

Article (refereed) - postprint

Bullock, James M.; Mallada Gonzalez, Laura; Tamme, Riin; Götzenberger, Lars; White, Steven M.; Pärtel, Meelis; Hooftman, Danny A.P.. 2017 A synthesis of empirical plant dispersal kernels. *Journal of Ecology*, 105 (1). 6-19. [10.1111/1365-2745.12666](https://doi.org/10.1111/1365-2745.12666)

© 2016 The Authors. *Journal of Ecology* © 2016 British Ecological Society

This version available <http://nora.nerc.ac.uk/515569/>

NERC has developed NORA to enable users to access research outputs wholly or partially funded by NERC. Copyright and other rights for material on this site are retained by the rights owners. Users should read the terms and conditions of use of this material at

<http://nora.nerc.ac.uk/policies.html#access>

This document is the author's final manuscript version of the journal article, incorporating any revisions agreed during the peer review process. Some differences between this and the publisher's version remain. You are advised to consult the publisher's version if you wish to cite from this article.

The definitive version is available at

<http://besjournals.onlinelibrary.wiley.com/hub/>

Contact CEH NORA team at

noraceh@ceh.ac.uk

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. Hooftman^{1,6}

1. NERC Centre for Ecology & Hydrology, Benson Lane, Wallingford, Oxfordshire, OX10 8BB, UK.

2. Evolution & Ecology Research Centre, School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, NSW 2052, Australia.

3. Institute of Ecology and Earth Sciences, University of Tartu, Lai 40, Tartu 51005, Estonia.

4. Institute of Botany, Czech Academy of Sciences, Dukelská 135, 37982 Třeboň, Czech Republic.

5. Wolfson Centre for Mathematical Biology, Mathematical Institute, Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG, UK.

6. Lactuca: Environmental Data Analyses and Modelling, Diemen, 1112NC, The Netherlands.

Correspondence author: James M. Bullock, jmbul@ceh.ac.uk

Running headline: A synthesis of empirical dispersal kernels

18 **Summary**

- 19 1. Dispersal is fundamental to ecological processes at all scales and levels of organisation but progress is
20 limited by a lack of information about the general shape and form of plant dispersal kernels. We
21 addressed this gap by synthesising empirical data describing seed dispersal and fitting general dispersal
22 kernels representing major plant types and dispersal modes.
- 23 2. A comprehensive literature search resulted in 107 papers describing 168 dispersal kernels for 144
24 vascular plant species. The data covered 63 families, all the continents except Antarctica, and the broad
25 vegetation types of forest, grassland, shrubland, and more open habitats (e.g. deserts). We classified
26 kernels in terms of dispersal mode (ant, ballistic, rodent, vertebrates other than rodents, vehicle or
27 wind), plant growth form (climber, graminoid, herb, shrub or tree), seed mass and plant height.
- 28 3. We fitted 11 widely-used probability density functions to each of the 168 datasets to provide a statistical
29 description of the dispersal kernel. The Exponential Power (ExP) and Log-sech (LogS) functions
30 performed best. Other 2-parameter functions varied in performance. For example, the Lognormal and
31 Weibull performed poorly, while the 2Dt and Power law performed moderately well. Of the single-
32 parameter functions, the Gaussian performed very poorly, while the Exponential performed better. No
33 function was among the best-fitting for all datasets.
- 34 4. For 10 plant growth form/dispersal mode combinations for which we had >3 datasets, we fitted ExP and
35 LogS functions across multiple datasets to provide generalised dispersal kernels. We also fitted these
36 functions to sub-divisions of these growth form/dispersal mode combinations in terms of seed mass (for
37 animal-dispersed seeds) or plant height (wind-dispersed) classes. These functions provided generally
38 good fits to the grouped datasets, despite variation in empirical methods, local conditions, vegetation
39 type and the exact dispersal process.
- 40 5. *Synthesis.* We synthesise the rich empirical information on seed dispersal distances to provide
41 standardised dispersal kernels for 168 case studies and generalised kernels for plant growth
42 form/dispersal mode combinations. Potential uses include: a) choosing appropriate dispersal functions

43 in mathematical models; b) selecting informative dispersal kernels for one's empirical study system; and
44 c) using representative dispersal kernels in cross-taxon comparative studies.

45

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

48 **Introduction**

49 Good estimates of plant dispersal distances are required by ecologists in their attempts to understand and
50 model processes such as local interactions (Bolker & Pacala 1997), species' ability to track climate change
51 (Travis *et al.* 2013), population dynamics in fragmented landscapes (Gilbert *et al.* 2014), invasions (Hastings
52 *et al.* 2005), metapopulation persistence (Ovaskainen & Cornell 2006), and evolutionary change
53 (Hallatschek & Fisher 2014). Furthermore, a recent systematic review concluded that limited dispersal
54 knowledge often compromises conservation planning (Driscoll *et al.* 2014). The holy grail for dispersal
55 ecologists is to develop mechanistic models of dispersal, which represent the physical processes involved in
56 seed movement. These allow predictions of seed dispersal *de novo* and avoid the great effort required to
57 measure dispersal directly (Bullock, Shea & Skarpaas 2006). Mechanistic models and have been realised for
58 wind dispersal (Nathan *et al.* 2011), and their predictions have been used in models of spatial population
59 dynamics (Skarpaas & Shea 2007; Soons & Bullock 2008; Bullock *et al.* 2012). While there is some progress
60 towards mechanistic representations of seed dispersal by certain other vectors, such as by vertebrates
61 (Bullock *et al.* 2011; Cortes & Uriarte 2013) or water (Thompson *et al.* 2014), we remain a long way from
62 mechanistic models applicable to every seed dispersal process.

63 While work towards mechanistic models should continue, remarkably little use has been made of
64 empirical plant dispersal kernels in elucidating general information for use by ecologists. Plant dispersal
65 data, gathered for a wide range of species across a variety of habitats, are a rich source of information on
66 dispersal patterns. Over 20 years ago, Willson (1993) presented an important analysis of seed dispersal
67 kernels, showing how dispersal distances vary according to plant life form (i.e. herb, shrub, tree) and
68 dispersal mode (i.e. ballistic, wind, vertebrate). Many more kernels have been published since 1993 and
69 statistical methods have moved on (Willson analysed kernels using least squares linear regression). Two
70 recent papers have synthesised empirical plant dispersal data, relating mean or maximum dispersal distance
71 to variables such as dispersal mode, plant growth form, seed release height, seed mass or terminal velocity
72 (Thomson *et al.* 2011; Tamme *et al.* 2014). These analyses are useful – for example Tamme *et al.* (2014)
73 provided R code to predict the maximum dispersal distance from simple plant traits – they offer no

74 information on the complete set of dispersal distances (i.e. the shape of the kernel), which is necessary for
75 many models of spatial dynamics. For example, modelled population spread rates are greatly influenced by
76 the form and extent of the tail of the dispersal kernel (Kot, Lewis & van den Driessche 1996; Caswell, Lensink
77 & Neubert 2003). There is great potential for synthesising published dispersal data, to provide ecologists
78 with general dispersal kernels for different plant types and dispersal modes.

79 It is common to fit statistical functions to dispersal data, providing so-called phenomenological
80 dispersal kernels, which are a useful summary of dispersal patterns for activities such as modelling
81 population spread or linking demography to dispersal patterns (Bullock, Shea & Skarpaas 2006). A variety
82 of functions have been implemented; Nathan *et al.* (2012) discuss 13 simple (i.e. one or two parameter)
83 functions that have been used in different studies. These functions have also been used in different ways,
84 with inconsistency in what data are fitted and how the function is interpreted (Cousens, Dytham & Law
85 2008; Nathan *et al.* 2012). A dispersal kernel is a probability density function of dispersal distances, and the
86 formulation depends on whether it represents the distribution of distances dispersed (the dispersal distance
87 kernel) or the distribution of the final locations of dispersers (the dispersal location kernel) (Nathan *et al.*
88 2012).

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

100 justification for the specific function(s) chosen. This suggests a lack of cohesion in approaches to modelling
101 dispersal, which is likely to constrain progress.

