



UNIVERSITY OF LEEDS

This is a repository copy of *Spatially explicit approach to estimation of total population abundance in field surveys.*

White Rose Research Online URL for this paper:

<http://eprints.whiterose.ac.uk/134548/>

Version: Accepted Version

Article:

Takashina, N, Kusumoto, B, Beger, M et al. (2 more authors) (2018) Spatially explicit approach to estimation of total population abundance in field surveys. *Journal of Theoretical Biology*, 453. pp. 88-95. ISSN 0022-5193

<https://doi.org/10.1016/j.jtbi.2018.05.013>

© 2018 Elsevier Ltd. Licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Spatially explicit approach to estimation of total population abundance in field surveys

Nao Takashina^{1*}, Buntarou Kusumoto², Maria Beger^{3,4}, Suren Rathnayake⁵, Hugh P.
Possingham^{3,6}

¹*Tropical Biosphere Research Center, University of the Ryukyus,
3422 Sesoko Motobu, Okinawa 905-0227, Japan*

²*Center for Strategic Research Project, University of the Ryukyus, 1 Senbaru, Nishihara, Okinawa
903-0213, Japan*

³*ARC Centre of Excellence for Environmental Decisions, School of Biological Sciences, The University of
Queensland, St Lucia, QLD 4072, Australia*

⁴*School of Biology, Faculty of Biological Sciences, University of Leeds, Leeds LS2 9JT, UK*

⁵*School of Biological Sciences, The University of Queensland, St Lucia, QLD 4072, Australia*

⁶*The Nature Conservancy, 4245 North Fairfax Drive Suite 100 Arlington, VA 22203-1606, USA*

Keywords: Field survey; population estimation; random sampling; spatial point process

*Corresponding author

Current address: Biodiversity and Biocomplexity Unit, Okinawa Institute of Science and Technology Graduate
University, Onna-son, Okinawa 904-0495, Japan

Email: nao.takashina@gmail.com

Abstract

The population abundance is fundamental in ecology and conservation biology, which provides essential information for predicting population dynamics and implementing conservation actions. While a range of approaches have been proposed to estimate population abundance based on existing data, data deficiency is ubiquitous. Where there is no information available, a population estimation will rely on labor intensive field surveys. Typically, time is one of the critical constraints in conservation and, often, management decisions must be made quickly under a data deficient situation. Hence, it is important to acquire a theoretical justification for survey method to meet a required estimation precision. However, there is no such a theory available in a spatially explicit context, while spatial considerations are critical to any field survey. Here, we develop a spatially explicit theory for population estimation, which allows us to examine the estimation precision under different survey designs and individual distribution patterns (e.g. random/clustered sampling and individual distribution). We demonstrate that the clustered sampling decreases the estimation precision when individuals form clusters, while sampling designs do not affect the estimation accuracy when individuals are distributed randomly. Regardless of the individual distribution, the estimation precision becomes higher with increasing total population abundance and the sampled fraction. Obtained insights provide theoretical bases for efficient field survey designs in information deficiency situations.

19 Introduction

20 Estimating the abundance of populations is important for ecological studies and conservation biol-
21 ogy [1–7], as is the role of ecosystem monitoring to observe changes in ecosystems [8–10]. In con-
22 servation, such knowledge helps one to estimate the risk of extinction of threatened species [11,12],
23 and to implement effective conservation actions [13].

24 While methods for statistically inferring population abundance with existing spatial data are
25 well developed [4–6, 14, 15], information on the abundance of threatened or rare species is often
26 rather limited and biased given budgetary constraints and access to remote sites [16, 17], requiring
27 further data collection or correction of data biases. For example, Reddy and Davalos [16] examined
28 an extensive data set of 1068 passerine birds in sub-Saharan Africa, and they found that data on
29 even well-known taxa are significantly biased to areas near cities and along rivers. Typically, time
30 is one of the critical constraints in conservation areas facing ongoing habitat loss and environmental
31 degradations [18]. In such cases, management decisions must be made quickly despite often having
32 only limited knowledge of a system [13, 19, 20]. On the other hand, for many ecological studies
33 and ecosystem monitoring programs, data must be accurate enough to be able to detect ecological
34 change [9]. Hence, given time and budgetary constraints and required precision of data, it is
35 desirable to set up an effective survey design to reduce time and effort of sampling.

36 Ultimately, we face trade-offs between data accuracy, time, and money. To tackle this trade-off
37 and provide generic insights to people designing a population survey, we need to handle different
38 sampling methods, choice of sampling unit scale, and data availability. However, most previous
39 approaches are spatially implicit (e.g., [5, 6, 14, 15, 21]), and it is therefore not straightforward to
40 compare the effect of different survey designs within a single theoretical framework applied. For
41 example, the negative binomial distribution (NBD) is frequently used to describe the underlying
42 individual distribution of a species. In the NBD, the parameter characterizing the degree of spatial
43 aggregation is scale dependent, and needs to be calibrated for each sampling unit scale. However,
44 this procedure is not intuitive and makes consistent comparison between survey designs difficult,
45 as the parameter characterizing aggregation is usually inferred from observed data rather than
46 biological mechanisms [14].

47 To develop generic insight into field survey performance under data deficient situations, we
48 develop a spatially explicit theory for population abundance estimation, which allows us to consis-
49 tently examine the estimation precision under various data collection schemes and different sampling
50 scales. Specifically, we examine simple random sampling and cluster sampling [22, 23] as popula-
51 tion sampling schemes. Cluster sampling reflects existing geographically biased sampling to some
52 extent, and hence, it is expected to give a general insight into prevalent field survey designs. These
53 sampling schemes are combined with spatial point processes (SPPs), a spatially explicit stochas-
54 tic model, to reveal effects of different survey designs as well as different individual distribution
55 patterns on the performance of population estimate. SPPs are widely applied in ecological studies
56 due to their flexibility, applicability to many ecological distribution, and availability of biological
57 interpretations [24–30]. Many examples come from studies of plant communities [24–26, 28, 29], but
58 others include studies of coral communities [31], and avian habitat selection to examine distribu-
59 tions of bird nests [32]. Although individual distributions often show clustering patterns in plant
60 and coral communities [25, 33–35], Bayard and Elphick [32] showed no statistical evidence of non-
61 random distributions in avian habitat selection at two salt marshes. Therefore, we examine both
62 clustering and random individual distribution patterns as example. By combining with sampling
63 strategies, we provide the general properties of "random/clustering sampling + random/clustering
64 individual distributions" without information on target species. Therefore, facing to a data defi-
65 cient situation, the best one can do is that merely assume if the species is randomly distributed or
66 forming clusters in space to develop sampling designs.

67 However, the method developed is general enough and suitable for any sampling of organism
68 or location used by an organism (e.g., nest and lek site) that is sedentary in space on a time scale
69 of the field survey where its spatial distributions can be described by SPPs. Hence, the results of
70 general sampling situation discussed may provide generic perspective of sampling designs.

