bioRxiv preprint doi: https://doi.org/10.1101/137158; this version posted May 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

A theory for ecological survey methods to map individual distributions

Nao Takashina[1*], Maria Beger[2,3], Buntarou Kusumoto[4], Suren Rathnayake[5], Hugh P. Possingham[2,6]

[1] *Tropical Biosphere Research Center, University of the Ryukyus, 3422 Sesoko Motobu, Okinawa 905-0227, Japan*
[2] *ARC Centre of Excellence for Environmental Decisions, School of Biological Sciences, The University of Queensland, St Lucia, QLD 4072, Australia*
[3] *School of Biology, Faculty of Biological Sciences, University of Leeds, Leeds LS2 9JT, UK*
[4] *Center for Strategic Research Project, University of the Ryukyus, 1 Senbaru, Nishihara, Okinawa 903-0213, Japan*
[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: Ecological survey, presence/absence map, spatial distribution, spatial point processes

Abstract

Spatially-explicit approaches have been widely recommended for various applications of ecosystem management. In practice, the quality of the data involved in the management decision-making, such as presence/absence or habitat maps, affects the management actions recommended, and therefore it is a key to management success. However, available data is often biased and incomplete. Although previous studies have advanced ways to effectively resolve data bias and missing data, there still remains a question about how we design the entire ecological survey to develop a dataset through field surveys. Ecological survey may inherently have multiple spatial scales to be determined beforehand, such as the spatial extent of the ecosystem under concern (observation window), the resolution to map the individual distributions (mapping unit), and the area of survey within each mapping units (sampling unit). In this paper, we develop a theory to understand ecological survey for mapping individual distributions applying spatially-explicit stochastic models. Firstly, we use spatial point processes to describe individual spatial placements drawn using either random or clustering processes. An ecological survey is then introduced with a set of spatial scales and individual detectability. Regardless of the spatial pattern assumed, the choice of mapping unit largely affects presence detection rate, and the fraction of individuals covered by the

*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

bioRxiv preprint doi: https://doi.org/10.1101/137158; this version posted May 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

presence-detected patches. Tradeoffs between these quantities and the resolution of the map are found, associated with an equivalent asymptotic behaviors for both metrics at sufficiently small and large mapping unit scales. Our approach enables us to directly discuss the effect of multiple spatial scales in the survey, and estimating the survey outcome such as the presence detection rate and the number of individuals the presence-detected patches hold. The developed theory may significantly facilitate management decision-making and inform the design of monitoring and data gathering.

1 Introduction

Understanding the spatial characteristics of ecosystems is one of the central challenges in ecology [1]. Such knowledge forms a prerequisite for effective ecosystem management due to an increasing need for spatially explicit approaches in fisheries and wildlife management [2–4] and for the establishment of terrestrial and marine protected areas [5–7].

In ecosystem management, the quality of the data involved in the management decision-making, such as presence/absence or habitat maps, affect the management actions recommended [8–10]. Therefore, creating an ecologically and statistically adequate dataset is key to management success. However, available data is often biased and incomplete [8, 9], due to, for example, different accessibility to sites [8], existence of the favored study sites [8], and imperfect detectability of individuals [11, 12]. These biases hinder the effective implementation of management actions, and may lead to perverse outcomes or wasted management resources. Hence it is important to discuss and benchmark the quality of the spatially explicit data that underlies management decisions.

There is a body of literatures to tackle the challenges of data gathering, including sampling designs for effectively allocating the survey effort under the time and budgetary constraints [13–15], methods for reducing the bias of occurrence data by estimating the detectability of species [12, 16–18], and mathematical theory for ecological sampling [19, 20]. Although these researches have significantly advanced our insight into ecosystem monitoring and ecological survey, there still remains a question about how we actually design the entire ecological survey to systematically develop dataset through a field survey, as the spatial scale issue, such as how to chose the resolution of a map, is often omitted. This is perhaps because many existing studies consider the space to be sampled implicitly. Presence/absence or habitat map is widely used in ecosystem management [16], where at least three different spatial scales may exist; the spatial extension of the ecosystem under concern, resolution to map the individual distributions, and minimum size of survey units. To systematically gather the spatial data, manager should explicitly take into consideration these three spatial scales, because the manner of the sampling and management outcomes depend on the resolution of a map. For example, in fisheries management, finely implemented fishing quota allocations may result in better management outcomes [7, 21], and this can be done with the distribution map with a high degree of resolution. However, surveying an area at a fine spatial scale is often impractically expensive in the large scale survey, and the choice of resolution itself faces a budgetary constraint. Hence, quantitatively estimating the performance of a

bioRxiv preprint doi: https://doi.org/10.1101/137158; this version posted May 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

sampling method in advance facilitates survey decision-making.

In this paper, we develop a theory of ecological survey method for systematically mapping individual distributions by making use of the spatial point processes (SPPs), a spatially explicit stochastic model. The SPPs is widely applied to the study of plant community [22–25], and also applied to coral community [26]. Therefore, they are potential target species of the developed theory. In this study, the SPPs describes individual spatial locations by two different processes accounting random or clustering patterns. An ecological survey is then introduced with a set of spatial scales and detectability of individuals. Our spatially-explicit approach is capable of revealing a series of questions important for ecological survey, such as effect of the choice of the spatial scales and spatial distribution patterns of individuals on accuracy of the distribution map. This knowledge enables one to determine the design of an ecological survey beforehand given accuracy of a map required. The developed theory may significantly facilitate management decision-making and give solid bases of data gathering.

