Predicting species' maximum dispersal distances from simple plant traits

RIIN TAMME,[1,5] LARS GÖTZENBERGER,[1,2] MARTIN ZOBEL,[1] JAMES M. BULLOCK,[3] DANNY A. P. HOOFTMAN,[3] ANTS KAASIK,[4] AND MEELIS PÄRTEL[1]

[1]*Institute of Ecology and Earth Sciences, University of Tartu, Lai 40, Tartu 51005 Estonia*
[2]*Institute of Botany, Czech Academy of Science, Dukelská 135, CZ-379 82 Třeboň, Czech Republic*
[3]*NERC Centre for Ecology and Hydrology, Maclean Building, Benson Lane, Wallingford OX10 8BB United Kingdom*
[4]*Institute of Ecology and Earth Sciences, University of Tartu, Vanemuise 46, Tartu 51014 Estonia*

Abstract. Many studies have shown plant species' dispersal distances to be strongly related to life-history traits, but how well different traits can predict dispersal distances is not yet known. We used cross-validation techniques and a global data set (576 plant species) to measure the predictive power of simple plant traits to estimate species' maximum dispersal distances. Including dispersal syndrome (wind, animal, ant, ballistic, and no special syndrome), growth form (tree, shrub, herb), seed mass, seed release height, and terminal velocity in different combinations as explanatory variables we constructed models to explain variation in measured maximum dispersal distances and evaluated their power to predict maximum dispersal distances. Predictions are more accurate, but also limited to a particular set of species, if data on more specific traits, such as terminal velocity, are available. The best model ($R^2 = 0.60$) included dispersal syndrome, growth form, and terminal velocity as fixed effects. Reasonable predictions of maximum dispersal distance ($R^2 = 0.53$) are also possible when using only the simplest and most commonly measured traits; dispersal syndrome and growth form together with species taxonomy data. We provide a function (dispeRsal) to be run in the software package R. This enables researchers to estimate maximum dispersal distances with confidence intervals for plant species using measured traits as predictors. Easily obtainable trait data, such as dispersal syndrome (inferred from seed morphology) and growth form, enable predictions to be made for a large number of species.

Key words: dispersal syndrome; growth form; migration; mixed-effects model; predictive model; seed mass; seed release height; taxonomy; terminal velocity; traits.

INTRODUCTION

Dispersal is the unidirectional movement of individuals away from their place of birth (Levin et al. 2003). For plants, this spatial movement occurs through vegetative growth or seed dispersal (here we use "seed" as a general term for a reproductive dispersing unit). While vegetative spread plays a great role in small-scale patterns of species distribution (Moora et al. 2009), seed dispersal is a process that shapes local populations and communities as well as large-scale distribution of species (Levin et al. 2003, Clobert et al. 2012).

In order to understand the role of dispersal at various spatial scales, it is of primary importance to know how far seeds disperse. However, studying seed dispersal and measuring dispersal distances for plants is challenging. Methods such as seed traps, direct observations, or genetic markers are used in the field (Bullock et al. 2006). Some studies employ mechanistic models to estimate dispersal distances using information on seed and plant traits as well as environmental conditions

(Soons and Ozinga 2005, Nathan et al. 2011, Bullock et al. 2012). Spatial patterns of seed dispersal are often depicted by a dispersal curve, i.e., a dispersal distance distribution, which displays the frequency (proportion or probability) of seeds reaching a given distance (Nathan et al. 2012). The shape of the dispersal curve can vary for different species or dispersal syndromes, but a feature of many dispersal curves is a flat tail, i.e., relatively rare long-distance dispersal events (Levin et al. 2003). However, such long-distance dispersal events have disproportionate importance considering their rarity, for example, facilitating rapid spread of species and connecting fragmented populations (Nathan et al. 2008).

Many studies have shown dispersal distances to be strongly related to life-history traits. For animals, several recent papers highlight the strong correlation between species traits and dispersal ability (Garrard et al. 2012, Hein et al. 2012, Stevens et al. 2012). For plants, it is known that dispersal distance is related to a dispersal syndrome (Willson 1993, Pärtel and Zobel 2007, Vittoz and Engler 2007). Dispersal syndrome may generally be deduced from seed morphology, which is assumed to facilitate dispersal by a specific vector (e.g.,

[5] E-mail: riin.tamme@ut.ee

animals, wind, water [Hughes et al. 1994, Thomson et al. 2010]). However, a multiplicity of vectors, as well as variation in the behavior of these vectors, makes it difficult to establish a direct association between dispersal syndrome and long-distance dispersal (Nogales et al. 2007). Even within a single dispersal syndrome, dispersal distances vary in magnitude (e.g., Augspurger 1986, Vittoz and Engler 2007, Muller-Landau et al. 2008, Thomson et al. 2011). Additional information about seed and plant characteristics may help in explaining seed dispersal distances. Models of seed dispersal often incorporate information about seed mass and shape (e.g., Augspurger 1986, Matlack 1987, Ernst et al. 1992). A recent global review of seed dispersal distances of over 300 species (Thomson et al. 2011) revealed that plant height is more important than seed mass in determining seed dispersal distances. Similarly, for wind-dispersed species, seed release height is often a good predictor of dispersal distance (Soons and Ozinga 2005).