71 **Methods**

72 In this analysis, we consider a situation where there is no prior spatial data available to infer the
73 distribution and abundance of a target species. We assume that our estimate of population size

74 is based only on field surveys where a fraction of sampled units α of the region of concern, W , is
75 surveyed using a sampling unit size, S (Fig. 1: Note we also use the notation R to represent region
76 in general. S is used when we specifically discuss the sampling unit.). We focus on a case where no
77 measurement error occurs in each sampling unit, suggesting that sampling units should be chosen
78 to ensure only trivial sampling errors in practice. It may vary for sampling in different systems.
79 For example, such an area may be larger for counting plant species compared to counting coral
80 species due to different visibility and accessibility of field surveys.

81 First, we introduce an estimator of population abundance, its expected value and variance,
82 which explicitly accounts for the effect of sampling unit size. These relevance to specific sampling
83 schemes and individuals distribution patterns are the main concern of this paper. Next, we explain
84 some basic properties of spatial point processes (SPPs), and models to describe spatial distribution
85 patterns of individuals. Using this framework, we test our analytical results formula for population
86 estimation.

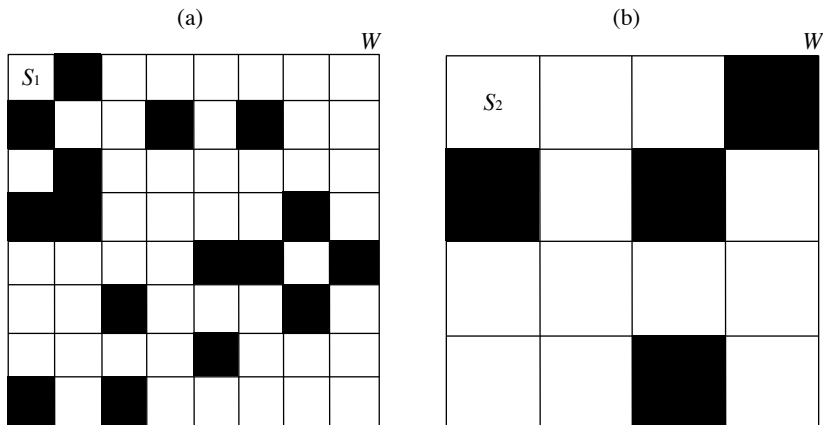


Figure 1: Example of simple random sampling with (a) smaller, and (b) larger sampling unit size, labeled S_1 and S_2 , respectively. The whole region of concern W is divided into sampling units with equal size, and a certain fraction α is randomly sampled (shaded unit) without replacement, where all sampling units have the equal probability of being chosen. Essentially, applying larger sampling units corresponds to a cluster sampling. The examples show the case of $\alpha = 0.25$.

87 Survey design

88 Given parameters specifying the survey design noted above, a simple random sampling (SRS)
89 without replacement [23] is conducted for collecting count data (Fig 1). In the SRS without
90 replacement, all the sampling units have an equal probability of being chosen. The number of
91 sampling units, N_t , and the sampled units, N_s , change with a sampling unit size, S . We assume all
92 the sampling units have an equal size. With larger sampling units, the degree of the geographical
93 sampling bias increases especially when the fraction of a sampled region is small (Fig 1). This
94 design corresponds to one-stage cluster sampling [23], where either all or none of the area within
95 the larger sampling units are in the sample. It is worth noting, however, that the degree of cluster
96 sampling is relative: any SRS can be considered to be cluster sampling if it is compared to SRS
97 with a smaller sampling unit size. In this article, we simply use these terms to imply that we are
98 using relatively *small* and *large* sampling units.

99 Population estimator

100 Following the data collection, we apply the unbiased linear estimator of the population abundance
101 in the region of concern W , $n(W)$ [22, 23],

$$\begin{aligned}\hat{n} | S &= \frac{N_t}{N_s} \sum_i^{N_s} y_i, \\ &= \frac{N_t}{N_s} \sum_k^{\infty} n_k k,\end{aligned}\tag{1}$$

102 where, $\hat{n} | S$ is the estimated population abundance given sampling unit size S , y_i is the number of
103 sampled individuals at the i th sampling trial, and n_k is the frequency of the sampled units holding
104 k individuals ($n_k = 0$ for large k because the number of individuals within each sampling unit is
105 finite). Note y_i and n_k change depending on the sampling unit size and underlying spatial point
106 patterns. In the SRS without replacement with the number of sampled units N_s , the frequency
107 n_k is only the random variable, following a multivariate hypergeometric distribution $p(n_k | S, N_s)$

108 with the mean $N_s p(k | S)$. Hence, the average population estimation \hat{n} is

$$\begin{aligned} \text{E}[\hat{n} | S] &= \frac{N_t}{N_s} \sum_k^{\infty} \text{E}[n_k | S] k, \\ &= N_t \text{E}[k | S]. \end{aligned} \quad (2)$$

109 The variance of the population estimate under the SRS without replacement is obtained by multi-
 110 plying the finite population correction (fpc) $:= (N_t - N_s)/(N_t - 1)$ [22] by the variance under the
 111 SRS with replacement:

$$\begin{aligned} \text{Var}[\hat{n} | S] &= (\text{fpc}) \left(\frac{N_t}{N_s} \right)^2 \left(\sum_k^{\infty} \text{Var}[n_k | S] k^2 + \sum_{\substack{k, k' \\ k \neq k'}}^{\infty} \text{Cov}[n_k n_{k'} | S] k k' \right), \\ &= \frac{N_t^2}{N_s} \left(\frac{N_t - N_s}{N_t - 1} \right) \text{Var}[k | S], \end{aligned} \quad (3)$$

112 where, the fact that the probability $p(n_k | S, N_s)$ follows a multinomial distribution with $\text{Var}[n_k | S] =$
 113 $N_s p(k | S, N_s) (1 - p(k | S, N_s))$ and $\text{Cov}[n_k n_{k'} | S] = -N_s p(k | S, N_s) p(k' | S, N_s)$ ($k \neq k'$) [36] are used.
 114 Therefore, the variance of the abundance estimate is determined by a constant multiplied by vari-
 115 ance of individual numbers in the sampling unit.

116 Spatial distribution of individuals

117 To account for explicit spatial distributions of individuals, we use spatial point processes (SPPs)
 118 [24, 29]. The underlying models used in our analysis are the homogeneous Poisson process and
 119 Thomas process, generating random and cluster distribution patterns of individuals, respectively.
 120 Properties of these processes are found in the literature (e.g., [24, 29, 37]) and, hence, we only
 121 introduce the properties relevant to our questions.

