

DISPERSAL PROCESSES DRIVING PLANT MOVEMENT: RANGE SHIFTS IN A CHANGING WORLD

A synthesis of empirical plant dispersal kernels

James M. Bullock^{1*}, Laura Mallada González¹, Riin Tamme^{2,3}, Lars Götzenberger⁴, Steven M. White^{1,5}, Meelis Pärtel³ and Danny A. P. Hooffman^{1,6}

¹NERC Centre for Ecology and Hydrology, Benson Lane, Wallingford Oxfordshire OX10 8BB, UK; ²Evolution and Ecology Research Centre, School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW 2052, Australia; ³Institute of Ecology and Earth Sciences, University of Tartu, Lai 40, 51005 Tartu, Estonia; ⁴Institute of Botany, Czech Academy of Sciences, Dukelská 135, 37982 Třeboň, Czech Republic; ⁵Wolfson Centre for Mathematical Biology, Mathematical Institute, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, UK; and ⁶Lactuca: Environmental Data Analyses and Modelling, 1112NC Diemen, The Netherlands

Summary

1. Dispersal is fundamental to ecological processes at all scales and levels of organization, but progress is limited by a lack of information about the general shape and form of plant dispersal kernels. We addressed this gap by synthesizing empirical data describing seed dispersal and fitting general dispersal kernels representing major plant types and dispersal modes.

2. A comprehensive literature search resulted in 107 papers describing 168 dispersal kernels for 144 vascular plant species. The data covered 63 families, all the continents except Antarctica, and the broad vegetation types of forest, grassland, shrubland and more open habitats (e.g. deserts). We classified kernels in terms of dispersal mode (ant, ballistic, rodent, vertebrates other than rodents, vehicle or wind), plant growth form (climber, graminoid, herb, shrub or tree), seed mass and plant height.

3. We fitted 11 widely used probability density functions to each of the 168 data sets to provide a statistical description of the dispersal kernel. The exponential power (ExP) and log-sech (LogS) functions performed best. Other 2-parameter functions varied in performance. For example, the log-normal and Weibull performed poorly, while the 2Dt and power law performed moderately well. Of the single-parameter functions, the Gaussian performed very poorly, while the exponential performed better. No function was among the best-fitting for all data sets.

4. For 10 plant growth form/dispersal mode combinations for which we had >3 data sets, we fitted ExP and LogS functions across multiple data sets to provide generalized dispersal kernels. We also fitted these functions to subdivisions of these growth form/dispersal mode combinations in terms of seed mass (for animal-dispersed seeds) or plant height (wind-dispersed) classes. These functions provided generally good fits to the grouped data sets, despite variation in empirical methods, local conditions, vegetation type and the exact dispersal process.

5. Synthesis. We synthesize the rich empirical information on seed dispersal distances to provide standardized dispersal kernels for 168 case studies and generalized kernels for plant growth form/dispersal mode combinations. Potential uses include the following: (i) choosing appropriate dispersal functions in mathematical models; (ii) selecting informative dispersal kernels for one's empirical study system; and (iii) using representative dispersal kernels in cross-taxon comparative studies.

Key-words: dispersal distance, dispersal location kernel, dispersal mode, exponential, exponential power, Gaussian, log-sech, plant height, probability density function, seed mass

Introduction

Good estimates of plant dispersal distances are required by ecologists in their attempts to understand and model processes such as local interactions (Bolker & Pacala 1997), species'

ability to track climate change (Travis *et al.* 2013), population dynamics in fragmented landscapes (Gilbert *et al.* 2014), invasions (Hastings *et al.* 2005), metapopulation persistence (Ovaskainen & Cornell 2006) and evolutionary change (Halatschek & Fisher 2014). Furthermore, a recent systematic review concluded that limited dispersal knowledge often compromises conservation planning (Driscoll *et al.* 2014). The

*Correspondence author. E-mail: jmbul@ceh.ac.uk

holy grail for dispersal ecologists is to develop mechanistic models of dispersal, which represent the physical processes involved in seed movement. These allow predictions of seed dispersal *de novo* and avoid the great effort required to measure dispersal directly (Bullock, Shea & Skarpaas 2006). Mechanistic models have been realized for wind dispersal (Nathan *et al.* 2011), and their predictions have been used in models of spatial population dynamics (Skarpaas & Shea 2007; Soons & Bullock 2008; Bullock *et al.* 2012). While there is some progress towards mechanistic representations of seed dispersal by certain other vectors, such as by vertebrates (Bullock *et al.* 2011; Cortes & Uriarte 2013) or water (Thompson *et al.* 2014), we remain a long way from mechanistic models applicable to every seed dispersal process.

While work towards mechanistic models should continue, remarkably little use has been made of empirical plant dispersal kernels in elucidating general information for use by ecologists. Plant dispersal data, gathered for a wide range of species across a variety of habitats, are a rich source of information on dispersal patterns. Over 20 years ago, Willson (1993) presented an important analysis of seed dispersal kernels, showing how dispersal distances vary according to plant life form (i.e. herb, shrub, tree) and dispersal mode (i.e. ballistic, wind, vertebrate). Many more kernels have been published since 1993 and statistical methods have moved on (Willson analysed kernels using least-squares linear regression). Two recent papers have synthesized empirical plant dispersal data, relating mean or maximum dispersal distance to variables such as dispersal mode, plant growth form, seed release height, seed mass or terminal velocity (Thomson *et al.* 2011; Tamme *et al.* 2014). These analyses are useful; for example, Tamme *et al.* (2014) provided R code to predict the maximum dispersal distance from simple plant traits. However, they offer no information on the complete set of dispersal distances (i.e. the shape of the kernel), which is necessary for many models of spatial dynamics. For example, modelled population spread rates are greatly influenced by the form and extent of the tail of the dispersal kernel (Kot, Lewis & van den Driessche 1996; Caswell, Lensink & Neubert 2003). There is great potential for synthesizing published dispersal data, to provide ecologists with general dispersal kernels for different plant types and dispersal modes.

It is common to fit statistical functions to dispersal data, providing so-called phenomenological dispersal kernels, which are a useful summary of dispersal patterns for activities such as modelling population spread or linking demography to dispersal patterns (Bullock, Shea & Skarpaas 2006). A variety of functions have been implemented; Nathan *et al.* (2012) discuss 13 simple (i.e. one or two parameter) functions that have been used in different studies. These functions have also been used in different ways, with inconsistency in what data are fitted and how the function is interpreted (Cousens, Dytham & Law 2008; Nathan *et al.* 2012). A dispersal kernel is a probability density function of dispersal distances, and the formulation depends on whether it represents the distribution of distances dispersed (the dispersal distance kernel) or the distribution of the final locations of dispersers (the dispersal location kernel) (Nathan *et al.* 2012).

The simplest functions used are the single-parameter Gaussian and exponential. The Gaussian distribution represents seeds moving by Brownian motion for a fixed period of time, while the exponential results from seeds moving randomly and having a certain probability of settling per unit of time. Because of these properties, these two functions are especially popular in mathematical developments of theory concerning spatial dynamics (O'Dwyer & Green 2010; Gilbert *et al.* 2014; Harsch *et al.* 2014). Many individual studies suggest, however, that the Gaussian and exponential do not represent real dispersal kernels well (Clark 1998; Bullock & Clarke 2000), and other functions have been proposed to capture better the high leptokurtosis and long (or even fat, i.e. not exponentially bounded) tails seen in such data. These functions, in general, do not have a mechanistic basis, but they can provide a good description of the dispersal kernel. While some studies have contrasted several functions in fitting to empirical data (Clark *et al.* 2005; Martinez & Gonzalez-Taboada 2009), generally only one or very few functions are fitted, with little justification for the specific function(s) chosen. This suggests a lack of cohesion in approaches to modelling dispersal, which is likely to constrain progress.

The primary aim of this study was to combine and synthesize published empirical seed dispersal data, using literature review and statistical fitting of dispersal functions, to provide general dispersal kernels representing major plant types and dispersal modes. These provide empirically based information about dispersal distances and patterns for plant groups. In doing so, we assess which of the simple functions used in the dispersal literature perform best in representing dispersal kernels across a wide range of species, ecosystems and study types.

Materials and methods

EMPIRICAL DISPERSAL DATA

We examined the literature for data describing seed dispersal kernels. Our initial source was our analysis (Tamme *et al.* 2014) of maximum dispersal distances, which brought together previous studies that summarized published dispersal information, and also did a search on ISI Web of Science (WoS) using the keywords: 'seed', 'dispersal' and 'distance'. We repeated this WoS search between 31 January 2012 (Tamme *et al.*'s end date) and 1 November 2014. All papers were examined for data describing seed distributions with distance from a source. Published data were included in our analyses according to the following criteria: (i) The data described dispersal from a discrete source in two-dimensional space, such as a single plant, a small group of plants or a single depot (e.g. in vertebrate dispersal studies). Linear sources or large patches were excluded as the resulting dispersal kernels could not reasonably be characterized as having a point source (different kernels could be fitted (Shaw *et al.* 2006), but comparison across data sets would be difficult); (ii) The data covered at least four distance locations, to provide a reasonable description of the kernel. In fact, of the data sets included none had only four points, one had five, the rest more; (iii) The data comprised specific distances ('traps') or distance classes ('bins') and at each of these a measure of seed density (absolute or relative), number of seeds, or the proportion of dispersed (i.e. all seeds found) or all (i.e. those counted or placed at the discrete source)

seeds. Lists of individual seed dispersal distances were included; (iv) If the data described seedlings/plants at distances rather than seeds, they were excluded as these are the outcomes of spatial variation in recruitment and survival as well as dispersal; (v) Data generated by fitted kernels (e.g. by inverse modelling) or mechanistic modelling were excluded as these assume an underlying statistical function; and (vi) Data projected by combining empirical information on vector movement and seed retention/deposition (e.g. bird movement and gut retention times) were included. This last decision allowed us to include a large number of data sets for vertebrate dispersal: three for climbers, one for herbs, 11 for shrubs and 17 for trees. However, dispersal estimated by this method was significantly further than when estimated by other methods (e.g. following dispersers or genetic markers). The mean dispersal distances (estimated using the fitted exponential power parameters) of trees were greater for data sets using seed retention methods (median of the mean distances = 3106 m. $n = 17$) than for those using other methods (median = 168 m. $n = 17$) (Kruskal–Wallis $P = 0.02$). There were too few data sets to carry out this test on other growth forms. Nevertheless, we retained these data sets and address this issue in the Discussion.

