ECOGRAPHY

Research

Developing generalized sampling schemes with known error properties: the case of a moving observer

Nao Takashina and Evan P. Economo

N. Takashina (https://orcid.org/0000-0002-9594-9264) ✉ (nao.takashina@gmail.com) and E. P. Economo, Biodiversity and Biocomplexity Unit, Okinawa Inst. of Science and Technology Graduate Univ., Onna-son, Okinawa, Japan.

Ecography
44: 293–306, 2021
doi: 10.1111/ecog.05198

Subject Editor: Miguel G. Matias
Editor-in-Chief: Miguel Araújo
Accepted 12 October 2020

Pattern in space and time is central to ecology, and adequately designed ecological sampling is needed to resolve those patterns, pursue ecological questions and design conservation strategies. Recently, there has been an explosion of various ecological data due to the proliferation of online data-sharing platforms, citizen science programs and new technology such as unmanned aerial vehicles (UAVs), but data reliability, consistency and the error properties of the sampling method are usually uncertain. While there are a number of standard survey protocols for different taxa, they often subjectively designed and standardization is meant to facilitate repeatability rather than produce a quantitative evaluation of the data (e.g. error properties). Here, we describe an ecological survey scheme consisting of an 'algorithm' to be followed in the field that will result in a standard set of data as well as the error properties of the data. While many such sampling schemes could be developed that target different types of organisms, we focus on one case of a moving observer attempting to detect a species in the field (e.g. a birder, UAV, etc.) with the goal of producing a presence–absence map. The multiscale model developed is spatially explicit and accommodates inherent survey tradeoffs such as sampling speed, detectability and map resolution. Given a set of sampling parameters, the model provides estimates of the total sampling time and map accuracy translated into the probability of false negative. Additionally it also provides an actual and sampled occupancy–area curve across mapping resolutions that can be utilized to discuss sampling effects. While the proposed sampling framework is simple, the same general approach could be adapted for other conditions to meet the needs of a particular taxon. If a set of 'canonical' sampling algorithms could be developed with known mathematical properties, it would enhance reliability and usage of ecological datasets.

Keywords: citizen science, ecological survey, occupancy–area curve, presence–absence map, UAVs

NORDIC SOCIETY OIKOS

www.ecography.org

Introduction

Documenting spatiotemporal patterns is a central need for ecology (Levin 1992) and ecological surveys provide a foundation for asking ecological questions and supports decision making for ecosystem management and conservation (Nichols and Williams 2006, McDonald-Madden et al. 2010, Danielsen et al. 2014). Although the recent progress in generating and consolidating large data have enabled us to reveal common ecological patterns across the globe and discuss ecosystem conservation at global scales (Orme et al. 2006, Pimm et al. 2014, Di Marco et al. 2019), our knowledge of biodiversity is still largely deficient due to geographical and taxonomic sampling biases (Böhm et al. 2013, Dulvy et al. 2014, Bland et al. 2015).

Over the last few decades, global ecological data have increasingly been available given rapid proliferation of online data-sharing platforms (e.g. Global Biodiversity Information Facility (GBIF) (GBIF.org 2019)) working synergistically with citizen science programs such as eBird (Sullivan et al. 2009), and emergent tools such as unmanned areal vehicles (UAVs) (Anderson and Gaston 2013, Bonney et al. 2014, Crutsinger et al. 2016, Chandler et al. 2017). In addition, UAVs have recently been introduced to the citizen science program, for instance, in Guyana to monitor agricultural fields (Cummings et al. 2017). These have been widely utilized in ecological studies and ecosystem management at scales that have previously not been feasible (Dickinson et al. 2012, Kobori et al. 2016, Chandler et al. 2017). However, there are a number of possible survey protocols (Stem et al. 2005, Burton et al. 2015), and often they are arbitrarily tailored to local situations (Danielsen et al. 2014). Since field observations are the basis for ecological datasets, our understanding of ecological phenomena is ultimately influenced by these observation methods. In addition, detailed surveillance procedures and potential data bias and error are often not explicitly mentioned, leaving quality of available data unclear. The use of ambiguous data risks mis-interpretation of ecological structure as surveillance methods such as the area and shape of sampling regions could alter observed patterns (Takashina et al. 2019).

The development of transparent and widely applicable survey methods with a theoretical basis would both improve data quality and maximize return on investment of time and resources. Given that monetary and time constraints are an issue for any conservation effort, ecological surveys must be undertaken in a cost-effective manner. As data quality significantly alters the conservation benefits (McDonald-Madden et al. 2010), incorporating surveillance time and data accuracy into the management planning facilitates conservation success (Chades et al. 2008, Bennett et al. 2018, Takashina et al. 2018b). Therefore, having ecological survey guidelines provide a quantitative benchmark of data quality and survey time and can improve the decision making of the survey design. Quantifying accuracy of available data also substantially helps to judge its utility given a specific ecological question and conservation (IUCN 2017).

There have been a number of sampling designs proposed and examined for different field systems, such as biodiversity estimation of spiders in tropical ecosystems (Coddington et al. 1991), termite assemblages in tropical forests (Jones and Eggleton 2000), aquatic invertebrate assemblages in three Tennessee River tributaries (Kerans et al. 1992), the marine coastal ecosystem monitoring by a remote device (Aguzzi et al. 2011), coral reef biodiversity (Reef Life Survey 2019) and many more. The standardization of sampling methods can lead to comparable datasets. On the other hand, sampling performance parameters, including the degree of error in the data for a given amount of sampling, usually remains unclear. There is also a large body of literature developing theory for aspects of survey protocols such as effort allocation under budgetary constraints to achieve optimal information gain (Field et al. 2005, Mackenzie and Royle 2005) or optimal surveillance for invasive species (Epanchin-Niell et al. 2012) in a spatially-implicit context. On the other hand, generalized spatially explicit survey schemes are largely deficient. Spatially explicit protocols are necessary since they allow us to explore detailed survey protocols and quantitatively discuss inherent trade-offs such as sampling path and map resolution. Previously, Baxter and Hamilton (2018) numerically performed how the species detection error via an UAV survey affects the population estimate and occupancy within 50×50 grid cells where individual distribution are explicitly modeled. Takashina et al. (2018a) developed a multi-scale mathematical framework to assess survey accuracy given a detectability on the presence–absence map where the survey region, resolution of the map and actual sampling region can be flexibly changed. These attempts focused on unveiling generic tradeoffs in the survey such as detectability, spatial resolution and sampling accuracy.

To ensure ecological sampling is conducted in a cost–effective manner with a specific sampling design, including sampling path, map resolution and sampling error, we need a theoretical framework that quantifies the sampling time and its accuracy in a spatially explicit setting. Regardless of observer considered (e.g. field ecologist, citizen scientist or UAV) and sampling method to detect individuals (e.g. ocular, plot, transect or image-based methods (Booth et al. 2005, 2006, Booth and Cox 2008)), a wide variety of ecological surveys seeking to observe 'events' (e.g. presence of a species, interaction between species, etc.) within a specific area and sampling period commonly consist of the four outcomes: 1) the event was not detected during the sampling period because the event did not occur, 2) the event occurred during the sampling period but was not detected (i.e. false negative), 3) detection of an event that occurred within the given sampling period, 4) false detection of an event. The component (4) has two types: false detection of (4-a) an event that did not occur; and (4-b) an event that occurred but it was not detected until the event of the false detection. The inaccuracy of data arises from false negatives and false positives. We propose that explicitly modeling these components under general situations and developing a solid insight into mathematical properties of the model can lead to the development

of a general survey protocol utilized by a variety of samplers and sampling methods.

Here we aim to develop a general spatially-explicit ecological survey framework to produce a presence–absence map as an output. We focus on cases involving a moving observer, such as a field ecologist, wildlife manager, UAV or citizen scientist seeking to create a presence–absence map within an arbitrary area. The sampling process involves an inherent error of individual detection, undermining map accuracy by introducing false negative and positives into the resulting map. Presence–absence maps are one of the example of a survey output, and these have a broad applicability of both ecology and conservation biology including the estimation of the species extinction risk (IUCN 2017), measuring beta diversity (Koleff et al. 2003) and understanding the impact of invasive species expansions (Sanders et al. 2003). Our framework naturally provides an occupancy probability and therefore we can also obtain the actual and sampled occupancy–area curve (Kunin 1998, Azaele et al. 2012) across mapping resolutions. Our goal is to demonstrate that, by carefully determining a parameters of a given sampling scheme, the survey framework can offer an quantitative estimation of expected total survey time and accuracy of the obtained presence–absence map measured by the false negative probability of absence region; the central trade-offs of ecological surveys. We then perform an in silico demonstration of how our method can be used to design sampling and provide theoretical predictions using several test datasets, including both simulated datasets and the actual spatial pattern of tree distributions from the 50 ha plot at Barro Colorado Island, Panama.