122 Homogeneous Poisson process

123 One of the simplest class of SPPs is the homogeneous Poisson process where the points (i.e. indi-
 124 viduals) are placed randomly within the region of concern and the number of points given in the

125 region R , $n(R)$, comes from a Poisson distribution with an average μ_R :

$$\text{Prob}(n(R) = k) = \frac{\mu_R^k}{k!} e^{-\mu_R}, \quad (k = 0, 1, \dots) \quad (4)$$

126 where, μ_R is known as the intensity measure [24, 29] defined by

$$\mu_R = \lambda \nu(R), \quad (5)$$

127 where, $\lambda := n(W)/\nu(W)$ is the intensity of individuals in the whole region W [29], and $\nu(R)$ is the
128 area of region R .

129 **Thomas process**

130 The Thomas process, characterizing the clustering pattern of individuals, belongs to the family of
131 Neyman-Scott processes [24, 29]. The Thomas process provides more general framework to address
132 spatial ecological patterns since most species are clumped in nature rather than random [38]. Even
133 though the model assumptions are minimal and does not assume a heterogeneous environment, it
134 creates patterns consistent with species that live in heterogeneous environment (e.g., [25, 28]). The
135 Thomas process is also amenable to an analytical approach, and therefore it is suitable to develop
136 mathematical understanding by minimizing model complexity [24, 25, 28–30]. The Thomas process
137 is obtained by the following three steps:

- 138 1. Parents are randomly placed according to the homogeneous Poisson process with a parent
139 intensity λ_p .
- 140 2. Each parent produces a random discrete number c of daughters, realized independently and
141 identically.
- 142 3. Daughters are scattered around their parents independently with an isotropic bivariate Gaus-
143 sian distribution with variance σ^2 , and all the parents are removed in the realized point
144 pattern.

145 The intensity of individuals for the Thomas process is [29]

$$\lambda_{th} = \bar{c}\lambda_p, \tag{6}$$

146 where, \bar{c} is the average number of daughters per parent. To allow population estimate comparisons
147 between the two SPPs, we chose the intensity of the Thomas process so as to have the same average
148 number of individuals within the region of concern W . Namely, the parameters λ_p and \bar{c} satisfy

$$\lambda_{th} = \bar{c}\lambda_p = \lambda. \tag{7}$$

149 We also assume that the number of daughters per parents c follows the Poisson distribution with
150 the average number \bar{c} .

151 Results

152 The total number of sampling units and sampled units are $N_t = \nu(W)/\nu(S)$ and $N_s = \lfloor \alpha N_t \rfloor$
153 respectively, where $\lfloor x \rfloor$ is the greatest integer not larger than x , and α is the fraction of sampled
154 units ($0 \leq \alpha \leq 1$). We are here interested in how the population estimates deviate from the true
155 value. Therefore, one of the quantities to show these effect may be

$$\frac{E[\hat{n} | S] \pm SE[\hat{n} | S]}{E[n(W)]}. \tag{8}$$

156 Note in the analysis below, we use $\lfloor \alpha N_t \rfloor = \alpha N_t$ for simplicity, but this approximation becomes
157 negligible when αN_t is sufficiently large.

158 Population estimation under the homogeneous Poisson distribution

159 For the homogeneous Poisson process, $\text{Var}[k|S]$ is equivalent to the variance of the Poisson process
160 with average $\lambda\nu(S)$. Therefore, by substituting this expression into Eq. (3) and with some algebra,

161 we obtain the SE of the population estimate of the homogeneous Poisson process

$$\text{SE}_{po}[\hat{n} | S] = \sqrt{n(W) \left(\frac{1}{\alpha} - 1 \right) \frac{N_t}{N_t - 1}}. \quad (9)$$

162 When the total number of sampling units is sufficiently large ($N_t \gg 1$), we obtain the simpler form

$$\text{SE}_{po}[\hat{n} | S] \simeq \sqrt{n(W) \left(\frac{1}{\alpha} - 1 \right)}. \quad (10)$$

163 Under such circumstances, the standard error of the abundance estimation is only the function of
 164 the expected population total existing in the concerned region $n(W)$ and the sampling fraction α ;
 165 and does not depend on the sampling unit size. Therefore, we can write $\text{SE}_{po}[\hat{n} | S] = \text{SE}_{po}[\hat{n}]$. Due
 166 to the term $n(W)^{1/2}$ in $\text{SE}_{po}[\hat{n} | S]$, the relative variation from its average decreases with the factor
 167 $(1/\alpha - 1)^{1/2} n(W)^{-1/2}$. These results were confirmed by numerical simulations, and they show good
 168 agreement with analytical results (Fig. 2).

169 Population estimation under the Thomas process

170 For the Thomas process, deriving a theoretical form of the variance of individuals given across
 171 sampling scales, $\text{Var}[k|S]$, is challenging, although the probability generating functional of the
 172 Thomas process is known, e.g., [29]. Instead, we apply an approximated pdf of the Thomas process
 173 to obtain an explicit form of $\text{Var}[k|S]$. By assuming that each daughter location has no correlation
 174 to its sisters locations, we derive the approximated pdf of the Thomas process (see Appendix for
 175 the detailed derivations):

$$p(n | S) = \sum_k \text{Po}(k, \lambda_p \nu(S^t)) \text{Po}(n, k \bar{c} p_d(S)). \quad (11)$$

176 where, $\text{Po}(k, \lambda)$ is the Poisson distribution with the intensity λ , and $p_d(S)$ is the probability that
 177 an individual daughter produced by a parent situated in the region, $S + S_{out}$, falls in S . S_{out} is
 178 the surrounding region of S where parents can potentially supply daughters to the region S (See
 179 Appendix for the detailed definition of S_{out}). This probability is determined by the dispersal kernel

