

## Quantifying seed dispersal kernels from truncated seed-tracking data

Ben T. Hirsch<sup>1,2\*</sup>, Marco D. Visser<sup>2,3</sup>, Roland Kays<sup>1,2</sup> and Patrick A. Jansen<sup>2,4,5</sup>

<sup>1</sup>New York State Museum, 3140 CEC, Albany, NY 12230, USA; <sup>2</sup>Smithsonian Tropical Research Institute, Unit 9100, Box 0948, DPO AA 34002-9898, USA; <sup>3</sup>Department of Experimental Plant Ecology, Radboud University Nijmegen, Heyendaalseweg 135, 6525 AJ Nijmegen, The Netherlands; <sup>4</sup>Centre for Ecosystem Studies, Wageningen University and Research Centre, PO Box 47, 6700 AA Wageningen, The Netherlands; and <sup>5</sup>Community and Conservation Ecology Group, University of Groningen, PO Box 11103, 9700 CC Groningen, The Netherlands

### Summary

1. Seed dispersal is a key biological process that remains poorly documented because dispersing seeds are notoriously hard to track. While long-distance dispersal is thought to be particularly important, seed-tracking studies typically yield incomplete data sets that are biased against long-distance movements.

2. We evaluate an analytical procedure developed by Jansen, Bongers & Hemerik (2004) to infer the tail of a seed dispersal kernel from incomplete frequency distributions of dispersal distances obtained by tracking seeds. This ‘censored tail reconstruction’ (CTR) method treats dispersal distances as waiting times in a survival analysis and censors nonretrieved seeds according to how far they can reliably be tracked. We tested whether CTR can provide unbiased estimates of long-distance movements which typically cannot be tracked with traditional field methods.

3. We used a complete frequency distribution of primary seed dispersal distances of the palm *Astrocaryum standleyanum*, obtained with telemetric thread tags that allow tracking seeds regardless of the distance moved. We truncated and resampled the data set at various distances, fitted kernel functions on CTR estimates of dispersal distance and determined how well this function approximated the true dispersal kernel.

4. Censored tail reconstruction with truncated data approximated the true dispersal kernel remarkably well but only when the best-fitting function (lognormal) was used. We were able to select the correct function and derive an accurate estimate of the seed dispersal kernel even after censoring 50–60% of the dispersal events. However, CTR results were substantially biased if 5% or more of seeds within the search radius were overlooked by field observers and erroneously censored. Similar results were obtained using additional simulated dispersal kernels.

5. Our study suggests that the CTR method can accurately estimate the dispersal kernel from truncated seed-tracking data if the kernel is a simple decay function. This method will improve our understanding of the spatial patterns of seed movement and should replace the usual practice of omitting nonretrieved seeds from analyses in seed-tracking studies.

**Key-words:** censored tail reconstruction, censored tail reconstruction, kernel, long-distance dispersal, seed dispersal, seed tracking, thread tag

### Introduction

Seed dispersal is an important process affecting population dynamics, gene flow, species diversity and biological invasions of plants (Janzen 1970; Connell 1971; Nathan & Muller-Lan-

dau 2000; Wright 2002; Jansen, Bongers & van der Meer 2008). In particular, seeds that disperse over relatively short distances typically have lower survival than those that disperse further away from conspecifics (Janzen 1970; Comita *et al.* 2010; Mangan *et al.* 2010). Describing the probability distribution of dispersal distances, the so-called dispersal kernel is crucial for understanding these biological processes (Nathan & Muller-Landau 2000; Jongejans, Skarpaas & Shea 2008).

\*Correspondence author. E-mail: hirschb@si.edu  
Correspondence site: <http://www.respond2articles.com/MEE/>

Estimates of long-distance dispersal (LDD), the tail of the dispersal kernel, are particularly important for modelling species invasions or predicting species responses to climate and habitat change (Clark *et al.* 1999; Cain, Milligan & Strand 2000; Levin *et al.* 2003).

A principal reason why seed dispersal remains relatively poorly understood is that dispersing seeds are notoriously hard to follow (Wang & Smith 2002). It is often difficult or impossible to individually tag and follow seeds because they are so small. Researchers typically use seed marking methods such as thread tags or radioisotopes and then attempt to relocate tagged seeds after dispersal (reviewed in: Forget & Wenny 2005). Although this has resulted in a much better understanding of when and how far seeds disperse, a certain proportion of the dispersed seeds in these studies are never recovered, thus their dispersal distances remain unrecorded (Wang & Smith 2002). Because seeds that disperse relatively far are the most difficult to find, LDD events are the least likely to be observed. The resulting bias against far-dispersed seeds is problematic because it misses the tail of the seed dispersal kernel, which is the most important portion of the distribution (Bullock & Clarke 2000). Seeds that disperse far are important for a host of ecological and evolutionary processes such as the spread of invasive species, metapopulation dynamics and maintenance of diversity (Portnoy & Willson 1993, Cain, Milligan & Strand 2000; Nathan *et al.* 2003; Soons & Bullock 2008).

In most seed-tracking studies, researchers search for dispersed tagged seeds within a predefined radius of the point of release (Fleming & Heithaus 1981; Howe 1990; De Steven 1994; Fragoso 1997; Jansen, Bongers & Hemerik 2004; Jansen, Bongers & van der Meer 2008). Seeds that disperse further than this search radius are automatically lost. Typically, these seeds are classified as 'missing' and omitted from the data set, thus dispersal distances greater than the search radius are not represented in the data set. Jansen, Bongers & Hemerik (2004) developed an approach for overcoming this problem, but this method has never been validated. This method, henceforth called 'censored tail reconstruction' (CTR) uses survival analysis to estimate the entire dispersal kernel based on the pattern observed at the beginning of the distribution. Instead of omitting missing seeds, this analysis assumes that all missing seeds dispersed beyond the search radius. Dispersal distances are treated as waiting times, while missing seeds are treated as observations censored at the search radius. The full dispersal kernel is then estimated by fitting a cumulative function to Kaplan–Meier probability estimates of dispersal distance.

Jansen, Bongers & Hemerik (2004) used the CTR method to estimate the dispersal kernel of *Carapa procera* seeds dispersed by scatter-hoarding rodents in French Guiana from data on thread-tracked seeds with incomplete recovery. To date, these two studies remain the only published applications of this method. Although the CTR method is arguably superior to the common practice of simply omitting missing seeds from a data set, the accuracy of this method in providing credible estimates of dispersal distances has not been tested. Jansen, Bongers & Hemerik (2004) fitted a Weibull function to the survivorship curves, but it is unknown whether this is the best-

fitting function in general. It is also unknown how sensitive the CTR method is to falsely censored seeds, for example seeds overlooked by field researchers within the search radius.

To test whether the CTR method produces accurate estimates of seed dispersal distributions, we used an existing, unpublished data set: the complete frequency distribution of seed dispersal distances for the rodent-dispersed palm *Astrocaryum standleyanum* (hereafter: *Astrocaryum*), which was obtained through seed tracking with telemetric thread tags (B.T. Hirsch, P.A. Jansen & R. Kays submitted). We truncated the data set at various distances to mimic search distances used in the field, fitted different dispersal kernel functions through CTR, used  $\Delta$ AIC values to select the best-fitting function and determined how well this function approximated the observed full dispersal kernel. Additionally, we quantified the effect of function selection and falsely censored seeds on the CTR results. These tests allowed us to evaluate the overall robustness of the CTR method and make recommendations about study design. Finally, we used the CTR method in conjunction with simulated dispersal kernels to test whether it can be used in studies of plant species with different shaped dispersal kernels.

