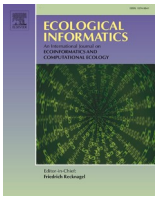
Ecological Informatics 75 (2023) 102018

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf



dispfit: An R package to estimate species dispersal kernels

António Proença-Ferreira [a,b,c,d,*], Luís Borda-de-Água [b,c,d], Miguel Porto [b,c,d], António Mira [a,e], Francisco Moreira [b,c,d,f], Ricardo Pita [a,e]

[a] UBC, Conservation Biology Lab, Department of Biology, School of Sciences and Technology, University of Évora, Mitra, 7002-554 Évora, Portugal
[b] CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Instituto Superior de Agronomia, Universidade de Lisboa, 1349-017 Lisboa, Portugal
[c] CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Campus de Vairão, Universidade do Porto, 4485-661 Vairão, Portugal
[d] BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, 4485-661 Vairão, Portugal
[e] MED – Mediterranean Institute for Agriculture, Environment and Development, CHANGE – Global Change and Sustainability Institute; Instituto de Investigação e Formação Avançada, Universidade de Évora, Pólo da Mitra, Apartado 94, 7006-554 Évora, Portugal
[f] REN Biodiversity Chair, CIBIO/InBIO-UP, Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, Campus Agrário de Vairão, Rua Padre Armando Quintas, 4485-601 Vairão, Portugal

ARTICLE INFO

Keywords:
Dispersal distance
Dispersal kernel
Distribution function
Model selection
Species movement

ABSTRACT

Dispersal of organisms is a ubiquitous aspect of the natural world, with wide implications across scales and organization levels. Interest in dispersal has risen sharply over the past 30 years, mostly due to the multiple and rapid global changes ecosystems face. Among the various aspects that may characterize a dispersion event, dispersal distance is considered a key descriptor in a wide variety of studies across taxonomic groups. Typically, dispersal distances are defined in the form of dispersal kernels describing the dispersal distance distribution according to probability density functions. Although numerous methods providing dispersal data exist, there is still a lack of intuitive and comprehensive approaches and tools to estimate dispersal kernels from such data. Here we present the *dispfit* package, an R software application developed to fill this gap. *dispfit* fits and compares different families of parameterized functions to describe and predict dispersal distances. It includes 9 well-known and commonly used distributions, computes goodness-of-fit and model selection statistics, and estimate each distribution's parameters, along with their first four moments (mean, standard deviation, skewness, and kurtosis). We describe the main functions included in *dispfit* and provide an example to illustrate the workflow of the typical analyses performed within the package. We believe that *dispfit* will critically contribute to improving the modelling of species' dispersal distances, thus enhancing the understanding of the ecological and evolutionary processes involving dispersal movement.

1. Introduction

Dispersal is defined as the movement of organisms from a source location (either after birth or breeding) to a different location where they might establish and potentially reproduce (Clobert et al., 2012; Matthysen, 2012; Nathan et al., 2008b). It has long been recognized as an important life-history trait affecting the dynamics and evolution of populations. Dispersal evolves from a combination of species and individual characteristics, and their responses to changes in biotic and abiotic environments (Benard and McCauley, 2008; Bowler and Benton, 2005). Theoretical and empirical advances on its causes, consequences, mechanisms, and patterns suggest that dispersal is tightly linked to other life-history traits (e.g. survival, niche breadth, trophic level, the timing of reproduction, mating strategies, fecundity), and can therefore be understood as a complex syndrome of traits, consisting of the integrated expression of many morphological, physiological, and behavioural aspects (Ronce and Clobert, 2012). The ecological and evolutionary consequences of dispersal and related life-history traits span over multiple levels (from individuals to whole ecosystems) and have major implications for conservation and environmental management (Driscoll et al.,

Abbreviations: pdf, probability density function.

* Corresponding author at: CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Instituto Superior de Agronomia, Universidade de Lisboa, 1349-017 Lisboa, Portugal.
E-mail address: antoniomiguelpferreira@gmail.com (A. Proença-Ferreira).

https://doi.org/10.1016/j.ecoinf.2023.102018
Received 24 August 2022; Received in revised form 3 February 2023; Accepted 4 February 2023
Available online 8 February 2023

2014). The rapid changes that most ecosystems are currently undergoing have increased the interest in understanding how dispersal shapes species' responses to those changes.

Among the multiple aspects that may characterize a dispersion event (e.g. mechanism, post-dispersal survival, gene flow, triggering time), the distance between the start and end points (i.e. net displacement) remains a key spatial descriptor of dispersal on which many studies rely on (Nathan et al., 2012). In particular, dispersal distance is considered a crucial parameter to understand, for instance, the potential for local adaptation and genetic differentiation, the structuring of metapopulation and metacommunity networks, the dynamics of invasive species, the shifting in species distribution ranges, and the species' ability to track environmental conditions such as landscape or climate change (Nathan et al., 2008a; Ronce, 2007). Including dispersal distance in studies focussing on these processes is thus of critical importance to improve their significance and predictive ability. In this context, the dispersal kernel concept has been central to analysing dispersal processes and spatial patterns. The dispersal kernel describes the distribution of dispersal distances based on probability density functions and is usually characterized by parameters describing the shape and form of those functions (Bullock et al., 2016). Currently, dispersal kernels are commonly incorporated in ecological and evolutionary studies dealing with dispersal across various taxonomic groups (Chapman et al., 2007; Clark et al., 1999; Coombs and Rodríguez, 2007; Krkošek et al., 2007), but are also needed to inform policy to proper habitat management and conservation (Driscoll et al., 2014; Jongejans et al., 2008; Saura et al., 2014). Also, there has been an increasing number of probability distribution functions proposed to describe dispersal kernels with different properties (e.g. Hovestadt et al., 2012). However, the choice of the most appropriate function to fit empirical data is still largely based on personal preference or experience, suggesting a lack of coherence in approaches to model dispersal distances (Bullock et al., 2016).

Despite their great popularity, there is a paradoxical absence of tools to fit and compare different dispersal kernel functions from empirical dispersal data in a straightforward fashion. This is somehow surprising, given the considerable technical developments in movement ecology for improving the accuracy and precision of dispersal distance measures (Driscoll et al., 2014), including capture-recapture (Serrano et al., 2019), isotope analysis (Hall and Beissinger, 2017), GPS and satellite tracking (Cadahía et al., 2010; Krieger et al., 2012) or molecular data (Bode et al., 2017). Here we present the `R` package *dispfit*, which was developed to fill this gap by providing an easy-to-use application to compare the fit of dispersal distance data to different families of parameterized functions commonly used to describe the relations between dispersal and distance. Package *dispfit* uses data obtained from dispersal distances measured directly (e.g. capture-recapture methods or other biologging techniques – such as radio, GPS, or satellite tracking) or indirectly (e.g. parentage analysis from genetic data). While many of the functions considered may be found scattered in multiple statistical software and packages [e.g. `R` packages *fitdistplus* (Delignette-Muller and Dutang, 2015) and *gamlss* (Rigby and Stasinopoulos, 2005), or program *CurveExpert* (Hyams, 2010)], *dispfit* specifically uses functions adapted to species dispersal, providing the necessary measures to describe and compare dispersal kernels, which are not directly obtained from other software sources. We describe the functionalities, usage, and performance of *dispfit*, and recommend its use in ecological and evolutionary studies requiring information on dispersal distance.

