

# Community distance sampling models allowing for imperfect detection and temporary emigration

YUICHI YAMAURA<sup>1,2,†</sup> AND J. ANDREW ROYLE<sup>3</sup>

<sup>1</sup>Department of Forest Vegetation, Forestry and Forest Products Research Institute, 1 Matsunosato, Tsukuba, Ibaraki 305-8687 Japan

<sup>2</sup>Fenner School of Environment and Society, Australian National University, Canberra, Australian Capital Territory 2601 Australia

<sup>3</sup>U.S. Geological Survey, Patuxent Wildlife Research Center, Laurel, Maryland 20708 USA

**Citation:** Yamaura, Y., and J. A. Royle. 2017. Community distance sampling models allowing for imperfect detection and temporary emigration. *Ecosphere* 8(12):e02028. 10.1002/ecs2.2028

**Abstract.** Recent developments of community abundance models (CAMs) enable us to analyze communities subject to imperfect detection. However, existing CAMs assume spatial closure, that is, that individuals are always present in the sampling plots, which is often violated in field surveys. Violation of this assumption, such as in the presence of spatial temporary emigration, can lead to the underestimates of detection probability and overestimates of population densities and diversity metrics. Here, we propose a model that simultaneously accommodates both temporary emigration and imperfect detection by integrating CAMs and a form of hierarchical distance sampling for open populations. Expected values of species richness are obtained via the summation of occupancy (or incidence) probabilities, based on species-level densities, across all species of the community. Simulations were used to examine the effects of spatial temporary emigration on the estimation of biological communities. We also applied the proposed model to empirical data and constructed area-based rarefaction curves accounting for temporary emigration. Simulation experiments showed that temporary emigration can decrease the local species richness ( $\alpha$  diversity) based on densities and increase the species turnover ( $\beta$  diversity). Raw species counts can overestimate or underestimate  $\alpha$  diversity in the presence of temporary emigration, but the specific biases depend on the values of detection and emigration probabilities. Our newly proposed model yielded unbiased estimates of  $\alpha$ ,  $\beta$ , and  $\gamma$  diversity in the presence of temporary emigration. The application to empirical data suggested that accounting for temporary emigration lowered area-based rarefaction curves because availability probabilities of individual species were estimated to be  $<1$ . Temporary emigration prevails in field surveys and has broad significance for understanding the ecology and function of biological communities and separation of imperfect detection and temporary emigration resolves long-standing issues in the use of count data. We therefore suggest that the consideration of temporary emigration would contribute to understanding the nature and role of biological communities.

**Key words:** additive partitioning; alpha ( $\alpha$ ) diversity; area-based rarefaction; beta ( $\beta$ ) diversity; closure assumption; gamma ( $\gamma$ ) diversity; hierarchical community model.

**Received** 30 July 2017; revised 6 October 2017; accepted 11 October 2017. Corresponding Editor: Lucas Joppa.

**Copyright:** © 2017 Yamaura and Royle. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

† **E-mail:** yamaura@ffpri.affrc.go.jp

## INTRODUCTION

Species richness is defined as the number of species in a specified area, and it is the simplest and most basic measure of species diversity (Whittaker 1975). Species richness varies in space

and time (Rosenzweig 1995), and its drivers are of primary interest in ecology (Huston 1994). As humans have transformed  $>75\%$  of Earth's ice-free land surface (Ellis and Ramankutty 2008) and appropriated  $>23\%$  of its terrestrial net primary productivity (Haberl et al. 2007), global

species diversity is now greatly threatened (Barnosky et al. 2011). Anthropogenic drivers of species richness or human-induced changes in species richness are therefore of prime interest in modern ecology (Newbold et al. 2016).

In this context, species richness has often been compared among sampling plots (MacArthur and MacArthur 1961, Huston 2014). One difficulty in doing so is the non-linear relationship between species richness and area (Connor and McCoy 1979), and it is therefore challenging to compare species richness among studies having sampling plots of different sizes and also even among sampling plots within the same study (Wiens 1989, Gotelli and Colwell 2001). For example, when comparing species richness among different land uses, how to control for differences in plot size is a major issue (cf. Newbold et al. 2015).

Recently, we have proposed a community modeling framework based on individual species-level abundance models (Yamaura et al. 2011, 2012). This community abundance modeling framework assumes that local abundance follows a Poisson distribution with expected abundance ( $\lambda$ ), and the probability that at least one individual of a species occurs is related directly to  $\lambda$  (Royle and Dorazio 2008):  $\Pr[N \geq 1] = 1 - \exp(-\lambda)$ . Under this model, we can link this expected abundance  $\lambda$  to the log-transformed size (area) of sampling plot  $j$  ( $A_j$ ) using the log link (Connor et al. 1997):  $\log(\lambda_j) = \beta_0 + 1 \times \log(A_j)$ . Here,  $\beta_0$  is the logarithmically transformed expected abundance when plot size is 1 ( $\lambda_j = \exp[\beta_0]$ , i.e., population density). By fixing the coefficient of a logarithmically transformed size at 1, we assume that population densities are constant irrespective of the area. However, we can relax this assumption using a free parameter  $\beta_1$  not constrained to be equal to 1 (i.e.,  $\beta_1 \times \log[A_j]$ ). We can also include other environmental covariates and associated coefficients into this linear predictor of  $\log(\lambda_j)$  (e.g.,  $\beta_2 \times x_2$ ). By expanding these equations into  $R$  species, we can therefore link expected species richness and area (Yamaura et al. 2016a, b) according to:

$$E[\text{Species richness}_j] = \sum_{i=1}^R [1 - \exp(-\lambda_{ij})] \quad (1)$$

where  $\lambda_{ij} = \exp(\beta_{0i} + 1 \times \log[A_j])$  is the expected abundance of species  $i$  in sampling plot  $j$  having size  $A_j$ . In the presence of imperfect detection of

individuals and species, the relevant parameters dictating  $\lambda_{ij}$  are estimated from repeated counts or capture–recapture data of detected species (Yamaura et al. 2016a, b). We again note that it is assumed that population densities of species constituting communities are constant irrespective of the area (by notation of  $1 \times \log[A_j]$ ) for simplicity (James and Wamer 1982, but see Yamaura et al. 2016a). This formulation allows us to construct area-based rarefaction curves (i.e., changes in species richness with area) given the scale-free parameters,  $\lambda_{ij}[A_j = 1] = \exp(\beta_{0i})$ : the population densities of individual species.

This development suggests the importance of accounting for variation in density among species comprising the community. Using the number of detected individuals (count data), our proposed community abundance models (CAMs) yield density estimates of individual species and species richness accounting for imperfect detection (Yamaura et al. 2011, 2012, 2016b), which is one of the major sources biasing density estimates (Kéry and Royle 2016). However, there is another major source of bias for mobile organisms: temporary emigration, which is due, in part, to movement of individuals off of the survey plot being sampled. Nichols et al. (2009) decomposed the detection process in a broad sense into three parts (Fig. 1). Given that home range of individuals only partially overlaps the sampling plots, individuals can be unavailable for sampling because they are out of the plot when the survey is conducted (Fig. 2). Following Nichols et al. (2009), we refer to the probability of individuals residing in the plots as the presence probability,  $p_p$ , and  $1 - p_p$  is the probability of temporary emigration (Table 1). This temporary emigration process due to movement of individuals is called spatial temporary emigration (Kéry and Royle 2016), and we say that individuals are spatially available when they are present in the sampling plot. Second, given that an individual is spatially available to be sampled (i.e., the individual is on the sampling plot), individuals can again be unavailable for the detection since they may not produce detection signals (cues). For example, birds may be sitting on the nest, and be silent; marine mammals may be below the water surface. This process which is unrelated to individuals moving about their home range is a random temporary emigration (Kéry and Royle 2016) and the associated probability can be denoted by

