}$ of the thinned processes satisfy*

$$\lim_{\beta \to 0} g_{a\beta}(\xi, \eta) = \bar{g}(\xi, \eta), \qquad\qquad \|\xi - \eta\| > 2h, \qquad \text{(DD.4)}$$

$$\lim_{\beta \to \infty} g_{a\beta}(\xi, \eta) = \tilde{g}(\xi, \eta)g_{\text{MatII}\infty}(\|\xi - \eta\|), \qquad \|\xi - \eta\| > h, \qquad \text{(DD.5)}$$

*where $\bar{g}$ is the pair correlation function of a Cox process with driving field $\Lambda$, $\tilde{g}$ is the pair correlation function of a Cox point process with driving field given by*

$$\tilde{\Lambda}(\xi) = \mathbb{1}(\Lambda(\xi) > 0) \qquad\qquad \text{(DD.6)}$$

*and $g_{\text{MatII}\infty}$ is the limit for $\beta \to \infty$ of the pair correlation function of a Matérn type II model with intensity $\beta$ before thinning. This pair correlation function takes the simple form*

$$g_{\text{MatII}\infty}(r) = 2\tau_h/\Gamma_h(r), \quad r > h. \qquad\qquad \text{(DD.7)}$$

*Example* 1 (continued). Let $X$ be the MTMCP described previously in this example. Then, using the approximation in Example 1, $\check{g}$ in Theorem DD.1 will be equal to $\tilde{g}$ in Theorem DD.2 for $r > 2h$, and as $g_{\text{MatII}\infty}(r) = 1$ for $r > 2h$, the limit of $g(r)$ and $g_a(r)$ for $\beta \to \infty$ will be identical.

Comparing Theorem DD.1 with Theorem DD.2, we observe that $g(r)$ and $g_a(r)$ show similar behaviour in the limits, in particular for $r > 2h$, whereas for $h < r < 2h$ it is more complicated to understand the relation. More precisely, for $\beta \to 0$ the limits are identical for $r > 2h$ and for $\beta \to \infty$ they show similar behaviour for MTMCPs, cf. Example 1.

## DD.2   Parameter estimation in Matérn thinned Matérn cluster processes

In this section, we present a thorough study of parameter estimation in Matérn thinned Matérn cluster processes (MTMCPs) that, due to space limitations, could not be presented in Paper D.

In principle, the inference is based on a standard estimation procedure, the so-called minimum contrast method (Diggle, 2003), where the difference between a non-parametric estimate of a second-order summary statistic (in this case the pair correlation function (pcf)) and the corresponding theoretical function is minimized. However, it is non-trivial to use the method, as estimation of the pcf is based on kernel smoothing, which does not capture the discontinuity of the theoretical function caused by the hard core distance in the MTMCPs, thus the estimate is

biased. Moreover, as discussed in Paper D, the theoretical summary statistics in the Matérn thinned Cox process models are hard to work with in practice, thus we need to substitute these with approximations. Hence, it is clearly crucial for the quality of the method that the bias introduced by kernel smoothing in the estimated pcf is controlled and that the quality of the approximations suggested in Paper D is sufficiently good. These questions are studied in this section. We start with a discussion of non-parametric estimation of the pcf.

### DD.2.1 Non-parametric estimation of the pair correlation function

We consider the estimation situation, where a single point pattern $X \cap W$ is observed in a bounded observation window $W$ with $|W| > 0$ in $\mathbb{R}^d$. There exist a number of estimators for the pcf $g$ and often these include an *edge-correction factor*, which is needed due to missing information for points in $W$ near the boundary, caused by the unobserved points outside $W$. A simpler alternative is to use *minus sampling*, where only the points in $W$ with a distance larger than $r$ to the boundary of $W$ are used. Using this method, not all information in $X \cap W$ is used in the estimate, but with sufficient data, this method may be preferred.

Common for all estimators of the pcf is the need for kernel smoothing. Therefore, let $k : \mathbb{R} \to \mathbb{R}^+$ be a kernel function, i.e. a probability density with compact support and let $k_\delta(\xi) \coloneqq k(\xi/\delta)/\delta$ be the corresponding kernel with bandwidth $\delta > 0$. Two of the edge-corrected kernel estimators for the pair correlation function in the stationary and isotropic case are given by

$$\hat{g}_i(r) = \frac{\hat{\rho}_i^{(2)}(r)}{\widehat{\rho^2}}, \qquad i = 1, 2, \tag{DD.8}$$

where

$$\hat{\rho}_1^{(2)}(r) = \frac{1}{2\pi r} \sum_{\xi, \eta \in X}^{\neq} \frac{\mathbb{1}(\xi, \eta \in W) k_\delta(r - \|\eta - \xi\|)}{|W \cap W_{\eta-\xi}|}, \tag{DD.9}$$

$$\hat{\rho}_2^{(2)}(r) = \frac{1}{2\pi} \sum_{\xi, \eta \in X}^{\neq} \frac{\mathbb{1}(\xi, \eta \in W) k_\delta(r - \|\eta - \xi\|)}{|W \cap W_{\eta-\xi}| \|\eta - \xi\|^{d-1}} \tag{DD.10}$$

and

$$\widehat{\rho^2} = n(X \cap W)(n(X \cap W) - 1)/|W|. \tag{DD.11}$$

Here $n(X \cap W)$ denotes the number of points in $X \cap W$. The estimators (DD.9) and (DD.10) (*translate*-edge-corrected estimators) were suggested in Fiksel (1988) with inspiration from edge-corrections for the $K$-function, discussed in Ohser and Stoyan (1981). The estimator (DD.10) is a modification of (DD.9), where the contribution from the interpoint distances is used instead of the distance $r$. This may reduce bias for small values of $r$. Discussions regarding estimates of $\rho^2$ may be found in Møller

and Waagepetersen (2004, p. 39). If $|W \cap W_\xi| > 0$, for all $\xi \in B(0, r)$, the estimators fulfil

$$\mathbb{E}\hat{\rho}_1^{(2)}(r) = \int_0^\infty k_\delta(r-s)\rho^{(2)}(s)(s/r)\,\mathrm{d}s, \qquad (DD.12)$$

$$\mathbb{E}\hat{\rho}_2^{(2)}(r) = \int_0^\infty k_\delta(r-s)\rho^{(2)}(s)\,\mathrm{d}s. \qquad (DD.13)$$

Estimation procedures based on kernel smoothing are sensitive towards the choice of bandwidth of the kernel. Large values of the bandwidth yield smooth estimates, but important details may be lost, whereas small values of the bandwidth yield noisy estimates. The optimal procedure for estimation of pcfs is therefore frequently discussed in the literature, and several simulation studies have been carried out, see e.g. discussions in Stoyan and Stoyan (1994, Section 15.4.2), Møller and Waagepetersen (2004, Section 4.3.5) and Illian et al. (2008, Section 4.3.3).

It is well-known, that a hard core property of point processes such as Matérn thinned Cox processes ($\rho^{(2)}(r) = 0$, $r < h$, where $h > 0$ is the hard core distance) is not utilized in a kernel based estimate $\hat{g}$ of the pcf. Thus, for comparison between the theoretical function $g$, which contains a discontinuity at $r = h$, and the smooth function $\hat{g}$, $g$ should be smoothed with the same kernel. From (DD.12) and (DD.13) it follows that $\hat{g}_1$ and $\hat{g}_2$ in (DD.8) should be compared to

$$\tilde{g}_1(r) = \int_0^\infty k_\delta(r-s)g(s)(s/r)\,\mathrm{d}s \quad \text{and} \quad \tilde{g}_2(r) = \int_0^\infty k_\delta(r-s)g(s)\,\mathrm{d}s, \qquad (DD.14)$$

respectively. Alternatively, if it is known prior to estimation that $g(r) = 0$, for $r < h$, $h > 0$, the bias due to the kernel estimation can be reduced by the so-called *reflection method* (Illian et al., 2008), where the following estimator is considered

$$\hat{g}_{\mathrm{ref}}(r) = \begin{cases} \hat{g}(r) + \hat{g}(h-r) & \text{for } r \geq h \\ 0 & \text{otherwise.} \end{cases} \qquad (DD.15)$$

## DD.2.2 Parametric inference for Matérn thinned Cox processes

For point processes with theoretical second-order characteristics in a reasonably tractable form, second-order moment estimation methods have become of great interest. These simulation-free methods have been developed as alternatives to computationally demanding Markov chain Monte Carlo simulation methods like maximum likelihood estimation. In a study of moment methods for stationary Cox processes by Dvořák and Prokešová (2012), it was concluded that for the majority of the considered Cox process cases, the minimum contrast method, in combination with the pcf, provides the best estimates. However, when the pcf does not have a closed form, the method becomes numerically more demanding.

In the minimum contrast method, a vector of parameters $\theta$ is estimated by minimizing the discrepancy measure

$$D(\theta) = \int_{r_1}^{r_2} (\hat{T}(s)^q - T(s;\theta)^q)^2\,\mathrm{d}s \qquad (DD.16)$$

between the theoretical summary function $T(\cdot;\theta)$ and its empirical non-parametric counterpart $\hat{T}$. The method performs best both in terms of speed and accuracy, if the theoretical function is available in a reasonable form and is sensitive to the variation of $\theta$.

For many Cox processes, the minimum contrast method can be applied both using the $K$-function and the pcf $g$, when these have closed form theoretical expressions (Møller and Waagepetersen, 2004, p. 182). The $K$-function has traditionally been used most frequently. However, recent studies report, that the estimation procedure using the pcf is more stable, see e.g. Dvořák and Prokešová (2012). In Diggle (2003, Section 6.1.1) it is recommended to use the tuning parameter $r_2 \leq 0.25$ for fitting point patterns in the unit square window using the $K$-function, and $q = 1/4$ for aggregated point patterns and $q = 1/2$ for regular point patterns. The parameter $r_1$ can be set to 0 or a small positive values, e.g. depending on the minimum inter-point distance between the observed points.