2. Package overview

The *dispfit* package is written in the R language and is freely available on GitHub. It can be used under `R` version 3.5.1 or higher (R Core Team, 2017), and may be installed using package *devtools*:

```
> devtools::install_github("https://github.com/apferreira/dispfit")
> library(dispfit)
```

The functions available in *dispfit* depend on packages *ggplot2* (Wickham, 2016), *msm* (Jackson, 2011), *numDeriv* (P. Gilbert and Varadhan, 2019), and *reshape* (Wickham, 2007).

2.1. Dispersal kernel formulation in dispfit

Assuming that a single point is the site of origin of all dispersers from a population, then the dispersal distance of each disperser is the Euclidean distance between the origin and its end point. The dispersal distances of all dispersers thus reflect a continuous parametric distribution, or probability density function (pdf), that characterizes the studied population. We can then define a dispersal kernel as the pdf of the distribution of the values of the Euclidean distances between the source and the final location of a dispersal event. There are several characterizations of a dispersal kernel, for instance Nathan et al. (2012) distinguish between "dispersal distance kernel, K_D", and "dispersal location kernel, K_L". We will follow the latter in this paper and briefly review these concepts in Box 1.

2.2. Distributions included in dispfit

The *dispfit* package fits 9 well-known distributions for estimating dispersal kernels (Clark et al., 1999; Nathan et al., 2012) (Table 1). The simplest functions considered are the single parameter Rayleigh and exponential, which are the most popular in mathematical developments for studying spatial dynamics theory (M. A. Gilbert et al., 2014). The remaining 7 functions are two-parameter distributions, often referred to as better depicting real dispersal kernels than the Rayleigh and exponential functions (Bullock and Clarke, 2000; Clark, 1998; Clark et al., 1999) (Table 1).

2.3. dispfit main functions

dispfit includes one main function, `dispersal.kernel`, which covers most of the functionalities of the package. This function fits one or more dispersal kernels with 1–2 parameters, by estimating the distribution of kernel parameters, a vector θ (with all elements $\theta_j > 0$), maximizing the likelihood function, defined as

$$\mathscr{L}(\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

with x_i being the n observations of variable X and $f(\bullet|\theta)$ the probability density function of the distribution. The function `dispersal.kernel` provides parameter estimates of each fitted distribution, as well as model selection statistics, including the Akaike Information Criteria (AIC), the Akaike Information Criteria for small sample sizes (AICc), the Bayesian Information Criteria (BIC), and the Akaike weights of the relative strength of each model (w_i) (Burnham et al., 2011; Burnham and Anderson, 2002). Finally, it also presents goodness-of-fit (GOF)

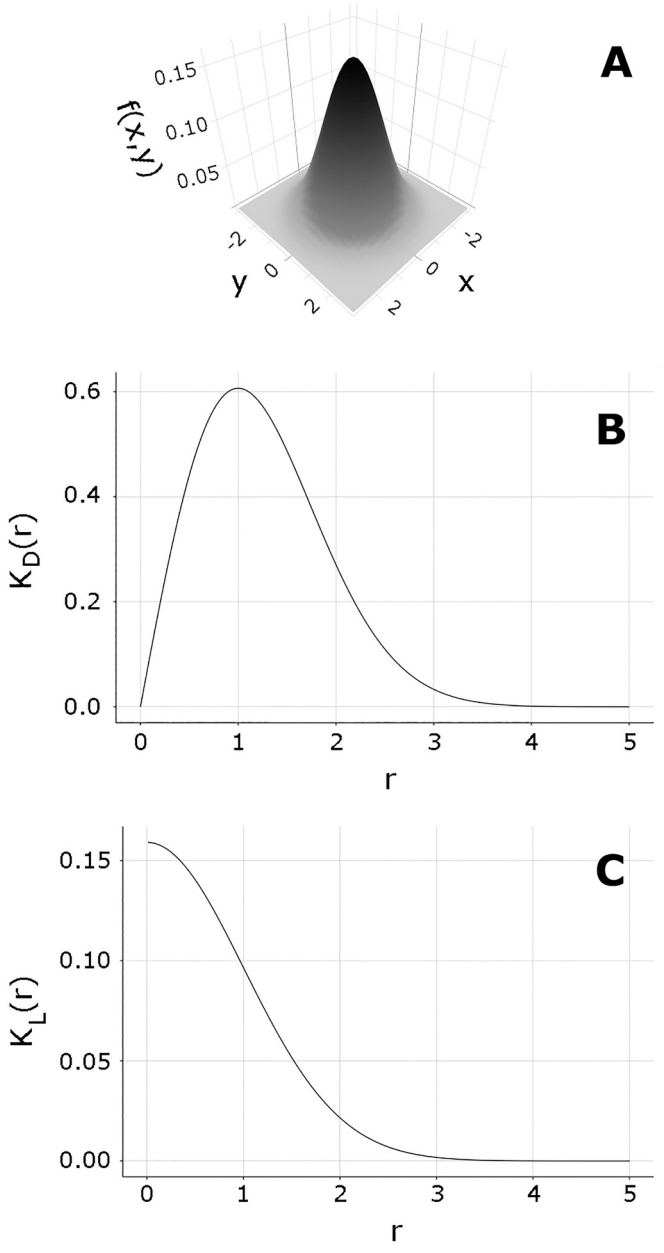


Fig. 1. Example of the distribution of dispersers following an isotropic binormal distribution centred in the origin: $f(x,y) = \frac{1}{2\pi\sigma^2} e^{\left(-\frac{1}{2\sigma^2}(x^2+y^2)\right)}$ (A). Under the assumption of isotropy and using polar coordinates, the above distribution can be written as a function of r as $f(r) = \frac{r}{\sigma^2} e^{\left(-\frac{r^2}{2\sigma^2}\right)}$, which is the Rayleigh distribution using the 'dispersal distance kernel', $k_D(r)$ (B). The 'dispersal distance kernel' and the 'dispersal location kernel' are related such that $k_D(r) = 2\pi r k_L(r)$, so that the example Rayleigh distribution may be converted in such a way that, if $a^2 = 2\sigma^2$, $k_L(r) = \frac{1}{\pi a^2} e^{-\left(\frac{r}{a}\right)^2}$ (C). See Box 1.

measures of the distributions, computing the Chi-squared and Kolmogorov-Smirnov two-sample (K—S) tests, and their associated *p*-values (Press, 1992; Sokal and Rohlf, 1995). If *p*-values are above the threshold defined by the user, then the estimated distribution does not differ significantly from the data and is therefore accepted as a good approximation. The GOF measures may be used either isolated or in tandem, allowing for more robust interpretations. Because kernel parameters are specific for each family of distributions, they cannot be directly compared among different families, even when using the same dataset (Austerlitz et al., 2004). Therefore, to compare quantitatively the shape of different distributions, `dispersal.kernel` also computes the first four moments of each distribution: mean μ, standard deviation σ, skewness α_3, and kurtosis α_4 (Clark et al., 1999) (see Table 2 for their definition and equations).

