Spatially explicit approach to population abundance estimation in field surveys

Nao Takashina[1]*, Buntarou Kusumoto[2], Maria Beger[3,4] Suren Rathnayake[5], 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*

Abstract

The abundance of species is a fundamental consideration in ecology and conservation biology. Although broad models have been proposed to estimate the population abundance using existing data, available data is often limited. With no information available, a population estimation will rely on time consuming field surveys. Typically, time is a critical constraint in conservation and often management decisions must be made quickly under the data limited situation. Depending on time and budgetary constraints, the required accuracy of field survey changes significantly. Hence, it is desirable to set up an effective survey design to minimize time and effort of sampling given required accuracy. We examine a spatially-explicit approach to population estimation using spatial point processes, enabling us to explicitly and consistently discuss various sampling designs. We find that the accuracy of abundance estimation varies with both ecological factors and survey design. Although the spatial scale of sampling does not affect estimation accuracy when the underlying individual distribution is random, it decreases with the sampled unit size if individuals tend to form clusters. These results are derived analytically and checked numerically. Obtained insights provide a benchmark to predict the quality of population estimation, and improve survey designs for ecological studies and conservation.

Introduction

Estimating the abundance of populations is important for ecological studies and conservation biology [1–7], as is the role of ecosystem monitoring to observe changes in ecosystems [8–10]. In

---

*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

conservation, such knowledge helps one to estimate the risk of extinction of species [11, 12], and to implement effective conservation actions [13].

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

Ultimately, we face a trade-off between accuracy, time, and money. To tackle this trade-off, and provide advice to people designing a spatial sampling approach, we need a theory that can handle different sampling methods, choice of sampling unit size, and fraction of sampled region. Most previous approaches are spatially implicit e.g., [5,6,14,15,21], and it is therefore not straightforward to compare the effect of different survey designs. For example, as in the previous studies just mentioned above, the negative binomial distribution (NBD) is the most frequently used distribution for describing the underlying biological distribution of a species. In the NBD, the parameter characterizing the degree of spatial aggregation is scale dependent, and needs to be calibrated for each sampling unit scale. However, this procedure is not intuitive, as the parameter characterizing aggregation is usually inferred by observed data rather than biological mechanisms [14].

To address this challenge, here we use a spatially explicit approach, which enables us to explicitly and consistently compare the effect of different sampling schemes across sampling unit scales. Specifically, we examine random sampling and cluster sampling [22, 23] as sampling schemes, due to their simplicity of implementation. Moreover, cluster sampling reflects existing geographically biased sampling to some extent. These sampling schemes are combined with spatial point processes (SPPs), a spatially-explicit stochastic model to reveal effects of different survey regimes as well as ecological factors on the performance of population estimate. SPPs are often applied in ecological studies due to their flexibility and availability of biological interpretations [24–28]. Many of their examples come from studies of plant communities [24, 25, 27, 28], but include also study of coral communities [29], where spatial aggregation patterns are typically observed [30, 31]. Hence, field surveys of plant and coral species are the potential applications in our framework.

Using a spatially-explicit approach, we revealed new insights into population estimation with different survey designs as well as ecological regimes. The accuracy of the abundance estimation through random sampling varies with both ecological factors, such as the spatial distribution pattern of individuals, and survey design. Although the spatial scale of sampling does not affect estimation accuracy when the underlying spatial pattern is random, it decreases with the size of sampling unit if individuals tend to form clusters. However, under both random and clustered distributions, the relative standard error of abundance estimation decreased with the number of individuals, and the fraction of a region sampled. Obtained these insights will help to improve field survey designs for both ecological studies and conservation planning.

2

Methods

In this analysis, we consider a situation where there is no prior spatial data available to infer the distribution and abundance of a target species. We assume that our estimate of population size is based only on field surveys where a sampled fraction α ($0 < \alpha \le 1$) of the region of concern, W, is surveyed using a sampling unit size, S (Fig. 1). We focus on a case where no sampling error occurs in each sampling unit, suggesting that a sampling unit size should be chosen to satisfy this requirement in practice, and it may vary for sampling in different systems. For example, such an area may be larger for counting plant species compared to counting coral species due to different visibility and accessibility.