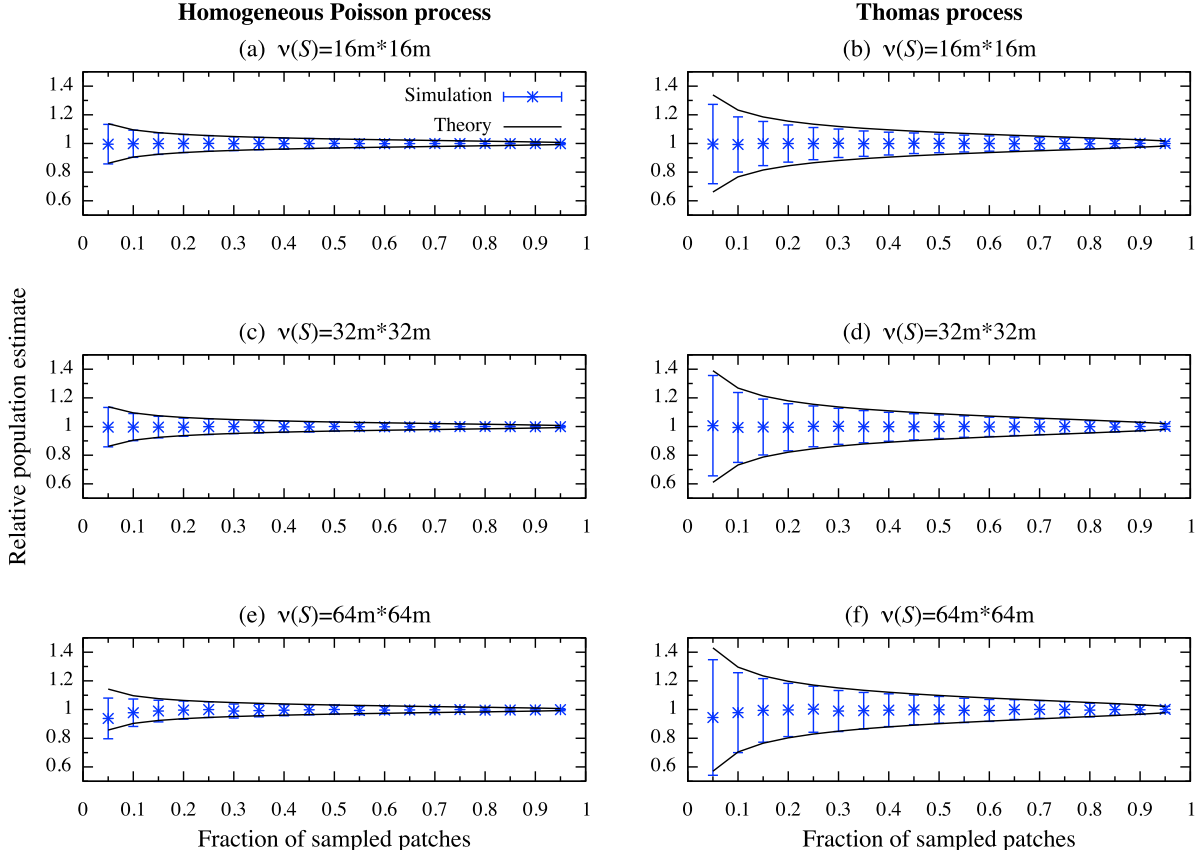


Figure 2: Relative value of the population estimate with the average individuals $E[n(W)] = 10^3$ under the three sampling scales. Larger sampling area implies more cluster sampling. Each panel shows relative average estimate \pm relative standard error (Eq. (8)) of simulation and theoretical results. Relative average estimate for theoretical results is omitted since it is an unbiased estimator. The parameter values used are $\bar{c} = 10$, $\sigma = 10$, and $\nu(W) = 2^{20}\text{m}^2$ ($1024\text{m} \times 1024\text{m}$).

180 (See Eq. (A.3) in Appendix), and therefore, closely related to dispersal distance of the species.
 181 Thomas [39] refers to the form of Eq. (11) as the *double Poisson distribution*, in derivations of her
 182 original Thomas model, in which spatial effects are implicitly described. On the other hand, Eq.
 183 (11) explicitly handles spatial effect, such as the size of sampling unit S and the effect of dispersal
 184 $p_d(S)$. Eq. (11) enables us to derive an approximated form of $SE_{th}[\hat{n}|S]$ (see Appendix for detailed
 185 derivation):

$$SE_{th}[\hat{n}|S] = SE_{po}[\hat{n}|S] \sqrt{\frac{\nu(S')}{\nu(S)} p_d(S) (1 + \bar{c} p_d(S))}. \quad (12)$$

186 This equation suggests that the standard error of the Thomas process, $SE_{th}[\hat{n}|S]$, is described
 187 by the multiplication of $SE_{po}[\hat{n}|S]$ and a term characterizing the degree of cluster of the Thomas
 188 process. Therefore, the similar discussions made for $SE_{po}[\hat{n}|S]$ can also be applied to $SE_{th}[\hat{n}|S]$.
 189 Especially, the effect of the expected population abundance $n(W)$ on the relative variation holds
 190 true in this situation. Eq. (12) suggest that increasing the average number of daughters, \bar{c} , increases
 191 the standard error. In addition, by definition of $p_d(S)$ Eq. (A.3), a smaller value of σ increases
 192 $p_d(S)$. Roughly speaking, a species with a large expected number of daughters, \bar{c} , and smaller
 193 dispersal distance of daughters, σ , form a high degree of clusters in individual distributions, and
 194 it increases the standard error of the population estimate $SE_{th}[\hat{n}|S]$. The approximated $SE_{th}[\hat{n}|S]$,
 195 Eq. (12), shows good agreement with the values obtained by the numerical simulations across
 196 sampling areas, although it shows slight deviations from the numerical values when the fraction of
 197 sampling patches is small (α is around 0.05-0.1; Fig. 2). Typically, increasing the sampling unit
 198 size (i.e., more clustered sampling) in population estimations increases the standard error, but it
 199 decreases with the fraction of sampled patches. We also confirmed the similar agreement between
 200 Eq. (12) and numerical simulations with different parameters (Fig. A.2).

201 Discussion

202 We examined a method for population estimation combined with spatial point processes (SPPs),
 203 spatially explicit model, to reveal effects of different survey regimes as well as individual distribu-

204 tion patterns on the precision of population estimates. By assuming the random and clustering
205 placements of individuals as underlying distribution patterns, we analytically show that the indi-
206 vidual distributions and sampling schemes, such as random sampling and cluster sampling, change
207 significantly the standard error of the abundance estimate. In our sampling framework, increasing
208 the sampling unit size corresponds to an increase of geographical bias of the sampling (i.e., cluster
209 sampling; see Survey design). Typically, we find that the standard error of the abundance estimate
210 is insensitive to the sampling unit size applied when the underlying individual distribution is the
211 homogeneous Poisson process. On the other hand, the Thomas process analysis suggests that popu-
212 lation estimate will result in less precise population estimates. Typically, under clustered individual
213 distributions, the standard error increases as the degree of clustering sampling increases. We also
214 show that the standard error of the population estimate increases with the parameter characteriz-
215 ing the degree of clustering of individual distributions. In addition, although for both individual
216 distribution patterns, our results show that the absolute value of the standard error increases with
217 the number of individuals, the relative standard error decreases with the factor proportional to
218 $n(W)^{-1/2}$.

219 In practice, simple random sampling with a fine sampling unit may not easily be conducted
220 due to time and budgetary constraints, and different accessibility to sites [16,23,40]. However, this
221 sampling scheme enables us to obtain more reliable data since extensive sampling in inaccessible
222 region may also lead to new discoveries [16]. Hence, this sampling scheme may be suitable for
223 many ecological studies and ecosystem monitoring projects which require estimations to capture
224 spatial and/or temporal patterns of the population. Alternatively, cluster sampling, which causes
225 a geographical sampling bias, is often the favored survey design practically since it is less expensive
226 and easy to implement [16,23]. Therefore, this survey design may be applied to managements where
227 a target species require quick conservation action at a cost of precision of data. Most importantly,
228 in line with the discussion of Takashina et al. [30], insights developed in the paper should be applied,
229 by clearly setting a feasible goal of population estimate with time and economic constraints, before
230 survey designs are developed.