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

108 **Methods**

109 EMPIRICAL DISPERSAL DATA

110 We examined the literature for data describing seed dispersal kernels. Our initial source was our analysis
111 (Tamme *et al.* 2014) of maximum dispersal distances, which brought together previous studies that
112 summarized published dispersal information, and also did a search on ISI Web of Science (WoS) using the
113 keywords: “seed”, “dispersal” and “distance”. We repeated this WoS search between 31st January 2012
114 (Tamme *et al.*'s end date) and 1st November 2014. All papers were examined for data describing seed
115 distributions with distance from a source. Published data were included in our analyses according to the
116 following criteria. i) The data described dispersal from a discrete source in 2-dimensional space, such as a
117 single plant, a small group of plants, or a single depot (e.g. in vertebrate dispersal studies). Linear sources
118 or large patches were excluded as the resulting dispersal kernels could not reasonably be characterised as
119 having a point source (different kernels could be fitted (Shaw *et al.* 2006), but comparison across datasets
120 would be difficult). ii) The data covered at least four distance locations, to provide a reasonable description
121 of the kernel. In fact of the datasets included none had only four points; one had five, the rest more. iii) The
122 data comprised specific distances (‘traps’) or distance classes (‘bins’) and at each of these a measure of seed
123 density (absolute or relative), number of seeds, or the proportion of dispersed (i.e. all seeds found) or all
124 (i.e. those counted or placed at the discrete source) seeds. Lists of individual seed dispersal distances were
125 included. iv) If the data described seedlings/plants at distances rather than seeds, they were excluded as
126 these are the outcome of spatial variation in recruitment and survival as well as dispersal. v) Data generated
127 by fitted kernels (e.g. by inverse modelling) or mechanistic modelling were excluded as these assume an
128 underlying statistical function. vi) Data projected by combining empirical information on vector movement
129 and seed retention/deposition (e.g. bird movement and gut retention times) were included. This last
130 decision allowed us to include a large number of datasets for vertebrate dispersal: three for climbers, one
131 for herbs, 11 for shrubs and 17 for trees. However, dispersal estimated by this method was significantly
132 further than when estimated by other methods (e.g. following dispersers or genetic markers). The mean
133 dispersal distances (estimated using the fitted Exponential Power parameters – see below) of trees were

134 greater for datasets using seed retention methods (median of the means distances = 3106 m. $n = 17$) than
135 for those using other methods (median = 168 m. $n = 17$) (Kruskal Wallis $p=0.02$). There were too few datasets
136 to carry out this test on other growth forms. Nevertheless, we retained these datasets, and address this
137 issue in the Discussion.

138 Data were extracted from tables or, in the majority of cases, from figures, in which case we used
139 the Datathief software (Tummers 2006). As the data were presented in a variety of forms, we needed to
140 represent all dispersal kernels in a common format for analysis and comparison. The most robust approach
141 was to convert all data into paired seed densities (i.e. m^{-2}) and distances, representing the so-called dispersal
142 location kernel (Nathan *et al.* 2012). This accommodated spatial discontinuities (i.e. not all distances had
143 associated data) and the common binning of data into distance classes. This gave us 107 papers presenting
144 168 dispersal kernels. Some papers reported multiple dispersal datasets, and these were either for different
145 plant species, for different vectors (in terms of the dispersal mode, e.g. wind vs vertebrate; or the vector
146 species) of the same species, or for the same species in different environments (e.g. forest vs clearing).
147 Replicate kernels for a species in the same situation (i.e. vector and environment the same) were analysed
148 together, accounting for replication. Occasionally different papers contained versions of the same dataset;
149 for these we selected the most comprehensive dataset.

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

154 We added supplementary data to each dispersal dataset. i) The dispersal mode: ant, ballistic,
155 rodent, vertebrates other than rodents (separated as rodents were a large class and other vertebrates are
156 more mobile), vehicle (cars, etc), wind. In this paper, we use the term dispersal mode (e.g. Willson *et al.*
157 1993), which is synonymous with the term dispersal syndrome as used by Tamme *et al.* (2014) and Thomson
158 *et al.* (2011). Wind-dispersed species were separated into those with an appendage that facilitates wind
159 dispersal such as samaras or plumes, vs those with simple seeds/fruits. Modes were taken from the original

160 publications. Each dataset had a specific, stated mode, although different datasets for the same species
161 sometimes represented different modes (reflecting that many species are dispersed by multiple vectors
162 (Bullock, Shea & Skarpaas 2006)). In the case of dispersal by vertebrates, some datasets represented
163 dispersal by a single species (30), while others represented dispersal by multiple species (42). ii) The growth
164 form: climber, graminoid, herb, shrub or tree. Classifications were taken, and sometimes simplified (e.g.
165 grasses and sedges into graminoids) to avoid too many classes, from the original publications. iii) Seed mass
166 (i.e. of the seed, not necessarily the dispersule) from Tamme *et al.* (2014), the original publication or online
167 sources, especially the Seed Information Database (Royal Botanic Gardens Kew 2015). iv) Plant height from
168 Tamme *et al.* (2014), the original publication or online sources. v) The plant family, the country and
169 continent, and the vegetation type; taken from the original paper. vi) The maximum distance to which
170 dispersal was measured. This represented a wide range; of the maximum measurement distances, the
171 median was 45m, the maximum 6,500m and the minimum 0.2m. But all these kernels showed clear distance
172 patterns (i.e. none showed a few similar densities close to the source, representing a very partial description
173 of the kernel).

174 FITTING DISPERSAL FUNCTIONS

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

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

$$199 \quad r^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad \text{eqn 1}$$

200 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
 201 mean value.

202 GENERALISING DISPERSAL KERNELS

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

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

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

$$219 \quad \text{mean}(ExP) = a \frac{\Gamma(3/b)}{\Gamma(2/b)} \quad \text{eqn 2}$$

220 (Γ is the gamma function). The LogS has an unspecified mean for $b>1$ (which was found for 60 datasets). We
221 used the ExP mean to compare the datasets in terms of how plant characteristics affected the ExP kernel.
222 We calculated the ExP mean dispersal distance for each of the 144 case studies with $r^2 \geq 0.7$, and tested for
223 differences in this mean according to four principal plant characteristics: growth form, dispersal mode, plant
224 height and seed mass. These were shown by Tamme et al. (2014) to be good predictors of measured
225 maximum dispersal distance, and they represent straightforward classifications of plants into types that
226 might be used in assigning dispersal kernels for modelling or other studies. We fitted linear models using
227 SAS Proc MIXED to the ExP means using all combinations of these four factors (seed mass and plant height
228 were \log_{10} transformed). Interaction terms were not included as the data were unbalanced and collinear.
229 The model with lowest AIC comprised Growth form + Dispersal Mode + Plant height (AIC= 405.8, $r^2=0.56$),
230 with one other within 4 AIC, i.e. Growth form + Dispersal Mode + Plant height + Seed mass (AIC = 405.9)
231 (see Table S2 for the full analysis). We also ran the best performing linear model combining classes which
232 might be expected to have similar dispersal kernels: graminoids and herbs (AIC = 410.1), climbers and trees
233 (405.8), rodents and other vertebrates (426.4), and vehicles and vertebrates (406.7). None of these reduced
234 the AIC so we kept the full set of growth forms and dispersal modes in further analyses.