First, we introduce an estimator of population abundance, its expected value and variance, which explicitly accounts for the effect of sampling unit size. Next, we explain some basic properties of spatial point processes (SPPs),and models to describe spatial distribution patterns of individuals. Using computer-generated spatial distribution patterns, we test our analytical results formula for population estimation.
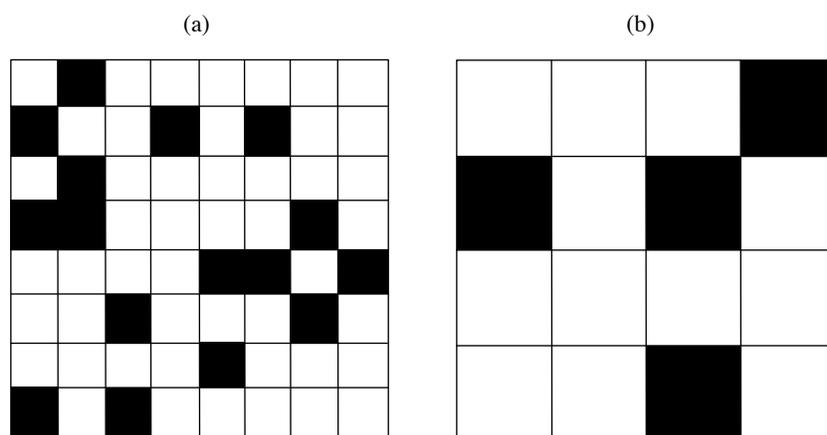
(a) (b)

Figure 1: Example of simple random sampling with (a) smaller, and (b) larger sampling unit size. The whole region of concern is divided into sampling unit with equal size, and a certain fraction α is randomly sampled (shaded unit) without replacement, where all sampling units have the equal probability to be chosen. No or sufficiently small measurement error is accompanied by each sampling trial. Essentially, applying larger sampling unit corresponds to a cluster sampling. The examples show the case of $\alpha = 0.25$.

Survey design

Given parameters specifying the survey design noted above, a simple random sampling (SRS) without replacement [23] is conducted for correcting count data (Fig 1). In the SRS without replacement, all the sampling units have an equal probability to be chosen and units already sampled once are not chosen again. The number of sampling units, N_t, and the sampled units, N_s, change with a sampling unit size, S. We assume all the sampling units have an equal size. As applying larger sampling unit sizes, the degree of the geographical sampling bias increases when the

3

fraction of a sampled region is small. Essentially, it corresponds to one-stage cluster sampling [23], where either all or none of the parcel of the area within the larger sampling unit are in the sample. It is worth noting, however, that the sampling unit size is arbitrarily chosen in ecological surveys; therefore the degree of cluster sampling is relative: any SRS can turn to be a cluster sampling if it is compared to SRS with a smaller sampling unit size. In this article, we simply use the simple random sampling and cluster sampling by implying using relatively *small* and *large* sampling units.

Population estimator

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

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

where, $\hat{n}|S$ is the estimated population abundance given sampling unit size S, y_i is the number of sampled individuals at ith sampling trial, and n_k is the frequency of the sampled units holding k individuals. Note y_i and n_k change depending on the sampling unit size and underlying spatial point patterns. In the SRS without replacement, the frequency n_k, given the number of sampled units N_s, is only the probability variable, following a multivariate hypergeometric distribution $p(n_k|S, N_s)$ with the mean $N_s p(k|S)$. Hence, the average population estimation $\hat{n}$ is

$$
\begin{aligned}
\mathrm{E}[\hat{n}|S] &= \frac{N_t}{N_s} \sum_{k}^{\infty} \mathrm{E}[n_k|S]k, \\
&= N_t \mathrm{E}[k|S].
\end{aligned}
\tag{2}
$$

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

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

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

Spatial distribution of individuals

To account for explicit spatial distributions of individuals, we use spatial point processes (SPPs) [24, 28]. The underlying models used in our analysis are the homogeneous Poisson process and

4

Thomas process, generating random and cluster distribution patterns of individuals, respectively. Properties of these processes are found in the literature (e.g., [24, 28, 32]) and, hence, we only introduce the properties relevant to our questions.

Homogeneous Poisson process

One of the simplest class of SPPs is the homogeneous Poisson process where the points (i.e. individuals) placement are randomly determined within the region of concern and the number of points given in the region R, $n(R)$, is according to the Poisson distribution with an average μ_R:

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

where, μ_R is known as the intensity measure [24, 28] defined by

$$\mu_R = \lambda \nu(R), \tag{5}$$

where, $\lambda := n(W)/\nu(W)$ is the point intensity of the whole region concerned W, and $\nu(R)$ is the area of region R.

Thomas process

The Thomas process, characterizing the clustering pattern of individuals, belongs to the family of Neyman-Scott processes [24, 28]. It provides a general framework to address spatial ecological patterns since almost species are clumped in nature [33]. In addition, the Thomas process is relatively amenable to an analytical approach [24, 25, 27, 28]. The Thomas process is obtained by the following three steps:

1. Parents are randomly placed according to the homogeneous Poisson process with a parent intensity λ_p.

2. Each parent produces a random discrete number c of daughters, realized independently and identically.

3. Daughters are scattered around their parents independently with an isotropic bivariate Gaussian distribution with variance σ^2, and all the parents are removed in the realized point pattern.