Data were extracted from tables or, in the majority of cases, from figures, in which case we used the DATATHIEF software (Tummers 2006). As the data were presented in a variety of forms, we needed to represent all dispersal kernels in a common format for analysis and comparison. The most robust approach was to convert all data into paired seed densities (i.e. m^{-2}) and distances, representing the so-called dispersal location kernel (Nathan *et al.* 2012). This accommodated spatial discontinuities (i.e. not all distances had associated data) and the common binning of data into distance classes. This gave us 107 papers presenting 168 dispersal kernels. Some papers reported multiple dispersal data sets, and these were either for different plant species, for different vectors (in terms of the dispersal mode, e.g. wind versus vertebrate; or the vector species) of the same species or for the same species in different environments (e.g. forest vs. clearing). Replicate kernels for a species in the same situation (i.e. vector and environment the same) were analysed together, accounting for replication. Occasionally, different papers contained versions of the same data set; for these, we selected the most comprehensive data set.

Most data sets, 125, were given as ‘bin’ data such that numbers or proportions of seeds were reported for contiguous distance classes. In these cases, the distance was assumed to be at the mid-point of the bin. Forty-one kernels were given as ‘trap’ data, with densities or seed numbers reported for discrete and non-contiguous distances. Only two data sets were given as individual seed distances.

We added supplementary data to each dispersal data set. (i) The dispersal mode: ant, ballistic, rodent, vertebrates other than rodents (separated as rodents were a large class and other vertebrates are more mobile), vehicle (cars, etc.) and wind. In this paper, we use the term dispersal mode (e.g. Willson 1993), which is synonymous with the term dispersal syndrome as used by Tamme *et al.* (2014) and Thomson *et al.* (2011). Wind-dispersed species were separated into those with an appendage that facilitates wind dispersal such as samaras or plumes, versus those with simple seeds/fruits. Modes were taken from the original publications. Each data set had a specific, stated mode, although different data sets for the same species sometimes represented different modes [reflecting that many species are dispersed by multiple vectors (Bullock, Shea & Skarpaas 2006)]. In the case of dispersal by vertebrates, some data sets represented dispersal by a single species (30), while others represented dispersal by multiple species (42); (ii) The growth form: climber, graminoid, herb, shrub or tree. Classifications were taken, and sometimes

simplified (e.g. grasses and sedges into graminoids) to avoid too many classes, from the original publications; (iii) Seed mass (i.e. of the seed, not necessarily the dispersule) from Tamme *et al.* (2014), the original publication or online sources, especially the Seed Information Database (Royal Botanic Gardens Kew 2015); (iv) Plant height from Tamme *et al.* (2014), the original publication or online sources; (v) The plant family, the country and continent, and the vegetation type; taken from the original paper; and (vi) The maximum distance to which dispersal was measured. This represented a wide range; of the maximum measurement distances, the median was 45 m, the maximum 6500 m and the minimum 0.2 m. But all these kernels showed clear distance patterns (i.e. none showed a few similar densities close to the source, representing a very partial description of the kernel).

FITTING DISPERSAL FUNCTIONS

We fitted 11 functions to each of the 168 empirical kernels (Table 1). These are 1- or 2-parameter functions commonly used in analysing dispersal data as summarized by Nathan *et al.* (2012), from whose list of functions we excluded only the general mixture function (describing a mixture of two unspecified functions) and the undefined version of the power law. Mixtures of functions can indeed provide useful descriptions of dispersal (e.g. Bullock & Clarke 2000), but they comprise more than two parameters which can lead to overfitting and a lack of generality. We used the forms of the functions as dispersal location kernels given by Nathan *et al.* (2012) and represent dispersal in two dimensions. As the data were densities, we multiplied these functions by a fitted parameter Q which is the number of seeds dispersed. In some cases, Q was reported in the publication, being the number at the source. But in many cases, Q was not reported or known with any accuracy (the number of seeds found is a poor estimate of Q because it is likely that not all dispersed seeds were detected by the sampling method), so Q was fitted in each case.

These functions were fitted to each data set using SAS Proc NLMIXED, which fits nonlinear mixed models by maximum likelihood. The shape parameter a , the scale parameter b , and Q were fixed effects while replicates, if present, were added as a random effect to the Q parameter (i.e. with common a and b values as we assumed these to be unvarying) – see eqn S1 in Supporting Information. We used Poisson errors following examination of the data and residuals. Zero densities were retained, except that data sets that ended with a series of zero densities (22 of the 168) at the furthest distances were truncated to the first zero density to avoid zero-inflation of the data. We assumed dispersal was isotropic in all cases, and indeed, all studies apart from one did not split data according to direction (Bullock & Clarke 2000; we merged data among the directions in this case). Fits of the functions were compared using AICc, a corrected AIC which relaxes the large sample assumption. As with AIC, this penalizes functions which have more parameters. The best-fitting function for each data set was that with lowest AICc, but following convention, we included in this best-fit group those with an $\Delta AICc$ (i.e. difference from the lowest AICc) ≤ 4 . The AICc only determines the best of the available functions, so goodness-of-fit was assessed by calculating Nakagawa & Schielzeth’s (2013) general r^2 :

$$r^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad \text{eqn 1}$$

where n is the number of observations, y_i is the i th observed value, \hat{y}_i is the i th predicted value, and \bar{y} is the mean value.

Table 1. The probability density functions [dispersal location kernels, taken from Nathan *et al.* (2012)] fitted to the 168 seed dispersal data sets, along with summaries of the goodness-of-fit to these data sets. Distance (in m) is given by d . Fitted parameters are the scale parameter a and the shape parameter b . Densities were seeds m^{-2} . Γ is the gamma function

Name	Probability density function	Number of the 168 data sets that have				
		$\Delta AICc \leq 4$ (best-fit group)	$\Delta AICc > 4$	Not converged	$r^2 \geq 0.7$	Median r^2
Log-sech (log-hyperbolic secant)	$\frac{1/(\pi^2 bd^2)}{(d/a)^{1/b} + (d/a)^{-1/b}}$	119	49	0	142	0.971
Exponential power	$\frac{b}{2\pi a^2 \Gamma(2/b)} \exp\left(-\frac{d^b}{a^b}\right)$	111	57	0	144	0.981
Power law	$\frac{(b-2)(b-1)}{2\pi a^2} \left(1 + \frac{d}{a}\right)^{-b}$	101	65	2	135	0.973
Logistic	$\frac{b}{2\pi a^2 \Gamma(2/b) \Gamma(1-(2/b))} \left(1 + \frac{d^b}{a^b}\right)^{-1}$	100	68	0	133	0.951
2Dt	$\frac{b-1}{\pi a^2} \left(1 + \frac{d^2}{a^2}\right)$	98	70	0	136	0.972
Gamma	$\frac{b}{2\pi a^2 \Gamma(b)} \left(\frac{d}{a}\right)^{b-2} \exp\left(-\frac{d}{a}\right)$	98	70	0	135	0.974
Inverse Gaussian (WALD)	$\frac{\sqrt{b}}{\sqrt{8\pi^3 d^5}} \exp\left(-\frac{b(d-a)^2}{2a^2 d}\right)$	88	77	0	123	0.953
Weibull	$\frac{1}{2\pi a^2} d^{b-2} \exp\left(-\frac{d^b}{a^b}\right)$	77	80	11	101	0.829
Exponential	$\frac{1}{2\pi a^2} \exp\left(-\frac{d}{a}\right)$	68	89	11	120	0.876
Log-normal	$\frac{1}{(2\pi)^{3/2} bd^2} \exp\left(-\frac{\ln(d/a)^2}{2b^2}\right)$	55	109	4	69	0.489
Gaussian	$\frac{1}{\pi a^2} \exp\left(-\frac{d^2}{a^2}\right)$	30	101	37	63	0.509

GENERALIZING DISPERSAL KERNELS

AICc values showed that the exponential power (ExP) and log-sech (LogS) were by far the most frequent functions in the best-fit group across the 168 data sets. We concluded these would be the two most appropriate functions for a general description of dispersal kernels according to plant characteristics. Using two functions allowed us check whether findings were independent of the exact function used. The ExP and LogS functions did not describe all data sets well, so we applied an arbitrary threshold of $r^2 \geq 0.7$ for inclusion of a case study in the further analyses using the ExP (144 of the 168 data sets) or LogS (142: this is not exactly the same group as that meeting this criterion for the ExP; 16 data sets were assigned uniquely to ExP or LogS).

We analysed whether the included studies for either function were a biased set in terms of plant characteristics of the full set of studies, and so whether our analyses using this set would give a biased understanding of dispersal kernels (Table S1). This was not the case for growth form, dispersal mode, plant height or seed mass, with the exception of a slight difference in plant heights for LogS, with excluded plants being taller (Kruskal–Wallis; $P = 0.043$; means 13.04 m vs. 8.95 m). We also used Kruskal–Wallis to assess whether the included studies represented a biased set of case studies in terms of the methodology – that is the maximum distance over which dispersal was measured. This was not the case for ExP ($P = 0.283$) or LogS ($P = 0.515$).