231 Here we investigate population estimation with assuming that no data is available and with

232 general ecological and sampling assumptions. However, our results provide generic insights into
233 ecological survey design such as how the sampling unit size used and individual distribution patterns
234 affect the precision of population estimation. Typically, it suggests that more clustered samplings
235 and/or more clustered individual distributions cause less precise population estimations, but the
236 precision improves with the fraction of sampled patches. For both ecological and conservation
237 applications in mind, our sampling framework is kept as general as possible. Therefore, it allows
238 one to further extend the framework to handle more complex situations where, for example, the
239 concerned region holds multiple sampling unit sizes or a budgetary constraint is explicitly taken
240 into consideration. Also, SPPs is not a only choice in our framework, but one can also use any
241 spatially explicit models as long as the model allows to calculate Eq. (3). Especially, for analytical
242 tractability, we focused on how individual distributions and sampling strategies affect the accuracy
243 of population estimate by assuming no or sufficiently small measurement error. Although many
244 empirical studies have adopted this assumption [41], imperfect detection is also frequently observed
245 even in sessile organisms such as plants (e.g. [42, 43]). Also, if searching time is fixed, chance of
246 imperfect detection would increase with survey area [44]. This indicates that the sampling unit size
247 should be chosen while taking the scale-dependency of the imperfect detection into account. Further
248 studies about how imperfect detection changes our predictions is highly beneficial for developing
249 robust survey designs.

250 **Acknowledgements**

251 We would like to thank T. Fung and B. Stewart-Koster for their thoughtful comments. NT and
252 BK were funded by the Program for Advancing Strategic International Networks to Accelerate the
253 Circulation of Talented Researchers of the Japan Society for the Promotion of Science, and they
254 acknowledge the support for coordinating the research program from Dr Yasuhiro Kubota and Dr
255 James D. Reimer.

Literature Cited

- [1] E. C. Pielou, An introduction to mathematical ecology, Wiley-Interscience, New York, 1969.
- [2] E. C. Pielou, Ecological Diversity, John Wiley and Sons Inc, New York, 1975.
- [3] D. L. Otis, K. P. Burnham, G. C. White, D. R. Anderson, Statistical Inference from Capture Data on Closed Animal Populations, *Source Wildl. Monogr.* (62) (1978) 3–135. doi:10.2307/2287873.
- [4] W. E. Kunin, Extrapolating species abundance across spatial scales, *Science* 281 (5382) (1998) 1513–1515. doi:10.1126/science.281.5382.1513.
- [5] F. He, K. J. Gaston, Estimating Species Abundance from Occurrence, *Am. Nat.* 156 (5) (2000) 553–559. doi:10.1086/303403.
- [6] F. L. He, K. J. Gaston, Occupancy-abundance relationships and sampling scales, *Ecography* (Cop.). 23 (4) (2000) 503–511. doi:10.1111/j.1600-0587.2000.tb00306.x.
- [7] K. H. Pollock, J. D. Nichols, T. R. Simons, G. L. Farnsworth, L. L. Bailey, J. R. Sauer, Large scale wildlife monitoring studies: Statistical methods for design and analysis, *Environmetrics* 13 (2) (2002) 105–119. doi:10.1002/env.514.
- [8] B. Goldsmith, *Monitoring for Conservation and Ecology*, Vol. 3, Chapman & Hall, 1991. doi:10.1016/0305-1978(91)90074-A.
- [9] D. B. Lindenmayer, G. E. Likens, Adaptive monitoring: a new paradigm for long-term research and monitoring, *Trends Ecol. Evol.* 24 (9) (2009) 482–486. doi:10.1016/j.tree.2009.03.005.
- [10] S. H. M. Butchart, M. Walpole, B. Collen, A. van Strien, J. P. W. Scharlemann, R. E. A. Almond, J. E. M. Baillie, B. Bomhard, C. Brown, J. Bruno, K. E. Carpenter, G. M. Carr, J. Chanson, A. M. Chenery, J. Csirke, N. C. Davidson, F. Dentener, M. Foster, A. Galli, J. N. Galloway, P. Genovesi, R. D. Gregory, M. Hockings, V. Kapos, J.-F. Lamarque, F. Leverington,

- 280 J. Loh, M. A. McGeoch, L. McRae, A. Minasyan, M. H. Morcillo, T. E. E. Oldfield, D. Pauly,
281 S. Quader, C. Revenga, J. R. Sauer, B. Skolnik, D. Spear, D. Stanwell-Smith, S. N. Stuart,
282 A. Symes, M. Tierney, T. D. Tyrrell, J.-C. Vie, R. Watson, Global Biodiversity: Indicators of
283 Recent Declines, *Science* 328 (5982) (2010) 1164–1168. doi:10.1126/science.1187512.
- 284 [11] D. I. MacKenzie, J. D. Nichols, J. E. Hines, M. G. Knutson, A. B. Franklin, Estimating site
285 occupancy, colonization, and local extinction when a species is detected imperfectly, *Ecology*
286 84 (8) (2003) 2200–2207. doi:10.1890/02-3090.
- 287 [12] M. A. McCarthy, S. J. Andelman, H. P. Possingham, Reliability of Relative Predictions in Pop-
288 ulation Viability Analysis, *Conserv. Biol.* 17 (4) (2003) 982–989. doi:10.1046/j.1523-1739.
289 2003.01570.x.
- 290 [13] H. S. Grantham, K. A. Wilson, A. Moilanen, T. Rebelo, H. P. Possingham, Delaying conser-
291 vation actions for improved knowledge: How long should we wait?, *Ecol. Lett.* 12 (4) (2009)
292 293–301. doi:10.1111/j.1461-0248.2009.01287.x.
- 293 [14] E. Kuno, Evaluation of statistical precision and design of efficient sampling for the population
294 estimation based on frequency of occurrence, *Res. Popul. Ecol. (Kyoto)*. 28 (1986) 305–319.
- 295 [15] J. A. Royle, J. D. Nichols, Estimating abundance from repeated presenceabsence data or point
296 counts, *Ecology* 84 (3) (2003) 777–790. doi:10.1890/0012-9658(2003)084[0777:EAFRPA]2.
297 0.CO;2.
- 298 [16] S. Reddy, L. M. Dávalos, Geographical sampling bias and its implications for conservation
299 priorities in Africa, *J. Biogeogr.* 30 (11) (2003) 1719–1727. doi:10.1046/j.1365-2699.2003.
300 00946.x.
- 301 [17] T. A. Gardner, J. Barlow, I. S. Araujo, T. C. Ávila-Pires, A. B. Bonaldo, J. E. Costa, M. C.
302 Esposito, L. V. Ferreira, J. Hawes, M. I. M. Hernandez, M. S. Hoogmoed, R. N. Leite, N. F.
303 Lo-Man-Hung, J. R. Malcolm, M. B. Martins, L. A. M. Mestre, R. Miranda-Santos, W. L.
304 Overal, L. Parry, S. L. Peters, M. A. Ribeiro, M. N. F. Da Silva, C. Da Silva Motta, C. A.