Although many studies have examined the relationship between plant traits and dispersal distances, few have used a macroecological approach, including species from different regions and with varying dispersal syndromes. There are currently three attempts aiming to generalize relationships between plant traits and dispersal distance across multiple species. Willson (1993) addressed the effect of dispersal syndrome on shapes of seed shadows within growth form classes. Vittoz and Engler (2007) did the same on the basis of a larger sample, but included mainly species from Central Europe, and created a more complex system of categories where dispersal syndromes were integrated with other plant traits. In the most recent study, Thomson et al. (2011) included the largest data set to date; they addressed the effect of plant height and seed mass on dispersal distances across all species and within particular dispersal syndromes. These three studies have provided extremely valuable information about the importance of dispersal syndrome as well as growth form in determining dispersal distances across species. However, no studies so far have explicitly evaluated the predictive power of these traits in estimating dispersal distances.

Here we apply a cross-validation statistical approach to determine how well various plant traits can predict dispersal distances by using 576 species from different regions. We use traits that are either widely available in databases or can be easily obtained from other sources, such as local floras. We focus on species' maximum dispersal distances to emphasize the importance of long-distance dispersal and the potential to reach a given site, but admit that maximum dispersal distance is significantly and strongly correlated with mean dispersal distance (Thomson et al. 2011). In addition to providing information on the relationship between plant traits and dispersal distances across species, we evaluate the possibility to use simple traits to predict dispersal distances for species that lack distance data. We also provide a freeware tool (dispeRsal) that can be used to predict dispersal distances from trait data for a large number of plant species.

METHODS

Data on seed dispersal distances and plant traits

To compile our data set, we used six previous studies that summarized published dispersal information and addressed a large number of species (Willson 1993, Hughes et al. 1994, Cain et al. 1998, Bullock and Clarke 2000, Ness et al. 2004, Vittoz and Engler 2007), as well as studies in which the dispersal distances were derived from models parameterized with field data (e.g., Sheldon and Burrows 1973, Augspurger 1986, Matlack 1987, Ernst et al. 1992, Stöcklin and Bäumler 1996, Kiviniemi and Telenius 1998, Jongejans and Telenius 2001, Soons and Ozinga 2005). We additionally searched the ISI Web of Science for papers published up until 31 January 2012, using the keywords "seed," "dispersal," and "distance." We included case studies that provided data about dispersal distances corresponding to individual species and specific dispersal syndromes.

From each study, we extracted dispersal syndrome and dispersal distance data. We distinguished five main dispersal syndromes: (1) wind dispersal with special mechanisms (e.g., plumes, wings), (2) animal dispersal (endo-, epi-, and synzoochory by vertebrates, including human), (3) ant dispersal, (4) ballistic dispersal, and (5) wind dispersal without special mechanisms (i.e., seeds do not have any special structures to enhance dispersal but are usually dispersed by abiotic factors). We did not further separate animal dispersal (e.g., endo-, epi-, or synzoochory) because of too few data points within each category. Since water dispersal distance data were available for only four species, we excluded this syndrome from the analyses. Dispersal distance data in the original studies were given as the longest observed maximum, 99th percentile, 90th percentile, mean, mode, and median distances. Additionally, we collected information on plant growth form from the case studies, or from databases LEDA (Kleyer et al. 2008) and PLANTS (*available online*).[6] Furthermore, we collected additional data on seed release height (LEDA; Kleyer et al. 2008), seed mass (Seed Information Database; *available online*)[7] and terminal velocity (LEDA; Kleyer et al. 2008).

We were able to compile data from 121 case studies for 576 species from 102 families (see Supplement 1). For each species, we use only the highest value reported in the literature as the maximum dispersal distance in the analyses. Roughly half of the data came from observational field studies, whereas the other half derived from modeling studies that were parameterized with field

---

[6] http://plants.usda.gov/
[7] http://data.kew.org/sid/

data. Our data set is global, but data from temperate regions in Europe and North America are more commonly represented (468 species in temperate regions compared to 108 in the tropics). The complete species set could not be used for all analyses since data on seed mass, release height, and terminal velocity were available for only 491, 292, and 247 species, respectively. For most of the species (519), we used the furthest dispersal distance given, which was the measured maximum distance or 99th percentile (for one species we used the 90th percentile). Similar to Thomson et al. (2011), we found in a linear regression that mean dispersal distance was a good predictor of the maximum distance (R^2 = 0.85) and therefore used the relationship ($\log_{10}$(maximum) = 0.795 + 0.984$\log_{10}$(mean)) to estimate the maximum for species for which we only had mean values (57 species). Maximum dispersal distances, seed release height, seed mass, and terminal velocity were $\log_{10}$-transformed for analyses.

Although trait data can explain plant ecological responses, taxonomy or phylogeny can add information about functional variation among species (Reinhart et al. 2012). Therefore, we also use taxonomy in our predictions. The taxonomic classification was obtained in two steps. First we used the Taxonstand (Cayuela et al. 2012) library in R (R Development Core Team 2012) to validate species names and to assign genera to families. Subsequent assignment of families to orders followed APG III for angiosperms (Angiosperm Phylogeny Group 2009) and Christenhusz et al. (2011) for gymnosperms.

Data analysis

To find the most important traits affecting maximum dispersal distances, as well as to find candidate models for predictions, we first analyzed the relationship between maximum dispersal distance (response variable) and different combinations of plant traits. Using a linear regression model, we first included all traits (dispersal syndrome, growth form, seed mass, seed release height, terminal velocity) as fixed effects. Since data for all traits was available for only 155 species, we additionally tested models that incorporated more species, but fewer traits. Specifically, we analyzed the relationship between maximum dispersal distance and (1) dispersal syndrome, growth form, and terminal velocity (247 species); (2) dispersal syndrome, growth form, seed mass, and seed release height (264 species); (3) dispersal syndrome, growth form, and seed release height (290 species); (4) dispersal syndrome, growth form, and seed mass (488 species); (5) dispersal syndrome and growth form (576 species). We also tested for interactions between fixed effects in the models. The statistical significance of fixed effects in the linear models was evaluated using Type II ANOVA. The linear model assumptions were evaluated visually and all the continuous data (maximum dispersal distance, seed mass, seed release height, terminal velocity) were $\log_{10}$-transformed for normality.