2 Methods

2.1 Models of spatial distribution of individuals

To develop a theory of ecological survey to map individual distributions, we explicitly model the spatial distribution patterns of individuals. Spatial point processes (SPPs) [22, 25] provide models to describe such patterns with high flexibility and analytical tractability [24]. Here, we apply the homogeneous Poisson process and the Thomas process, a family of the Neyman-Scott process (Fig. 1).

One of the simplest SPPs is the homogeneous Poisson process where the points (i.e. individuals) are randomly distributed and the number of points of a given region A, $N(A)$, is according to the Poisson distribution with an average μ_A:

$$\mathrm{Prob}(N(A) = k) = \frac{\mu_A^k}{k!} e^{-\mu_A}, \qquad (k = 0, 1, \dots) \tag{1}$$

where, μ_A is also regarded as the intensity measure [22, 25] described as

$$\mu_A = \lambda \nu(A), \tag{2}$$

where, $\lambda = (\textit{total points})/(\textit{area of concerned region } A)$ is the intensity in the given region, and $\nu(A)$ is the area of A.

The Neyman-Scott process [22, 25] provides us more general framework to analyze spatial ecological data and characterize the clustering pattern of individuals [22–25]. By the following three steps, the Neyman-Scott process is obtained:

- Parents are randomly placed according to the homogeneous Poisson process with an intensity λ_p.
- Each parent produces a random discrete number c of daughters, realized independently and identically for each parent.

bioRxiv preprint doi: https://doi.org/10.1101/137158; this version posted May 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

- Daughters are scattered around their parents independently with an identical spatial probability density, $f(\mathbf{y})$, and all the parents are removed in the realized point pattern.

The intensity of the Neyman-Scott process is [25]

$$\lambda = \bar{c}\lambda_p, \tag{3}$$

where, $\bar{c}$ is the average number of daughters per parent. The probability generating functional (pgfl) of the number of daughters within a given region of the Neyman-Scott process is [22,25]

$$G(v) = \exp\left(-\lambda_p \int_{\mathbf{R}^d} \left[1 - G_n\left(\int_{\mathbf{R}^d} v(\mathbf{x}+\mathbf{y})f(\mathbf{y})d\mathbf{y}\right)\right] d\mathbf{x}\right), \tag{4}$$

where, $G_n\left(\int_{\mathbf{R}^d} v(\mathbf{x}+\mathbf{y})f(\mathbf{y})d\mathbf{y}\right)$ is the probability generating function (pgf) of the random number c, the number of daughters per parent.

The Thomas process is a special case of the Neyman-Scott process, where $f(\mathbf{y})$ is an isotropic bivariate Gaussian distribution with the variance σ^2 [25]. We also assume that the number of daughters per parent follows the Poisson distribution with the average number, $\bar{c}$. The pgfl of the Thomas process, Eq. (4), within a given region A is obtained by substituting the pgf of the number of daughters per parent G_n in Eq. (4). It is obtained, by the given assumptions, as

$$\begin{aligned} G_n\left(\int_{\mathbf{R}^d} v(\mathbf{x}+\mathbf{y})f(\mathbf{y})d\mathbf{y}\right) &= \sum_{k=0}^{\infty}\left(\int_{\mathbf{R}^d} v(\mathbf{y})f(\mathbf{y}-\mathbf{x})d\mathbf{y}\right)^k \frac{\bar{c}^k}{k!}e^{-\bar{c}}, \\ &= \exp\left[-\bar{c}\left(1 - \int_{\mathbf{R}^d} v(\mathbf{y})f(\mathbf{y}-\mathbf{x})d\mathbf{y}\right)\right], \\ &= \exp\left[-\bar{c}(1-t)\left(\int_A f(\mathbf{y}-\mathbf{x})d\mathbf{y}\right)\right], \end{aligned} \tag{5}$$

where, to obtain the last line, $v(\mathbf{y}) = 1-(1-t)\mathbf{1}_A(\mathbf{y})$ is used, and here $\mathbf{1}_A(\mathbf{y})$ is the indicator function. Therefore, the pgfl of the number of daughters within the region A of the Thomas process is

$$G(t) = \exp\left(-\lambda_p \int_{\mathbf{R}^2}\left[1 - \exp\left\{-\bar{c}(1-t)\left(\int_A k(\|\mathbf{x}-\mathbf{y}\|)d\mathbf{y}\right)\right\}\right] d\mathbf{x}\right), \tag{6}$$

where, $k(\|\mathbf{x}-\mathbf{y}\|)$ is an isotropic bivariate Gaussian distribution with variance σ^2,

$$k(\|\mathbf{x}-\mathbf{y}\|) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma^2}\right). \tag{7}$$

In order to reasonably compare the results of Thomas process with those of the homogeneous Poisson process, we chose the intensity of the Thomas process so as to have, on average, the same number of points within the concerned region. Namely, the parameters λ_p and $\bar{c}$ satisfy

$$\bar{c}\lambda_p = \lambda, \tag{8}$$

where, the left hand side (lhs) is the intensity of the Thomas process and the right hand side (rhs) is the intensity of the homogeneous Poisson.

bioRxiv preprint doi: https://doi.org/10.1101/137158; this version posted May 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