235 We therefore fitted general ExP and LogS dispersal functions to the $r^2 \geq 0.7$ case studies grouped
236 according to growth form and dispersal mode. For both ExP and LogS 10 form/mode primary combinations
237 had a reasonable number (>3) of case studies to fit general functions (Table S3); we considered ≤ 3 cases to
238 be too affected by the particularities of the individual case studies. Preliminary analyses showed an
239 alternative approach using plant height or seed mass as modifiers of the a and/or b parameters was not
240 effective and led to poor model convergence. To use the information provided by these extra variables,
241 where there were sufficient datasets we also fitted functions to subdivisions of the primary form/mode
242 groups, using the variable most likely to be important for that group: seed mass for animal-dispersed groups
243 and plant height for wind-dispersed groups (Table 3). While seed mass rather than plant height would be
244 expected to affect animal dispersal, both might be important for wind dispersal. Thomson *et al.* (2014)
245 found that plant height is a much more important predictor of dispersal distance than seed mass, so we
246 chose this as the grouping criterion for wind-dispersed seeds. The number and bounds of the subdivisions
247 were arbitrary, but based on the number of datasets and the distribution of values of these variables, using
248 similar subdivisions across the groups for comparability (Table 3). For both ExP & LogS we fitted Q , a and b
249 to the density data, with each case and each replicate within a case was included as random factor that
250 affected Q . Thus we searched for common a and b values across all cases, in accordance with the hypothesis
251 that all studies within a specific growth form/dispersal mode (+ seed mass or plant height class) combination
252 followed the same underlying probability density function.

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

258 **Results**

259 FITTING DISPERSAL FUNCTIONS TO 168 DATASETS

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

268 Of the 11 functions, the Exponential Power, 2Dt, Logistic, Gamma & Log-sech converged for all
269 datasets, but the other functions did not always converge (Table 1). This poor convergence was only
270 substantial for the Gaussian (37 datasets), which also showed poorest fit with $\Delta AICc \leq 4$ for only 30 datasets.
271 The other one parameter function, the Exponential, had a much better performance, being in the best-fit
272 group in 68 cases. The Lognormal and the Weibull also performed rather poorly. The other functions –
273 Exponential power, 2Dt, Power law, WALD, Logistic, Log-sech and Gamma – all performed better, but all fell
274 outside the best-fit group in many cases. Interestingly, the WALD, although based on a mechanistic
275 representation of dispersal by wind (Katul *et al.* 2005), was in the best-fit group of only 15 of the 55 wind
276 dispersal datasets and by this metric performed better than only the Exponential, Gaussian and Lognormal.
277 The r^2 values showed generally very good fit with high values (in terms of the median value and number of
278 datasets for which $r^2 \geq 0.7$; Table 1) for most functions except the Gaussian and Lognormal, and the patterns
279 of r^2 values followed those of the AICc values. Fig. 1 shows some example plots of data with the fitted
280 functions. Tables S4 and S5 present the fitted parameter values for all well-fitting functions for each of the
281 168 datasets and the supplementary data describing the species and study system.

282 The Exponential power (Exp; $\Delta AICc \leq 4$ for 111 cases) and Log-sech (LogS; 119) clearly outperformed
283 the other functions (but note these numbers are $\ll 168$). The Exp can reduce to an Exponential ($b=1$) or

284 Gaussian ($b=2$), and can fit a fat tail ($b<1$) or a thin tail ($b>1$). Of the 144 datasets selected as showing good
285 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
286 fat-tailed. The LogS is always fat-tailed (power-law tail) and flattens (for $b<1$) or decreases ($b>1$) close to 0
287 distance. Of the 142 datasets with good general fit to the LogS, 90 had $b<1$ and 52 $b>1$.

288 Although the case study composition of the best-fit group differed among the functions, there was
289 no bias in comparison to the full set of case studies in terms of dispersal mode ($\chi^2 = 45.9$, $df = 40$, $p=0.24$),
290 growth form ($\chi^2 = 21.7$, $df = 20$, $p=0.36$), plant height (Kruskal Wallis $H = 9.6$, $df = 10$, $p=0.48$) or seed mass
291 ($H = 12.5$, $df = 10$, $p=0.25$). We also asked whether the study design affected the set of best-fit functions
292 (e.g. was a function allowing a fatter tail less likely to fit kernels measured over shorter distances?), but the
293 maximum distance over which dispersal was measured did not differ among the best-fit groups ($H = 15.4$,
294 $df = 10$, $p=0.12$).

295 GENERALISED DISPERSAL KERNELS

296 The fitting of the ExP and LogS functions across grouped datasets provided generalised dispersal kernels for
297 a number of well-studied plant growth form/dispersal mode combinations (Table 3, Fig. 2). Across the 10
298 combinations both the ExP and LogS gave similar and generally good fits, with r^2 ranging from 0.39-0.97
299 (median 0.73, $r^2 < 0.7$ for 3 combinations) for the former and 0.32-0.97 (median 0.78, $r^2 < 0.7$ for 3) for the
300 latter; although the patterns of r^2 values across the form/mode combinations differed between ExP and
301 LogS (note that it is not appropriate to compare the fits of the ExP and LogS using AIC as they were fitted to
302 different sets of case studies). Similar results were seen for the sub-divisions according to seed mass or
303 plant height classes, with $r^2 < 0.7$ for 12 of 38 groups across the ExP and LogS functions.

304 Median dispersal distances of the fitted functions were in the order Trees (ExP mean of the medians
305 = 20.9 m) > Shrubs (3.05 m) > Herbs and Graminoids (0.38 m). Vertebrates (excluding rodents; ExP mean of
306 the medians = 22.5 m) dispersed seeds further than Rodents (5.45 m), which dispersed seeds similar
307 distances to Wind for dispersules with an appendage (8.75 m), while dispersal by Ants (0.87 m), Ballistic
308 (0.45 m) and Wind for dispersules with no appendage (0.20 m) resulted in the shortest median distances.
309 The same patterns were found for the 95th percentile and for the LogS function (Table 3). While these

310 means are somewhat confounded as growth form and dispersal mode were not found in all possible
311 combinations, more specific comparisons show the same patterns. For Trees, dispersal distances are of the
312 order Vertebrate > Wind (with appendage) > Rodent. For Shrubs, Vertebrate > Ant. However, for Herbs,
313 Wind-dispersal of dispersules with appendages did not result in longer dispersal distances than by Wind
314 without appendages, Ant or Ballistic.

315 Taller Herbs, Graminoids or Trees had greater median and 95th percentile distances than shorter
316 plants within the same growth form/dispersal mode combination, and these differences were substantial
317 (Table 3; Fig 2c,e,j; Fig S1c,e,j) and consistent between the ExP and LogS fits. Seed mass did not show a
318 consistent pattern of effects on animal-dispersed kernels. For the ExP fits, lighter seeds were dispersed
319 further for Ant-dispersed Herbs (Fig. 2a), Vertebrate-dispersed Shrubs (Fig. 2g) and Rodent-dispersed Trees
320 (Fig. 2h), but the pattern was reversed for Vertebrate-dispersed Trees (Fig. 2i) and there was little difference
321 between seed mass classes for Ant-dispersed Shrubs (Fig. 2f). Furthermore, the patterns changed to some
322 degree when using the LogS function in that this indicated heavier seeds dispersed further in Ant-dispersed
323 Herbs (Fig. S1a) and Ant-dispersed Shrubs (Fig. S1f) and no pattern for Rodent-dispersed Trees (Fig. S1h).
324 This indicates a strong pattern for plant height effects on wind dispersal, but an inconsistent pattern for
325 seed mass effects on animal dispersal.

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

333 **Discussion**

334 In this paper we present standardised dispersal kernels for 168 case studies representing a range of plant
335 types across a wide geographic range, as well as generalised kernels for well-studied plant growth
336 form/dispersal mode combinations. While it would be preferable to determine dispersal kernels directly in
337 the system(s) one is studying, dispersal is notoriously difficult and resource-consuming to measure (Bullock,
338 Shea & Skarpaas 2006). It is indicative of this difficulty that in contrast to our collation of 168 dispersal
339 datasets for 144 species, the COMPADRE database of plant demographic matrices currently includes 637
340 species and 6242 matrices (<http://www.compadre-db.org/Data/Compadre>; accessed 02/06/16) and has
341 many more to be added. Potential uses of our synthesis of empirical dispersal data include: 1) Choosing
342 appropriate dispersal functions in generic modelling studies; 2) Selecting informative dispersal kernels for
343 one's empirical study system; 3) Using representative dispersal kernels in cross-taxon comparative studies.
344 Below we expand on each of these uses and discuss points that arise, including those relating to data quality.