For the classical MCPs and TPs, the parameter estimates can be found using the default settings in `matclust.estpcf` and `thomas.estpcf` functions in the `spatstat` library (Baddeley et al., 2015) of R. Note that for these processes, $g$ does not depend on $\mu$, thus $\theta = (\kappa, R)$ or $\theta = (\kappa, \sigma)$, respectively. Subsequently, $\mu$ is estimated using $\hat{\mu} = \hat{\rho}/\hat{\kappa}$.

In order to estimate the parameters in the Matérn thinned Cox process models, we propose to use the minimum contrast method in combination with the approximated theoretical pair correlation function $g_a$ and an estimator $\hat{g}$, using either

$$\int_h^{r_2} (\hat{g}_{\text{ref}}(s)^q - g_a(s;\theta)^q)^2 \, ds \quad \text{or} \quad \int_0^{r_2} (\hat{g}(s)^q - \tilde{g}_a(s;\theta)^q)^2 \, ds, \tag{DD.17}$$

where $\hat{g}_{\text{ref}}$ and $\tilde{g}_a(s;\cdot)$ are the reflected or smooth versions of the functions as defined in Section DD.2.1. Alternatively, to compare $\hat{g}(\cdot)$ with $g_a(\cdot;\theta)$ directly, a large value of $r_1$ must be chosen to avoid the kernel-effect, e.g. $r_1 = h + \delta$ (the hard-core distance plus the kernel bandwidth). No standard recommendations are available for the method in combination with the pcf, but following the simulation study of Cox processes by Dvořák and Prokešová (2012), we shall use $r_2 = 0.25$ and $q = 1/4$.

For Matérn thinned Neyman-Scott processes, the simple expressions for $\rho_a$ and $g_a$ in (2.17) and (2.18), allow for a simplification of $D(\theta)$ as the dimension of $\theta$ can be reduced. We have

$$g_a(r) = \frac{2\Gamma_h(r)\rho_a - 2(1 - (1 - \rho_a \tau_h) \exp(\tilde{b}(r)/\tilde{a}))}{\rho_a^2 \Gamma_h(r)(\Gamma_h(r) - \tau_h)} \tag{DD.18}$$

and

$$\kappa = -\log(1 - \rho_a \tau_h)/\tilde{a} \tag{DD.19}$$

where $\tilde{a} = a/\kappa$ and $\tilde{b}(r) = b(r)/\kappa$ do not depend on $\kappa$. Hence, replacing $\rho_a$ with an estimate of the intensity $\hat{\rho}$, allows $\kappa$ to be estimated separately. For MTMCPs and MTTPs, assuming $h$ is known or estimated be the minimum inter-point distance, we can therefore minimize $D(\theta)$, where $\theta = (\mu, R)$ and $\theta = (\mu, \sigma)$, respectively, and subsequently estimate $\kappa$ using (DD.19).

### DD.2.3 Simulation study

Simulations and estimation of parameters were performed using the `spatstat` library of R, where standard functions were used for MCPs and TPs, and similar adjusted functions were used for MTMCPs. For each point process model and each combination of parameters, we generated 500 independent realizations in the unit square window $W = [0,1]^2$ and re-estimated the parameters using the minimum contrast method. Simulation of the processes are straightforward for MTMCPs and MCPs, using larger windows, dependent on $R$ and $h$. For the Thomas process, see default settings in `spatstat`. Estimation of parameters is explained thoroughly in the following sections. To assess the performance of parameter estimation in MTMCPs on middle-sized to large point patterns, exhibiting different degree of clustering and regularity, we chose sixteen combinations of parameter values and compared the results to similar MCPs and TPs. The parameter $h$ was assumed to be known. Alternatively, this parameter could have been estimated prior to the other parameters using the minimum inter-point distance. For each of eight combinations of the intensity $\rho$, cluster intensity $\kappa$ and cluster radius $R$, four models were compared;

- **Highly regular MTMCP**, with parameters $(\kappa, R, \mu, h)$, where $h = R/3$ and $\mu$ were chosen according to formula (13) and (15) in Paper D, to obtain the desired intensity $\rho$.

- **Slightly regular MTMCP**, with parameters $(\kappa, R, \mu, h)$, where $h = R/5$ and $\mu$ were chosen according to formula (13) and (15) in Paper D, to obtain the desired intensity $\rho$.

- **MCP**, with parameters $(\kappa, R, \mu)$, where $\mu = \kappa/\rho$.

- **TP**, with parameters $(\kappa, \sigma, \mu)$, where $\mu = \kappa/\rho$ and $\sigma$ where chosen in agreement with the parameter $R$, such that

$$\int_{\mathbb{R}^2} k_\sigma(\|\xi\|)\mathbb{1}(\|\xi\| \leq R)\,\mathrm{d}\xi \approx 95\%,$$

where $k_\sigma(\cdot)$ denotes the kernel of the TP.

In Table DD.1 an overview of the parameter values for the MTMCPs, MCPs and TPs is given. Sample realizations are shown in Figure DD.1 for model I-II and in Figure DD.5 in the Appendix for all the models. Approximated pcfs for MTMCPs (red stippled lines) obtained by (2.18) and theoretical pcfs for MCPs and TPs (blue full drawn and stippled lines, respectively) are shown in Figure DD.2 for model I-II and in Figure DD.6-DD.8 in the Appendix for all the models. The grey areas represent the effect from the parameter $\mu$ on the pcfs of the MTMCPs. The lower bound corresponds to the pcf of a Boolean model ($\mu \to \infty$) and the upper bound corresponds to the pcf of the MCP model ($\mu \to 0$), see Section DD.1.
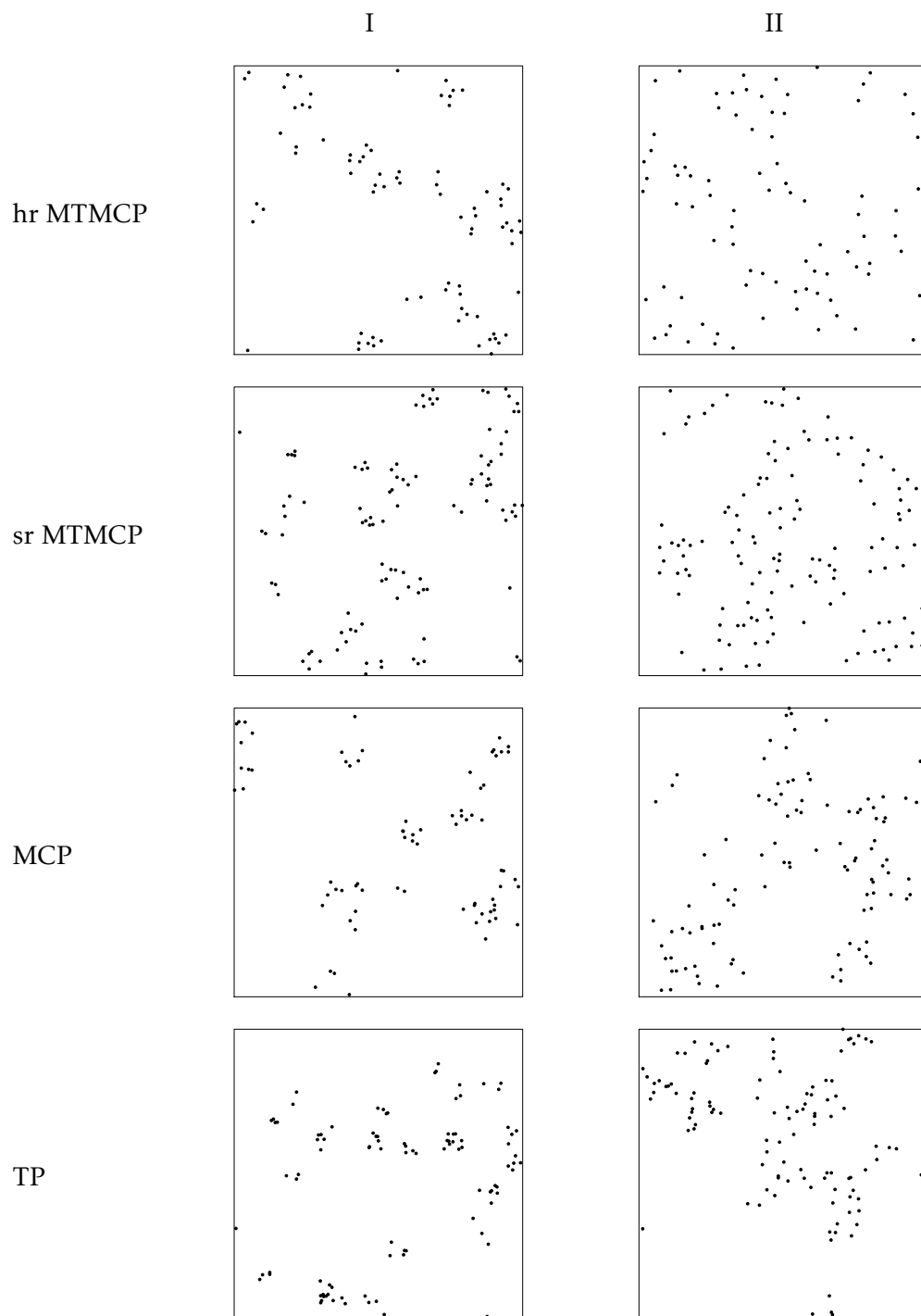
**Figure DD.1:** Sample realizations of model I-II for four types of processes; highly regular MTMCPs, slightly regular MTMCPs, MCPs and TPs.