## Methodology

### SEED DISPERSAL DATA

The data set with seed dispersal distances that we used for our test was collected on Barro Colorado Island (BCI), Panama, a 1560-ha island protected and administered by the Smithsonian Tropical Research Institute (9°10'N, 79°51'W). BCI is covered with primary and secondary semi-deciduous moist tropical forest. Annual rainfall averages 2600 mm with an average temperature of 27 °C. The dry season generally lasts from mid-December to May (Terrestrial-Environmental Sciences Program of the Smithsonian Tropical Research Institute).

The study species, *Astrocaryum standleyanum*, is a Neotropical arborescent palm occurring from Costa Rica to Ecuador. Trees annually produce 3–6 pendulous infructescences with up to 1500 ovoid fruits in total. The local fruiting period is from March to the beginning of July (De Steven *et al.* 1987). The fresh weight of the 2–3 cm seeds averages 9.6 g (Wright *et al.* 2010). *Astrocaryum* depends on scatter-hoarding by rodents for seed dispersal, in particular on agoutis *Dasyprocta punctata* (Smythe 1989; Galvez *et al.* 2009), 2–4 kg caviomorph rodents that bury the seeds in the soil as food reserves for periods of food scarcity (Smythe 1978, 1989).

A complete frequency distribution of seed dispersal distances of *Astrocaryum* was obtained by placing 589 tagged seeds at 52 seed stations across a ~25-ha area in the centre of BCI. Each seed had a telemetric thread tag and a black nylon-coated stainless steel leader wire tied to a 3.8-g radiotransmitter with 20-cm wire antenna (Advanced Telemetry Systems, Isanti, MN, USA; Hirsch *et al.* submitted). Dispersal distance was measured of 423 seeds removed from seed stations by animals, and no differences in removal rate or dispersal distance were found between seeds with and without transmitters

(Hirsch *et al.* submitted). Removed seeds were located by sight or with a hand-held telemetry receiver (Yaesu VR-500) and three-element Yagi antenna. The transmitter was occasionally bitten off of the seed, but 97% of seed tags were recovered intact (Hirsch *et al.* submitted). Dispersal distance and direction were measured with measuring tape and a compass. If a seed dispersed more than 20 m, a hand-held GPS receiver (Garmin 60CSx GPS) was used to measure the dispersal distance. We used the primary dispersal distances obtained from the above study to formulate our empirical dispersal kernel (e.g. no secondary dispersal events were included).

#### DISPERSAL KERNEL FITTING

We fitted four commonly used dispersal kernels in their one-dimensional form (i.e. probability density functions) directly to the distribution of dispersal distances from the 417 radio-tracked seeds: (i) lognormal, (ii) Weibull, (iii) exponential and (iv) 1DT (Table 1). All are simple decay functions in which larger dispersal distances are less frequent than any shorter dispersal distance, as is commonly assumed in seed dispersal studies. We used the function *optim* in R 2.10 (R Development Core Team 2010) to search for the parameter values in each of the four probability density functions that maximized the likelihood  $L$  of the observed distances ( $d$ );

$$L(d | p) = \prod_{i=1}^{i=n} f(d_i, p),$$

where  $d$  is a vector of  $n$  observed dispersal distances,  $p$  a set of parameters corresponding to one of the probability density function  $f$ . We used Akaike information criterion (AIC; Akaike 1974) to determine which function fitted the observed data best.

#### CENSORED TAIL RECONSTRUCTION

The CTR method (cf. Jansen, Bongers & Hemerik 2004; Jansen, Bongers & van der Meer 2008) uses survival analy-

**Table 1.** Dispersal kernel functions fit to empirical seed dispersal distances in the palm *Astrocaryum standleyanum*, ranked by fit.  $\Delta$ AIC values denote the difference in AIC scores between the current model and the best-fitting model

Rank	Function	Model	$\Delta$ AIC
1	Lognormal <sup>*,†,‡</sup>	$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$	0.00
2	Weibull <sup>§,¶</sup>	$f(x) = \begin{cases} \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}; & x \geq 0 \\ 0; & x < 0 \end{cases}$	19.70
3	Exponential <sup>  ,***</sup>	$f(x) = \lambda e^{-\lambda x}$	28.17
4	1DT <sup>‡,††,‡‡</sup>	$f(x) = \frac{2xb}{a} \left(1 + \frac{x^2}{a}\right)^{-b+1}$	38.82

\*Greene & Johnson (1989).

†Skarpaas, Shea & Bullock (2005).

‡Bullock, Shea & Skarpaas (2006).

§Jansen, Bongers & van der Meer (2008).

¶Tufto, Engen & Hindar (1997).

\*\*Austerlitz *et al.* (2004).

††Jones & Muller-Landau (2008).

‡‡Clark *et al.* (1999).

sis to estimate the dispersal kernel, assuming that missing seeds have travelled beyond the search radius. CTR treats the retrieval of a dispersed seed as an event, observed dispersal distances as failure times, and missing seeds as events censored at a given dispersal distance, that is, the radius of the area in which dispersed seeds were searched. Kaplan–Meier survival analysis is used to calculate the survivorship function to which a standard dispersal kernel can then be fitted and used to predict the tail of the distribution.

The steps used in CTR are:

1. Collect data on seed dispersal distance using thread tags or similar methods as appropriate for the study system. The search radius and the number of seeds lost (i.e. moving further than this distance) should be recorded.
2. Estimate the Kaplan–Meier survivorship curve, treating dispersal distance as time, and including all missing seeds as observations censored at the search radius.
3. Fit probability density functions to the K–M survivorship curve. We test four functions used in previous studies here, but any other appropriate decay function can be used.
4. Use the AIC selection procedure to determine which probability function best fits the data.

We provide an example R-code which can be used as a guide to conduct CTR analyses in Appendix S1. The R-code uses the packages ‘survival’ (Therneau & Lumley 2009) and ‘fdr-tool’ (Strimmer 2011).

To estimate the accuracy of CTR, we estimated the difference between CTR-derived dispersal kernels fitted on truncated data and the ‘true’ empirical dispersal kernel for *Astrocaryum*. Truncated data were obtained by assuming that all *Astrocaryum* seeds dispersed beyond a given search radius were missing. We then compared the average distance of the 95% percentile of the dispersal kernel between the CTR and empirical results. We used the 95th percentile criteria as a measure for LDD because of its use in previous studies (e.g. Nathan *et al.* 2003).

We evaluated the sensitivity of the CTR method to three potential sources of bias: (i) the probability density function used, (ii) the size of the search radius (or proportion of seeds which fall within the search radius) and (iii) the proportion of seeds overlooked by observers within the search radius (falsely censored seeds). We estimated confidence intervals ( $\alpha = 0.05$ ) for each measure of bias (detailed below) using a nonparametric bootstrap (Efron & Tibshirani 1993). Confidence intervals were calculated as the 2.5 and 97.5 percentiles of the bootstrapped estimates.