Methods

Let us assume that an ecological survey is carried out in a region W where individuals of a target species are distributed randomly or tend to form intraspecific clusters (Supplementary material Appendix A Fig. S1). The region W is surveyed by an observer such as field biologist, citizen scientist or UAV (henceforth, 'observer'). The purpose of this section is to introduce a method to quantify the ability of each observer, and to introduce a survey protocol that can be widely used in ecosystem management and monitoring programs. In this framework, key outputs of the ecological survey are the expected total survey time and map accuracy in terms of the probability of false negative. We summarized the key parameters introduced below in Table 1.

Table 1. Definition of central parameters.

Symbol	Parameter
W	Observation window
M	Mapping resolution
S	Sampling resolution
$\nu(R)$	Area of region R
N_M	Number of mapping units
λ_{dist}	Intensity of individuals of distribution pattern dist
ε	probability of sampling error
t_s	sampling stopping time

Quantifying the ability of the observer

Various observers can be represented by the framework discussed by changing the ability level of an observer, and these are characterized by the detection error of an individual ε and sampling intensity. The sampling intensity reflects the average fractional area searched within a concerned region until a given sampling time t_s by an observer. The sampling intensity is a product of various factors such as the sampling trajectory, sampling method and the speed to follow the sampling trajectory (Fig. 1). Although the sampling intensity can be uniquely determined for each observer provided a sampling trajectory and its speed, multiple definitions can be possible since the above-mentioned factors are not independent but interact each other. Here, we use an intuitive definition by introducing the sampling unit S: average region that an observer surveys at a given time unit and speed; for example, a citizen scientist surveys region S_1 every 5 min with the ground speed 25 m min^{-1} or region S_2 every 5 min with the ground speed 30 m min^{-1}, and a UAV surveys region S_1 every 10 s with the flight speed 3 m s^{-1}. Then, sampling intensity by time t_s is comprised of sampling efficiency $\alpha(t)$ and the area (represented by $\nu(\cdot)$) of sampling unit S given a sampling speed:

$$\text{Sampling intensity} = \frac{\nu(S)\int_0^{t_s}\alpha(t)dt}{\nu(\text{concerned region})}$$

The sampling efficiency measures the redundancy of a sampling trajectory and it satisfies $0 < \int_0^{t_s}\alpha(t)dt \le t_s$ where the equality occurs under the ideal sampling trajectory. It gives the highest sampling intensity, and the smaller sampling intensity occurs when the sampling trajectory causes redundant sampling (i.e. trajectory of Fig. 1, top). An easy way to minimize sampling redundancy is to subdivide each mapping unit into sampling units, and sample each unit at each time unit in a predetermined trajectory where overlap does not occur. In this study, we will adopt this approach. The integral becomes the summation in case of the discrete–time survey. Each individual sampling is associated with the sampling error that may depend on the sampling speed: too fast sampling speed compared to the ability of an observer may cause a high detection error while the effect of decreasing the sampling speed may be diminished at a certain speed. However, we do not require to model this relationship for our purpose, and we arbitrary determine the sampling unit S and sampling error ε.

A general protocol for ecological surveys by a moving observer

We here introduce a general field survey protocol (Fig. 2). We use a discrete time in the presentation that may be easier

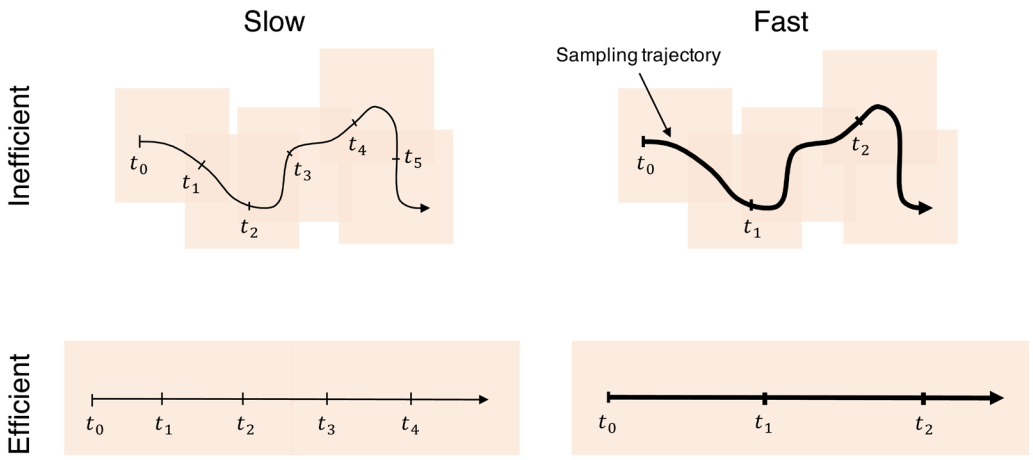

Figure 1. Schematic diagrams of the region sampled (colored region). Sampling trajectory (arrows) and the sampling speed (e.g. m s^{-1}; curve width) to follow the trajectory determine the sampling intensity. The labels t_i represents time steps of the sampling on the trajectories. Faster (thick curve) and efficient sampling trajectory increases the area sampled per unit time, than slower (thin curve) and inefficient sampling. However, faster sampling speed may cause larger sampling error.

to handle in a field survey practice but its generalization is straightforward. The survey protocol follows the two rules:

1. Define the resolution of map M (mapping unit).
2. Each mapping unit is sampled with the sampling unit S, efficiency $a(t)$ and detection error ε with an arbitrary trajectory as in Fig. 1.

By these two basic rules we can conduct various survey protocols other than the protocol we present below. Since the size of mapping unit M is unique, the number of mapping units is $N_M = \nu(W)/\nu(M)$. Since we focus on the survey to obtain a presence–absence map, we further introduce the following two steps:

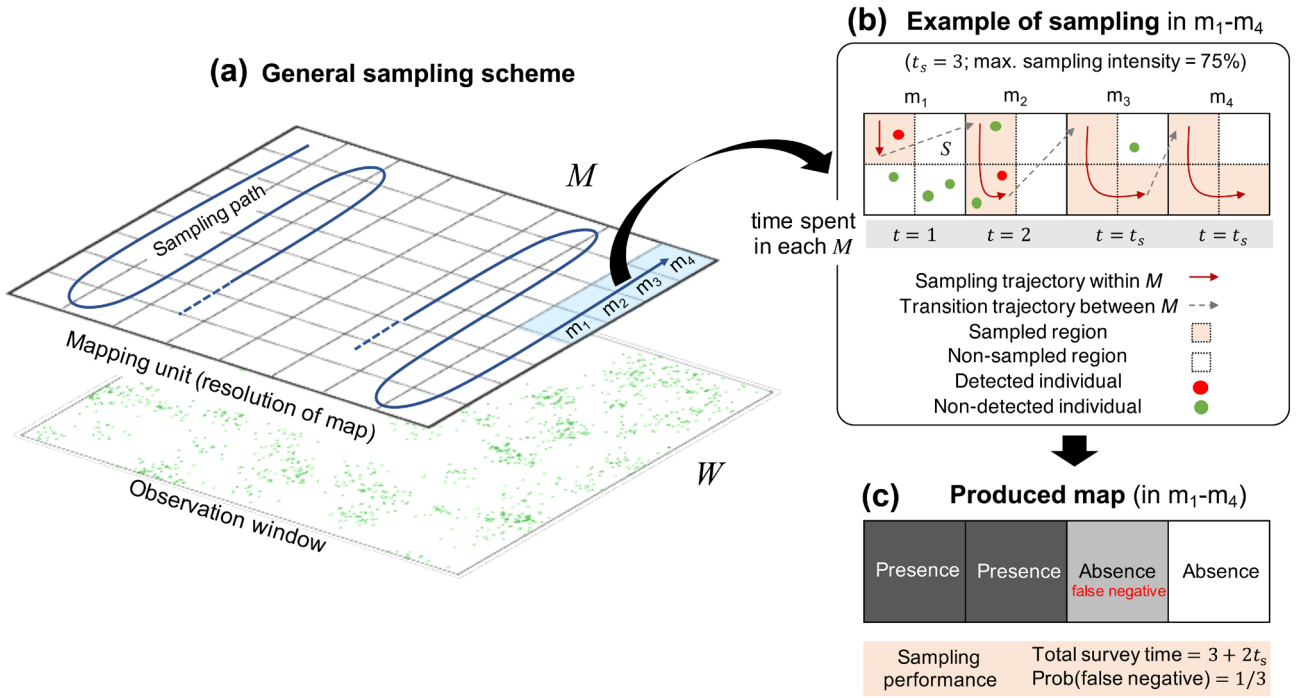

Figure 2. General survey scheme proposed to create a presence–absence map with observation window where the target population is surveyed, mapping unit and sampling unit (W, M, S). By determining (a) a sampling path to visit all mapping units and (b) sampling trajectory within each mapping unit based on sampling unit, we can create (c) a presence–absence map and calculate the expected total survey time and the probability of false negative. An example of the sampling and the resulting presence–absence map shown here is only the last four mapping units (labeled m_1, m_2, m_3 and m_4). In this example, there are 64 mapping units and each mapping unit composed of 4 sampling units.