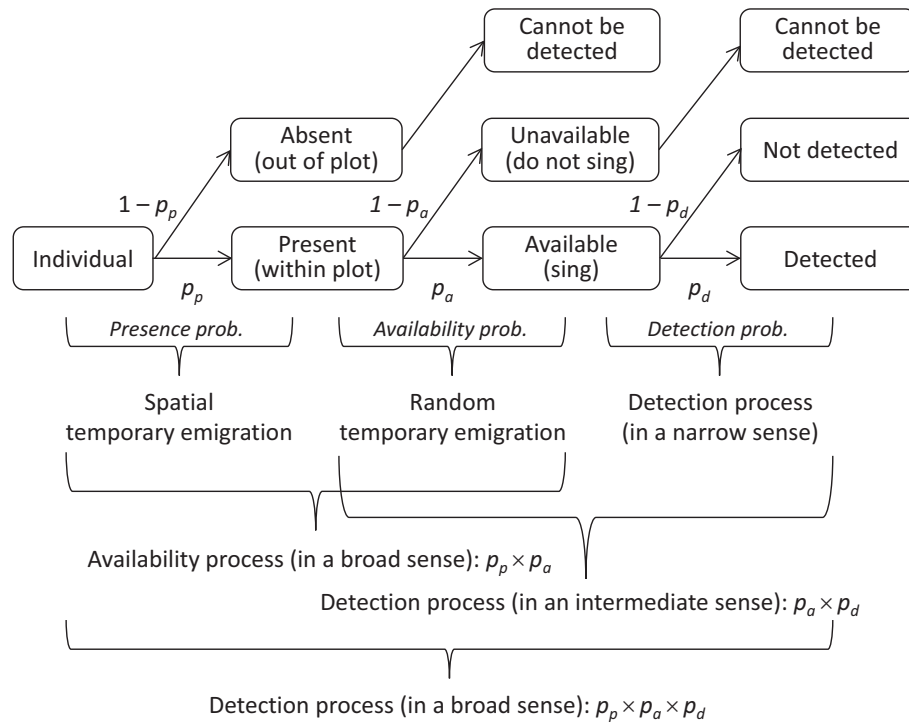


Fig. 1. Decomposition of processes associated with detecting individuals during field surveys. An individual whose home range or territory overlaps with the sampling plot is detected by the surveyor via three steps (Nichols et al. 2009). Individual can be available for detection with probability  $p_d$  given that it is in the plot and produces signals (song in an acoustic survey). These two processes required for the individual detection (spatial and random temporary emigration: Kéry and Royle 2016) are denoted by the probabilities of  $p_p$  and  $p_a$ , and we call their product “availability” in a broad sense ( $p_p \times p_a = \phi$ ). Which parameters of the detection probabilities ( $p_d$  or  $p_a \times p_d$  or  $p_p \times p_a \times p_d$ ) are associated with the field survey depends on the survey methods employed.

$1 - p_a$  (Nichols et al. 2009) where  $p_a$  is the probability of availability related to signaling (Table 1). We call the combined process of spatial and random temporary emigrations as an availability process in a broad sense (Fig. 1), and the corresponding probability is  $\phi (= p_p \times p_a)$ . We also call its complement ( $1 - \phi$ ) as temporary emigration in a broad sense. Finally, an individual that is both present on the plot and available for detection through signals can be detected by the surveyor with probability  $p_d$  (Nichols et al. 2009). The problem of temporary emigration can propagate to inference about not only density but also occupancy (e.g., Tyre et al. 2003); in distance sampling, random temporary emigration is known to violate the assumption that all birds on the census line (i.e., zero distance) are perfectly detected (denoted by  $g[0] = 1$ ), and can cause substantial bias in density estimates (Buckland et al. 2004).

Here, we propose an approach to overcome this problem in a community context by integrating count-based CAMs and open hierarchical distance sampling (HDS) that allows for temporary emigration (Kéry and Royle 2016). Although our proposed model does not specify population dynamics, the open-population version of HDS allows the population size to change during the course of multiple surveys (Kéry and Royle 2016). Our model separately estimates detection probability ( $p_d$ ) and availability probability in a broad sense (denoted by  $\phi = p_p \times p_a$ ) and produces density estimates of individual species unbiased by temporary emigration. This is possible because detection probability  $p_d$  can be estimated by a single visit with distance sampling given the assumption of perfect detection at the zero distance. Multiple visits to individual sites allow for inference about availability probability.

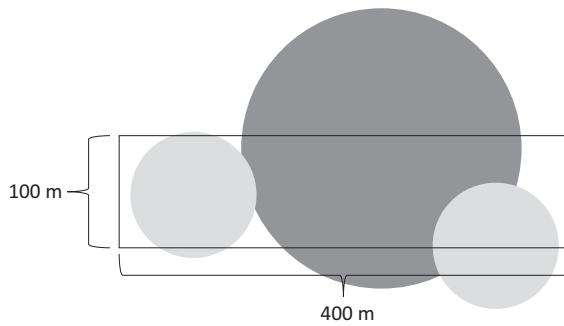


Fig. 2. Schematic illustration of partial availability of bird individuals during the survey. A rectangular sampling plot is assumed, and its size follows the field survey used by the model application. Territory size differs among the species; species like coal tit *Periparus aster* can have small territories (1 ha), some of which may be well incorporated by the plot (shown by the light-gray circles). But other species like woodpeckers can have larger territories only partially overlapping with the plot (dark gray), and they can be unavailable (out of the plot) when the survey is conducted.

We used simulation to show how temporary emigration affects the estimation of biological communities, including  $\alpha$ ,  $\beta$ , and  $\gamma$  diversity, and that our model can yield unbiased estimates. We then applied our model to empirical data and constructed area-based rarefaction curves accounting for temporary emigration. In the next section, we begin with the overview of the temporary emigration process and next develop the model.

## METHODS

### *Temporary emigration as a component of the detection process*

Among spatial and random temporary emigration processes, we are primarily interested in the spatial temporary emigration because this process is prevalent in studies on mobile organisms. For example, observed individuals whose territories

partially overlap with the sampling plots are frequently excluded from the analysis (e.g., Loman and von Schantz 1991), and James and Wamer (1982) excluded 23% (11/48) of the observed species in order to construct rarefaction curves since only fractions of their territories were included in the studied plots. Hanski and Haila (1988) showed that male individuals of chaffinch *Fringilla coelebs* can spend a long time out of their singing territories for foraging. Haila (1988) also discussed the problems of density calculation in small habitat fragments when birds can forage in the surrounding areas. Stratford and Robinson (2005) suggested that tropical birds have substantially larger territories and home ranges compared to temperate birds and thus should be more susceptible to biases induced by spatial temporary emigration.

Ideally, the three processes involved with imperfect detection should be explicitly addressed when estimating population densities and community structure. However, our ability to resolve all three components of the detection process depends on the protocols being employed (see also *Discussion*). If we assume  $p_a = 1$ , and when we use distance sampling on repeated sampling bouts, we can resolve spatial temporary emigration ( $p_p$ ) from probability of detection ( $p_d$ ). Our study is on this line of model development in a community context. However, in that case, it is not possible to explicitly estimate all three components. Rather, we can only estimate directly  $p_d$  and the product  $p_p \times p_a (= \phi$  in our model).

In the presence of spatial temporary emigration of mobile individuals, density can be defined as the instantaneous number of individuals residing in the area of interest (i.e., based on their snapshot location) divided by the respective area (Buckland et al. 2001, Royle and Dorazio 2008). If we do not separate  $p_p$  from  $p_d$ , then nominal detection probability from repeated counts includes  $p_{p'}$  and the number of individuals that

Table 1. Three quantities associated with detecting individuals.

