Fast dispersal simulation using bivariate quantiles

F. H. Evans[a]**, A. J. Diggle**[a]

[a]*Department of Agriculture and Food Western Australia, Locked Bag 4, Bentley Delivery Centre,
Western Australia, 6983*
Email: *art.diggle@agric.wa.gov.au*

Abstract: Spatial-temporal models of the spread of invasive species can require dispersal of large numbers of individuals from many locations at recurrent times, making them slow to execute. We present a fast algorithm for simulating dispersal of large numbers of individuals. The algorithm is stochastic and can be applied using any bivariate probability density function as the dispersal kernel. It achieves computational efficiency while still allowing the simulation of rare and important long-distance dispersals by combining different approaches for within and outside the tail of the dispersal kernel. The tail is specified by a given bivariate quantile, where the q-th bivariate quantile is defined to be the contour of equiprobability within which a proportion $0 < q < 1$ of dispersing individuals will settle.

We provide a method for finding bivariate quantiles that can be applied to any bivariate dispersal kernel derived from independent densities for distance and direction of dispersal. To illustrate this approach, we show how the Cauchy distribution can be used to produce isotropic and anisotropic bivariate dispersal kernels by assuming that the direction of dispersal is either random or takes a von Mises distribution.

We show that the algorithm is considerably faster than generating individual random samples from a bivariate dispersal kernel. It also performs better for larger grid sizes, and when there are larger numbers of individuals to be spread, than an approach that generates samples from a Binomial distribution for each grid cell using the probability of dispersal to that cell. The degree of computational efficiency achieved by the algorithm compared to the Binomial approach depends upon the speed with which random samples can be generated from the tail of the bivariate dispersal kernel used.

Keywords: fast dispersal algorithm, simulation, spread, spatial

1 INTRODUCTION

Deterministic models, such as diffusion models (Fisher, 1937; Skellam, 1951) can under-estimate simulation of spread because they do not allow rare long-distance dispersals (Hastings et al., 2005). Stochastic models can simulate rare long-distance dispersals by moving each individual according to a random sample drawn from a probability density function (PDF) describing the position of dispersed individuals relative to the origin of dispersal, called the dispersal kernel (Mollison, 1977). However, a stochastic spread simulation can require the spread of millions of individuals from each cell in a grid-based model, and these spread events recur regularly during the simulation, so models can be slow to run.

Using a bivariate dispersal kernel, faster spread modelling is commonly achieved by integrating the PDF to calculate the probability of dispersal to each cell in the grid from the originating cell, and then multiplying that probability by the total number N of individuals to be spread (e.g., Colbach and Sache, 2001). The disadvantage is that individuals can not spread past a certain distance, precluding simulation of rare but important long-distance events. Rare events can be simulated if rounding is avoided and instead the number of individuals dispersing to each cell is simulated as a random sample from a Binomial distribution $P(N, p)$, where p is the probability of dispersal to that cell. However, this approach can be slow for large grid sizes.

We present a fast algorithm for generating large numbers of dispersals over a spatial grid. It achieves computational efficiency while still allowing the simulation of rare long-distance dispersals by combining integration of the density over grid cells with generation of individual samples in the tail of the distribution, where the tail of the distribution is specified using a given bivariate quantile. We define the term *bivariate quantile* similarly to the usual use of the term quantile in univariate statistics, so that the q-th bivariate quantile defines the contour of equiprobability within which a proportion q of dispersing individuals will settle. The algorithm can be used with any bivariate PDFs derived from independent distributions for the distance and location of spread. We show how to derive such bivariate distributions, where the dispersal direction can be fully or only partially random, and how to locate their bivariate quantiles. We compare the speed of our algorithm to generating individual random samples and the Binomial approach described above and show when it is faster than its alternatives.