Building the predictive models

Following the data analysis, we chose five groups of candidate models for testing their predictive power in estimating dispersal distances. The group 1 of candidate models included dispersal syndrome, growth form, and terminal velocity as explanatory factors. Group 2 included dispersal syndrome, growth form, seed mass, and seed release height as explanatory variables. Group 3 included dispersal syndrome, growth form, and seed release height. Group 4 included dispersal syndrome, growth form, and seed mass. Group 5 included only dispersal syndrome and growth form. We considered different groups to represent different sets of plant traits and to assess whether accurate predictions are also possible without information on terminal velocity, for which data are much scarcer than for release height, seed mass, or growth form. Models with terminal velocity only apply to species with wind or ballistic dispersal syndrome. We did not include any models with interactions for our predictions due to too few (or no) data points for some of the trait combinations when data were divided into two groups in testing the predictive power of the models. Furthermore, we set up models that accounted for taxonomic structure (order, family) in the data by using the nested taxonomy of the species as a random variable in linear mixed-effect models (Pinheiro and Bates 2000). Altogether we fitted 15 different models (three for each group) that were compared by Akaike information criterion (AIC) and Bayesian information criterion (BIC) values. An additional 23 models with random slope effect (Schielzeth and Forstmeier 2009) and genus-level taxonomy were also fitted.

To test the ability of these models to predict seed maximum dispersal distances accurately, we randomly split our data set into two parts. Two-thirds of the data was used to fit the predictive models and we used these parameter values to predict maximum dispersal distances for the remaining species from their traits. Next, a major axis (Type II) regression between the observed and predicted maximum dispersal distances ($\log_{10}$-transformed values) was conducted, calculating R^2, intercept, and slope of this regression. For each model, the results are presented as averages of 999 runs, each with a new split of the data into model and test groups. Because of varying number of species and different fixed effects, AIC and BIC comparison of models are not valid between groups. Therefore, we use R^2 from a major axis regression between the observed and predicted distances for a more accurate comparison of all models.

All analyses were conducted in the software R (R Development Core Team 2012). We used the lm function in the stats package (R Development Core Team 2012) for fitting linear regression models and the lme function in the nlme package (Pinheiro et al. 2011) for fitting mixed-effect models. For the Type II ANOVA test, we used the ANOVA function in the car package

(Fox and Weisberg 2011). In testing the accuracy of predictions, the major axis regression estimates were calculated in the package smatr (Warton et al. 2011).

RESULTS

Depending on the model and the set of species used, different factors were important in explaining maximum dispersal distances (see Appendix A: Tables A1–A6 for the descriptive model results). In the full model (model $R^2 = 0.51$, $P < 0.001$, Table A1), only dispersal syndrome and terminal velocity were significant. The best model (highest model R^2) included dispersal syndrome and terminal velocity (model $R^2 = 0.61$, $P < 0.001$, Table A2). However, both of these models included mostly herbs and wind-dispersed species. In the simplest model ($R^2 = 0.51$, $P < 0.001$, Table A6), including only dispersal syndrome and growth form as explanatory variables and all 576 species, both dispersal syndrome and growth form were statistically significant. On average, maximum dispersal distances increase from wind (no special mechanisms) < ballistic < ant < wind (special mechanisms) < animal (Appendix A: Fig. A1). Considering growth form across all species, maximum dispersal distances increase from herb < shrub < tree (Fig. A2).

The R^2 for the regression between observed and predicted values ranged between 0.45 and 0.60, with the best predictive model having dispersal syndrome, growth form, and terminal velocity as fixed effects (Fig. 1A). Naturally, this model cannot be used to predict dispersal distances for animal or ant dispersal. However, even a simple linear model incorporating only categorical variables (growth form and dispersal syndrome) predicted dispersal distances with a reasonable accuracy (predictions $R^2 = 0.50$; Fig. 1B). For predictions, the models that incorporated taxonomic structure as a random variable generally performed as well or better than the simple linear models (Table 1) but not in all cases (i.e., R^2 lower or equal, see Appendix B: Table B1 for the full set of models). Regarding the relationship between observed and predicted values, we found that intercepts were higher than 0 and slopes of the regression lines were <1 for all tested models. The average intercept estimates ranged between 0.30 and 0.44 and the slope estimates between 0.59 and 0.74. Hence, predicted values were, in general, overestimates for species with short-distance dispersal and underestimates for species with long-distance dispersal.

To allow researchers to calculate maximum dispersal distances with confidence intervals (CI) for their species of interest, we provide the code for a function dispeRsal to be run in R (see Supplement 2 for the necessary .rda file as well as instructions). The function will work with any species set for which at least dispersal syndrome is available. There are no constraints on the taxonomical affinity of the species, but prediction for species whose order and family match the ones that are used in the predictive models will be on average more accurate

(Table 1). This is because only for those species the random effect of taxonomy can be considered when predicting dispersal distances. However, for species either lacking the taxonomic information or having no matching taxonomical information, dispersal distances are still predicted from the fixed effects (i.e., traits).