#### Function selection

We tested the sensitivity of CTR for function selection by comparing dispersal distance estimates derived with the CTR method for each of four previously defined probability density functions. These four functions were chosen because they have commonly been used in prior studies of seed dispersal (Table 1). We compared the CTR-derived distance of the

95% percentile of the dispersal kernel using four mathematical functions vs. the empirical results. Through bootstrapping we also tested how often AIC yielded the true (or nontruncated) dispersal model among the four candidate models fit to truncated data.

### Search radius

A typical search radius used in previous studies is 20 m (Howe 1990; De Steven 1994); however, it is unknown whether using such a radius with the CTR method can yield accurate results. Here we created multiple truncated data sets based on the empirical *Astrocaryum* dispersal kernel with search radii ranging between 1.6 and 134.5 m which corresponded to 0–90% of seeds falling outside the search radius. We then tested the effectiveness of the CTR method using these various search radii (19 different radii were evaluated in total, each increasing radius corresponded to a 5% increase in the proportion of seeds recovered). We used the difference between the observed ‘true’ 95th percentile of dispersal distance and the CTR-derived results as an estimate of bias. Bias ( $\varepsilon$ ) was calculated as  $\varepsilon = \mu_{\text{ctr}} - \mu$ , where  $\mu$  is the ‘true’ observed LDD distance and  $\mu_{\text{ctr}}$  is the CTR-derived LDD measure. Here we report the absolute proportional bias  $|\varepsilon|/\mu$ .

### False-censoring or overlooking seeds

The CTR method is based on the assumption that all seeds not recovered within the search radius were dispersed beyond this radius. To determine how robust the CTR method is to violations of this assumption, we evaluated bias when 0–50% of the seeds were overlooked (in 11 equally spaced steps, each step corresponding to a 5% increase in the proportion of overlooked seeds). This was done by randomly removing a given percentage of the seeds from a truncated data set (truncated at 20 m) and treating them as censored. We used the same measure of bias ( $|\varepsilon|/\mu$ ) for this analysis.

### ROBUSTNESS ANALYSIS

To evaluate how sensitive CTR is to the specific shape of the distribution, we ran the above analyses for a variety of simulated seed distributions that had the same sample size and scale as the empirical distribution (Appendix S2). We used a Monte Carlo type simulation to generate dispersal distributions through random number generation from each of the four tested probability density functions.

## Results

Primary dispersal distances of the 417 *Astrocaryum* seeds ranged between 0.15 and 132.5 m (mean 14.7 m, median 7.5 m). Of the four dispersal kernel functions, the lognormal fitted the observed dispersal data best (Table 1, Fig. 1). Compared with the estimated 95th quantile distance calculated from the empirical data set (58.6 m), the lognormal distribution was more

accurate (estimate = 58.4 m, 95% CI = 44.4–74.8 m) than the Weibull (34.7 m, 34.7–37.5 m), exponential (34.9 m, 32.0–37.6 m) and 1DT distributions (97.6 m, 73.9–139.8 m) (Fig. 2). These analyses demonstrate that the CTR method is highly sensitive to the mathematical function that is fitted on the Kaplan–Meier survivorship estimates. When the best-fitting model (lognormal) was used with the CTR method, the derived results were very similar to the empirical results (Fig. 2).

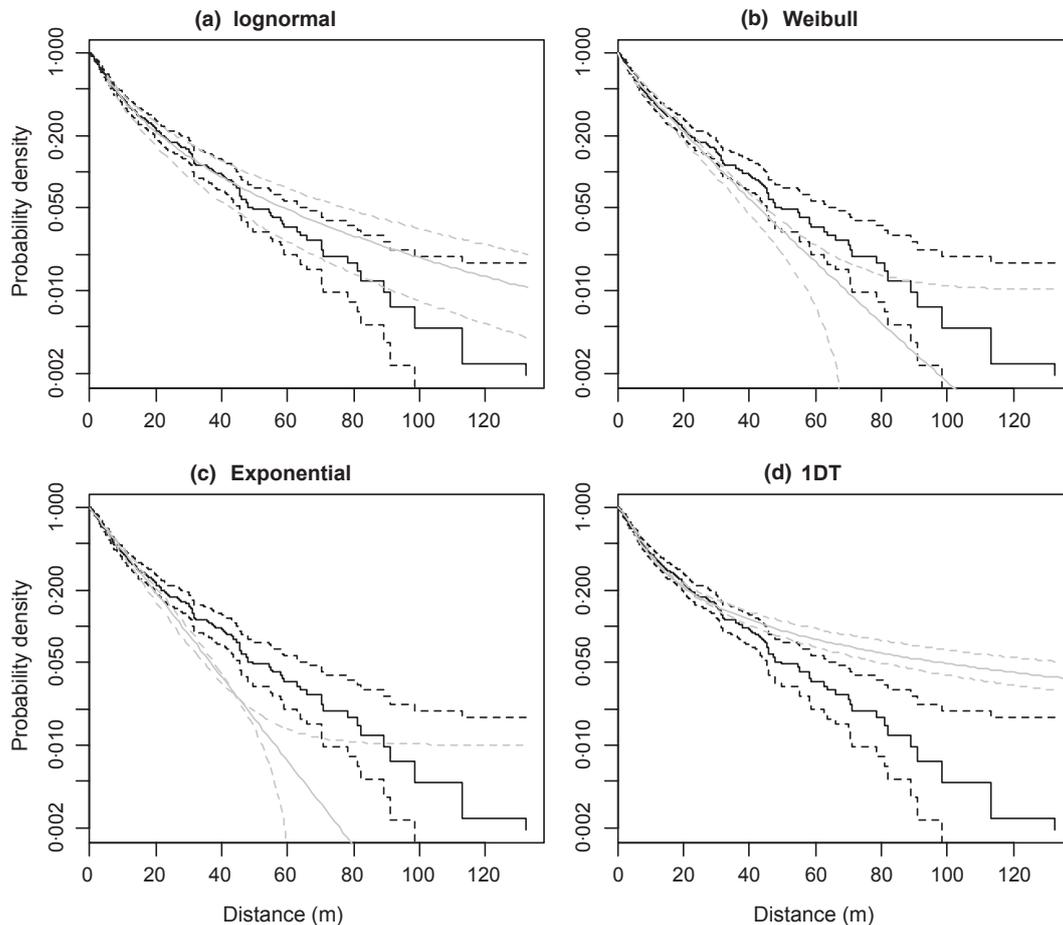
Given the importance of model choice, the AIC approach was an important step in selecting the best model for our CTR-derived data set. The lognormal function, which gave the best overall fit to the full observed data set and provided the least bias, was selected 97% of times based on its  $\Delta\text{AIC}$  score (after 1000 resampled data sets). Similar results were obtained with simulated data (Table S1). In the simulation, a noncritical issue arose when the curve could be approximated equally well by two different models (for example, the Weibull with a shape parameter of 1 is the equivalent of the exponential), but model predictions were essentially identical in such cases. These results indicate that the AIC model selection can effectively select the CTR-derived function that corresponds with the true dispersal kernel in the evaluated cases.

