Author/s:
Cousens, RD;Hughes, BD;Mesgaran, MB

Title:
Why we do not expect dispersal probability density functions based on a single mechanism to fit real seed shadows

Date:
2018-05-01

Citation:
Cousens, R. D., Hughes, B. D. & Mesgaran, M. B. (2018). Why we do not expect dispersal probability density functions based on a single mechanism to fit real seed shadows. Journal of Ecology, 106 (3), pp.903-906. https://doi.org/10.1111/1365-2745.12891.

PROFESSOR ROGER DAVID COUSENS (Orcid ID : 0000-0003-2192-0183)

Article type : Forum

Editor : Shurong Zhou

FORUM

Why we do not expect dispersal probability density functions based on a single mechanism to fit real seed shadows

Roger D. Cousens[1*], Barry D. Hughes[2] and Mohsen B. Mesgaran[1]

[1]School of BioSciences and [2]School of Mathematics & Statistics, The University of Melbourne, Victoria 3010, Australia

---

Summary

1. Bullock et al. (2017) have suggested that the theory behind the WALD (Wald Analytical Long Distance) model for wind dispersal from a point source needs to be re-examined. This is on the basis that an inverse Gaussian probability density function (pdf) does not provide the best fit to seed shadows around individual source plants known to be dispersed by wind.
2. We present two reasons why we would not necessarily expect any of the standard mechanistically-derived pdfs to fit real seed shadows any better than empirical functions.
3. Firstly, the derivation of "off-the-shelf" pdfs such as the Gaussian, exponential and inverse Gaussian involves only one of the processes and factors that together generate a real seed shadow. It is implausible to expect that a single-process model, no matter how sophisticated in detail, will capture the behaviour of an entire, complex system, which may

involve a number of sequential random processes, or a superposition of parallel random processes, or both.

4. Secondly, even if there is only one process involved and we have a perfect model for that process, the basic parameters of the model would be difficult to pin down precisely. Moreover, these parameters are unlikely to remain constant over a dispersal season, so that effectively we observe the outcome of a linear combination of dispersal events with different parameter values, constituting a form of averaging over the parameters of the distribution. Simple examples show that averaging a pdf over its parameters can lead to a pdf from an entirely different class.
5. *Synthesis and applications*. The failure of the inverse Gaussian model to fit seed shadow data is not in itself a reason to doubt the validity of the WALD model for movement of particles through the air under specified environmental conditions. A greater awareness is needed of the differences between the WALD and the inverse Gaussian (or Wald) and the purposes for which they are used. The complexity of dispersing populations of seeds means that any of the standard mechanistically-derived pdfs will actually be merely empirical in this context. Shape and flexibility of a pdf is far more important for adequately describing data than some perceived higher status.

*Correspondence author. E-mail rcousens@unimelb.edu.au

Introduction

In their recent meta-analysis of seed dispersal data, Bullock et al. (2017) observed that the inverse Gaussian probability density function (pdf) does not fit real dispersal kernel data (empirical seed shadows) for wind-dispersed species very well, despite it being related to the WALD model for dispersal by wind (Katul et al. 2005). Because of this they suggest that the theory behind the WALD model needs to be revisited. We explain why, regardless of whether or not the WALD itself is a good model, this deduction is not logical. But first it is necessary to review some important definitions, since inconsistency in nomenclature has, unfortunately, resulted in some confusion.

The WALD model (Katul et al. 2005) was derived as a simplification of a highly complex mechanistic model of movement of unattached propagules from a point source by wind. Katul et al. (2005) showed that it was reasonable at predicting frequency distributions of controlled-release dispersal distance data, using parameter estimates based only on plant traits and extant weather conditions. Since their model has the same functional form as the inverse Gaussian distribution (often also called the Wald, although strictly the latter is only a special case of the inverse Gaussian: Seshadri 1993, 1999) they named their model the Wald Analytical Long-Distance (WALD) model. The inverse

Gaussian (or the Wald) and the WALD are *not* synonymous even though the derivation of both involves the first passage time problem for Brownian motion with drift (Schrödinger 1915).