DISCUSSION

Information about dispersal distances is relevant for understanding a multitude of ecological processes and for addressing several conservation issues (Trakhtenbrot et al. 2005, McConkey et al. 2012). Using a large data set, we have shown that dispersal syndrome, growth form, and terminal velocity explain over 60% of the variation in maximum dispersal distances. Similarly, the power to predict maximum dispersal distances for a model including dispersal syndrome, growth form, and terminal velocity is 61%. Unfortunately, terminal velocity data are not available for many plant species and also these predictions are applicable only to species with wind or ballistic dispersal syndrome. Therefore, as an alternative, we fitted and tested the predictive power of more general models that include more species. Even so, these simple and easily obtainable traits explained maximum dispersal distances relatively well. For example, dispersal syndrome together with plant growth form described more than one-half of the variation in maximum dispersal distances and the predictive power of this model increased to 53% when taxonomy data were included. Our cross-validation approach demonstrated strong predictive power for all of our models (predictions $R^2 > 0.45$). Our main conclusion from these results is that it is indeed possible to predict dispersal distances for a large number of species using easily obtainable data for a few traits.

The importance of dispersal syndrome in shaping dispersal distance patterns is well known in ecology and therefore our models always included this trait. Species that have special adaptations to enhance dispersal travel longer distances than species without such mechanisms (Willson 1993, Vittoz and Engler 2007, Thomson et al. 2011). Generally, species specialized to ballistic or ant dispersal have low maximum dispersal distances (on average below 10 m) while wind dispersal with special structures and animal dispersal result in much longer distances (see Appendix A: Fig. A1; also Vittoz and Engler 2007, Thomson et al. 2011). Variation in dispersal syndromes can have significant consequences on the distribution and assembly of plant species. For example, Seidler and Plotkin (2006) found dispersal syndrome to determine the spatial aggregation of species in a tropical forest. Similarly, Drezner et al. (2001) showed the importance of dispersal syndrome to determine the species distribution patterns in a riparian zone. Additionally, dispersal syndrome can determine the vulnerability of species to global change (Montoya et al. 2008).
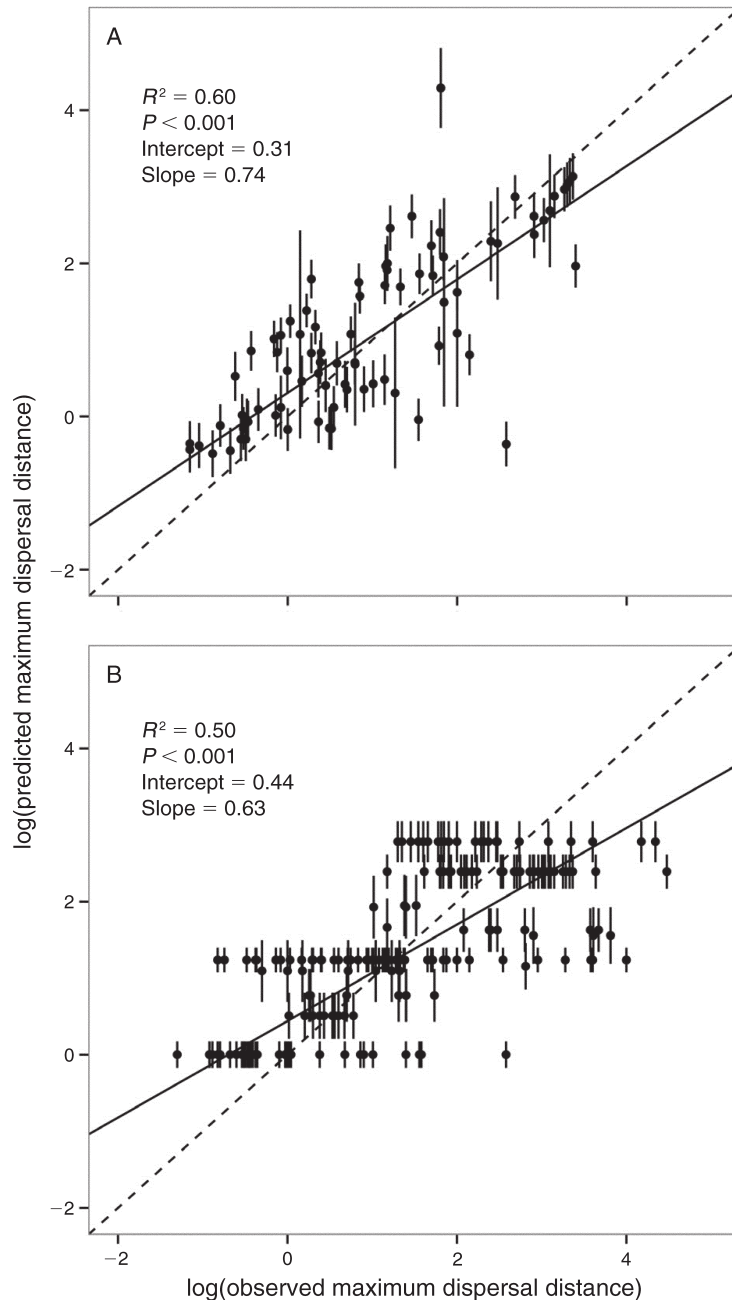
FIG. 1. Example plots of regressions of predicted dispersal distance on observed dispersal distance (measured in meters; values were log-transformed) for two different predictive models. The dashed line depicts the slope of unity with intercept zero, and the solid line depicts the major axis regression slope. Points represent different species, and bars around points depict confidence intervals of predicted values. Values of R^2, P, intercept, and slope are for the major axis regression. (A) Simple linear regression model with dispersal syndrome, growth form, and terminal velocity as explanatory variables. (B) Simple linear regression model with dispersal syndrome and growth form as explanatory variables.