### THE EFFECT OF SEARCH RADIUS

Bias rapidly decreased as a higher proportion of seeds are recovered (i.e. search radius increased) up to a point, and then levelled off (Fig. 3). In fact, there is little improvement in the estimate when more than 50% of seeds are recovered, which would have been accomplished with a 7.5-m search radius for *Astrocaryum* seeds. The simulated results also show acceptable bias when a large proportion of the seeds are recovered ( $\geq 50\%$ ), which suggests that this result is robust (Fig. S1).

### THE EFFECT OF OVERLOOKING SEEDS (FALSE-CENSORING)

The CTR method was highly sensitive to false-censoring of seeds within the search radius (Fig. 4). The CTR method worked well when  $< 5\%$  of the seeds that dispersed within the search radius were overlooked, but bias was substantial when larger proportions were overlooked. For example, starting from the 5% threshold of falsely censored seeds, the proportional bias increases exponentially from 29.4% (Fig. 4.) Simulations showed that the strength of the bias due to falsely censored seeds depended on the ‘fatness’ of the tail of the distribution (the proportion of seeds dispersing long distances); ‘fat-tailed’ distributions, such as the lognormal, were relatively sensitive to overlooked seeds (Fig. S2). This shows that the assumption of the CTR method that all nonrecovered seeds dispersed beyond the search radius is critical and that the effect of violating this assumption is much greater than the effect of using a relatively small search radius.



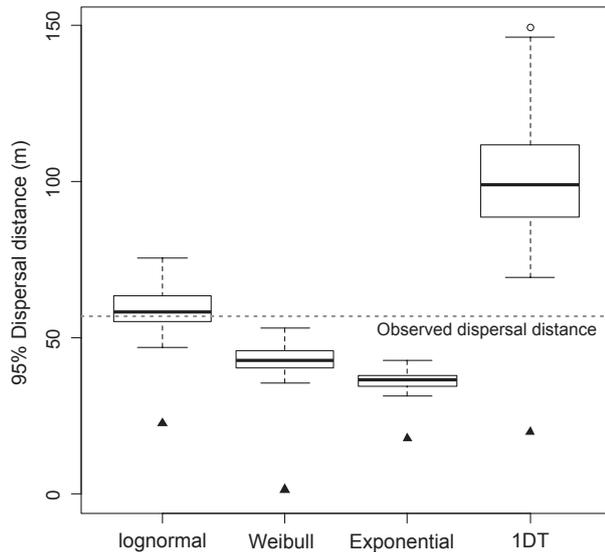
**Fig. 1.** Alternative dispersal kernels of *Astrocaryum standleyanum* (grey lines), fitted to incomplete seed dispersal data through censored tail reconstruction (CTR), compared with the Kaplan–Meier survivorship curve showing the true distribution of seed dispersal distances (black lines). The curves show the probability (or proportion) of seeds dispersing beyond any given distance. In these examples, the alternative dispersal kernels were fitted to data sets truncated at 20 m ( $\approx 75\%$  seeds recovered). Solid lines indicate median estimates, and dashed lines indicate the 95% confidence intervals. The lognormal distribution showed the best fit to the full data set (Table 1) and provided the best approximation of the dispersal distances. Note the log-scale of the probability.

## Discussion

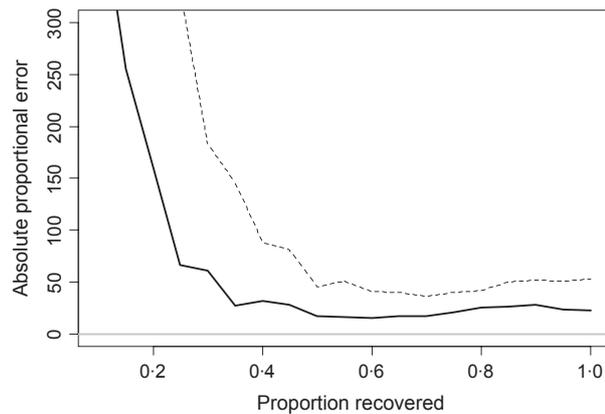
Seed-tracking studies typically classify nonrecovered seeds as missing observations, which produces an inherent bias against longer distance seed movement. Here we used a full-dispersal data set, obtained with telemetric seed tags, to evaluate an alternative method for handling these nonrecovered seeds: CTR (Jansen, Bongers & Hemerik 2004; Jansen, Bongers & van der Meer 2008). We found that the CTR method can produce excellent approximations of the true dispersal kernel as long as 50% or more of the seeds dispersed are recovered and  $< 5\%$  of the seeds dispersed within the search radius are overlooked (Fig. 3). In all cases evaluated, the CTR method approximated the true dispersal kernel better than the standard practice of omitting nonretrieved seeds from the data set.

The ability to accurately predict dispersal distance at a given percentile using the CTR method was greatly affected by the choice of function that was fitted to the survival estimates. However, even when using truncated data, it was generally possible to choose the ‘correct’ function with the use of the

AIC selection method. This appears to be independent of the shape of the kernel, as demonstrated in our simulation results (Table S1). We advise researchers to take care in selecting a set of dispersal models from which to conduct model selection as appropriate functions will vary across dispersal systems. Also note that in some systems, complex multimodal kernels exist (e.g. Russo, Portnoy & Augspurger 2006). These cannot be described with commonly used simple (decay with distance) seed dispersal functions (Cousens, Dytham & Law 2008). Researchers should also be aware that problems can arise when estimating 2D seed density using the best-fitting kernel obtained from 1D data, as not all 1D models are equally appropriate for translation to two dimensions. Depending on the precise mathematical formulation, some 1D kernels (e.g. the exponential) that allow for nonzero predictions at the origin will result in infinite densities at the point zero ( $a/2\pi r = \infty$  when  $r = 0$ ) when translated to 2D. A list of suitable two-dimensional density kernels which can be freely translated from 1 to 2 dimensions are listed in the study by Cousens, Dytham & Law (2008); table 5.2).

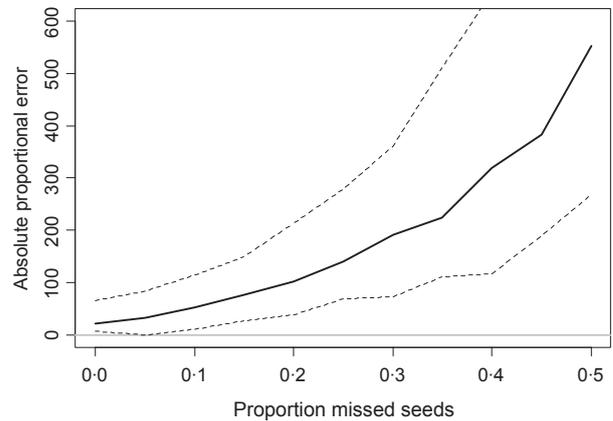


**Fig. 2.** Estimates of the tail of the dispersal kernel of *Astrocarium standleyanum* obtained through censored tail reconstruction (CTR) with incomplete seed dispersal data for four alternative kernel functions, compared with the empirically observed value. Values shown are the average 95th percentile dispersal distances after 1000 bootstraps for each of the functions. Estimated 95th percentile dispersal distances when not conducting a CTR correction are shown as solid black triangles.