The WALD is only "analytical" in its original context of the movement of particles under specified wind conditions. An inverse Gaussian can be fit to natural seed shadow data, without wind dispersal being implied at all, as just another phenomenological model that may happen to describe the data. Using the original formulation of the WALD and assuming that it is reasonable to expect a real seed shadow to be approximated by propagules behaving as they do from experimental releases, it is then possible to obtain "quasi-mechanistic" (Jongejans et al. 2008) estimates of the WALD parameters.

We consider that there are two main reasons why it is illogical to argue that the WALD should be reconsidered simply because an inverse Gaussian does not fit seed shadow data for wind dispersed propagules.

1. The standard mechanistically-derived pdfs do not describe all the mechanisms

The standard so-called "mechanistic" pdfs used to describe dispersal are derived from quite restricted assumptions about the motion of populations of objects: all individuals start from a single point source, travel under narrowly-defined environmental conditions and are identical in the properties that determine their movement. This is the case for the various pdfs derived from Brownian motion, random walks and from wind dispersal. But we know that these assumptions are inappropriate for the population of seeds dispersing from a plant over its lifetime, which is what we require for making inferences about spatial population dynamics.

We should distinguish between (i) models merely intended to summarise positions of naturally dispersed seeds from (ii) those intended to improve our understanding of movement processes – which we may test using artificial releases. One model may not be sufficient for both purposes. Even though we may, for simplicity, base simulations of a mechanistic movement model on mean parameter values justified by arguments that levels of variation are low or of little consequence (e.g. Nathan et al. 2002), this does not necessarily mean that we expect these assumptions to be valid for every real seed in a real location under a real set of environmental conditions. Every model is simplistic to some extent, and this is certainly the case for current dispersal pdfs. A mechanistic model of *post-release* movement, such as the WALD, describes just one component of a system that may be highly complex (Cousens et al. 2008). We give two illustrations.

(a) Even if our mechanistic model is perfectly fitted to the particular aspect of the dispersal process that the model is based on, the final dispersed position of a seed may be the sum of a number of sequential random displacement processes. For example, a fruit may burst explosively, producing randomly distributed seed landing points in still air conditions. Any wind present at the time of liberation, strong winds subsequent to landing depending where the seed falls, and rain can all provide additional displacement components. The pdfs of sums of several independent random variables do not typically belong to the same family as any of the constituent random variables. This issue has been extensively explored in the probability literature (Feller 1971, Gnedenko & Kolmogorov 1968, Hughes 1995, Zolotarev 1986). If we consider a class of random variables for which the densities of each differ only in terms of a centering parameter and a horizontal scale parameter, the class is called "stable" if the sum of two independent random variables from the class also belongs to the class. Although the classes of stable heavy-tailed random variables with infinite variance have also been definitively determined, it has been established that the normal (or Gaussian) pdf is the unique stable class with finite variance. Adding two independent normal variables produces a normal variable (its mean being the sum of the means of the constituents, and its variance the sum of the variances of the constituents). In contrast, for example, it is easily demonstrated by integration that the sum of two independent random variables with identical one-dimensional Laplace densities $(k/2)\exp(-k|x|)$ produces a random variable with density $(k/4)(1+k|x|)\exp(-k|x|)$, which looks very different to a Laplace density.

(b) Instead of seed displacements reflecting a number of concatenated individual displacement events for each seed, there may be displacements by multiple vectors, leading to a superposition of dispersals by each vector. Significant dispersal by a second vector not well modelled by the preferred mechanistic model for the first vector also compromises the ability of the mechanistic model to describe the full dispersal process. Clark et al. (1998) showed that the combination of two exponentially-dispersed sub-populations, representing long- and short-distance dispersal units, produced spatial population dynamics resembling those of a fat-tailed distribution rather than an exponential.

