Bullock, James M.; Hooftman, Danny A.P.; Tamme, Riin; Götzenberger, Lars; Pärtel, Meelis; Mallada Gonzalez, Laura; White, Steven M. 2018. **All dispersal functions are wrong, but many are useful: a response to Cousens et al**. *Journal of Ecology*, 106 (3). 907-910, which has been published in final form at https://doi.org/10.1111/1365-2745.12890

All dispersal functions are wrong, but many are useful: a response to Cousens et al.

James M. Bullock[1]*, Danny A.P. Hooftman[1,2], Riin Tamme[3], Lars Götzenberger[4], Meelis Pärtel[3], Laura Mallada González[1], and Steven M. White[1,5]

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

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

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.

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

Running headline: Useful dispersal functions

Summary

1. To address the lack of information about the shape and extent of real dispersal kernels, Bullock *et al.* (2017) synthesized empirical information on seed dispersal distances. Testing the fit of a variety of probability density functions, they found no function was the best-fitting for all datasets but some outperformed others. Cousens, Hughes and Mesgaran (2017) focus on their specific finding of the generally poor fit of the WALD function to wind dispersal data and use this to argue that mechanistically derived functions would not be expected to fit data particularly well.
2. We agree in part with this argument and discuss the issues that may lead to poor fit, including the simplifying assumptions of the WALD and the complexity of the dispersal process. We explain the fundamental linkage between the mechanistic form of the WALD and the derived function used for fitting to data.
3. We demonstrate however, that the logic that a mechanistically based function could fit to data is valid, under the hypothesis that it encompasses the key processes determining the dispersal kernel. This argument is supported by the facts that: a) our analyses and others have shown the WALD performs well in a number of cases; and b) the WALD is the best fitting function for an example in which we simulate dispersal data using a realistic representation of variability in the wind dispersal process.
4. *Synthesis.* While there are reasons that mechanistically derived functions may not fit well to empirical data, they do in some empirical and simulated cases and this suggests they can capture the dispersal behaviour of real systems. Mechanistic functions should be explored along with other more general functions when describing empirical data to investigate their simplifying assumptions and to add to our arsenal of functions for analysing dispersal data. Analyses using these functions are critical if we are to move from simply describing the system in which the data were gathered to gaining more general insights into dispersal and predicting its consequences.

Key-words: dispersal kernel, inverse Gaussian, probability density function, prediction, seed dispersal, WALD**,** wind dispersal

Introduction

Our synthesis of plant dispersal studies (Bullock *et al.* 2017) analysed the rich empirical information on seed dispersal distances from studies on a wide variety of plants across many ecosystems worldwide. A major aspect of our study was the fitting of a number of widely used probability density functions to these data sets, and a comparison of their performance. We found that many of these straightforward functions described the empirical data well, but the performance of alternative functions varied. No single function fitted all data sets well, but certain functions – the exponential power and log-sech – were the best performing on average. We then fitted these two functions to data sets that represented combinations of dispersal mode and plant growth form. These functions fit the combined data sets well despite variation among studies in empirical methods, local conditions, vegetation type and the exact dispersal process. The analysis of alternative dispersal kernels and presentation of generalized kernels for growth form/dispersal mode groups provides a rich resource for ecologists, and we described ways by which this improved information might enrich spatial ecology.

We take this opportunity to correct typographical errors in our paper in the equations for: the 2Dt, which should be $\frac{b-1}{\pi a^2}\left(1+\frac{d^2}{a^2}\right)^{-b}$; the gamma, which should be $\frac{1}{2\pi a^2\Gamma(b)}\left(\frac{d}{a}\right)^{b-2}\exp\left(-\frac{d}{a}\right)$; and the Weibull, which should be $\frac{b}{2\pi a^2}d^{b-2}\exp\left(-\frac{d^b}{a^b}\right)$.