**Fig. 3.** Effect of censoring on the accuracy of censored tail reconstruction (CTR). Shown is the proportional error ( $100 \times \text{abs}\{\varepsilon/\mu\}$ ) as a function of the proportion of seeds recovered within the search radius. Lines indicate median error after 1000 bootstraps (solid line) and 95% CI (dashed lines). In the case of *Astrocarium standleyanum*, the proportions 0.2, 0.4, 0.6 and 0.8 correspond to the fraction of seeds dispersed at least 3.3, 6.8, 12 and 21 m.

The CTR method worked surprisingly well even when sampling a small part of the full dispersal kernel, producing accurate results even when 50% of seeds fell outside the search radius. For the *Astrocarium* in our study, this could have been met with a 7.5-m search radius (assuming no seeds are overlooked). Given that the shape of seed dispersal kernels can vary between years and between species (Greene *et al.* 2004), we recommend researchers choose a search radius that includes at least 50% of their tagged seeds. We also encourage further tests



**Fig. 4.** Effect of false-censoring on the accuracy of censored tail reconstruction (CTR). Error is shown as a function of the proportion of seeds that is overlooked within the search radius. Lines indicate median error after 1000 bootstraps (solid line) and 95% CI (dashed lines).

of the CTR method on different plant species and in systems where dispersal occurs at different spatial scales. Even if the 50% cut-off cannot be applied to any and all study systems, our results provide guidelines for the experimental design of future seed-tracking studies. Our results suggest that the traditional search radius of 20–30 m is sufficient for use with the CRT method if seed dispersal is on a similar scale as *Astrocarium*.

We found that error resulting from falsely censored seeds within the search area is a much larger concern than the proportion of seeds censored. Overlooked seeds can greatly distort the results, and the accuracy of the CTR method is extremely sensitive to these observer errors. If seeds in a given study system are easily overlooked, or if the search radius is too large to efficiently find >95% of seeds that fall within the area, the CTR method could lead to large overestimations of long distance dispersal. In addition, if the seeds in a given study system are completely destroyed when eaten, this may have a similar effect as overlooked seeds. The CTR method can only be used in systems where eaten seeds can be retrieved or where seeds are never immediately consumed. Depending on how easy it is to overlook seeds with a particular tracking method, a trade-off could exist between the size of the search area and the amount of overlooked seeds. We suggest that researchers choose a search radius and tracking method that yields low rates of overlooked seeds. We also feel that it would be useful for researchers to empirically test the efficiency of their field crew in detecting seeds to ensure that they are within the range recommended by our sensitivity analysis.

Comparing CTR-derived estimates of seed dispersal kernels vs. the true kernel showed that the CTR method can accurately estimate the dispersal kernel using truncated seed-tracking data. It should also be possible to reanalyse data from previously published studies to extract complete dispersal kernels, provided that the search radius is reported and that the search was full and reliable. Our results show that the CTR method can be used in conjunction with standard tagging methods to

adequately approximate complete seed dispersal kernels by collecting enough data over a smaller area to characterize the scale and shape of the relationship. For example, CTR would be ideal in conjunction with radioisotope labelling because Geiger counters allow retrieval of a very high proportion of cached seeds in a given area (Vander Wall 1997). These radioisotope labels can also be used to recover the seed coat of eaten seeds. Low-tech seed-tagging methods such as thread tags and fluorescent marking are typically much more economically feasible than methods that allow the measurement of complete seed dispersal kernels, such as genetics or radiotelemetry. Using these methods along with the CTR would allow researchers around the globe to obtain credible dispersal kernels from more plant species, thus extending our understanding of seed dispersal, and plant ecology in general. CTR should entirely replace the traditional practice of simply ignoring missing seeds.

## Acknowledgements

We thank Eelke Jongejans and two anonymous reviewers for valuable comments to an earlier version of the manuscript. This study was supported by funding from the National Science Foundation (NSF-DEB 0717071 to RWK) and the Netherlands Organization for Scientific Research (grants W85-239 and 863-07-008 to P.A.J.). M.D.V. acknowledges funding from the Smithsonian Tropical Research Institute fellowship programme.