- 305 Peres, The cost-effectiveness of biodiversity surveys in tropical forests, *Ecol. Lett.* 11 (2) (2008)
306 139–150. [doi:10.1111/j.1461-0248.2007.01133.x](https://doi.org/10.1111/j.1461-0248.2007.01133.x).
- 307 [18] A. E. Camaclang, M. Maron, T. G. Martin, H. P. Possingham, Current practices in the
308 identification of critical habitat for threatened species, *Conserv. Biol.* 29 (2) (2015) 482–492.
309 [doi:10.1111/cobi.12428](https://doi.org/10.1111/cobi.12428).
- 310 [19] M. Bode, K. A. Wilson, T. M. Brooks, W. R. Turner, R. A. Mittermeier, M. F. McBride,
311 E. C. Underwood, H. P. Possingham, Cost-effective global conservation spending is robust to
312 taxonomic group., *Proc. Natl. Acad. Sci. U. S. A.* 105 (17) (2008) 6498–6501. [doi:10.1073/
313 pnas.0710705105](https://doi.org/10.1073/pnas.0710705105).
- 314 [20] A. Hastings, Timescales and the management of ecological systems, *Proc. Natl. Acad. Sci.* [doi:](https://doi.org/10.1073/pnas.1604974113)
315 [10.1073/pnas.1604974113](https://doi.org/10.1073/pnas.1604974113).
- 316 [21] J. L. Green, J. B. Plotkin, A statistical theory for sampling species abundances, *Ecol. Lett.*
317 10 (11) (2007) 1037–1045. [doi:10.1111/j.1461-0248.2007.01101.x](https://doi.org/10.1111/j.1461-0248.2007.01101.x).
- 318 [22] J. Rice, *Mathematical Statistics and Data Analysis*, 3rd Edition, Vol. 72, Thomson Higher
319 Education, Belmont, 2007. [doi:10.2307/3619963](https://doi.org/10.2307/3619963).
- 320 [23] S. Lohr, *Sampling: design and analysis*, Nelson Education, 2009.
- 321 [24] N. A. C. Cressie, *Statistics for Spatial Data*, John Wiley & Sons, New York, 1993.
- 322 [25] J. B. Plotkin, M. D. Potts, N. Leslie, N. Manokaran, J. Lafrankie, P. S. Ashton, Species-
323 area curves, spatial aggregation, and habitat specialization in tropical forests., *J. Theor. Biol.*
324 207 (1) (2000) 81–99. [doi:10.1006/jtbi.2000.2158](https://doi.org/10.1006/jtbi.2000.2158).
- 325 [26] K. Shimatani, Y. Kubota, Spatial analysis for continuously changing point patterns along
326 a gradient and its application to an *Abies sachalinensis* population, *Ecol. Modell.* 180 (2-3)
327 (2004) 359–369. [doi:10.1016/j.ecolmodel.2004.04.036](https://doi.org/10.1016/j.ecolmodel.2004.04.036).
- 328 [27] N. Picard, C. Favier, A Point-Process Model for Variance-Occupancy-Abundance Relation-
329 ships, *Am. Nat.* 178 (3) (2012) 383–396. [doi:10.1086/661249](https://doi.org/10.1086/661249).

- 330 [28] S. Azaele, S. J. Cornell, W. E. Kunin, Downscaling species occupancy from coarse spatial
331 scales, *Ecol. Appl.* 22 (3) (2012) 1004–1014. doi:[10.1890/11-0536.1](https://doi.org/10.1890/11-0536.1).
- 332 [29] S. N. Chiu, D. Stoyan, W. S. Kendall, J. Mecke, *Stochastic Geometry and Its Applications*,
333 John Wiley & Sons, New York, 2013.
- 334 [30] N. Takashina, M. Beger, B. Kusumoto, S. Rathnayake, H. Possingham, A theory for ecological
335 survey methods to map individual distributions, *Theor. Ecol.* doi:[https://doi.org/10.1007/
336 s12080-017-0359-7](https://doi.org/10.1007/s12080-017-0359-7).
- 337 [31] S. Muko, K. Shimatani, Y. Nozawa, *Spatial analyses for nonoverlapping objects with size
338 variations and their application to coral communities* (2014).
- 339 [32] T. S. Bayard, C. S. Elphick, Using Spatial Point-Pattern Assessment to Understand the Social
340 and Environmental Mechanisms that Drive Avian Habitat Selection, *Auk* 127 (3) (2010) 485–
341 494. doi:[10.1525/auk.2010.09089](https://doi.org/10.1525/auk.2010.09089).
- 342 [33] L. R. Taylor, I. P. Woiwod, J. N. Perry, The Density-Dependence of Spatial Behaviour and
343 the Rarity of Randomness, *J. Anim. Ecol.* 47 (2) (1978) pp. 383–406. doi:[10.2307/3790](https://doi.org/10.2307/3790).
- 344 [34] R. Condit, P. S. Ashton, P. Baker, S. Bunyavejchewin, C. V. S. Gunatilleke, I. A. U. N.
345 Gunatilleke, S. P. Hubbell, R. B. Foster, A. Itoh, J. V. LaFrankie, H. S. Lee, E. Losos,
346 N. Manokaran, R. Sukumar, T. Yamakura, Spatial patterns in the distribution of tropical
347 tree species., *Science* 288 (1982) (2000) 1414–1418. doi:[10.1126/science.288.5470.1414](https://doi.org/10.1126/science.288.5470.1414).
- 348 [35] M. Beger, G. P. Jones, P. L. Munday, Conservation of coral reef biodiversity: A comparison
349 of reserve selection procedures for corals and fishes, *Biol. Conserv.* 111 (1) (2003) 53–62.
350 doi:[10.1016/S0006-3207\(02\)00249-5](https://doi.org/10.1016/S0006-3207(02)00249-5).
- 351 [36] Multinomial distribution. Encyclopedia of Mathematics. URL:
352 http://www.encyclopediaofmath.org/index.php?title=Multinomial_distribution&oldid=28544
353 Data Accessed: Mar 18 2018.

