Title: A Bayesian approach to inferring dispersal kernels with incomplete mark-recapture data

Author: Akira Terui*
Department of Biology, the University of North Carolina at Greensboro, Greensboro, NC 27412 USA
* hanabi0111@gmail.com

Abstract

Dispersal is a fundamental ecological process that links populations, communities and food webs in space. However, dispersal is tremendously difficult to study in the wild because we must track individuals dispersing in a landscape. One conventional method to measure animal dispersal is a mark-recapture technique. Despite its usefulness, this approach has been recurrently criticized because it is virtually impossible to survey all possible ranges of dispersal in nature. Here, I propose a novel Bayesian model to better estimate dispersal parameters from incomplete mark-recapture data. The dispersal-observation coupled model, DOCM, can extract information from both recaptured and unrecaptured individuals, providing less biased estimates of dispersal parameters. Simulations demonstrated the usefulness of DOCM under various sampling designs. I also suggest extensions of the DOCM to accommodate more realistic scenarios. Application of the DOCM may, therefore, provide valuable insights into how individuals disperse in the wild.

Keywords: movement, statistical inference, spatial ecology, Markov chain Monte Carlo, simulation

Introduction

Ecological entities rarely exist independently. Dispersal – any movement of organisms across space – links populations (Hanski 1999; Hanski & Ovaskainen 2000; Terui *et al.* 2018a; Terui *et al.* 2014b), communities (Leibold *et al.* 2004; Terui & Miyazaki 2016) and food webs (Nakano *et al.* 1999; Nakano & Murakami 2001; Spiller *et al.* 2010; Terui *et al.* 2018b) that are otherwise isolated from one another. Ecologists have been intrigued by the spatial process due to its implications for critical applied issues, such as the metapopulation persistence of endangered species in fragmented landscapes (Hanski 1999). More recently, accumulating evidence suggests that dispersal is highly plastic and context-dependent with significant consequences for landscape-level dynamics (Bonte & de la Pena 2009; Bonte *et al.* 2012; Cote *et al.* 2011; Cote *et al.* 2013; Fronhofer *et al.* 2018; Fronhofer *et al.* 2017; Little *et al.* 2019; Terui *et al.* 2017). For example, Fronhofer *et al.* (2018) have shown that context-dependent dispersal in experimental landscapes has stabilizing effects on local food webs coupled via dispersal. Therefore, there is an increasing awareness that an in-depth understanding of dispersal is critical to biodiversity forecasting during rapid environmental changes. Nevertheless, dispersal is inherently difficult to study in the wild (Nathan 2001). There have been many attempts to track dispersal in natural systems (Clobert *et al.* 2012; Comte & Olden 2018; Nathan *et al.* 2008), but linking dispersal processes with specific ecological factors has been challenged by the incomplete nature of field observations. As such, a detailed analysis of dispersal is, to some extent, biased towards small-scale controlled experiments, limiting our ability to infer spatial dynamics at large spatial scales.

One conventional method to measure animal dispersal in the wild is a mark-recapture technique. Although mark-recapture studies can provide valuable insights into how individuals move across space, there are some serious problems when applying this method in nature. First, it is virtually impossible to survey all possible range of dispersal in a landscape (Fujiwara *et al.* 2006; Gowan & Fausch 1996; Schwalb *et al.* 2010; Terui *et al.* 2014a). Consequently, a substantial portion of individuals can leave

behind the study area, causing serious underestimation of dispersal parameters. Second, even when marked individuals remained in the study area, imperfect detection of marked individuals may pose a challenge to infer dispersal processes (Pepino *et al.* 2012; Rodriguez 2002). To date, several statistical models have been proposed to overcome these difficulties (Fujiwara *et al.* 2006; Pepino *et al.* 2012; Rodriguez 2002). For example, Rodriguez (2002) developed a general class of dispersal models that describe how marked individuals are recaptured through dispersal and sampling processes. However, these models are implicit about unrecaptured individuals and/or have limited extendibility to more complex models that capture plastic and context-dependent dispersal. Hence, there is a need to develop a new class of statistical models that have a greater extension capacity.

Bayesian inference provides a flexible statistical framework that may open the opportunity to overcome challenges in utilizing mark-recapture data (Kéry & Schaub 2012; Terui *et al.* 2017). Here, I introduce a novel Bayesian model that integrates dispersal and observation processes into a single coupled model. The dispersal-observation coupled model, DOCM, can extract information from both recaptured and unrecaptured individuals. Consequently, the model can provide less biased estimates of ecological parameters. In this study, I demonstrate that the usefulness of DOCM using simulated test datasets produced under various sampling designs and discuss its extension capacity to more realistic models.

Model

I consider a situation in which a virtual ecologist conducts a mark-recapture study in a one-dimensional space (e.g., a stream). They choose a section with length L for the mark-recapture study (i.e., the observation section) and divide it into subsections with length l. The number of subsections is thus $L\,l^{-1}$. In each subsection, virtual ecologists perform an initial capture survey and assign a subsection ID to each individual to locate them. After marking individuals uniquely, captured individuals are released into the center of the subsection where they were caught. Then, released individuals disperse freely for a certain period and a recapture survey occurs in the observation section. Since the observation section is a finite domain, individuals can leave this area. Also, only survived individuals may be recaptured with some probability even when marked individuals stay in the observation section. Thus, to be recaptured, individuals must (1) stay in the observation section, (2) survive until being recaptured, and (3) be detected if they survive and remain in the observation section. To represent this data-producing process, I propose the following modeling framework that integrates dispersal and observation processes (Figure 1).

Dispersal model. I first model the dispersal process. Let μ_i and x_i be locations at initial capture and recapture sessions, respectively, for individual i. The variables μ_i and x_i may be expressed as the distance from the center of the capture/recaptured subsection to either end of the observation section (e.g., the downstream end of the study section in streams). I assume the location variable at recapture x_i to follow a Laplace distribution, a dispersal kernel commonly used in the dispersal literature (Nathan *et al.* 2012; Rodriguez 2002):

$$f_l(x_i, \mu_i, \delta) = \frac{1}{2\delta}\exp\left(-\frac{|x_i-\mu_i|}{\delta}\right) \quad (1)$$

The parameter δ is the average dispersal distance. Equation 1 illustrates that the recapture location x_i is conditional on the release location μ_i and the dispersal parameter δ (Figure 1).

Observation model. After dispersal, marked individuals are subject to an imperfect observation process. Let y_i be the variable representing a recapture history for individual i ($y_i = 1$ if recaptured; otherwise 0). The response variable y_i can be modeled as a realization of a Bernoulli distribution:

$$y_i|\psi_i s\xi \sim Bernoulli(\psi_i s\xi) \quad (2)$$

where ψ_i is the probability that individual i moves to the subsection of recapture (recaptured individuals) or stays in the observation section (unrecaptured individuals), s is the survival probability between the time points of release and recapture, and ξ is the detection probability during a recapture survey. The parameters s and ξ can be isolated if an independent dataset to estimate detection probability, e.g., multiple-pass removal data, is available (Dorazio *et al.* 2005). Otherwise, the two parameters need to be condensed into recapture probability ϕ $(= s\xi)$ such that:

$$y_i|\psi_i \phi \sim Bernoulli(\psi_i \phi) \quad (3)$$

Here, I couple the observation and dispersal models by describing ψ_i as a function of the dispersal parameter δ and release location μ_i. Specifically, ψ_i is denoted as:

$$\psi_i = \begin{cases} \int_{x_i-\frac{l}{2}}^{x_i+\frac{l}{2}} f_l(x_i, \mu_i, \delta)\, dx_i \; for\; recaptured\; individuals \\ \int_0^L f_l(x_i, \mu_i, \delta)\, dx_i \; for\; unrecaptured\; individuals \end{cases} \quad (4)$$

Recaptured individuals are known to be present at the subsection of recapture, so the range of integration is given as $x_i - \frac{l}{2}$ to $x_i + \frac{l}{2}$ in equation 4 (i.e., from one end to another end of the subsection). This expression gives the probability of movement from the release location μ_i to the subsection of recapture given the dispersal parameter δ. For unrecaptured individuals, equation 4 accounts for two important facts: (1) a greater value of the dispersal parameter decreases the probability of remaining in the observation section (ψ_i) and (2) the release location μ_i influences ψ_i (i.e., individuals released near the edge of the observation section are more likely to emigrate; Figure 1b). Key parameters in the DOCM were summarized in Table 1.