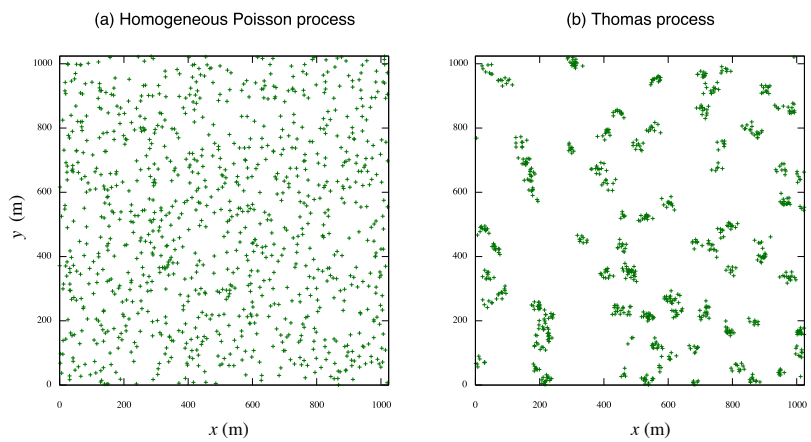

Figure 1: Example of point patterns within a observation window 1024m × 1024m. (a) Homogeneous point process with the intensity $\lambda = 10^{-3}$; (b) Thomas process with the same intensity value as the homogeneous Poisson process $\lambda_p \bar{c} = \lambda$, where $\lambda_p = 10^{-4}$ and $\bar{c} = 10$. The variance of the bivariate normal distribution $\sigma^2 = 100$. See the text for the interpretations of the parameters.

2.2 Design of ecological survey

2.2.1 Survey rules and basic properties

Let us consider the situation where an ecological survey takes place for the purpose of creating a presence/absence map of a given region. A presence/absence map is characterized by its various spatial resolutions: the spatial scale of the *mapping unit* defines the resolution of the map, and the spatial scale of the *sampling unit* determines the sampling density within each mapping unit (Fig. 2). We assume the following three key sampling rules.

- There are three spatial scales to consider in the conceptual design of a survey program. The spatial scale of ecological survey conducted W, the resolution of map M (i.e., scale of the mapping unit), and unit scale of survey in each mapping unit, S, are arbitrary determined by the managers, but single resolutions are allowed for each of the spatial scales.
- Every mapping unit is assessed.
- The whole region of a mapping unit is recorded as presence if at least one individual is detected, otherwise recorded as absence (Fig. 3).

Through the second and third assumptions, changing the scale of the mapping unit affects the presence detection rate (Fig. 3). Although using a larger mapping unit could improve the probability a mapping unit with individual recorded as presence, accuracy of presence/absence map decreases due to its large resolution.

bioRxiv preprint doi: https://doi.org/10.1101/137158; this version posted May 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

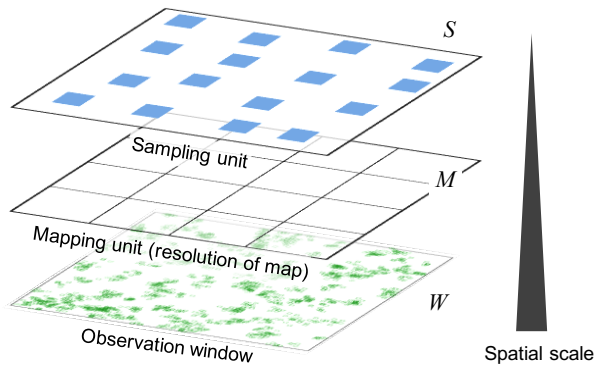

Figure 2: Multiple spatial scales in ecological survey. Each scale is arbitrary determined by managers.

2.2.2 Modeling the ecological survey

Here, we model the ecological survey with the three main assumptions listed above. Let, on average, N individuals of a species be distributed over a given window W, which is the region under concern. The manner of individual distribution is according to either the homogeneous Poisson process or the Thomas process. The resolution of the presence/absence map is defined by the scale of mapping unit M, and every mapping unit is sampled with the sampling density, S (Fig. 2). The survey is associated by the sampling error for each individual at a rate $\gamma := 1-\beta$, which is the rate at which individuals are not detected despite being present, and where, β is the detectability of an individual. The areas of each mapping unit is $1,\ 2,\ 4, \ldots,$ or 2^n times smaller than the area of a given window W, to assure the first assumption above. Let $\nu(X)$ be the area of a region X. With the definitions detailed above, we obtain

$$\nu(M) = \nu(W)/2^m, \qquad (m = 0, 1, \ldots, n) \tag{9}$$

where, the superscript m represents the number of subdivisions of the window W. From Eq. (9), the number of mapping units within a given window W, is

$$[\text{Number of mapping units}] = \nu(W)/\nu(M) = 2^m. \tag{10}$$

As the record for each mapping unit is based on an survey within the mapping unit, we obtain

$$\nu(S) = \alpha\nu(M) \leq \nu(M), \quad (0 < \alpha \leq 1) \tag{11}$$

where, α is the sampling density within a mapping unit. Combining Eq. (9) and Eq. (11), we obtain

$$\nu(S) \leq \nu(M) \leq \nu(W). \tag{12}$$

Let the intensity of the points within a given window W be [22, 25]

$$\lambda = \frac{N}{\nu(W)}. \tag{13}$$

bioRxiv preprint doi: https://doi.org/10.1101/137158; this version posted May 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