The intensity of the Thomas process is [28]

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

where, $\bar{c}$ is the average number of daughters per parent. For the sake of comparison between the population estimation of the two SPPs under the same average number of individuals, we chose the intensity of the Thomas process so as to have the same average individuals within the region of concern W. Namely, the parameters λ_p and $\bar{c}$ satisfy

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

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

Results

The total number of sampling units and sampled number is $N_t = \nu(W)/\nu(S)$ and $N_s = \lfloor \alpha N_t \rfloor$, where $\lfloor x \rfloor$ is the maximum integer not larger than x. We are here interested in how the estimations of population total deviate from the true value, and its relative degree compared to the true value. Therefore, one of the quantities to show these effect may be

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

Note in the analysis below, we use $\lfloor \alpha N_t \rfloor = \alpha N_t$ for simplicity, but this effect is rather small when αN_t is sufficiently large.

Population estimation under the homogeneous Poisson distribution

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

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

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

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

Under such circumstances, the standard error of the abundance estimation is only the function of the expected population total existing in the concerned region $n(W)$ and the sampling fraction α; and does not depend on the sampling unit size. Therefore, we can write $\mathrm{SE}_{po}[\hat{n}|S] = \mathrm{SE}_{po}[\hat{n}]$. Due to the term $n(W)^{1/2}$ in $\mathrm{SE}_{po}[\hat{n}|S]$, the relative variation from its average decreases with the factor $(1/\alpha - 1)^{1/2}n(W)^{-1/2}$. These results are confirmed by numerical simulations, and they show good agreement with analytical results (Fig. 2, 3).

Population estimation under the Thomas process

For the Thomas process, deriving an analytical form of the variance of individuals given a spatial scale of sampling $\mathrm{Var}[k|S]$ is challenging, although the probability generating functional of the Thomas process is known, e.g., [28]. Here we apply an approximated probability distribution of the Thomas process for the sake of obtaining an explicit form of $\mathrm{SE}_{th}[\hat{n}|S]$. Assuming that all parents within the region S', where parents can potentially supply daughters to a region S, given the same number of daughters, we derive an approximated $\mathrm{SE}_{th}[\hat{n}|S]$ as (Appendix).

$$\mathrm{SE}_{th}[\hat{n}|S] = \mathrm{SE}_{po}[\hat{n}|S] + \sqrt{n(W)\frac{N_t}{N_t-1}\left(\frac{1}{\alpha}-1\right)\bar{c}p_d(S)}, \tag{11}$$

where, $p_d(S)$ is the probability that an individual daughter produced by a parent situated a surrounding region of S, S', falls in the region S (Appendix). Therefore, the standard error of the