345 CHOOSING APPROPRIATE DISPERSAL FUNCTIONS

346 Many functions are used to describe empirical dispersal kernels and the choice for a particular study often
347 seems to arise from personal preference or experience. The Gaussian and Exponential have statistically
348 mechanistic basis, in that they describe a movement process (Petrovskii & Morozov 2009). Some other
349 functions have a similar philosophy. For example, the 2Dt is a continuous mixture of Gaussian kernels with
350 variance parameters distributed as the inverse of a Gamma distribution (Clark *et al.* 1999). However, the
351 functions used are often chosen as 2-parameter equations allowing high leptokurtosis and long (sometimes
352 fat) tails, with little consideration of the underlying mechanism. This suggests that no one function will be
353 the best as they are generally simple (to allow fitting) statistical descriptions of a complex process. However,
354 few studies compare the fit of multiple functions to dispersal data. Martinez and Gonzalez-Taboada (2009)
355 fitted Exponential, 2Dt, Log-normal and Weibull functions (as well as a Weibull-Exponential mixture) to a
356 number of vertebrate and wind-dispersed trees and shrubs in a forest system and found different functions
357 performed better for different species, with no relation between dispersal mode and best-fit function. Clark
358 *et al.* (2005) fitted Exponential, Gaussian, Inverse Power (i.e. a Power Law function) and 2Dt functions to

359 dispersal data for different trees in a forest plot. They found Gaussian and 2Dt functions best fitted wind-
360 dispersed seeds, while the Inverse Power best fitted the bird- and monkey-dispersed seeds. Greene *et al.*
361 (2004), again considering trees from a variety of studies, found the Lognormal performed better than the
362 2Dt or Weibull.

363 For this study, we opted for an objective selection of dispersal functions, using 11 listed in a review
364 by Nathan *et al.* (2012). This approach gave insights into the performance of different functions when fitted
365 to a wide range of dispersal kernels. The Log-sech (LogS) and Exponential Power (Exp) showed the best fits
366 to the datasets. It is difficult to pinpoint why these two functions perform best, but this finding suggests
367 these might be useful general functions to use in models. However, these functions have quite different
368 properties and histories of use in dispersal studies. The LogS has been used very rarely, just in one study of
369 dispersal of birds (Van Houtan *et al.* 2010). The LogS becomes the Cauchy for $b=1$ (Nathan *et al.* 2012), which
370 is occasionally used in seed dispersal studies (Seri, Maruvka & Shnerb 2012; Munoz *et al.* 2013). The LogS
371 also has unattractive properties, such as no mean value for $b \geq 1$, nor does it have moments. The Exp has
372 been widely used on a variety of taxa (Nathan *et al.* 2012) and on many seed dispersal data since Clark *et*
373 *al.* (1998). It has useful properties such as a mean value (eqn 2), has moments and it reduces to the
374 Exponential or Gaussian for certain values of b . An interesting finding was that of the 144 datasets showing
375 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
376 estimated for the major growth form/dispersal mode groupings had $b < 1$, as did 17 of the 19 plant
377 height/seed mass subdivisions of these groupings. As an illustration of the influence of the b parameter, 57
378 of the 168 datasets had both the Exp and the Exponential in the best-fit group and had $b < 1$ for the Exp (10
379 had $b > 1$). The median value of the Exp b in these 57 datasets was 0.445 and the 95th percentile (as a measure
380 of the tail) of the fitted function was a median of 234% more than that of the fitted Exponential. Fat-tailed
381 kernels are a popular concept in dispersal ecology (Klein, Lavigne & Gouyon 2006; Nathan *et al.* 2008), but
382 this property of the Exp allows an explicit test whether empirical kernels are indeed fat-tailed. Because of
383 these properties, the Exp might be more generally useful than the LogS.

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

397 Our analysis suggests the Gaussian is a very poor representation of the dispersal process. This point
398 has long been raised in the dispersal literature (Wallace 1966), but our analysis provides an objective
399 affirmation. While its mathematical properties make the Gaussian popular in mathematical representations
400 of dispersal, the low kurtosis and thin tail can lead to inaccurate predictions about dispersal distances and
401 spatial dynamics (Clark *et al.* 1998; Klein, Lavigne & Gouyon 2006). The second one-parameter function, the
402 Exponential, has had a similarly negative press to the Gaussian (Bullock & Clarke 2000; Nathan *et al.* 2012),
403 but it performed much better than the Gaussian. While it did not perform as well as almost all the 2-
404 parameter functions (but better than the Lognormal - Table 1), our analyses suggest the Exponential has
405 some credence as a simple function to use both as a comparator for more complex functions when fitting
406 data, and as a straightforward function in mathematical modelling studies. It might be argued that the good
407 fit in many cases may reflect data quality or inadequate sampling at long distances, in that the tail of the
408 kernel may not be well described. This point is countered by our analysis showing that the distance over
409 which dispersal was measured in a study did not differ among the best-fit groups for the set of functions.

410 This does raise the issue that while the tail of the dispersal kernel is often discussed in the literature and
411 has a clear meaning in mathematical formulations of the kernel (Klein, Lavigne & Gouyon 2006), it is not
412 clear what exactly the tail is in empirical measures of dispersal, or what a sufficient measurement of this tail
413 looks like. Portnoy & Willson (1993) defined the tail as “the set of distance categories beyond the last clear
414 mode of the data”. By this definition, visual examination of each dataset suggested each represented a
415 kernel tail to some degree (e.g. Fig. 1).

416 SELECTING INFORMATIVE DISPERSAL KERNELS

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

429 Furthermore, the ExP and LogS functions gave somewhat different generalised kernels, as
430 illustrated by variation in the predicted medians and 95th percentiles. This partially reflects the intrinsic
431 differences in the functions – the LogS has a power-law tail and a weak effect of distance close to the source,
432 whereas the ExP represents a smoother decline with distance (Nathan *et al.* 2012). But the differences also
433 reflect uncertainty in our analyses due to variation among the datasets within each group in terms of the
434 ecology of each system and the data-gathering approach. Our parametrisation of both functions will allow
435 researchers to use them in combination and represent this uncertainty in the kernel.

436 Each individual case study, and thus each generalised kernel, relates to a single dispersal mode.
437 Multiple dispersal agents may be involved in the dispersal of seeds from a plant or of an individual seed.
438 Thus the ‘total dispersal kernel’ (Nathan *et al.* 2008) of a plant or population may require combining kernels
439 for multiple dispersal modes. To do so one should convolve (Neubert & Parker 2004) the relevant dispersal
440 distance probability density functions (pdf) – i.e. the dispersal location pdf/ $2\pi d$. Surprisingly, only three of
441 the 107 studies considered dispersal by multiple modes: ballistic dispersal followed by ant dispersal
442 (Beaumont, Mackay & Whalen 2009); and dispersal by vehicles vs that by wind (Arnold 1981; Bullock *et al.*
443 2003). The same approach could be used if, say, different animal species disperse the seed and one has
444 individual kernels for each animal vector (Lehouck *et al.* 2009).

445 USING REPRESENTATIVE DISPERSAL KERNELS

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

456 This potential raises the issue of the how dispersal kernels vary according to plant and vector
457 characteristics. It should be noted that analysis of the differences among classifications are not the main
458 focus of this paper, and the classifications are used more to give ecologists well-defined groups from which
459 to draw dispersal kernels relevant to their needs. We fitted generalised ExP and LogS functions for growth
460 form and dispersal syndrome combinations and were able to sub-divide these further according to plant
461 height (wind dispersal) or seed mass (animal dispersal) classes. These functions were fitted to groupings of

462 datasets each of which encompassed a wide range of systems, a diversity of data-gathering methods and
463 variety in the plants and dispersal vectors studied. For example, of the 30 papers (59 datasets) assessing
464 non-rodent vertebrate dispersal, the majority studied dispersal by small to medium sized birds (15 papers,
465 some papers studied more than one group), such as thrushes, tanagers or mockingbirds. Larger birds such
466 as toucans or cassowaries were covered in six papers, primates such as tamarins or spider monkeys in four
467 and medium-sized omnivorous mammals such as possums and martens in three. Bears, deer, fruit bats,
468 iguanas and a large freshwater fish were studied in one paper each, and two papers studied a broad,
469 unspecified group of vertebrate dispersers. Thus, there is inherent variation among the datasets within each
470 group. Despite this, the functions fitted to each group had generally high r^2 values, being >0.7 in the great
471 majority of cases and being significant at $p < 0.05$ in all. Thus we have confidence in the value of these
472 generalised functions for wider use. The collecting of further dispersal datasets would allow further
473 subdivision of datasets according to key variables such as vertebrate type, local wind speed, size of wind-
474 dispersal appendage, or vegetation type.