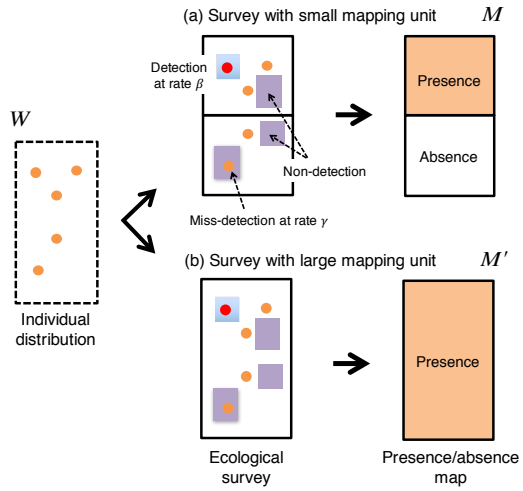

Figure 3: Ecological survey scheme within the observation window W. (a) Survey with a small mapping unit M; (b) Survey with a larger mapping unit M'. Although the same individual distribution (left panel) and distributions of sampling units (center panel) are used, obtained presence/absence maps resulted in the survey (right panels) can be different since the probability to detect a presence patch varies with the size of mapping unit.

As we noted above, the parameters for the Thomas process are chosen so as to satisfy Eq. (8).

3 Assessing the accuracy of presence/absence map

Given the spatial point pattern, sampling density, α, detectability of an individual, β, and scale of mapping unit, M, we calculate two main quantities of the ecological survey. That is, the presence detection rate (PDR), and the fraction of individuals covered within presence-detected patches (FIC). The presence detection rate reflects the accuracy of the obtained map, and the FIC holds information that connects the presence detection rate with population abundance. Hence, later may also be important for conservation. The presence detection rate is obtained by

$$\begin{aligned} \mathrm{E}_i[\mathrm{PDR}|\beta, S, M] &= \frac{\sum_{k=1}^{\infty} p(N=k|\beta, S)}{\sum_{k=1}^{\infty} p(N=k|M)}, \\ &= \frac{1-p(0|\beta, S)}{1-p(0|M)}, \end{aligned} \tag{14}$$

where, i indicates the underlying point pattern. Since the FIC is not easy to derive analytically for the Thomas process, we only provide an analytical expression of the FIC for the homogeneous Poisson process, and give numerical results for the Thomas process.

bioRxiv preprint doi: https://doi.org/10.1101/137158; this version posted May 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

3.1 Ecological survey with individual distributions based on the homogeneous Poisson process

Where individuals are distributed in space based on the homogeneous Poisson process, presence detection rate from Eq. (14) is

$$\mathrm{E}_{po}[\mathrm{PDR}|\beta, S, M] = \frac{1 - e^{-\beta\lambda\nu(S)}}{1 - e^{-\lambda\nu(M)}} = \frac{1 - e^{-\alpha\beta\lambda\nu(M)}}{1 - e^{-\lambda\nu(M)}}, \tag{15}$$

where, equality $\nu(S) = \alpha\nu(M)$ is used. Eq. (15) has rather simple form and, thus, we can easily see the parameter dependence. The intensity of the points λ (Eq. 13) defines the average number of individuals existing within a given the observation window, W, and since $d\mathrm{E}_{po}/d\lambda \geq 0$, $\mathrm{E}_{po}[\mathrm{PDR}]$ increases as the average number of individuals increase, and vice versa. Especially, when the intensity becomes $\lambda \to \infty$, $\mathrm{E}_{po}[\mathrm{PDR}]$ becomes 1 regardless of the scale of mapping units. Intuitively, as the sampling density α and detectability β increase, $\mathrm{E}_{po}[\mathrm{PDR}]$ increases, and vice versa. The asymptotic behaviors of Eq. (15) are easily obtained as:

$$\lim_{M\to 0} \mathrm{E}_{po}[\mathrm{PDR}|\beta, S, M] \simeq \alpha\beta, \tag{16}$$

$$\lim_{M\to W} \mathrm{E}_{po}[\mathrm{PDR}|\beta, S, M] \simeq 1, \tag{17}$$

where, W needs to be large enough to satisfy $e^{-\alpha\beta\lambda\nu(W)} \simeq 0$. These results show good agreement with the numerical results (Fig. 4a).

For the homogeneous Poisson process, we can derive an analytical form of the average number of individuals covered within presence-detected patches (IC) as follows:

$$\begin{aligned} \mathrm{E}_{po}[\mathrm{IC}|\beta, S, M] &= 2^m \sum_{k=1}^{\infty} \{1 - (1 - \alpha\beta)^k\} k \frac{(\lambda\nu(M))^k}{k!} e^{-\lambda\nu(M)}, \\ &= 2^m \lambda\nu(M) \left\{1 - (1 - \alpha\beta) e^{-\alpha\beta\lambda\nu(M)}\right\}, \end{aligned} \tag{18}$$

where, on the first line of rhs, 2^m is the number of mapping units within the given window W, inside of the curly brackets is the probability that none of k points are detected by a survey given a mapping unit M, and the remaining term is the expected number of points within the mapping unit. The FIC is obtained by dividing Eq. (18) by the intensity measure Eq. (2) of the observation window, μ_W,

$$\mathrm{E}_{po}[\mathrm{FIC}|\beta, S, M] = 1 - (1 - \alpha\beta) e^{-\alpha\beta\lambda\nu(M)}. \tag{19}$$

The dependences of the parameters λ, α, and β are qualitatively the same as those of Eq. (15). In addition, the asymptotic behaviors of Eq. (19) are equivalent to Eqs. (16) and (17). Fig. (4b) confirms the analytical evaluations of $\mathrm{E}_{po}[\mathrm{FIC}]$.