3. Define the maximum sampling time in each mapping unit t_s.
4. Go to next mapping unit when a first individual is observed, a false detection occurs or sampling time reach the stopping time t_s.

In the following mathematical descriptions, we present the situation where the possibility of false positive detection is negligible for simplicity. The integration of false positive detection is straightforward, but it requires further assumptions of causes of the false detection, one more probability variable and the calculations of two extra average times to estimate a total sampling time compared to the discussion below. We leave technical details in the Supplementary material Appendix D, and we will present only results in the main text.

Given the sampling protocols 1–4, the probability of false negative is calculated probabilistically

$$P\big(\text{No detection of individual by time } t_s \text{ within a mapping unit with individual(s)}\big)$$

and the expected total survey time t_{samp} is described by

$$\begin{aligned} t_{\text{samp}} = N_M t_s \{ & P(\text{No individual exists in a mapping unit}) \\ & + P(\text{No detection of individual by time } t_s \\ & \qquad \text{within a mapping unit with individual(s)}\} \\ & + N_M E[\text{Time to detect first individual}] \\ & \times P(\text{Detection of an individual by time } t_s) \\ & + (N_M - 1) E[\text{Traveling time from one} \\ & \qquad \text{mapping unit to another}] \end{aligned}$$

where, P(Detection of an individual by time t_s) is the (sampled) occupancy probability of the presence–absence map. Traveling times from one location to the next patch may be independent each other and therefore it may be linearly scaled to the number of mapping units, and we omit in the following discussion. Let us define the random variable $X(t)$ be the accumulated number of detected individuals at time t in a given mapping unit, $E[t]$ be the average time that the variable $X(t) = 0$ switches to $X(t) > 0$, and the random variable Y be the number of individual in a given mapping unit. Then the above expressions are formally described as

$$P(X(t_s) = 0, Y > 0) \tag{1a}$$

$$\begin{aligned} t_{\text{samp}} = N_M t_s \{ & P(Y = 0) + P(X(t_s) = 0, Y > 0)\} \\ & + N_M E[t] P(X(t_s) > 0, Y > 0) \end{aligned} \tag{1b}$$

When the sampling time is fixed within each mapping unit (fixed-time sampling), the sampling time is simply $t \times$ (number of sampling units). Obviously, the probability of false negative is the same as the sampling method mentioned above. The occupancy–area curve is obtained by calculating $P(X(t_s) > 0, Y > 0)$ across mapping resolutions.

Takashina et al. (2018a) previously obtained the probability $P(X(R) > 0 | Y > 0)$ under a fixed-area survey, where $X(R)$ is the number of individuals found in the fixed survey region, R, and they regarded it to the presence mapped fraction. Since the probability of false negative in our protocol is the same as that of the fixed-area survey, we also have the relationship $P(X(t_s) > 0 | Y > 0) = P(X(R) > 0 | Y > 0)$. Therefore, we can use the following relationship derived in Takashina et al. (2018a)

$$P(X(t_s) > 0 \,|\, Y > 0) = \frac{1 - P(X(t_s) = 0)}{P(Y > 0)} \tag{2}$$

Case studies using simulated and real-world datasets

Following the theoretical analysis, we aimed to test and demonstrate the survey method using experimental surveys. Although one could do this with new real field surveys, we chose to perform in silico experiments by simulating the survey of target datasets because through simulation many replicates can be run with different conditions, and the target data are known a priori so errors can be assessed. As test cases, we simulated artificial datasets with different clustering properties, and used real species data from the well known Barro Colorado Island (BCI) 50 ha plot dataset (ForestGEO 2020).

Simulated individual distributions

In the theoretical analysis, we assume random or clustering individual distributions. Here we outline the process to simulate individual distribution data with these properties. More detailed discussion is found in the Supplementary material Appendix A.

The random and clustering individual distributions are generated by applying the theory of spatial point processes that accommodate stochasticity in individual distributions in region W. Specifically, random individual distributions are generated by a homogeneous Poisson process, and a Thomas process, widely applied to characterize intraspecific aggregation patterns (Plotkin et al. 2000, Morlon et al. 2008, Azaele et al. 2012, May et al. 2018, Takashina et al. 2018a, 2019), describes clustering spatial patterns. The parameters do not change with a spatial scale considered, and aggregation of these processes can realize well-established community patterns across scales such the species–area relationship (Plotkin et al. 2000, Takashina et al. 2019).

In the homogeneous Poisson process, the number of individuals X in a given region R with area $\nu(R)$ and the intensity λ_{po} follows the Poisson distribution

$$P(X = k) = \frac{(\lambda_{\text{po}} \nu(R))^k}{k!} e^{-\lambda_{\text{po}} \nu(R)}$$

Similarly, we can define the intensity of the Thomas process λ_{th}, and in the analysis, we set $\lambda_{th}=\lambda_{po}$ to satisfy the average numbers in a given area are the same under both individual distributions.

Setting sampling details

We set $\nu(S)$, $\nu(M)$ and $\nu(W)$ to be the form 2^n ($n=0, 1, 2\ldots$) for numerical convenience (e.g. we set $\nu(W)=2^4$ km^2). Also, we present the normalized sampling time under fixed- and non-fixed time samplings for the presentation convenience. For normalizing the total sampling time, we simply divide Eq. 1b by $N_M t_{max}$, where t_{max} is the sampling time that the sampling intensity becomes 1: $\nu(S)t_{max}/\nu(M)=1$. Namely, this normalized value measures the sampling time relative to the time for the exhaustive survey. Then we perform the survey proposed when the average population density is 10, 100 or 1000 km^{-2} for both random and clustering individual distributions. We provide examples of realized point patterns for both random and clustering individual distributions in Supplementary material Appendix A Fig. S1 with a parameter set used in the following analysis. We also examine two probabilities of the sampling error $\varepsilon=0.1$ and 0.3. We run 10^5-time simulations for each scenario to quantify the sampling average and 5 and 95 percentile of the simulations. To simulate an occupancy–probability curve, we set t_s so as to satisfy the sampling intensity is 0.5: $\nu(S)t_s/\nu(M)=0.5$, where the area of mapping resolutions are $\nu(M)=\{2^{-11}, 2^{-10}, \ldots, 2^4\}$. For surveys with false detection, we will focus on the sampling time and the probabilities of false negative and positive when the probability of the sampling error is $\varepsilon=0.1$ as the results are qualitatively similar to the surveys without false detection.

An empirical dataset: plant species in the 50-ha forest plot at Barro Colorado Island, Panama

We will demonstrate application of the sampling design and its expected sampling time and the probability of false negative using the well known vascular plant data at Barro Colorado Island (BCI) where highly spatially-resolved location data is available in a well-defined region with the area $\nu(W)=1000\times 500$ m. We examine the dataset of *Zanthoxylum panamense* (n = 463) and *Cecropia obtusifolia* (n = 362) independently as test cases. We will show, via Ripley's K function, that these species show a comparable pattern to the random or clustering individual distribution at a given distance. Note that we are using the BCI data here as a 'true' pattern which we 'sample' using in silico surveys by a moving observer. The BCI plot is already exhaustively sampled with low error through standard protocols (ForestGEO 2020) and we are not suggesting those protocols should be changed. Rather, the point of this exercise is to test sampling methods in silico on a real-world high quality dataset as a target.

Setting sampling details

To make the proposed sampling protocol fully available, we require equal-sized mapping units with an area $\nu(M)$ consisting of equal-sized sampling units with an area $\nu(S)$. In this example, we choose the area of mapping resolution and sampling resolution $\nu(M)=32.05\times 32.05$ m and $\nu(S)=4.05\times 4.05$ m, respectively, that gives the total number of mapping units 1024 and sampling units within each mapping unit 64. We set the probability of sampling error to be 10% ($\varepsilon=0.1$). Here we do not need to specify an actual time to spend in each mapping unit, to calculate the fractional time as in the theoretical analysis. However, the actual total sampling time is readily recovered once the time unit is specified. We perform 10^4-time simulations for each scenario to quantify the sampling average and 5 and 95 percentile of the simulations.

Results

Theoretical properties of the proposed method

Here, we demonstrate how false negative Eq. 1a and the total survey time Eq. 1b are derived. By doing so, we can show these forms can be greatly simplified, and make the underlying tradeoffs clear. Later on this chapter, we perform numerical simulations for the sake of verification of our analysis and visualizations. Here, without loss of generality, we assume that the time step is discrete and the sampling trajectory is ideal (schematic image is provided in Fig. 2b), and therefore the sampling intensity until time t_s ($\leq t_{max}$) is represented by $\nu(S)t_s/\nu(M)$. In Supplementary material Appendix C, we also asymptotic behavior when the mapping unit becomes very small.

Random individual distribution

Here we overview the derivations of each probability required for Eq. 1a, 1b, and technical details are left in Supplementary material Appendix B. Later, we will show the same manner will immediately follow in the case of clustering individual distributions.