|      |    | $\kappa$ | $\rho$ | $R$ | $\sigma$ | $h = R/\cdot$ | $\mu$ (MTMCP) | $\mu$ (MCP, TP) | $p_{ret}(\%)$ |
|------|----|----------|--------|-----|----------|---------------|---------------|-----------------|---------------|
| I    | hr | 25       | 100    | .05 | .02      | 3             | 5.4           | 4               | 75            |
|      | sr |          |        |     |          | 5             | 4.4           |                 | 91            |
| II   | hr |          |        | .10 | .04      | 3             | 6.7           |                 | 60            |
|      | sr |          |        |     |          | 5             | 4.7           |                 | 85            |
| III  | hr |          | 150    | .05 | .02      | 3             | 10.0          | 6               | 60            |
|      | sr |          |        |     |          | 5             | 7.0           |                 | 86            |
| IV   | hr |          |        | .10 | .04      | 3             | 17.5          |                 | 34            |
|      | sr |          |        |     |          | 5             | 7.7           |                 | 78            |
| V    | hr | 50       | 200    | .05 | .02      | 3             | 5.7           | 4               | 70            |
|      | sr |          |        |     |          | 5             | 4.5           |                 | 89            |
| VI   | hr |          |        | .10 | .04      | 3             | 11.4          |                 | 35            |
|      | sr |          |        |     |          | 5             | 5.1           |                 | 78            |
| VII  | hr |          | 300    | .05 | .02      | 3             | 11.4          | 6               | 53            |
|      | sr |          |        |     |          | 5             | 7.2           |                 | 83            |
| VIII | hr |          | (∗)    | .10 | .04      | 3             | 22.8          |                 | 20            |
|      | sr |          |        |     |          | 5             | 8.9           |                 | 67            |

**Table DD.1:** Parameters of the MTMCPs and similar MCPs and TPs in all parametric models considered in this study. The index I–VIII corresponds to eight parametric MCP and TP models, whereas the abbreviations *hr* and *sr* is used for *highly regular* MTMCPs and *slightly regular* MTMCPs, respectively, in short for the choice of *h*. For all the MTMCP models, $p_{ret}$ denotes the retention probability. In VIII hr (∗) marks that the desired intensity could not be obtained as $\rho \rightarrow 239$ (approximately) for $\mu \rightarrow \infty$. Instead, $\mu = 22.8$ (two times the value for case VI hr) yields $\rho \approx 228$.

### DD.2.4 Preliminary studies of Matérn thinned cluster processes

To find an (optimal) strategy for parameter estimation in MTMCPs models, we considered several pcf estimates and three methods for the discrepancy measure: large lower $r$ ($r_1 = h + \delta$) (M1), the reflection method (M2) and the convolution method (M3), as described in Section DD.2.2.

The modified pcf estimator $\hat{g}_2$, see (DD.8) and (DD.10), clearly improved the bias of the estimator near zero, which was important in M1 and M2. However, no differences were found for M3. The choice of smoothing bandwidth was important both in terms of estimating parameters in M1 and for some cases of M2, as well as the behaviour of $\hat{g}$ in M2, as e.g. artificial jumps appeared using default values. The methods M2 and M3 were almost equally good, when using $\hat{g}_2$ as the pcf estimator. To summarize, preliminary simulations indicate that a reasonable strategy for parameter estimation was obtained by the *pcf.ppp* function using: default values of $r$ (equidistant values between 0 and 0.25), the Epanechnikov kernel, fixed kernel bandwidth $\delta = h$ (i.e. smooth Epanechnikov bandwidth $bw = h/\sqrt{5}$), *translate* for the choice of edge correction, *divisor = d* (i.e. using $\hat{g}_2$), $q = 1/4$, combined with either M2 to obtain $\hat{g}$ similar to $g$ for visualization or M3 to estimate the parameters in the MTMCP model. In the following section, this set-up is used.
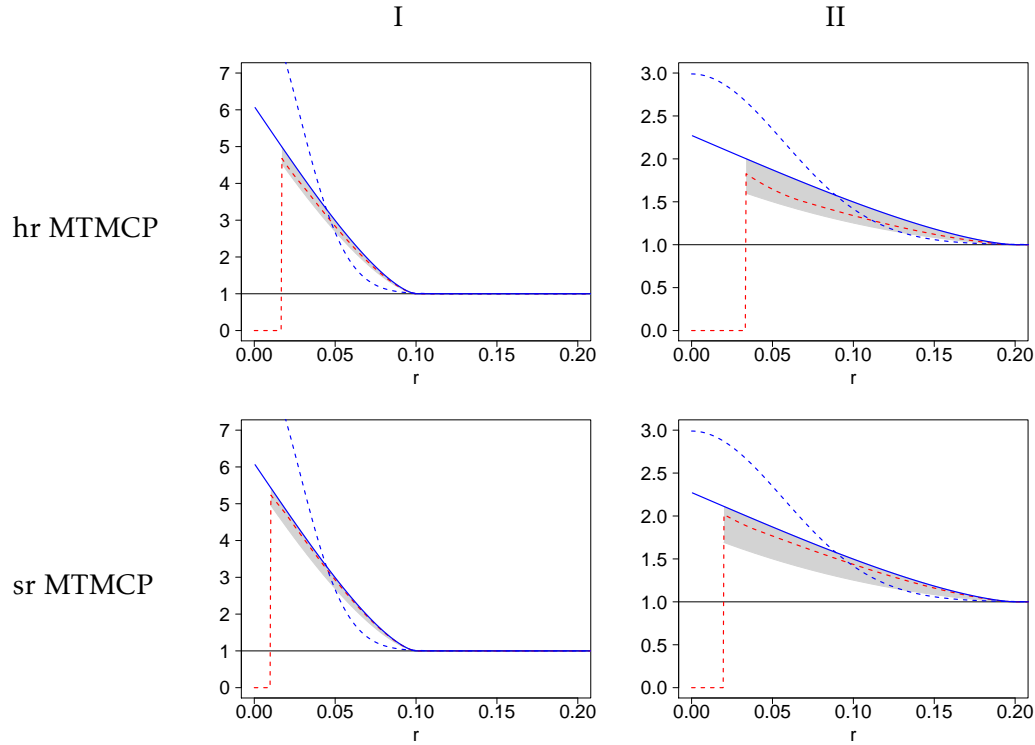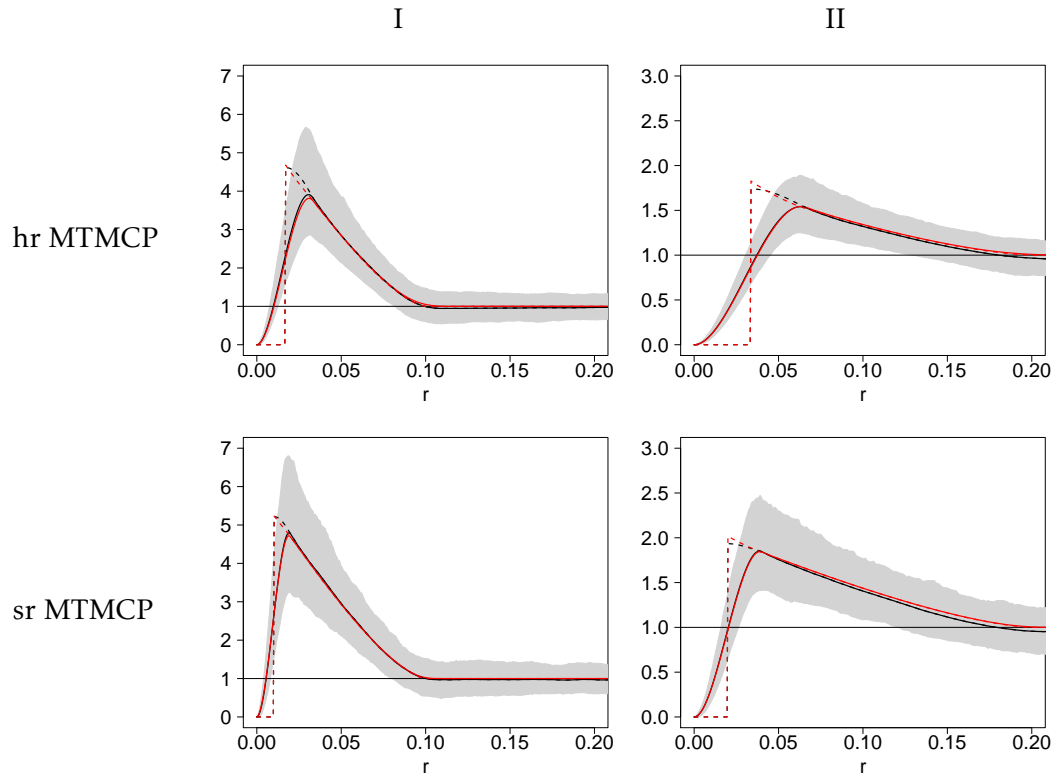
**Figure DD.2:** The approximated pcfs $g_a$ (red stippled lines) of MTMCPs are compared to the theoretical pcfs for the similar MCPs and TPs (blue full drawn and stippled lines, respectively) in model I-II. The grey areas represent the effect from the parameter $\mu$ on the pcfs of the MTMCPs. The lower bound corresponds to the pcf of a Boolean model ($\mu \to 0$) and the upper bound corresponds to the pcf of the MCP model ($\mu \to \infty$).