6
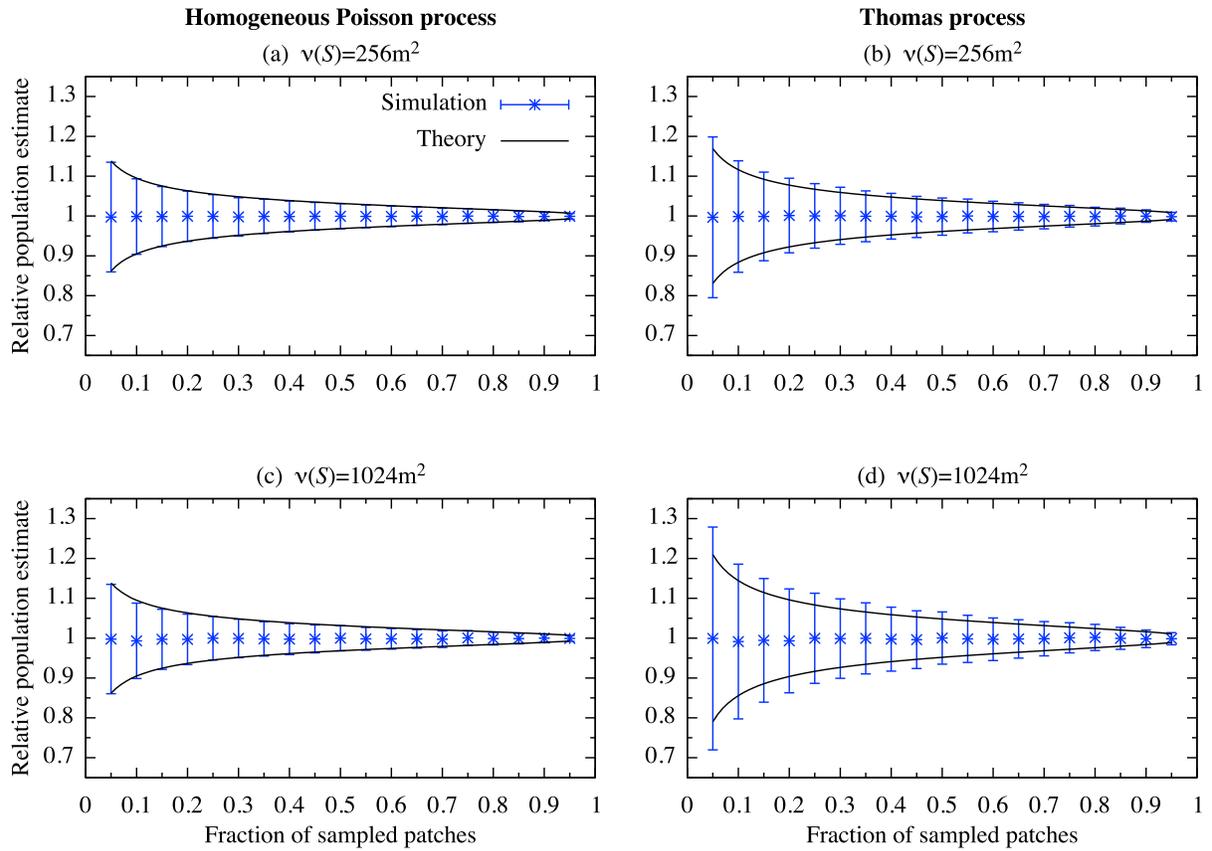
Homogeneous Poisson process

Thomas process
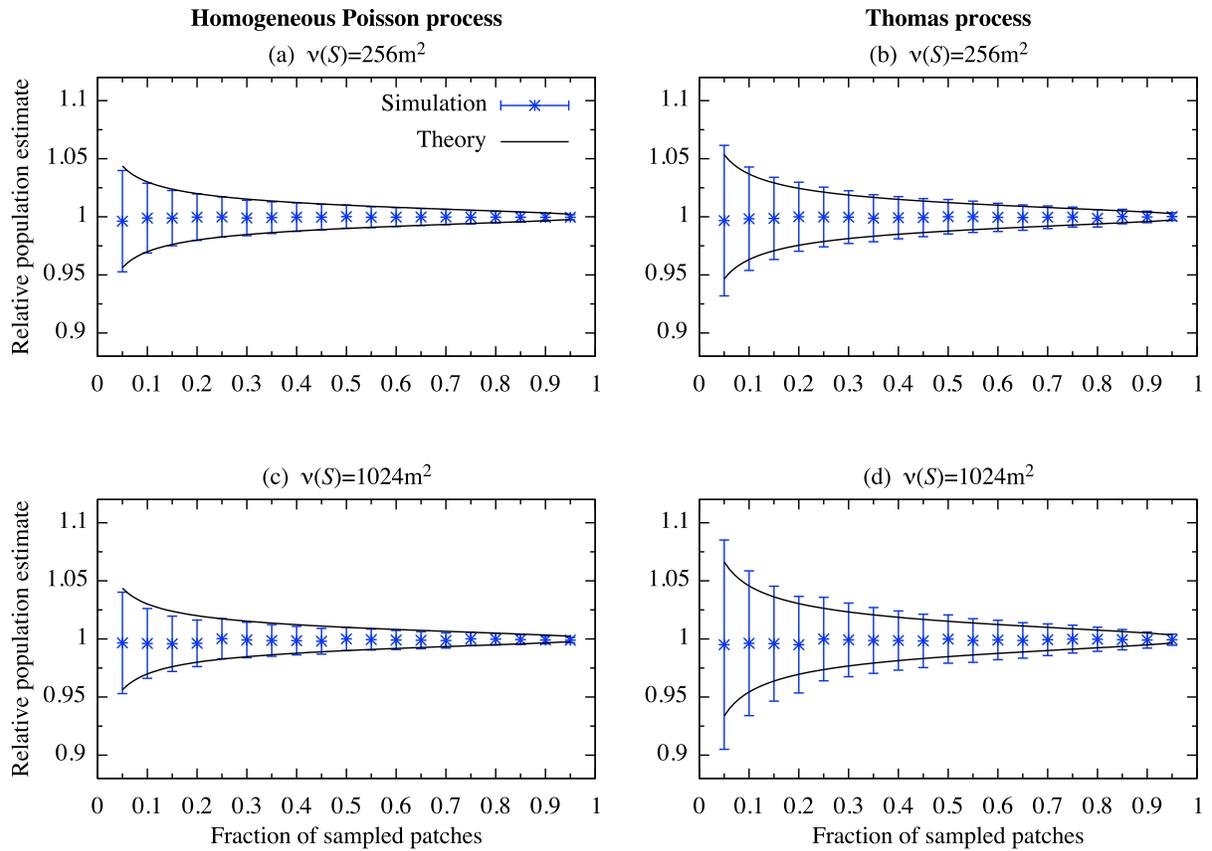

Figure 2: Relative value of the population estimate with the average individuals $E[n(W)] = 10^3$. The parameter values used are $\bar{c} = 10$, $\sigma = 10$, and $\nu(W) = 2^{20}\text{m}^2$ (1024m×1024m).