Pointwise confidence intervals based on profile-likelihood (Venzon and Moolgavkar, 1988) are estimated for each kernel parameter. Confidence envelopes for predicted values are computed by generating a set of distributions based on a cross-sequence value of the sequence of one of the parameter values ranging from the lower to the upper confidence intervals of each parameter (i.e. for each value of the sequence of parameter *a*, distributions are calculated for all values of the sequence of parameter *b*). For each predicted value, that calculation's minimum and maximum outcomes are then defined as the lower and upper confidence envelopes. The `R` function `plot` draws the tested distributions, yielding a visual comparison of their shapes, while the function `predict` will return a table of predicted values according to the best fitting models.

3. Worked example

The following example demonstrates the usefulness and functionality of *dispfit*. We simulated a dataset with a known distribution and then tested the performance of *dispfit* in identifying the original distribution from which the dataset was generated.

Input data should be a vector representing the observed dispersal distances. For this example, we created a random vector of 200 values from a log-normal distribution, with parameters mean $\lambda = 5$ and SD = 1, using the function `rlnorm`, from the *stats* R package (we used the `set.seed` function so that the data is reproducible):

```
> set.seed(1111)
> simulated.data <- rlnorm(200, meanlog = 5, sdlog = 1)
```

Next, the data can be fitted to some or all of the available distributions by using the function `dispersal.kernel`:

```
> kernel.test <- dispersal.kernel(simulated.data, distribution = "all", order.by = "AICc")
```

In this example, all the available distributions were fitted and the results ordered according to their minimum AICc value, but the user may choose the distributions to be fitted as well as the criterion to rank the fitted distributions.

```
> kernel.test
```

Two summary tables are printed by calling the resulting object. Firstly, a table listing the fitted distributions with the corresponding information criteria values and GOF tests is presented (Table 3). Model selection should take into consideration both the model selection criteria and the GOF tests. In this example, four of the distributions (Weibull, gamma, exponential, and Rayleigh) were rejected by the Chi-squared test and two by the KS test (exponential and Rayleigh), when 0.05 is defined as the threshold for a good approximation. From the accepted distributions (log-normal, Wald, geometric, generalized normal, and 2Dt), the log-normal distribution was chosen as providing the best fit describing the data according to all information criteria, with an AICc difference from the second best distribution above 2, indicating

Box 1
Dispersal kernel formulation in *dispfit*

We briefly review the concepts of "dispersal distance kernel, K_D", and "dispersal location kernel, K_L" (Nathan et al., 2012). Although it is very well known, we derive the radial expression of K_D (r) in the special case of an isotropic binormal distribution because it illustrates an important difference between K_D and K_L: when plotted as a function of a radial distance, K_D (r) peaks at $r \neq 0$ and it goes through zero at the origin, a characteristic that is somehow counter-intuitive, and often underappreciated.

Following Nathan et al. (2012), we assume that a single point is the origin site of the dispersers. Therefore, the dispersal distance of each disperser is the Euclidean distance between the origin and its end point. Assuming that the distribution of the distance of the dispersers (K_D) follows an isotropic binormal distribution centred in the origin then, in cartesian coordinates (x, y), it can be expressed as:

$$\mathrm{K_D}(\mathrm{x},\mathrm{y}) = \frac{1}{2\pi\sigma^2}\mathrm{e}^{\left(-\frac{1}{2\sigma^2}\left(\mathrm{x}^2+\mathrm{y}^2\right)\right)} \tag{1}$$

This is shown in Fig. 1-A and could represent, for instance, the idealized distribution of the distances travelled by seeds of a tree located at the origin ($x = 0, y = 0$), where the surface maximum occurs. Often, it is more convenient to use polar coordinates, with distance from the origin $r > 0$ and angle $0 < \theta \leq 2\pi$, such that $x = r\cos\theta$ and $y = r\sin\theta$. In this case, to obtain the corresponding K_D (r, θ) one needs to calculate the Jacobian of the transformation, which in this case is r (e.g., Rice, 2007). Substituting x and y in Eq. (1) by their expressions in terms of r and θ, and multiplying by the Jacobian, the K_D (r, θ) is:

$$\mathrm{K_D}(\mathrm{r},\theta) = \frac{\mathrm{r}}{2\pi\sigma^2}\mathrm{e}^{\left(-\frac{r^2}{2\sigma^2}\right)} \tag{2}$$

Because we assumed isotropy, the angle θ is irrelevant and we can integrate Eq. (2) over θ to obtain $K_D(r)$:

$$\mathrm{K_D}(\mathrm{r}) = \int_0^{2\pi} \frac{\mathrm{r}}{2\pi\sigma^2}\mathrm{e}^{\left(-\frac{r^2}{2\sigma^2}\right)}\mathrm{d}\theta$$

or

$$\mathrm{K_D}(\mathrm{r}) = \frac{\mathrm{r}}{\sigma^2}\mathrm{e}^{\left(-\frac{r^2}{2\sigma^2}\right)} \tag{3}$$

Eq. (3) is the Rayleigh distribution. An important feature of this distribution is that it goes through zero at the origin and, consequently, the maximum occurs for $r \neq 0$ (Fig. 1-B), even when the corresponding two-dimensional distribution has its maximum at the origin (as seen in Fig. 1-A). It may be more intuitive to understand this result if we consider that the data are being collected in rings of equal radius, Δr, as illustrated in Fig. 2. In general, a small number of points are located at rings far from the origin (because few propagules get that far), thus K_D attains small values and for very large distances it is, for practical purposes, equal to zero. When we move towards the origin ($x = 0$, $y = 0$) the number of points in each ring increases, thus contributing to an increase of K_D, but, simultaneously, the area of each ring becomes increasingly smaller, as illustrated in Fig. 2. From these two opposite trends (an increase in the number of points but a decrease in the area of the rings) results a maximum of K_D for an intermediate value of r. The continuous limits assumed in the above equations imply that there is a maximum somewhere when $r \neq 0$, and that K_D ($r = 0$) $= 0$ (see Fig. 2).

Finally, Nathan et al. (2012) defined the dispersal location kernel, K_L (r) in two dimensions and assuming isotropy, as:

$$\mathrm{K_L}(\mathrm{r}) = \frac{1}{2\pi\mathrm{r}}\mathrm{K_D}(\mathrm{r})$$

or, for the used Rayleigh distribution, if $a^2 = 2\sigma^2$,

$$\mathrm{K_L}(\mathrm{r}) = \frac{1}{\pi\,\mathrm{a}^2}\,\mathrm{e}^{-\left(\frac{\mathrm{r}}{\mathrm{a}}\right)^2} \tag{4}$$

shown in Fig. 1-C. By dividing K_D by $2\pi r$, dispersal readily becomes understood in terms of number of points per unit of area, i.e., a density of points. The two kernels are then characterized as the probability that a movement from a source arrives at a distance r, $K_D(r)$, and of the density of movements at distance r, $K_L(r)$ (Peart, 1985).