Evaluation of model performance. To evaluate the performance of the DOCM, I generated test datasets under different sampling designs. Specifically, I focused on the following design factors that are related to sampling efforts in the field: (1) the number of individuals marked and released N (100, 500, and 1000 individuals), (2) the length of the observation section L (500 and 1000 m) and (3) the length of an individual subsection or resolution l (20 and 50 m) (Figure 2). In addition, I considered variation in the recapture probability ϕ (0.25, 0.50, and 0.75). I considered all possible combinations of N, L, l, *and* ϕ (36 combinations) when generating test datasets.

Under each sampling design, I produced 100 test datasets with different values of δ, which was drawn from a uniform distribution (range: 10 – 300 m). Each independent dataset was generated as follows. First, N individuals were assigned randomly to $\frac{L}{l}$ subsections (Figure 3a). These "marked" individuals were released at the center of the captured subsection, which was recorded as release location μ_i. Second, released individuals relocate themselves along a one-dimensional space according to a known dispersal kernel as $x_{i,true}|\, \mu_i, \delta \sim Laplace(\mu_i, \delta)$ (Figure 3b). Individuals were considered to remain in the observation section if true recapture location $x_{i,true}$ was within a range of 0 – L m. Then, remained

individuals were recaptured with recapture probability ϕ (Figure 3c). When recaptured, the true recapture location $x_{i,true}$ was rounded to a location value at the center of the recapture subsection x_i to mimic real field data (Figure 3c). For unrecaptured individuals, x_i was recorded as "NA".

I estimated average dispersal distance δ and recapture probability ϕ using the DOCM and a simple dispersal model. The simple dispersal model is a "control" that does not model the observation process and the average dispersal distance was estimated as $x_i|\mu_i, \delta \sim Laplace(\mu_i, \delta)$. The estimates of average dispersal distances were compared between the models. Meanwhile, the estimated recapture probability ϕ was compared with the proportion of individuals recaptured ($\frac{n}{N}$, where n is the number of recaptured individuals) in the test dataset used to estimate ϕ because the simple dispersal model does not estimate ϕ. Finally, I assessed the accuracy (i.e., the closeness of the median estimate to the true parameter) and precision (i.e., the 95% credible interval [CI]) of the estimated parameters.

The Bayesian models were fitted to the test datasets to estimate δ and ϕ. Vague priors were assigned to the parameters: a half-Cauchy distribution for δ (scale = 500) and a uniform distribution for ϕ (range: 0 – 1). Three Markov chain Monte Carlo (MCMC) chains were run with 4500 iterations, 1500 burn-ins, and 3 thin numbers, resulting in a total of 1500 MCMC samples. Convergence was assessed by whether the R–hat indicator of each parameter had reached a value near 1. All statistical analysis was conducted using R 3.5.1 (RCoreTeam 2019) and JAGS 4.3.0 (Plummer 2003). A sample of JAGS scripts for the DOCM was provided in Box 1. R and JAGS scripts used in simulations will be made available at Github upon publication.

Results and discussion

Model performance. The DOCM performed well under various sampling designs. Figure 4 shows the relationship between the true and estimated values of δ (denoted as δ_{true} and δ_{est}, respectively) when the recapture probability ϕ was 0.50. The parameter estimates from the DOCM were always closer to the true values (compare red and black lines in Figure 4) compared with those derived from the simple dispersal model without the observation process. The degree of improvement was significant. While 95% CIs of the simple dispersal model tended not to include δ_{true}, the DOCM was more likely to encompass the true values especially when the observation length was long enough ($L = 1000$ m). Similarly, the DOCM provided less biased estimates of recapture probability ϕ, a composite of survival and detection probabilities (Figure 5). The estimated ϕ was higher than the proportion of individuals recaptured in the test dataset ($\frac{n}{N}$, where n is the number of recaptured individuals) because it was corrected for permanent emigration. However, as the δ_{true} increases, the DOCM became underestimating the parameters, though the degree of bias is better than the simple dispersal model. This pattern was apparent when the observation section was short relative to δ_{true} and is caused by the substantial number of individuals leaving behind the observation section. For each of the parameters, these results were qualitatively similar irrespective of ϕ_{true}, although higher values of ϕ_{true} led to the narrower range of 95% CIs for δ_{est} as more individuals were recaptured. Detailed results with different values of ϕ_{true} were provided in Figures S1 – S4.

The number of individuals marked (N), observation section length (L) and spatial resolution (l) had distinct effects on δ_{est} and ϕ_{est}. The number of individuals marked had a clear influence on the precision of the parameter estimates. The 95% CIs of δ_{est} and ϕ_{est} (error bars in Figures 4 and 5) became narrower clearly when N increased from 100 to 500 individuals. A further increase in N, however, did not improve the precision of the parameter estimates. Increasing N did not contribute to improving the

accuracy of the parameter estimates (i.e., the closeness to δ_{true} and ϕ_{true}). In contrast, the length of the observation section L was more influential on the accuracy of δ_{est} while having little influence on the precision of the parameter estimates (Figures 4 and 5). Increased L improved the accuracy of δ_{est} because long-distance dispersers were more likely to be recaptured. Neither accuracy nor precision was improved when the spatial resolution of sampling (smaller l) increased.

Usefulness and limitations. The DOCM worked well under various sampling designs, proving its usefulness to infer dispersal processes in the wild. The DOCM can extract information from both recaptured and unrecaptured individuals, thereby improving the accuracy of parameter estimates. The DOCM, therefore, represents a promising tool to study dispersal processes. To apply the DOCM, users must obtain the following data: (1) individuals must be marked uniquely or by release subsection; (2) release location (μ_i); (3) recapture location (x_i); (3) spatial resolution of subsection length (l); (4) observation section length (L). These are a common dataset obtained through a mark-recapture study, so no additional work may be required to use the DOCM. Furthermore, if users have an independent estimate of detection probability ξ through multiple-pass removal (Dorazio *et al.* 2005) or other methods, it is also possible to estimate the true survival rate s that is corrected for permanent emigration (Terui *et al.* 2017). However, there are caveats when interpreting the results. As stated above, the estimated dispersal parameter δ_{est} and recapture probability ϕ_{est} can be biased when the average dispersal distance (δ_{true}) exceeds ca. $\frac{L}{5}$. This happened because a significant portion of individuals may have left the observation section. Practically, the estimated dispersal parameter δ_{est} may be used to determine whether the average dispersal distance exceeds the threshold. In cases where $\delta_{est} > \frac{L}{5}$, users shall acknowledge the potential bias in parameter estimates.

It is important to emphasize that different design factors (L, N, l) had different effects on the parameter estimates, corroborating the previous findings by Pépino *et al.* (2016). The results indicate that increasing the length of the observation section is most effective to increase the estimation accuracy of model parameters (the closeness to the true value). This is reasonable because increasing the length of the observation section is the only way to catch long-distance dispersers. In contrast, increasing the number of individuals marked is more important to improve the precision of the dispersal parameters (i.e., the range of 95% CI). Therefore, I recommend users paying close attention to the length of the observation section L and the number of individuals marked N when designing a mark-recapture study. Spatial resolution l had minimal influence on the accuracy and precision of parameter estimates, so this design component may be determined based on the biology of a study species.