### DD.2.5 Comparison between the Matérn thinned Matérn cluster processes, Matérn cluster processes and Thomas processes

In this section, we describe the performance of parameter estimation of the MTMCPs compared to the performance for MCPs and TPs.

The obtained estimates of the pcfs in the MTMCPs are shown together with the approximated pcfs $g_a$ (red stippled lines) in Figure DD.3 for model I-II and in Figure DD.6-DD.8 for the remaining models. The estimates are represented by the grey areas, which mark point-wise central 95%-regions, based on 500 simulations. The average pcfs, based on these simulations, are also shown (black full drawn lines), the reflected average pcfs (black stippled lines) as well as the smooth versions of $g_a$ (red full drawn lines). Thus the stippled lines illustrate the functions expected to be closest to the true theoretical pcf $g$ and verify, that the proposed approximation $g_a$ captures most of the true behaviour of $g$. On the other hand, the full drawn lines illustrates the average functions compared in the estimation procedure.

Main results are illustrated using the relative deviations of the parameters $\kappa$, $\mu$ and $R$ in Figure DD.4 for model I-II and Figure DD.9-DD.10 in the Appendix for all models. Furthermore, the results are summarized in Table DD.2 and DD.3

**Figure DD.3:** Estimates of the pcfs in the highly regular MTMCPs and slightly regular MTMCPs in model I-II are shown together with the approximated pcfs $g_a$ (red stippled lines). The estimates are represented by the grey areas, which mark point-wise central 95%-regions, based on 500 simulations. The average pcfs, based on these simulations, are also shown (black full drawn lines), the reflected average pcfs (black stippled lines) as well as the smooth versions of $g_a$ (red full drawn lines).

in the Appendix. The tables show relative mean biases $1/N \sum_i^N (\hat{\theta}_i - \theta)/\theta$ (RMBs) of the estimators and relative mean squared errors $1/N \sum_i^N (\hat{\theta}_i - \theta)^2/\theta^2$ (RMSEs). All the statistics were obtained from the middle 95% of the estimates from the 500 replications. Neglecting the 5% of the most extreme estimates was inspired by the procedure in a simulation study in Dvořák and Prokešová (2012). The motivation is, that in practical applications, if such estimates were encountered, they would be identified easily as unrealistic and practitioners would probably alter the optimization methods e.g. by introducing limits of the parameters.

The figures and tables show that for the majority of cases considered in this study, parameter estimation for MTMCPs is or almost is as efficient as parameter estimation for the simpler standard MTPs and TPs.

Common for all four processes (highly regular MTMCPs, slightly regular MTM-CPs, MCPs and TPs) is that estimation of $\kappa$, $\mu$ and $R$ or $\sigma$ depends much on the level of clustering ($R$ or $\sigma$). The highly clustered processes (model I, III, V and VII) result in much more accurate estimates than the slightly clustered processes (model II, IV, VI and VIII). For each of the eight models (I-VIII), all four processes the parameter $\kappa$
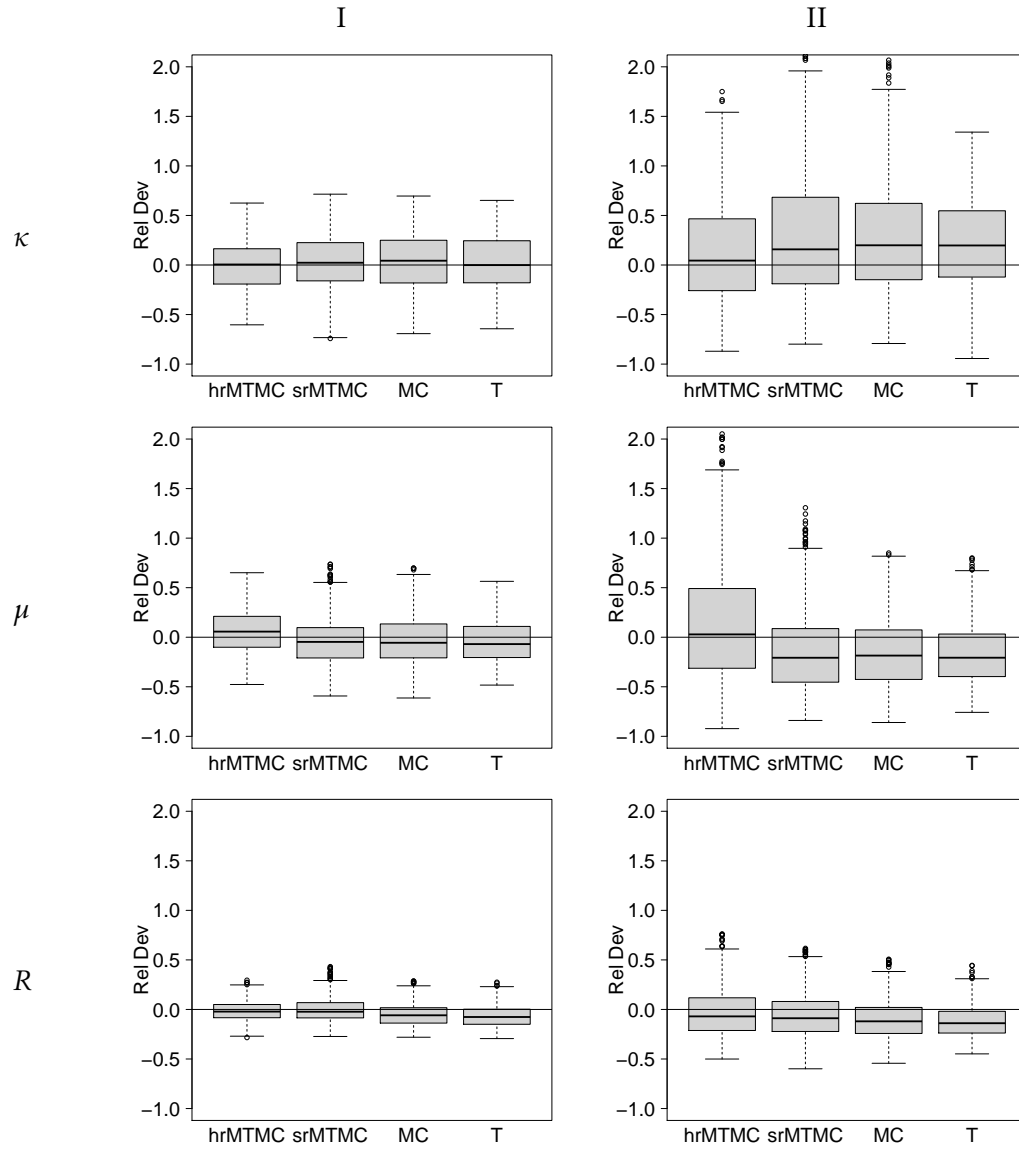
**Figure DD.4:** The relative deviation of the estimates from the true parameters $\kappa$, $\mu$ and $R$ for highly regular MTMCPs, slightly regular MTMCPs, MCPs and TPs in model I and II.

is almost equally well estimated. For the highly clustered processes (model I, III, V and VII), $\mu$ and $R$ are estimated equally well for all four processes, with estimation for highly regular MTMCPs being slightly less accurate. For the remaining four cases, apart from model II for $\mu$ and model II and VI for $R$, estimation of $\mu$ and $R$ for slightly regular MTMCPs is somewhat less accurate, whereas for highly regular MTMCPs, estimation of $\mu$ and $R$ is much less accurate.

The overall conclusion is that parameter estimation of slightly regular MTMCPs works well, relative to the estimation of MCPs and TPs. For hightly regular MTMCPs it seems to be more difficult to estimate some of the parameters, in particular if there is a low level of clustering.

### DD.2.6  Discussion

In this simulation study, we have found that parameter estimation in the MTMCP model is almost as simple and efficient, as standard estimation procedures for MCPs and TPs, using the minimum contrast method. Thus, the models introduced in Paper D can be used in practical applications, and furthermore, the proposed approximations seem to represent the true theoretical pcfs sufficiently accurately to be used for parameter estimation. Additionally, we have found specific recommendations for the use of the minimum contrast method to fit a MTMCP model to observed point patterns using parameters and discrepancy measure as recommended in Section DD.2.4. In particular, using a convolution of the theoretical (approximated) pcf identical to the one used in the estimated pcf is recommended for a robust parameter estimation procedure.

However, for some cases of highly regular MTMCPs (model IV hr, VI hr and VIII hr), further investigation is needed for proper estimation of the parameters in the model, in particular for the parameter $\mu$. One might be able to overcome the problems for the highly regular MTMCPs, by using prior knowledge of the data, requiring certain bounds on the parameters or by choosing other tuning parameters in the minimum contrast method. On the other hand, the minimum contrast method performs best, if the function under consideration is sensitive to the variation of the parameters. For the complicated highly regular and slightly clustered cases (model IV hr, VI hr and VIII hr), it can be seen from Figure-DD.6-DD.8 that this is not true for high values of the parameter $\mu$, as the function almost hit the limit function for $\mu \rightarrow \infty$. The problematic cases have high values of $\mu$ and low values of $p_{ret}$ - see Table DD.1. One may rise the question how important correct estimation of $\mu$ is, if the behaviour of the thinned process does not depend much on $\mu$, from a certain level of $\mu$. The sample realizations shown in Figure DD.5 illustrate how difficult it is to distinguish the clusters in the slightly clustered cases. Thus, it is not surprising that parameter estimation is more difficult in these cases.

In general, when fitting a Cox process to data, one often has replicated realizations of the model. This might improve the parameter estimation, as it will prevent extreme values of the parameter estimates.