The mean distance of the ExP is a function of both parameters:

$$\text{mean(ExP)} = a \frac{\Gamma(3/b)}{\Gamma(2/b)}, \tag{eqn 2}$$

(Γ is the gamma function). The LogS has an unspecified mean for $b > 1$ (which was found for 60 data sets). We used the ExP mean to compare the data sets in terms of how plant characteristics affected the ExP kernel. We calculated the ExP mean dispersal distance for each of the 144 case studies with $r^2 \geq 0.7$ and tested for differences in this mean according to four principal plant characteristics: growth form, dispersal mode, plant height and seed mass. These were shown

by Tamme *et al.* (2014) to be good predictors of measured maximum dispersal distance, and they represent straightforward classifications of plants into types that might be used in assigning dispersal kernels for modelling or other studies. We fitted linear models using SAS Proc MIXED to the ExP means using all combinations of these four factors (seed mass and plant height were \log_{10} transformed). Interaction terms were not included as the data were unbalanced and collinear. The model with lowest AIC comprised Growth form + Dispersal Mode + Plant height (AIC = 405.8, $r^2 = 0.56$), with one other within 4 AIC, that is Growth form + Dispersal Mode + Plant height + Seed mass (AIC = 405.9) (see Table S2 for the full analysis). We also ran the best-performing linear model combining classes which might be expected to have similar dispersal kernels: graminoids and herbs (AIC = 410.1), climbers and trees (405.8), rodents and other vertebrates (426.4), and vehicles and vertebrates (406.7). None of these reduced the AIC so we kept the full set of growth forms and dispersal modes in further analyses.

We therefore fitted general ExP and LogS dispersal functions to the $r^2 \geq 0.7$ case studies grouped according to growth form and dispersal mode. For both ExP and LogS 10 form/mode, primary combinations had a reasonable number (>3) of case studies to fit general functions (Table S3); we considered ≤ 3 cases to be too affected by the particularities of the individual case studies. Preliminary analyses showed an alternative approach using plant height or seed mass as modifiers of the a and/or b parameters was not effective and led to poor model convergence. To use the information provided by these extra variables, where there were sufficient data sets we also fitted functions to subdivisions of the primary form/mode groups, using the variable most likely to be important for that group: seed mass for animal-dispersed groups and plant height for wind-dispersed groups (Table 3). While seed mass rather than plant height would be expected to affect animal dispersal, both might be important for wind dispersal. Thomson *et al.* (2011) found that plant height is a much more important predictor of dispersal distance than seed mass, so we chose this as the grouping criterion for wind-dispersed seeds. The number and bounds of the subdivisions were arbitrary, but based on

the number of data sets and the distribution of values of these variables, using similar subdivisions across the groups for comparability (Table 3). For both ExP and LogS, we fitted Q , a and b to the density data, with each case and each replicate within a case was included as random factor that affected Q . Thus, we searched for common a and b values across all cases, in accordance with the hypothesis that all studies within a specific growth form/dispersal mode (+ seed mass or plant height class) combination followed the same underlying probability density function.

We calculated percentiles of the dispersal distances for each fit of the ExP and LogS using the integrals of these functions formulated in terms of dispersal distances (i.e. the dispersal distance kernel rather than the location kernel given in Table 1 – see Nathan *et al.* (2012), whereby the distance kernel = the location kernel/ $2\pi d$). For long-tailed kernels such as these, the median is a good measure of the average dispersal distance and the 95th percentile summarizes the tail (Tamme *et al.* 2014).

Results

FITTING DISPERSAL FUNCTIONS TO 168 DATA SETS

The 168 data sets covered 63 families, all the continents except Antarctica (30 countries), and the broad vegetation types of forest (100 data sets), grassland (46), shrubland (6) and more open habitats (16; e.g. desert, clearfell, urban). The classification of data sets among dispersal modes and growth forms was uneven (Table 2). While the distribution of dispersal modes in nature is not known in detail, Jordano (2000) estimated *ca.* 40–90% of species in forests and 20–50% in shrublands are vertebrate-dispersed, while frugivory is uncommon in many other vegetation types. This suggests a bias towards measurements of vertebrate dispersal (43% of data sets), as well as towards forest ecosystems [which account for only 31% of land cover world-wide (FAO 2010)].

Of the 11 functions, the exponential power, 2Dt, logistic, gamma and log-sech converged for all data sets, but the other functions did not always converge (Table 1). This poor convergence was only substantial for the Gaussian (37 data sets), which also showed poorest fit with $\Delta\text{AICc} \leq 4$ for only 30 data sets. The other one parameter function, the exponential, had a much better performance, being in the best-fit group in 68 cases. The log-normal and the Weibull also performed rather poorly. The other functions – exponential power, 2Dt, power law, WALD, logistic, log-sech and gamma – all performed better, but all fell outside the best-fit group in many cases. Interestingly, the WALD, although based on a

mechanistic representation of dispersal by wind (Katul *et al.* 2005), was in the best-fit group of only 15 of the 55 wind dispersal data sets and by this metric performed better than only the exponential, Gaussian and log-normal. The r^2 values showed generally very good fit with high values (in terms of the median value and number of data sets for which $r^2 \geq 0.7$; Table 1) for most functions except the Gaussian and log-normal, and the patterns of r^2 values followed those of the AICc values. Figure 1 shows some example plots of data with the fitted functions. Tables S4 and S5 present the fitted parameter values for all well-fitting functions for each of the 168 data sets and the supplementary data describing the species and study system.

The exponential power (ExP; $\Delta\text{AICc} \leq 4$ for 111 cases) and log-sech (LogS; 119) clearly outperformed the other functions (but note these numbers are $\ll 168$). The ExP can reduce to an exponential ($b = 1$) or Gaussian ($b = 2$), and can fit a fat tail ($b < 1$) or a thin tail ($b > 1$). Of the 144 data sets selected as showing good general fit to the ExP (i.e. $r^2 \geq 0.7$), 117 had a $b < 1$, and 27 had $b > 1$, suggesting that a majority of kernels are fat-tailed. The LogS is always fat-tailed (power-law tail) and flattens (for $b < 1$) or decreases ($b > 1$) close to 0 distance. Of the 142 data sets with good general fit to the LogS, 90 had $b < 1$ and 52 $b > 1$.

Although the case study composition of the best-fit group differed among the functions, there was no bias in comparison with the full set of case studies in terms of dispersal mode ($\chi^2 = 45.9$, d.f. = 40, $P = 0.24$), growth form ($\chi^2 = 21.7$, d.f. = 20, $P = 0.36$), plant height (Kruskal–Wallis $H = 9.6$, d.f. = 10, $P = 0.48$) or seed mass ($H = 12.5$, d.f. = 10, $P = 0.25$). We also asked whether the study design affected the set of best-fit functions (e.g. was a function allowing a fatter tail less likely to fit kernels measured over shorter distances?), but the maximum distance over which dispersal was measured did not differ among the best-fit groups ($H = 15.4$, d.f. = 10, $P = 0.12$).

GENERALIZED DISPERSAL KERNELS

The fitting of the ExP and LogS functions across grouped data sets provided generalized dispersal kernels for a number of well-studied plant growth form/dispersal mode combinations (Table 3, Fig. 2). Across the 10 combinations, both the ExP and LogS gave similar and generally good fits, with r^2 ranging from 0.39 to 0.97 (median 0.73, $r^2 < 0.7$ for three combinations) for the former and 0.32–0.97 (median 0.78, $r^2 < 0.7$ for

Table 2. The distribution of the 168 dispersal data sets among growth forms and dispersal modes

Dispersal mode	Ant	Ballistic	Rodent	Vehicle	Vertebrate (excl. rodent)	Water	Wind (appendage)	Wind (no appendage)	Total
Growth form									
Climber	0	1	0	0	3	0	2	0	6
Graminoid	1	0	0	1	0	1	0	9	12
Herb	11	13	1	2	3	0	10	13	53
Shrub	6	2	0	0	15	0	1	4	28
Tree	0	3	12	0	38	0	16	0	69
Total	18	19	13	3	59	1	29	26	168

Table 3. The a and b values for the exponential power and log-sech functions fitted to case studies grouped by growth form and dispersal mode, and subdivided where possible by seed mass (animal-dispersed) or plant height (wind-dispersed). These subdivisions show the actual data ranges (see text) and so are discontinuous. The goodness-of-fit (r^2) and number of studies (n) are given with median and 95th dispersal distances of the fitted kernel. See Figs 2 and S1 for plots of the fits