Model extension. The DOCM can be extended in two distinct ways. First, the dispersal model can capture the further complexity of the dispersal process. A growing body of evidence suggests that populations are composed of "resident" and "mobile" individuals with different behavioral and/or phenotypic characteristics (Clobert *et al.* 2012; Clobert *et al.* 2009; Cote *et al.* 2008; Cote *et al.* 2011; Cote *et al.* 2013; Terui *et al.* 2017). Such a linkage between dispersal and individual-level traits can be modeled using the following expression:

$$x_i | \mu_i, \delta_i \sim Laplace(\mu_i, \delta_i) \quad (5a)$$

$$\log(\delta_i) = \alpha + \beta z_i \quad (5b)$$

where α is the intercept, β the regression coefficient and z_i the linear predictor representing an individual trait. Expressed differently, equation 5 can be written as:

$$f_l(x_i, \mu_i, \alpha, \beta, z_i) = \frac{1}{2\exp(\alpha+\beta z_i)} \exp\left(-\frac{|x_i-\mu_i|}{\exp(\alpha+\beta z_i)}\right) \quad (6)$$

This model connects the trait variable z_i with the dispersal parameter δ_i by estimating α and β. In this model, individuals follow different dispersal kernels according to their ecological trait(s), such as body size. If the variable z is a random variable that follows a normal distribution with a mean μ_z and standard deviation σ_z, $g(z, \mu_z, \sigma_z)$, then the composite dispersal kernel $h(x_i, \mu_i, \alpha, \beta)$ is:

$$h(x_i, \mu_i, \alpha, \beta) = \int f_l(x_i, \mu_i, \alpha, \beta, z) g(z, \mu_z, \sigma_z)\, dz \quad (7)$$

If z is a binary variable drawn from a Bernoulli distribution with a success probability p, the composite dispersal kernel is:

$$h(x_i, \mu_i, \alpha, \beta) = p\left[\frac{1}{2\exp(\alpha+\beta)} \exp\left(-\frac{|x_i-\mu_i|}{\exp(\alpha+\beta)}\right)\right] + (1-p)\left[\frac{1}{2\exp(\alpha)} \exp\left(-\frac{|x_i-\mu_i|}{\exp(\alpha)}\right)\right] \quad (8)$$

Equations 7 and 8 can be interpreted as an extension of a mixture Laplace dispersal kernel, in which a certain proportion of individuals are assigned randomly to a resident or mobile component in the model (Rodriguez 2002). The difference with a mixture Laplace dispersal kernel is that the above equations are explicit regarding "who is resident or mobile" as the expected dispersal distance for individual *i* (δ_i) is related to ecological traits via the regression parameters. Terui *et al.* (2017) used this modeling framework to assess the effects of parasite infection on the dispersal of a stream fish species. It is important to note that there are many other dispersal kernels, such as a mixture of Gaussian dispersal kernels (Comte & Olden 2018; Nathan *et al.* 2012; Skalski & Gilliam 2000). Users may choose appropriate dispersal kernels given the ecology of a study species. I also point interested readers to Nathan *et al.* (2012) for dispersal kernels in two dimensional systems as another extension of the dispersal model.

Second, the observation model can also be extended to account for individual-level variability in recapture probability ϕ. Survival and detection probabilities may vary among individuals and ignoring this complexity could cause biased estimates of dispersal parameters. The simplest way to account for the variability is to model ϕ_i as a random variable drawn from a Beta distribution:

$$\phi_i \sim Beta(\epsilon, \nu) \quad (9)$$

This allows the model to account for individual-level variation in recapture probability ϕ. If there are hypothesized predictors that could influence the recapture probability (e.g., habitat structure), such effects can be modeled as:

$$\phi_i = Beta(\epsilon_i, \nu_i) \quad (10a)$$
$$\epsilon_i = \pi\theta_i \quad (10b)$$
$$\nu_i = \pi(1-\theta_i) \quad (10c)$$
$$\text{logit}(\theta_i) = \alpha_\theta + \beta_\theta z_i \quad (10d)$$

where θ_i is the expected recapture probability, π the dispersion parameter, α_θ the intercept and β_θ the regression coefficient. Therefore, the DOCM can deal with the complexity of field data.

Conclusion. Dispersal is a fundamental process that drives the ecology and evolution of various organisms (Clobert *et al.* 2012) and quantifying dispersal is a critical task to forecast spatial dynamics of ecological systems (Hanski 1999; Hanski & Ovaskainen 2000; Terui *et al.* 2017). Although great strides have been made in how to quantify dispersal in the wild (e.g., genotyping) (Comte & Olden 2018; Morrissey & Ferguson 2011), direct measurements of dispersal still provide essential information for an understanding of spatial processes (Comte & Olden 2018; Kadoya & Inoue 2015; Terui *et al.* 2017). The Bayesian implementation of the DOCM provides extensive flexibility in the model formulation, offering a generic framework to study dispersal in the wild. Accurate inference of dispersal processes with sophisticated statistical models may enhance our ability to manage ecosystems in a changing world.

Acknowledgments

I would like to express my sincere appreciation for all the help and kindness I have received from everyone who has supported my work.

References

Bonte D, de la Pena E (2009) Evolution of body condition‐dependent dispersal in metapopulations. Journal of Evolutionary Biology 22: 1242-1251.

Bonte D, Van Dyck H, Bullock JM, Coulon A, Delgado M, Gibbs M, Lehouck V, Matthysen E, Mustin K, Saastamoinen M (2012) Costs of dispersal. Biological Reviews 87: 290-312.

Clobert J, Baguette M, Benton TG, Bullock JM (2012) Dispersal ecology and evolution. Oxford University Press, Oxford

Clobert J, Galliard L, Cote J, Meylan S, Massot M (2009) Informed dispersal, heterogeneity in animal dispersal syndromes and the dynamics of spatially structured populations. Ecology Letters 12: 197-209.

Comte L, Olden JD (2018) Fish dispersal in flowing waters: A synthesis of movement- and genetic-based studies. Fish and Fisheries 19: 1063-1077.

Cote J, Dreiss A, Clobert J (2008) Social personality trait and fitness. Proceedings of the Royal Society B-Biological Sciences 275: 2851-2858.

Cote J, Fogarty S, Brodin T, Weinersmith K, Sih A (2011) Personality-dependent dispersal in the invasive mosquitofish: group composition matters. Proceedings of the Royal Society B: Biological Sciences 278: 1670-1678.

Cote J, Fogarty S, Tymen B, Sih A, Brodin T (2013) Personality-dependent dispersal cancelled under predation risk. Proceedings of the Royal Society of London B: Biological Sciences 280: 20132349.

Dorazio RM, Jelks HL, Jordan F (2005) Improving removal‐based estimates of abundance by sampling a population of spatially distinct subpopulations. Biometrics 61: 1093-1101.

Fronhofer EA, Legrand D, Altermatt F, Ansart A, Blanchet S, Bonte D, Chaine A, Dahirel M, De Laender F, De Raedt J, di Gesu L, Jacob S, Kaltz O, Laurent E, Little CJ, Madec L, Manzi F, Masier S, Pellerin F, Pennekamp F, Schtickzelle N, Therry L, Vong A, Winandy L, Cote J (2018) Bottom-up and top-down control of dispersal across major organismal groups. Nature Ecology & Evolution 2: 1859-1863.

Fronhofer EA, Nitsche N, Altermatt F (2017) Information use shapes the dynamics of range expansions into environmental gradients. Global Ecology and Biogeography 26: 400-411.

Fujiwara M, Anderson KE, Neubert MG, Caswell H (2006) On the Estimation of Dispersal Kernels from Individual Mark-Recapture Data. Environmental and Ecological Statistics 13: 183-197.

Gowan C, Fausch KD (1996) Mobile brook trout in two high-elevation Colorado streams: re-evaluating the concept of restricted movement. Canadian Journal of Fisheries and Aquatic Sciences 53: 1370-1381.

Hanski I (1999) Metapopulation ecology. Oxford University Press, Oxford

Hanski I, Ovaskainen O (2000) The metapopulation capacity of a fragmented landscape. Nature 404: 755-758.

Kadoya T, Inoue T (2015) Spatio‐temporal pattern of specific gravity of mangrove diaspore: implications for upstream dispersal. Ecography 38: 472-479.

Kéry M, Schaub M (2012) Bayesian population analysis using WinBUGS: a hierarchical perspective. Academic Press, Waltham, MA

Leibold MA, Holyoak M, Mouquet N, Amarasekare P, Chase JM, Hoopes MF, Holt RD, Shurin JB, Law R, Tilman D, Loreau M, Gonzalez A (2004) The metacommunity concept: a framework for multi-scale community ecology. Ecology Letters 7: 601-613.