bioRxiv preprint doi: https://doi.org/10.1101/137158; this version posted May 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

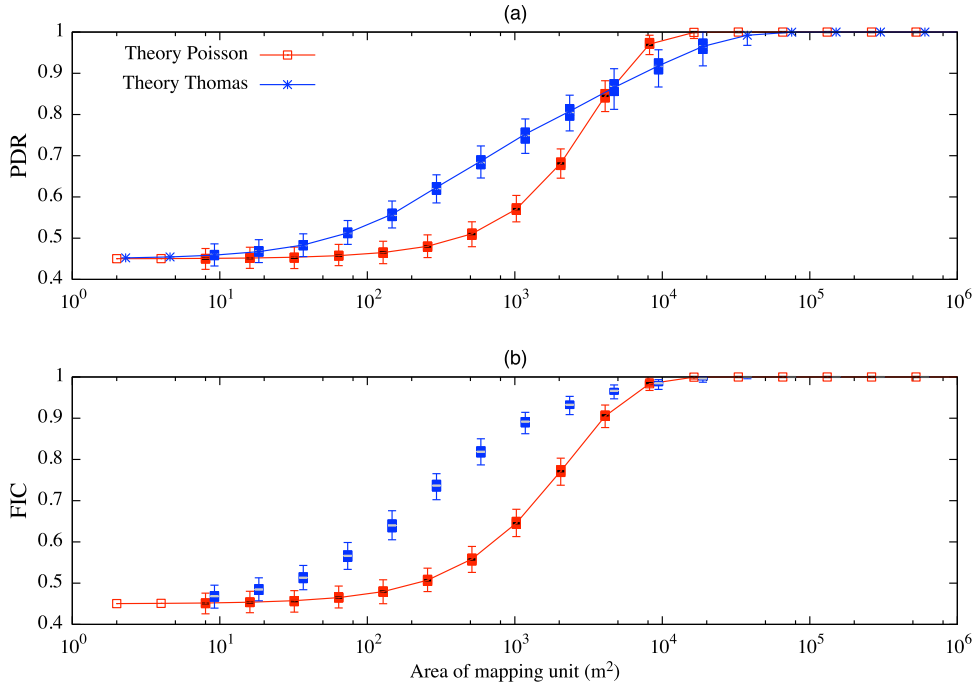

Figure 4: (color online) Analytical and simulated (candlestick) values of (a) the presence detection rate (PDR); and (b) the fraction of individuals covered within presence-detected patches (FIC) across mapping unit scales. x-axis is the area of mapping unit (m^2). Each candlestick shows, from the bottom, 5, 25, 50, 75, and 95 percentile values of 1000 simulation trials. The values of the sampling density and detectability are $\alpha = 0.5$ and $\beta = 0.9$, respectively. The other parameter values are the same as in Fig. (1).

3.2 Ecological survey with individual distributions based on the Thomas process

Here we consider the situation where individuals are distributed according to the Thomas process. By Eq. (14), we calculate the presence detection rate for the Thomas process:

$$\mathrm{E}_{th}[\mathrm{PDR}|\beta, S, M] = \frac{1 - p_{th}(0|\beta, S)}{1 - p_{th}(0|M)}, \tag{20}$$

where, the probability of each event of the Thomas process is obtained by the pgfl Eq. (4): $p_{th}(n|A) = 1/n!(d^n G(t)/dt^n)|_{t=0}$. Therefore, $p_{th}(0|A)$ is

$$p_{th}(0|A) = \exp\left(-\lambda_p \int_{\mathbf{R}^2} \left[1 - \exp\left\{-\bar{c}\left(\int_A \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma^2}\right) d\mathbf{x}\right)\right\}\right] d\mathbf{y}\right). \tag{21}$$

Specifically, the second term inside the square brackets for $p_{th}(0|\beta, S)$ becomes $\exp(-\alpha\beta\bar{c}\int_M \frac{1}{2\pi\sigma^2}\exp(-\|\mathbf{x}-\mathbf{y}\|^2/2\sigma^2)d\mathbf{x})$, due to the sampling density and the detectability. Although Eq. (20) and (21) are not easy to interpret, we can calculate its asymptotic

bioRxiv preprint doi: https://doi.org/10.1101/137158; this version posted May 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

behaviors as

$$\lim_{M\to 0} \mathrm{E}_{th}[\mathrm{PDR}|\beta, S, M] \simeq \alpha\beta, \tag{22}$$

$$\lim_{M\to W} \mathrm{E}_{th}[\mathrm{PDR}|\beta, S, M] \simeq 1, \tag{23}$$

They are equivalent to the asymptotic behaviors of the homogeneous Poisson process Eqs. (16) and (17). Fig. (4a) plots both analytical and numerical results. Analytical values show the good agreement with numerical results.

To obtain an explicit form for FIC of the Thomas process is cumbersome as the pgfl of the Thomas process Eq. (6) is rather complex. Therefore, we only show the numerical value for the FIC of the Thomas process (Fig. 4b). The FIC for the Thomas process increases faster than Eq. (15) as the mapping scale increases. The asymptotic behavior shows similar trends to the other results.