When individuals are distributed randomly (i.e. via homogeneous Poisson process), the probability of false negative is directly obtained by calculating the probability of 1) mis-detection given individual encounter (hitting), and 2) encounter no species (non-hitting) until time t_s given individual existing in a mapping unit as follows (Supplementary material Appendix B)

$$P_{po}(X(t_s)=0, Y>0)=e^{-\lambda_{po}\nu(S)t_s(1-\varepsilon)}-e^{-\lambda_{po}\nu(M)} \quad (3)$$

where, the subscript po indicates the underlying individual distribution (Poisson process). The probability of false negative can be reduced more efficiently by increasing the sampling stopping time t_s when the intensity λ_{po} and sampling unit $\nu(S)$ are larger and detection error ε is smaller. Equation 3 immediately gives us the probabilities of $P_{po}(X(t_s)>0|Y>0)$, $P_{po}(X(t_s)>0, Y>0)$ and $P_{po}(X(t_s)=0, Y>0)$ (Supplementary material Appendix B) where the occupancy probability of the presence–absence map has the form

$$P_{\mathrm{po}}(X(t_s)>0, Y>0) = 1 - e^{-\lambda_{\mathrm{po}}\nu(S)t_s(1-\varepsilon)} \quad (4)$$

These probabilities allow us to calculate the total survey time (Supplementary material Appendix B)

$$t_{\mathrm{samp}}^{\mathrm{po}} = N_M \sum_{t=0}^{t_s-1} P_{\mathrm{po}}(X(t)=0) = N_M \frac{1-e^{-\lambda_{\mathrm{po}}\nu(S)t_s(1-\varepsilon)}}{1-e^{-\lambda_{\mathrm{po}}\nu(S)(1-\varepsilon)}} \quad (5)$$

This suggests intuitive characteristics of ecological survey: the total sampling time is proportional to the number of mapping units N_M, an effect of a mapping resolution M. As in the case of Eq. 3, the sampling time is reduced more efficiently by increment the sampling stopping time t_s when the factors λ_{po}, $\nu(S)$ and $1-\varepsilon$ are larger.

Clustered individual distribution

The same discussion follows when individuals tend to create clusters in space which can be described by the Thomas process. The form of zero probability is given in Supplementary material Appendix A Eq. A3, and the probability of false negative and the occupancy probability are immediately obtained by using the relationship in Eq. 2.

Provided the false negative probability, the sampling time has the following form:

$$t_{\mathrm{samp}}^{\mathrm{th}} = N_M \sum_{t=0}^{t_s-1} P_{\mathrm{th}}(X(t)=0) \quad (6)$$

where, th indicates the underlying individual distribution (Thomas process) as above. This has the same form as the first line of Eq. 5 although we could not further simplify.

Case studies: using 'in silico' surveys to test the theory using example datasets as targets

Simulated datasets: random and clustering individual distributions

Here we generate random and clustering individual distribution patterns in a manner presented in Supplementary material Appendix A. We perform numerical simulation to verify the mathematical analyses above.

Expected sampling performance without false detection

Provided by simulated individual distributions, we can conduct ecological survey on computer and quantify its performance. Figure 3 shows that the theoretical values have a good agreement with numerical values. As predicted by the discussion above, the reduction of the normalized sampling time (top) and the probability of false negative (bottom) given an increment of the sampling stopping time t_s is larger when the intensity λ_{po} is larger. This is also true for the clustering individual distributions. In addition, the normalized sampling time is nearly linearly scaled by the normalized sampling time when the intensity λ_{po} and λ_{th} are small as the theoretical discussion suggests. The same line of discussions hold true with the different probability of sampling error ($\varepsilon=0.3$; Supplementary material Appendix A Fig. S1) and sampling protocols with $\varepsilon=0.1$ and 0.3 (see Supplementary material Appendix A–D for plots with ($\nu(M)$, $\nu(S)=2^{-4}\times 2^{-4}$ km^2, $2^{-6}\times 2^{-6}$ km^2), Fig. S2; and ($\nu(M)$, $\nu(S)=2^{-3}\times 2^{-3}$ km^2, $2^{-5}\times 2^{-5}$ km^2), Fig. S3). For a comparison, we also plot the total survey time under the fixed-time survey (Fig. 3 top) where

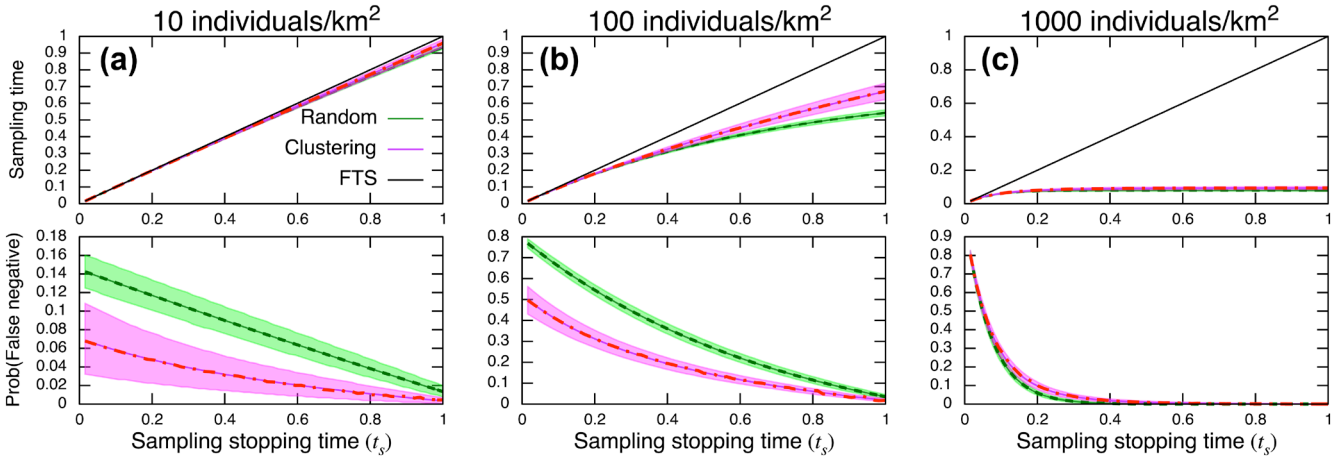

Figure 3. Normalized sampling time and false negative given normalized stopping time t_s with the probability of sampling error $\varepsilon=0.1$. For each individual distribution scenarios, the numerical average (lines) and its theoretical value (dashed) are provided. The shaded area are between 5 and 95 percentiles of a 10^5-time numerical simulation. FTS represents the fixed-time survey. The scales of surveys are are $\nu(W)=4\times 4$ km, $\nu(M)=2^{-3}\times 2^{-3}$ km, $\nu(S)=2^{-6}\times 2^{-6}$ km and the probability of sampling error is $\varepsilon=0.1$. The parameter values of the clustered (Thomas) process are $\lambda_{\mathrm{th}}=5$, $c=20$, $\sigma_{\mathrm{th}}=0.1$. Supplementary material Appendix A Fig. S1 for example realizations of individual distributions with the intensities $\lambda_{\mathrm{po}}=\lambda_{\mathrm{th}}=100$.

sampling time in each mapping unit is fixed to t_s (i.e. the total survey time is $N_M t_s$).

Figure 4 shows (a) actual and sampled occupancy–area curves across mapping unit $\nu(M)$ and (b) relative probability of sampled and actual that measures the agreement of sampled estimation. We observe that the probability of occurrence asymptotically converges to $\lambda_{\text{dist}}\nu(S)t_s(1-\varepsilon)$ and 1, when the area of the mapping unit approaches 0, and when the area of mapping unit becomes sufficiently large, respectively. At intermediate sizes of mapping unit, randomly distributed individuals show higher probabilities occurrence than spatially aggregated individuals regardless of the actual or sampled curve (Fig. 4a). The relative probability has the form of $P(X(t_s) > 0 | Y > 0)$, and the the agreement is decreased with the size of mapping unit (Fig. 4b). When the sampling intensity is fixed across mapping unit scales, this is equivalent to the presence mapped fraction, $P(X(R) > 0 | Y > 0)$ where $X(R)$ is the number of individuals in the region R (Takashina et al. 2018a) and it asymptotically converges to sampling intensity × error rate when the area of the mapping unit approach 0 (Takashina et al. 2018a); $\nu(S)t_s(1-\varepsilon)$ in our context.

Expected sampling performance with a possibility of false detection

Here we briefly demonstrate that the similar results can be obtained in ecological survey with a false positive detection in line with the result in Fig. 3 ($(\nu(M), \nu(S)) = (2^{-3} \times 2^{-3}$ km^2, $2^{-6} \times 2^{-6}$ km^2), the probability of sampling error is $\varepsilon = 0.1$). We assume that the number of false positive detection after sampling the area $\nu(S)t$ follows a Poisson distribution with an intensity λ_{fp}: Po($\lambda_{fp}\nu(S)t$) (Supplementary material Appendix D). Qualitatively similar patterns are realized for the normalized sampling time (Fig. 5; top) and the probability of false negative (Fig. 5; middle). Intuitively, the probability of false negative increases with the sampling time (Fig. 5; bottom) and it becomes negligible when the number of individuals are sufficiently large (Fig. 5c; bottom).