A recent study by Thomson et al. (2011) showed the importance of plant height in determining dispersal distances as opposed to seed mass that is generally deemed more important (Greene and Johnson 1993, Xiao et al. 2005, Muller-Landau et al. 2008). According to our cross-validation analyses, models with seed mass had higher predictive power (higher R^2 values) than models where seed mass was absent (Table 1, compare groups 2 vs. 3, 3 vs. 4, and 4 vs. 5). However, all of our models also included plant growth form (tree, shrub, or herb), which acts as a proxy for plant height and influences dispersal distance patterns. In addition, plant

TABLE 1. Overview of the models used to predict maximum seed dispersal distances.

Group	n	Fixed effects	Random effects	AIC	BIC	R^2	df	Intercept	Slope
1	247	DS, GF, TV		650.2	681.5	0.60	80	0.31	0.74
1	247	DS, GF, TV	order	649.9	684.6	0.60	80	0.30	0.74
1	247	DS, GF, TV	order/family	649.7	688.0	0.59	80	0.34	0.72
2	264	DS, GF, RH, SM		749.6	785.0	0.49	86	0.34	0.64
2	264	DS, GF, RH, SM	order	739.4	778.4	0.52	86	0.31	0.65
2	264	DS, GF, RH, SM	order/family	736.4	778.8	0.51	86	0.31	0.66
3	290	DS, GF, RH		833.6	866.4	0.45	95	0.37	0.59
3	290	DS, GF, RH	order	821.6	858.0	0.49	95	0.35	0.61
3	290	DS, GF, RH	order/family	817.3	857.3	0.48	95	0.34	0.61
4	488	DS, GF, SM		1307.1	1344.7	0.53	161	0.41	0.66
4	488	DS, GF, SM	order	1301.0	1342.8	0.54	161	0.40	0.67
4	488	DS, GF, SM	order/family	1297.5	1343.4	0.54	161	0.40	0.66
5	576	DS, GF		1549.5	1584.3	0.50	190	0.44	0.63
5	576	DS, GF	order	1543.4	1582.5	0.51	190	0.44	0.63
5	576	DS, GF	order/family	1530.8	1574.2	0.52	190	0.43	0.64

Notes: The models were assigned to one of five groups according to their fixed effects. From each group, a simple linear model and the linear mixed-effects models, with order or family nested in order as a random effect, are represented here. For each group, n denotes the number of species used in the analyses and predictions. For all the models, $P < 0.001$. Values of R^2, degrees of freedom (df), intercept, and slope are for the major axis regression (MA) between predicted and observed dispersal distances (average of 999 runs). Note that because of different fixed effects, Akaike information criterion (AIC) and Bayesian information criterion (BIC) comparison of models are only valid within groups. Abbreviations are: DS, dispersal syndrome; GF, growth form; SM, seed mass; RH, seed release height; and TV, terminal velocity.

height and growth form already account for some variation in seed size (e.g., taller plants have generally larger seeds [Moles et al. 2004]). Our results suggest that seed size variation within growth forms is still an important predictor of seed dispersals distances. Additionally, seed mass may have a stronger impact on dispersal by affecting establishment success (Moles and Westoby 2004) rather than influencing dispersal distance.

An important insight here is that, depending on the set of species used and the trait data available, different traits are important in describing dispersal distance patterns. For example, when mostly wind-dispersed herbs are included (Appendix A: Table A1), terminal velocity is more important than seed release height or seed mass. This is probably because the species under comparison have relatively similar heights and seeds.

We envision several research themes that could benefit from data on plant dispersal distances. First, dispersal distance estimates allow the testing of theories that relate metapopulation or metacommunity dynamics to dispersal (e.g., Poschlod et al. 1998, Johst et al. 2002, Honnay et al. 2005, Soons et al. 2005, Higgins 2009). This knowledge can also be used in designing nature conservation areas based on increasing landscape connectivity (Couvreur et al. 2004, Hooftman and Bullock 2012). Second, knowledge about dispersal distances enables assessment of the potential role of dispersal limitation in shaping local communities and driving extinctions, as well as in recolonizations (Zobel 1997). Dispersal distance data could also help to understand why communities might be incomplete, i.e., having a relatively large "dark diversity" (the absent portion of the habitat-specific species pool [Pärtel et al. 2011]). Third, at a biogeographic scale, incorporating dispersal distance information into species distribution

models could improve predictions of projected species ranges under different scenarios of global change (Lavergne et al. 2006, Parolo and Rossi 2008, Vittoz et al. 2009, Buckley et al. 2010, Guisan and Rahbek 2011, Bullock et al. 2012). Finally, information on dispersal distances may help to predict rates of spread of invasive species (Bradley et al. 2010).

Although the predictive power for estimating dispersal distances is high, dispersal is a complex process and caveats apply to the estimations. The accuracy of the predictions is to some extent influenced by the methodology that has been used to obtain the observed data. Methods that involve searching within a restricted area (e.g., seed traps) may underestimate long-distance dispersal (Bullock et al. 2006). Certain methods, such as tracking animal vectors or genetic markers may provide better estimates of long-distance dispersal (Nathan and Muller-Landau 2000). Also, it is important to note that our predictive models are most valuable for species (or other taxonomic organization level) comparisons and do not include intra-specific variability in dispersal distances (Bullock et al. 2003, Russo et al. 2006, Norghauer et al. 2011). If a seed is dispersed by non-standard mechanisms (i.e., other than suggested by its dispersal syndrome) or undergoes multiple dispersal events, this may lead to long-distance events that are not well predicted by the putative dispersal syndrome (Bullock et al. 2006). Due to limited dispersal distance data available in the literature, we currently have less data for species from tropical regions and with more specific dispersal syndromes (e.g., endozoochory), which leads to less precise estimations for these species. In general, our approach will become more accurate as more data on plant traits become available.