4 Discussion

By explicitly accounting for the spatial distribution patterns of individuals through spatial point processes (SPPs) and multiple spatial scales of field survey, we develop a theory for ecological survey to map individual distributions. The theory quantifies two metrics, the presence detection rate (PDR) and the fraction of individuals covered by the presence-detected patches (FIC), and thus allows us to predict the outcome of an ecological survey under certain survey designs. When both the sampling density α and the detectability within mapping unit β are not equal to 1, we find a tradeoff between the value of the PDR/FIC and the resolution of the map. The PDR and FIC show the equivalent asymptotic behaviors for both the homogeneous Poisson process and the Thomas process where $\alpha\beta$ and 1 are the outcomes of the small and large limit of mapping units, M, respectively. The fine limit of all these asymptotic behaviors are understood as follows: as the mapping unit scale goes to sufficiently small, each mapping unit can hold at most one individual. In such a situation, the probability to detect the single individual is clearly $\alpha\beta$. The asymptotic behavior suggests that there is a certain scale of the mapping unit above or below which the performance of an ecological survey does not change. Thus, in practice, we need to choose a scale of the mapping unit between these limits. Since the SPPs we applied does not account for the active migration of individuals, the theory is most relevant for species with a relatively sessile life history, such as plants or marine sessile invertebrate species.

Spatial extension of the ecosystem that SPPs accounting individual aggregations describes could be large enough to cover a wide range of spatial scales. For example, Azaele et al. [24] showed that a Thomas model fitted to the distribution map of British rare vascular plant species (see the detailed description of the data set [27]) with three coarse resolutions (40000, 10000, and 2500 km^2) can outperform many existing spatially-implicit models in terms of the down-scaling predictions of the species occupancy probability. In addition, Grilli et al. [28] showed that a special case of the Poisson clustering processes, a group of the point processes where parents locations are followed by a Poisson process [25] such as the

bioRxiv preprint doi: https://doi.org/10.1101/137158; this version posted May 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Neyman-Scott process, recovers the species-area relationship at a local scale to continental scale as predicted by various existing models (e.g., [29]). Hence, even though we used a observation window $\nu(W) = 1024\text{m} \times 1024\text{m}$ as an example, it can be generalized by changing its scale and the sampling intensity. In addition, it is worth noting that albeit individuals of most species are typically aggregated [30, 31] the Thomas process could be approximated by the homogeneous Poisson process under a certain condition: when the intensity of individuals is large, the PDR of the Thomas process comes close to that of the homogeneous Poisson process ($\bar{c}\lambda_p = \{10^{-2},\ 10^{-1}\}$ in Fig. A.1). This is due to increased parent intensity decreasing spatial heterogeneity over the region concerned, suggesting potential applicability of the simpler model to an abundant ecosystem.

For simplicity, we consider a situation where each mapping unit is sampled with the same sampling density, α, and detectability, β, and the location of the sampled unit within a mapping unit is chosen randomly. These are rather idealized assumptions and may be further generalized. For example, it may be reasonable to assume that the sampling density, α, and the detectability, β, become almost 1 at a certain fine scale of the mapping unit. Although such a fine scale may not be achieved because of budgetary constraints, explicitly taking into account the spatial effect on α and β gives us better understanding about the fine scale of asymptotic behavior. In practice, the location of the sampling unit may be determined by more strategic manner depending on ones purpose. Indeed, previous studies had proposed several sampling strategies which emphasize, for example, a spatially contiguous placement of the sampling units to correctly capture ecological patterns (e.g. [32]), a systematic placement to efficiently reflect spatially structured ecological processes [33,34], or a representative design for major environmental gradients to maximize per effort information of organism's distribution [35, 36]. While these strategies have been compared empirically using actual dataset (e.g. [34]), the developed theory in this paper may provide a theoretical base to evaluate the effectiveness and efficiency of such purpose-dependent sampling strategies.

In practice, the developed theory for ecological survey should be, to an extent, complemented by an estimation of the existing number of individuals within given the observation window, W since the intensity affects PDR and FIC (Fig. A.1). An estimation of the population abundance could be done by using historical or surrogate data. Statistical and theoretical methods such as species distribution modeling [37] estimating the occurrence of plant species across scale [24,38] or predicting the population abundance in a coral reef environment [39] may complement these methods. Conducting a pilot survey is one alternative way to estimate the population abundance with a required estimation accuracy. Takashina *et al.* [40] recently developed a framework for the pilot sampling providing a required minimum sampling effort to satisfy the required accuracy. Complemented by these steps, the theory developed here has a potential to significantly improve survey frameworks.

Acknowledgements

We thank M. Akasaka, B. Stewart-Koster, T. Fung, R.A. Chisholm, L.R. Carrasco and S. Azaele for their thoughtful comments and discussions. NT and BK were funded by the

bioRxiv preprint doi: https://doi.org/10.1101/137158; this version posted May 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Program for Advancing Strategic International Networks to Accelerate the Circulation of Talented Researchers of the Japan Society for the Promotion of Science. They acknowledge the support for coordinating the research program from Dr Yasuhiro Kubota and Dr James D. Reimer.

References

[1] S. A. Levin, The problem of pattern and scale in ecology, Ecology 73 (6) (1992) 1943–1967.

[2] E. K. Pikitch, C. Santora, E. a. Babcock, A. Bakun, R. Bonfil, D. O. Conover, P. Dayton, P. Doukakis, D. Fluharty, B. Heneman, E. D. Houde, J. Link, Ecosystem-Based Fishery Management, Science 300 (July) (2004) 2003–2003. doi:10.1126/science.1106929.

[3] C. White, B. S. Halpern, C. V. Kappel, Ecosystem service tradeoff analysis reveals the value of marine spatial planning for multiple ocean uses., Proc. Natl. Acad. Sci. U. S. A. 109 (12) (2012) 4696–4701. doi:10.1073/pnas.1114215109.