Quantity	Symbol	Description	Complement ( $1 - p$ )
Presence probability	$p_p$	Probability that individual resides in the plot	Spatial temporary emigration
Availability probability	$p_a$	Probability that individual produces detection signal given $p_p$	Random temporary emigration
Detection probability	$p_d$	Probability that individual is detected by surveyor given presence and available	Imperfect detection (in a narrow sense)



can be present in the plots (“super-population size”) is estimated (Nichols et al. 2009). This leads to the overestimation of the population densities since the denominator calculating densities is not the size of sampling plots but the larger effective sampling area (Chandler et al. 2011). Hutto et al. (1986) therefore suggested that actual density cannot be calculated reliably due to the unknown effective sampling area in the fixed-radius point counts.

The concept of spatial temporary emigration is manifested as a closure assumption in models of abundance and occupancy estimation in which population size or occupancy status is assumed to be constant over the survey period (Royle 2004, MacKenzie et al. 2006). Spatial temporary emigration violates this assumption. To safely ignore the problem of temporary emigration for density estimation, sampling plots must be sufficiently large relative to the home range size of individuals (Royle and Dorazio 2008). However, in the context of sampling communities, home range size differs among species, and large plots suited for species with large home range sizes cannot usually be attained due to logistics.

#### *The model: open-population CAM*

We integrate CAM and open HDS to simultaneously deal with both imperfect detection and temporary emigration. We note that sampling methods other than distance sampling can be used such as removal sampling (Amundson et al. 2014), capture–recapture protocols (Yamaura et al. 2016a), and multiple counts in a single visit (Chandler et al. 2011); in any case, multiple visits to individual sampling plots are required to deal with temporary emigration. We develop our model using distance sampling because this method is a well-known method and commonly practiced in field studies. We can apply distance sampling when we measure and record the distance from the transect line or point to the individuals; nevertheless, distance sampling has been rarely used in CAMs (Sollmann et al. 2016). The model is composed of three hierarchical levels, and the first one is the process describing the super-population size of species  $i$  for site  $j$  ( $M_{ij}$ ), and we assume that  $M_{ij}$  follows a Poisson process (or negative binomial distribution; Sollmann et al. 2016):

$$M_{ij} \sim \text{Poisson}(\lambda_{ij}) \quad (2)$$

where  $\lambda_{ij}$  is an expected value and will be represented by a log-linear predictor as a function of relevant covariates. Conceptually, this quantity  $M_{ij}$  is the population size of species  $i$  that are ever exposed to sampling at site  $j$ , that is, which have at least some part of their home range within the sampled plot. We can include random effects in the predictor to consider variation not explained by the Poisson distribution (Yamaura et al. 2012).

The next level deals with the emigration process, and we assume that the number of individuals exposed to sampling (present in the sampling plots and signaling presence) during visit  $k$  ( $N_{ijk}$ ) follows a binomial process with  $M_{ij}$  as the number of trials and success rate (availability probability in the broad sense)  $\phi_i (= p_p \times p_a)$ :

$$N_{ijk} \sim \text{Binomial}(M_{ij}, \phi_i). \quad (3)$$

Although we assume that every individual of the same species has constant  $\phi_i$ , this probability can vary among individuals depending on the location of home ranges relative to the sampling plots (Fig. 2). However, in the context of a single-species hierarchical model with the removal sampling, Chandler et al. (2011) showed that density estimates can be unbiased when  $\phi_i$  varies among individuals but is modeled as constant. Bias is expected to decrease with the number of sampling sites and visits, which is relevant with the concept pooling robustness (Burnham et al. 1980).

The third level of the hierarchical model is the detection process, describing the number of individuals detected at visit  $k$  in distance classes (up to  $D$ ),  $\mathbf{y}_{ijk} = (y_{ijk1}, y_{ijk2}, \dots, y_{ijkD})$ :

$$\mathbf{y}_{ijk} \sim \text{Multinomial}(N_{ijk}, \boldsymbol{\pi}_{ij}) \quad (4)$$

where  $\boldsymbol{\pi}_{ij}$  (cell probabilities of the multinomial distribution) describes the distribution of observed detections among distance classes (discrete distance classes are commonly used; Buckland et al. 2001). These detection and availability parameters can depend on covariates (Kéry and Royle 2016, Sollmann et al. 2016). We note that availability and detection probabilities are separately estimated given this data structure and model formulation. Variations in number of available individuals exposed to the detection ( $N_{ijk}$ ) among the visits provide information on the availability probability  $\phi_i$  (Eq. 3), and variations in number

of detected individuals among the distance classes ( $\mathbf{y}_{ijk}$ ) given the total number of detected individuals at the individual visits (expected to be  $\lambda_{ij} \times \phi_i \times \sum_{d=1}^D \pi_{ij}$ ) provide the information necessary to estimate detection probability (parameters describing  $\pi_{ij}$  [Eq. 4]). In other words, multiple visits to the sampling sites are required to estimate availability probability, while detection probability can be estimated via a single visit using distance sampling. Nevertheless, multiple visits are expected to increase the accuracy of parameter estimates of distance sampling models. We also note that detection probability of individual species is described by the scale parameters ( $\sigma_i$ ) of the detection model which dictates the rate of decay of detection probability as a function of distance (Buckland et al. 2001, Kéry and Royle 2016). We assume that detection probability is equal to 1 at the zero distance for individuals that are available for detection during any particular visit (see *Discussion* for the relaxation of this assumption).

For each parameter (e.g., intercepts in the linear predictor of  $\lambda_{ij}$ ), species-level parameters (suitably transformed) are governed by a single normal distribution with community-level hyper-parameters (mean and SD), and the existence of undetected species throughout the survey is considered by use of the method of data augmentation (Yamaura et al. 2016b).

### Simulation experiments

We conducted simulation experiments to examine the effects of imperfect detection and temporary emigration on community analysis with 30 sampling sites. There was a single covariate ( $x_j$ ) that can affect abundance of individual species, which ranged from  $-1$  to  $1$ , and their specific values were equally spaced among the sites (i.e.,  $-1.00, -0.93, \dots, 0.93, 1.00$ ). The regional species richness (the number of species comprising the regional meta-community,  $R$ ) was fixed at 40 across the 30 sampling sites, and these species had 1.0 mean abundance (super-population size) and 1.0 standard deviation ( $\exp[\mu_{\beta 0}] = \sigma_{\beta 0} = 1$ ), and mean slope values 1.0 for  $x_j$  in the linear predictor of  $\lambda_{ij}$  ( $\mu_{\beta 1} = \sigma_{\beta 1} = 1$ ). This situation indicates that species richness increases with  $x_j$ . We then addressed five cases with different mean values of detection probability represented by the scale parameter ( $\bar{\sigma}$ ) of the half-normal detection model

(Buckland et al. 2001, Kéry and Royle 2016) and availability probability ( $\bar{\phi}$ : Table 2). These five cases comprised the combinations of high and low detection and availability probabilities, and an intermediate case. Species-specific parameters of detection and availability probabilities ( $\sigma_i$  and  $\phi_i$ ) are log- and logit-transformed normal random variables. We then conducted virtual surveys with three visits to each site since three visits have been shown to produce good performance of the single-species and community  $N$ -mixture (closed population) models (Yamaura 2013, Yamaura et al. 2016b). We collected distance sampling data with rectangular plots up to 50 m maximum distance and 10 m width distance classes (five distance classes). Although we used rectangular plots as in our empirical data, we can adopt the circular plots (point transects) using the corresponding distributional function of distances (Buckland et al. 2001, Kéry and Royle 2016). The variations of scale parameters represent the situation in which detection probability quickly declines ( $\bar{\sigma} = 10$ ) and hardly declines ( $\bar{\sigma} = 100$ ) from zero to the maximum 50 m distance. These two scale parameters represent 0.25 ( $\bar{\sigma} = 10$ ) and 0.96 ( $\bar{\sigma} = 100$ ) plot-level detection probabilities (Table 2).

We applied the open-population CAM to the distance sampling data to infer the responses of community-level densities and species richness to the covariate. We augmented the observed data sets with  $m$  ( $= 80 - S$ ) potential species with zero count histories, where  $S$  was the number of detected species. We defined expected densities ( $\text{dens}_{ij}$ ) of the plots as  $\lambda_{ij} \times \phi_i \times w_i$

Table 2. Five simulation scenarios with different detection and availability probability.