The MTMCP is a more sophisticated model than the MP and the TP, as it

contains both clustering and regularity properties, however it should be noted that the complexity of the process is reduced by assuming that the hard core distance $h$ is known and in this case MTMCP contains the same number of unknown parameters as the MCP and the TP. Alternatively in applications, $h$ could be estimated as the minimum inter-point distance. This, of course, introduces additional error.

Above, we have compared the MTMCPs to cluster processes with the same number of clusters $\kappa$ and intensity $\rho$, to study the applicability of the new models. Based on the results in this study, it could be relevant to construct and investigate a set-up, where prior retention probabilities are chosen, and the other parameters are chosen accordingly.

## DD.3   Appendix

*Proof of Theorem DD.1.* Let $\rho_\beta$ and $\rho_\beta^{(2)}$ be the intensity and product density of the Matérn thinned Cox process with driving field $\beta\Lambda$. To proof (DD.1), we use that the retention probabilities $p_{\text{ret}|\beta\Lambda}(\xi)$ and $p_{\text{ret}|\beta\Lambda}^{(2)}(\xi,\eta)$ in (2.14)-(2.15) in Section 2.3.1 converge to one for $\beta \to 0$, if $\|\xi - \eta\| > h$, which is a consequence of l'Hôpitals rule. More explicitly,

$$\lim_{\beta\to 0} p_{\text{ret}|\beta\Lambda}(\xi) = \lim_{\beta\to 0} \frac{1 - \exp(-\beta\Omega_\xi)}{\beta\Omega_\xi} = \lim_{\beta\to 0} \frac{\Omega_\xi \exp(-\beta\Omega_\xi)}{\Omega_\xi} = 1 \qquad (\text{DD.20})$$

and

$$\lim_{\beta\to 0} p_{\text{ret}|\beta\Lambda}^{(2)}(\xi,\eta) = 2\lim_{\beta\to 0} \left( \frac{1 - \exp(-\beta\Omega_\xi)}{\beta^2\Omega_\xi\Omega_{\eta\setminus\xi}} - \frac{1 - \exp(-\beta\Omega_{\xi\cup\eta})}{\beta^2\Omega_{\xi\cup\eta}\Omega_{\eta\setminus\xi}} \right) = 1 \qquad (\text{DD.21})$$

By dominated convergence, we therefore have

$$\lim_{\beta\to 0} \rho_\beta(\xi)/\beta = \lim_{\beta\to 0} \mathbb{E}\left( p_{\text{ret}|\beta\Lambda}(\xi)\Lambda(\xi) \right) = \mathbb{E}(\Lambda(\xi)) \qquad (\text{DD.22})$$

and

$$\lim_{\beta\to 0} \rho_\beta^{(2)}(\xi,\eta)/\beta^2 = \lim_{\beta\to 0} \mathbb{E}(p_{\text{ret}|\beta\Lambda}^{(2)}(\xi,\eta)\Lambda(\xi)\Lambda(\eta)) = \mathbb{E}(\Lambda(\xi)\Lambda(\eta)), \qquad (\text{DD.23})$$

thus

$$\lim_{\beta\to 0} g_\beta(\xi,\eta) = \lim_{\beta\to 0} \frac{\rho_\beta^{(2)}(\xi,\eta)}{\rho_\beta(\xi)\rho_\beta(\eta)} = \frac{\mathbb{E}(\Lambda(\xi)\Lambda(\eta))}{\mathbb{E}\Lambda(\xi)\mathbb{E}\Lambda(\eta)} = \bar{g}(\xi,\eta). \qquad (\text{DD.24})$$

For (DD.2), note that $p_{\text{ret}|\beta\Lambda}^{(2)}(\xi,\eta) = p_{\text{ret}|\beta\Lambda}(\xi)p_{\text{ret}|\beta\Lambda}(\eta)$ if $\|\xi - \eta\| > 2h$. By dominated convergence, we have

$$\lim_{\beta\to\infty} \rho_\beta(\xi) = \lim_{\beta\to\infty} \mathbb{E}\left( \frac{1 - \exp(-\beta\Omega_\xi)}{\beta\Omega_\xi} \beta\Lambda(\xi)\mathbb{1}(\Lambda(\xi) > 0) \right)$$
$$= \mathbb{E}\left( \Lambda(\xi)/\Omega_\xi \mathbb{1}(\Lambda(\xi) > 0) \right) = \mathbb{E}\check{\Lambda}(\xi), \qquad (\text{DD.25})$$

and

$$\lim_{\beta \to \infty} \rho_\beta^{(2)}(\xi, \eta) = \lim_{\beta \to \infty} \mathbb{E}\left(\tfrac{(1-\exp(-\beta\Omega_\xi))(1-\exp(-\beta\Omega_\eta))}{\beta\Omega_\xi\beta\Omega_\eta}\beta^2\Lambda(\xi)\Lambda(\eta)\mathbb{1}(\Lambda(\xi),\Lambda(\eta) > 0)\right)$$

$$= \mathbb{E}\left(\Lambda(\xi)/\Omega_\xi\Lambda(\eta)/\Omega_\eta\mathbb{1}(\Lambda(\xi) > 0, \Lambda(\eta) > 0)\right) = \mathbb{E}(\check{\Lambda}(\xi)\check{\Lambda}(\eta)),$$
$$(DD.26)$$

thus

$$\lim_{\beta \to \infty} g_\beta(\xi, \eta) = \lim_{\beta \to \infty} \frac{\rho_\beta^{(2)}(\xi, \eta)}{\rho_\beta(\xi)\rho_\beta(\eta)} = \frac{\mathbb{E}(\check{\Lambda}(\xi)\check{\Lambda}(\eta))}{\mathbb{E}\check{\Lambda}(\xi)\mathbb{E}\check{\Lambda}(\eta)} = \check{g}(\xi, \eta). \qquad (DD.27)$$

The last equalities in (DD.24) and (DD.27) use the fact that the product density of an inhomogeneous Poisson point process factorizes. □

*Proof of Theorem DD.2.* Let $\rho_{a\beta}$ and $\rho_{a\beta}^{(2)}$ be the approximated intensity and product density of the Matérn thinned Cox process with driving field $\beta\Lambda$. The proof of (DD.4) is similar to the proof in Theorem DD.1, where we use that the approximative retention probabilities $p_{a|\beta\Lambda}(\xi)$ and $p_{a|\beta\Lambda}^{(2)}(\xi, \eta)$ converge to one for $\beta \to 0$, if $\|\xi - \eta\| > 2h$. For (DD.5), we have by dominated convergence

$$\lim_{\beta \to \infty} \rho_{a\beta}(\xi) = \lim_{\beta \to \infty} \mathbb{E}\left(\frac{1-\exp(-\beta\Lambda(\xi)\tau_h)}{\beta\Lambda(\xi)\tau_h}\beta\Lambda(\xi)\mathbb{1}(\Lambda(\xi) > 0)\right)$$

$$= \mathbb{E}(1/\tau_h\mathbb{1}(\Lambda(\xi) > 0)) = \mathbb{E}(\check{\Lambda}(\xi))/\tau_h, \qquad (DD.28)$$

and with

$$p_{a|\beta\Lambda}^{(2)}(\xi, \eta) = \tfrac{\Gamma_h(r)(2-\exp(-\beta\Lambda(\xi)\tau_h)-\exp(-\beta\Lambda(\eta)\tau_h))-2\tau_h(1-\exp(-\beta(\Lambda(\xi)+\Lambda(\eta))\Gamma_h(r)/2))}{\beta^2\tau_h\Gamma_h(r)(\Gamma_h(r)-\tau_h)\Lambda(\xi)\Lambda(\eta)},$$
$$(DD.29)$$