Little CJ, Fronhofer EA, Altermatt F (2019) Dispersal syndromes can impact ecosystem functioning in spatially structured freshwater populations. Biology Letters 15: 20180865.

Morrissey MB, Ferguson MM (2011) Individual variation in movement throughout the life cycle of a stream-dwelling salmonid fish. Molecular Ecology 20: 235-248.

Nakano S, Miyasaka H, Kuhara N (1999) Terrestrial-aquatic linkages: Riparian arthropod inputs alter trophic cascades in a stream food web. Ecology 80: 2435-2441.

Nakano S, Murakami M (2001) Reciprocal subsidies: dynamic interdependence between terrestrial and aquatic food webs. Proceedings of the National Academy of Sciences of the United States of America 98: 166-170.

328 Nathan R (2001) The challenges of studying dispersal. Trends in Ecology & Evolution 16: 481-483.
329 Nathan R, Getz WM, Revilla E, Holyoak M, Kadmon R, Saltz D, Smouse PE (2008) A movement
330 ecology paradigm for unifying organismal movement research. Proceedings of the National
331 Academy of Sciences of the United States of America 105: 19052-19059.
332 Nathan R, Klein E, Robledo-Arnuncio JJ, Revilla E (2012) Dispersal kernels: review. In: J Clobert, M
333 Baguette, TG Benton, JM Bullock (eds) Dispersal Ecology and Evolution. Oxford University
334 Press, Oxford, UK, pp. 187-210
335 Pepino M, Rodriguez MA, Magnan P (2012) Fish dispersal in fragmented landscapes: a modeling
336 framework for quantifying the permeability of structural barriers. Ecological Applications 22:
337 1435-1445.
338 Pépino M, Rodríguez MA, Magnan P (2016) Assessing the detectability of road crossing effects in
339 streams: mark–recapture sampling designs under complex fish movement behaviours. Journal of
340 Applied Ecology 53: 1831-1841.
341 Plummer M (2003) JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling
342 Proceedings of the 3rd international workshop on distributed statistical computing. Vienna,
343 Austria.
344 RCoreTeam (2019) R: A Language and Environment for Statistical Computing. https://www.R-
345 project.org/ 25 August, 2019
346 Rodriguez MA (2002) Restricted movement in stream fish: The paradigm is incomplete, not lost. Ecology
347 83: 1-13.
348 Schwalb AN, Garvie M, Ackerman JD (2010) Dispersion of freshwater mussel larvae in a lowland river.
349 Limnology and Oceanography 55: 628-638.
350 Skalski GT, Gilliam JF (2000) Modeling diffusive spread in a heterogeneous population: A movement
351 study with stream fish. Ecology 81: 1685-1700.
352 Spiller DA, Piovia-Scott J, Wright AN, Yang LH, Takimoto G, Schoener TW, Iwata T (2010) Marine
353 subsidies have multiple effects on coastal food webs. Ecology 91: 1424-1434.
354 Terui A, Ishiyama N, Urabe H, Ono S, Finlay JC, Nakamura F (2018a) Metapopulation stability in
355 branching river networks. Proceedings of the National Academy of Sciences 115: E5963-E5969.
356 Terui A, Miyazaki Y (2016) Three ecological factors influencing riverine fish diversity in the Shubuto
357 River system, Japan: habitat capacity, habitat heterogeneity and immigration. Limnology 17: 143-
358 149.
359 Terui A, Miyazaki Y, Yoshioka A, Kadoya T, Jopp F, Washitani I (2014a) Dispersal of larvae of
360 *Margaritifera laevis* by its host fish. Freshwater Science 33: 112-123.
361 Terui A, Miyazaki Y, Yoshioka A, Kaifu K, Matsuzaki SS, Washitani I (2014b) Asymmetric dispersal
362 structures a riverine metapopulation of the freshwater pearl mussel *Margaritifera laevis*. Ecology
363 and Evolution 4: 3004-3014.
364 Terui A, Negishi JN, Watanabe N, Nakamura F (2018b) Stream Resource Gradients Drive Consumption
365 Rates of Supplemental Prey in the Adjacent Riparian Zone. Ecosystems 21: 772-781.
366 Terui A, Ooue K, Urabe H, Nakamura F (2017) Parasite infection induces size-dependent host dispersal:
367 consequences for parasite persistence. Proceedings of the Royal Society B: Biological Sciences
368 284: 20171491.
369
370

Box 1 Sample JAGS script for the dispersal-observation coupled model

```
model{
  # Priors
  theta ~ dscaled.gamma(500, 1)
  phi ~ dunif(0, 1)

  # Observation model
  for (i in 1:Nsample){
    Y[i] ~ dbern(psi[i]*phi)
    psi[i] <- pdexp(UL[i], Mu[i], theta) - pdexp(DL[i], Mu[i], theta)
  }

  # Dispersal model
  for(i in 1:Nsample){
    X[i] ~ ddexp(Mu[i], theta)
  }
  delta <- 1/theta

}
```

391 **Table 1** Key parameters used in the dispersal-observation coupled model (DOCM)
392

Parameter	Interpretation
δ	Average dispersal distance
μ	Release location
ψ	Probability of moving from the release location μ to the recapture subsection (recaptured individuals); Probability of remaining in the observation section (unrecaptured individuals)
ϕ	Recapture probability
s	True survival probability
ξ	Detection probability

393
394

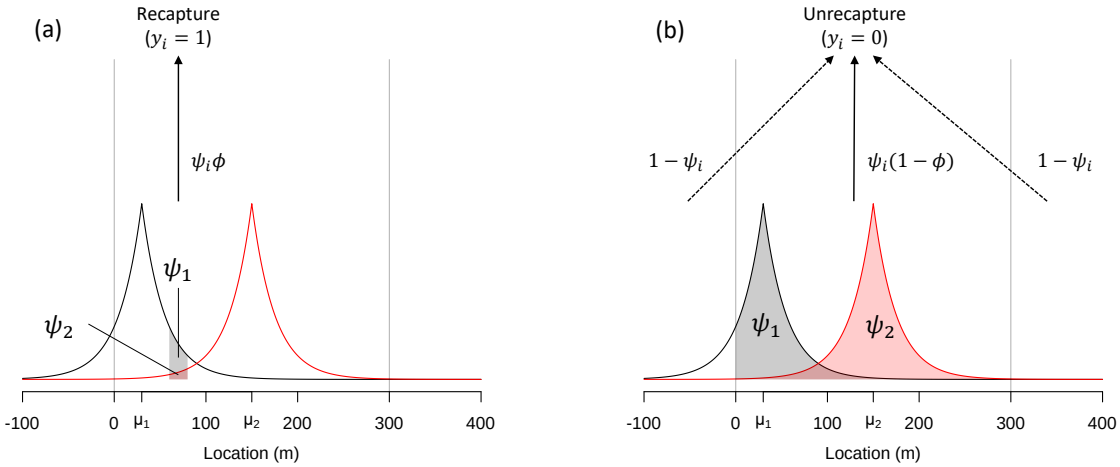
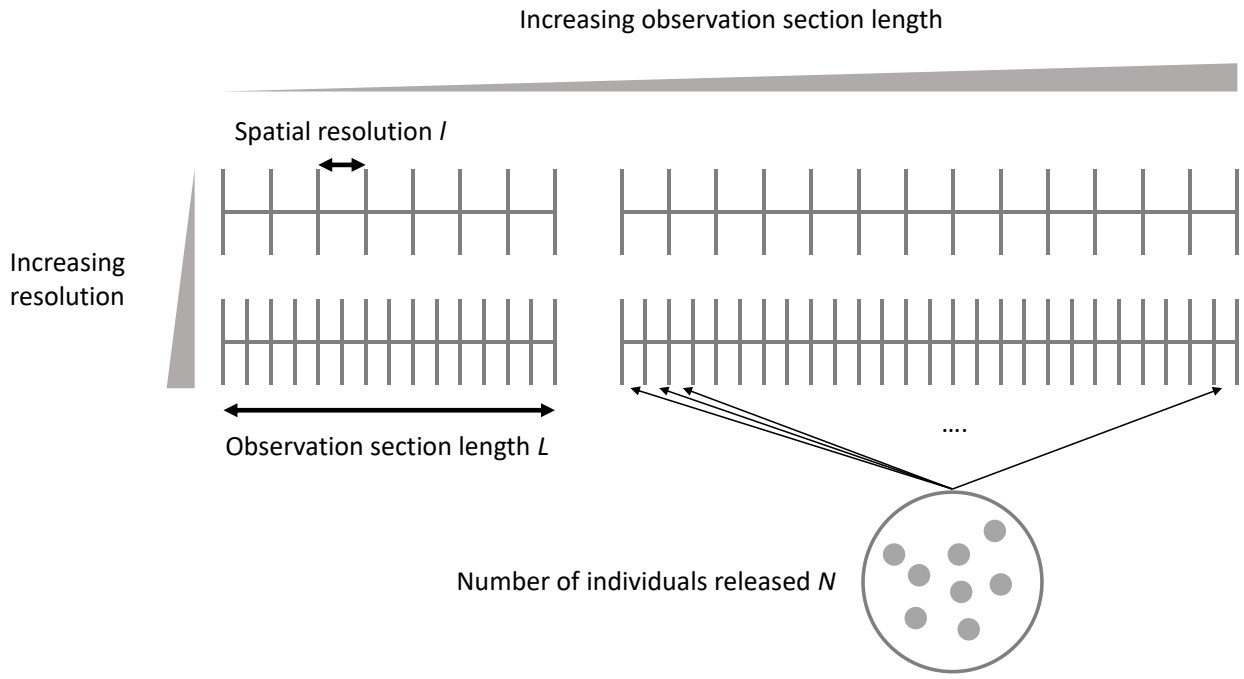