Case	$\bar{\sigma}$	$\bar{p}_d$	$\bar{\phi}$
1	100	0.96	0.9
2	100	0.96	0.1
3	10	0.25	0.9
4	10	0.25	0.1
5	40	0.79	0.5

*Notes:* Values of 100 and 10 for indicate high and low detection probability, respectively. Plot-level detection probabilities  $\bar{p}_d$  corresponding to  $\bar{\sigma}$  are also shown. Values of 0.9 and 0.1 for  $\bar{\phi}$  indicate high and low availability probability, respectively. The last case represents the intermediate scenario. Species variations of  $\sigma$  and  $\phi$  were produced by normal distributions via log and logit transformations, respectively. We used values of 0.5 for the SD of  $\sigma$  and 1.0 for SD of  $\phi$  at log and logit scales, respectively.

where  $\lambda_{ij}$  is an expected value of super-population size, and  $w_i$ , the data augmentation variable, is a random variable that follows a Bernoulli distribution and indicates whether species  $i$  is included in the community (Yamaura et al. 2016b). For the detected species, this variable takes the value of 1 (with posterior probability 1.0) but can have the value of 0 for undetected species. We estimated site-specific community densities by summing  $\text{dens}_{ij}$  across the species. Although we did not divide these quantities by the corresponding plot size, we treated these as densities since they were abundances for a constant unit size. Density-based expected local species richness ( $\alpha$  diversity) was estimated by summing the occupancy probability ( $\psi_{ij}$ ) derived from the local densities across the species:

$$\psi_{ij} = 1 - \exp(-\text{dens}_{ij}). \quad (5)$$

We obtained the overall occupancy probability across 30 sites for individual species ( $\psi_i$ ) as the complement of the products of probability that a site is not occupied:

$$\psi_i = 1 - \prod_{j=1}^{30} (1 - \psi_{ij}). \quad (6)$$

The summation of  $\psi_i$  across the species is the expected overall species richness across the sampling plots ( $\gamma$  diversity), and we partitioned  $\gamma$  diversity into  $\beta$  and  $\alpha$  diversity by subtracting the mean expected  $\alpha$  diversity from  $\gamma$  diversity (additive partitioning; Crist et al. 2003).

To show the effects of making use of repeated distance sampling data and therefore considering temporary emigration, we reduced the distance sampling data to simple counts and applied the conventional  $N$ -mixture CAM that does not consider temporary emigration (Yamaura et al. 2012, 2016b). The  $N$ -mixture CAM thus serves as an explicit model for the super-population size of individuals ever available for sampling during the repeated surveys (Nichols et al. 2009, Chandler et al. 2011). This conventional CAM is based on simple repeated count data and is suggested to be a promising approach to dealing with community-level abundance (Iknayan et al. 2014). In the application of  $N$ -mixture CAM to the temporary emigration data, the number of detected individuals for species  $i$  on site  $j$  at visit  $k$  ( $y_{ijk}$ )

is  $y_{ijk} \sim \text{Binomial}(p_{\text{cap},i} N_{ij})$  where  $p_{\text{cap},i}$  is the probability of an individual being encountered, averaged over all distance classes. However, as described above, because the  $N$ -mixture model does not consider the emigration process (Eq. 3) and regards  $N_{ij}$  as a realization of the underlying Poisson process ( $N_{ij} \sim \text{Poisson}[\lambda_{ij}]$ ), the nominal detection probability ( $p_{\text{cap},i}$ ) is  $p_{\text{cap},i} \times \phi_i$ . Estimates of  $N_{ij}$  are therefore super-population sizes. It is again noted that the objective of the comparison is to provide an alternative analysis ( $N$ -mixture model), which is sensible when only total counts are obtained (i.e., absent distance data), for our proposed open-population community model. Another modeling option is to fit a basic distance sampling model based on the aggregate counts in each distance class, summed across the visits ( $\mathbf{y}_{ij} = \sum_k \mathbf{y}_{ijk}$ ). This quantity follows the multinomial distribution:  $\mathbf{y}_{ij} \sim \text{Multinomial}(K \times N_{ij}, \boldsymbol{\pi}_{ij})$ . Although we did not fit this model to our data, as in the  $N$ -mixture model, the nominal detection probability would include the emigration probability ( $\phi_i$ ) and also produce estimates of the super-population size.

We obtained the same estimates described above from this closed-population CAM except that densities were calculated as  $\lambda_{ij} \times w_i$  (availability probability [ $\phi_i$ ] was not considered). We compared the inferred responses of community densities and species richness to the covariate and  $\alpha$ ,  $\beta$ , and  $\gamma$  diversity from open- and closed-population CAMs (using posterior distributions) with those of naïve counts and known true (expected) values on closure and non-closure (open) assumptions. These true values were obtained by the known true values of  $\lambda_{ij}$  used to generate detection data with and without  $\phi_i$  for values on open and closure assumptions, respectively.

We fitted the models using Markov chain Monte Carlo (MCMC), using conventional uninformative priors, 10,000 burn-in, and 100,000 iterations, and with three chains. We truncated community- and species-level  $\sigma$  in the range of 2.7–403 to enhance the convergence. Model fitting was conducted using JAGS ver. 3.2.0 (Plummer 2012) with jagsUI ver. 1.3.7 (Kellner 2015) and R ver. 3.2.3 (R Core Team 2015). We also used vegan ver. 2.3-0 (Oksanen et al. 2015) for the additive partitioning of count data. We assessed the chain convergence with the Gelman-Rubin

statistic of all parameters ( $<1.1$ ), and conducted an additional 100,000 iterations until convergence was attained using the function `autojags` in `jagsUI`. We replicated the simulations 100 times for each of five cases with different detection probability and availability probabilities. We obtained median values of the posterior distributions, and their mean values were used to represent the data across the 100 replicates.

#### *Model application to the empirical data*

We applied the open-population CAM to distance sampling data from a survey of birds which we previously used to develop and fit the  $N$ -mixture (closed population) CAM (Yamaura et al. 2012). The study area ( $15 \times 15$  km) was in the northern part of Kitakami highland, Iwate prefecture, northern Japan ( $39^{\circ}50'$  N,  $141^{\circ}19'$  E). This area (200–1000 m asl) was covered by forests, which were dominated by deciduous natural forests and larch *Larix leptolepis* plantations. Original data were collected to examine the abundance and species richness of breeding bird communities in four open (pasture, meadow, young planted forest, and abandoned clear-cut) and two forested habitats (mature planted forest and natural old-growth). We established five sampling plots each of which was 4 ha in size ( $100 \times 400$  m; Fig. 1) for each habitat type (total of 30 plots). Sampling plots were spaced at least 600 m apart except for meadow plots. Line transect surveys were conducted by one person (Y.Y.) during 19 May–28 June 2009, and each plot was visited in the morning (sunrise to 10:00 a.m.) five times on different days. We recorded the number of individuals of each species in the plots and their distances from the transect line.

We fitted the open-population CAM to the data and derived the abundance in each habitat of size  $h$  (ha) accounting for temporary emigration as follows:  $\text{abund}_{ijh} = (\lambda_{ij} \times \phi_i \times w_i/4) \times \text{area}_h$  where  $\lambda_{ij}$  is an expected super-population size of species  $i$  for habitat  $j$  in the original plot size (4 ha), and the quantity in parentheses indicates the expected density (per ha) accounting for temporary emigration ( $\phi_i$ ). We substituted this term instead of  $\lambda_{ij}$  in Eq. 1, and obtained expected species richness for each of six habitats at 53 different plot sizes from 0.01 to 10 ha. Species richness of songbirds was expected to change across this range of plot size (Yamaura et al. 2016a). We then constructed the

area-based rarefaction curves for six habitat types based on these estimates. Following the framework of additive partitioning, we also obtained mean values of species richness across six habitats as  $\alpha$  diversity at respective plot sizes. Expected species richness across six habitats ( $\gamma$  diversity) was estimated using Eqs. 5, 6, in which we first estimated the probability that each species occupied at least one of the six habitats at respective sizes and then summed these probabilities across the species. We similarly constructed rarefaction curves from the closed-population CAM ignoring temporary emigration. The data were composed of 47 observed species, and we augmented with 30 potential species. We fitted the model with the same settings as in the simulations (e.g., numbers of burn-in and iterations, software), and convergence was determined for the relevant community-level parameters.