475 Because the datasets varied in terms of distance classes, maximum distance, seed densities and
476 sampling methods, fitting functions across all the data with a and b varying according to group would not
477 have been sensible or feasible. However, comparisons of the fitted curves and the dispersal quantiles
478 allowed tentative conclusions about differences among these groups. Our data suggested average dispersal
479 distances vary among growth forms in the order Trees > Shrubs > Herbs, Graminoids, and among dispersal
480 modes in the order Vertebrates (excluding rodents) > Rodents, Wind (with appendage) > Wind (without
481 appendage), Ant, Ballistic. Taller plants within groups disperse seeds further by wind, but seed mass has an
482 inconsistent relationship with dispersal distance (see also Thomson *et al.* (2011)). Variation in whether or
483 not lighter-seeded species disperse further than heavier-seeded species may arise if larger seeds are
484 dispersed by larger, more mobile animals (Nathan *et al.* 2008) and also where a varying number of seeds
485 are dispersed together in a fruit (Jordano 2000).

486 Our findings build on the analysis by Willson (1993), who carried out log-linear regression analysis
487 of collated dispersal kernels (i.e. fitting an Exponential function). Interestingly, she concluded wind-

488 dispersed herbs with seed appendages dispersed further than herbs using ballistic or ant dispersal, and
489 wind-dispersed trees and shrubs dispersed further than those dispersed by vertebrates. These findings
490 contradict ours and probably reflect the current availability of more data (60 papers compared to our 107)
491 as well as the more sophisticated analytical methods that are now available. However, this does
492 demonstrate the need to continue gathering dispersal data directly as it is likely conclusions will develop as
493 more case studies become available.

494 MEASURING SEED DISPERSAL

495 Any statistical synthesis of multiple datasets is bound by the number and quality of the studies used. The
496 168 datasets represent a great effort by the researchers involved and we hope this synthesis does them
497 justice by using these studies to provide general information for the wider use of dispersal kernels in
498 ecological research. However, the findings of this paper, especially the generalised dispersal kernels, are
499 likely to be much improved upon if further dispersal data are gathered. The fitted kernels provided here can
500 also be used to inform direct seed dispersal studies. Skarpaas, Shea and Bullock (2005) detailed a Monte
501 Carlo approach to designing seed dispersal studies efficiently using information on the likely dispersal
502 kernel, which could be provided by case studies and generalised kernels given in this paper.

503 While the relatively small number of studies is generally limiting, we can identify three particular
504 areas of research that require focussed activity. One is to use methods that allow better characterisation of
505 the tail, such as molecular markers or tracking animal dispersers (Bullock, Shea & Skarpaas 2006). The
506 variety of methods used to gather data is likely to lead to uncertainty. While we could not analyse this
507 formally due to the fact that methods used are strongly linked to the characteristics of the system studied,
508 analysis of biases due to methods would be useful. We did find that estimates of tree seed dispersal by
509 vertebrates were affected by the methods used. A number of studies combined measures of seed retention,
510 such as gut passage time, with data on movement by the animal vector. Dispersal distances estimated by
511 this method were in general further than dispersal measured by other methods, such as following
512 vertebrate vectors, seed trapping or using molecular markers. It is unclear however, whether this represents
513 a bias in the data as methods are usually chosen to match the researcher's understanding of the system,

514 such as knowledge that seeds are being carried a long way by vertebrate dispersers (see Cortes & Uriarte
515 2013). While biases according to method are possible, they do not undermine the aim of this paper, which
516 is to synthesise existing information of empirical dispersal kernels.

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

530

531 **Acknowledgements**

532 This research was supported by SCALES EU–FP7–226852, CEH project NEC05264, the Estonian Ministry of
533 Education and Research (IUT20-29) and the European Regional Development Fund (Centre of Excellence
534 EcolChange). We thank David Westcott and Patrick Jansen for providing more detailed dispersal data for
535 their pond apple and palm systems respectively, and to all the researchers who worked hard to generate
536 dispersal data across the World.

537

538 **Data accessibility**

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

541 **References**

- 542 Arnold, R.M. (1981) Population dynamics and seed dispersal of *Chaenorrhinum minus* on railroad cinder
543 ballast. *The American Midland Naturalist*, **106**, 80-91.
- 544 Baeten, L., Davies, T.J., Verheyen, K., Van Calster, H. & Vellend, M. (2015) Disentangling dispersal from
545 phylogeny in the colonization capacity of forest understorey plants. *Journal of Ecology*, **103**, 175-
546 183.
- 547 Beaumont, K.P., Mackay, D.A. & Whalen, M.A. (2009) Combining distances of ballistic and
548 myrmecochorous seed dispersal in *Adriana quadripartita* (Euphorbiaceae). *Acta Oecologica-
549 International Journal of Ecology*, **35**, 429-436.
- 550 Bocedi, G., Palmer, S.C.F., Pe'er, G., Heikkinen, R.K., Matsinos, Y.G., Watts, K. & Travis, J.M.J. (2014)
551 RangeShifter: a platform for modelling spatial eco-evolutionary dynamics and species' responses
552 to environmental changes. *Methods in Ecology and Evolution*, **5**, 388-396.
- 553 Bolker, B. & Pacala, S.W. (1997) Using moment equations to understand stochastically driven spatial
554 pattern formation in ecological systems. *Theoretical Population Biology*, **52**, 179-197.
- 555 Bullock, J.M. (2012) Plant dispersal and the velocity of climate change. *Dispersal Ecology and Evolution*
556 (eds J. Clobert, M. Baguette, T.G. Benton & J.M. Bullock), pp. 366-377. Oxford University Press,
557 Oxford.
- 558 Bullock, J.M., Mallada González, L, Tamme, R., Götzenberger, R., White, S.M., Pärtel, M. & Hooftman,
559 D.A.P. (2016) Data from: A synthesis of empirical plant dispersal kernels. *Dryad Digital Repository*
560 <http://dx.doi.org/10.5061/dryad.mq2ff>
- 561 Bullock, J.M. & Clarke, R.T. (2000) Long distance seed dispersal by wind: measuring and modelling the tail
562 of the curve. *Oecologia*, **124**, 506-521.
- 563 Bullock, J.M., Galsworthy, S.J., Manzano, P., Poschlod, P., Eichberg, C., Walker, K. & Wichmann, M.C.
564 (2011) Process-based functions for seed retention on animals: a test of improved descriptions of
565 dispersal using multiple data sets. *Oikos*, **120**, 1201-1208.