Figure 1 Graphical representation of the dispersal-observation coupled model (DOCM). The black and red lines are the examples of Laplace dispersal kernels for individual 1 and 2 released at different locations (average dispersal distance $\delta = 25$ m for both kernels). Vertical gray lines indicate the observation section (0 – 300 m for this example). Shaded areas denote ψ_i that represents the probability that an individual moves from the release location μ_i to the recapture subsection for recaptured individuals (a) or the probability that an individual stays in the observation section for unrecaptured individuals (b). Individuals released at different locations (μ_1 and μ_2) have different values of ψ_i. After the dispersal process, individuals are subject to incomplete recapture surveys, by which individuals may be detected with the recapture probability ϕ if they remained in the observation section. Note that the recapture probability ϕ is a composite of survival and detection probabilities.



408
409 **Figure 2** Schematic diagram of sampling designs. Three design factors were considered: (1) the number
410 of individuals released N, (2) observation section length L, and (3) spatial resolution l.
411

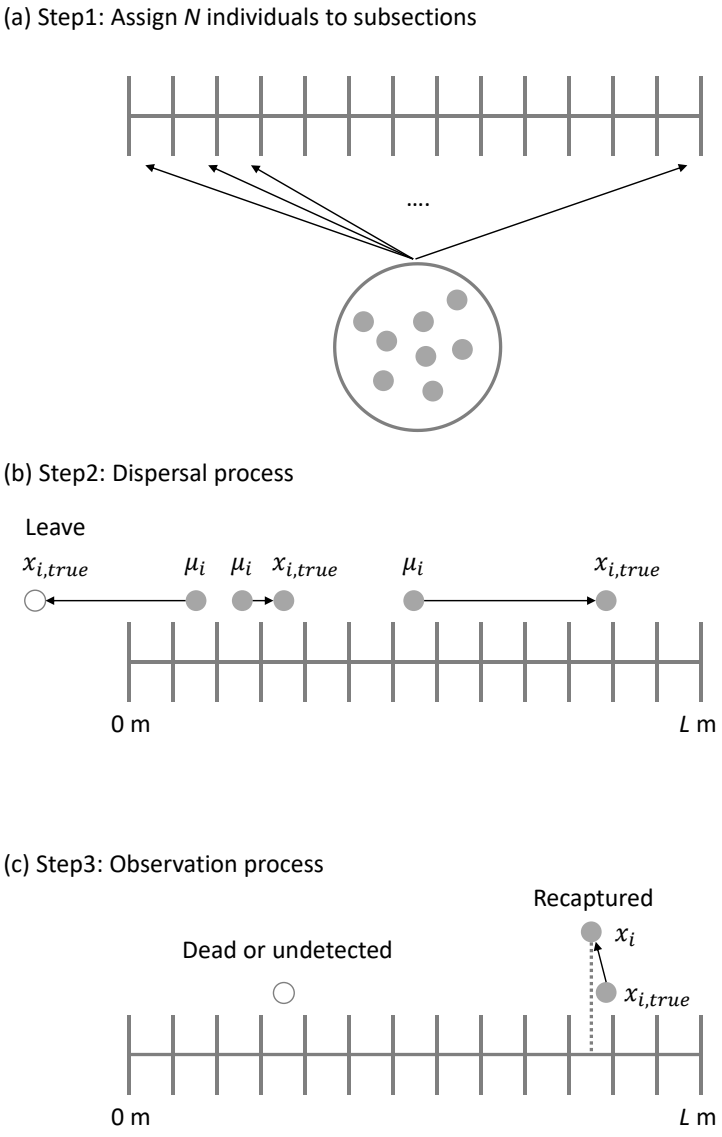


Figure 3 Procedure used to generate test datasets. (a) In step 1, N individuals were randomly distributed in the observation section, resembling a marking process in the field. (b) In step 2, marked individuals move freely from the center of the released subsection (μ_i) according to a known dispersal kernel (in this case, a Laplace distribution with a mean dispersal distance δ). Since the observation section is a finite domain, individuals can leave the observation section and may never be recaptured (open circle in the figure). Individuals were considered to remain in the observation section when true location after dispersal $x_{i,true}$ was within a range of $0 - L$ m. (c) In step 3, individuals that remained in the observation section were subject to an incomplete recapture survey. Remained individuals were recaptured with recapture probability ϕ, and recapture location x_i was recorded as the center of the recapture subsection (the vertical dotted line in the figure). A certain proportion of remained individuals $(1 - \phi)$ may not be recaptured because they may have died or undetected (open circle in the figure).