## RESULTS

#### *Analysis of simulated data*

Under the high detectability and availability scenario ( $\bar{\sigma} = 100$ ,  $\bar{\phi} = 0.9$ ), which can be considered as an ideal situation (nearly perfect detection and high availability), there were few differences among the true, estimated, and observed values of community-level densities and species richness against covariate ( $x$ ), and additive partitioning (Fig. 3a–c). Nevertheless, even with this high availability, relaxation of the closure assumption slightly decreased expected community densities and local species richness. Low availability ( $\bar{\phi} = 0.1$ ) greatly decreased true community densities and local species richness compared with estimates obtained under the closure assumption, but did not greatly decrease overall species richness ( $\gamma$  diversity);  $\beta$  diversity was therefore increased (Fig. 3d–f). The open-population CAM correctly recovered true values of community densities and species richness. The closed-population CAM estimated the appropriate super-population sizes (Fig. 3). Therefore, the presence of temporary emigration, when ignored, led to the overestimation and underestimation of  $\alpha$  and  $\beta$  diversity, respectively. The most information-poor scenario ( $\bar{\sigma} = 10$ ,  $\bar{\phi} = 0.1$ ) was challenging for the closed-population CAM, yielding very diffuse posterior distributions and hence wide CIs (Fig. 3j–l). Bias of raw counts of individuals (index of densities) and local species richness



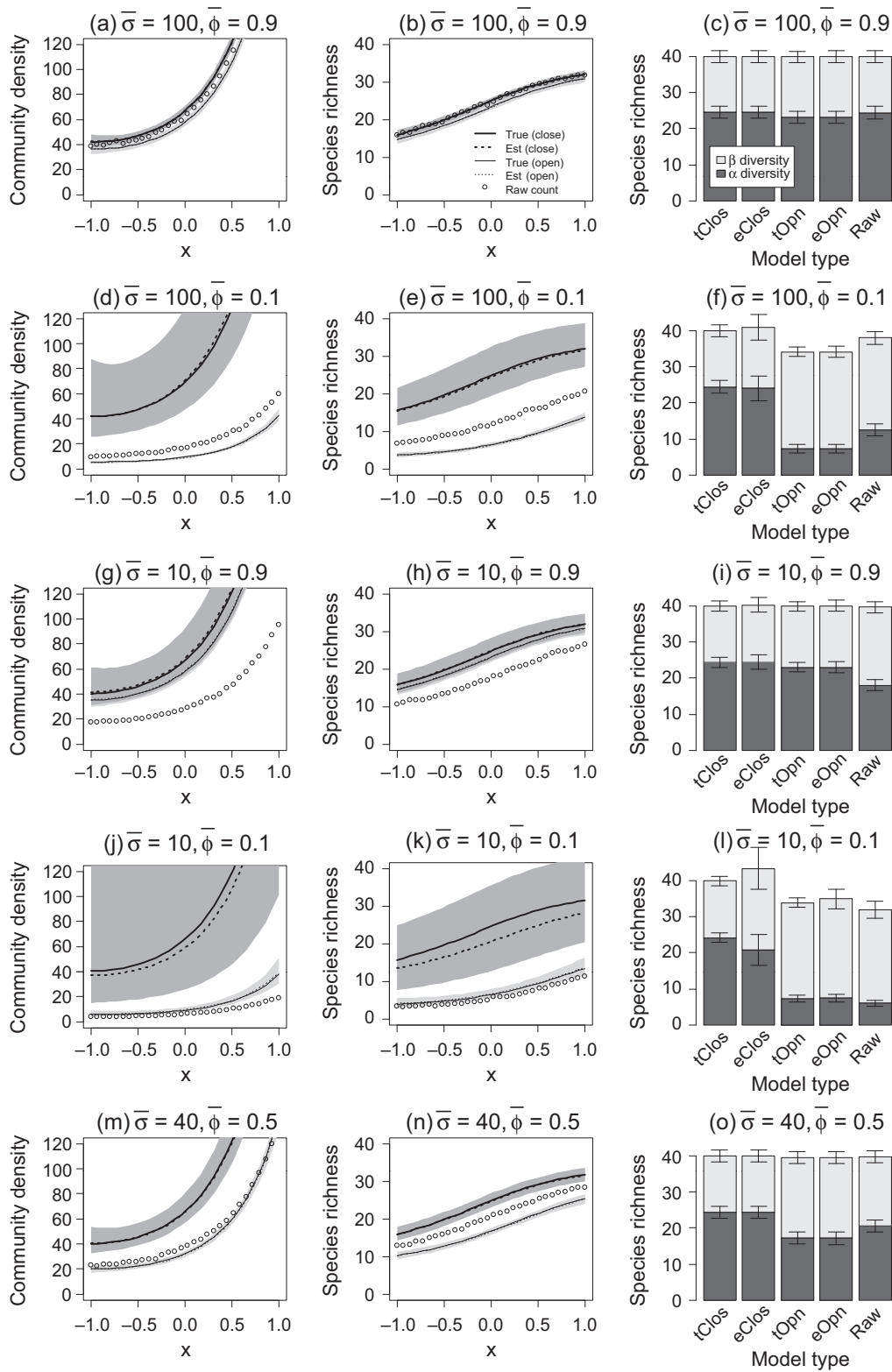


Fig. 3. True and estimated community-level densities and species richness for five cases subject to different

(Fig. 3. Continued)

detection and availability probabilities. Left column: True and estimated values of community densities are shown in relation to the single covariate,  $x_j$ . Central column: species richness and  $x_j$ . True values under closure and open assumptions are shown by solid lines using the expected values, while estimates are shown by the median (dotted or broken lines) and 95% CIs (dark-gray and light-gray colors for closed and open populations, respectively). Raw count indicates the number of detected individuals (maximum numbers over the three visits were summed across the species) and species during the three replicated visits at each site. Mean values are used to represent the data across the 100 replicates. Right column: partitioning of  $\gamma$  diversity (overall species richness) across 30 sampling sites into  $\alpha$  (mean species richness) and  $\beta$  diversity (differences between  $\gamma$  and  $\alpha$  diversity). tClos and tOpn indicate the true values of expected (true) species richness based on the closed- and open-population assumptions. eClos and eOpn indicate estimates of closed- and open-population community abundance models, respectively. Vertical bars indicate SDs of through the 100 replications.

depended on the situation; high detectability and low availability ( $\bar{\sigma} = 100$ ;  $\bar{\phi} = 0.1$ ) made raw counts higher than true values of community densities and species richness under the open community data-generating model (Fig. 3d–f). Low detectability and high availability ( $\bar{\sigma} = 10$ ;  $\bar{\phi} = 0.9$ ) instead made raw counts lower than true values of community densities and species richness (Fig. 3g–i). In the most information-poor scenario with low detectability and availability ( $\bar{\sigma} = 10$ ,  $\bar{\phi} = 0.1$ ), and the intermediate scenario ( $\bar{\sigma} = 40$ ,  $\bar{\phi} = 0.5$ ), the biases of raw counts of local species richness were smaller than those of the above scenarios (Fig. 3j–o).