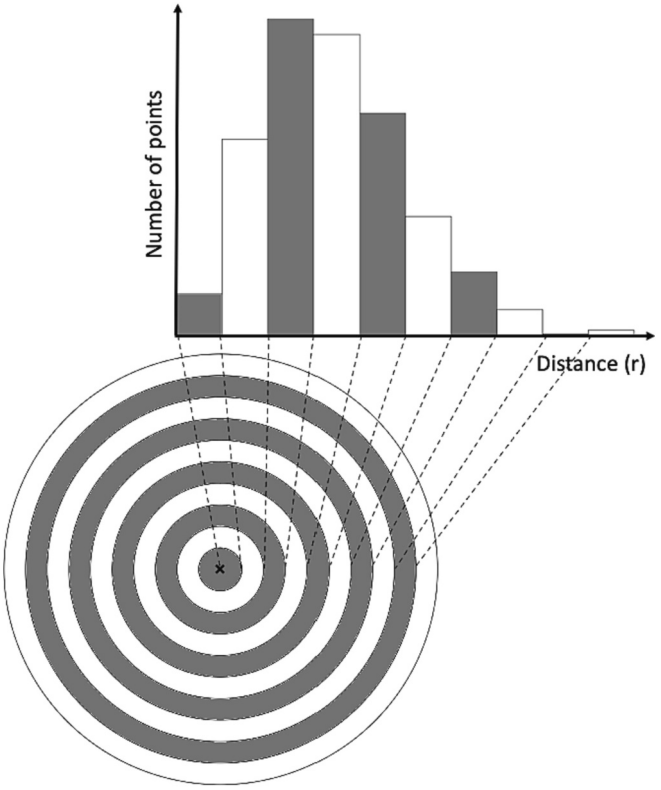


Fig. 2. A histogram showing the distribution of dispersal distance by class. For example, seeds dispersing from a tree (represented by **X**, the origin), will fall within the area surrounding that tree. Data points will then be the seeds collected in each ring of equal Δr, and a histogram of these data will show the number of seeds collected in each ring. When we start collecting far from the origin, we have a small number of points (because few seeds get that far). When moving towards the origin, we will find more points, so the bars representing the number of points also increase. However, because Δr is constant, when moving towards the origin the area of the rings also becomes smaller. Therefore, close to the origin, the area of the rings is so small that the number of points collected within a ring is very low (even if the density of points is high). From the opposite trends in the increase in the number of points towards the origin, but a reduction in the area size of rings, results a maximum for a value of r different from zero.

strong support (see Burnham et al., 2011).

A second table with the estimated parameters and moments for each fitted distribution is also printed by calling the object (e.g. `kernel.test$distribution.parameters`). Note that, in the specific case of function `rlnorm`, the `meanlog` parameter (μ) is equal to the logarithm of Parameter 1 and the `sdlog` parameter (σ) is equal to Parameter 2 of the log-normal distribution from the *dispfit* package. In this example, the estimated parameters for the log-normal distribution were very close to the parameters used to simulate the data, such that if Parameter 1 = 185.94 (95% Confidence Intervals: 159.80–216.40) then $\mu = 5.22$ (CI95: 5.07–5.38), and Parameter 2 = $\sigma = 1.09$ (CI95: 0.99–1.20) (Table 4).

The fitted distributions can be plotted using the `plot function`, to aid in a visual interpretation of the results:

```
> plot(kernel.test, fit.criteria = "AIC", criteria.
dif = 4, envelopes = TRUE, plot.data = TRUE)
```

For this example, we selected the three distributions with an AICc difference from the top model lower than 4 (log-normal, Wald and geometric) and chose to plot the lower and upper 95% confidence *envelopes* for all selected distributions, although other criteria may be chosen by the user (Fig. 3). Together with the fitted distributions, the plot function also displays the kernel density of the distribution of the original dataset (see black line in Fig. 3). Since the plot is generated using the *ggplot2* package, some visual characteristics can be modified using that package's syntax (e.g. changing the colour of the lines, setting another background, or modifying the axis labels and title).

Finally, the `predict` function may be used to estimate values predicted by the selected models. Similarly to the `plot` function, the `predict` function will return a table with the dispersal probability and corresponding upper and lower confidence envelopes of a set of values (from 1 to the maximum value in the original data):

```
> predict(kernel.test, fit.criteria = "AIC",
criteria.dif = 4, envelopes = TRUE)
```

4. Discussion

Accurate estimates of dispersal kernels are needed to understand and model processes such as (meta)population dynamics and persistence in fragmented landscapes (Morales et al., 2010), invasion spread and direction (Wilson et al., 2009), environmental and climate change impacts on species (Travis et al., 2013), community assemblages (Hill et al.,

Table 1
Families of distributions implemented in the *dispfit* R software package. For each kernel, the $K_L(r)$ "dispersal location kernel" and the $K_D(r)$ "dispersal density kernel" used by *dispfit* to fit the supplied data are presented, along with its general properties. Notice that $K_D(r) = 2\pi r K_L(r)$. Notation was adapted from Nathan et al., 2012. In the package outputs, parameters a and b are always referred to as Parameter 1 and Parameter 2, respectively. Depending on the distribution, a or b might refer to location, shape, or scale parameters (see Forbes et al., 2011 for a detailed explanation). Except where noted, a and b are equal to the respective parameters of each distribution defined in Forbes et al., 2011. r is the dispersal distance.

Kernel	$K_L(r)$	$K_D(r)$	Properties and notes
Rayleigh	$\frac{1}{\pi a^2} e^{-\left(\frac{r}{a}\right)^2}$	$\frac{2r}{a^2} e^{-\left(\frac{r}{a}\right)^2}$	Characterized by a single parameter, a, where $a = \sigma\sqrt{2}$, with σ being the standard deviation of the isotropic bivariate normal distribution.
Exponential	$\frac{1}{2\pi a^2} e^{-\frac{r}{a}}$	$\frac{r}{a^2} e^{-\frac{r}{a}}$	Characterized by a single parameter, a, where $a = \gamma$, i.e., the exponential scale parameter.
Generalized Normal	$\frac{b}{2\pi a^2 \Gamma\left(\frac{2}{b}\right)} e^{-\left(\frac{r}{a}\right)^b}$	$\frac{br}{a^2 \Gamma\left(\frac{2}{b}\right)} e^{-\left(\frac{r}{a}\right)^b}$	Also called Exponential Power distribution. Γ is the gamma function (Abramowitz and Stegun, 1965). The parameter a is the scale parameter, while b is the shape parameter, that affects the "fatness" of the tail of the distribution. When $b < 1$ the distribution is fat-tailed, while when $b > 1$ the distribution is thin-tailed, meaning fewer long-distance dispersal events. It's worth noting that when $b = 2$ this pdf is equal to the Rayleigh distribution, and when $b = 1$ it's equal to the Exponential distribution.
Bivariate Student's t (2*Dt*)	$\frac{(b-1)}{\pi a^2}\left[1+\left(\frac{r}{a}\right)^2\right]^{-b}$	$\frac{2r(b-1)}{a^2}\left[1+\left(\frac{r}{a}\right)^2\right]^{-b}$	Defined for values of $a > 0$ and $b > 1$. It has a leptokurtic (i.e. fat-tail) shape and is concave near the origin. It's described to be the best fit for seed dispersal in several tree species (Clark et al., 1999).
Geometric	$\frac{(b-2)(b-1)}{2\pi a^2}\left(1+\frac{r}{a}\right)^{-b}$	$\frac{r(b-2)(b-1)}{a^2}\left(1+\frac{r}{a}\right)^{-b}$	Defined for all values of $a > 0$ and $b > 2$. It shows a fat tail, being less leptokurtic than the Weibull distribution, but more than the 2*Dt*.