When examining the relationship between predicted and observed dispersal values in cross-validation, we

found that predicted maximum dispersal distances are overestimated for species with short-distance dispersal and underestimated for species with long-distance dispersal (see Fig. 1). A detailed analysis of the data revealed that over- and underestimations in predicted distances resulted from two conditions, depending on the explanatory variables in the model. First, the response variable (i.e., seed dispersal distance) showed distributions with different skew and kurtosis among dispersal syndromes and growth forms, which leads to inexact estimation, as observed in our data. Solving this issue would require transformation within groups (or combinations thereof) prior to calculating the predictive models so that the data are of equal shape in distribution within the groups. However, we consider the described models to be the most reasonable compromise to maintain the general approach of the predictive models without having to integrate complex specializations. Second, for models using continuous explanatory variables, ordinary least squares regression between two variables that both have a measurement error underestimates the regression slope. Type II regression could be applied, but its use for prediction is discouraged (Legendre and Legendre 1998, Smith 2009). Also, there are no statistical routines that implement Type II regression in a generalized way to make it suitable for our predictive models. Thus, there is a possibility for further improvements of our predictive models as the statistical methods develop.

With the freeware tool dispeRsal provided here (Supplement 2), researchers are able to estimate dispersal distances for their own set of plant species. Furthermore, the tool also offers the possibility to calculate confidence intervals for the estimated dispersal distances, which is a useful option to account for the uncertainty of the current predictive models. Using confidence intervals will allow researchers to use a range of dispersal distances instead of just one predicted value. With dispeRsal it is also possible to predict a species' maximum dispersal distance for each of a variety of dispersal syndromes, to account for the fact that a species can have several dispersal syndromes (Hintze et al. 2013) and that species can disperse by other mechanisms than predicted from their seed morphology (Bullock et al. 2006).

In summary, easily obtainable trait data, such as dispersal syndrome and growth form, as well as species taxonomy, determine a great part of the variation in dispersal distances. This knowledge allows predicting seed dispersal distances from simple plant traits. In particular, the predictive powers of models that include only simple categorical traits such as dispersal syndrome and growth form are promising. Together with the growing body of available data for plant dispersal traits (e.g., D^3 database; Hintze et al. 2013), the ability to estimate dispersal distances from simple traits will allow predictions to be made for a large number of species. The availability of estimates of dispersal distances of different plant species should enable researchers to study the consequences of changing landscapes and climate more accurately, especially when considering processes at relatively large scales, such as migration and the interconnectivity of populations and communities.

LITERATURE CITED

Angiosperm Phylogeny Group. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. Botanical Journal of the Linnean Society 161:105–121.

Augspurger, C. K. 1986. Morphology and dispersal potential of wind-dispersed diaspores of neotropical trees. American Journal of Botany 73:353–363.

Bradley, B. A., D. M. Blumenthal, D. S. Wilcove, and L. H. Ziska. 2010. Predicting plant invasions in an era of global change. Trends in Ecology and Evolution 25:310–318.

Buckley, L. B., M. C. Urban, M. J. Angilletta, L. G. Crozier, L. J. Rissler, and M. W. Sears. 2010. Can mechanism inform species' distribution models? Ecology Letters 13:1041–1054.

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

Bullock, J. M., I. L. Moy, S. J. Coulson, and R. T. Clarke. 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., K. Shea, and O. Skarpaas. 2006. Measuring plant dispersal: an introduction to field methods and experimental design. Plant Ecology 186:217–234.

Bullock, J. M., S. M. White, C. Prudhomme, C. Tansey, R. Perea, and D. A. P. Hooftman. 2012. Modelling spread of British wind-dispersed plants under future wind speeds in a changing climate. Journal of Ecology 100:104–115.

Cain, M. L., H. Damman, and A. Muir. 1998. Seed dispersal and the Holocene migration of woodland herbs. Ecological Monographs 68:325–347.

Cayuela, L., I. Granzow-de la Cerda, F. S. Albuquerque, and D. J. Golicher. 2012. TAXONSTAND: an R package for species names standardisation in vegetation databases. Methods in Ecology and Evolution 3:1078–1083.

Christenhusz, M. J., J. L. Reveal, A. Farjon, M. F. Gardner, R. R. Mill, and M. W. Chase. 2011. A new classification and linear sequence of extant gymnosperms. Phytotaxa 19:55–70.

Clobert, J., M. Baguette, T. G. Benton, and J. M. Bullock. 2012. Dispersal ecology and evolution. Oxford University Press, New York, New York, USA.

Couvreur, M., B. Christiaen, K. Verheyen, and M. Hermy. 2004. Large herbivores as mobile links between isolated nature reserves through adhesive seed dispersal. Applied Vegetation Science 7:229–236.

Drezner, T. D., P. L. Fall, and J. C. Stromberg. 2001. Plant distribution and dispersal mechanisms at the Hassayampa River Preserve, Arizona, USA. Global Ecology and Biogeography 10:205–217.

Ernst, W. H. O., E. M. Veenendaal, and M. M. Kebakile. 1992. Possibilities for dispersal in annual and perennial grasses in a savanna in Botswana. Vegetatio 102:1–11.

Fox, J., and S. Weisberg. 2011. An {R} companion to applied regression. Second edition. Sage, Thousand Oaks, California, USA.

Garrard, G. E., M. A. McCarthy, P. A. Vesk, J. Q. Radford, and A. F. Bennett. 2012. A predictive model of avian natal dispersal distance provides prior information for investigating response to landscape change. Journal of Animal Ecology 81:14–23.