- 354 [37] J. Illian, A. Penttinen, H. Stoyan, D. Stoyan, *Statistical Analysis and Modelling of Spatial*
355 *Point Patterns*, Vol. 76, John Wiley & Sons, Chichester, 2008. doi:10.1002/9780470725160.
- 356 [38] R. Condit, P. S. Ashton, P. Baker, S. Bunyavejchewin, S. Gunatilleke, N. Gunatilleke, S. P.
357 Hubbell, R. B. Foster, A. Itoh, J. V. LaFrankie, H. S. Lee, E. Losos, N. Manokaran, R. Suku-
358 mar, T. Yamakura, Spatial patterns in the distribution of tropical tree species, *Science*
359 288 (1982) (2000) 1414–1418. doi:10.1126/science.288.5470.1414.
- 360 [39] M. Thomas, A generalization of Poisson’s binomial limit for use in ecology, *Biometrika* 36(1/2)
361 (1949) 18–25.
- 362 [40] H. Possingham, I. Ball, S. Andelman, Mathematical methods for identifying representative
363 reserve networks, in: *Quant. Methods Conserv. Biol.*, Springer New York, 2000, pp. 291–306.
364 doi:10.1007/0-387-22648-6_17.
- 365 [41] K. F. Kellner, R. K. Swihart, Accounting for imperfect detection in ecology: A quantitative
366 review (2014). doi:10.1371/journal.pone.0111436.
- 367 [42] G. Chen, M. Kéry, J. Zhang, K. Ma, Factors affecting detection probability in plant distribution
368 studies, *J. Ecol.* 97 (6) (2009) 1383–1389. doi:10.1111/j.1365-2745.2009.01560.x.
- 369 [43] G. Chen, M. Kéry, M. Plattner, K. Ma, B. Gardner, Imperfect detection is the rule rather
370 than the exception in plant distribution studies, *J. Ecol.* 101 (1) (2013) 183–191. doi:10.
371 1111/1365-2745.12021.
- 372 [44] K. D. Clarke, M. Lewis, R. Brandle, B. Ostendorf, Non-detection errors in a survey of per-
373 sistent, highly-detectable vegetation species, *Environ. Monit. Assess.* 184 (2) (2012) 625–635.
374 doi:10.1007/s10661-011-1991-0.

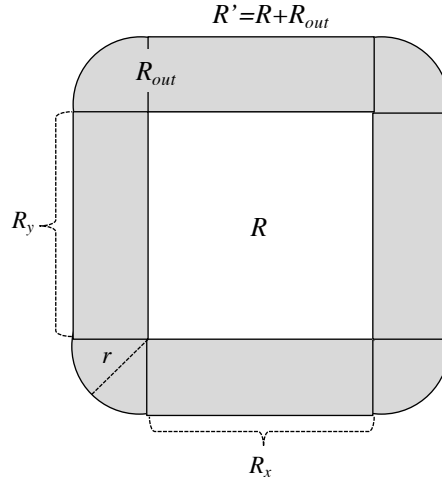


Figure A.1: R is the concerned region with area $R_x \times R_y$. Parents outside R with a distance less than r from the edges of R (parents in R_{out}) may also contribute to the number of daughters in the concerned region R . The whole region where parents can supply daughters to R is $R' = R + R_{out}$.

375 Appendix

376 Derivations of an approximated pdf of the Thomas process

377 Here, we derive an approximated form of the probability distribution function (pdf) of the Thomas
 378 process. For this purpose, we firstly introduce two regions R' and R_{out} . Let R' be the region where
 379 a parent potentially supplies the daughters to the region R . Then R' is decomposed into two regions
 380 $R' = R + R_{out}$, where R_{out} is the surrounding region of R and satisfies with $R' \setminus R$ (Fig. A.1). Here,
 381 we approximate the probability that n individuals fall in the region R with k' individuals produced
 382 by parents in R' by the binomial distribution, though sisters (i.e., daughters share a same parent)
 383 locations depend on its parent location. Under this assumption, the probability that n individuals

384 found in region R is described

$$\begin{aligned}
p(n|R) &= \underbrace{\sum_k \frac{(\lambda_p \nu(R'))^k}{k!} e^{-\lambda_p \nu(R')}}_{\text{no. parents in } R'} \underbrace{\sum_{k'} \binom{k'}{n} p_d(R)^n (1-p_d(R))^{k'-n} \sum_{k' \in \mathcal{K}} \prod_i \frac{\bar{c}_i^{k'_i}}{k'_i!} e^{-\bar{c}}}_{\text{Prob}(n \text{ daughters fall in } R \text{ provided } k' \text{ daughters produced by parents in } R')} \quad (\text{A.1}) \\
&= \sum_k \frac{(\lambda_p \nu(R'))^k}{k!} e^{-\lambda_p \nu(R')} \sum_{k'} \binom{k'}{n} p_d(R)^n (1-p_d(R))^{-n} e^{-\bar{c}k} \sum_{k' \in \mathcal{K}} \prod_i \frac{\{\bar{c}(1-p_d(R))\}^{k'_i}}{k'_i!},
\end{aligned}$$

385 where, $k' = k'_1 + \dots + k'_k$ and k'_i is the number of daughters produced by parent i . $\sum_{k' \in \mathcal{K}}$ runs all
386 the combinations of k' satisfies $\sum_i k'_i = k'$. As one can easily see $\sum_{k' \in \mathcal{K}} k'! \prod_i \{\bar{c}(1-p)\}^{k'_i} / k'_i!$ is
387 the coefficient of expansion of $(\lambda_1 + \dots + \lambda_k)^{k'_1 + \dots + k'_k}$, where we set $\lambda_1 = \dots = \lambda_k = \bar{c}(1-p_d(R))$.

388 Therefore, Eq. (A.1) becomes

$$\begin{aligned}
p(n|R) &= \sum_k \frac{(\lambda_p \nu(R'))^k}{k!} e^{-\lambda_p \nu(R')} \sum_{k'} \frac{1}{(k'-n)!n!} p_d(R)^n (1-p_d(R))^{-n} e^{-\bar{c}k} \sum_{k' \in \mathcal{K}} k'! \prod_i \frac{\{\bar{c}(1-p_d(R))\}^{k'_i}}{k'_i!}, \\
&= \sum_k \frac{(\lambda_p \nu(R'))^k}{k!} e^{-\lambda_p \nu(R')} \frac{1}{n!} p_d(R)^n (1-p_d(R))^{-n} e^{-\bar{c}k} \sum_{k'} \frac{(\bar{c}k(1-p))^{k'}}{(k'-n)!}, \\
&= \sum_k \frac{(\lambda_p \nu(R'))^k}{k!} e^{-\lambda_p \nu(R')} \frac{(\bar{c}k(1-p_d(R)))^n}{n!} p_d(R)^n (1-p)^{-n} e^{-\bar{c}k} \sum_{k'} \frac{(\bar{c}k(1-p))^{k'-n}}{(k'-n)!}, \\
&= \sum_k \frac{(\lambda_p \nu(R'))^k}{k!} e^{-\lambda_p \nu(R')} \frac{(\bar{c}k)^n}{n!} p_d(R)^n e^{-\bar{c}k} e^{\bar{c}k(1-p_d(R))}, \\
&= \sum_k \frac{(\lambda_p \nu(R'))^k}{k!} e^{-\lambda_p \nu(R')} \frac{(\bar{c}k p_d(R))^n}{n!} e^{-\bar{c}k p_d(R)}, \\
&= \sum_k \text{Po}(k, \lambda_p \nu(R')) \text{Po}(n, k \bar{c} p_d(R)). \quad (\text{A.2})
\end{aligned}$$