(*continued on next page*)

Table 1 (*continued*)

Kernel	$K_L\,(r)$	$K_D\,(r)$	Properties and notes
Lognormal	$\frac{1}{(2\pi)^{3/2} b\, r^2} e^{-\frac{\log\left(\frac{r}{a}\right)^2}{2b^2}}$	$\frac{1}{\sqrt{2\pi}\, b\, r} e^{-\frac{\log\left(\frac{r}{a}\right)^2}{2b^2}}$	Defined for all positive values of a and b.
Wald	$\frac{\sqrt{b}}{\sqrt{8\pi^3 r^5}} e^{-\frac{b(r-a)^2}{2a^2 r}}$	$\frac{\sqrt{b}}{\sqrt{2\pi\, r^3}} e^{-\frac{b(r-a)^2}{2a^2 r}}$	Also termed as Inverse Gaussian distribution, is defined for all positive values of a and b.
Weibull	$\frac{b}{2\pi a^b} r^{b-2} e^{-\left(\frac{r}{a}\right)^b}$	$\frac{b}{a^b} r^{b-1} e^{-\left(\frac{r}{a}\right)^b}$	Defined for all positive real numbers of a and b. The distribution is thin-tailed when $b > 1$ and fat-tailed otherwise. Like the Generalized Normal distribution, when $b = 2$ it degenerates to the Rayleigh distribution. However, when $b = 1$ it is not equal to the Exponential distribution. The Weibull distribution has a highly leptokurtic shape (Austerlitz et al., 2004).
Gamma	$\frac{1}{2\pi a^2\, \Gamma(b)} \left(\frac{r}{a}\right)^{b-2} e^{-\frac{r}{a}}$	$\frac{1}{a^b\, \Gamma(b)} r^{b-1} e^{-\frac{r}{a}}$	Defined for all positive values of a and b. It has an exponential-like tail and equals the Exponential distribution when $b = 2$.

Table 2
Equations used by the *dispfit* R software package to calculate the first four moments (if they are defined) of the implemented $K_L\,(r)$ kernels (Austerlitz et al., 2004; Clark et al., 1999; Press, 1992) (see Table 1 for details).

Name	Equation	Definition
Mean (1st moment)	$\mu = 2\pi \int_0^\infty r^2\, K_L\,(r)\, dr$	Arithmetic mean of a distribution.
Standard Deviation (2nd moment)	$\sigma = \sqrt{2\pi \int_0^\infty r(r-\mu)^2\, K_L\,(r)\, dr}$	The degree of spread about the mean.
Skewness (3rd moment)	$\alpha_3 = \frac{2\pi \int_0^\infty r^4\, K_L\,(r)\, dr}{\sigma^3}$	Degree of asymmetry. The higher the value, the longer the tail is to the right of the central maximum, and so the higher the long-distance dispersal.
Kurtosis (4th moment)	$\alpha_4 = \frac{2\pi \int_0^\infty r^5\, K_L\,(r)\, dr}{\sigma^4}$	Degree of peakedness. A distribution with a relatively high peak is called leptokurtic.

2017), gene flow (Browne et al., 2018), and evolutionary change (Ochocki and Miller, 2017). While a variety of functions have been proposed and implemented to estimate dispersal kernels, researchers still often rely on a single or only a few distributions to fit observed data on dispersal distances, even though dispersal kernels are likely to vary greatly among and within study species and systems (Bullock et al., 2016). This inhibits proper identification of the best dispersal kernel given the data, which in turn may affect the predictive ability of models incorporating dispersal kernels to investigate ecological and evolutionary processes (Bowler and Benton, 2005), impacting the necessary knowledge to address ecological management and conservation (Driscoll et al., 2014).

The R package *dispfit* presented here provides a useful and effective tool to implement and compare the fit of 9 well-known dispersal kernels from observed data, thereby offering a straightforward framework for kernel distribution selection that improves inferences on the fundamental processes affected by species dispersal distance. In addition, if available data exist, *dispfit* may be particularly convenient for assessing possible variations in the distribution of species dispersal, by fitting separate dispersal kernels according to individual covariates (e.g. age class, or sex) and environmental contexts (in space and time), as well as for comparing differences in how multiple species (or traits) may disperse and distribute under environmental change. Also, because the variety of methods to obtain data on dispersal distance (e.g. capture-recapture, habitat occupancy, telemetry, molecular information) may lead to some uncertainty (Driscoll et al., 2014), it may be of interest to assess how dispersal kernels may be affected by different sampling methods. These and many other ecological, evolutionary, and methodological research questions regarding species dispersal distance can be easily explored in *dispfit*, by taking advantage of the standardized outputs produced, allowing direct comparisons among fitted distributions.

Although *dispfit* considers a large set of distributions, we acknowledge that many more dispersal distributions are available and used in the literature (Capdevila et al., 2018; Nathan et al., 2012). While some of these consist of special cases of more general distributions already implemented in *dispfit* (e.g. exponential function), other functions (e.g. general mixture function) were not included in the current version of the package. Although mixture functions have been referred to provide useful descriptions of dispersal distance (e.g. Bullock and Clarke, 2000), they also comprise increased dimensionality, which can lead to overfitting and lack of generality (Bullock et al., 2017). However, future versions of the package may include these and other distributions that can be useful for dispersal kernel estimation. Likewise, future developments of *dispfit* will incorporate methods to deal with missed long-distance dispersal events that may occur to locations outside a predefined study area (e.g. Barrowclough, 1978; Terui, 2020) and that may lead to biased estimates of dispersal, or other vital rates such as survival (Barrowclough, 1978; Zimmerman et al., 2007). Therefore, we advise researchers to acknowledge these assumptions and limitations in their studies, even when not explicitly accounting for them in the models. Also, in cases where the form and extent of the tail of the dispersal kernel are characterized by extreme value distributions (Bullock et al., 2016; García and Borda-de-Água, 2017; Tung et al., 2017), we recommend the use of other existing tools that already allow for the fitting of extreme value distributions [see for e.g. R packages *evd* (Stephenson, 2002) or *extRemes* (Gilleland and Katz, 2016)].

Overall, package *dispfit* offers an intuitive approach based on simple functions for researchers interested in having representative and informative dispersal kernels for a given population. It can be applied in ecological and evolutionary research dealing with species dispersal, but its applications might be further expanded to a wider array of studies and disciplines, including species migration, geomorphology (e.g. transport of sediments), or even epidemiology. We consider that its effectiveness for quantitative and comparative proposes may contribute

Table 3
Comparison of model selection indexes provided by the *dispfit* R software package, calculated for the example random dataset obtained with R function `rlnorm`, as detailed in the text. K-S stands for Kolmogorov-Smirnov two-sample test.