- 566 Bullock, J.M., Moy, I.L., Coulson, S.J. & Clarke, R.T. (2003) Habitat-specific dispersal: environmental effects
567 on the mechanisms and patterns of seed movement in a grassland herb *Rhinanthus minor*.
568 *Ecography*, **26**, 692-704.
- 569 Bullock, J.M., Shea, K. & Skarpaas, O. (2006) Measuring plant dispersal: an introduction to field methods
570 and experimental design. *Plant Ecology*, **186**, 217-234.
- 571 Bullock, J.M., White, S.M., Prudhomme, C., Tansey, C., Perea, R. & Hooftman, D.A.P. (2012) Modelling
572 spread of British wind-dispersed plants under future wind speeds in a changing climate. *Journal of*
573 *Ecology*, **100**, 104-115.
- 574 Caswell, H., Lensink, R. & Neubert, M.G. (2003) Demography and dispersal: Life table response
575 experiments for invasion speed. *Ecology*, **84**, 1968-1978.
- 576 Clark, C.J., Poulsen, J.R., Bolker, B.M., Connor, E.F. & Parker, V.T. (2005) Comparative seed shadows of
577 bird-, monkey-, and wind-dispersed trees. *Ecology*, **86**, 2684-2694.
- 578 Clark, J.S. (1998) Why trees migrate so fast: confronting theory with dispersal biology and the paleorecord.
579 *The American Naturalist*, **152**, 204-224.
- 580 Clark, J.S., Fastie, C., Hurtt, G., Jackson, S.T., Johnson, C., King, G.A., Lewis, M., Lynch, J., Pacala, S.,
581 Prentice, C., Schupp, E.W., Webb, T. & Wyckoff, P. (1998) Reid's paradox of rapid plant migration -
582 Dispersal theory and interpretation of paleoecological records. *Bioscience*, **48**, 13-24.
- 583 Clark, J.S., Silman, M., Kern, R., Macklin, E. & HilleRisLambers, J. (1999) Seed dispersal near and far:
584 Patterns across temperate and tropical forests. *Ecology*, **80**, 1475-1494.
- 585 Cortes, M.C. & Uriarte, M. (2013) Integrating frugivory and animal movement: a review of the evidence
586 and implications for scaling seed dispersal. *Biological Reviews*, **88**, 255-272.
- 587 Cousens, R.D., Dytham, C. & Law, R. (2008) *Dispersal in Plants: a Population Perspective*. Oxford University
588 Press, Oxford.
- 589 Driscoll, D.A., Banks, S.C., Barton, P.S., Ikin, K., Lentini, P., Lindenmayer, D.B., Smith, A.L., Berry, L.E., Burns,
590 E.L., Edworthy, A., Evans, M.J., Gibson, R., Heinsohn, R., Howland, B., Kay, G., Munro, N., Scheele,

591 B.C., Stirnemann, I., Stojanovic, D., Sweaney, N., Villaseñor, N.R. & Westgate, M.J. (2014) The
592 trajectory of dispersal research in conservation biology. Systematic review. *PLoS ONE*, **9**, e95053.

593 Engler, R. & Guisan, A. (2009) MigClim: Predicting plant distribution and dispersal in a changing climate.
594 *Diversity and Distributions*, **15**, 590-601.

595 FAO (2010) Global Forest Resources Assessment 2010. FAO Forestry Paper 163.

596 Gilbert, M.A., White, S.M., Bullock, J.M. & Gaffney, E.A. (2014) Spreading speeds for stage structured plant
597 populations in fragmented landscapes. *Journal of Theoretical Biology*, **349**, 135-149.

598 Greene, D.F., Canham, C.D., Coates, K.D. & Lepage, P.T. (2004) An evaluation of alternative dispersal
599 functions for trees. *Journal of Ecology*, **92**, 758-766.

600 Hallatschek, O. & Fisher, D.S. (2014) Acceleration of evolutionary spread by long-range dispersal.
601 *Proceedings of the National Academy of Sciences of the United States of America*, **111**, E4911-
602 E4919.

603 Harsch, M.A., Zhou, Y., HilleRisLambers, J. & Kot, M. (2014) Keeping Pace with Climate Change: Stage-
604 Structured Moving-Habitat Models. *The American Naturalist*, **184**, 25-37.

605 Hastings, A., Cuddington, K., Davies, K.F., Dugaw, C.J., Elmendorf, S., Freestone, A., Harrison, S., Holland,
606 M., Lambrinos, J., Malvadkar, U., Melbourne, B.A., Moore, K., Taylor, C. & Thomson, D. (2005) The
607 spatial spread of invasions: new developments in theory and evidence. *Ecology Letters*, **8**, 91-101.

608 Jones, F.A. & Muller-Landau, H.C. (2008) Measuring long-distance seed dispersal in complex natural
609 environments: an evaluation and integration of classical and genetic methods. *Journal of Ecology*,
610 **96**, 642-652.

611 Jordano, P. (2000) Fruits and frugivory. *Seeds- the ecology of regeneration in plant communities* (ed. M.
612 Fenner). CABI, Wallingford.

613 Katul, G.G., Porporato, A., Nathan, R., Siquiera, M., Soons, M.B., Poggi, D., Horn, H.S. & Levin, S.A. (2005)
614 Mechanistic analytical models for long-distance seed dispersal by wind. *The American Naturalist*,
615 **166**, 368-381.

616 Kjellsson, G. (1985) Seed fate in a population of *Carex pilulifera* L .1. Seed dispersal and ant-seed
617 mutualism. *Oecologia*, **67**, 416-423.

618 Klein, E.K., Lavigne, C. & Gouyon, P.-H. (2006) Mixing of propagules from discrete sources at long distance:
619 comparing a dispersal tail to an exponential. *BMC Ecology*, **6**, 1-12.

620 Kot, M., Lewis, M.A. & van den Driessche, P. (1996) Dispersal data and the spread of invading organisms.
621 *Ecology*, **77**, 2027-2042.

622 Lehouck, V., Spanhove, T., Demeter, S., Groot, N.E. & Lens, L. (2009) Complementary seed dispersal by
623 three avian frugivores in a fragmented Afromontane forest. *Journal of Vegetation Science*, **20**,
624 1110-1120.

625 Martinez, I. & Gonzalez-Taboada, F. (2009) Seed dispersal patterns in a temperate forest during a mast
626 event: performance of alternative dispersal kernels. *Oecologia*, **159**, 389-400.

627 Munoz, F., Beeravolu, C.R., Pelissier, R. & Couteron, P. (2013) Do spatially-implicit estimates of neutral
628 migration comply with seed dispersal data in tropical forests? *PLoS ONE*, **8**, e72497.

629 Nakagawa, S. & Schielzeth, H. (2013) A general and simple method for obtaining R² from generalized
630 linear mixed-effects models. *Methods in Ecology and Evolution*, **4**, 133-142.

631 Nathan, R., Katul, G., Bohrer, G., Kuparinen, A., Soons, M., Thompson, S., Trakhtenbrot, A. & Horn, H.
632 (2011) Mechanistic models of seed dispersal by wind. *Theoretical Ecology*, **4**, 113-132.

633 Nathan, R., Klein, E., Robledo-Arnuncio, J.J. & Revilla, E. (2012) Dispersal kernels: review. *Dispersal ecology
634 and evolution* (eds J. Clobert, M. Baguette, T.G. Benton & J.M. Bullock). Oxford University Press,
635 Oxford.

636 Nathan, R., Schurr, F.M., Spiegel, O., Steinitz, O., Trakhtenbrot, A. & Tsoar, A. (2008) Mechanisms of long-
637 distance seed dispersal. *Trends in Ecology & Evolution*, **23**, 638-647.

638 Neubert, M.G. & Parker, I.M. (2004) Projecting rates of spread for invasive species. *Risk Analysis*, **24**, 817-
639 831.

640 O'Dwyer, J.P. & Green, J.L. (2010) Field theory for biogeography: a spatially explicit model for predicting
641 patterns of biodiversity. *Ecology Letters*, **13**, 87-95.