Growth form	Dispersal mode	Exponential Power (Exp)						Log-sech (LogS)					
		a	b	r^2	n	Percentile distances (m)		a	b	r^2	n	Percentile distances (m)	
						50th	95th					50th	95th
Herb	Ant	0.5281	1.2762	0.743	9	0.629	1.572	0.4580	0.3859	0.442	12	0.458	1.222
	0.7–8 mg	2.5×10^{-6}	0.1888	0.929	3	0.561	6.838	0.2305	0.3528	0.871	5	0.231	0.565
	10–36 mg	0.3726	1.1615	0.939	6	0.499	1.305	0.4667	0.4726	0.620	7	0.467	1.552
Herb	Ballistic (all <10 mg & <1 m)	0.0917	0.6349	0.390	10	0.470	1.757	1.0634	0.8319	0.455	12	1.063	4.926
	Wind + appendage	4.7×10^{-5}	0.2336	0.879	10	0.388	3.623	0.1253	0.6893	0.752	10	0.125	0.723
Herb	1–3.5 m	1×10^{-8}	0.1423	0.645	5	0.985	17.91	0.0197	1.4989	0.996	5	0.020	0.868
	0.1–0.8 m	0.0030	0.3454	0.686	5	0.406	2.499	0.1286	0.6547	0.804	5	0.129	0.679
	Wind no appendage (all <1 m tall)	4.2×10^{-6}	0.2069	0.881	12	0.205	2.223	0.1297	0.9075	0.901	10	0.130	1.303
	Wind no appendage	3.0×10^{-8}	0.1597	0.430	8	0.190	2.908	0.0571	0.8171	0.749	8	0.057	0.258
Graminoid	1.75–3 m	31.985	1.0141	0.235	3	45.93	123.3	31.771	0.5193	0.839	3	31.77	119.0
	0.06–0.65 m	1×10^{-8}	0.1549	0.974	5	0.126	2.015	0.1420	1.5777	0.819	5	0.142	7.836
Shrub	Ant	0.1716	0.5940	0.963	7	1.116	4.368	0.8923	0.6126	0.966	6	0.892	4.235
	7–9 mg	0.1915	0.6272	0.974	4	1.024	3.858	0.7483	0.5883	0.982	4	0.748	3.339
Shrub	29–40 mg	0.2680	0.6568	0.861	3	1.229	4.485	1.5088	0.8458	0.868	2	1.509	12.95
	Vertebrate	1×10^{-8}	0.1339	0.996	18	4.974	99.39	28.659	0.8208	0.831	18	28.66	130.1
	0.2–3 mg	1×10^{-8}	0.1161	0.879	11	375.7	9471	29.216	0.8369	0.870	11	29.22	245.2
Tree	31–69 mg	1×10^{-8}	0.1264	0.973	2	26.01	569.7	25.220	0.8415	0.980	2	25.22	214.2
	5000–10 500 mg	0.0008	0.3122	0.997	5	0.249	1.693	0.0584	0.8088	0.861	5	0.058	0.456
Tree	Rodent	0.1507	0.4171	0.703	11	5.449	20.29	8.4496	0.7566	0.873	11	8.450	57.83
	195–950 mg	2.0615	0.6538	0.809	5	9.594	35.12	7.9526	0.7733	0.744	5	7.953	56.79
Tree	2420–18 800 mg	0.3212	0.5035	0.811	6	4.190	18.58	8.9256	0.7269	0.803	6	8.926	56.64
	Vertebrate	1×10^{-8}	0.1246	0.690	39	40.01	897.1	49.997	1.3989	0.807	39	50.00	1751
Tree	0.3–15.1 mg	1×10^{-8}	0.1288	0.781	10	14.96	317.9	15.511	1.2950	0.740	10	15.51	417.2
	31–180 mg	1×10^{-8}	0.1237	0.867	11	49.89	1132	10.014	1.7790	0.264	9	10.01	921.8
Tree	200–800 mg	1×10^{-8}	0.1238	0.804	9	48.68	1103	97.668	0.7489	0.624	13	97.67	655.7
	1000–113 700 mg	1.8875	0.3410	0.198	9	285.2	1778	22.922	1.0219	0.439	7	22.92	307.9
Tree	Wind + appendage	0.5602	0.4289	0.720	21	17.11	86.57	19.709	0.5853	0.317	20	19.71	87.27
	30–46 m	1.3437	0.4654	0.364	7	25.98	122.7	29.601	0.9953	0.7031	6	29.60	371.6
Tree	5–15 m	2.7825	0.8346	0.580	14	6.663	20.87	3.4581	0.3988	0.4835	14	3.458	9.531

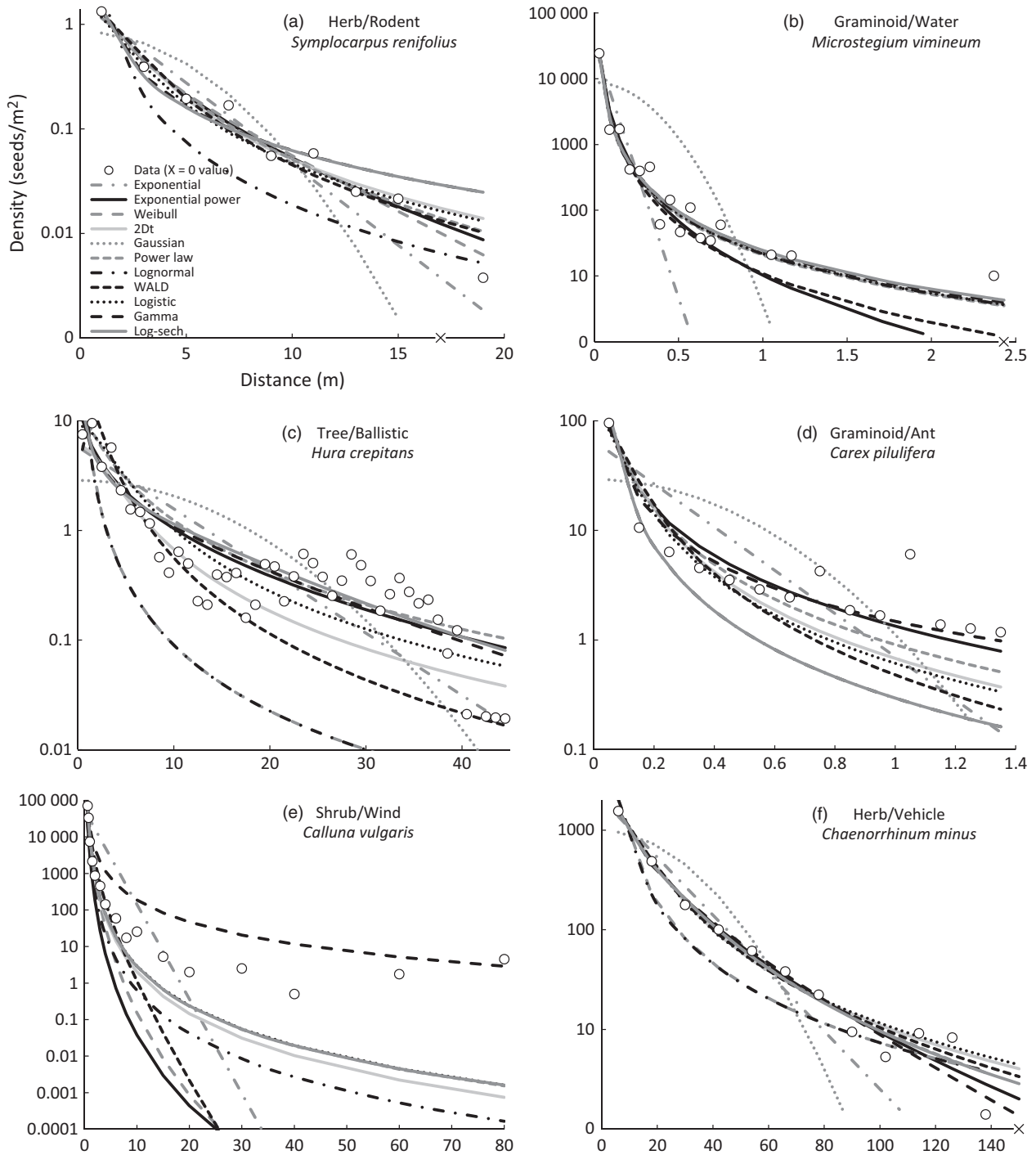


Fig. 1. Six examples of the 168 data sets with the 11 fitted dispersal functions (see Table 1), showing \log_{10} seed density against distance. The examples are chosen to reflect the less common growth form/dispersal mode groupings, which therefore were not among the ten groups (Table 3, Fig. 2) for which general functions were fitted. Data sources: (a) Wada & Uemura (1994), (b) Tekiel & Barney (2013), (c) Swaine & Beer (1977), (d) Kjellsson (1985), (e) Bullock & Clarke (2000), (f) Arnold (1981). Where plotted, the y-axis is not continuous between 0 and the next tick.

3) for the latter; although the patterns of r^2 values across the form/mode combinations differed between ExP and LogS (note that it is not appropriate to compare the fits of the ExP and LogS using AIC as they were fitted to different sets of case studies). Similar results were seen for the subdivisions according to seed mass or plant height classes, with $r^2 < 0.7$ for 12 of 38 groups across the ExP and LogS functions.

Median dispersal distances of the fitted functions were in the order trees (ExP mean of the medians = 20.9 m) > shrubs (3.05 m) > herbs and graminoids (0.38 m). Vertebrates (excluding rodents; ExP mean of the medians = 22.5 m) dispersed seeds further than rodents (5.45 m), which dispersed seeds similar distances to wind for dispersules with an appendage (8.75 m), while dispersal by ants (0.87 m), ballistic (0.45 m)