	AIC	Delta AIC	AICc	Delta AICc	BIC	Delta BIC	wi	Chi-squared value	Chi-squared significance	K-S value	K-S significance
Log-Normal	5521.16508	0	5521.22599	0	5527.76171	0	0.73324881	29.406801	0.34145474	0.07901554	0.57188932
Wald	5524.59847	3.43339055	5524.65938	3.43339055	5531.19511	3.43339055	0.13173467	35.9006321	0.11748657	0.08882979	0.42895325
Geometric	5524.97887	3.81378773	5525.03978	3.81378773	5531.5755	3.81378773	0.10891756	29.1231457	0.35495313	0.09126984	0.3931462
Generalized Normal	5528.18504	7.01996139	5528.24596	7.01996139	5534.78168	7.01996139	0.0219223	36.4421684	0.1059597	0.0711658	0.70240902
2Dt	5531.50143	10.3363506	5531.56234	10.3363506	5538.09807	10.3363506	0.00417582	33.8740036	0.16974068	0.12282609	0.11096344
Weibull	5549.34754	28.1824646	5549.40846	28.1824646	5555.94418	28.1824646	5.57E-07	86.8411055	3.33E-08	0.09471503	0.34159592
Gamma	5550.6644	29.499317	5550.72531	29.499317	5557.26103	29.499317	2.88E-07	150.826674	0	0.10512821	0.2252158
Exponential	5625.25037	104.085291	5625.27057	104.044579	5628.54869	100.786974	1.87E-23	25,472.8757	0	0.24180203	1.82E-05
Rayleigh	5875.32328	354.1582	5875.34348	354.117488	5878.6216	350.859882	9.32E-78	8.4184E+10	0	0.42089744	1.33E-15

Table 4
Estimates of parameter and moments for each distribution calculated by the *dispfit* R software package for the example random dataset obtained with R function `rlnorm`, as detailed in the text.

	Parameter 1	Parameter 1 lower CI	Parameter 1 upper CI	Parameter 2	Parameter 2 lower CI	Parameter 2 upper CI	Mean	Standard Deviation	Skewness	Kurtosis
Log-Normal	185.937023	159.798618	216.399506	1.08890836	0.99023214	1.20486987	336.389171	507.161449	7.94997725	211.03066
Wald	331.60259	275.211655	416.388932	154.163652	125.917084	186.418459	331.60259	486.335392	678.299046	32.2646667
Geometric	194.841192	127.06945	308.432555	4.08634381	3.52358111	4.95926705	358.709996	39.4755846	Inf	Inf
Generalized Normal	4.98045587	1.06514581	15.3739986	0.41010331	0.33399696	0.4974134	326.494973	391.228386	6.60270657	42.8791378
2Dt	117.619944	91.2459872	151.256623	1.60885183	1.48222177	1.77351144	617.783877	Inf	Inf	Inf
Weibull	321.560993	274.379398	375.236681	0.94351222	0.84967892	1.04099532	330.180957	350.127053	2.18286586	7.28336875
Gamma	331.6612	268.23901	418.048935	0.99915816	0.83735732	1.18255234	331.381993	331.521567	2.00084238	6.00505531
Exponential	165.690975	150.462823	183.047805	NA	NA	NA	165.690975	165.690975	6	24
Rayleigh	528.101853	493.5199	566.919448	NA	NA	NA	468.018082	244.64391	13.3716823	43.4273298

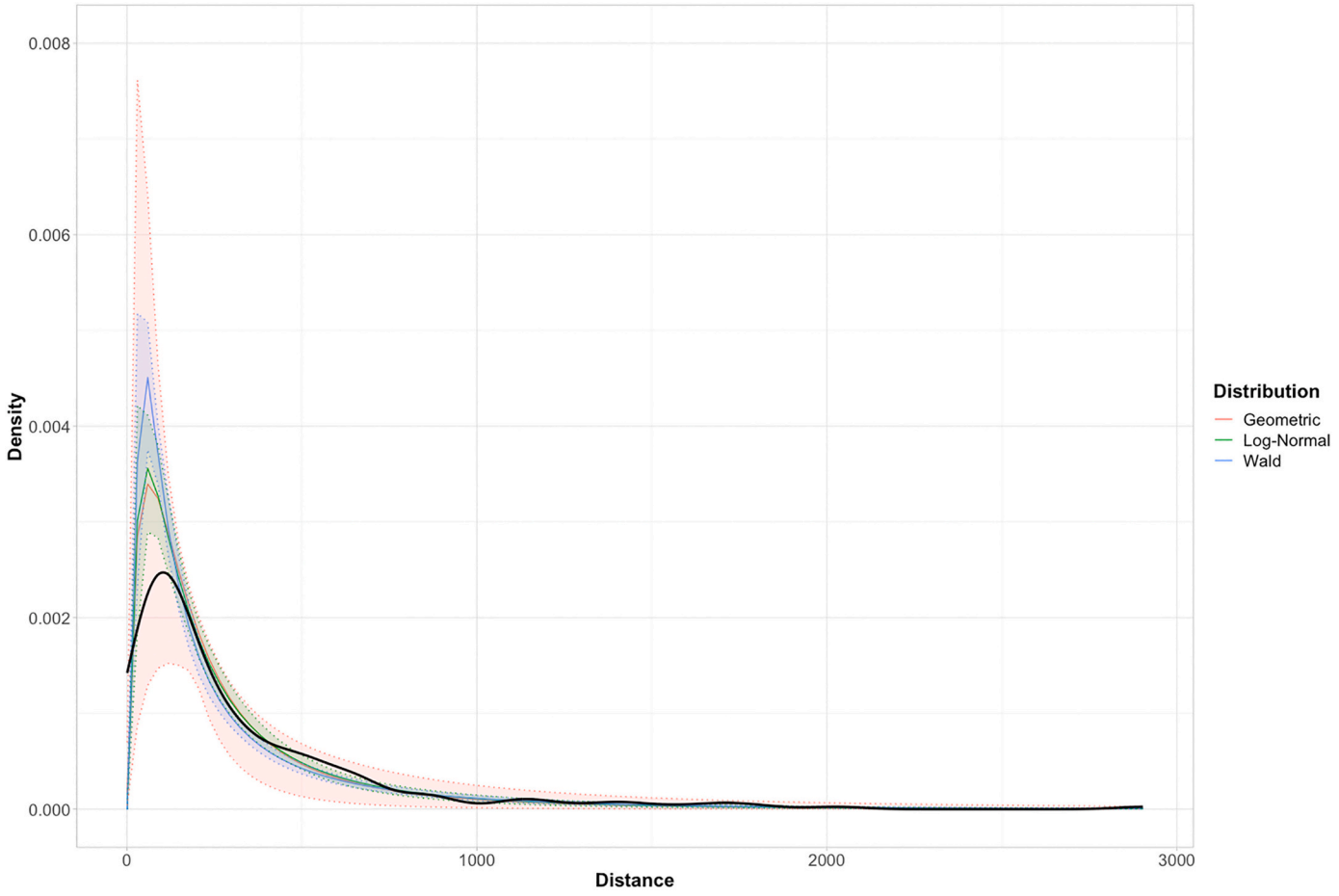


Fig. 3. Distributions with an AICc <4 fitted with *dispfit* R software package for the simulated dataset obtained with R function `rlnorm` (see text for details). Each colour matches a distribution, as detailed in the legend. Since the distributions yielded similar estimates, they appear almost overlaid. Solid lines show the distributions with parameters estimated in Table 4. Dotted lines show the 95% confidence *envelopes* of each distribution, as detailed in the text. The black line shows the real density or distribution of the analysed data.

to better synthesising and set the research agenda on the empirical-based estimation of dispersal kernels.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

This paper and package were supported by the Fundação para a Ciência e a Tecnologia (FCT) under projects NETPERSIST (PTDC/AAG-MAA/ 3227/2012) and MATEFRAG (PTDC/BIA-BIC/6582/2014). APF was supported by FCT grant SFRH/BD/109242/2015. LBA was financed under the Norma Transitória – L57/2016/ CP1440/CT0022. MP was

supported by national funds through FCT in the scope of Norma Transitória – DL57/2016/CP1440/CT0017. RP was supported by FCT through a research contract under the Portuguese Decree-Law nr. 57/2016. We thank one anonymous reviewer for important suggestions which greatly improved the paper.