Greene, D. F., and E. A. Johnson. 1993. Seed mass and dispersal capacity in wind-dispersed diaspores. Oikos 67:69–74.

Guisan, A., and C. Rahbek. 2011. SESAM—a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. Journal of Biogeography 38:1433–1444.

Hein, A. M., C. Hou, and J. F. Gillooly. 2012. Energetic and biomechanical constraints on animal migration distance. Ecology Letters 15:104–110.

Higgins, K. 2009. Metapopulation extinction risk: dispersal's duplicity. Theoretical Population Biology 76:146–155.

Hintze, C., F. Heydel, C. Hoppe, S. Cunze, A. König, and O. Tackenberg. 2013. D^3: the dispersal and diaspore database—baseline data and statistics on seed dispersal. Perspectives in Plant Ecology, Evolution and Systematics 15:180–192.

Honnay, O., H. Jacquemyn, B. Bossuyt, and M. Hermy. 2005. Forest fragmentation effects on patch occupancy and population viability of herbaceous plant species. New Phytologist 166:723–736.

Hooftman, D. A. P., and J. M. Bullock. 2012. Mapping to inform conservation: a case study of changes in semi-natural habitats and their connectivity over 70 years. Biological Conservation 145:30–38.

Hughes, L., M. Dunlop, K. French, M. R. Leishman, B. Rice, L. Rodgerson, and M. Westoby. 1994. Predicting dispersal spectra—a minimal set of hypotheses based on plant attributes. Journal of Ecology 82:933–950.

Johst, K., R. Brandl, and S. Eber. 2002. Metapopulation persistence in dynamic landscapes: the role of dispersal distance. Oikos 98:263–270.

Jongejans, E., and A. Telenius. 2001. Field experiments on seed dispersal by wind in ten umbelliferous species (Apiaceae). Plant Ecology 152:67–78.

Kiviniemi, K., and A. Telenius. 1998. Experiments on adhesive dispersal by wood mouse: seed shadows and dispersal distances of 13 plant species from cultivated areas in southern Sweden. Ecography 21:108–116.

Kleyer, M., et al. 2008. The LEDA Traitbase: a database of life-history traits of the Northwest European flora. Journal of Ecology 96:1266–1274.

Lavergne, S., J. Molina, and M. Debussche. 2006. Fingerprints of environmental change on the rare mediterranean flora: a 115-year study. Global Change Biology 12:1466–1478.

Legendre, P., and L. Legendre. 1998. Numerical ecology. Elsevier, Amsterdam, The Netherlands.

Levin, S. A., H. C. Muller-Landau, R. Nathan, and J. Chave. 2003. The ecology and evolution of seed dispersal: a theoretical perspective. Annual Review of Ecology, Evolution, and Systematics 34:575–604.

Matlack, G. R. 1987. Diaspore size, shape, and fall behavior in wind-dispersed plant species. American Journal of Botany 74:1150–1160.

McConkey, K. R., S. Prasad, R. T. Corlett, A. Campos-Arceiz, J. F. Brodie, H. Rogers, and L. Santamaria. 2012. Seed dispersal in changing landscapes. Biological Conservation 146:1–13.

Moles, A. T., D. S. Falster, M. R. Leishman, and M. Westoby. 2004. Small-seeded species produce more seeds per square metre of canopy per year, but not per individual per lifetime. Journal of Ecology 92:384–396.

Moles, A. T., and M. Westoby. 2004. Seedling survival and seed size: a synthesis of the literature. Journal of Ecology 92:372–383.

Montoya, D., M. A. Zavala, M. A. Rodriguez, and D. W. Purves. 2008. Animal versus wind dispersal and the robustness of tree species to deforestation. Science 320:1502–1504.

Moora, M., M. Öpik, K. Zobel, and M. Zobel. 2009. Understory plant diversity is related to higher variability of vegetative mobility of coexisting species. Oecologia 159:355–361.

Muller-Landau, H. C., S. J. Wright, O. Calderon, R. Condit, and S. P. Hubbell. 2008. Interspecific variation in primary seed dispersal in a tropical forest. Journal of Ecology 96:653–667.

Nathan, R., G. G. Katul, G. Bohrer, A. Kuparinen, M. B. Soons, S. E. Thompson, A. Trakhtenbrot, and H. S. Horn. 2011. Mechanistic models of seed dispersal by wind. Theoretical Ecology 4:113–132.

Nathan, R., E. Klein, J. J. Robledo-Arnuncio, and E. Revilla. 2012. Dispersal kernels: review. Pages 187–210 in J. Clobert, M. Baguette, T. G. Benton, and J. M. Bullock, editors. Dispersal ecology and evolution. Oxford University Press, Oxford, UK.

Nathan, R., and H. C. Muller-Landau. 2000. Spatial patterns of seed dispersal, their determinants and consequences for recruitment. Trends in Ecology and Evolution 15:278–285.

Nathan, R., F. M. Schurr, O. Spiegel, O. Steinitz, A. Trakhtenbrot, and A. Tsoar. 2008. Mechanisms of long-distance seed dispersal. Trends in Ecology and Evolution 23:638–647.

Ness, J. H., J. L. Bronstein, A. N. Andersen, and J. N. Holland. 2004. Ant body size predicts dispersal distance of ant-adapted seeds: implications of small-ant invasions. Ecology 85:1244–1250.

Nogales, M., D. P. Padilla, C. Nieves, J. C. Illera, and A. Traveset. 2007. Secondary seed dispersal systems, frugivorous lizards and predatory birds in insular volcanic badlands. Journal of Ecology 95:1394–1403.