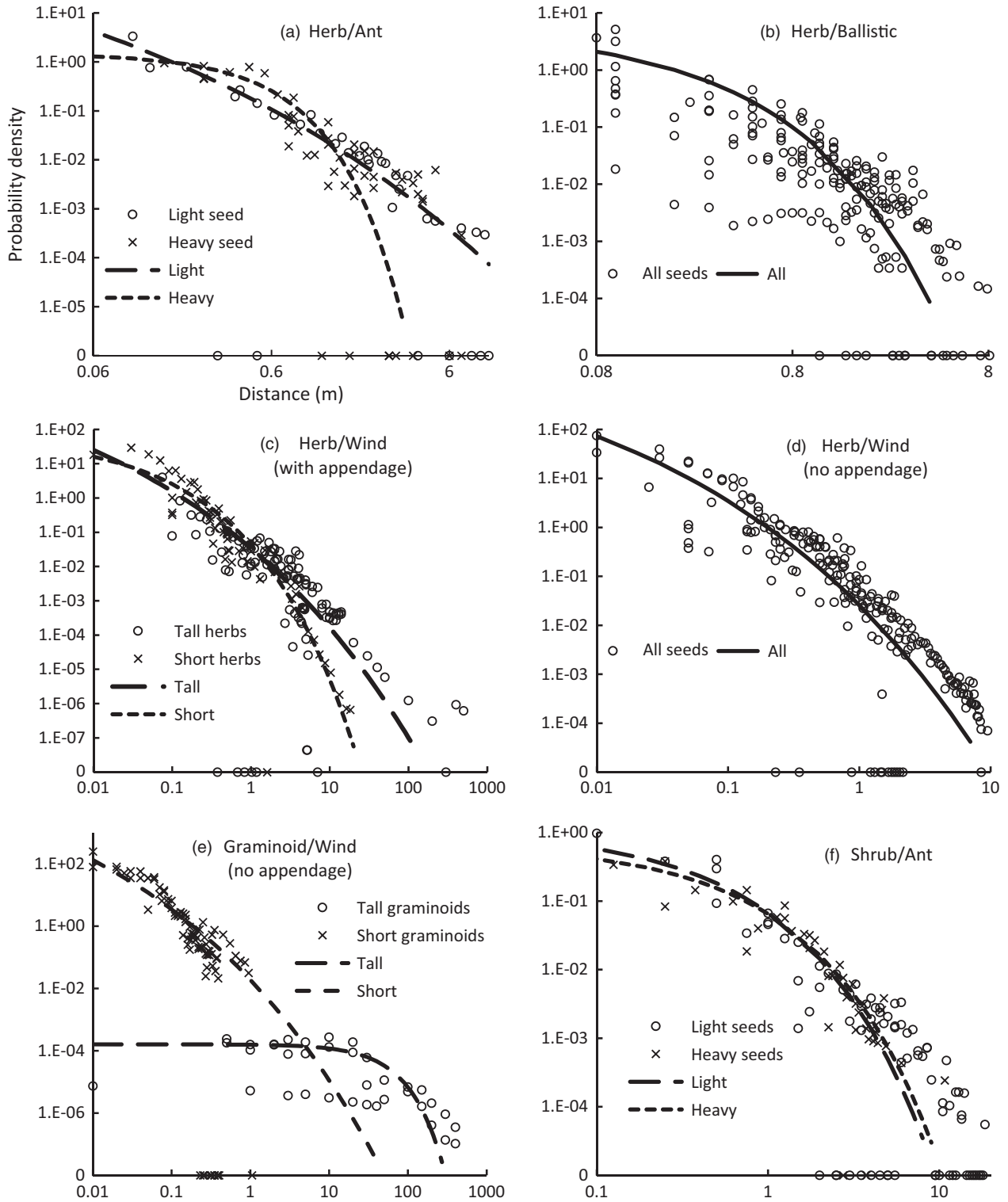


Fig. 2. Generalized dispersal kernels for all 10 growth form/dispersal mode combinations for which there were sufficient data sets. Where possible, the data sets were also split into tall versus short plants (wind-dispersed) or light versus heavy seeds (animal-dispersed). The plots show the data and the fitted exponential power functions (plots for the log-sech function are shown in the SI). In contrast to Fig. 1, the data are the probability densities, calculated by dividing the measured seed density by the individual Q (total seed number) value estimated for each data set while fitting the function. Both axes are logged to gain best visibility of the data, which cover a large range in both dimensions. The y-axis is not continuous between 0 and the next tickmark. Further information is in Table 3.

and wind for dispersules with no appendage (0.20 m) resulted in the shortest median distances. The same patterns were found for the 95th percentile and for the LogS function (Table 3).

While these means are somewhat confounded as growth form and dispersal mode were not found in all possible combinations, more specific comparisons show the same patterns.

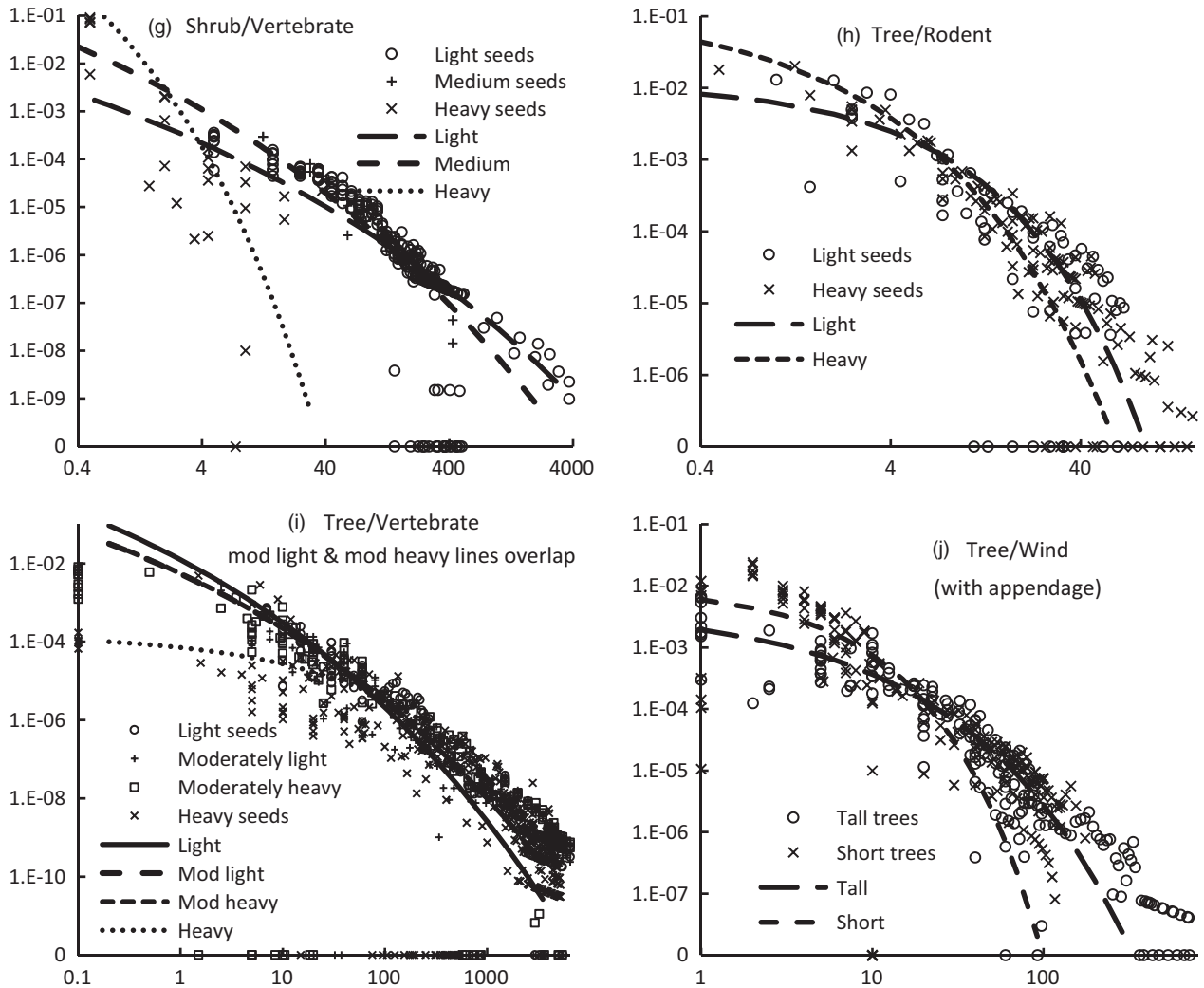


Fig. 2. continued

For trees, dispersal distances are of the order vertebrate > wind (with appendage) > rodent. For shrubs, vertebrate > ant. However, for herbs, wind-dispersal of dispersules with appendages did not result in longer dispersal distances than by wind without appendages, ant or ballistic.

Taller herbs, graminoids or trees had greater median and 95th percentile distances than shorter plants within the same growth form/dispersal mode combination, and these differences were substantial (Table 3; Fig. 2c,e,j; Fig. S1c,e,j) and consistent between the ExP and LogS fits. Seed mass did not show a consistent pattern of effects on animal-dispersed kernels. For the ExP fits, lighter seeds were dispersed further for ant-dispersed herbs (Fig. 2a), vertebrate-dispersed shrubs (Fig. 2g) and rodent-dispersed trees (Fig. 2h), but the pattern was reversed for vertebrate-dispersed trees (Fig. 2i) and there was little difference between seed mass classes for ant-dispersed shrubs (Fig. 2f). Furthermore, the patterns changed to some degree when using the LogS function in that this indicated heavier seeds dispersed further in ant-dispersed herbs (Fig. S1a) and ant-dispersed shrubs (Fig. S1f) and no pattern for rodent-dispersed trees (Fig. S1h). This indicates a strong

pattern for plant height effects on wind dispersal, but an inconsistent pattern for seed mass effects on animal dispersal.

The two functions ExP and LogS described similar dispersal kernels for each combination of data sets, as described above. However, these functions have different intrinsic shapes (Nathan *et al.* 2012) and were fitted to slightly different data sets. The modelled median and 95th percentile dispersal distances therefore differed between the ExP and LogS fitted to each combination of data sets (Table 1). This difference for the median (in terms of the absolute % difference of the LogS value from the ExP value) ranged from 3% to 476%, with a median of 48%. However, one function did not give a consistently higher or lower median value than the other.

Discussion

In this paper, we present standardized dispersal kernels for 168 case studies representing a range of plant types across a wide geographic range, as well as generalized kernels for well-studied plant growth form/dispersal mode combinations.

While it would be preferable to determine dispersal kernels directly in the system(s) one is studying, dispersal is notoriously difficult and resource-consuming to measure (Bullock, Shea & Skarpaas 2006). It is indicative of this difficulty that in contrast to our collation of 168 dispersal data sets for 144 species, the COMPADRE data base of plant demographic matrices currently includes 637 species and 6242 matrices (<http://www.compadre-db.org/Data/Compadre>; accessed 02/06/16) and has many more to be added. Potential uses of our synthesis of empirical dispersal data include the following: (i) choosing appropriate dispersal functions in generic modelling studies; (ii) selecting informative dispersal kernels for one's empirical study system; and (iii) using representative dispersal kernels in cross-taxon comparative studies. Below we expand on each of these uses and discuss points that arise, including those relating to data quality.