In short, then, the value of a mechanistic model in fitting field data rests not only on the aptness of the model for the dispersal process that it captures, but on the absence of significant contribution from other sequential and/or concurrent dispersal mechanisms. So it is hardly surprising that a real seed shadow may be totally different in shape from a highly simplistic model of one component (e.g. Neubert & Parker 2004). It is worth noting that the WALD "mechanistically based" density fitted Aleppo Pine dispersal data much better when it was combined with other functions describing the local environment (Schurr et al. 2008). Moreover, we do not expect some of the modelled mechanisms to be appropriate in a particular circumstance, so in itself the fact that a pdf is "mechanistically-based" does not necessarily give it some sort of higher status than a purely empirical pdf. For example, we do not expect most animals dispersing seeds to follow a random walk or wind-dispersed seeds to be under Brownian motion. Ergo, it does not add any additional validity to (say) a multi-generational spatial population dynamics model or to a simulation model used to determine optimal sampling strategies (Skarpaas & Shea 2007), for the dispersal function to be a "mechanistic" pdf. General shape or mathematical simplicity may be more relevant.

2. Even assigning parameters to a mechanism is problematic

Propagules mature over periods of several weeks during which time dispersal vector behaviour varies considerably; they are commonly dispersed from highly distributed sources depending on plant architecture; the minimum force required for release will depend on their maturity; the actual release force may depend on previous forces applied to the propagule and its maturity at that time; propagules from the same plant vary considerably in their physical attributes and therefore in their capacity for movement; their movement may well be impeded by the maternal plant and by other plants along its trajectory; and different propagules may be dispersed by different vectors as we have already noted above. And so on.

The simplest way to attempt to account for this underlying variability is to make the parameters of the chosen dispersal model randomly variable between dispersal events, as, for example, Clark et al. (1999) have done in developing their 2Dt model. To illustrate this point in Fig. 1, for a large number of isotropic independent individual dispersal events in the plane governed by a single dispersal law with a random parameter, the resultant total dispersal is not well described by any parameter choice in the individual dispersal law, but fits well to an empirical law with few parameters unrelated to the law that governs individual displacements. It follows from Fig. 1 that although lack of fit of a mechanistically-based pdf *might* reflect additional mechanisms that need to be accounted for, thus advancing our ecological understanding, an alternative explanation is highly likely: that lack of fit tells us little about how adequately we have modelled the mechanisms, merely indicating that the environment (hence the model's parameters) varies considerably.

Conclusions

It is fallacious to consider a mechanistic model for one process as a failure if it does not fit a pdf resulting from all processes combined. Our view is that the search for a pdf to describe whole plant seed shadow data should not be constrained by preconceptions about desirable traits (such as a fat tail) or mechanisms. We would expect each seed shadow to be context-specific, so as Bullock et al. (2017) point out we would not expect to find a universal one-fits-all pdf. The decision should logically be based on the data. We may need several flexible functions to choose between.

The paradox, then, is how to compare shapes and scales of dispersal across contexts if we fit different equations. One can calculate moments, such as the mean and leptokurtosis for the function best fitting each data set; for some data non-parametric statistical "kernel density estimates" could be used to calculate distance quantiles (cf. definition of long-distance dispersal: Jordano et al. 2017). Whatever approach is taken, it is probably more important right now to focus on obtaining data that are up to the task of adequately describing the true population of dispersal distances; many published data sets are not. All pdfs, thin- or fat-tailed, will provide poor estimates of what really happens at extreme distances if there are few data in that region, no matter what up-to-date statistical methods are used in each case or how many data sets there are in a meta-analysis. This problem would be more readily appreciated if measures of parameter confidence were more widely examined, rather than just the estimates themselves.

Acknowledgements

MBM was funded from Australian Research Council grant DP140100608. We thank James Bullock for supplying his SAS code. The authors have no conflicts of interest to declare.

Authors' Contributions

RDC identified the issue and drafted the outline of the arguments; BDH contributed mathematical theory and critical restructuring; together they wrote and edited the text. MBM wrote code, designed and generated simulations and graphs. All authors contributed critically to the drafts and gave final approval for publication.