The same line of discussions hold true with other sets of mapping and sampling resolutions (see Supplementary material Appendix A–D for plots with $(\nu(M), \nu(S)) = (2^{-4} \times 2^{-4}$ km^2, $2^{-6} \times 2^{-6}$ km^2), Fig. S4; and $(\nu(M), \nu(S)) = 2^{-3} \times 2^{-3}$ km^2, $2^{-5} \times 2^{-5}$ km^2), Fig. S5).

Field dataset: plant species in Barro Colorado Island (BCI)

Here, we apply our sampling protocol to the spatial distribution of two vascular plants, *Zanthoxylum panamense* and *Cecropia obtusifolia* in the 50-ha (1000 × 500 m) plot at BCI (Fig. 6a–b). This section is intended to develop a connection to well-known dataset in spatial ecology, although we don't argue it is practically useful for this particular study system (e.g. BCI is already thoroughly sampled for all species, and thus single species surveys with optimized protocols are not necessary). The BCI data is taken as a target that we attempt to sample using our simulated surveys.

Spatial aggregation of BCI data

We first quantify spatial aggregation pattern of the data points to make the comparison to theoretical results clear. Ripley's K function summarizes such a spatial statistics of data points $(s_1, s_2, \ldots, s_N)$ up to the distance t within a concerned region of area A_0 (Ripley 1977, Chiu et al. 2013)

$$\hat{K}(t) = \frac{1}{\hat{\lambda}} \sum_{\substack{i,j \\ i \neq j}} \frac{1}{w(s_i, s_j)} \frac{I(\| s_i - s_j \| < t)}{N} \tag{7}$$

where, $\hat{\lambda} = N / A_0$ is the data density with the number of data points N, $w(s_i, s_j)$ is the edge-correction factor, the proportion of a circle centered at s_i with the radius $\|s_i - s_j\|$ overlapping with A_0, and I is the indicator function ($I(\cdot) = 0$ if $(\cdot)$ is true, 0 otherwise). For example, if the value of $\hat{K}(t)$ of any individual distributions is larger than that of a random individual distribution at distance t, it indicates that the concerned distribution is more aggregated than the random distribution up to the distance t, and vice versa. This is a relative concept and the same discussion is made for a comparison to the aggregated (Thomas) process used in the preceding analysis. Hence, *Zanthoxylum panamense* shows similar pattern as

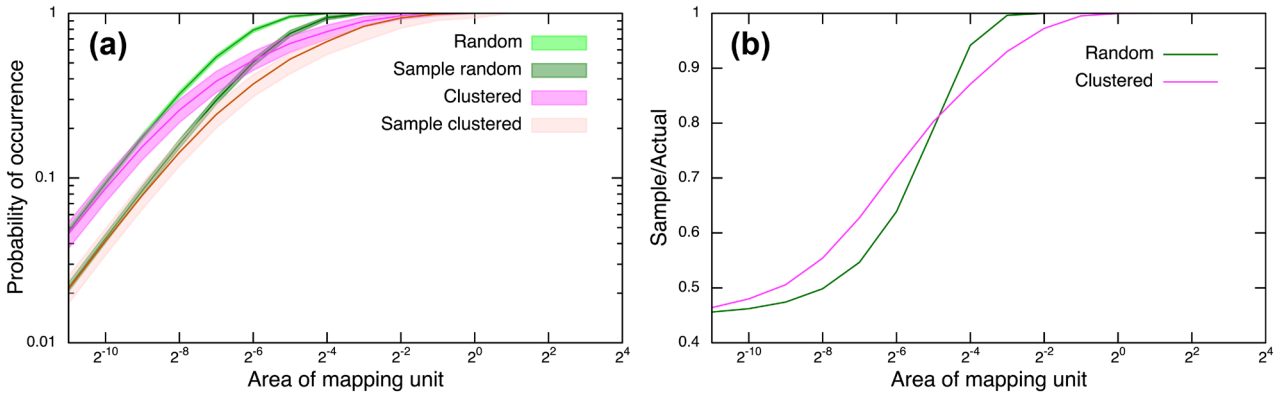

Figure 4. (a) Occupancy–area curves and (b) relative occupancy probability of sampled and actual values provided the sampling intensity 0.5. The shaded areas are between 5 and 95 percentiles of a 10^5-time numerical simulation, and the lines are the theoretical value. The intensity of both distributions λ_{dist} are 100. The other parameters used are the same as in Fig. 3.

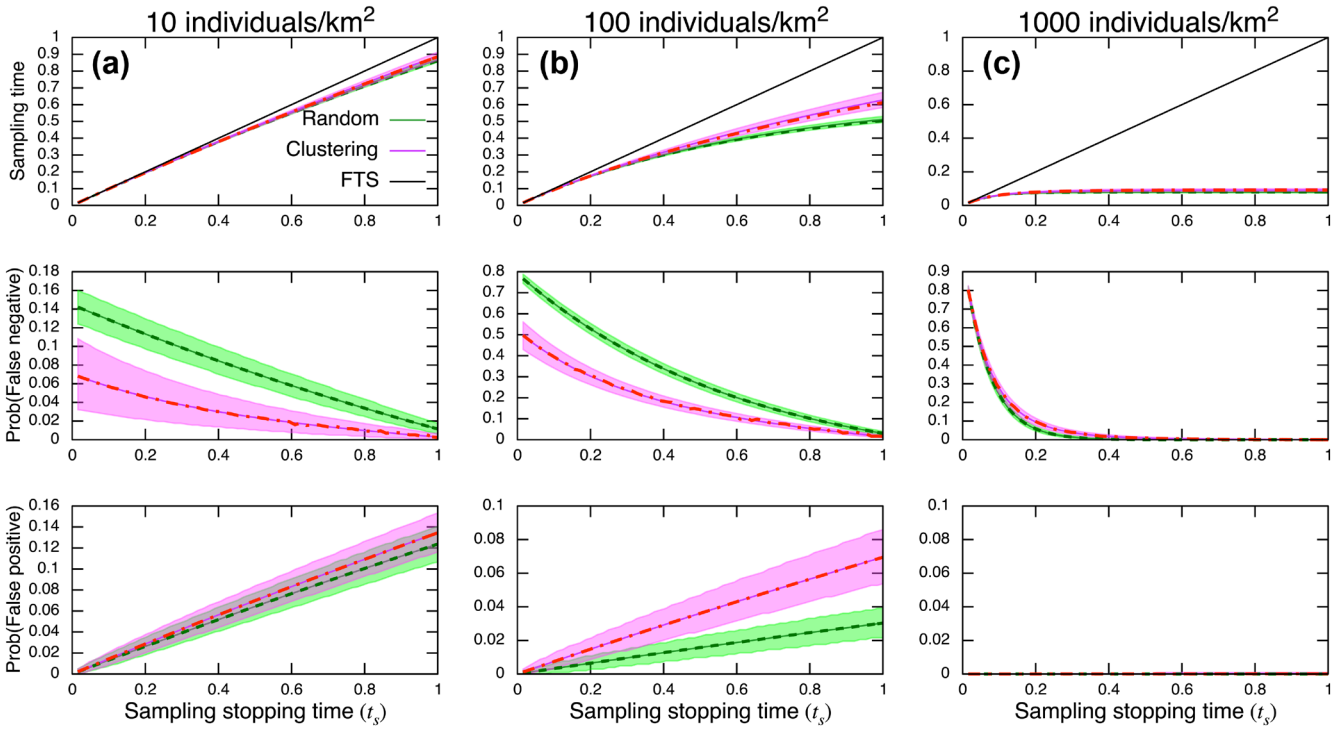

Figure 5. Normalized sampling time, false negative and false positive given normalized stopping time t_s with the probability of sampling error $\varepsilon = 0.1$. For each individual distribution scenarios, the numerical average (lines) and its theoretical value (dashed) are provided. The shaded area are between 5 and 95 percentiles of a 10^5-time numerical simulation. FTS represents the fixed-time survey. The scales of surveys are are $\nu(W) = 4 \times 4$ km, $\nu(M) = 2^{-3} \times 2^{-3}$ km, $\nu(S) = 2^{-6} \times 2^{-6}$ km. The intensity of false positive is $\lambda_{fp} = 10$. The parameters for the points generations are the same as in Fig. 3.

random distribution but with more aggregation at shorter distances, and *Cecropia obtusifolia* shows more aggregation than the Thomas process at shorter distance but less aggregation at a larger distances.

Expected sampling performance in BCI and decision making for sampling design

Once we define the sampling design based on definition above, it is straightforward to compute and quantify the sampling performance discussed in the preceding section (Fig. 7). The top panels of Fig. 7 shows one of realized presence–absence map when the sampling stopping time is $t_s = 0.78$.

How can we utilize these theoretical predictions? To consider this, first let us discuss the situation where we employ a fixed-time sampling. For example, when the time taken to sample a unit area $\nu(S) = 4.05 \times 4.05$ m is, on average, 30 s, then the expected time of the sampling is proportional to the number of the total sampling units (1024×64); 546 h.