CHOOSING APPROPRIATE DISPERSAL FUNCTIONS

Many functions are used to describe empirical dispersal kernels, and the choice for a particular study often seems to arise from personal preference or experience. The Gaussian and exponential have statistically mechanistic basis, in that they describe a movement process (Petrovskii & Morozov 2009). Some other functions have a similar philosophy. For example, the 2Dt is a continuous mixture of Gaussian kernels with variance parameters distributed as the inverse of a gamma distribution (Clark *et al.* 1999). However, the functions used are often chosen as 2-parameter equations allowing high leptokurtosis and long (sometimes fat) tails, with little consideration of the underlying mechanism. This suggests that no one function will be the best as they are generally simple (to allow fitting) statistical descriptions of a complex process. However, few studies compare the fit of multiple functions to dispersal data. Martinez & Gonzalez-Taboada (2009) fitted exponential, 2Dt, log-normal and Weibull functions (as well as a Weibull-exponential mixture) to a number of vertebrate and wind-dispersed trees and shrubs in a forest system and found different functions performed better for different species, with no relation between dispersal mode and best-fit function. Clark *et al.* (2005) fitted exponential, Gaussian, inverse power (i.e. a power law function) and 2Dt functions to dispersal data for different trees in a forest plot. They found Gaussian and 2Dt functions best fitted wind-dispersed seeds, while the inverse power best fitted the bird- and monkey-dispersed seeds. Greene *et al.* (2004), again considering trees from a variety of studies, found the log-normal performed better than the 2Dt or Weibull.

For this study, we opted for an objective selection of dispersal functions, using 11 listed in a review by Nathan *et al.* (2012). This approach gave insights into the performance of different functions when fitted to a wide range of dispersal kernels. The log-sech (LogS) and exponential power (ExP) showed the best fits to the data sets. It is difficult to pinpoint why these two functions perform best, but this finding suggests these might be useful general functions to use in models. However, these functions have quite different properties

and histories of use in dispersal studies. The LogS has been used very rarely, just in one study of dispersal of birds (Van Houtan *et al.* 2010). The LogS becomes the Cauchy for $b = 1$ (Nathan *et al.* 2012), which is occasionally used in seed dispersal studies (Seri, Maruvka & Shnerb 2012; Munoz *et al.* 2013). The LogS also has unattractive properties, such as no mean value for $b \geq 1$, nor does it have moments. The ExP has been widely used on a variety of taxa (Nathan *et al.* 2012) and on many seed dispersal data since Clark *et al.* (1998). It has useful properties such as a mean value (eqn 2), has moments and it reduces to the exponential or Gaussian for certain values of b . An interesting finding was that of the 144 data sets showing a good fit of the ExP (i.e. $r^2 > 0.7$), 117 had $b < 1$, which indicates a fat tail. Similarly, nine of the 10 ExP functions estimated for the major growth form/dispersal mode groupings had $b < 1$, as did 17 of the 19 plant height/seed mass subdivisions of these groupings. As an illustration of the influence of the b parameter, 57 of the 168 data sets had both the ExP and the exponential in the best-fit group and had $b < 1$ for the ExP (10 had $b > 1$). The median value of the ExP b in these 57 data sets was 0.445 and the 95th percentile (as a measure of the tail) of the fitted function was a median of 234% more than that of the fitted exponential. Fat-tailed kernels are a popular concept in dispersal ecology (Klein, Lavigne & Gouyon 2006; Nathan *et al.* 2008), but this property of the ExP allows an explicit test whether empirical kernels are indeed fat-tailed. Because of these properties, the ExP might be more generally useful than the LogS.

Our suggestion that no one function would be expected to fit all data sets well is supported by the fact that no function was always in the best-fit group for each data set. However, there was a strong differentiation among the functions in how well they were represented in the best-fit groups. Certain functions are very popular in the wider dispersal literature, such as the log-normal (Greene *et al.* 2004), which performed very badly here; or the 2Dt (Jones & Muller-Landau 2008) which performed fairly well. It is particularly interesting that the WALD performed poorly for wind-dispersed data sets. The WALD is an analytical formulation of a wind-dispersal model (Katul *et al.* 2005) and has been used widely as a mechanistic model to generate dispersal kernels (Skarpaas & Shea 2007; Bullock *et al.* 2012). While its poor performance when used in this paper as a fitted function does not negate such uses, it does suggest that more validation is needed. The WALD has a density mode at a distance >0 and the fact that many of the 55 wind-dispersed data sets showed monotonically declining density with distance may explain its poor performance. It would be useful to revisit the WALD theory to analyse why it fails, such as the simplifying assumptions used upon the Langevin and the Fokker-Planck equations (Katul *et al.* 2005).

Our analysis suggests the Gaussian is a very poor representation of the dispersal process. This point has long been raised in the dispersal literature (Wallace 1966), but our analysis provides an objective affirmation. While its mathematical properties make the Gaussian popular in mathematical representations of dispersal, the low kurtosis and thin tail can lead to inaccurate predictions about dispersal distances and spatial

dynamics (Clark *et al.* 1998; Klein, Lavigne & Gouyon 2006). The second one-parameter function, the exponential, has had a similarly negative press to the Gaussian (Bullock & Clarke 2000; Nathan *et al.* 2012), but it performed much better than the Gaussian. While it did not perform as well as almost all the 2-parameter functions (but better than the log-normal – Table 1), our analyses suggest the exponential has some credence as a simple function to use both as a comparator for more complex functions when fitting data, and as a straightforward function in mathematical modelling studies. It might be argued that the good fit in many cases may reflect data quality or inadequate sampling at long distances, in that the tail of the kernel may not be well described. This point is countered by our analysis showing that the distance over which dispersal was measured in a study did not differ among the best-fit groups for the set of functions. This does raise the issue that while the tail of the dispersal kernel is often discussed in the literature and has a clear meaning in mathematical formulations of the kernel (Klein, Lavigne & Gouyon 2006), it is not clear what exactly the tail is in empirical measures of dispersal, or what a sufficient measurement of this tail looks like. Portnoy & Willson (1993) defined the tail as ‘the set of distance categories beyond the last clear mode of the data’. By this definition, visual examination of each data set suggested each represented a kernel tail to some degree (e.g. Fig. 1).

SELECTING INFORMATIVE DISPERSAL KERNELS

As dispersal is intrinsic to plant life histories, governing local and regional population dynamics, genetic structuring, evolutionary processes and community dynamics to name a few, knowledge of dispersal kernels should be a key aim in many studies. The relative scarcity of such information is therefore limiting in ecology. The individual case studies and the generic dispersal kernels presented in this paper are therefore a resource to help address this limitation. A researcher might choose one or more case studies that match to some extent the characteristics of their study system and/or a generalized dispersal kernel which does the same. While we provide generalized kernels for 10 growth form/dispersal mode combinations, there are a small number of case studies representing 13 other combinations. The potential uses of such kernels are multiple. If one has the luxury of choice among multiple case studies, selection could follow alternative criteria, such as fine-scale measures of short-distance dispersal when studying competition or the Janzen–Connell effect versus less detail, but measures over long distances, to assess ability to spread or persist in fragmented landscapes.

Furthermore, the ExP and LogS functions gave somewhat different generalized kernels, as illustrated by variation in the predicted medians and 95th percentiles. This partially reflects the intrinsic differences in the functions – the LogS has a power-law tail and a weak effect of distance close to the source, whereas the ExP represents a smoother decline with distance (Nathan *et al.* 2012). But the differences also reflect uncertainty in our analyses due to variation among the data

sets within each group in terms of the ecology of each system and the data-gathering approach. Our parameterization of both functions will allow researchers to use them in combination and represent this uncertainty in the kernel.

Each individual case study, and thus each generalized kernel, relates to a single dispersal mode. Multiple dispersal agents may be involved in the dispersal of seeds from a plant or of an individual seed. Thus, the ‘total dispersal kernel’ (Nathan *et al.* 2008) of a plant or population may require combining kernels for multiple dispersal modes. To do so, one should convolve (Neubert & Parker 2004) the relevant dispersal distance probability density functions (pdf) – that is the dispersal location pdf/ $2\pi d$. Surprisingly, only three of the 107 studies considered dispersal by multiple modes: ballistic dispersal followed by ant dispersal (Beaumont, Mackay & Whalen 2009); and dispersal by vehicles versus that by wind (Arnold 1981; Bullock *et al.* 2003). The same approach could be used if, say, different animal species disperse the seed and one has individual kernels for each animal vector (Lehouck *et al.* 2009).

USING REPRESENTATIVE DISPERSAL KERNELS

Empirical dispersal kernels are useful for multispecies studies. These might involve modelling how fast species may spread, or potential distributions, under a changing climate. Current approaches use a small number of empirical data sets (Bullock 2012) or simple, predetermined dispersal functions, such as the exponential (Engler & Guisan 2009; Bocedi *et al.* 2014). A broader range of empirically determined kernels, applicable to a wide range of species, should allow more realistic and representative forecasts. Another use would be to represent dispersal in comparative studies. While plant demographic data have been used in comparative life-history analyses (Salguero-Gomez *et al.* 2016), dispersal information has not been included. Other comparative analyses use traits related to dispersal ability, such as plant height, seed mass and dispersal mode (Baeten *et al.* 2015). The use of information on the kernel itself should help better integrate a more rounded understanding of dispersal into such studies.