- 642 Ovaskainen, O. & Cornell, S.J. (2006) Asymptotically exact analysis of stochastic metapopulation dynamics
643 with explicit spatial structure. *Theoretical Population Biology*, **69**, 13-33.
- 644 Petrovskii, S. & Morozov, A. (2009) Dispersal in a statistically structured population: fat tails revisited. *The*
645 *American Naturalist*, **173**, 278-289.
- 646 Portnoy, S. & Willson, M.F. (1993) Seed dispersal curves - behavior of the tail of the distribution.
647 *Evolutionary Ecology*, **7**, 25-44.
- 648 Royal Botanic Gardens Kew (2015) Seed Information Database (SID). Version 7.1. <http://data.kew.org/sid/>.
- 649 Salguero-Gomez, R., Jones, O.R., Jongejans, E., Blomberg, S.P., Hodgson, D.J., Mbeau-Ache, C., Zuidema,
650 P.A., de Kroon, H. & Buckley, Y.M. (2016) Fast-slow continuum and reproductive strategies
651 structure plant life-history variation worldwide. *Proceedings of the National Academy of Sciences*
652 *of the United States of America*, **113**, 230-235.
- 653 Seri, E., Maruvka, Y.E. & Shnerb, N.M. (2012) Neutral dynamics and cluster statistics in a tropical forest.
654 *The American Naturalist*, **180**, E161-E173.
- 655 Shaw, M.W., Harwood, T.D., Wilkinson, M.J. & Elliott, L. (2006) Assembling spatially explicit landscape
656 models of pollen and spore dispersal by wind for risk assessment. *Proceedings of the Royal Society*
657 *B-Biological Sciences*, **273**, 1705-1713.
- 658 Skarpaas, O. & Shea, K. (2007) Dispersal patterns, dispersal mechanisms, and invasion wave speeds for
659 invasive thistles. *The American Naturalist*, **170**, 421-430.
- 660 Skarpaas, O., Shea, K. & Bullock, J.M. (2005) Optimizing dispersal study design by Monte Carlo simulation.
661 *Journal of Applied Ecology*, **42**, 731-739.
- 662 Soons, M.B. & Bullock, J.M. (2008) Non-random seed abscission, long-distance wind dispersal and plant
663 migration rates. *Journal of Ecology*, **96**, 581-590.
- 664 Swaine, M.D. & Beer, T. (1977) Explosive seed dispersal in *Hura crepitans* L. (Euphorbiaceae). *New*
665 *Phytologist*, **78**, 695-708.
- 666 Tamme, R., Götzenberger, L., Zobel, M., Bullock, J.M., Hooftman, D.A.P., Kaasik, A. & Pärtel, M. (2014)
667 Predicting species' maximum dispersal distances from simple plant traits. *Ecology*, **95**, 505-513.

- 668 Tekiela, D.R. & Barney, J.N. (2013) Quantifying *Microstegium vimineum* seed movement by non-riparian
669 water dispersal using an ultraviolet-marking based recapture method. *PLoS ONE*, **8**, e63811.
- 670 Thompson, S.E., Assouline, S., Chen, L., Trahktenbrot, A., Svoray, T. & Katul, G.G. (2014) Secondary
671 dispersal driven by overland flow in drylands: Review and mechanistic model development.
672 *Movement Ecology*, **2**, 7.
- 673 Thomson, F.J., Moles, A.T., Auld, T.D. & Kingsford, R.T. (2011) Seed dispersal distance is more strongly
674 correlated with plant height than with seed mass. *Journal of Ecology*, **99**, 1299-1307.
- 675 Trakhtenbrot, A., Katul, G.G. & Nathan, R. (2014) Mechanistic modeling of seed dispersal by wind over
676 hilly terrain. *Ecological Modelling*, **274**, 29-40.
- 677 Travis, J.M.J., Delgado, M., Bocedi, G., Baguette, M., Bartoń, K., Bonte, D., Boulangeat, I., Hodgson, J.A.,
678 Kubisch, A., Penteriani, V., Saastamoinen, M., Stevens, V.M. & Bullock, J.M. (2013) Dispersal and
679 species' responses to climate change. *Oikos*, **122**, 1532-1540.
- 680 Tummers, B. (2006) Datathief III. <http://datathief.org/>.
- 681 Van Houtan, K.S., Bass Jr, O.L., Lockwood, J. & Pimm, S.L. (2010) Importance of estimating dispersal for
682 endangered bird management. *Conservation Letters*, **3**, 260-266.
- 683 Wada, N. & Uemura, S. (1994) Seed dispersal and predation by small rodents on the herbaceous
684 understory plant *Symplocarpus renifolius*. *The American Midland Naturalist*, **132**, 320-327.
- 685 Wallace, B. (1966) On the dispersal of *Drosophila*. *The American Naturalist*, **100**, 551-563.
- 686 Westcott, D.A., Bentrupperbaumer, J., Bradford, M.G. & McKeown, A. (2005) Incorporating patterns of
687 disperser behaviour into models of seed dispersal and its effects on estimated dispersal curves.
688 *Oecologia*, **146**, 57-67.
- 689 Willson, M.F. (1993) Dispersal mode, seed shadows, and colonization patterns. *Vegetatio*, **108**, 261-280.

Table 1 The probability density functions (dispersal location kernels, taken from Nathan *et al.* (2012)) fitted to the 168 seed dispersal datasets, along with summaries of the goodness of fit to these datasets. 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 datasets that have:				Median r^2
		$\Delta AICc \leq 4$ (best-fit group)	$\Delta AICc > 4$	not converged	$r^2 \geq 0.7$	
Log-sech (log-hyperbolic secant)	$\frac{1/(\pi^2 b a^2)}{(d/a)^{1/b} + (d/a)^{-1/b}}$	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
Lognormal	$\frac{1}{(2\pi)^{3/2} b d^2} \exp\left(-\frac{\ln(d/a)^2}{2b^2}\right)$	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

Table 2 The distribution of the 168 dispersal datasets 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 sub-divided where possible by seed mass (animal-dispersed) or plant height (wind-dispersed). These sub-divisions 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)	
						50 th	95 th					50 th	95 th
Herb	Ant	0.5281	1.2762	0.743	9	0.629	1.572	0.4580	0.3859	0.442	12	0.458	1.222
	0.7-8mg	2.5×10^{-6}	0.1888	0.929	3	0.561	6.838	0.2305	0.3528	0.871	5	0.231	0.565
	10-36mg	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 <10mg & <1m)	0.0917	0.6349	0.390	10	0.470	1.757	1.0634	0.8319	0.455	12	1.063	4.926
Herb	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
	1-3.5m	1×10^{-8}	0.1423	0.645	5	0.985	17.91	0.0197	1.4989	0.996	5	0.020	0.868
	0.1-0.8m	0.0030	0.3454	0.686	5	0.406	2.499	0.1286	0.6547	0.804	5	0.129	0.679
Herb	Wind no appendage (all <1m 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
Graminoid	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
	1.75-3m	31.985	1.0141	0.235	3	45.93	123.3	31.771	0.5193	0.839	3	31.77	119.0
	0.06-0.65m	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
	7-9mg	0.1915	0.6272	0.974	4	1.024	3.858	0.7483	0.5883	0.982	4	0.748	3.339
	29-40mg	0.2680	0.6568	0.861	3	1.229	4.485	1.5088	0.8458	0.868	2	1.509	12.95
Shrub	Vertebrate	1×10^{-8}	0.1339	0.996	18	4.974	99.39	28.659	0.8208	0.831	18	28.66	130.1
	0.2-3mg	1×10^{-8}	0.1161	0.879	11	375.7	9471	29.216	0.8369	0.870	11	29.22	245.2
	31-69mg	1×10^{-8}	0.1264	0.973	2	26.01	569.7	25.220	0.8415	0.980	2	25.22	214.2
	5000-10500mg	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
	195-950mg	2.0615	0.6538	0.809	5	9.594	35.12	7.9526	0.7733	0.744	5	7.953	56.79
	2420-18800mg	0.3212	0.5035	0.811	6	4.190	18.58	8.9256	0.7269	0.803	6	8.926	56.64
Tree	Vertebrate	1×10^{-8}	0.1246	0.690	39	40.01	897.1	49.997	1.3989	0.807	39	50.00	1751
	0.3-15.1mg	1×10^{-8}	0.1288	0.781	10	14.96	317.9	15.511	1.2950	0.740	10	15.51	417.2
	31-180mg	1×10^{-8}	0.1237	0.867	11	49.89	1132	10.014	1.7790	0.264	9	10.01	921.8
	200-800mg	1×10^{-8}	0.1238	0.804	9	48.68	1103	97.668	0.7489	0.624	13	97.67	655.7
	1000-113700mg	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
	30-46m	1.3437	0.4654	0.364	7	25.98	122.7	29.601	0.9953	0.7031	6	29.60	371.6
	5-15m	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 datasets 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 and Uemura (1994), b) Tekieli and Barney (2013), c) Swaine and Beer (1977), d) Kjellsson (1985), e) Bullock and Clarke (2000), f) (Arnold (1981)). Where plotted, the y axis is not continuous between 0 and the next tick.