389 where, $\text{Po}(k, \lambda)$ is the poisson distribution with the intensity λ and $p_d(R)$ is the probability that an
390 individual daughter produced by a parent within R' falls in R . Since a parent location is randomly
391 chosen in R' , we calculate $p_d(R)$ as follows

$$p_d(R) = \frac{1}{\nu(R')} \int_{R'} \int_R \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma^2}\right) d\mathbf{x}d\mathbf{y}, \quad (\text{A.3})$$

392 where, \mathbf{x} and \mathbf{y} are location in R and R' , respectively. By referring Fig. A.1, $\nu(R')$ is calculated as

$$\nu(R') = (2r + R_x)(2r + R_y) - r^2(4 - \pi), \quad (\text{A.4})$$

393 where, r is the distance that on average a fraction u of daughters scattered by the parent (placed
 394 center) are covered. r is calculated by converting the expression of the isotropic bivariate gaussian
 395 on cartesian coordinates, $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx dy 1/(2\pi\sigma^2) \exp\{-(x^2 + y^2)/(2\sigma^2)\}$, to the one on the polar
 396 coordinates, and solving about r

$$r = \sqrt{-2\sigma^2 \log(1 - u)}, \quad (\text{A.5})$$

397 where, in the analysis, we set $u = 0.99$ (i.e., 99% of daughters fall within this distance).

398 **Standard error of the Thomas process**

399 Using Eq. (A.2), we calculate the first moment and the second moment of the point number k in
 400 region R

$$\text{E}[n(R)] = \lambda_p \bar{c} p_d(R) \nu(R') = \lambda_p \bar{c} \nu(R), \quad (\text{A.6})$$

$$\text{E}[n(R)^2] = \lambda_p \bar{c} p_d(R) \nu(R') (1 + \bar{c} p_d(R) + \lambda_p \bar{c} p_d(R) \nu(R')). \quad (\text{A.7})$$

401 Using Eqs (3), (9), (A.6), and (A.7) and the fact $\lambda_p \bar{c} = \lambda = n(W)/\nu(W)$, $N_t = \nu(W)/\nu(S)$, and
 402 $N_s = \alpha N_t$, we calculate Eq. (12) as follows:

$$\begin{aligned} \text{SE}_{th}[\hat{X}|S] &= \sqrt{\lambda_p \bar{c} p_d(S) \nu(S') (1 + \bar{c} p_d(S)) \frac{N_t^2}{N_s} \left(\frac{N_t - N_s}{N_t - 1} \right)}, \\ &= \sqrt{n(W) \left(\frac{1}{\alpha} - 1 \right) \frac{N_t}{N_t - 1} \frac{\nu(S')}{\nu(S)} p_d(S) (1 + \bar{c} p_d(S))}, \\ &= \text{SE}_{po}[\hat{X}|S] \sqrt{\frac{\nu(S')}{\nu(S)} p_d(S) (1 + \bar{c} p_d(S))}. \end{aligned} \quad (\text{A.8})$$

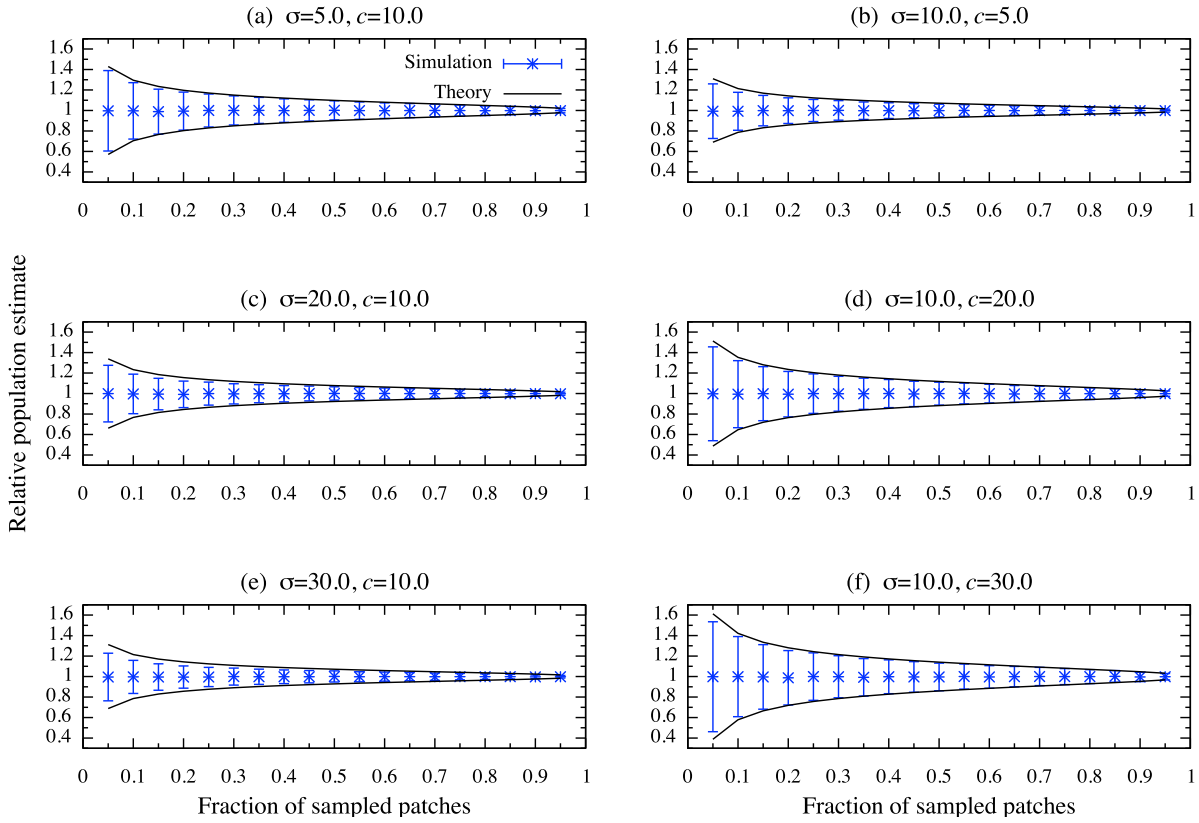


Figure A.2: Relative value of the population estimate with the average individuals $E[n(W)] = 10^3$ with different parameters. Sampling area is $32\text{m} \times 32\text{m}$. Each panel shows relative average estimate \pm relative standard error (Eq. (8)) of simulation and theoretical results. Relative average estimate for theoretical results is omitted since it is an unbiased estimator. Total area is $\nu(W) = 2^{20}\text{m}^2$ ($1024\text{m} \times 1024\text{m}$).