This potential raises the issue of the how dispersal kernels vary according to plant and vector characteristics. It should be noted that analysis of the differences among classifications is not the main focus of this paper, and the classifications are used more to give ecologists well-defined groups from which to draw dispersal kernels relevant to their needs. We fitted generalized ExP and LogS functions for growth form and dispersal syndrome combinations and were able to subdivide these further according to plant height (wind dispersal) or seed mass (animal dispersal) classes. These functions were fitted to groupings of data sets each of which encompassed a wide range of systems, a diversity of data-gathering methods and variety in the plants and dispersal vectors studied. For example, of the 30 papers (59 data sets) assessing non-rodent vertebrate dispersal, the majority studied dispersal by small- to medium-sized birds (15 papers, some papers studied more than one group), such as thrushes, tanagers or mockingbirds.

Larger birds such as toucans or cassowaries were covered in six papers, primates such as tamarins or spider monkeys in four and medium-sized omnivorous mammals such as possums and martens in three. Bears, deer, fruit bats, iguanas and a large freshwater fish were studied in one paper each, and two papers studied a broad, unspecified group of vertebrate dispersers. Thus, there is inherent variation among the data sets within each group. Despite this, the functions fitted to each group had generally high r^2 values, being >0.7 in the great majority of cases and being significant at $P < 0.05$ in all. Thus, we have confidence in the value of these generalized functions for wider use. The collecting of further dispersal data sets would allow further subdivision of data sets according to key variables such as vertebrate type, local wind speed, size of wind-dispersal appendage or vegetation type.

Because the data sets varied in terms of distance classes, maximum distance, seed densities and sampling methods, fitting functions across all the data with a and b varying according to group would not have been sensible or feasible. However, comparisons of the fitted curves and the dispersal quantiles allowed tentative conclusions about differences among these groups. Our data suggested average dispersal distances vary among growth forms in the order trees $>$ shrubs $>$ herbs, graminoids, and among dispersal modes in the order vertebrates (excluding rodents) $>$ rodents, wind (with appendage) $>$ wind (without appendage), ant, ballistic. Taller plants within groups disperse seeds further by wind, but seed mass has an inconsistent relationship with dispersal distance (see also Thomson *et al.* 2011). Variation in whether or not lighter seeded species disperse further than heavier seeded species may arise if larger seeds are dispersed by larger, more mobile animals (Nathan *et al.* 2008) and also where a varying number of seeds are dispersed together in a fruit (Jordano 2000).

Our findings build on the analysis by Willson (1993), who carried out loglinear regression analysis of collated dispersal kernels (i.e. fitting an exponential function). Interestingly, she concluded wind-dispersed herbs with seed appendages dispersed further than herbs using ballistic or ant dispersal, and wind-dispersed trees and shrubs dispersed further than those dispersed by vertebrates. These findings contradict ours and probably reflect the current availability of more data (60 papers compared to our 107) as well as the more sophisticated analytical methods that are now available. However, this does demonstrate the need to continue gathering dispersal data directly as it is likely conclusions will develop as more case studies become available.

MEASURING SEED DISPERSAL

Any statistical synthesis of multiple data sets is bound by the number and quality of the studies used. The 168 data sets represent a great effort by the researchers involved, and we hope this synthesis does them justice by using these studies to provide general information for the wider use of dispersal kernels in ecological research. However, the findings of this paper, especially the generalized dispersal kernels, are likely

to be much improved upon if further dispersal data are gathered. The fitted kernels provided here can also be used to inform direct seed dispersal studies. Skarpaas, Shea & Bullock (2005) detailed a Monte Carlo approach to designing seed dispersal studies efficiently using information on the likely dispersal kernel, which could be provided by case studies and generalized kernels given in this paper.

While the relatively small number of studies is generally limiting, we can identify three particular areas of research that require focussed activity. One is to use methods that allow better characterization of the tail, such as molecular markers or tracking animal dispersers (Bullock, Shea & Skarpaas 2006). The variety of methods used to gather data is likely to lead to uncertainty. While we could not analyse this formally due to the fact that methods used are strongly linked to the characteristics of the system studied, analysis of biases due to methods would be useful. We did find that estimates of tree seed dispersal by vertebrates were affected by the methods used. A number of studies combined measures of seed retention, such as gut passage time, with data on movement by the animal vector. Dispersal distances estimated by this method were in general further than dispersal measured by other methods, such as following vertebrate vectors, seed trapping or using molecular markers. It is unclear however, whether this represents a bias in the data as methods are usually chosen to match the researcher's understanding of the system, such as knowledge that seeds are being carried a long way by vertebrate dispersers (see Cortes & Uriarte 2013). While biases according to method are possible, they do not undermine the aim of this paper, which is to synthesize existing information of empirical dispersal kernels.

The second research area concerns our finding that dispersal of trees and by vertebrates are favourite study systems. Studies on other dispersal modes and growth forms would provide much needed data for relatively understudied dispersal systems; two in particular are dispersal of seeds by humans or by water. Finally, little is known about how dispersal varies among habitats. In Fig. S2, we examine the seven species in our analysis for which kernels were measured in different habitat types (usually in different studies, but through the same dispersal mode). The study methods and habitat contrasts were too inconsistent to allow patterns to be discerned. It is likely that habitat type and structure will affect the dispersal process (Westcott *et al.* 2005; Trakhtenbrot, Katul & Nathan 2014) and so more formal contrasts of kernels among habitat types would allow characterization of intraspecific variation in dispersal. Furthermore, synthesis of such data would be aided by more complete presentation of the gathered data in papers. Many papers we used presented data in graphs or other summary forms (e.g. summing across replicates). Analyses would have more power if data are published in their raw form, and we would encourage researchers to do so.

Acknowledgements

This research was supported by SCALES EU-FP7-226852, CEH project NEC05264, the Estonian Ministry of Education and Research (IUT20-29) and

the European Regional Development Fund (Centre of Excellence EcolChange). We thank David Westcott and Patrick Jansen for providing more detailed dispersal data for their pond apple and palm systems, respectively, and to all the researchers who worked hard to generate dispersal data across the World.

Data accessibility

The dispersal data for each of the 168 data sets are available from the Dryad Digital Repository: <http://dx.doi.org/10.5061/dryad.mq2ff> (Bullock et al. 2016).