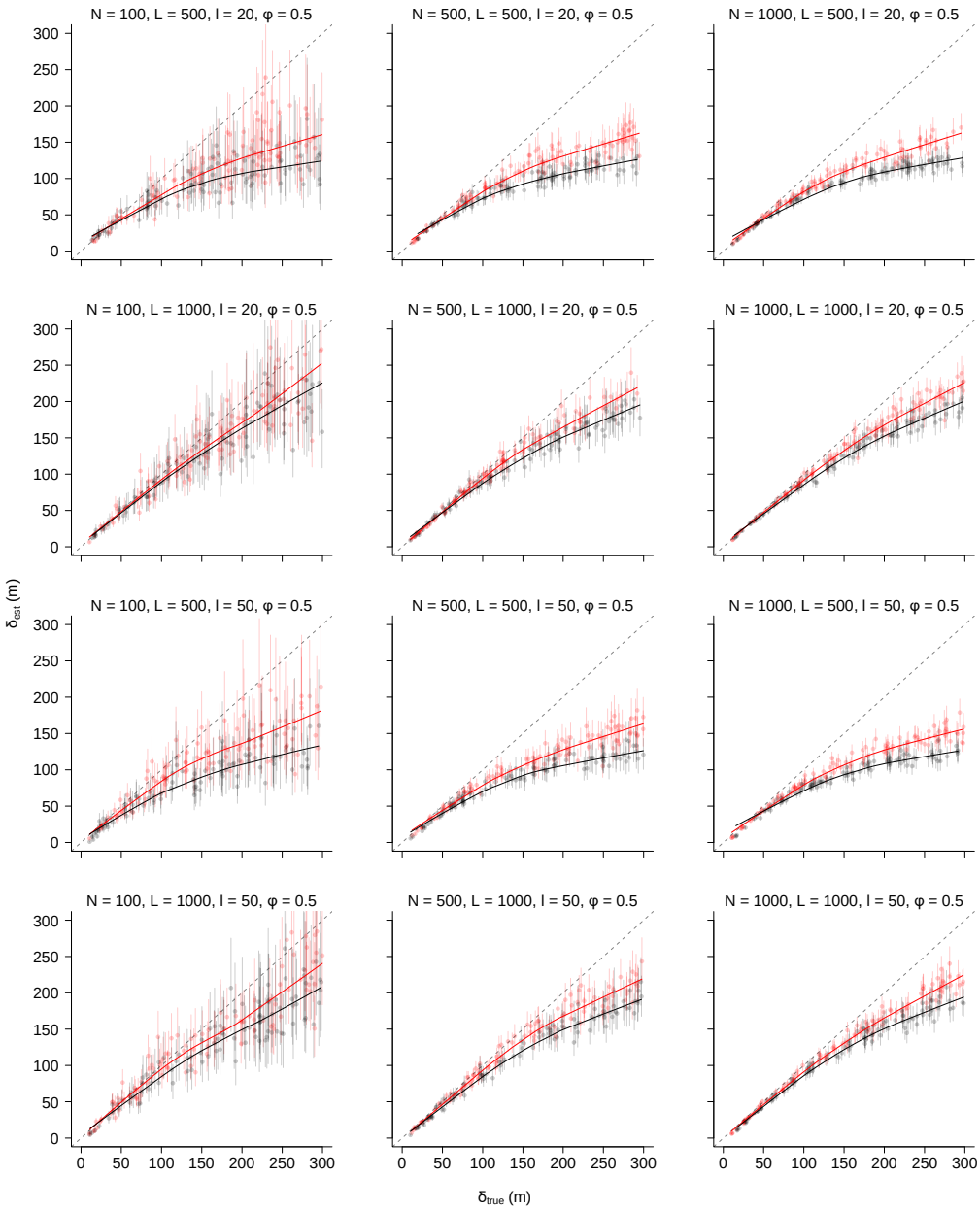


Figure 4 The relationship between true (*y*-axis) and estimated (*x*-axis) average dispersal distances when the true recapture probability $\phi_{true} = 0.5$. Red points are the median estimates from the DOCM, while grey points showing the median estimates from the simple dispersal model. Error bars are 95% credible intervals. Gray broken lines denote a 1:1 relationship. Different panels are estimates under different sampling designs, and the values of sampling design factors are shown on the top of each panel. Red and gray solid lines are smooths for the DOCM and the simple dispersal model, respectively.

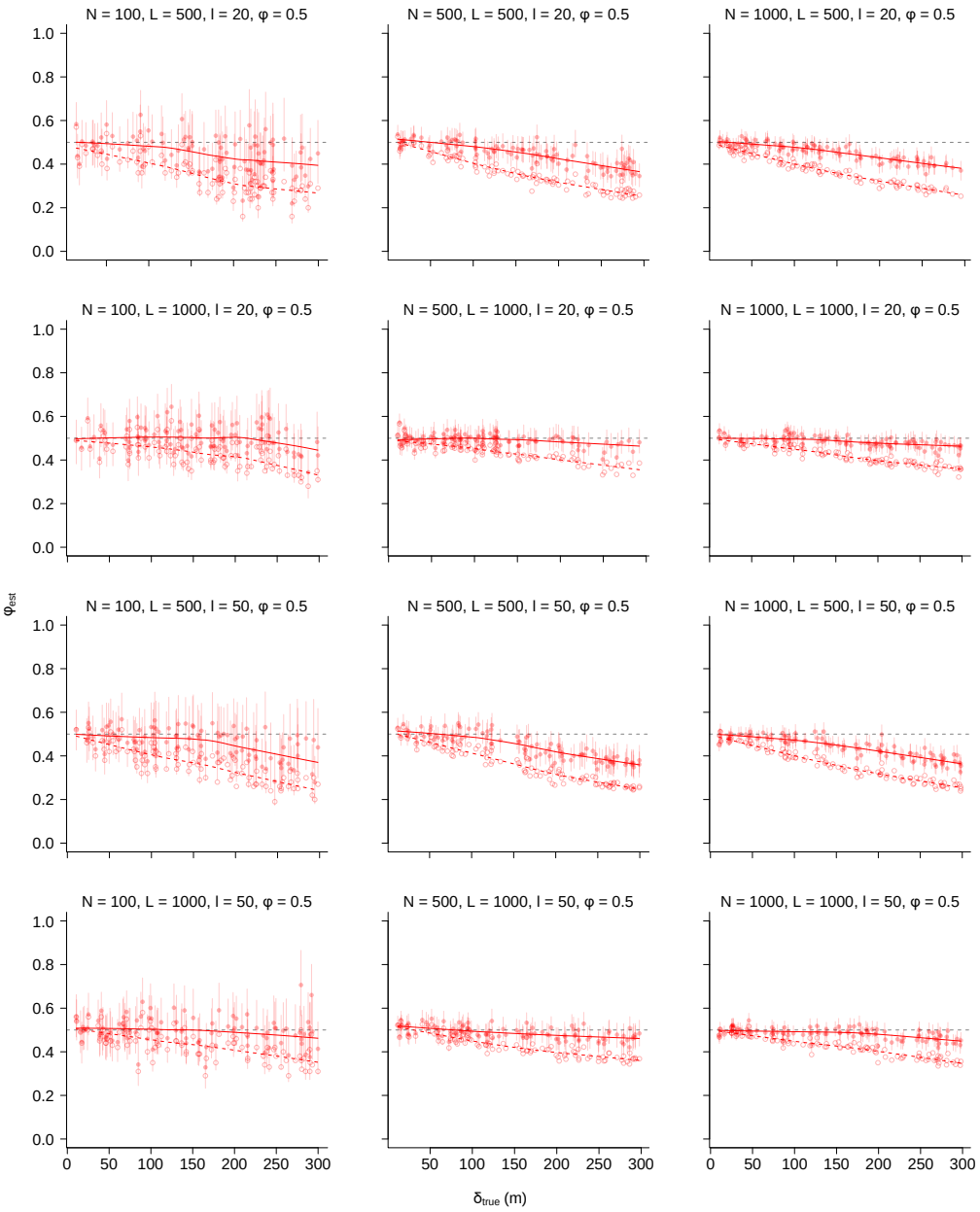


Figure 5 The relationship between the estimated recapture probability ϕ_{est} and true average dispersal distance δ_{true} when the true recapture probability $\phi_{true} = 0.5$. Red filled points are the median estimates from the DOCM while red open points denoting the proportion of individuals recaptured for each test dataset. Error bars are 95% credible intervals. Grey broken lines denote the true recapture probability ϕ_{true}. Different panels are estimates under different sampling designs, and the values of sampling design factors are shown on the top of each panel. Solid and broken red lines are smooths for the DOCM estimates and the proportion of individuals recaptured, respectively.