Among other probability density functions, our analysis included the WALD, which is based on a mechanistic description of seed dispersal by wind (Katul *et al.* 2005). We noted the fact that when fitted to datasets describing seed dispersal by wind, the WALD performed poorly compared with many other functions, in that it was among the best-fitting functions in relatively few cases. Cousens, Hughes and Mesgaran (2017) focus on this specific aspect of our paper and suggest one would not expect the WALD to fit empirical data particularly well. They give two closely interlinked reasons for this, which can be summarised as an argument that the simplifications of the WALD make it unlikely to fit the complexity of real data. The WALD is based on an assumption of a single seed release height and unvarying environmental conditions (including wind speed) during the dispersal period. We do not disagree with this argument in general – in fact we made a similar argument in our paper. But, we show below how Cousens, Hughes and Mesgaran (2017) over-simplify the issues and therefore unnecessarily downplay the utility of mechanistically based functions in describing empirical data.

Why the WALD might not fit real dispersal kernels

The WALD function is based upon simplifications to an idealised three-dimensional Lagrangian stochastic dispersal model for the trajectories of air particles having no mass in turbulent flows, where the drift and diffusion terms are determined by assuming a high Reynolds number and well-mixed conditions, modelled by a generalised Fokker-Planck equation (Thomson 1987; Katul *et al.* 2005). The final function form for the WALD is derived to make further simplifying assumptions, which we discuss below. These simplifications result in an inverse Gaussian distribution, which is considerably more useful to ecologists than some cumbersome stochastic differential equation which retains the full complexity of dispersal by wind.

The equation given in our paper (see also Nathan *et al.* (2012)) is the re-parameterised WALD suitable for fitting to dispersal data by finding solutions for the parameters *a* and *b*, whereby the probability density of seeds at distance $d = \frac{\sqrt{b}}{\sqrt{8\pi^3 d^5}} \exp\left(-\frac{b(d-a)^2}{2a^2 d}\right)$, which in this form is the 2-dimensional dispersal location kernel (see Bullock *et al.* 2017). This is derived by Katul *et al.* (2005)) from the mechanistic model, which allows calculation of a dispersal kernel from measures of plants and the environment. Specifically, $a = \frac{H\overline{U}}{F}$ and $b = \left(\frac{H}{\overline{\sigma}}\right)^2$, where *H* is the seed release height, *F* is the seed terminal velocity, $\overline{U}$ is the mean wind speed at the height of seed release and $\overline{\sigma}$ is a turbulent flow parameter (Katul *et al.* 2005; Bullock *et al.* 2012). Since the WALD is mechanistically derived and parameterised by plant traits and environmental variables, one may use these readily available data to predict dispersal and spread without *a priori* obtained dispersal data (Bullock *et al.* 2012; Hemrová *et al.* 2017). This means the fitted function is fundamentally linked to the theory of the mechanistic model.

As is clear in our paper, we agree that there are good reasons why the WALD may not fit empirical data well, but these are several. Cousens, Hughes and Mesgaran (2017) give one suggestion. In our paper we suggested two additional and equally valid mechanisms by which the model might not fit to empirical data. Considering the underlying theory, simplifying assumptions in the WALD include (Katul *et al.* 2005): flow is vertically homogeneous; seed terminal velocity is achieved instantly after seed release; the seed settling time is assumed to be much longer than the vertical velocity integral timescale; and the simplifications to the Thomson (1987) model, including Gaussian fluctuations and the use of Kolmogorov scaling within the inertial subrange to arrive at the diffusion coefficient.

Many of the coefficients in the WALD are averaged, which allows the full equations to be simplified from the underlying equations. One such coefficient is the mean wind speed, given by $\bar{U}$ in Katul *et al.* (2005). There is variation over a season in the wind speed a falling seed might experience, as Cousens, Hughes and Mesgaran (2017) state, but it will also vary over the time that the seed takes to fall and hence $\bar{U}$ could be modelled as a function of time, which would result in an intractable non-closed-form equation. Another example is the seed release height, as this will vary naturally. As Cousens, Hughes and Mesgaran (2017) suggest, one could sum up all the possible release heights of individual seeds and the corresponding WALDs to get a new kernel, which is hard to work with. Or one might convolve some distribution of seed release heights (e.g. a Gaussian) with a WALD. This might fit better, but there are now extra parameters and one could go on like this with a large number of possible combinations.