we obtain

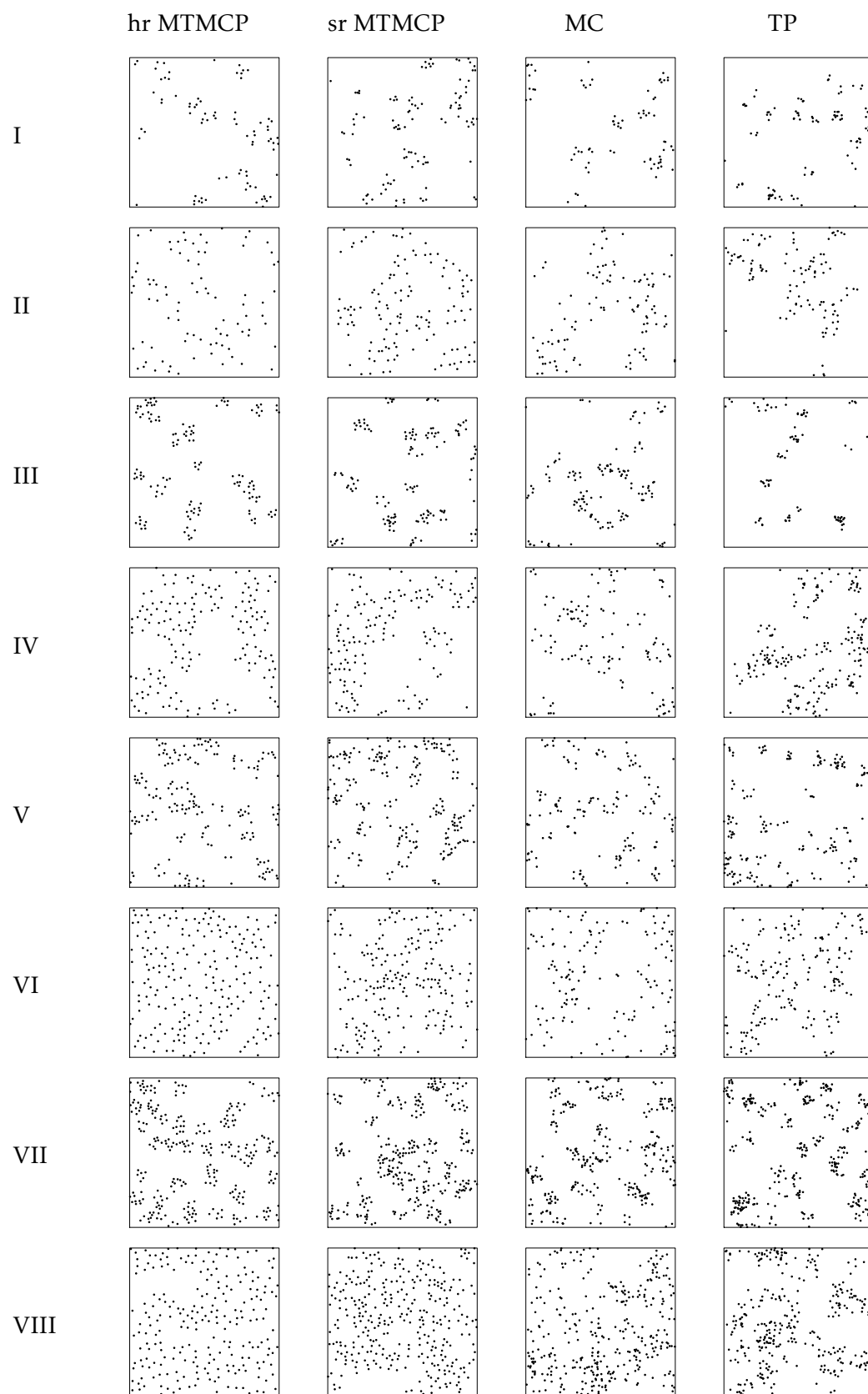$$\lim_{\beta \to \infty} \rho_{a\beta}^{(2)}(\xi, \eta) = \lim_{\beta \to \infty} \mathbb{E}\left(p_{a|\beta\Lambda}^{(2)}(\xi, \eta)\beta^2\Lambda(\xi)\Lambda(\eta)\mathbb{1}(\Lambda(\xi),\Lambda(\eta) > 0)\right)$$

$$= \mathbb{E}(2/(\tau_h\Gamma_h(r))\mathbb{1}(\Lambda(\xi),\Lambda(\eta) > 0)) = \mathbb{E}(\check{\Lambda}(\xi)\check{\Lambda}(\eta))2/(\tau_h\Gamma_h(r)), \quad (DD.30)$$

where $r = \|\xi - \eta\|$, thus
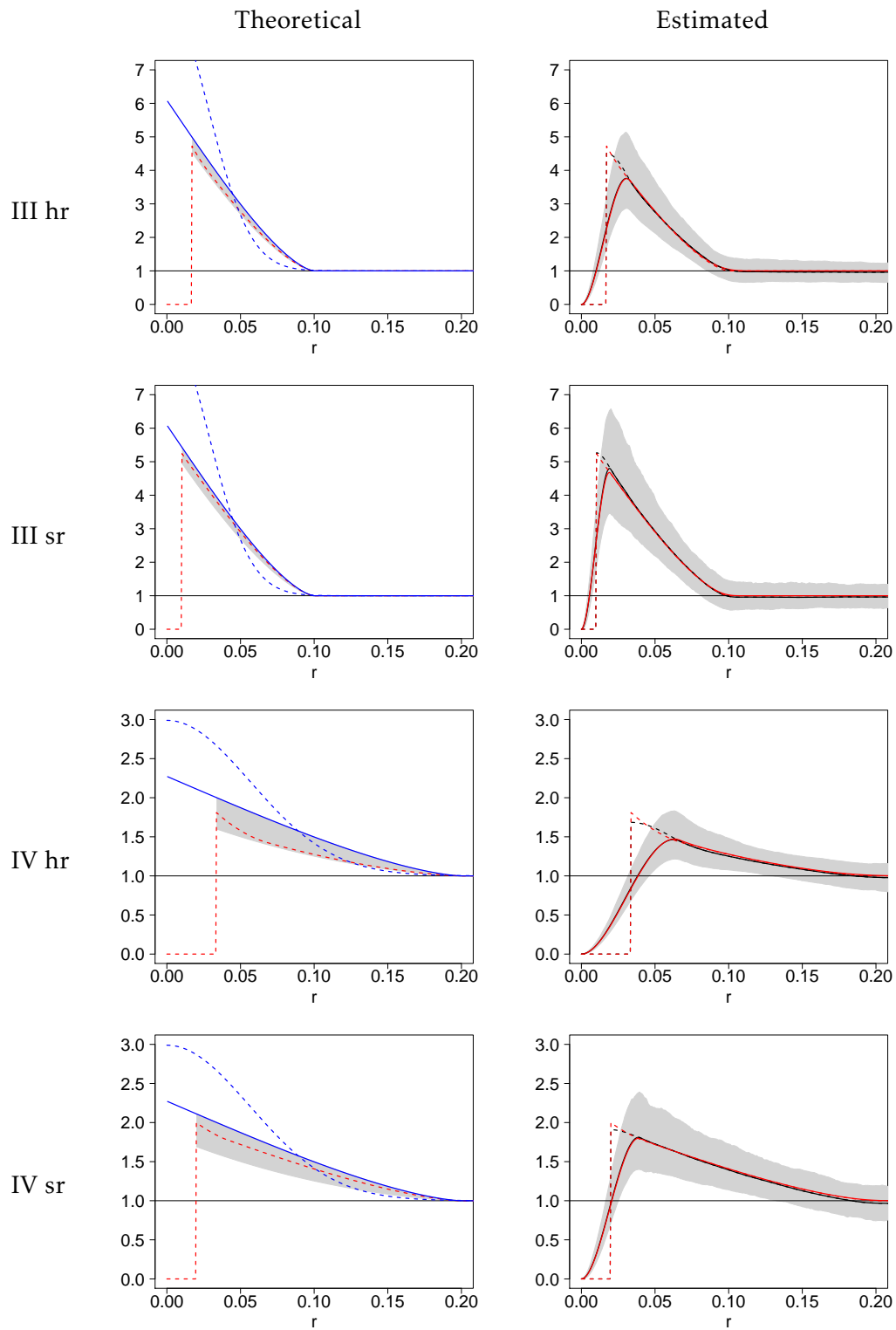
$$\lim_{\beta \to \infty} g_{a\beta}(\xi, \eta) = \lim_{\beta \to \infty} \frac{\rho_{a\beta}^{(2)}(\xi, \eta)}{\rho_{a\beta}(\xi)\rho_{a\beta}(\eta)} = \frac{\mathbb{E}(\check{\Lambda}(\xi)\check{\Lambda}(\eta))}{\mathbb{E}\check{\Lambda}(\xi)\mathbb{E}\check{\Lambda}(\eta)}\frac{2\tau_h}{\Gamma_h(r)}$$

$$= \check{g}(\xi, \eta)g_{\text{MatII}\infty}(\|\xi - \eta\|). \qquad (DD.31)$$

□

**Figure DD.5:** Sample realizations of all models I-VIII in Table DD.1 for highly regular MTMCPs, slightly regular MTMCPs, MCPs, and TPs.

Theoretical                    Estimated



**Figure DD.6:** Theoretical and estimated pcfs for model III-IV. See Figure DD.2 and DD.3 for a detailed description.
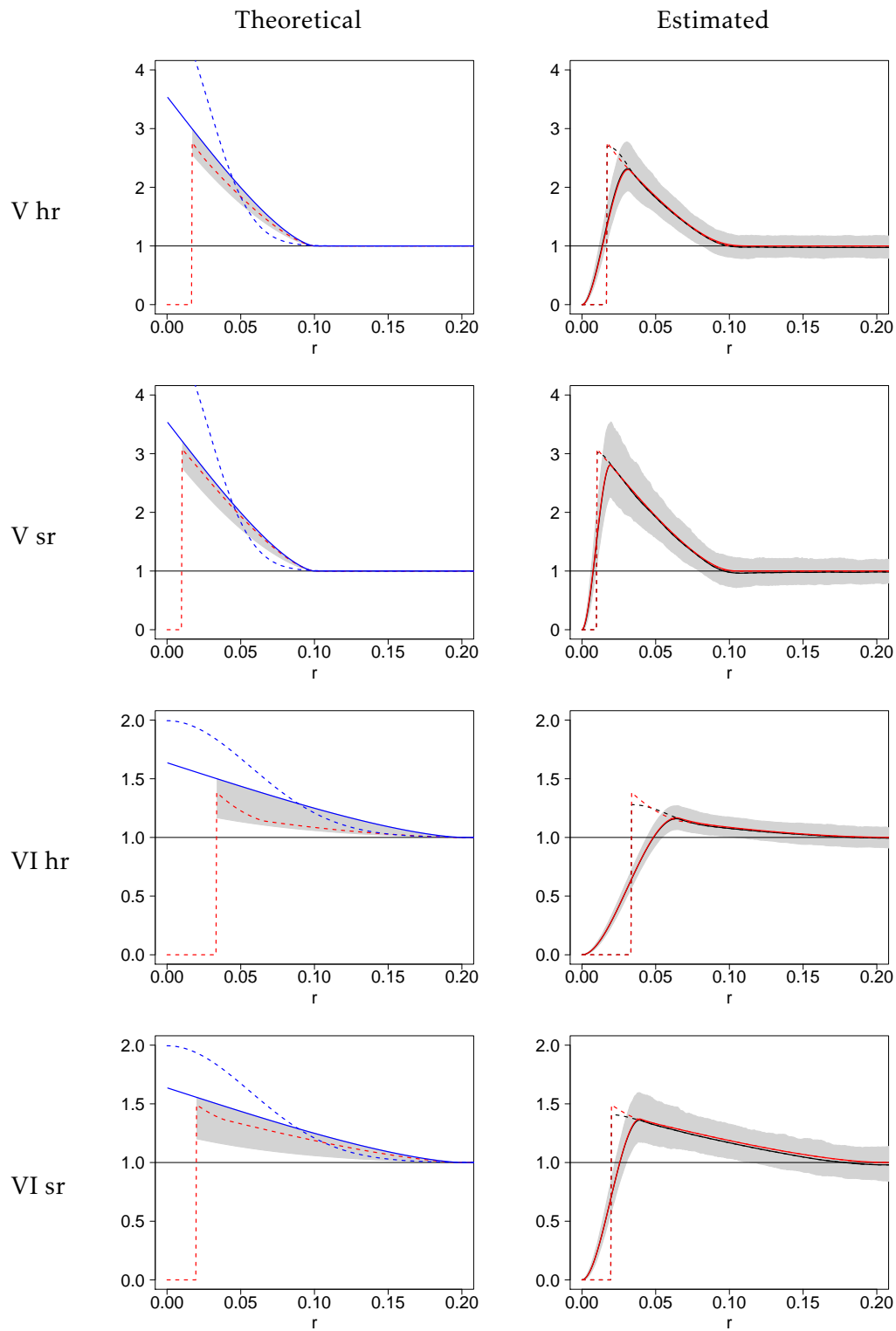
**Figure DD.7:** Theoretical and estimated pcfs for model V-VI. See Figure DD.2 and DD.3 for a detailed description.