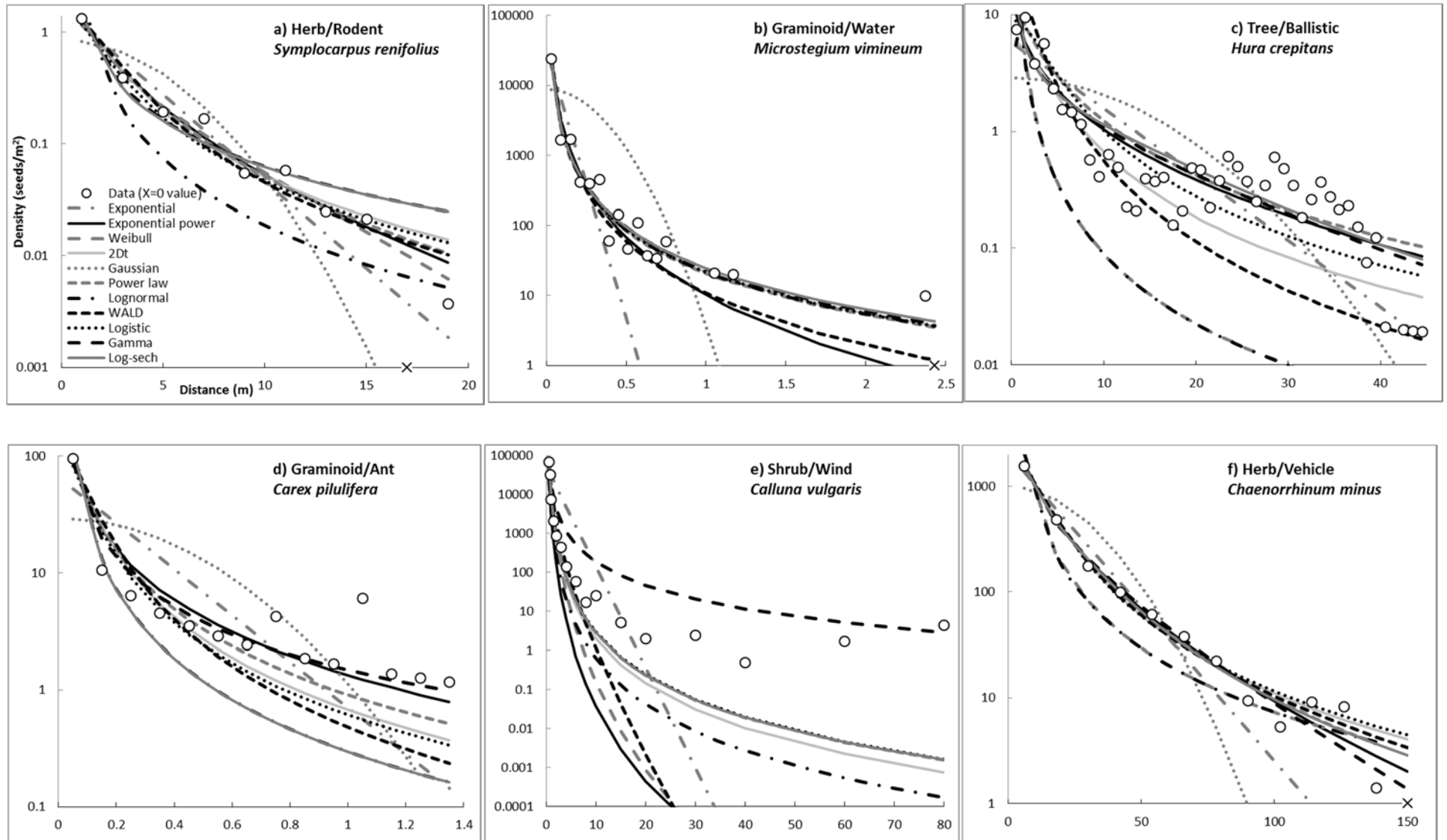


Fig 2. Generalised dispersal kernels for all 10 growth form/dispersal mode combinations for which there were sufficient datasets. Where possible, the datasets were also split into tall vs short plants (wind-dispersed) or light vs heavy seeds (animal-dispersed). The plots show the data and the fitted 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 dataset 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.

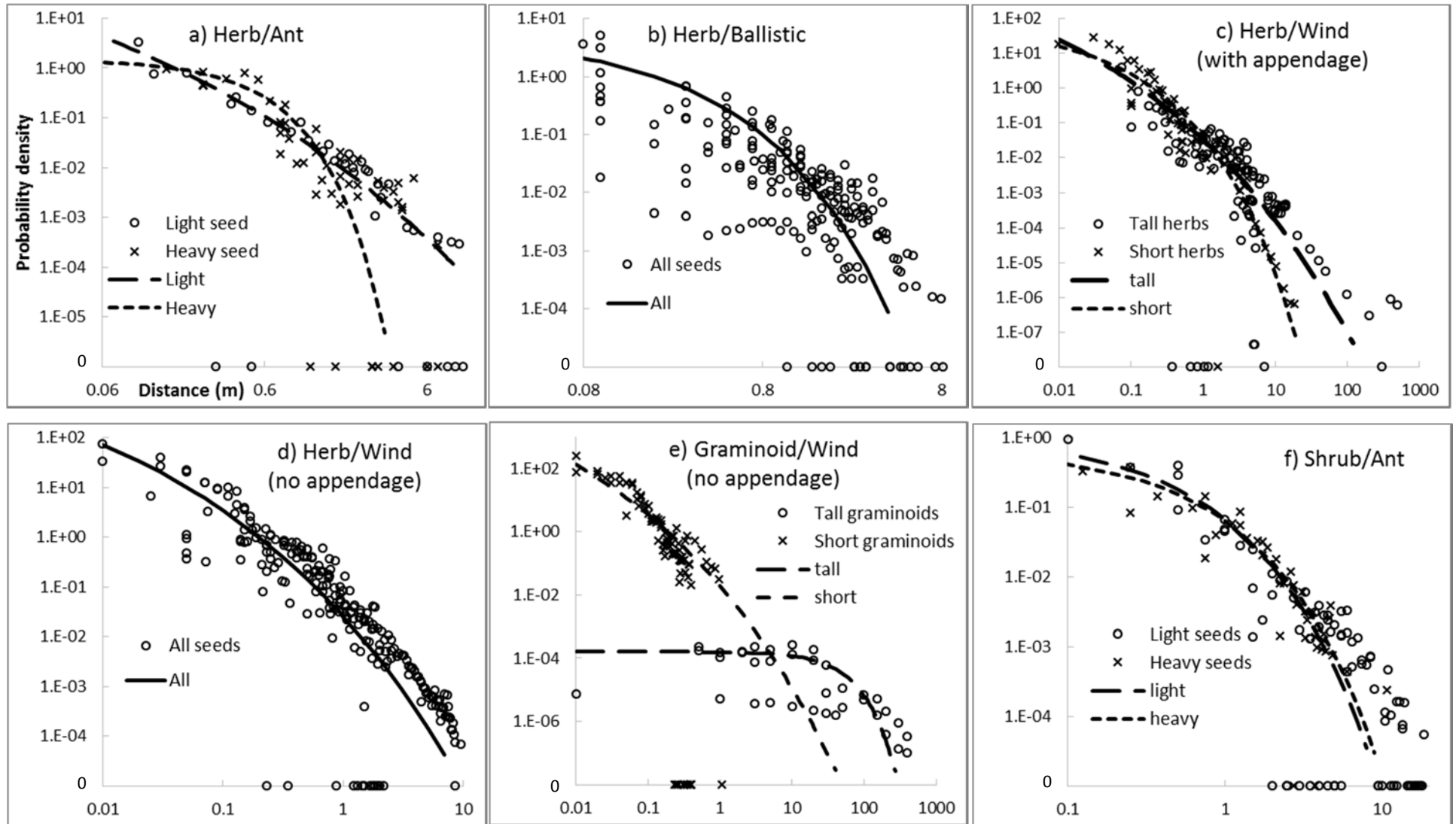


Fig. 2. continued