## References

- Akaike, H. (1974) New look at statistical-model identification. *IEEE Transactions on Automatic Control*, **19**, 716–723.
- Austerlitz, F., Dick, C.W., Dutech, C., Klein, E.K., Oddou-Muratario, S., Smouse, P.E. & Sork, V.L. (2004) Using genetic markers to estimate the pollen dispersal curve. *Molecular Ecology*, **13**, 937–954.
- Bullock, J.M. & Clarke, R.T. (2000) Long distance seed dispersal by wind: measuring and modeling the tail of the curve. *Oecologia*, **124**, 506–521.
- Bullock, J.M., Shea, K. & Skarpaas, O. (2006) Measuring plant dispersal: an introduction to field methods and experimental design. *Plant Ecology*, **186**, 217–234.
- Cain, M.L., Milligan, B.G. & Strand, A.E. (2000) Long-distance seed dispersal in plant populations. *American Journal of Botany*, **87**, 1217–1227.
- Clark, J.S., Silman, M., Kern, R., Macklin, E. & HilleRisLambers, J. (1999) Seed dispersal near and far: patterns across temperate and tropical forests. *Ecology*, **80**, 1475–1494.
- Comita, L.S., Muller-Landau, H.C., Aguilar, S. & Hubbell, S.P. (2010) Asymmetric density dependence shapes species abundance in a tropical tree community. *Science*, **329**, 330–332.
- Connell, J.H. (1971) On the role of natural enemies in preventing competitive exclusion in some marine mammals and in rain forest trees. *Dynamics of Populations* (eds P.J. Boer & G.R. Gradwel), pp. 298–310. PUDOC, Wageningen.
- Cousens, R., Dytham, C. & Law, R. (2008) *Dispersal in Plants*. Oxford University Press, Oxford.
- De Steven, D. (1994) Tropical tree seedling dynamics: recruitment patterns and their population consequences for three canopy species. *Journal of Tropical Ecology*, **10**, 369–383.
- De Steven, D., Windsor, D.M., Putz, F.E. & de Leon, B. (1987) Vegetative and reproductive phenologies of a palm assemblage in Panama. *Biotropica*, **19**, 342–356.
- Efron, B. & Tibshirani, R. (1993) *An Introduction to the Bootstrap*. Chapman & Hall, Boca Raton, Florida.
- Fleming, T.H. & Heithaus, E.R. (1981) Frugivorous bats, seed shadows, and the structure of tropical forests. *Biotropica*, **13**, 45–53.
- Forget, P.-M. & Wenny, D.G. (2005) A review of methods used to study seed removal and secondary seed dispersal. *Seed Fate: Predation, Secondary Dispersal, and Seedling Establishment* (eds P.-M. Forget, J.E. Lambert, P.E. Hulme & S.B. Vander Wall), pp. 379–394. CAB International, Wallingford.
- Fragoso, J.M.V. (1997) Tapir-generated seed shadows: scale-dependent patchiness in the Amazon rain forest. *Journal of Ecology*, **85**, 519–529.
- Galvez, D., Kranstauer, B., Kays, R.W. & Jansen, P.A. (2009) Scatter hoarding by the Central American agouti: a test of optimal cache spacing theory. *Animal Behaviour*, **78**, 1327–1333.
- Greene, D.F. & Johnson, E.A. (1989) A model of wind dispersal of winged or plumed seeds. *Ecology*, **70**, 339–347.
- Greene, D.F., Canham, C.D., Coates, K.D. & Lepage, P.T. (2004) An evaluation of alternative dispersal functions for trees. *Journal of Ecology*, **92**, 758–766.
- Howe, H.F. (1990) Seed dispersal by birds and mammals: implications for seedling demography. *Reproductive Ecology of Tropical Forest Plants* (eds K.S. Bawa & M. Hadley), pp. 191–218. Man and the Biosphere Parthenon/UNESCO, Rome.
- Jansen, P.A., Bongers, F. & Hemerik, L. (2004) Seed mass and mast seeding enhance dispersal by a Neotropical scatter-hoarding rodent. *Ecological Monographs*, **74**, 569–589.
- Jansen, P.A., Bongers, F. & van der Meer, P.J. (2008) Is farther seed dispersal better? Spatial patterns of offspring mortality in three rainforest tree species with different dispersal abilities. *Ecography*, **31**, 43–52.
- Jansen, P.A., Elschot, K., Verkerk, P.J. & Wright, S.J. (2010) Seed predation and defleshing in the agouti-dispersed palm *Astrocaryum standleyanum*. *Journal of Tropical Ecology*, **26**, 1–8.
- Janzen, D.H. (1970) Herbivores and the number of tree species in tropical forests. *The American Naturalist*, **104**, 501–527.
- Jones, A.F. & Muller-Landau, H.C. (2008) Measuring long-distance seed dispersal in complex natural environments: an evaluation and integration of classical and genetic methods. *Journal of Ecology*, **96**, 642–652.
- Jongejans, E., Skarpaas, O. & Shea, K. (2008) Dispersal, demography and spatial population models for conservation and control management. *Perspectives in Plant Ecology, Evolution, and Systematics*, **9**, 153–170.
- Levin, S.A., Muller-Landau, H.C., Nathan, R. & Chave, J. (2003) The ecology and evolution of seed dispersal: a theoretical perspective. *Annual Review of Ecology, Evolution and Systematics*, **34**, 575–604.
- Mangan, S.A., Schnitzer, S.A., Herre, E.A., Mack, K., Valencia, M., Sanchez, E. & Bever, J.D. (2010) Negative plant-soil feedback predicts relative species abundance in a tropical forest. *Nature*, **466**, 752–756.
- Nathan, R. & Muller-Landau, H.C. (2000) Spatial patterns of seed dispersal, their determinants and consequences for recruitment. *Trends in Ecology & Evolution*, **15**, 278–285.
- Nathan, R., Perry, G., Cronin, J.T., Strand, A.E. & Cain, M.L. (2003) Methods for estimating long-distance dispersal. *Oikos*, **103**, 261–273.
- Portnoy, S. & Willson, M.F. (1993) Seed dispersal curves: behavior of the tail of the distribution. *Evolutionary Ecology*, **7**, 25–44.
- R Development Core Team (2010) *R: A Language and Environment for Statistical Computing*. R foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Russo, S.E., Portnoy, S. & Augspurger, C.K. (2006) Incorporating animal behavior into seed dispersal models: implications for seed shadows. *Ecology*, **87**, 3160–3174.
- Skarpaas, O., Shea, K. & Bullock, J.M. (2005) Optimizing dispersal study design by Monte Carlo simulation. *Journal of Applied Ecology*, **42**, 731–739.
- Smythe, N. (1978) The natural history of the Central American agouti (*Dasyprocta punctata*). *Smithsonian Contributions to Zoology*, **257**, 1–52.
- Smythe, N. (1989) Seed survival in the palm *Astrocaryum standleyanum*: evidence for dependence upon its seed dispersers. *Biotropica*, **21**, 50–56.
- Soons, M.B. & Bullock, J.M. (2008) Non-random seed abscission, long-distance wind dispersal and plant migration rates. *Journal of Ecology*, **96**, 581–590.
- Strimmer, K. (2011) *fdrtool: estimation and control of (local) false discovery Rates*. URL <http://CRAN.R-project.org/package=fdrtool>.
- Therneau, T. & Lumley, T. (2009) *Survival analysis, including penalized likelihood*. URL <http://CRAN.R-project.org/package=survival>.
- Tufto, J., Engen, S. & Hindar, K. (1997) Stochastic dispersal processes in plant populations. *Theoretical Population Biology*, **52**, 16–26.
- Vander Wall, S.B. (1997) Dispersal of singleleaf piñon pine (*Pinus monophylla*) by seedcaching rodents. *Journal of Mammalogy*, **78**, 181–191.
- Wang, B.C. & Smith, T.B. (2002) Closing the seed dispersal loop. *Trends in Ecology and Evolution*, **17**, 379–385.
- Wright, J.S. (2002) Plant diversity in tropical forests: a review of mechanisms of species coexistence. *Oecologia*, **130**, 1–14.
- Wright, J.S., Kitajima, K., Kraft, N.J.B., Reich, P.B., Wright, I.J., Bunker, D.E., Condit, R., Dalling, J.W., Davies, S.J., Diaz, S., Engelbrecht, B.M.J., Harms, K.E., Hubbell, S.P., Marks, C.O., Ruiz-Jaen, M.C., Salvador, C.M.

& Zanne, A.E. (2010) Functional traits and the growth-mortality trade-off in tropical trees. *Ecology*, **91**, 3364–3674.

Received 12 July 2011; accepted 29 November 2011  
Handling Editor: Robert Freckleton

## Supporting Information

Additional Supporting Information may be found in the online version of this article.

**Appendix S1.** Example R code for conducting a CTR analysis using generated data.

**Appendix S2.** Results from simulated distributions.

**Fig. S1.** Effect of search radius on the bias of the CTR method applied to four simulated datasets.

**Fig. S2.** Effect of overlooking seeds on the bias of the method applied to four simulated datasets.

**Table S1.** The use of AIC to identify the distribution of a truncated dataset showed high accuracy, except for equivalent models (Exponential-Weibull).