References

Abramowitz, M., Stegun, I.A., 1965. Handbook of Mathematical Functions. Courier Corporation.

Austerlitz, F., Dick, C.W., Dutech, C., Klein, E.K., Muratorio, S.O., Smouse, P.E., Sork, V. L., 2004. Using genetic markers to estimate the pollen dispersal curve. Mol. Ecol. 13, 937–954. https://doi.org/10.1111/j.1365-294X.2004.02100.x.

Barrowclough, G.F., 1978. Sampling bias in dispersal studies based on finite area. Bird-Banding 49, 333–341. https://doi.org/10.2307/4512392.

Benard, M.F., McCauley, S.J., 2008. Integrating across life-history stages: consequences of Natal habitat effects on dispersal. Am. Nat. 171, 553–567. https://doi.org/10.1086/587072.

Bode, M., Williamson, D.H., Harrison, H.B., Outram, N., Jones, G.P., 2017. Estimating dispersal kernels using genetic parentage data. Methods Ecol. Evol. 9, 490–501. https://doi.org/10.1111/2041-210X.12922.

Bowler, D.E., Benton, T.G., 2005. Causes and consequences of animal dispersal strategies: relating individual behaviour to spatial dynamics. Biol. Rev. 80, 205–225. https://doi.org/10.1017/S1464793104006645.

Browne, L., Ottewell, K., Sork, V.L., Karubian, J., 2018. The relative contributions of seed and pollen dispersal to gene flow and genetic diversity in seedlings of a tropical palm. Mol. Ecol. 27, 3159–3173. https://doi.org/10.1111/mec.14768.

Bullock, J.M., Clarke, R.T., 2000. Long distance seed dispersal by wind: measuring and modelling the tail of the curve. Oecologia 124, 506–521. https://doi.org/10.1007/PL00008876.

Bullock, J.M., Mallada González, L., Tamme, R., Götzenberger, L., White, S.M., Pärtel, M., Hooftman, D.A.P., 2016. A synthesis of empirical plant dispersal kernels. J. Ecol. 105, 6–19. https://doi.org/10.1111/1365-2745.12666.

Bullock, J.M., Hooftman, D.A.P., Tamme, R., Götzenberger, L., Pärtel, M., Mallada González, L., White, S.M., 2017. All dispersal functions are wrong, but many are useful: A response to Cousens et al. J. Ecol. 105, 6. https://doi.org/10.1111/1365-2745.12890.

Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, 2nd ed. Springer-Verlag.

Burnham, K.P., Anderson, D.R., Huyvaert, K.P., 2011. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. Behav. Ecol. Sociobiol. 65, 23–35. https://doi.org/10.1007/s00265-010-1029-6.

Cadahía, L., López-López, P., Urios, V., Negro, J.J., 2010. Satellite telemetry reveals individual variation in juvenile Bonelli's eagle dispersal areas. Eur. J. Wildl. Res. 56, 923–930. https://doi.org/10.1007/s10344-010-0391-z.

Capdevila, P., Linares, C., Aspillaga, E., Riera, J.L., Hereu, B., 2018. Effective dispersal and density-dependence in mesophotic macroalgal forests: insights from the Mediterranean species Cystoseira zosteroides. PLoS One 13, e0191346. https://doi.org/10.1371/journal.pone.0191346.

Chapman, D.S., Dytham, C., Oxford, G.S., 2007. Modelling population redistribution in a leaf beetle: an evaluation of alternative dispersal functions. J. Anim. Ecol. 36–44 https://doi.org/10.1111/j.1365-2656.2006.01172.x.

Clark, J.S., 1998. Why trees migrate so fast: confronting theory with dispersal biology and the Paleorecord. Am. Nat. 152, 204–224. https://doi.org/10.1086/286162.

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. https://doi.org/10.1890/0012-9658(1999)080[1475:SDNAFP]2.0.CO;2.

Clobert, J., Baguette, M., Benton, T.G., Bullock, J.M., 2012. Dispersal Ecology and Evolution, 1st ed. Oxford University Press, Oxford, United Kingdom.

Coombs, M.F., Rodríguez, M.A., 2007. A field test of simple dispersal models as predictors of movement in a cohort of lake-dwelling brook charr. J. Anim. Ecol. 76, 45–57. https://doi.org/10.1111/j.1365-2656.2006.01188.x.

Delignette-Muller, M.L., Dutang, C., 2015. Fitdistrplus: an R package for fitting distributions. J. Stat. Softw. 64, 1–34. https://doi.org/10.18637/jss.v064.i04.

Driscoll, D.A., Banks, S.C., Barton, P.S., Ikin, K., Lentini, P., Lindenmayer, D.B., Smith, A. L., Berry, L.E., Burns, E.L., Edworthy, A., Evans, M.J., Gibson, R., Heinsohn, R., Howland, B., Kay, G., Munro, N., Scheele, Ben C., Stirnemann, I., Stojanovic, D., Sweaney, N., Villaseñor, N.R., Westgate, M.J., 2014. The trajectory of dispersal research in conservation biology. Systematic review. PLoS One 9, e95053. https://doi.org/10.1371/journal.pone.0095053.

Forbes, C., Evans, M., Hastings, N., Peacock, B., 2011. Statistical Distributions. John Wiley & Sons.

García, C., Borda-de-Água, L., 2017. Extended dispersal kernels in a changing world: insights from statistics of extremes. J. Ecol. 105, 63–74. https://doi.org/10.1111/1365-2745.12685.

Gilbert, P., Varadhan, R., 2019. numDeriv: Accurate Numerical Derivatives.

Gilbert, M.A., White, S.M., Bullock, J.M., Gaffney, E.A., 2014. Spreading speeds for stage structured plant populations in fragmented landscapes. J. Theor. Biol. 349, 135–149. https://doi.org/10.1016/j.jtbi.2014.01.024.

Gilleland, E., Katz, R.W., 2016. extRemes 2.0: an extreme value analysis package in R. J. Stat. Softw. 72 https://doi.org/10.18637/jss.v072.i08.

Hall, L.A., Beissinger, S.R., 2017. Inferring the timing of long-distance dispersal between rail metapopulations using genetic and isotopic assignments. Ecol. Appl. 27, 208–218. https://doi.org/10.1002/eap.1432.