**Figure DD.8:** Theoretical and estimated pcfs for model VII-VIII. See Figure DD.2 and DD.3 for a detailed description.

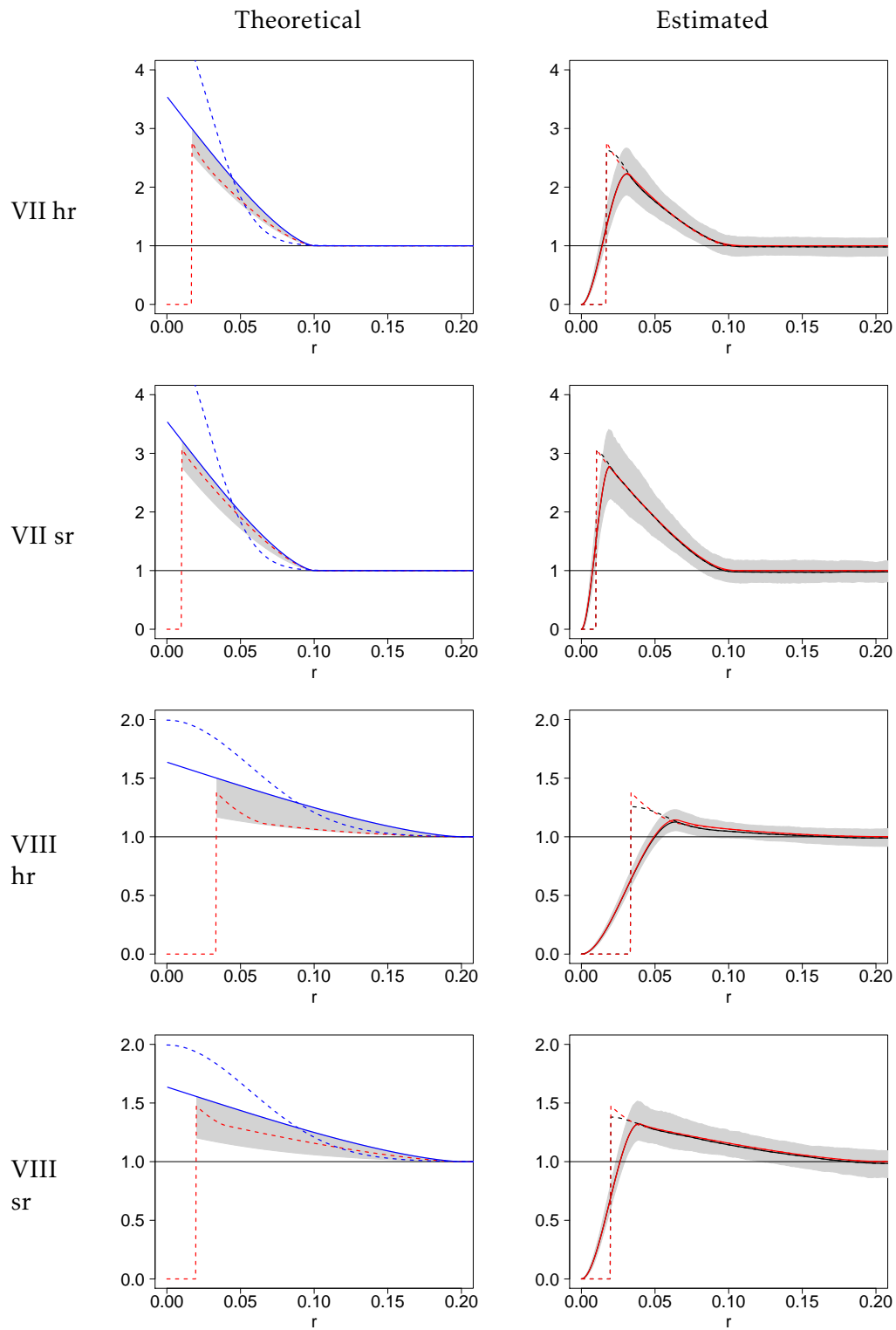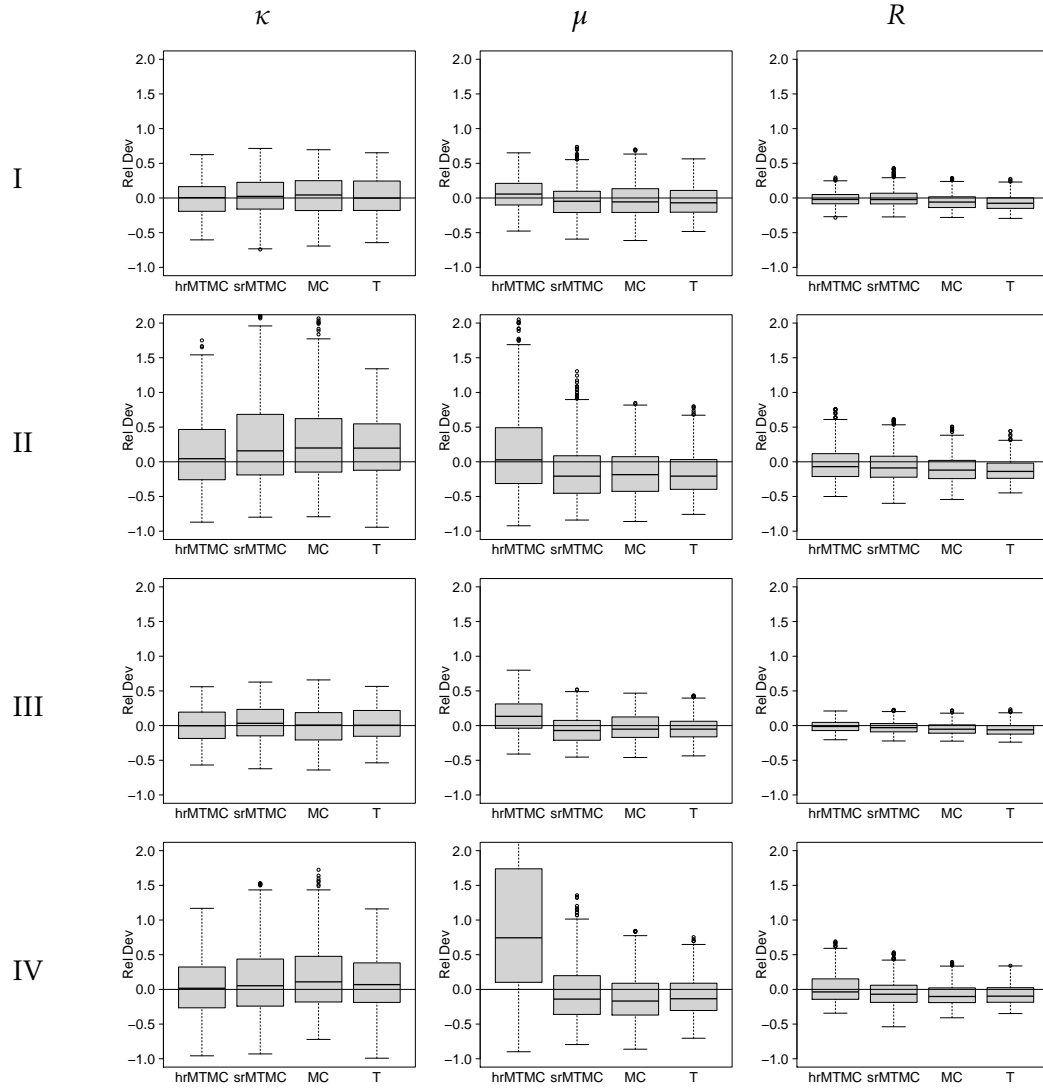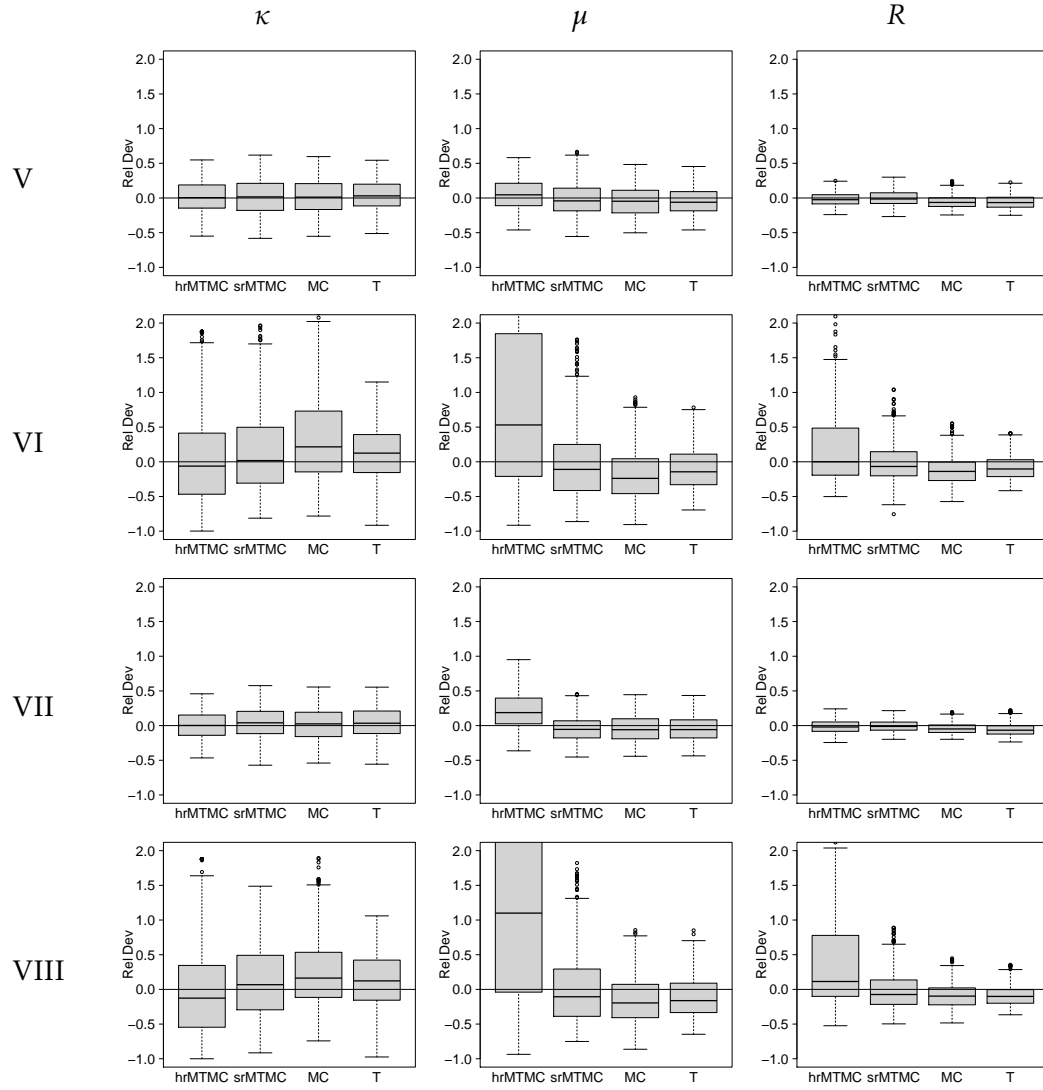**Figure DD.9:** Relative deviation of the estimates from the true parameters $\kappa$, $\mu$ and $R$ for all models I-IV in Table DD.1.