Why the WALD does fit real dispersal kernels, sometimes

Despite these issues, there are good reasons for fitting the WALD to data. The logic that a mechanistically based function might fit well to data is valid, as it is hoped that it encompasses the relevant processes determining the dispersal kernel and so captures the dispersal kernel. Thus, clear hypotheses are set up about the determinants of the realised kernel. Indeed, the WALD has been used by others when fitting functions to data. In the original paper proposing the WALD, Katul *et al.* (2005) proposed and implemented fitting it to measured dispersal kernels, while also introducing the assumptions in doing so which are being discussed here. The WALD has been tested in some studies in which alternative functions are compared. Studying a wind-dispersed tree, Norghauer, Nock and Grogan (2011) found the WALD and Weibull functions gave comparable and better fits to dispersal data than the lognormal. Lara-Romero *et al.* (2014) fitted functions to seedling data for two herbs using inverse modelling and found the WALD was at least as good a fit to the data as the 2Dt, exponential power and lognormal. It should also be clarified that in our study the WALD was by no means a poor fit to wind dispersal data in all cases, belying the implication that it will never fit empirical data well. Of the 55 wind dispersal data sets, the WALD was in the best-fit group for 15, and had an r^2>0.9 for 25.

One way to examine the ability of the WALD to describe dispersal data from a varying environment is to draw dispersal distances from WALD functions representing variation in parameter values, and then assess how well a single WALD fits these data in turn. Cousens, Hughes and Mesgaran (2017) do this, but their example is unrealistic. They make draws from a WALD with the parameters *a* and *b* varying independently each time "according to a uniform distribution of several orders of magnitude". In reality, these parameters are unlikely to vary either uniformly or over such wide ranges as they are based on plant and environmental variables (e.g. seed release height, wind speed), which are likely to be more tightly distributed and closer to the mean. Furthermore the parameters are correlated, i.e. both reflect wind conditions and plant height, and so do not vary independently. In Fig. 1 we develop a more realistic example, using data from Bullock *et al.* (2012) in which we showed that variation in horizontal wind speed over a season follows a Weibull distribution. Generating a dispersal data set using a WALD sampled over a Weibull distribution of wind speeds, we find that the WALD is a better fit (Fig. 1) than the log-sech or the other functions that we investigated in Bullock *et al.* (2017). This shows that it is sensible to ask the question whether a WALD fits wind dispersal data. We would note however, that if multiple parameters of the WALD (e.g. *H, F, U*) were allowed to vary over realistic distributions and convolved with the WALD, then the resulting distribution might take on a number of forms, and not necessarily the WALD.

Conclusion

In Bullock *et al.* (2017) we showed that it is possible to summarise the complex and variable dispersal process using simple functions over a large number of empirical data sets. We found none of the functions we used gave best fit overall, suggesting no single function captures the dispersal process intrinsically. Mechanistically based functions may fail to describe such data for reasons set out by us in the original paper, by Cousens, Hughes and Mesgaran (2017), and expanded upon here. We advocate however that these functions are explored along with other more general functions when describing empirical data both to assess whether their simplifying assumptions are valid when tested in the real world and to add to our arsenal of possible functions for analysing data. Parametric summaries of dispersal data are critical if we are to use the past and ongoing work of ecologists in gathering dispersal data for more than simply describing the system in which the data were gathered.

Prediction in ecology aims both to explain systems and to forecast, or anticipate, future changes (Mouquet *et al.* 2015). In line with both aims, our paper synthesized dispersal information and provided general dispersal functions. These are of use to researchers who may either not have the necessary data to model their system or may not be interested in case specific kernels. These general and better validated kernel functions would be useful, for example, in species distribution modelling (Miller & Holloway 2015), analysing spatial networks (Marleau, Guichard & Loreau 2014) and predicting responses to climate change (Santini *et al.* 2016).

Acknowledgements

This research was supported by CEH project NEC06429, the Estonian Ministry of Education and Research (IUT20-29) and the European Regional Development Fund (Centre of Excellence EcolChange).