Hill, M.J., Heino, J., Thornhill, I., Ryves, D.B., Wood, P.J., 2017. Effects of dispersal mode on the environmental and spatial correlates of nestedness and species turnover in pond communities. Oikos 126, 1575–1585. https://doi.org/10.1111/oik.04266.

Hovestadt, T., Bonte, D., Dytham, C., Poethke, H.J., 2012. Evolution and emergence of dispersal kernels—A brief theoretical evaluation. In: Clobert, J., Baguette, M., Benton, T.G., Bullock, J.M. (Eds.), Dispersal Ecology and Evolution. Oxford University Press, Oxford, United Kingdom.

Hyams, D.G., 2010. CurveExpert Software.

Jackson, C., 2011. Multi-state models for panel data: the msm package for R. J. Stat. Softw. 38, 1–28. https://doi.org/10.18637/jss.v038.i08.

Jongejans, E., Skarpaas, O., Shea, K., 2008. Dispersal, demography and spatial population models for conservation and control management. Perspect. Plant Ecol. Evol. Syst. 9, 153–170. https://doi.org/10.1016/j.ppees.2007.09.005.

Krieger, J., Grandy, R., Drew, M.M., Erland, S., Stensmyr, M.C., Harzsch, S., Hansson, B. S., 2012. Giant robber crabs monitored from space: GPS-based telemetric studies on Christmas Island (Indian Ocean). PLoS One 7, e49809–e49814. https://doi.org/10.1371/journal.pone.0049809.

Krkošek, M., Lauzon-Guay, J.-S., Lewis, M.A., 2007. Relating dispersal and range expansion of California Sea otters. Theor. Popul. Biol. 71, 401–407. https://doi.org/10.1016/j.tpb.2007.01.008.

Matthysen, E., 2012. Multicausality of dispersal: A review. In: Clobert, J., Baguette, M., Benton, T.G., Bullock, J.M. (Eds.), Dispersal Ecology and Evolution. Oxford University Press, Oxford, United Kingdom.

Morales, J.M., Moorcroft, P.R., Matthiopoulos, J., Frair, J.L., Kie, J.G., Powell, R.A., Merrill, E.H., Haydon, D.T., 2010. Building the bridge between animal movement and population dynamics. Philos. Trans. R. Soc. B 365, 2289–2301. https://doi.org/10.1098/rstb.2010.0082.

Nathan, R., Getz, W.M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D., Smouse, P.E., 2008a. A movement ecology paradigm for unifying organismal movement research. Proc. Natl. Acad. Sci. U. S. A. 105, 19052–19059. https://doi.org/10.1073/pnas.0800375105.

Nathan, R., Schurr, F., Spiegel, O., Steinitz, O., Trakhtenbrot, A., Tsoar, A., 2008b. Mechanisms of long-distance seed dispersal. Trends Ecol. Evol. 23, 638–647. https://doi.org/10.1016/j.tree.2008.08.003.

Nathan, R., Klein, E.K., Robledo-Arnuncio, J.J., Revilla, E., 2012. Dispersal kernels: review. In: Clobert, J., Baguette, M., Benton, T.G., Bullock, J.M. (Eds.), Dispersal Ecology and Evolution. Oxford University Press, Oxford, United Kingdom.

Ochocki, B.M., Miller, T.E.X., 2017. Rapid evolution of dispersal ability makes biological invasions faster and more variable. Nat. Commun. 8, 1–8. https://doi.org/10.1038/ncomms14315.

Peart, D.R., 1985. The quantitative representation of seed and pollen dispersal. Ecology 66, 1081–1083.

Press, W.H., 1992. Numerical Recipes in C. Cambridge University Press.

R Core Team, 2017. R: A Language and Environment for Computing.

Rigby, R.A., Stasinopoulos, D.M., 2005. Generalized additive models for location, scale and shape. Appl. Stat. 54, 507–554. https://doi.org/10.1111/j.1467-9876.2005.00510.x.

Ronce, O., 2007. How does it feel to be like a rolling stone? Ten questions about dispersal evolution. Annu. Rev. Ecol. Evol. Syst. 38, 231–253. https://doi.org/10.1146/annurev.ecolsys.38.091206.095611.

Ronce, O., Clobert, J., 2012. Dispersal syndromes. In: Clobert, J., Baguette, M., Benton, T.G., Bullock, J.M. (Eds.), Dispersal Ecology and Evolution. Oxford University Press, Oxford, United Kingdom.

Saura, S., Bodin, Ö., Fortin, M.-J., 2014. Stepping stones are crucial for species' long-distance dispersal and range expansion through habitat networks. J. Appl. Ecol. 51, 171–182. https://doi.org/10.1111/1365-2664.12179.

Serrano, F., Pita, R., Mota Ferreira, M., Beja, P., Segurado, P., 2019. Landscape connectivity affects individual survival in unstable patch networks: the case of a freshwater turtle inhabiting temporary ponds. Freshw. Biol. 65, 540–551. https://doi.org/10.1111/fwb.13449.

Sokal, R.R., Rohlf, F.J., 1995. Biometry: The Principles and Practice of Statistics in Biological Research, 3rd ed. W. H. Freeman and Company, New York.

Stephenson, A.G., 2002. evd: Extreme Value Distributions. R News 2, 0.

Terui, A., 2020. Modeling dispersal using capture–recapture data: A comparison of dispersal models. Ecol. Res. 35, 686–699. https://doi.org/10.1111/1440-1703.12168.

Travis, J.M.J., Delgado, M., Bocedi, G., Baguette, M., Bartoń, K., Bonte, D., Boulangeat, I., Hodgson, J.A., Kubisch, A., Penteriani, V., Saastamoinen, M., Stevens, V.M., Bullock, J.M., 2013. Dispersal and species' responses to climate change. Oikos 122, 1532–1540. https://doi.org/10.1111/j.1600-0706.2013.00399.x.

Tung, S., Mishra, A., Shreenidhi, P.M., Sadiq, M.A., Joshi, S., Sruti, V.R.S., Dey, S., 2017. Simultaneous evolution of multiple dispersal components and kernel. Oikos 127, 34–44. https://doi.org/10.1111/oik.04618.

Venzon, D.J., Moolgavkar, S.H., 1988. A method for computing profile-likelihood-based confidence intervals. J. R. Stat. Soc.: Ser. C: Appl. Stat. 37, 87–94. https://doi.org/10.2307/2347496.

Wickham, H., 2007. Reshaping data with the reshape package. J. Stat. Softw. 21, 1–20. https://doi.org/10.18637/jss.v021.i12.

Wickham, H., 2016. ggplot2: elegant graphics for data analysis. Springer-Verlag, New York. https://doi.org/10.1007/978-3-319-24277-4.
Wilson, J.R.U., Dormontt, E.E., Prentis, P.J., Lowe, A.J., Richardson, D.M., 2009. Something in the way you move: dispersal pathways affect invasion success. Trends Ecol. Evol. 24, 136–144. https://doi.org/10.1016/j.tree.2008.10.007.
Zimmerman, G.S., Gutiérrez, R.J., Lahaye, W.S., 2007. Finite study areas and vital rates: sampling effects on estimates of spotted owl survival and population trends. J. Appl. Ecol. 44, 963–971. https://doi.org/10.1111/j.1365-2664.2007.01343.x.