[4] N. Takashina, A. Mougi, Maximum sustainable yields from a spatially-explicit harvest model, J. Theor. Biol. 383 (21) (2015) 87–92.

[5] C. R. Margules, R. L. Pressey, Systematic conservation planning., Nature 405 (6783) (2000) 243–53. arXiv:j.1744-7429.2008.00471.x, doi:10.1038/35012251.

[6] T. Yamakita, H. Yamamoto, M. Nakaoka, H. Yamano, K. Fujikura, K. Hidaka, Y. Hirota, T. Ichikawa, S. Kakehi, T. Kameda, S. Kitajima, K. Kogure, T. Komatsu, N. H. Kumagai, H. Miyamoto, K. Miyashita, H. Morimoto, R. Nakajima, S. Nishida, K. Nishiuchi, S. Sakamoto, M. Sano, K. Sudo, H. Sugisaki, K. Tadokoro, K. Tanaka, Y. Jintsu-Uchifune, K. Watanabe, H. Watanabe, Y. Yara, N. Yotsukura, Y. Shirayama, Identification of important marine areas around the Japanese Archipelago: Establishment of a protocol for evaluating a broad area using ecologically and biologically significant areas selection criteria, Mar. Policy 51 (2015) 136–147. doi:10.1016/j.marpol.2014.07.009.

[7] N. Takashina, M. L. Baskett, Exploring the effect of the spatial scale of fishery management, J. Theor. Biol. 390 (2016) 14–22. doi:10.1016/j.jtbi.2015.11.005.

[8] 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.

[9] K. A. Wilson, M. I. Westphal, H. P. Possingham, J. Elith, Sensitivity of conservation planning to different approaches to using predicted species distribution data, Biol. Conserv. 122 (1) (2005) 99–112. doi:10.1016/j.biocon.2004.07.004.

bioRxiv preprint doi: https://doi.org/10.1101/137158; this version posted May 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

[10] P. De Ornellas, E. J. Milner-Gulland, E. Nicholson, The impact of data realities on conservation planning, Biol. Conserv. 144 (7) (2011) 1980–1988. doi:10.1016/j.biocon.2011.04.018.

[11] A. Moilanen, Implications of empirical data quality to metapopulation model parameter estimation and application, Oikos 96 (2002) 516–530. doi:10.1034/j.1600-0706.2002.960313.x.

[12] M. A. McCarthy, J. L. Moore, W. K. Morris, K. M. Parris, G. E. Garrard, P. A. Vesk, L. Rumpff, K. M. Giljohann, J. S. Camac, S. S. Bau, T. Friend, B. Harrison, B. Yue, The influence of abundance on detectability, Oikos 122 (5) (2013) 717–726. doi:10.1111/j.1600-0706.2012.20781.x.

[13] S. A. Field, A. J. Tyre, H. P. Possingham, Optimizing allocation of monitoring efford under economic and observational constraints, J. Wildl. Manag. 69 (2) (2005) 473–482. doi:10.2193/0022-541X(2005)069[0473:OAOMEU]2.0.CO;2.

[14] D. I. Mackenzie, J. A. Royle, Designing occupancy studies: General advice and allocating survey effort (2005). doi:10.1111/j.1365-2664.2005.01098.x.

[15] R. S. Epanchin-Niell, R. G. Haight, L. Berec, J. M. Kean, A. M. Liebhold, Optimal surveillance and eradication of invasive species in heterogeneous landscapes, Ecol. Lett. 15 (8) (2012) 803–812. doi:10.1111/j.1461-0248.2012.01800.x.

[16] A. J. Tyre, B. Tenhumberg, S. A. Field, D. Niejalke, K. Parris, H. P. Possingham, Improving precision and reducing bias in biological surveys: Estimating false-negative error rates (2003). doi:10.1890/02-5078.

[17] B. A. Wintle, M. A. McCarthy, K. M. Parris, M. A. Burgman, Precision and bias of methods for estimating point survey detection probabilities (2004). doi:10.1890/02-5166.

[18] G. E. Garrard, S. A. Bekessy, M. A. McCarthy, B. A. Wintle, When have we looked hard enough? A novel method for setting minimum survey effort protocols for flora surveys, Austral Ecol. 33 (8) (2008) 986–998. doi:10.1111/j.1442-9993.2008.01869.x.

[19] E. C. Pielou, An introduction to mathematical ecology, Wiley-Interscience, New York, 1969.

[20] E. C. Pielou, Ecological Diversity, John Wiley and Sons Inc, New York, 1975.

[21] M. Bode, J. Sanchirico, P. Armsworth, Returns from matching management resolution to ecological variation in a coral reef fishery, Proc. R. Soc. B. 283 (1826) (2016) 20152828.

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

bioRxiv preprint doi: https://doi.org/10.1101/137158; this version posted May 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

[23] 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.

[24] 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.

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

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

[27] S. Hartley, W. E. Kunin, J. J. Lennon, M. J. O. Pocock, Coherence and discontinuity in the scaling of species' distribution patterns., Proc. Biol. Sci. 271 (1534) (2004) 81–88. doi:10.1098/rspb.2003.2531.

[28] J. Grilli, S. Azaele, J. R. Banavar, A. Maritan, Spatial aggregation and the species-area relationship across scales, J. Theor. Biol. 313 (2012) 87–97. arXiv:arXiv:1209.3591v1, doi:10.1016/j.jtbi.2012.07.030.