Homogeneous Poisson process

(a) $\nu(S)$=256m^2

Thomas process

(b) $\nu(S)$=256m^2

(c) $\nu(S)$=1024m^2

(d) $\nu(S)$=1024m^2

Figure 3: Relative value of the population estimate with the average individuals $\mathrm{E}[n(W)] = 10^4$. The parameter values used are the same as in Fig. 2.

Thomas process, $\mathrm{SE}_{th}[\hat{n}|S]$, can be described by the sum of the standard error of the Homogeneous Poisson process and a term characterizing the degree of cluster of the Thomas process. The second term of Eq. (11) increases with the degree of the clustering. Because $p_d(S)$ approaches to 0 as $\nu(S)$ becomes small and approaches to 1 as $\nu(S)$ becomes large (see Eq. (A.2) and Fig. A.2 in Appendix), Eq. (11) can be described by simpler form when $p_d(S) \sim 0$ and $p_d(S) \sim 1$. Especially, if $\bar{c}p_d(S) \sim 0$ and $N_t \gg 1$, Eq. (11) coincides with $\mathrm{SE}_{po}[\hat{n}]$, where $\bar{c}$. Similarly, as $p_d(S)$ increases with the sampling unit size, applying a geographically biased sampling increases the standard error. It is the clear difference from the result obtained under the random spatial pattern of individuals. Biologically speaking, species with a high fecundity, $\bar{c}$, and smaller dispersal distance of daughters, σ, increases the $\mathrm{SE}_{th}[\hat{n}|S]$, and vice versa.

The approximated $\mathrm{SE}_{th}[\hat{n}|S]$, Eq. (11), shows good agreement with the values obtained by the numerical simulations, except where the fraction of the sampled patch is small (Fig. 2, 3). However, its relative deviation from the numerical values are decreased as the average number of individuals $\mathrm{E}[n(W)]$ increases (Fig. 3).

Discussion

We examined a method for spatially-explicit population estimation combined with spatial point processes (SPPs) to reveal effects of different survey regimes as well as ecological factors on the performance of population estimate. By assuming the random and clustering placements of individuals as underlying ecological distribution patterns, that is, the homogeneous Poisson and Thomas process, we analytically show that the individual distributions and sampling schemes, such as random sampling and cluster sampling, change significantly the standard error of the abundance estimate. In our sampling framework, increasing the sampling unit size corresponds to an increase of geographical bias of the sampling. Typically, we find that the standard error of the abundance estimate is insensitive to the sampling unit size applied when the underlying ecological distribution is the homogeneous Poisson process. It is however individuals of most species are typically aggregated [30, 34] unless the abundance of a species is low [34]. Instead, when the underlying ecological pattern is described by the Thomas process, the standard error is increased with the size of sampling unit. That is, under clustered ecological distributions, the standard error is increased as the degree of clustering sampling increases. We also show that the standard error of the population estimate increases with the parameter controlling the degree of clustering of individual distributions $\bar{c}p_d(S)$ (Eq. 11). In addition, although for both ecological distribution patterns our results suggest that absolute value of the standard error increases with the number of individuals, the relative standard error decreases with the factor proportional to $n(W)^{-1/2}$. This means that for a species with a smaller number of individuals, the accuracy of the population estimation becomes lower. Although the homogeneous Poisson process may be a crude assumption for a distribution pattern of individuals, its sampling outcomes coincide with that of the Thomas process when the number of sampling unit is large and $\bar{c}p_d(S) \ll 1$. This occurs if the fecundity of a species of concern $\bar{c}$ is sufficiently small, or the probability that an individual produced by a parent within a region S', where a parent potentially supplies her daughter to a sampling unit S, is sufficiently small.

In practice, simple random sampling may not easily be conducted due to time consuming, costly, and different accessibility to site [16, 23, 35]. However, simple random sampling may be a reasonable option when available information on the underlying species distribution patterns is limited [23], and it may enable us to obtain more reliable data since extensive samplings in inaccessible region

may also lead to new discoveries [16]. Alternatively, cluster sampling, which causes a geographical sampling bias but less expensive, is of often the favored survey design [16,23]. This may be suitable for the data collection of a species needed for quick conservation action at a cost of accuracy of data.

The central assumption in the analysis is that measurement error (i.e. imperfect detection) does not occur or is sufficiently small. Although many empirical studies have adopted this assumption [36], imperfect detection is also frequently observed even in sessile organisms like plants (e.g. [37, 38]). If searching time is fixed, rates of imperfect detection can be expected to increase with survey area [39]. This indicates that the sampling unit size should be chosen while taking the scale-dependency of the imperfect detection into account. Further studies are needed for more quantitative arguments about this issue.