#### Analysis of empirical data

Estimates of community- and species-level scale parameters ( $\sigma$ ) of the distance sampling CAM

reached the upper limits ( $>300$ ). This means that detection probabilities were estimated to be nearly 1 within the plots for all species, indicating that all the variations in counts among the visits were caused by temporary emigration. This was a result of the detection frequency not declining with distance, and the assumption of perfect detection at the zero distance. The estimate of community-level mean availability was 0.13 and species-level estimates ranged from 0.02 (blue-and-white flycatcher *Cyanoptila cyanomelana*) to 0.86 (common skylark *Alauda arvensis*); these values were almost the same as detection probability estimates obtained from the closed-population CAM.

Estimates of super-population size varied across six habitats, and this variation did not have clear associations with availability estimates (Fig. 4a), that is, every species had small super-

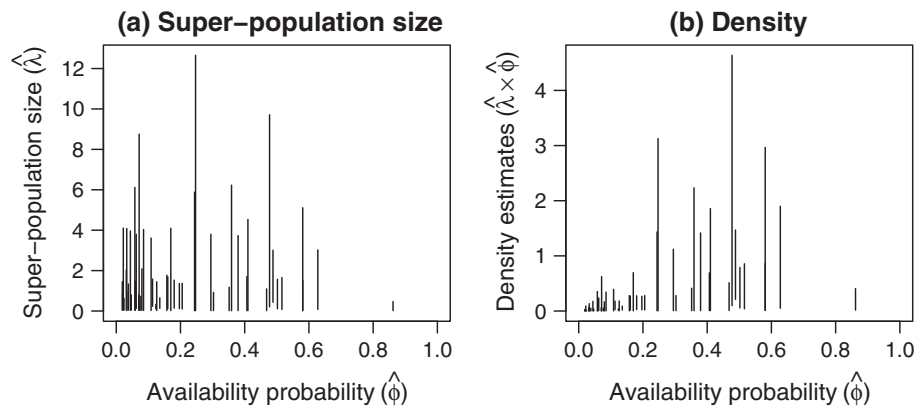


Fig. 4. Estimates of super-population size and density for individual species in relation to availability probability. Individual vertical lines indicate ranges (maximum and minimum values) of (a) super-population size and (b) density estimates across six habitat types for individual species, and they were shown against individual species' availability estimates. Estimates of super-population size were obtained by dividing the density estimates (posterior medians) by availability estimates.

population sizes in certain habitat types ( $\sim 0$ ) and high values in other habitats. However, since low availabilities made maximum density estimates decrease, species with higher availabilities had higher values of maximum estimated density and more variation in density estimates (Fig. 4b). Area-based rarefaction curves produced from the open-population CAM were lower than those of the closed-population CAM (Fig. 5a, b). Habitat-specific super-population sizes at community levels (mean values across species) were also almost the same in both CAMs, and relative relationships of rarefaction curves among six habitats remained unchanged (Fig. 5a, b). Estimated regional species richness was also the same in the two CAMs ( $\sim 50$ ). Because overall species richness

quickly reached the estimated regional species richness in the closed-population CAM, the difference between overall and mean species richness ( $\beta$  diversity) showed a unimodal shape (Fig. 5c). However, in the open-population CAM, three diversity metrics ( $\alpha$ ,  $\beta$ , and  $\gamma$ ) monotonically increased with area over the considered range of the area (Fig. 5d).

## DISCUSSION

### *Effects of temporary emigration on density and diversity metrics*

Temporary emigration and imperfect detection are long-standing issues in the analysis of count data (Hutto et al. 1986, Hutto 2016, 2017, Marques

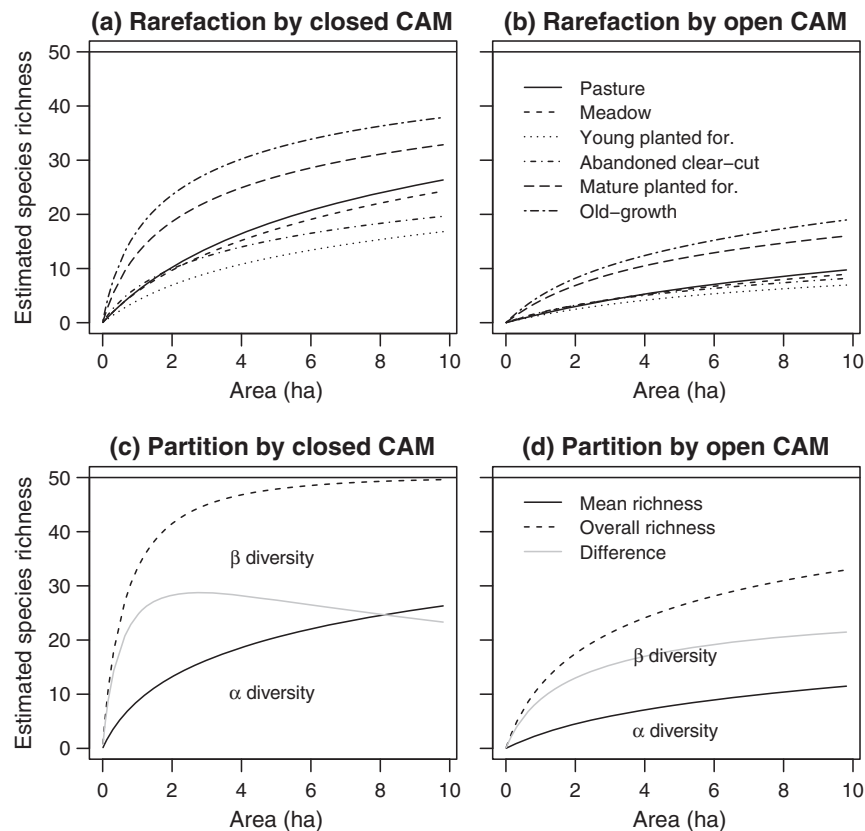


Fig. 5. Area-based rarefaction curves for bird species richness in six habitat types and their additive partitioning produced by two community abundance models. (a, b) Rarefaction curves are shown using the median values of posterior distributions of expected species richness. Both models yielded the same estimates of the regional species richness (upper horizontal lines). (c, d) Mean and overall species richness across the six habitat types ( $\alpha$  and  $\gamma$  diversity, respectively) derived from the rarefaction curves. The difference between  $\alpha$  and  $\gamma$  diversity was  $\beta$  diversity and also shown by gray lines.

et al. 2017). We demonstrated that temporary emigration as well as imperfect detection can seriously compromise the analysis of biological communities when we are primarily interested in density and density-based expected diversity metrics (Eqs. 5–6). When the closure assumption is violated, the  $N$ -mixture (closed population) model infers the number of individuals whose home ranges overlap with the sampling plots (i.e., the super-population size). When this quantity may be linked to the size of sampling plots, that is, it may be treated as the density, and the resultant density and diversity metrics can be greatly biased. Issues of the biased estimates would be important when detection probability can vary among sampling plots, which results in the confounding of the variations of diversity metrics with those of detection probabilities (Ruiz-Gutiérrez et al. 2010, Mc New and Handel 2015). Availability can also be expected to depend on covariates; territory size is likely to be small in resource-rich habitats (Marshall and Cooper 2004, Haché et al. 2012). Variations in availability among sites would be considered by treating the availability probability  $\phi_i$  as the function of covariates as in the detection probability of  $N$ -mixture model (Kéry 2008, Yamaura 2013). Biases involved with temporary emigration are likely to be large when sampling plots are small compared to home range of organisms.

We applied the open-population CAM to empirical repeated distance sampling data, and almost perfect detection in the plots was implied by the parameter estimates. We consider that this result is reasonable since it is known that detection probability is high (>90%) up to 50 m (Schieck 1997, Alldredge et al. 2007), which was the maximum distance of our distance sampling data. Therefore, our model application yielded density estimates on the basis of nearly perfect detection and imperfect availability. Imperfect availability led to a reduction in estimated density, and lowered the rarefaction curves based on density. We suggest that existing models only dealing with imperfect detection (Yamaura et al. 2011, 2012, 2016b) would confound detection probability and availability probability, and therefore, abundance estimates cannot be treated as density estimates in many cases (particularly when home range is larger than plot size) unless temporary emigration is explicitly modeled.