References

- Arnold, R.M. (1981) Population dynamics and seed dispersal of *Chaenorrhinum minus* on railroad cinder ballast. *The American Midland Naturalist*, **106**, 80–91.
- Baeten, L., Davies, T.J., Verheyen, K., Van Calster, H. & Vellend, M. (2015) Disentangling dispersal from phylogeny in the colonization capacity of forest understorey plants. *Journal of Ecology*, **103**, 175–183.
- Beaumont, K.P., Mackay, D.A. & Whalen, M.A. (2009) Combining distances of ballistic and myrmecochorous seed dispersal in *Adriana quadripartita* (Euphorbiaceae). *Acta Oecologica-International Journal of Ecology*, **35**, 429–436.
- Bocedi, G., Palmer, S.C.F., Pe'er, G., Heikkinen, R.K., Matsinos, Y.G., Watts, K. & Travis, J.M.J. (2014) RangeShifter: a platform for modelling spatial eco-evolutionary dynamics and species' responses to environmental changes. *Methods in Ecology and Evolution*, **5**, 388–396.
- Bolker, B. & Pacala, S.W. (1997) Using moment equations to understand stochastically driven spatial pattern formation in ecological systems. *Theoretical Population Biology*, **52**, 179–197.
- Bullock, J.M. (2012) Plant dispersal and the velocity of climate change. *Dispersal Ecology and Evolution* (eds J. Clobert, M. Baguette, T.G. Benton & J.M. Bullock), pp. 366–377. Oxford University Press, Oxford, UK.
- 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.
- Bullock, J.M., Shea, K. & Skarpas, O. (2006) Measuring plant dispersal: an introduction to field methods and experimental design. *Plant Ecology*, **186**, 217–234.
- Bullock, J.M., Moy, I.L., Coulson, S.J. & Clarke, R.T. (2003) Habitat-specific dispersal: environmental effects on the mechanisms and patterns of seed movement in a grassland herb *Rhinanthus minor*. *Ecography*, **26**, 692–704.
- Bullock, J.M., Galsworthy, S.J., Manzano, P., Poschlod, P., Eichberg, C., Walker, K. & Wichmann, M.C. (2011) Process-based functions for seed retention on animals: a test of improved descriptions of dispersal using multiple data sets. *Oikos*, **120**, 1201–1208.
- Bullock, J.M., White, S.M., Prudhomme, C., Tansey, C., Perea, R. & Hooftman, D.A.P. (2012) Modelling spread of British wind-dispersed plants under future wind speeds in a changing climate. *Journal of Ecology*, **100**, 104–115.
- Bullock, J.M., Mallada González, L., Tamme, R., Götzenberger, R., White, S.M., Pärtel, M. & Hooftman, D.A.P. (2016) Data from: A synthesis of empirical plant dispersal kernels. *Dryad Digital Repository*, <http://dx.doi.org/10.5061/dryad.mq2ff>
- Caswell, H., Lensink, R. & Neubert, M.G. (2003) Demography and dispersal: life table response experiments for invasion speed. *Ecology*, **84**, 1968–1978.
- Clark, J.S. (1998) Why trees migrate so fast: confronting theory with dispersal biology and the paleorecord. *The American Naturalist*, **152**, 204–224.
- Clark, J.S., Fastie, C., Hurr, G. et al. (1998) Reid's paradox of rapid plant migration – dispersal theory and interpretation of paleoecological records. *BioScience*, **48**, 13–24.
- 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.
- Clark, C.J., Poulsen, J.R., Bolker, B.M., Connor, E.F. & Parker, V.T. (2005) Comparative seed shadows of bird-, monkey-, and wind-dispersed trees. *Ecology*, **86**, 2684–2694.
- Cortes, M.C. & Uriarte, M. (2013) Integrating frugivory and animal movement: a review of the evidence and implications for scaling seed dispersal. *Biological Reviews*, **88**, 255–272.
- Cousens, R.D., Dytham, C. & Law, R. (2008) *Dispersal in Plants: A Population Perspective*. Oxford University Press, Oxford, UK.
- Driscoll, D.A., Banks, S.C., Barton, P.S. et al. (2014) The trajectory of dispersal research in conservation biology. Systematic review. *PLoS ONE*, **9**, e95053.
- Engler, R. & Guisan, A. (2009) MigClim: predicting plant distribution and dispersal in a changing climate. *Diversity and Distributions*, **15**, 590–601.
- FAO (2010) Global Forest Resources Assessment 2010. FAO Forestry Paper 163.
- Gilbert, M.A., White, S.M., Bullock, J.M. & Gaffney, E.A. (2014) Spreading speeds for stage structured plant populations in fragmented landscapes. *Journal of Theoretical Biology*, **349**, 135–149.
- 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.
- Hallatschek, O. & Fisher, D.S. (2014) Acceleration of evolutionary spread by long-range dispersal. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, E4911–E4919.
- Harsch, M.A., Zhou, Y., HilleRisLambers, J. & Kot, M. (2014) Keeping pace with climate change: stage-structured moving-habitat models. *The American Naturalist*, **184**, 25–37.
- Hastings, A., Cuddington, K., Davies, K.F. et al. (2005) The spatial spread of invasions: new developments in theory and evidence. *Ecology Letters*, **8**, 91–101.
- Jones, F.A. & 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.
- Jordano, P. (2000) Fruits and frugivory. *Seeds- the Ecology of Regeneration in Plant Communities* (ed. M. Fenner), pp. 125–165. CAB International, Wallingford, UK.
- Katul, G.G., Porporato, A., Nathan, R., Siquiera, M., Soons, M.B., Poggi, D., Horn, H.S. & Levin, S.A. (2005) Mechanistic analytical models for long-distance seed dispersal by wind. *The American Naturalist*, **166**, 368–381.
- Kjellsson, G. (1985) Seed fate in a population of *Carex pilulifera* L. Seed dispersal and ant-seed mutualism. *Oecologia*, **67**, 416–423.
- Klein, E.K., Lavigne, C. & Gouyon, P.-H. (2006) Mixing of propagules from discrete sources at long distance: comparing a dispersal tail to an exponential. *BMC Ecology*, **6**, 1–12.
- Kot, M., Lewis, M.A. & van den Driessche, P. (1996) Dispersal data and the spread of invading organisms. *Ecology*, **77**, 2027–2042.
- Lehoucq, V., Spanhove, T., Demeter, S., Groot, N.E. & Lens, L. (2009) Complementary seed dispersal by three avian frugivores in a fragmented Afrotropical forest. *Journal of Vegetation Science*, **20**, 1110–1120.
- Martinez, I. & Gonzalez-Taboada, F. (2009) Seed dispersal patterns in a temperate forest during a mast event: performance of alternative dispersal kernels. *Oecologia*, **159**, 389–400.
- Munoz, F., Beeravolu, C.R., Pelissier, R. & Coutron, P. (2013) Do spatially-implicit estimates of neutral migration comply with seed dispersal data in tropical forests? *PLoS ONE*, **8**, e72497.
- Nakagawa, S. & Schielzeth, H. (2013) A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, **4**, 133–142.
- Nathan, R., Schurr, F.M., Spiegel, O., Steinitz, O., Trakhtenbrot, A. & Tsoar, A. (2008) Mechanisms of long-distance seed dispersal. *Trends in Ecology & Evolution*, **23**, 638–647.
- Nathan, R., Katul, G., Bohrer, G., Kuparinen, A., Soons, M., Thompson, S., Trakhtenbrot, A. & Horn, H. (2011) Mechanistic models of seed dispersal by wind. *Theoretical Ecology*, **4**, 113–132.
- Nathan, R., Klein, E., Robledo-Arnuncio, J.J. & Revilla, E. (2012) Dispersal kernels: review. *Dispersal Ecology and Evolution* (eds J. Clobert, M. Baguette, T.G. Benton & J.M. Bullock), pp. 187–210. Oxford University Press, Oxford, UK.
- Neubert, M.G. & Parker, I.M. (2004) Projecting rates of spread for invasive species. *Risk Analysis*, **24**, 817–831.
- O'Dwyer, J.P. & Green, J.L. (2010) Field theory for biogeography: a spatially explicit model for predicting patterns of biodiversity. *Ecology Letters*, **13**, 87–95.
- Ovaskainen, O. & Cornell, S.J. (2006) Asymptotically exact analysis of stochastic metapopulation dynamics with explicit spatial structure. *Theoretical Population Biology*, **69**, 13–33.
- Petrovskii, S. & Morozov, A. (2009) Dispersal in a statistically structured population: fat tails revisited. *The American Naturalist*, **173**, 278–289.
- Portnoy, S. & Willson, M.F. (1993) Seed dispersal curves – behavior of the tail of the distribution. *Evolutionary Ecology*, **7**, 25–44.
- Royal Botanic Gardens Kew (2015) Seed Information Database (SID). Version 7.1. <http://data.kew.org/sid/>.
- Salguero-Gomez, R., Jones, O.R., Jongejans, E., Blomberg, S.P., Hodgson, D.J., Mbeau-Ache, C., Zuidema, P.A., de Kroon, H. & Buckley, Y.M. (2016) Fast-slow continuum and reproductive strategies structure plant life-history variation worldwide. *Proceedings of the National Academy of Sciences of the United States of America*, **113**, 230–235.

- Seri, E., Maruvka, Y.E. & Shnerb, N.M. (2012) Neutral dynamics and cluster statistics in a tropical forest. *The American Naturalist*, **180**, E161–E173.
- Shaw, M.W., Harwood, T.D., Wilkinson, M.J. & Elliott, L. (2006) Assembling spatially explicit landscape models of pollen and spore dispersal by wind for risk assessment. *Proceedings of the Royal Society B-Biological Sciences*, **273**, 1705–1713.
- Skarpaas, O. & Shea, K. (2007) Dispersal patterns, dispersal mechanisms, and invasion wave speeds for invasive thistles. *The American Naturalist*, **170**, 421–430.
- Skarpaas, O., Shea, K. & Bullock, J.M. (2005) Optimizing dispersal study design by Monte Carlo simulation. *Journal of Applied Ecology*, **42**, 731–739.
- 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.
- Swaine, M.D. & Beer, T. (1977) Explosive seed dispersal in *Hura crepitans* L. (Euphorbiaceae). *New Phytologist*, **78**, 695–708.
- Tamme, R., Götzenberger, L., Zobel, M., Bullock, J.M., Hooftman, D.A.P., Kaasik, A. & Pärtel, M. (2014) Predicting species' maximum dispersal distances from simple plant traits. *Ecology*, **95**, 505–513.
- Tekiela, D.R. & Barney, J.N. (2013) Quantifying *Microstegium vimineum* seed movement by non-riparian water dispersal using an ultraviolet-marking based recapture method. *PLoS ONE*, **8**, e63811.
- Thompson, S.E., Assouline, S., Chen, L., Trahtenbrot, A., Svoray, T. & Katul, G.G. (2014) Secondary dispersal driven by overland flow in drylands: review and mechanistic model development. *Movement Ecology*, **2**, 7.
- Thomson, F.J., Moles, A.T., Auld, T.D. & Kingsford, R.T. (2011) Seed dispersal distance is more strongly correlated with plant height than with seed mass. *Journal of Ecology*, **99**, 1299–1307.
- Trahtenbrot, A., Katul, G.G. & Nathan, R. (2014) Mechanistic modeling of seed dispersal by wind over hilly terrain. *Ecological Modelling*, **274**, 29–40.
- Travis, J.M.J., Delgado, M., Bocedi, G. *et al.* (2013) Dispersal and species' responses to climate change. *Oikos*, **122**, 1532–1540.
- Tummers, B. (2006) Datathief III. <http://datathief.org/>.
- Van Houtan, K.S., Bass, O.L. Jr, Lockwood, J. & Pimm, S.L. (2010) Importance of estimating dispersal for endangered bird management. *Conservation Letters*, **3**, 260–266.
- Wada, N. & Uemura, S. (1994) Seed dispersal and predation by small rodents on the herbaceous understory plant *Symplocarpus renifolius*. *The American Midland Naturalist*, **132**, 320–327.
- Wallace, B. (1966) On the dispersal of *Drosophila*. *The American Naturalist*, **100**, 551–563.
- Westcott, D.A., Bentrupperbaumer, J., Bradford, M.G. & McKeown, A. (2005) Incorporating patterns of disperser behaviour into models of seed dispersal and its effects on estimated dispersal curves. *Oecologia*, **146**, 57–67.
- Willson, M.F. (1993) Dispersal mode, seed shadows, and colonization patterns. *Vegetatio*, **108**, 261–280.

Received 28 June 2016; accepted 22 August 2016

Handling Editor: Mark Rees

Supporting Information

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

Table S1. Tests of whether the 142 datasets selected for fitting generalised log-sech or the 144 selected for the exponential power were biased sub-sets of the complete set of 168 datasets in terms of plant characteristics.

Table S2. AIC values for linear mixed model analysis of how means of the exponential power fitted to each of the 144 case studies for which $r^2 \geq 0.7$ are affected by all combinations of the variables growth form, dispersal mode, plant height and seed mass.

Table S3 The number of the 144/142 selected datasets for Exp/LogS falling within in each combination of growth form and dispersal mode.

Figure S1. Generalised dispersal kernels for all 10 growth form/dispersal mode combinations for which there were sufficient multi-species datasets.

Figure S2. The seven species/dispersal mode combinations for which there were multiple datasets.

Table S4. Details of the 168 datasets along with supplementary information and the source of the dispersal data.

Table S5. Parameter values for the best-fitting functions describing the dispersal location kernels for each dataset, following the numbering in Table S3.