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

```

# Example R code for conducting a CTR analysis using generated data
# Hirsch, Ben T, Visser, Marco D, Kays, Roland W, Jansen, Patrick A.
# Nijmegen June 2011
# Revised August 2011

##### load dependancies#####
# Code requires package fdrtool & survival to be in library
# otherwise use e.g. install.packages("fdrtool") first
require(fdrtool);require(survival)

##### Create data #####
# Next step is to generate example data "radiotagged distance", stored as object
x
# set random seed
set.seed(2011)
# generate data from lognormal distribution, 500 tracked seeds
# meanlog=log(50), sdlog=log(3)
x=rlnorm(500,log(50),log(3))

# truncate data after 20 units to create "tracked distances"
# with 20 m search radius
xtrunc=x[x<20]

# Prepare data for CTR
CTRdata=data.frame(
# all seeds that went beyond 20 meters are treated as censored events
# (distance > 20 meter)
d=c(xtrunc,rep(20,500-length(xtrunc))),
# classify events, found seeds = 1, censored seeds = 0
evnt=c(rep(1,length(xtrunc)),rep(0,500-length(xtrunc)))
# fitsurvival function
CTR_function=survfit(Surv(CTRdata$d, event=CTRdata$evnt) ~ 1)
# return survival probabilities (P) corresponding to distances (D)
P=summary(CTR_function)$surv;D=summary(CTR_function)$time

##### Define dispersal kernels #####
# these kernels are then fit through OLS (Ordinary Least Squares) to objects P &
D

# log normal
SSLN=function(param){
Ex=1-plnorm(D,meanlog=param[1],sdlog=param[2])
sum((Ex-P)^2)
}

# Weibull
SSW=function(param){
Ex=1-pweibull(D,shape=param[1],scale=param[2])
sum((Ex-P)^2)
}

# exponential
SSEX=function(param){
Ex=1-pexp(D,rate=param)
sum((Ex-P)^2)
}

```

```

}

# Normal
SSN=function(param){
Ex=1-phalfnorm(D,theta=param[1])
sum((Ex-P)^2)
}

##### Obtain kernels with reconstructed tails #####

#Fit each model to the censored data and store

fitLN=optim(c(1,1),SSLN)
LNpsave=c(fitLN$par[1],fitLN$par[2])

fitW=optim(c(1.2,55),SSW)
WBpsave=c(fitW$par[1],fitW$par[2])

# Note: above the OLS function was optimized with the Nelder-Mead algorithm
# however this algorithm is optimal for optimization problems of 2 Dimensions
# or greater. The quasi-Newton method 'BFGS' is better suited for 1 D (or 1
# parameter) problems. Alternatively the function 'optimize' can be used
# however result will be the same either way.

fitN=optim(c(0.05),SSN,method="BFGS")
Npsave=c(fitN$par[1])

fitEX=optim(c(0.01),SSEX,method="BFGS")
EXpsave=c(fitEX$par[1])

# choose best model based on AIC score

OLSscores=c(fitLN$value,fitW$value,
fitN$value,fitEX$value)

# vector with number of parameters for each model
pars=c(2,2,1,1)
# calculating AIC from sum of squares
AICscores=(500*log(OLSscores/500) + 2*pars)

##### FINAL #####

#selecting bestfitting model
bestfit=c("LN","WB","T","N","EX")[which(AICscores==min(AICscores))]

#checking difference in between estimated and generating kernel
par(cex.axis=0.9,cex.lab=1.1,las=1,mar=c(4,5,1,1),mfrow=c(2,1))
# density plots
curve(dlnorm(x,log(50),log(3)),0,150,col='grey',xlab="distance",
ylab="probabilty density",lwd=2)
curve(dlnorm(x,fitLN$par[1],fitLN$par[2]),col='black',add=T,lty='dashed',lwd=2)
legend(100,0.010,legend=c('True','CTR derived'),lty=c('solid','dashed'),
col=c('grey','black'),bty='n',lwd=2)

```

```
# probability P of dispersal beyond distance D
curve(1-plnorm(x,log(50),log(3)),0,150,col='grey',xlab="D",
ylab="P",lwd=2)
curve(1-
plnorm(x,fitLN$par[1],fitLN$par[2]),col='black',add=T,lty='dashed',lwd=2)
legend(100,0.90,legend=c('True', 'CTR derived'),lty=c('solid','dashed'),
col=c('grey','black'),bty='n',lwd=2)
```

## Appendix S2: Results from simulated distributions.

We tested how robust the results in the manuscript (and the CTR method in general) are to the specific shape of the dispersal distribution by applying it to simulated data with four different distributions: lognormal, Weibull, exponential, and 1DT. Datasets of dispersal distances were generated from each of the distributions using pseudo-random number generation in R (R development core team 2011). The randomly generated datasets had the same sample size and median dispersal distance as the originally evaluated empirical distribution (thus only differed in the shape of the distribution). We treated the simulated data exactly as the empirical data in the main text. From each, we created multiple truncated datasets (N= 1000), used the CTR method to estimate the dispersal kernel, and evaluated the bias resulting from 1) model choice, 2) search radius, and 3) proportion of overlooked seeds within the search radius.

1) *Model choice*. We tested how often AIC based selection on truncated datasets yielded the 'generating model' (the model from which the simulated dataset was actually created). The results (Table S1) show that the AIC procedure selected the generating model in the majority of cases. Only in cases where the generating model can be approximated equally well by two models (as is the case with Weibull and the exponential) will AIC model selection give some problems. This is non-critical as in these cases selection of a different yet practically equivalent model will not increase bias as the shape is equally well quantified by the wrongly selected but alternative model. Note that the Weibull can approximate the shape of the exponential and Gaussian.

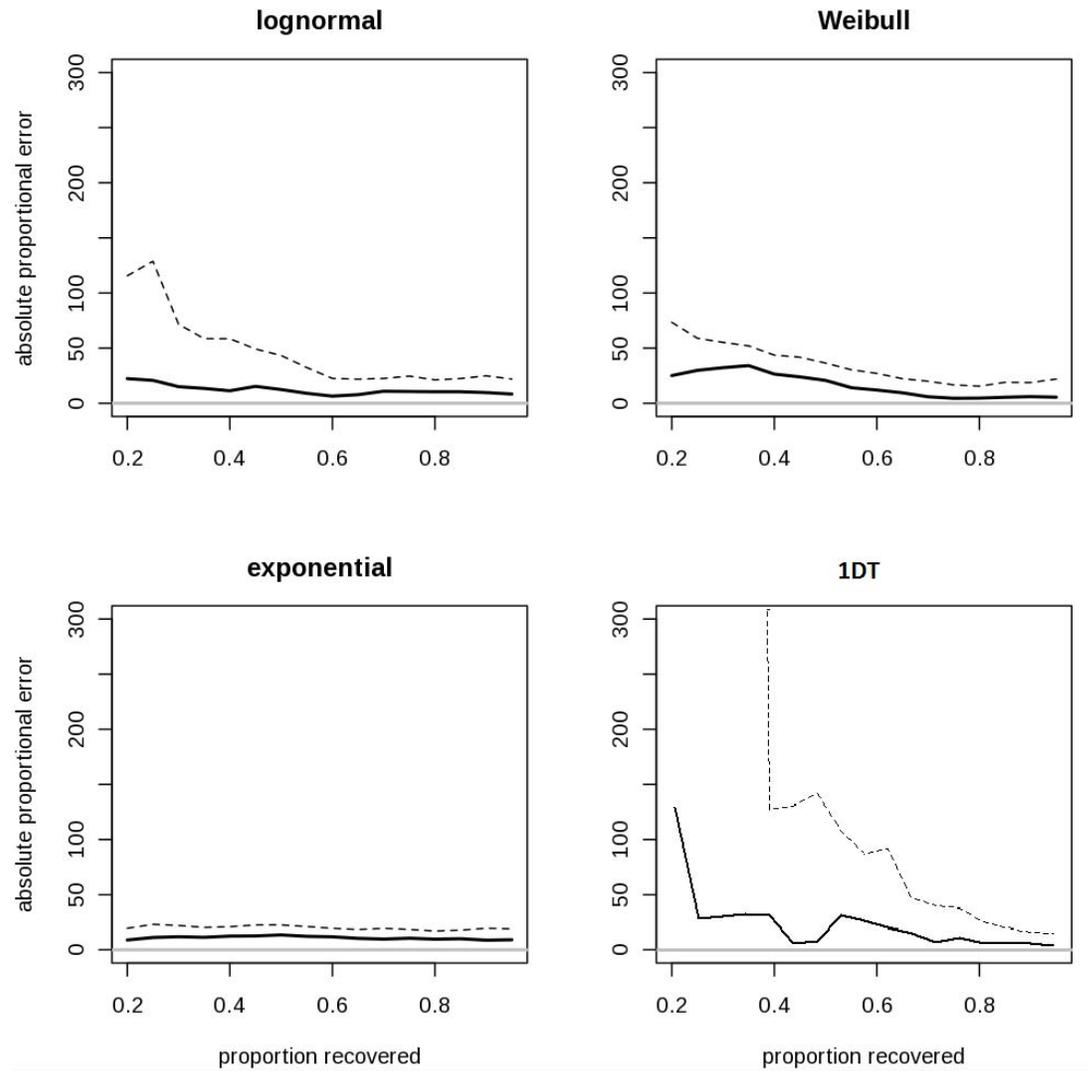
**Table S1.** The use of AIC to identify the distribution of a truncated dataset showed high accuracy, except for equivalent models (Exponential-Weibull). Bold face indicates the proportion of the time when the generating model was selected with the simulations as best fitting model.