References

Bullock, J.M., Mallada González, L., Tamme, R., Götzenberger, L., White, S.M., Pärtel, M. & Hooftman, D.A.P. (2017) A synthesis of empirical plant dispersal kernels. *Journal of Ecology,* **105,** 6-19.

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

Cousens, R.D., Hughes, B.D. & Mesgaran, M.B. (2017) Why we do not expect dispersal probability density functions based on a single mechanism to fit real seed shadows. *Journal of Ecology*.

Hemrová, L., Bullock, J.M., Hooftman, D.A.P., White, S.M. & Münzbergová, Z. (2017) Drivers of plant species' potential to spread: the importance of demography versus seed dispersal. *Oikos***.**

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

Lara-Romero, C., Robledo-Arnuncio, J.J., Garcia-Fernandez, A. & Iriondo, J.M. (2014) Assessing intraspecific variation in effective dispersal along an altitudinal gradient: a test in two Mediterranean high-mountain plants. *PLoS ONE,* **9,** 10.

Marleau, J.N., Guichard, F. & Loreau, M. (2014) Meta-ecosystem dynamics and functioning on finite spatial networks. *Proceedings of the Royal Society B-Biological Sciences,* **281,** 9.

Miller, J.A. & Holloway, P. (2015) Incorporating movement in species distribution models. *Progress in Physical Geography,* **39,** 837-849.

Mouquet, N., Lagadeuc, Y., Devictor, V., Doyen, L., Duputié, A., Eveillard, D., Faure, D., Garnier, E., Gimenez, O., Huneman, P., Jabot, F., Jarne, P., Joly, D., Julliard, R., Kéfi, S., Kergoat, G.J., Lavorel, S., Le Gall, L., Meslin, L., Morand, S., Morin, X., Morlon, H., Pinay, G., Pradel, R., Schurr, F.M., Thuiller, W. & Loreau, M. (2015) Predictive ecology in a changing world. *Journal of Applied Ecology,* **52,** 1293-1310.

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

Norghauer, J.M., Nock, C.A. & Grogan, J. (2011) The importance of tree size and fecundity for wind dispersal of big-leaf mahogany. *PLoS ONE,* **6,** 12.

Santini, L., Cornulier, T., Bullock, J.M., Palmer, S.C.F., White, S.M., Hodgson, J.A., Bocedi, G. & Travis, J.M.J. (2016) A trait-based approach for predicting species responses to environmental change from sparse data: how well might terrestrial mammals track climate change? *Global Change Biology,* **22,** 2415-2424.

Thomson, D.J. (1987) Criteria for the selection of stochastic models of particle trajectories in turbulent flows. *Journal of Fluid Mechanics,* **180,** 529-556.

Fig. 1. A illustration that dispersal data generated from a WALD probability density function with variation in parameter values are in turn fitted well by a WALD function with a single value for each parameter. We used the WALD to model dispersal mechanistically for the wind dispersed orchid *Himantoglossum hircinum*, as parameterised by Bullock *et al.* (2012) from measured plant and environmental characteristics. In that study, variation in wind speed through the dispersal season followed a Weibull distribution (r^2>0.99). To represent variation in the wind speed experienced by seeds as they are released from the plant, we drew 10,000 wind speeds from the fitted Weibull and used each to parameterise a WALD, and then drew a single dispersal distance from each individual WALD. We counted the number of seeds in 0.25 m distance bins: this bin size was selected as it represented well the shape of the resulting dispersal kernel (especially the non-zero mode), without giving an excessive number of bins. We then fitted the 11 probability density functions described by Bullock *et al.* (2017) to this kernel, using the dispersal distance kernel formulation (Nathan *et al.* 2012). The WALD fit best, having the lowest AIC and a r^2 (calculated as in Bullock *et al.* (2017)) of 0.981. The figure illustrates: the generated dispersal data, which we curtail at 10 m (encompassing 96% of individual dispersal distances) for this graph to aid clarity; the fitted WALD and 2Dt, which were the best and second best fitting functions respectively; and the power exponential and log-sech, which (Bullock *et al.* (2017)) showed fit well to data generally, but in this case did not perform particularly well.