Norghauer, J. M., C. A. Nock, and J. Grogan. 2011. The importance of tree size and fecundity for wind dispersal of big-leaf mahogany. PLoS ONE 6:e17488.

Parolo, G., and G. Rossi. 2008. Upward migration of vascular plants following a climate warming trend in the Alps. Basic and Applied Ecology 9:100–107.

Pärtel, M., R. Szava-Kovats, and M. Zobel. 2011. Dark diversity: shedding light on absent species. Trends in Ecology and Evolution 26:124–128.

Pärtel, M., and M. Zobel. 2007. Dispersal limitation may result in the unimodal productivity-diversity relationship: a new explanation for a general pattern. Journal of Ecology 95:90–94.

Pinheiro, J. C., and D. M. Bates. 2000. Mixed-effect models in S and S-Plus. Springer, New York, New York, USA.

Pinheiro, J., D. Bates, S. DebRoy, Sarkar, D. and R Development Core Team. 2011. nlme: linear and nonlinear mixed effects models. R package version 3.1-102. http://cran.r-project.org/web/packages/nlme/index.html

Poschlod, P., S. Kiefer, U. Tränkle, S. Fischer, and S. Bonn. 1998. Plant species richness in calcareous grasslands as affected by dispersability in space and time. Applied Vegetation Science 1:75–90.

R Development Core Team. 2012. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Reinhart, K. O., G. W. T. Wilson, and M. J. Rinella. 2012. Predicting plant responses to mycorrhizae: integrating evolutionary history and plant traits. Ecology Letters 15:689–695.

Russo, S. E., S. Portnoy, and C. K. Augspurger. 2006. Incorporating animal behaviour into seed dispersal models: implications for seed shadows. Ecology 87:3160–3174.

Schielzeth, H., and W. Forstmeier. 2009. Conclusions beyond support: overconfident estimates in mixed models. Behavioral Ecology 20:416–420.

Seidler, T. G., and J. B. Plotkin. 2006. Seed dispersal and spatial pattern in tropical trees. PLoS Biology 4:e344.

Sheldon, J. C., and F. M. Burrows. 1973. The dispersal effectiveness of the achene-pappus units of selected Compositae in steady winds with convection. New Phytologist 72: 665–675.

Smith, J. 2009. Can regression discontinuity help answer an age-old question in education? The effect of age on elementary and secondary school achievement. BE Journal of Economic Analysis and Policy 9:1–30.

Soons, M. B., J. H. Messelink, E. Jongejans, and G. W. Heil. 2005. Habitat fragmentation reduces grassland connectivity for both short-distance and long-distance wind-dispersed forbs. Journal of Ecology 93:1214–1225.

Soons, M. B., and W. A. Ozinga. 2005. How important is long-distance seed dispersal for the regional survival of plant species? Diversity and Distributions 9:165–172.

Stevens, V. M., A. Trochet, H. Van Dyck, J. Clobert, and M. Baguette. 2012. How is dispersal integrated in life histories: a quantitative analysis using butterflies. Ecology Letters 15:74–86.

Stöcklin, J., and E. Bäumler. 1996. Seed rain, seedling establishment and clonal growth strategies on a glacier foreland. Journal of Vegetation Science 7:45–56.

Thomson, F. J., A. T. Moles, T. D. Auld, and R. T. Kingsford. 2011. Seed dispersal distance is more strongly correlated with plant height than with seed mass. Journal of Ecology 99: 1299–1307.

Thomson, F. J., A. T. Moles, T. D. Auld, D. Ramp, S. Q. Ren, and R. T. Kingsford. 2010. Chasing the unknown: predicting seed dispersal mechanisms from plant traits. Journal of Ecology 98:1310–1318.

Trakhtenbrot, A., R. Nathan, G. Perry, and D. M. Richardson. 2005. The importance of long-distance dispersal in biodiversity conservation. Diversity and Distributions 11:173–181.

Vittoz, P., N. Dussex, J. Wassef, and A. Guisan. 2009. Diaspore traits discriminate good from weak colonisers on high-elevation summits. Basic and Applied Ecology 10:508–515.

Vittoz, P., and R. Engler. 2007. Seed dispersal distances: a typology based on dispersal modes and plant traits. Botanica Helvetica 117:109–124.

Warton, D., R. Duursma, D. Falster, and S. Taskinen. 2011. smatr: (standardised) major axis estimation and testing routines. R package version 3.2.4. http://CRAN.R-project.org/package=smatr

Willson, M. F. 1993. Dispersal mode, seed shadows, and colonization patterns. Vegetatio 108:261–280.

Xiao, Z., Z. Zhang, and Y. Wang. 2005. Effects of seed size on dispersal distance in five rodent-dispersed fagaceous species. Acta Oecologica 28:221–229.

Zobel, M. 1997. The relative role of species pools in determining plant species richness. An alternative explanation of species coexistence? Trends in Ecology and Evolution 12:266–269.

SUPPLEMENTAL MATERIAL

Appendix A

Overview of the results for linear models to explain maximum dispersal distances using different combinations of plant traits, and histograms of maximum dispersal distance data for different dispersal syndromes and growth forms (*Ecological Archives* E095-045-A1).

Appendix B

Overview of the predictive power of models with random and fixed effects included in the models (*Ecological Archives* E095-045-A2).

Supplement 1

Data about seed dispersal distances and related traits for 576 plant species used in the analyses, and references for data sources (*Ecological Archives* E095-045-S1).

Supplement 2

Function dispeRsal for predicting seed dispersal distances in R, and instructions for using dispeRsal (*Ecological Archives* E095-045-S2).