Here we investigate the population estimation across sampling unit sizes under random or clustered ecological distribution patterns. Our results provide general insight into ecological survey design such as how the sampling unit size used and ecological spatial distribution patterns affect the estimation accuracy. For both ecological and conservation application in mind, our sampling framework is kept as general as possible. Therefore, it allows one to further extend our framework to handle more complex situations where, for example, the concerned region holds multiple institutions with different sampling unit sizes or a budgetary constraint is explicitly taken into consideration.

Acknowledgements

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) 1513–1315. 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, J. Loh, M. A. McGeoch, L. McRae, A. Minasyan, M. H. Morcillo, T. E. E. Oldfield, D. Pauly, S. Quader, C. Revenga, J. R. Sauer, B. Skolnik, D. Spear, D. Stanwell-Smith, S. N. Stuart, A. Symes, M. Tierney, T. D. Tyrrell, J.-C. Vie, R. Watson, Global Biodiversity: Indicators of Recent Declines, Science 328 (5982) (2010) 1164–1168. doi:10.1126/science.1187512.

[11] D. I. MacKenzie, J. D. Nichols, J. E. Hines, M. G. Knutson, A. B. Franklin, Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly, Ecology 84 (8) (2003) 2200–2207. doi:10.1890/02-3090.

[12] M. A. McCarthy, S. J. Andelman, H. P. Possingham, Reliability of Relative Predictions in Population Viability Analysis, Conserv. Biol. 17 (4) (2003) 982–989. doi:10.1046/j.1523-1739.2003.01570.x.

[13] H. S. Grantham, K. A. Wilson, A. Moilanen, T. Rebelo, H. P. Possingham, Delaying conservation actions for improved knowledge: How long should we wait?, Ecol. Lett. 12 (4) (2009) 293–301. doi:10.1111/j.1461-0248.2009.01287.x.

[14] E. Kuno, Evaluation of statistical precision and design of efficient sampling for the population estimation based on frequency of occurrence, Res. Popul. Ecol. (Kyoto). 28 (1986) 305–319.

[15] J. A. Royle, J. D. Nichols, Estimating abundance from repeated presenceabsence data or point counts, Ecology 84 (3) (2003) 777–790. doi:10.1890/0012-9658(2003)084[0777:EAFRPA]2.0.CO;2.

[16] S. Reddy, L. M. Dávalos, Geographical sampling bias and its implications for conservation priorities in Africa, J. Biogeogr. 30 (11) (2003) 1719–1727. doi:10.1046/j.1365-2699.2003.00946.x.

[17] T. A. Gardner, J. Barlow, I. S. Araujo, T. C. Ávila-Pires, A. B. Bonaldo, J. E. Costa, M. C. Esposito, L. V. Ferreira, J. Hawes, M. I. M. Hernandez, M. S. Hoogmoed, R. N. Leite, N. F. Lo-Man-Hung, J. R. Malcolm, M. B. Martins, L. A. M. Mestre, R. Miranda-Santos, W. L. Overal, L. Parry, S. L. Peters, M. A. Ribeiro, M. N. F. Da Silva, C. Da Silva Motta, C. A. Peres, The cost-effectiveness of biodiversity surveys in tropical forests, Ecol. Lett. 11 (2) (2008) 139–150. doi:10.1111/j.1461-0248.2007.01133.x.

[18] A. E. Camaclang, M. Maron, T. G. Martin, H. P. Possingham, Current practices in the identification of critical habitat for threatened species, Conserv. Biol. 29 (2) (2015) 482–492. doi:10.1111/cobi.12428.

[19] M. Bode, K. A. Wilson, T. M. Brooks, W. R. Turner, R. A. Mittermeier, M. F. McBride, E. C. Underwood, H. P. Possingham, Cost-effective global conservation spending is robust to taxonomic group., Proc. Natl. Acad. Sci. U. S. A. 105 (17) (2008) 6498–6501. doi:10.1073/pnas.0710705105.

[20] A. Hastings, Timescales and the management of ecological systems, Proc. Natl. Acad. Sci.doi:10.1073/pnas.1604974113.

[21] J. L. Green, J. B. Plotkin, A statistical theory for sampling species abundances, Ecol. Lett. 10 (11) (2007) 1037–1045. doi:10.1111/j.1461-0248.2007.01101.x.

[22] J. Rice, Mathematical Statistics and Data Analysis, 3rd Edition, Vol. 72, Thomson Higher Education, Belmont, 2007. doi:10.2307/3619963.

[23] S. Lohr, Sampling: design and analysis, Nelson Education, 2009.

[24] N. A. C. Cressie, Statistics for Spatial Data, John Wiley & Sons, New York, 1993.

[25] J. B. Plotkin, M. D. Potts, N. Leslie, N. Manokaran, J. Lafrankie, P. S. Ashton, Species-area curves, spatial aggregation, and habitat specialization in tropical forests., J. Theor. Biol. 207 (1) (2000) 81–99. doi:10.1006/jtbi.2000.2158.