2 DERIVATION OF BIVARIATE PROBABILITY DENSITY FUNCTIONS FOR SIMULATING SPREAD

This section applies transformations between Cartesian and polar coordinates to derive bivariate PDFs that can be used as dispersal kernels. Without loss of generality, we assume that the origin of dispersal is the origin of the (x, y) space. If an individual disperses from the origin to a point (x, y) in the plane, we can determine the radial distance it has travelled $r = \sqrt{x^2 + y^2}$, and the direction it has travelled $\theta = \arctan(y/x)$. Given a bivariate probability density function $p(x, y)$, $p(r, \theta)$ is obtained by the standard change of variables approach (Stirzaker, 1999). Using the inverse transformations $x = r \cos \theta$ and $y = r \sin \theta$, the determinant of the Jacobian matrix $J_{r,\theta} = r$, then

$$p(r, \theta) = r p(x, y). \tag{1}$$

Leptokurtic distributions are better for predicting dispersal (Bullock and Clarke, 2000; Klein et al., 2006; Kot et al., 1996; Nathan and Muller-Landau, 2000). An example is the fat-tailed Cauchy distribution, which has been used to simulate dispersal of plant disease (Diggle et al., 2002; Xu and Ridout, 1988) and invasive species (Pitt et al., 2009). We therefore use the Cauchy distribution to demonstrate the derivation of isotropic and anisotropic bivariate dispersal kernels and their bivariate quantiles.

2.1 Isotropic Cauchy distribution

The PDF for the bivariate isotropic Cauchy distribution is given by

$$p(x, y | b) = \frac{b}{2\pi (x^2 + y^2 + b^2)^{3/2}}, \tag{2}$$

where b is the scale parameter (Achim and Kuruoglu, 2005). Any radial slice of the bivariate isotropic Cauchy distribution with scale b gives a univariate Cauchy distribution with mean 0 and scale b. Applying

(a) Probability density function.

(b) Bivariate quantiles at $q = 0.1$, 0.5 and 0.9.
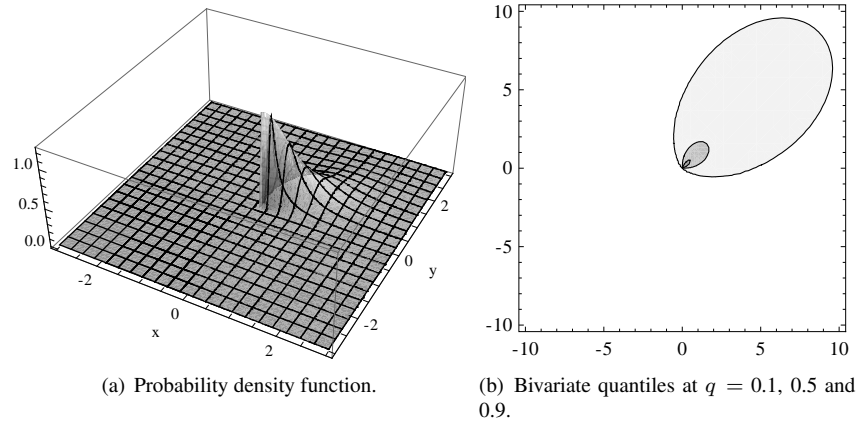
Figure 1. The Cauchy von Mises product distribution with $b = 1$, $\mu = \pi/4$ and $\kappa = 10$.

Equation 1 gives

$$p(r, \theta | b) = \frac{rb}{2\pi \left(x^2 + y^2 + b^2\right)^{3/2}}. \tag{3}$$

Because r and θ are assumed independent, $p(r, \theta) = p(r)p(\theta)$, and $p(\theta)$ is uniform over $[0, 2\pi]$ since the distribution is isotropic. Thus, the PDF for dispersal distance is given by

$$p(r|b) = \frac{br}{\left(b^2 + r^2\right)^{3/2}}. \tag{4}$$

The corresponding PDF for dispersal distance derived from a bivariate normal distribution is the Rayleigh density (Reza, 1994, pp. 214-215).