Supporting Information for:

Title: A Bayesian approach to inferring dispersal kernels with incomplete mark-recapture data

Author: Akira Terui*
Department of Biology, the University of North Carolina at Greensboro, Greensboro, NC 27412 USA
*hanabi0111@gmail.com

This PDF file includes:

Figures S1 – S4

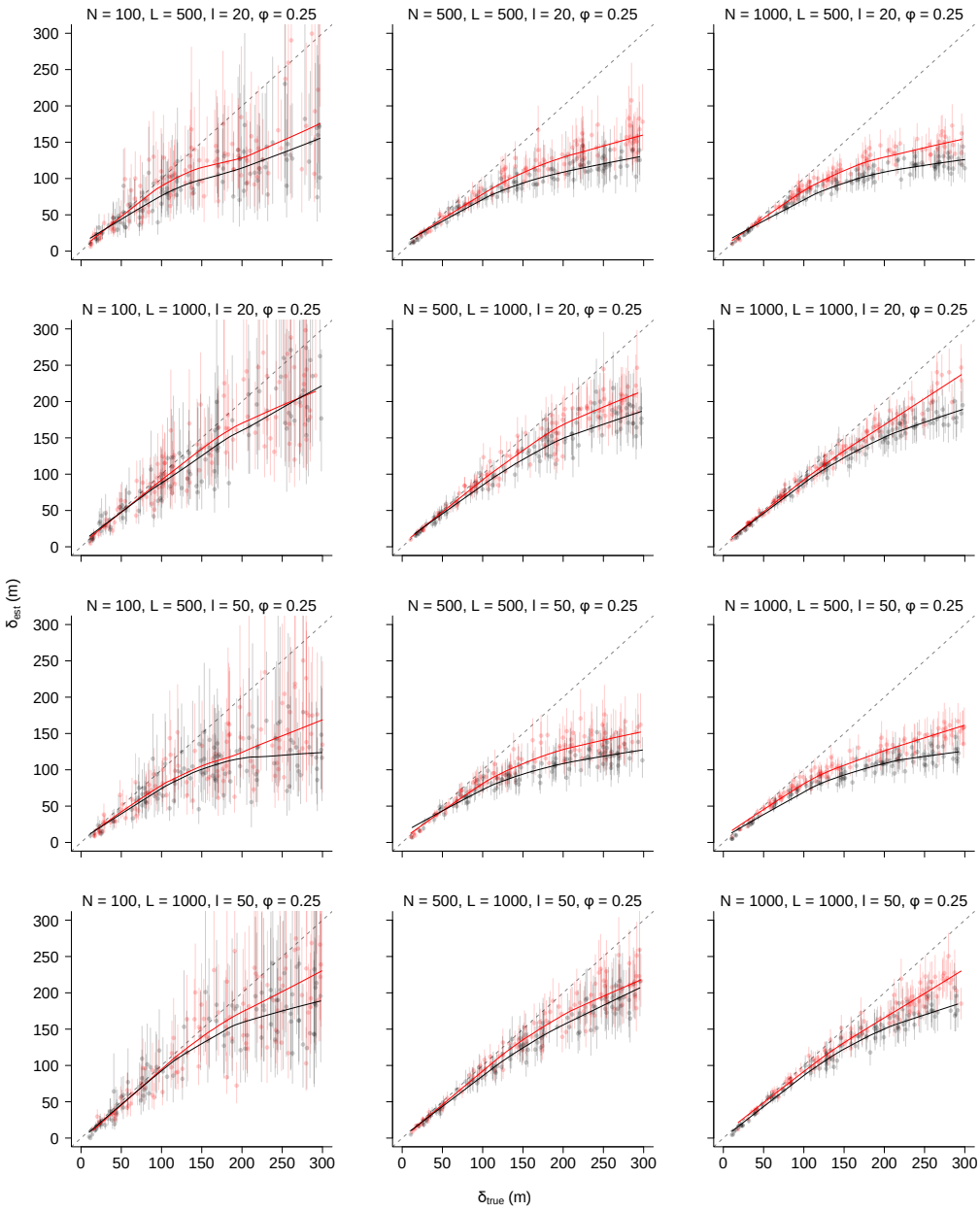


Figure S1 The relationship between true (δ_{true}; y-axis) and estimated (δ_{est}; x-axis) average dispersal distances when the true recapture probability $\phi_{true} = 0.25$. Red points are the median estimates from the DOCM while grey points showing the median estimates from the simple dispersal model. Error bars are 95% credible intervals. Gray broken lines denote a 1:1 relationship. Different panels are estimates under different sampling designs, and the values of sampling design factors are shown on the top of each panel. Red and gray solid lines are smooths for the DOCM and the simple dispersal model, respectively.

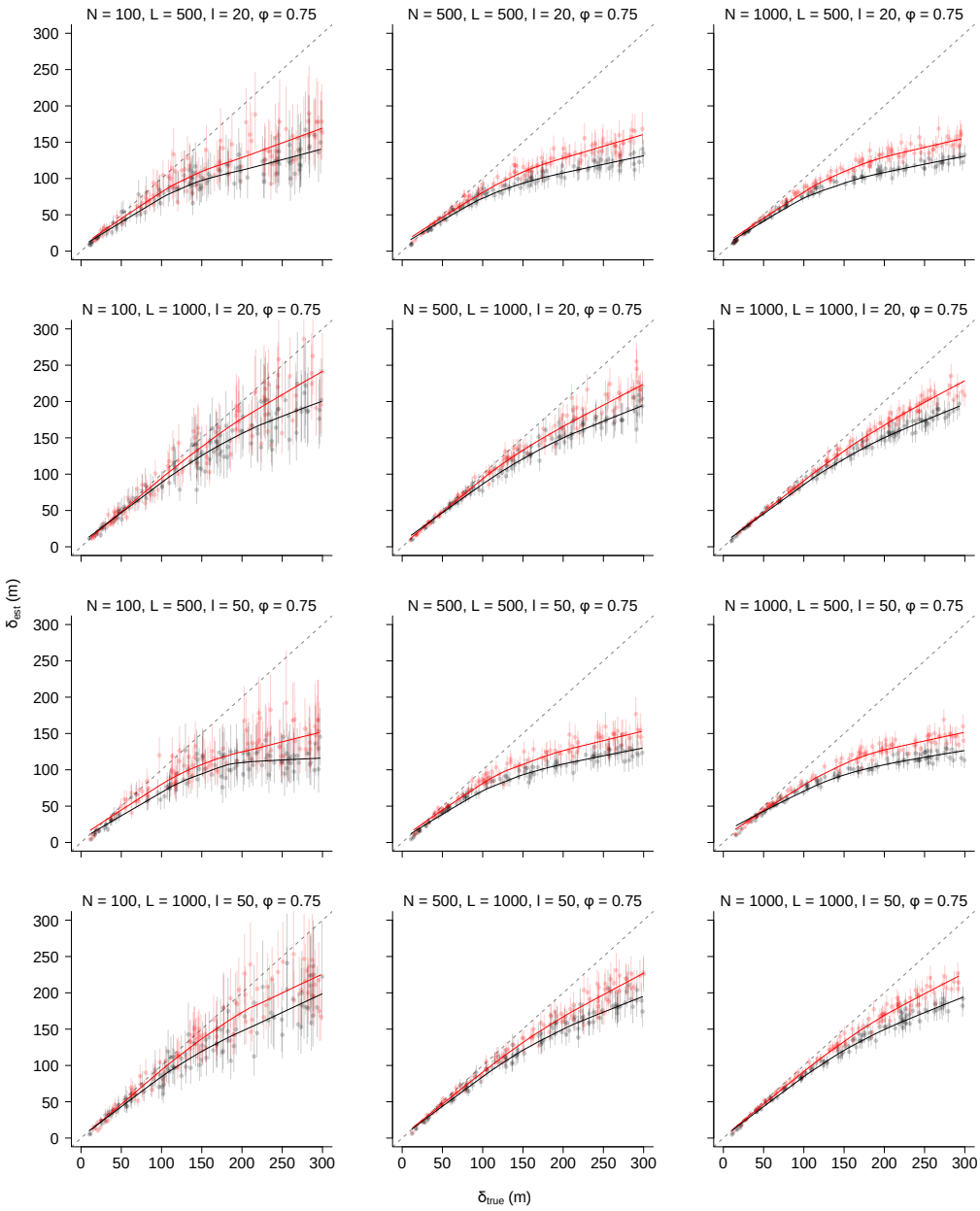


Figure S2 The relationship between true (*y*-axis) and estimated (*x*-axis) average dispersal distances when the true recapture probability $\phi_{true} = 0.75$. See Figure S1 for captions.