Next, we employ the sampling methods developed. Let us assume that we have a budgetary constraint that limits the maximum allowable sampling time. For example, let us define the maximum allowable sampling time to be 350 h. Then, our theoretical predictions provide a sampling stopping time, t_s, that enables us to produce a presence–absence map within this time limit and an expected sampling error in the map. Such a sampling stopping time is obtained from Fig. 7a–b, where the sampling stopping time gives the sampling time 0.64 (approx. 350/546). This sampling time is realized when the sampling stopping time is about $t_s = 0.78$ (23.4 s) in Fig. 7a and $t_s = 0.73$ (21.9 s) in Fig. 7b. Using these sampling stopping time, we can obtain the expected error contents in the realized map from Fig. 7c–d: 0.22 in Fig. 7c and 0.28 in Fig. 7d, respectively. Our framework does not include the transition time from one mapping unit to another, so, in practice, the sampling time should be taken smaller than 0.64 not to exceed the total allowable sampling time.

It is also possible to set a required map accuracy first and then calculate the sampling time, by following our discussion above backwards.

Discussion

The capacity for ecosystem surveys is rapidly growing worldwide, but can we get more out of our efforts and limited resources with smarter sampling strategies? We developed a theoretical framework for designing surveys with known error properties using as an example the case of a moving observer (such as field biologist, citizen scientist or UAV) attempting to estimate a presence–absence map. This allows the calculation of the expected total sampling time and probability of false negative given a set parameters. Therefore, this survey protocol enables us to estimate the central trade-offs of

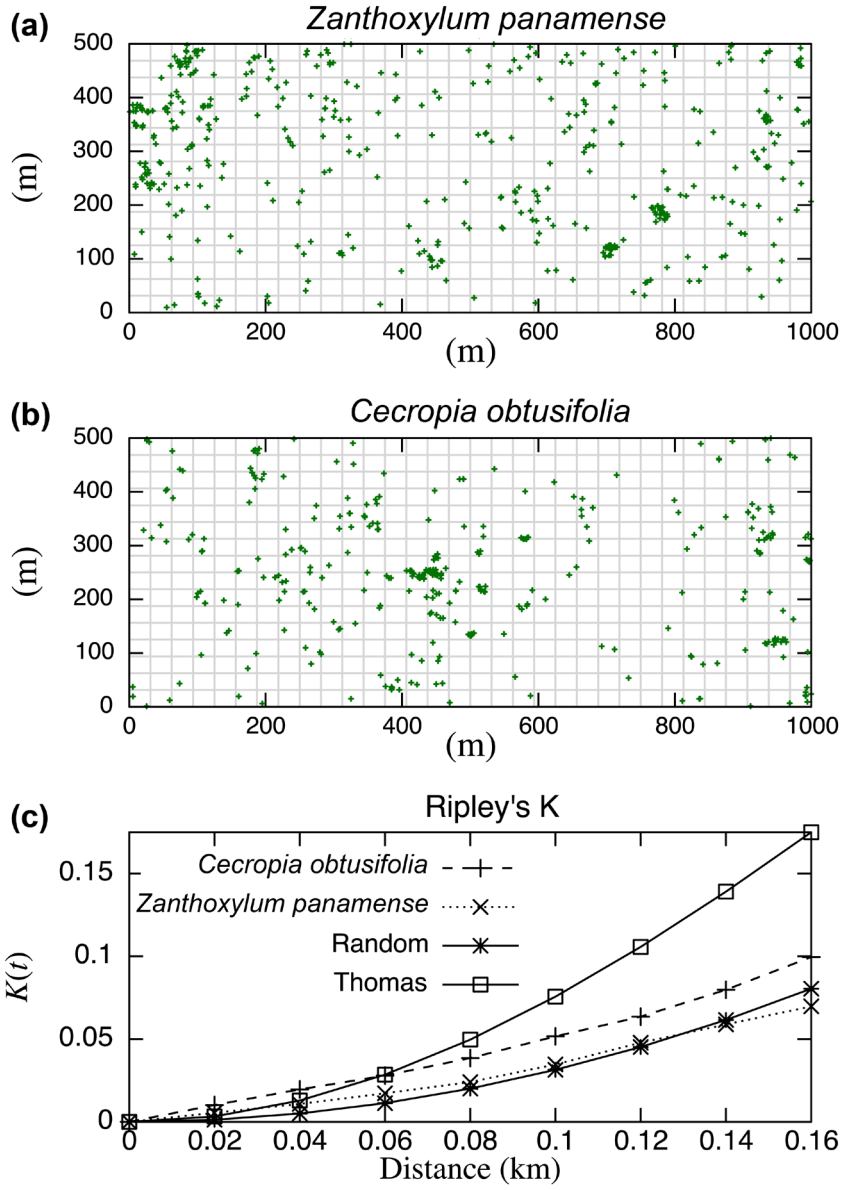

Figure 6. Distribution of (a) *Zanthoxylum panamense* (n = 463) and (b) *Cecropia obtusifolia* (n = 362). Subdivided regions represents the mapping unit used in this example (32.05 × 32.05 m). (c) Ripley's *K* function (Eq. 7) in the 50-ha plot at Barro Colorado Island.

existing ecosystem surveys: feasibility/cost and the accuracy of survey results. In particular, we showed that the performance can be specified by a set of multiple probabilities (Eq. 1a, 1b), and hence, we can discuss the effect of changing the ability of observer (e.g. detectability) or sampling design (e.g. mapping resolution) on the performance. Additionally, the framework provides the actual and sampled occupancy–area curves across mapping resolutions that enables us to estimate the sampling artifact. We also demonstrated the development of sampling design, the sampling performance and its interpretation for sampling decision using computational survey with example plant species distribution data from Barro Colorado Island.

In the application to simulated species distribution data, and empirical BCI data, we demonstrated the design of sampling and its decision-making process based on the theoretical prediction. In the example, we discussed a decision-making process to optimize the sampling accuracy under a budgetary constraint that limits the total sampling time. Figure 8 summarizes such a decision-making process to achieve an effective survey design given a maximum-possible sampling time. First, we estimate the number of hours that we can afford to carry out the survey from the budget available. We then quantify the observer's ability, design the survey protocol and set the required map accuracy in terms of the probability of false negative. Given a set of parameters, we can calculate the number of hours and map accuracy using the proposed framework. If the survey is not feasible given the budgetary constraint or does not meet required survey accuracy, we redefine the survey parameters and protocols