[29] R. H. MacArthur, E. O. Wilson, The theory of island biogeography, Vol. 1, Princeton University Press, 1967. doi:10.2307/1796430.

[30] 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.

[31] 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.

[32] B. Güler, A. Jentsch, I. Apostolova, S. Bartha, J. B. M.G., G. Campetella, R. Canullo, J. Házi, J. Kreyling, J. Pottier, G. Szabó, T. Terziyska, E. Uurlu, C. Wellstein, Z. Zimmermann, J. Dengler, How plot shape and spatial arrangement affect plant species richness counts: implications for sampling design and rarefaction analyses, J. Veg. Sci.doi:10.1111/jvs.12411,2016.

[33] D. Bellhouse, Biometrika, Biometrika 64 (3) (1977) 605–611.

[34] I. J. Fiske, E. M. Bruna, Alternative spatial sampling in studies of plant demography: Consequences for estimates of population growth rate, Plant Ecol. 207 (2) (2010) 213–225. doi:10.1007/s11258-009-9666-4.

[35] A. N. Gillison, K. R. W. Brewer, The use of gradient directed transects or gradsects in natural resource surveys, J. Environ. Manage. 20 (April) (1985) 103–127.

bioRxiv preprint doi: https://doi.org/10.1101/137158; this version posted May 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

[36] H. M. Neave, R. B. Cunningham, T. W. Norton, H. A. Nix, Preliminary evaluation of sampling strategies to estimate the species richness of diurnal, terrestrial birds using Monte Carlo simulation, Ecol. Modell. 95 (1) (1997) 17–27. doi:10.1016/S0304-3800(96)00016-6.

[37] S. Phillips, R. Anderson, R. Schapire, Maximum entropy modeling of species geographic distributions, Ecol. Modell. 190 (3-4) (2006) 231–259. doi:10.1016/j.ecolmodel.2005.03.026.

[38] F. He, K. J. Gaston, Estimating Species Abundance from Occurrence, Am. Nat. 156 (5) (2000) 553–559. doi:10.1086/303403.

[39] C. Mellin, S. Andrefouet, D. Ponton, Spatial predictability of juvenile fish species richness and abundance in a coral reef environment, Coral Reefs 26 (4) (2007) 895–907. doi:10.1007/s00338-007-0281-3.

[40] N. Takashina, B. Kusumoto, M. Beger, S. Rathnayake, H. P. Possingham, Spatially Explicit Approach To Population Abundance Estimation In Field Surveys, bioRxivdoi:https://doi.org/10.1101/131037.

bioRxiv preprint doi: https://doi.org/10.1101/137158; this version posted May 12, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

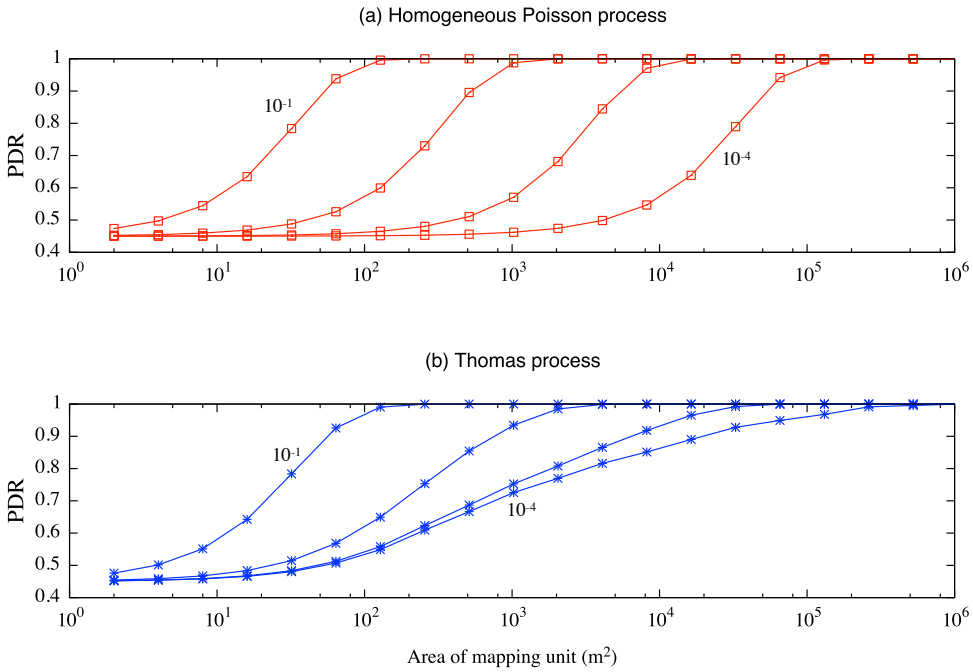

Figure A.1: Effect of the intensity (λ, $\bar{c}\lambda_p$) in the observation window, W, on the theoretical presence detection rate (PDR), Eqs. (15), (20). The intensity of the Thomas process is manipulated by changing the parent intensity λ_p. Individual distribution patters are according to the (a) Homogeneous Poisson process and (b) Thomas process. For the Thomas process, the PDR curves converge as the intensity becomes small, and come close to the corresponding curve of the homogeneous Poisson process as the intensity of the Thomas process increases. This is an effect that the increased parents intensity decreases spatial heterogeneity over the concerned region. For both panels, the order of the intensity monotonically decreases from left to right.