Data Accessibility

No empirical data are reported in this study. Fig. 1 is for illustration only and was generated by computer simulation. Model code available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.70pt4 (Cousens, 2017)

References

Bullock, J. M., González, L. M., 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.

Clark, J.S., Fastie, C., Hurtt, G., Jackson, S.T., Johnson, C., King, G.A., Lewis, M., Lynch, J., Pacala, S., Prentice, C., Schupp, E.W., Webb, T. & Wyckoff, P. (1998) Reid's paradox of rapid plant migration: Dispersal theory and interpretation of paleoecological records. *Bioscience*, **48**, 13-24.

Clark, J.S., Silman, M., Kern, R., Macklin, E. & HilleRisLambers, J. (1999) Seed dispersal near and far: patterns across temperate and tropical forests. *Ecology*, **80**, 1475-1494.

Cousens, R., Dytham, C. & Law, R. (2008) *Dispersal in Plants: A Population Perspective*. Oxford University Press, Oxford.

Feller, W. (1971) *An Introduction to Probability Theory and its Applications, Vol. 2, 2nd edition.* Wiley, New York.

Gnedenko, B.V. & Kolmogorov, A.N. (1968) *Limit Distributions for Sums of Independent Random Variables, revised edition.* Addison-Wesley, Reading, Massachusetts.

Hughes, B.D. (1995) *Random Walks and Random Environments*, Vol. 1. Oxford University Press.

Jongejans, E., Skarpaas, O. & Shea, K. (2008) Dispersal, demography and spatial population models for conservation and control management. *Perspectives in Plant Ecology, Evolution and Systematics*, **9**, 153-170.

Jordano, P. (2017) What is long-distance dispersal? And a taxonomy of dispersal events. *Journal of Ecology*, **105**, 75-84.

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.

Nathan, R., Katul, G.G., Horn, H.S., Thomas, S.M., Oren, R., Avissar, R., Pacala, S.W. & Levin, S.A. (2002) Mechanisms of long-distance dispersal of seeds by wind. *Nature*, **418**, 409-413.

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

Schrödinger, E. (1915) Zur Theorie der Fall- und Steigversuche an Teilchen mit Brownscher Bewegung. *Physikalische Zeitschrift*, **16**, 289–295.

Schurr, F.M., Steinitz, O. & Nathan, R. (2008) Plant fecundity and seed dispersal in spatially heterogeneous environments: models, mechanisms and estimation. *Journal of Ecology*, **96**, 628-641.