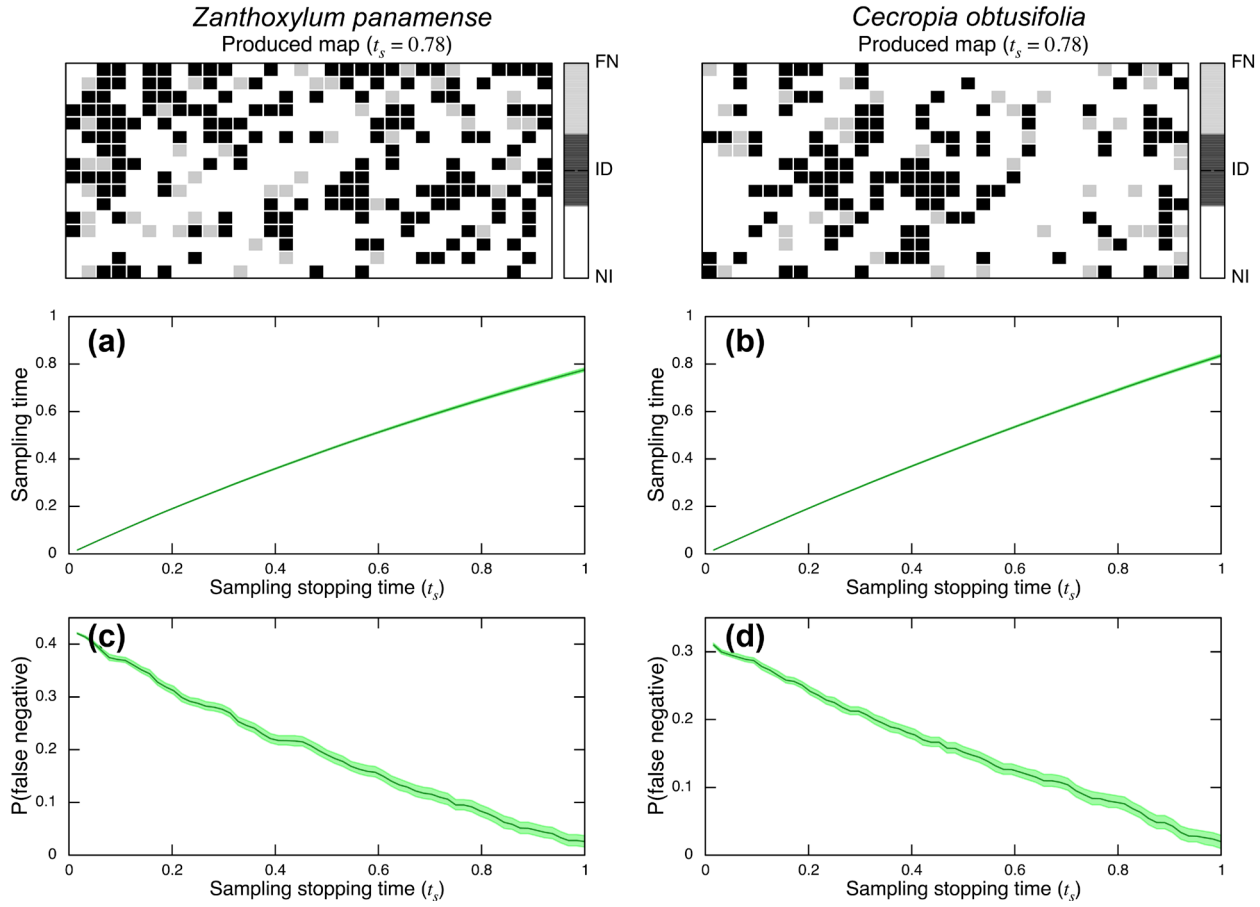

Figure 7. Expected sampling outcomes of two plant species (left) *Zanthoxylum panamense* and (right) *Cecropia obtusifolia* in the presented sampling method in 50-ha plot at Barro Colorado Island. Top panels show that one realized presence–absence map with sampling stopping time $t_s = 0.78$. The legends on the colorbar represents FN: false negative; ID: individual detection; NI: no individual exists. The second (a, b) and third (c, d) panels show estimated sampling time, and map accuracy given a sampling stopping time, t_s, respectivley, calculated by 10^4-time numerical simulations. The shaded areas are between 5 and 95 percentile in the simulations.

such as the sampling speed, mapping resolution, M and even required map accuracy and repeat until these predetermined requirements will be met. We can also use this workflow to find an optimal survey protocols once we define the objective function. For example, when our objective is to find the finest map resolution given a budgetary constraint and required map accuracy, then the optimal sampling protocol given the constraint can be found by repeating the same process as above, with the two conditions $t_{\text{samp}} < T$ and $P(\text{false negative}) \leq p$ satisfied, and find the smallest possible map resolution.

It is intuitive that false positive detection reduces total sampling time as it ends the sampling in a current mapping unit and starts a sampling in a new mapping unit. Also, it reduces the probability of false negative as false negative no longer occurs once a false detection is realized in a mapping unit. Our results suggest that the effects of false positive detection becomes significant in a survey with a larger sampling stopping time t_s and the smaller population size. Although integration of false positive detection is mathematically straightforward (Supplementary material Appendix D), modelling actual false positive detection is not simple: it may be caused by miss-classification of other species, objects, geography, etc. and be highly dependent on the nature of the observer. In the Supplementary material Appendix D, we applied to a simple phenomenological approach where we assume the number of false positive detection follows a probability distribution provided a sampling area. Yet simple, this can be a powerful method to consider false positive detection in our survey framework. Investigating relationship between the surveyed area and the number of false detection enables us to heuristically estimate the above-mentioned probability distribution.

In a situation where we know nothing about the abundance and spatial distribution of a species beforehand, it would not be possible to use theory to calculate an optimal design. However, this could be mitigated by a strategy to conduct a pilot sampling to estimate spatial parameters (Takashina et al. 2018b), or use a surrogate species and expert opinion (commonly used in ecosystem management) to inform the model (Caro and O'Doherty 1999, Martin et al. 2005). Also, the data accuracy can be estimated a posteriori once species data becomes available, and accumulating data may offer a better

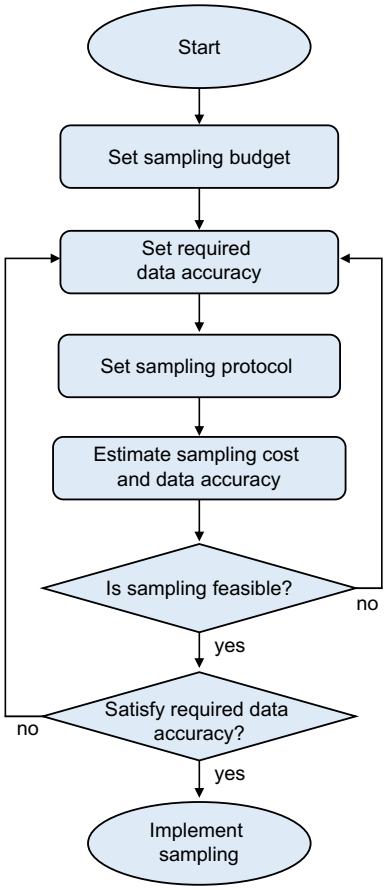

Figure 8. Flowchart of proposed survey design to achieve an effective sampling given a budgetary constraint.

estimate provided the detailed survey method. Hence, providing details of the sampling protocol used, if any, along with data will have benefits for future surveys.

This theoretical exercise can also be useful for developing general survey framework to obtain presence–absence maps. Discussion under a common framework, definitions and knowledge facilitates our understanding about sampling method and it helps to make a better management decision making (McDonald-Madden et al. 2010). For example, the IUCN Red List is one of the well-developed and widely-applied frameworks of quantitatively assessing the status of endangered species (IUCN 2017) where four criteria, generality, precision, realism and simplicity are thought to be key components (Keith et al. 2015). The IUCN Red List guides to assign the risk of extinction of species to the categories (e.g. Critically Endangered, Endangered or Vulnerable) based on quantitative criteria such as the reduction of the population and the geographic range (IUCN 2017). The assessment protocol first published in 1994 (IUCN 2017). Since then it has been under active development over e.g. criteria itself and terminologies used (Keith et al. 2013, 2015, IUCN 2017, Bland et al. 2018), and it has adopted to many countries and multiple ecosystem types (Nicholson et al. 2009). Our survey theory is developed in alignment with the need of such a common, general and quantitative framework. Seeking an opportunity to develop and describe a survey methods under the common terminologies and definitions would improve the overall ecological data. Moreover, discussion under the common framework may promote understanding and performance of the protocol itself. For example, there are various sampling trajectories within each mapping unit possible, such as spatially balanced sampling (Stevens and Olsen 2004, Robertson et al. 2013, Curran et al. 2020) and clustered sampling (Takashina et al. 2018b). These can be incorporated into our model, and these exercises will further promote development of a better sampling scheme but within a common framework. Also, sampling error is an emergent property of interactions of complex mechanisms, such as experience and the sampling speed. Establishing a common framework provides well-defined sampling problems that can boost future researches to improve the sampling protocol.

Sampling autonomously via drones is an emerging approach to ecological surveys (Crutsinger et al. 2016). One advantage of such devices, and why we feature them prominently in this paper, is the potential to program them with behavioral algorithms more easily than human behavior can be enforced. In actuality, however this technology is still mostly in the future. Other than a simple routines such as a fixed-time survey (Fig. 3) combined with image-based methods (Booth et al. 2005, 2006, Booth and Cox 2008) or via a ground control system, it may not be feasible for automated UAVs to conduct the presented survey protocols since it requires the realtime pixel-data processing to detect individuals on the fly. However, given an accelerating speed of hardware development and even faster software innovation (Crutsinger et al. 2016), UAVs are likely to be applicable to the protocol presented and more complex protocols in the near future. In fact, UAVs have already been applied to ecological sampling as an alternative to field sampling (Kattenborn et al. 2019). However, regardless of the current capacity of the technology, a theoretical development of the ecological survey method is essential to drive the innovation of the ecology-specific software (Crutsinger et al. 2016), and this will eventually promote ecological data reliability.

Here, we developed a theoretical framework for ecological surveys motivated by the increasing ecological survey intensity and the advantages of benchmarking data accuracy. The theoretical treatment of such a basic survey method and quantifying the survey performance will be a basis for the further discussion and development of more complex and inclusive survey approaches. For example, there is a phase to develop the sampling design under the presented sampling protocol, and a phase for citizen scientists to receive training in survey method (Danielsen et al. 2014). Also, ecological data is often subjected to further process (e.g. data input and classification, and image analysis for pixel data (Booth and Cox 2008, Ancin-Murguzur et al. 2019)) to utilize as input data in various research questions. These factors may reduce the available sampling time. Considering these multiple steps as a single framework will be necessary and it reduces the uncertainty of ecological sampling programs.

Data accessibility

The dataset we used is available at: <https://repository.si.edu/handle/10088/784>. We used the 2005 census data found in bci5.txt and species is labeled ZANTP1 and CECROB, respectively.

Acknowledgements – We would like to thank B. Kusumoto for his thoughtful comments on the manuscript. We are also grateful to Kenneth Dudley and three anonymous reviewers for their valuable comments.
Funding – NT was supported by Grant-in-Aid for the Japan Society for the Promotion of Science (JSPS) Fellows. Financial support was provided by JSPS (no. 17K15180 to EPE). EPE and NT were additionally supported by subsidy funding to OIST.
Conflicts of interest – We have no conflict of interest to declare.