The emigration formulated by the open-population CAM includes not only spatial movement of individuals (spatial temporary emigration), but also other stochastic processes that lead to individuals being unavailable for detection (Kéry and Royle 2016). This random temporary emigration (Fig. 1) may somewhat contribute to the underestimation of densities. Specifically, we formulated the density as the product of super-population size and availability probability ( $\lambda \times \phi$ ); following the notation of Fig. 1, our model dealt with the nominal detection probability as  $p_d$  rather than  $p_a \times p_d$  using distance sampling (given the perfect detection at the zero distance). Availability probability ( $\phi$ ) and associated density therefore are equivalent to  $p_p \times p_a$  and  $\lambda \times p_p \times p_a$ , respectively. The use of other sampling methods including removal sampling or multiple counts in a single visit (Chandler et al. 2011) would incorporate this unavailability ( $p_a$ ) into nominal detection probability ( $= p_a \times p_d$ ), and yield more unbiased density estimates ( $= \lambda \times p_p$ ). We also note that Amundson et al. (2014) combined removal sampling and distance sampling to deal with  $p_a$  and  $p_d$  simultaneously.

For example, in point counts using distance sampling, we may conduct distance sampling multiple times in the single visits and record the time and distance of the detection simultaneously. In this case, the series of the equations dictating the individual counts (Eqs. 2–4) would be changed as follows:

$$M_{ij} \sim \text{Poisson}(\lambda_{ij}) \quad (7)$$

$$N_{ijk} \sim \text{Binomial}(M_{ij}, \phi_i). \quad (8)$$

$$L_{ijkl} \sim \text{Binomial}(N_{ijk}, \theta_i) \quad (9)$$

$$\mathbf{y}_{ijkl} \sim \text{Multinomial}(L_{ijkl}, \boldsymbol{\pi}_{ij}) \quad (10)$$

where  $L_{ijkl}$  is the number of individuals spatially available and producing signals at sub-time  $l$  during visit  $k$ , and  $\theta_i$  is the associated probability of random temporary availability. Separation of these two temporary emigration processes in the transect survey may be achieved by recording other ancillary information of the detection process such as the time when individuals are detected (Borchers and Cox 2017).



*When are we primarily interested in density rather than abundance (super-population size) in a community context?*

When should we deal with spatial temporary emigration separately from imperfect detection by investing additional sampling effort? Individuals exclusively residing in the sampling plots are expected to play greater roles in biological communities and ecosystems than individuals infrequently visiting the sampling plots. Our proposed modeling framework would be an option to properly account for these differences, and species richness is treated as a function of densities of species comprising the community and area. Ordinary species richness (number of detected species in the plot) does not account for temporary emigration but rather implicitly enumerates the number of species whose individuals overlap their home range with sampling plots. Although species richness is suggested to be related to ecological functions performed by communities (e.g., Philpott et al. 2009), the derivative of the species richness (and occupancy probability)–area curves decreases as area increases, and differences in species richness among the habitats can be obscured depending on the area (plot size; Appendix S1). In such case, we suggest that ecosystem functions performed by organisms would be more directly linked to population densities or other related diversity metrics accounting for spatial availability.

In this study, we suggested that three parameters are related to species richness. First is the regional species richness: the number of species in regional communities which can occur in the habitats being sampled. We note that the regional species richness is different than  $\gamma$  diversity, which can be substantially smaller than regional species richness when the total size of the sampling plots is small (Iknayan et al. 2014, Yamaura et al. 2016b). Second is the mean population densities across species, and the third is the variation in density among the species. These quantities are community-level parameters estimated in hierarchical community models and can be scale-free parameters. For example, we can assume that these parameters do not change greatly when the regional biota is shared by the sampling plots, and underlying environments dictating habitat quality of constituent species are similar. In other words, if we have the estimates

of population densities for individual species, we can estimate and compare species richness across areas using area-based rarefaction curves.

In the CAMs, species variation in population densities (represented by the standard deviation [SD] of  $\beta_{0i}$  in Eq. 1) affects rarefaction curves differently depending on the area (or plot size). For small areas, larger SDs increase species richness while, in large areas, larger SDs decrease species richness (Appendix S1). Rarefaction curves computed for different habitat types can therefore cross at a certain plot size when different habitats have different levels of variability in population density among species (SD of  $\beta_{0i}$ ; Appendix S1). Rarefaction curves can also cross when regional species richness (or species pool of a meta-community; Iknayan et al. 2014) differs (Appendix S1). This suggests the significance of comparing species richness across areas even when different habitats are surveyed with the same plot sizes.

We developed our modeling framework focused on community-level densities and species richness ( $\alpha$ ,  $\beta$ , and  $\gamma$  diversity); however, as imperfect detection can impact the analysis of other diversity metrics such as the Shannon index (Yamaura et al. 2016b), variety of diversity metrics would also be affected by this prevalent ecological process of temporary emigration. Since temporary emigration prevails in field surveys and has broad significance for the ecology and function of populations and communities, we suggest the importance of accommodating not only imperfect detection but also temporary emigration in the analysis of biological communities.

## ACKNOWLEDGMENTS

We thank C. Amundson, E. F. Connor, and M. Kéry who provided useful comments on this paper. The manuscript was greatly improved from the comments from C. Handel and an anonymous reviewer. We used the supercomputer of AFFRIT, MAFF, Japan (High Performance Cluster Computing System), to conduct some of the simulation experiments. Ethics statement: This study did not conduct a field survey, but retrieved the data originally collected in our previous study. Relevant data and codes are uploaded as the Supporting Information (Data S1). Funding statement: Y. Yamaura was supported by JSPS KAKENHI (Grant Numbers JP26292074 and JP16KK0176). Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