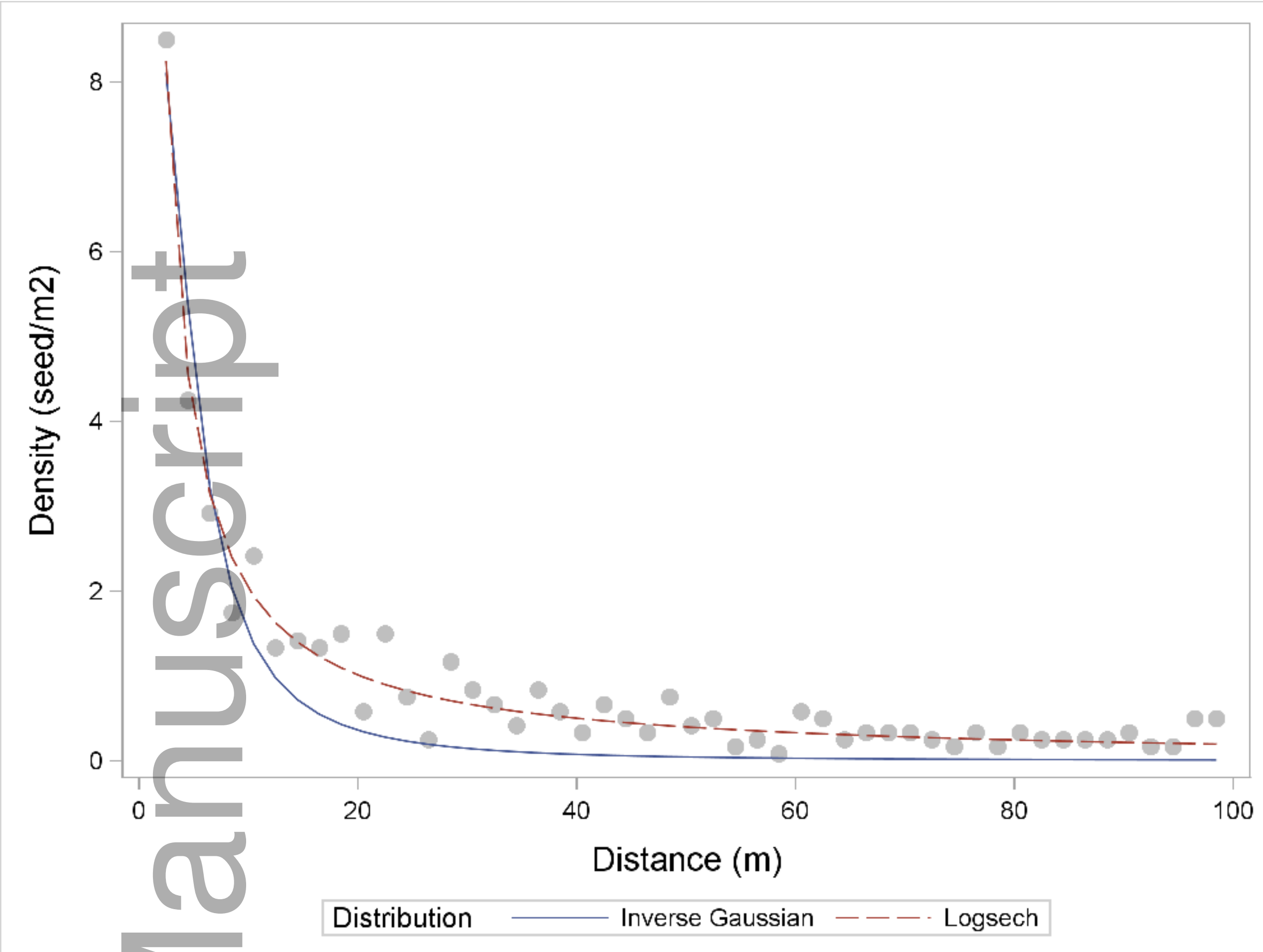
Seshadri, V. (1993) *The Inverse Gaussian Distribution: A Case Study in Exponential Families*. Oxford University Press, Oxford.

Seshadri, V. (1999) *The Inverse Gaussian distribution. Lecture Notes in Statistics Vol. 137.* Berlin: Springer-Verlag, New York.

Skarpaas, O. & Shea, K. (2007) Dispersal patterns, dispersal mechanisms, and invasion wave speeds for invasive thistles. *American Naturalist*, **107**, 421-430.

Zolotarev, V.M. (1986) *One-Dimensional Stable Distributions.* Translations of Mathematical Monographs, Vol. 65. American Mathematical Society, Providence, Rhode Island.

Fig.1. Empirical illustration of the fact that the distribution of a population of distances, each drawn from a pdf with variable parameters, is not expected to follow the same pdf. Using Matlab, 100,000 distance values were drawn from a one-dimensional inverse Gaussian distribution with both parameters allowed to vary each time according to a uniform distribution of several orders of magnitude; angles were drawn from a uniform distribution. Numbers that would be “caught” in circular traps at 2 m intervals along 12 radial transects were sampled; ten two-dimensional distributions were fitted to the density vs distance data in SAS as per Bullock et al. (2017). In this particular case, the log-sech (dashed line) was the best fit according to the Akaike Information Coefficient; the inverse Gaussian (solid line) was only seventh best and shows clear systematic lack of fit. Each dot represents the mean of 12 values.



jec_12891_f1.png