References

Aguzzi, J. et al. 2011. The new seafloor observatory (OBSEA) for remote and long-term coastal ecosystem monitoring. – Sensors 11: 5850–5872.
Ancin-Murguzur, F. J. et al. 2019. Efficient sampling for ecosystem service supply assessment at a landscape scale. – Ecosyst. People 15: 33–41.
Anderson, K. and Gaston, K. J. 2013. Lightweight unmanned aerial vehicles will revolutionize spatial ecology. – Front. Ecol. Environ. 11: 138–146.
Azaele, S. et al. 2012. Downscaling species occupancy from coarse spatial scales. – Ecol. Appl. 22: 1004–1014.
Baxter, P. W. and Hamilton, G. 2018. Learning to fly: integrating spatial ecology with unmanned aerial vehicle surveys. – Ecosphere 9: e02194.
Bennett, J. R. et al. 2018. When to monitor and when to act: value of information theory for multiple management units and limited budgets. – J. Appl. Ecol. 55: 2102–2113.
Bland, I. M. et al. 2018. Developing a standardized definition of ecosystem collapse for risk assessment. – Front. Ecol. Environ. 16: 29–36.
Bland, L. M. et al. 2015. Predicting the conservation status of data-deficient species. – Conserv. Biol. 29: 250–259.
Böhm, M. et al. 2013. The conservation status of the world's reptiles. – Biol. Conserv. 157: 372–385.
Bonney, R. et al. 2014. Next steps for citizen science. – Science 343: 1436–1437.
Booth, D. T. and Cox, S. E. 2008. Image-based monitoring to measure ecological change in rangeland. – Front. Ecol. Environ. 6: 185–190.
Booth, D. T. et al. 2005. Image analysis compared with other methods for measuring ground cover. – Arid Land Res. Manage. 19: 91–100.
Booth, D. T. et al. 2006. Point sampling digital imagery with 'Samplepoint'. – Environ. Monit. Assess. 123: 97–108.
Burton, A. C. et al. 2015. Wildlife camera trapping: a review and recommendations for linking surveys to ecological processes . – J. Appl. Ecol. 52: 675–685.
Caro, T. M. and O'Doherty, G. 1999. On the use of surrogate species in conservation biology. – Conserv. Biol. 13: 805–814.
Chades, I. et al. 2008. When to stop managing or surveying cryptic threatened species. – Proc. Natl Acad. Sci. USA 105: 13936–13940.
Chandler, M. et al. 2017. Contribution of citizen science towards international biodiversity monitoring. – Biol. Conserv. 213: 280–294.
Chiu, S. N. et al. 2013. Stochastic geometry and its applications. – Wiley.
Coddington, I. A. et al. 1991. Designing and testing sampling protocols to estimate biodiversity in tropical ecosystems. – In: Dudley, E. C. (ed.), Unity Evol. Biol. Proc. Fourth Int. Congr. Syst. Evol. Biol.,Dioscorides Press, pp. 44–60.
Crutsinger, G. M. et al. 2016. The future of UAVs in ecology: an insider perspective from the Silicon Valley drone industry. – J. Unmanned Veh. Syst. 4: 161–168.
Cummings, A. R. et al. 2017. Developing a UAV-based monitoring program with indigenous peoples. – J. Unmanned Veh. Syst. 5: 115–125.
Curran, J. F. et al. 2020. Combining spatially balanced sampling, route optimisation and remote sensing to assess biodiversity response to reclamation practices on semi-arid well pads. – Biodiversity doi: 10.1080/14888386.2020.1733085
Danielsen, F. et al. 2014 A multicountry assessment of tropical resource monitoring by local communities. – Bioscience 64: 236–251.
Di Marco, M. et al. 2019. Wilderness areas halve the extinction risk of terrestrial biodiversity. – Nature 573: 582–585.
Dickinson, J. L. et al. 2012. The current state of citizen science as a tool for ecological research and public engagement. – Front. Ecol. Environ. 10: 291–297.
Dulvy, N. K. et al. 2014. Extinction risk and conservation of the world's sharks and rays. – Elife 3: e00590.
Epanchin-Niell, R. S. et al. 2012. Optimal surveillance and eradication of invasive species in heterogeneous landscapes. – Ecol. Lett. 15: 803–812.
Field, S. A. et al. 2005. Optimizing allocation of monitoring efford under economic and observational constraints. – J. Wildl. Manage. 69: 473–482.
ForestGEO 2020. ForestGEO home page. – <https://forestgeo.si.edu> accessed 20 July 2020.
GBIF.org 2019. GBIF home page. – <www.gbif.org> accessed 1 March 2019.
IUCN 2017. Guidelines for using the IUCN Red List categories and criteria. Ver. 13. – Prepared by the Standards and Petitions Subcommittee, <www.iucnredlist.org/documents/RedListGuidel> accessed 1 March 2019.
Jones, D. T. and Eggleton, P. 2000. Sampling termite assemblages in tropical forests: testing a rapid biodiversity assessment protocol. – J. Appl. Ecol. 37: 191–203.
Kattenborn, T. et al. 2019. UAV data as alternative to field sampling to map woody invasive species based on combined Sentinel-1 and Sentinel-2 data. – Remote Sens. Environ. 227: 61–73.
Keith, D. A. et al. 2013. Scientific foundations for an IUCN Red List of ecosystems. – PLoS One 8: e62111.
Keith, D. A. et al. 2015. The IUCN Red List of ecosystems: motivations, challenges and applications. – Conserv. Lett. 8: 214–226.
Kerans, J. R. et al. 1992. Aquatic invertebrate assemblages: spatial and temporal differences among sampling protocols. – J. North Am. Benthol. Soc. 11: 377–390.
Kobori, H. et al. 2016. Citizen science: a new approach to advance ecology, education and conservation. – Ecol. Res. 31: 1–19.
Koleff, P. et al. 2003. Measuring beta diversity for presence–absence data. – J. Anim. Ecol. 72: 367–382.
Kunin, W. E. 1998. Extrapolating species abundance across spatial scales. – Science 281: 1513–1515.

Levin, S. A. 1992. The problem of pattern and scale in ecology. – Ecology 73: 1943–1967.
Mackenzie, D. I. and Royle, J. A. 2005. Designing occupancy studies: general advice and allocating survey effort. – J. Appl. Ecol. 42: 1105–1114.
Martin, T. G. et al. 2005. The power of expert opinion in ecological models using Bayesian methods: impact of grazing on birds. – Ecol. Appl. 15: 266–280.
May, F. et al. 2018. mobsim: an r package for the simulation and measurement of biodiversity across spatial scales. – Methods Ecol. Evol. 9: 1401–1408.
McDonald-Madden, E. et al. 2010. Monitoring does not always count. – Trends Ecol. Evol.25: 547–550.
Morlon, H. et al. 2008. A general framework for the distance-decay of similarity in ecological communities. – Ecol. Lett. 11: 904–917.
Nichols, J. D. and Williams, B. K. 2006. Monitoring for conservation. – Trends Ecol. Evol. 21: 668–673.
Nicholson, E. et al. 2009. Assessing the threat status of ecological communities. – Conserv. Biol. 23: 259–274.
Orme, C. D. L. et al. 2006. Global patterns of geographic range size in birds. – PLoS Biol. 4: e208.
Pimm, S. L. et al. 2014. The biodiversity of species and their rates of extinction, distribution and protection. – Science 344: 1246752.
Plotkin, J. B. et al. 2000. Species–area curves, spatial aggregation and habitat specialization in tropical forests. – J. Theor. Biol. 207: 81–99.
Reef Life Survey 2019. Survey methods. – <https://reeflifesurvey.com/methods> accessed 12 December 2019.
Ripley, B. D. 1977. Modelling spatial patterns. – J. R. Stat. Soc. Ser. B 39: 172–212.
Robertson, B. L. et al. 2013. BAS: balanced acceptance sampling of natural resources. – Biometrics 69: 776–784.
Sanders, N. J. et al. 2003. Community disassembly by an invasive species. – Proc. Natl Acad. Sci. USA 100: 2474–2477.
Stem, C. et al. 2005. Monitoring and evaluation in conservation: a review of trends and approaches. – Conserv. Biol. 19: 295–309.
Stevens, D. L. and Olsen, A. R. 2004. Spatially balanced sampling of natural resources. – J. Am. Stat. Assoc. 99: 262–278.
Sullivan, B. L. et al. 2009. eBird: a citizen-based bird observation network in the biological sciences. – Biol. Conserv. 142: 2282–2292.
Takashina, N. et al. 2018a. A theory for ecological survey methods to map individual distributions. – Theor. Ecol. 11: 213–223.
Takashina, N. et al. 2018b. Spatially explicit approach to estimation of total population abundance in field surveys. – J. Theor. Biol. 453: 88–95.
Takashina, N. et al. 2019. A geometric approach to scaling individual distributions to macroecological patterns. – J. Theor. Biol. 461: 170–188.