## LITERATURE CITED

- Allredge, M. W., T. R. Simons, and K. H. Pollock. 2007. Factors affecting aural detections of songbirds. *Ecological Applications* 17:948–955.
- Amundson, C. L., J. A. Royle, and C. M. Handel. 2014. A hierarchical model combining distance sampling and time removal to estimate detection probability during avian point counts. *Auk* 131:476–494.
- Barnosky, A. D., et al. 2011. Has the Earth's sixth mass extinction already arrived? *Nature* 471:51–57.
- Borchers, D. L., and M. J. Cox. 2017. Distance sampling detection functions: 2D or not 2D? *Biometrics* 73:593–602.
- Buckland, S. T., D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas. 2001. Introduction to distance sampling: estimating abundance of biological populations. Oxford University Press, Oxford, UK.
- Buckland, S. T., D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas. 2004. Advanced distance sampling: estimating abundance of biological populations. Oxford University Press, Oxford, UK.
- Burnham, K. P., D. R. Anderson, and J. L. Laake. 1980. Estimation of density from line transect sampling of biological populations. *Wildlife Monographs* 72:3–202.
- Chandler, R. B., J. A. Royle, and D. I. King. 2011. Inference about density and temporary emigration in unmarked populations. *Ecology* 92:1429–1435.
- Connor, E. F., E. Hosfield, D. A. Meeter, and X. Niu. 1997. Tests for aggregation and size-based sample-unit selection when sample units vary in size. *Ecology* 78:1238–1249.
- Connor, E. F., and E. D. McCoy. 1979. The statistics and biology of the species–area relationships. *American Naturalist* 113:791–833.
- Crist, T. O., J. A. Veech, J. C. Gering, and K. S. Summerville. 2003. Partitioning species diversity across landscapes and regions: a hierarchical analysis of  $\alpha$ ,  $\beta$ , and  $\gamma$  diversity. *American Naturalist* 162:734–743.
- Ellis, E. C., and N. Ramankutty. 2008. Putting people in the map: anthropogenic biomes of the world. *Frontiers in Ecology and the Environment* 6:439–447.
- Gotelli, N. J., and R. K. Colwell. 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters* 4:379–391.
- Haberl, H., K. H. Erb, F. Krausmann, V. Gaube, A. Bondeau, C. Plutzer, S. Gingrich, W. Lucht, and M. Fischer-Kowalski. 2007. Quantifying and mapping the human appropriation of net primary production in earth's terrestrial ecosystems. *Proceedings of the National Academy of Sciences* 104:12942–12947.
- Haché, S., M.-A. Villard, and E. M. Bayne. 2012. Experimental evidence for an ideal free distribution in a breeding population of a territorial songbird. *Ecology* 94:861–869.
- Haila, Y. 1988. Calculating and miscalculating density: the role of habitat geometry. *Ornis Scandinavica* 19:88–92.
- Hanski, I. K., and Y. Haila. 1988. Singing territories and home ranges of breeding Chaffinches: visual observation vs. radio-tracking. *Ornis Fennica* 65:97–103.
- Huston, M. A. 1994. Biological diversity: the coexistence of species on changing landscapes. Cambridge University Press, Cambridge, UK.
- Huston, M. A. 2014. Disturbance, productivity, and species diversity: empiricism vs. logic in ecological theory. *Ecology* 95:2382–2396.
- Hutto, R. L. 2016. Should scientists be required to use a model-based solution to adjust for possible distance-based detectability bias? *Ecological Applications* 26:1287–1294.
- Hutto, R. L. 2017. Reply to Marques et al. (2017): how to best handle potential detectability bias. *Ecological Applications* 27:1699–1702.
- Hutto, R. L., S. M. Pletschet, and P. Hendricks. 1986. A fixed-radius point count method for nonbreeding and breeding season use. *Auk* 103:593–602.
- Iknayan, K. J., M. W. Tingley, B. J. Furnas, and S. R. Beissinger. 2014. Detecting diversity: emerging methods to estimate species diversity. *Trends in Ecology & Evolution* 29:97–106.
- James, F. C., and N. O. Wamer. 1982. Relationships between temperate forest bird communities and vegetation structure. *Ecology* 63:159–171.
- Kellner, K. 2015. jagsUI: a wrapper around rjags to streamline JAGS analyses. <https://cran.r-project.org/web/packages/jagsUI/>
- Kéry, M. 2008. Estimating abundance from bird counts: binomial mixture models uncover complex covariate relationships. *Auk* 125:336–345.
- Kéry, M., and J. A. Royle. 2016. Applied hierarchical modeling in ecology: analysis of distribution, abundance and species richness using R and BUGS. Volume 1. Prelude and static models. Academic Press, San Diego, California, USA.
- Loman, J., and T. von Schantz. 1991. Birds in a farmland – more species in small than in large habitat island. *Conservation Biology* 5:176–187.
- MacArthur, R. H., and J. W. MacArthur. 1961. On bird species diversity. *Ecology* 42:594–598.
- MacKenzie, D. I., J. D. Nichols, J. A. Royle, K. H. Pollock, L. L. Bailey, and J. E. Hines. 2006. Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence. Academic Press, Amsterdam, The Netherlands.

- Marques, T. A., et al. 2017. Model-based approaches to deal with detectability: a comment on Hutto (2016a). *Ecological Applications* 27:1694–1698.
- Marshall, M. R., and R. J. Cooper. 2004. Territory size of a migratory songbird response to caterpillar density and foliage structure. *Ecology* 85:432–445.
- Mc New, L. B., and C. M. Handel. 2015. Evaluating species richness: biased ecological inference results from spatial heterogeneity in detection probabilities. *Ecological Applications* 25:1669–1680.
- Newbold, T., et al. 2015. Global effects of land use on local terrestrial biodiversity. *Nature* 520:45–50.
- Newbold, T., et al. 2016. Has land use pushed terrestrial biodiversity beyond the planetary boundary? A global assessment. *Science* 353:288–291.
- Nichols, J. D., L. Thomas, and P. B. Conn. 2009. Inferences about landbird abundance from count data: recent advances and future directions. Pages 201–235 in D. L. Thomson, E. G. Cooch, and M. J. Conroy, editors. *Modeling demographic processes in marked populations*. Springer US, New York, New York, USA.
- Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, and H. Wagner. 2015. *vegan: community ecology R package*. <https://cran.r-project.org/web/packages/vegan/>
- Philpott, S. M., O. Soong, J. H. Lowenstein, A. L. Pulido, D. T. Lopez, D. F. B. Flynn, and F. DeClerck. 2009. Functional richness and ecosystem services: bird predation on arthropods in tropical agroecosystems. *Ecological Applications* 19:1858–1867.
- Plummer, M. 2012. JAGS version 3.2.0 user manual. <http://mcmc-jags.sourceforge.net/>
- R Core Team. 2015. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rosenzweig, M. L. 1995. *Species diversity in space and time*. Cambridge University Press, Cambridge, UK.
- Royle, J. A. 2004. *N*-mixture models for estimating population size from spatially replicated counts. *Biometrics* 60:108–115.
- Royle, J. A., and R. M. Dorazio. 2008. *Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations and communities*. Academic Press, Amsterdam, The Netherlands.
- Ruiz-Gutiérrez, V., E. F. Zipkin, and A. A. Dhondt. 2010. Occupancy dynamics in a tropical bird community: unexpectedly high forest use by birds classified as non-forest species. *Journal of Applied Ecology* 47:621–630.
- Schieck, J. 1997. Biased detection of bird vocalizations affects comparisons of bird abundance among forested habitats. *Condor* 99:179–190.
- Sollmann, R., B. Gardner, K. A. Williams, A. T. Gilbert, and R. R. Veit. 2016. A hierarchical distance sampling model to estimate abundance and covariate associations of species and communities. *Methods in Ecology and Evolution* 7:529–537.
- Stratford, J. A., and W. D. Robinson. 2005. Gulliver travels to the fragmented tropics: geographic variation in mechanisms of avian extinction. *Frontiers in Ecology and the Environment* 3:85–92.
- Tyre, A. J., B. Tenhumberg, S. A. Field, D. Nijjalke, K. Parris, and H. P. Possingham. 2003. Improving precision and reducing bias in biological surveys: estimating false-negative error rates. *Ecological Applications* 13:1790–1801.
- Whittaker, R. H. 1975. *Communities and ecosystems*. Second edition. MacMillan, New York, New York, USA.
- Wiens, J. A. 1989. *The ecology of bird communities*. Volume 1. Foundations and patterns. Cambridge University Press, Cambridge, UK.
- Yamaura, Y. 2013. Confronting imperfect detection: behavior of binomial mixture models under varying circumstances of visits, sampling sites, detectability, and abundance, in small-sample situations. *Ornithological Science* 12:73–88.
- Yamaura, Y., E. F. Connor, J. A. Royle, K. Itoh, K. Sato, H. Taki, and Y. Mishima. 2016a. Estimating species–area relationships by modeling abundance and frequency subject to incomplete sampling. *Ecology and Evolution* 6:4836–4848.
- Yamaura, Y., M. Kéry, and J. A. Royle. 2016b. Study of biological communities subject to imperfect detection: bias and precision of community *N*-mixture abundance models in small-sample situations. *Ecological Research* 31:289–305.
- Yamaura, Y., J. A. Royle, K. Kuboi, T. Tada, S. Ikeno, and S. Makino. 2011. Modelling community dynamics based on species-level abundance models from detection/nondetection data. *Journal of Applied Ecology* 48:67–75.
- Yamaura, Y., J. A. Royle, N. Shimada, S. Asanuma, T. Sato, H. Taki, and S. Makino. 2012. Biodiversity of man-made open habitats in an underused country: a class of multispecies abundance models for count data. *Biodiversity and Conservation* 21:1365–1380.

## SUPPORTING INFORMATION

Additional Supporting Information may be found online at: <http://onlinelibrary.wiley.com/doi/10.1002/ecs2.2028/full>