**Figure DD.10:** Relative deviation of the estimates from the true parameters $\kappa$, $\mu$ and $R$ for all models V-VIII in Table DD.1.

| RMB | $\kappa$ | $\rho$ | $R$ | $\sigma$ | hrMTMC | srMTMC | MCP | TP |
|---|---|---|---|---|---|---|---|---|
| $\hat{\kappa}$ | 25 | 100 | .05 | .02 | -.005 | .043 | .044 | .022 |
| | | | .10 | .04 | .100 | .186 | .229 | .178 |
| | | 150 | .05 | .02 | -.001 | .007 | .009 | .020 |
| | | | .10 | .04 | .099 | .162 | .217 | .113 |
| | 50 | 200 | .05 | .02 | .002 | .013 | .039 | .031 |
| | | | .10 | .04 | .095 | .150 | .349 | .158 |
| | | 300 | .05 | .02 | .009 | .031 | .027 | .009 |
| | | | .10 | .04 | .069 | .124 | .244 | .119 |
| $\hat{\mu}$ | 25 | 100 | .05 | .02 | .074 | -.029 | -.044 | -.040 |
| | | | .10 | .04 | .095 | -.055 | -.100 | -.131 |
| | | 150 | .05 | .02 | .129 | -.007 | -.033 | -.041 |
| | | | .10 | .04 | $\approx 1$ | -.089 | -.123 | -.090 |
| | 50 | 200 | .05 | .02 | .085 | -.017 | -.048 | -.037 |
| | | | .10 | .04 | $\approx 1$ | .004 | -.172 | -.109 |
| | | 300 | .05 | .02 | .218 | -.011 | -.032 | -.015 |
| | | | .10 | .04 | $\approx 2$ | .046 | -.119 | -.106 |
| $\hat{R}$ or $\hat{\sigma}$ | 25 | 100 | .05 | .02 | .009 | .001 | -.051 | -.059 |
| | | | .10 | .04 | -.028 | -.019 | -.075 | -.100 |
| | | 150 | .05 | .02 | .005 | .003 | -.042 | -.051 |
| | | | .10 | .04 | -.010 | -.041 | -.082 | -.082 |
| | 50 | 200 | .05 | .02 | .005 | -.012 | -.059 | -.054 |
| | | | .10 | .04 | $\approx 2 \times 10^2$ | .025 | -.118 | -.091 |
| | | 300 | .05 | .02 | -.008 | -.006 | -.049 | -.036 |
| | | | .10 | .04 | $\approx 5 \times 10^2$ | .071 | -.080 | -.086 |

**Table DD.2:** Relative mean bias.

| RMSE | $\kappa$ | $\rho$ | $R$ | $\sigma$ | hrMTMC | srMTMC | MCP | TP |
|------|------|------|------|------|--------|--------|-----|-----|
| $\hat{\kappa}$ | 25 | 100 | .05 | .02 | .074 | .086 | .098 | .073 |
|  |  |  | .10 | .04 | .260 | .355 | .449 | .227 |
|  |  | 150 | .05 | .02 | .058 | .070 | .073 | .057 |
|  |  |  | .10 | .04 | .209 | .257 | .295 | .166 |
|  | 50 | 200 | .05 | .02 | .054 | .061 | .075 | .054 |
|  |  |  | .10 | .04 | .501 | .356 | .598 | .192 |
|  |  | 300 | .05 | .02 | .041 | .061 | .052 | .050 |
|  |  |  | .10 | .04 | .540 | .327 | .371 | .160 |
| $\hat{\mu}$ | 25 | 100 | .05 | .02 | .074 | .058 | .057 | .042 |
|  |  |  | .10 | .04 | .295 | .225 | .202 | .132 |
|  |  | 150 | .05 | .02 | .072 | .050 | .048 | .032 |
|  |  |  | .10 | .04 | $\approx 5$ | .171 | .140 | .097 |
|  | 50 | 200 | .05 | .02 | .068 | .052 | .053 | .039 |
|  |  |  | .10 | .04 | $\approx 5$ | .316 | .187 | .107 |
|  |  | 300 | .05 | .02 | .106 | .048 | .035 | .037 |
|  |  |  | .10 | .04 | $\approx 12$ | .362 | .153 | .094 |
| $\hat{R}$ or $\hat{\sigma}$ | 25 | 100 | .05 | .02 | .018 | .013 | .012 | .017 |
|  |  |  | .10 | .04 | .076 | .053 | .054 | .042 |
|  |  | 150 | .05 | .02 | .007 | .011 | .010 | .013 |
|  |  |  | .10 | .04 | .048 | .043 | .034 | .029 |
|  | 50 | 200 | .05 | .02 | .013 | .010 | .014 | .013 |
|  |  |  | .10 | .04 | $\approx 1 \times 10^6$ | .112 | .064 | .035 |
|  |  | 300 | .05 | .02 | .007 | .009 | .009 | .010 |
|  |  |  | .10 | .04 | $\approx 3 \times 10^5$ | .140 | .038 | .029 |

**Table DD.3:** Relative mean squared error.

# Bibliography

Baddeley, A., E. Rubak, and R. Turner (2015). *Spatial Point Patterns: Methodology and Applications with R*. Chapman & Hall/CRC.

Diggle, P. J. (2003). *Statistical Analysis of Spatial Point Patterns. Second edition*. Oxford University Press, New York.

Dvořák, J. and M. Prokešová (2012). Moment estimation methods for stationary spatial Cox processes - A comparison. *Kybernetika* (5), 1007–1026.

Fiksel, T. (1988). Edge-corrected density estimators for point processes. *Statistics 19*(1), 67–75.

Illian, J., A. Penttinen, H. Stoyan, and D. Stoyan (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*, Volume 70. Wiley, Chichester.

Møller, J. and R. P. Waagepetersen (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC, Boca Raton, FL.

Ohser, J. and D. Stoyan (1981). On the second-order and orientation analysis of planar stationary point processes. *Biometrical Journal 23*(6), 523–533.

Stoyan, D. and H. Stoyan (1994). *Fractals, Random Shapes, and Point Fields: Methods of Geometrical Statistics*. Wiley Chichester.