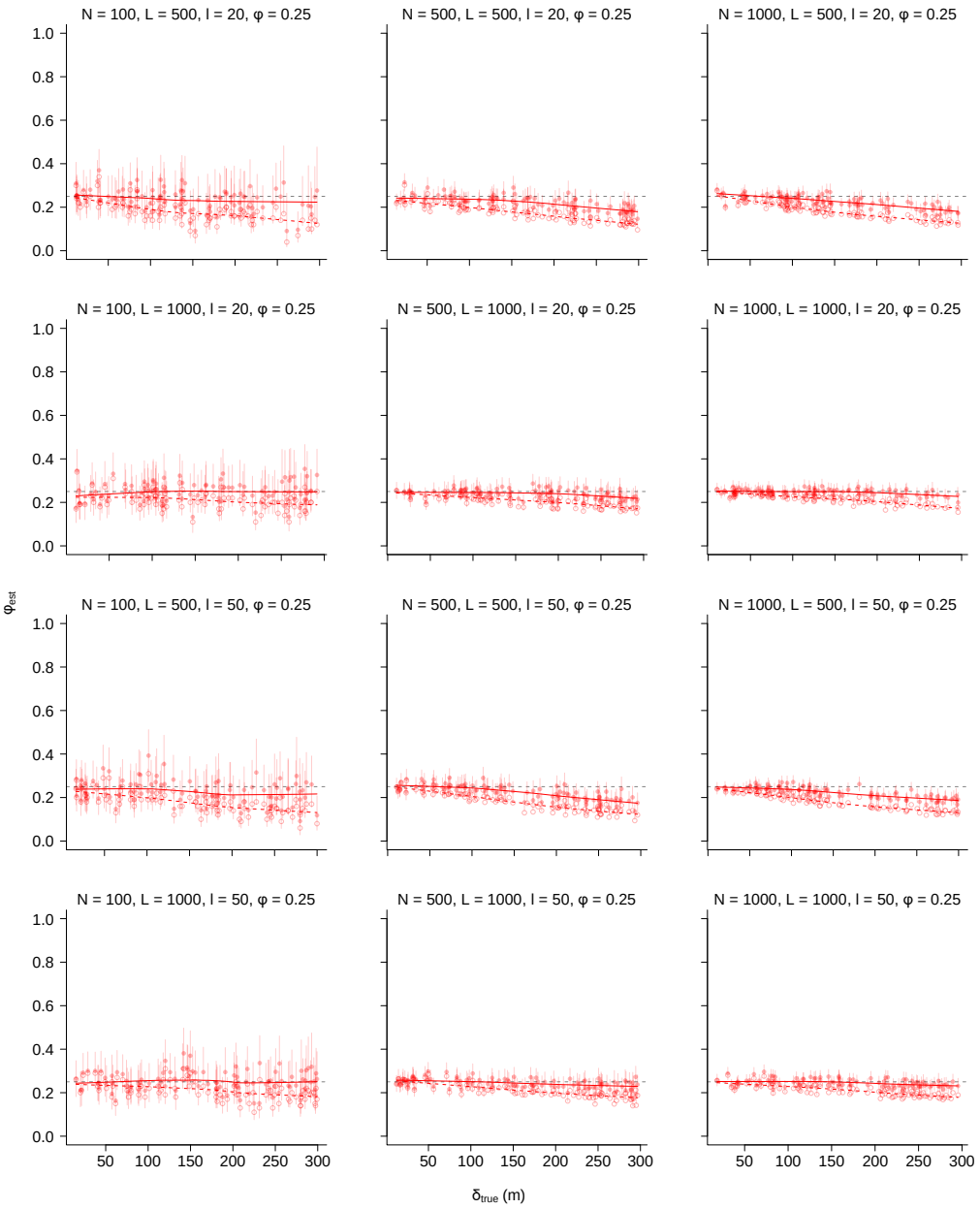


Figure S3 The relationship between the estimated recapture probability ϕ_{est} and true average dispersal distance δ_{true} when the true recapture probability $\phi_{true} = 0.25$. Red filled points are the median estimates from the DOCM while red open points denoting the proportion of individuals recaptured for each test dataset. Error bars are 95% credible intervals. Grey broken lines denote the true recapture probability ϕ_{true}. Different panels are estimates under different sampling designs, and the values of sampling design factors are shown on the top of each panel. Solid and broken red lines are smooths for the DOCM estimates and the proportion of individuals recaptured, respectively.

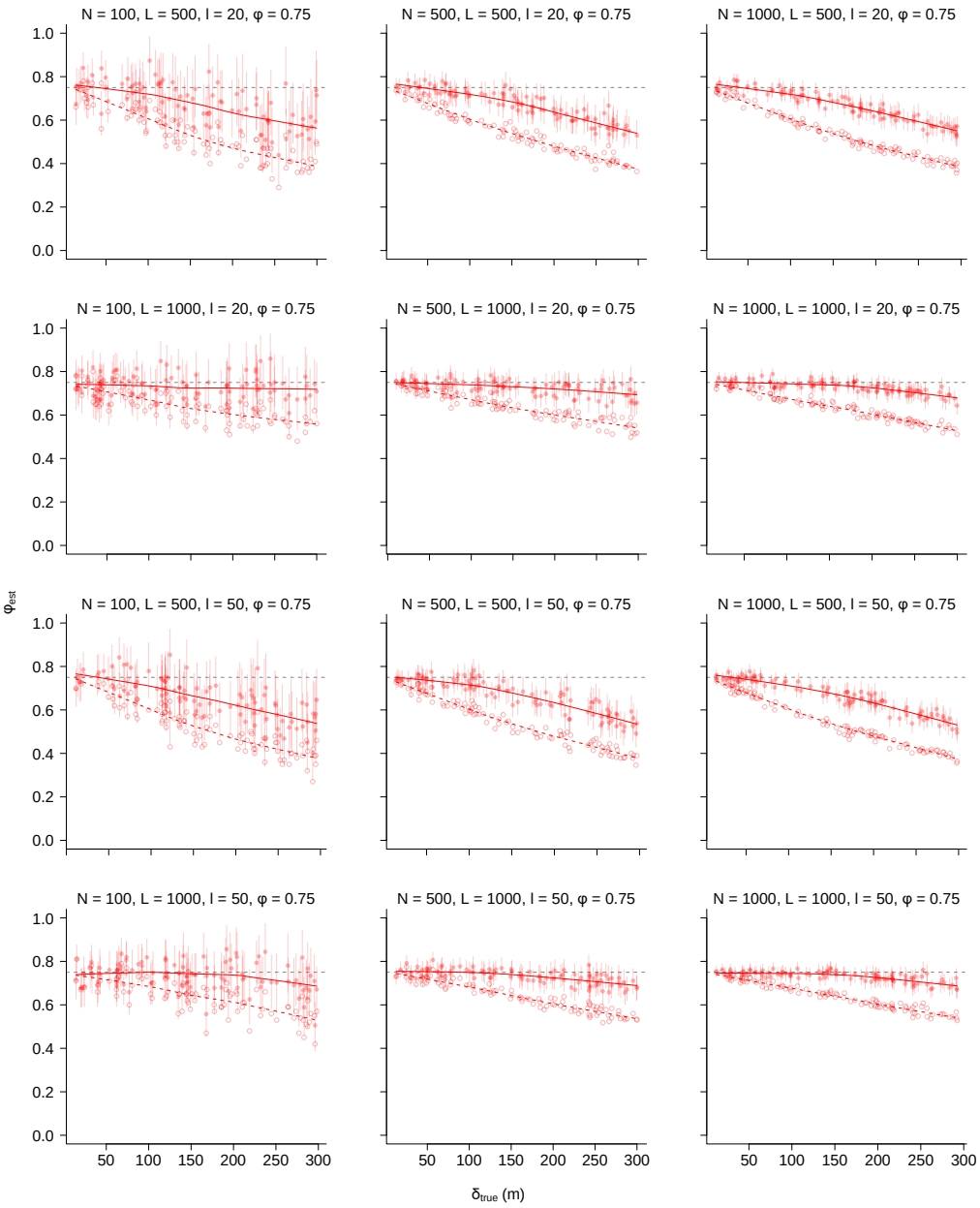


Figure S4 The relationship between the estimated recapture probability ϕ_{est} and true average dispersal distance δ_{true} when the true recapture probability $\phi_{true} = 0.75$. See Figure S3 for captions.