[26] N. Picard, C. Favier, A Point-Process Model for Variance-Occupancy-Abundance Relationships, Am. Nat. 178 (3) (2012) 383–396. doi:10.1086/661249.

[27] S. Azaele, S. J. Cornell, W. E. Kunin, Downscaling species occupancy from coarse spatial scales, Ecol. Appl. 22 (3) (2012) 1004–1014. doi:10.1890/11-0536.1.

[28] S. N. Chiu, D. Stoyan, W. S. Kendall, J. Mecke, Stochastic Geometry and Its Applications, John Wiley & Sons, New York, 2013.

[29] S. Muko, K. Shimatani, Y. Nozawa, Spatial analyses for nonoverlapping objects with size variations and their application to coral communities (2014).

[30] R. Condit, P. S. Ashton, P. Baker, S. Bunyavejchewin, S. Gunatilleke, N. Gunatilleke, S. P. Hubbell, R. B. Foster, A. Itoh, J. V. LaFrankie, H. S. Lee, E. Losos, N. Manokaran, R. Sukumar, T. Yamakura, Spatial patterns in the distribution of tropical tree species, Science 288 (1982) (2000) 1414–1418. doi:10.1126/science.288.5470.1414.

[31] M. Beger, G. P. Jones, P. L. Munday, Conservation of coral reef biodiversity: A comparison of reserve selection procedures for corals and fishes, Biol. Conserv. 111 (1) (2003) 53–62. doi:10.1016/S0006-3207(02)00249-5.

[32] J. Illian, A. Penttinen, H. Stoyan, D. Stoyan, Statistical Analysis and Modelling of Spatial Point Patterns, Vol. 76, John Wiley & Sons, Chichester, 2008. doi:10.1002/9780470725160.

[33] R. Condit, P. S. Ashton, P. Baker, S. Bunyavejchewin, C. V. S. Gunatilleke, I. A. U. N. Gunatilleke, S. P. Hubbell, R. B. Foster, A. Itoh, J. V. LaFrankie, H. S. Lee, E. Losos, N. Manokaran, R. Sukumar, T. Yamakura, Spatial patterns in the distribution of tropical tree species., Science 288 (1982) (2000) 1414–1418. `doi:10.1126/science.288.5470.1414`.

[34] L. R. Taylor, I. P. Woiwod, J. N. Perry, The Density-Dependence of Spatial Behaviour and the Rarity of Randomness, J. Anim. Ecol. 47 (2) (1978) pp. 383–406. `doi:10.2307/3790`.

[35] H. Possingham, I. Ball, S. Andelman, Mathematical methods for identifying representative reserve networks, in: Quant. Methods Conserv. Biol., Springer New York, 2000, pp. 291–306. `doi:10.1007/0-387-22648-6_17`.

[36] K. F. Kellner, R. K. Swihart, Accounting for imperfect detection in ecology: A quantitative review (2014). `doi:10.1371/journal.pone.0111436`.

[37] G. Chen, M. Kéry, J. Zhang, K. Ma, Factors affecting detection probability in plant distribution studies, J. Ecol. 97 (6) (2009) 1383–1389. `doi:10.1111/j.1365-2745.2009.01560.x`.

[38] G. Chen, M. Kéry, M. Plattner, K. Ma, B. Gardner, Imperfect detection is the rule rather than the exception in plant distribution studies, J. Ecol. 101 (1) (2013) 183–191. `doi:10.1111/1365-2745.12021`.

[39] K. D. Clarke, M. Lewis, R. Brandle, B. Ostendorf, Non-detection errors in a survey of persistent, highly-detectable vegetation species, Environ. Monit. Assess. 184 (2) (2012) 625–635. `doi:10.1007/s10661-011-1991-0`.

Appendix: Approximated first and second moment of the Thomas process

Here, we derive an approximated probability distribution function (pdf) given area R of the Thomas process. Let the whole region potentially contribute to the region R be $R' = R + R_{out}$, where R_{out} is the region where parents outside the region, R, can potentially supply daughters to the region R (Fig. A.1). Here. We assume that all parents give the same number of daughters within the small region, corresponding to an assumption that the variance in the numbers of daughter between parents is small. By the procedures to obtain the Thomas process, the probability n individuals found given region R may be described as follows:

$$
\begin{aligned}
p(n|R) &= \underbrace{\sum_k \frac{(\lambda_p \nu(R'))^k}{k!} e^{-\lambda_p \nu(R')}}_{\text{\# parents in } R'} \quad \underbrace{\sum_{k'} \frac{(\bar{c}k)^{k'}}{k'!} e^{-\bar{c}k}}_{\text{\# daughters by parents in } R'} \quad \underbrace{\binom{k'}{n} p_d(R)^n (1 - p_d(R))^{k'-n}}_{\text{prob}(n \text{ points fall in } R)}, \\
&= \sum_{k,k'} \frac{(\lambda_p \nu(R'))^k}{k!} e^{-\lambda_p \nu(R')} \frac{(\bar{c}k)^{k'}}{k'!} e^{-\bar{c}k} \frac{k'!}{(k'-n)! n!} p_d(R)^n (1 - p_d(R))^{k'-n}, \\
&= \frac{e^{-\lambda_p \nu(R')}}{n!} (\bar{c} p_d(R))^n \sum_{k,k'} \frac{(\lambda_p \nu(R'))^k}{k!} e^{-\bar{c}k} k^n \frac{(\bar{c}k)^{k'-n}}{(k'-n)!} (1 - p_d(R))^{k'-n}, \quad \text{(A.1)} \\
&= \frac{e^{-\lambda_p \nu(R')}}{n!} (\bar{c} p_d(R))^n \sum_k \frac{(\lambda_p \nu(R') e^{-c p_d(R)})^k}{k!} k^n, \\
&= \sum_k \frac{(\lambda_p \nu(R'))^k}{k!} e^{-\lambda_p \nu(R')} \frac{(k\bar{c}p)^n}{n!} e^{-k\bar{c}p_d(R)}, \\
&= \sum_k \text{Po}(k, \lambda_p \nu(R')) \text{Po}(n, k\bar{c}p_d(R)),
\end{aligned}
$$

where, $p_d(R)$ is the probability that an individual daughter produced by a parent within R' falls in R:

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

By referring Fig. A.1, $\nu(R')$ is calculated as

$$
\nu(R') = (2r + R_x)(2r + R_y) - r^2(4 - \pi), \quad \text{(A.3)}
$$

where, r is the distance that on average a fraction u of daughters scattered by the parent (placed center) are covered. r is calculated by converting the expression of the isotropic bivariate gaussian 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 coordinates, and solving about r

$$
r = \sqrt{-2\sigma^2 \log(1 - u)}, \quad \text{(A.4)}
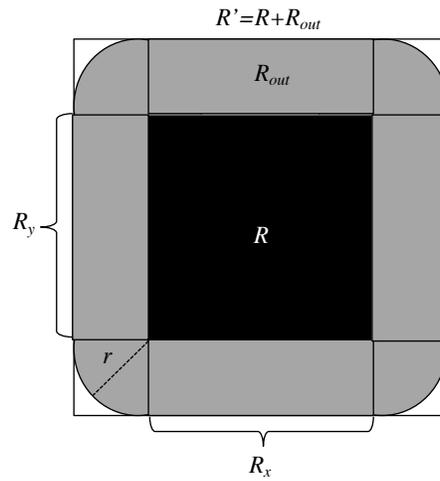$$

14

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 $R' = R + R_{out}$.
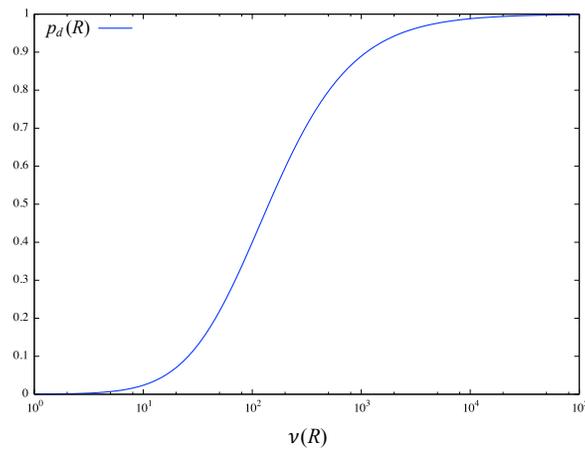

Figure A.2: Probability that an individual daughter produced by a parent within R' falls in R. The parameter values are the same as used in analysis. Namely, $u = 0.99$, $\sigma = 10$, and these give a value $r = 30.35$.

where, in the analysis, we set $u = 0.99$. An example functional form of Eq. (A.2) is shown in Fig. A.2.

Using Eq. (A.1), we calculate the first moment and the second moment of point number k in region R as

$$\mathrm{E}[n(R)] = \lambda_p \bar{c} p_d(R) \nu(R') = \lambda_p \bar{c} \nu(R), \tag{A.5}$$

$$\mathrm{E}[n(R)^2] = \lambda_p \bar{c} p_d(R) \nu(R')(1 + \bar{c} p_d(R) + \lambda_p \bar{c} p_d(R) \nu(R')). \tag{A.6}$$

Using Eqs (3), (A.5), and (A.6) and the assumption $\lambda_p \bar{c} = \lambda = n(W)/\nu(W)$, we calculate Eq. (11).