Generating model	AIC selected model (after 1000 resamples)			
	Lognormal	Weibull	Exponential	1DT
Lognormal	<b>98%</b>	1%	0%	1%
Weibull	1%	<b>98%</b>	1%	0%
Exponential	0%	93%	<b>7%</b>	0%
1DT	4%	0%	0%	<b>96%</b>

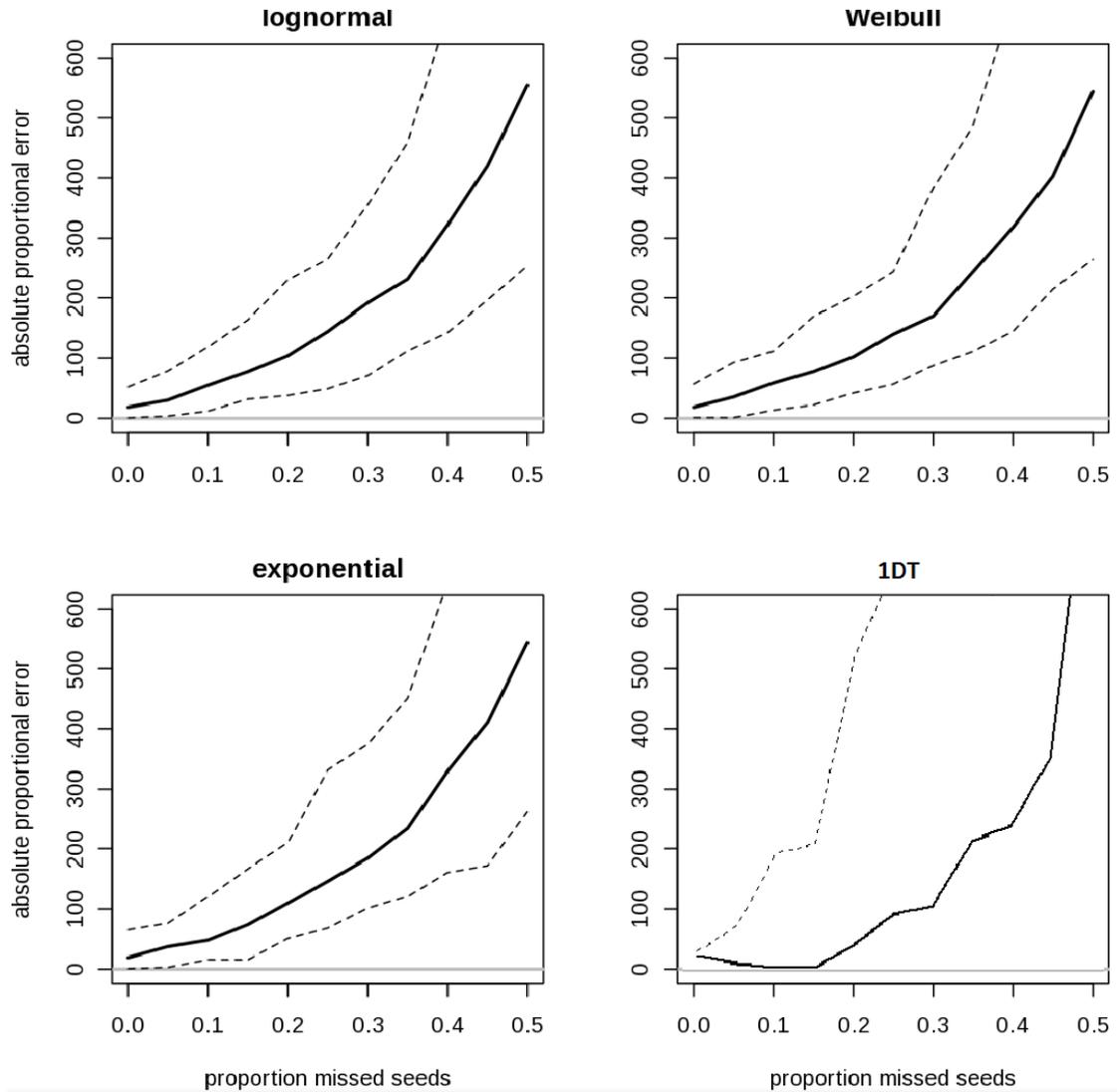
2) *Search radius*. The effects of dispersal kernel shape on bias related to search radius was evaluated by varying the search radii so that 20-80% of seeds fell outside, and then applying the CTR method. In general the results show that bias will increase when the generating distribution has a larger tail (Figure S2).

3) *False-censoring*. The effect of overlooking seeds within the search radius and including those seeds as censored observations was evaluated by varying the proportion of seeds overlooked for each simulated dataset from 0 and 50%. This was done by randomly removing 0 - 50% of the seeds from a truncated dataset and treating them as censored for each of the simulated datasets. In general, the results show a similar pattern as the effect of search radius; bias increases when the generating distribution has a larger tail (Figure S2). This demonstrates the robustness of the CTR method to violations of this assumption for different dispersal kernel shapes.

**Fig S1:** Effect of search radius on the bias of the CTR method applied to four simulated datasets. Bias is plotted against the proportion of tagged seeds recovered. Solid black lines indicate median bias for 1000 simulated datasets; dashed lines indicate 95% CI (calculated as the 2.5 and 97.5 percentiles).



**Fig S2:** Effect of overlooking seeds on the bias of the method applied to four simulated datasets. Plots show an increase in bias with the proportion of overlooked seeds. Solid black lines indicate median bias for 1000 simulated datasets, dashed lines indicate 95% CI (calculated as the 2.5 and 97.5 percentiles).



## References

R Development Core Team (2010). R: A language and environment for statistical computing. R foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.