For dispersal kernels, we require that the cumulative density function (CDF) $P(x, y | b)$ increases with distance from the origin. It is found by first integrating over θ and then over r as follows:

$$
\begin{aligned}
P(x, y | b) &= \left. \int_0^{2\pi} \int_0^{r'} p(r|b) \, p(\theta) \, dr d\theta \right|_{r' = \sqrt{x^2 + y^2}} \\
&= \int_0^{2\pi} \frac{1}{2\pi} \int_0^{\sqrt{x^2 + y^2}} p(r|b) \, dr \\
&= \frac{b}{\sqrt{b^2 + x^2 + y^2}}, \tag{5}
\end{aligned}
$$

Having derived the cumulative density, the q-th bivariate quantile, $0 < q < 1$, can be found by locating all points (x, y) such that $P(x, y | b) = q$.

2.2 Cauchy von Mises product distribution

We now construct a fat-tailed anisotropic distribution where dispersal distance follows the density $p(r|b)$ in Equation 4, but the angle of dispersal follows a von Mises distribution. The von Mises distribution is continuous on $[0, 2\pi]$, centered on a mean μ and with a second parameter κ that determines how concentrated the distribution is around the mean. The von Mises PDF is given by

$$p(\theta | \mu, \kappa) = \frac{e^{\kappa \cos(\theta - \mu)}}{2\pi I_0(\kappa)}, \tag{6}$$

where $I_0(\kappa)$ is the modified Bessel function of order 0. The bivariate PDF is found by taking

$$
\begin{aligned}
p(x, y|b, \mu, \kappa) &= \frac{1}{r} p(r, \theta|b, \mu, \kappa) \\
&= \frac{1}{r} p(r|b) p(\theta|\mu, \kappa) \\
&= \frac{br \left(b^2 + r^2\right)^{3/2}}{r} \frac{e^{\kappa \cos(\theta - \mu)}}{2\pi I_0(\kappa)}.
\end{aligned}
\tag{7}
$$

Then, substituting $x = r\cos\theta$ and $y = r\sin\theta$ gives:

$$
p(x, y|b, \mu, \kappa) = \frac{b e^{\kappa \cos(\mu - \arctan(x/y))}}{2\pi \left(b^2 + x^2 + y^2\right)^{3/2} I_0(\kappa)}.
\tag{8}
$$

The shape of the density is shown in Figure 1(a). Because the distribution is anisotropic, the bivariate quantiles are not circular. For the isotropic Cauchy dispersal kernel, the double integral in the calculation of the cumulative density function could be performed analytically. This is no longer the case when the direction of dispersal follows a von Mises distribution. However, we can still calculate the value of the CDF at any point (x, y) and thus determine the locations of the bivariate quantiles for values $0 < q < 1$. The CDF can be formulated using the change of variables approach as

$$
P(x, y|b) = \int \int_0^{r'} p(r|b)\, p(\theta|\mu, \kappa)\, drd\theta \Bigg|_{r' = \sqrt{x^2 + y^2}}
\tag{9}
$$

We now determine the angles over which to integrate. The point (x, y) corresponds to a contour of equiprobability such that $p(x, y|b, \mu, \kappa) = d$. For each r such that $0 < r < r'$, the contour of equiprobability will intersect the circle defined by r in two places, corresponding to the minimum and maximum angles over which we integrate the angular density. We can change the order of integration in Equation 9 so that

$$
P(x, y|b, \mu, \kappa) = \int_0^{r'} \int_{\theta_{\min}}^{\theta_{\max}} p(r|b)\, p(\theta|\mu, \kappa)\, d\theta dr \Bigg|_{r' = \sqrt{x^2 + y^2}}
\tag{10}
$$

where $\theta_{\min}$ and $\theta_{\max}$ must be found for each r such that $0 < r < r'$.

For a given r, the transformation between Cartesian and polar coordinates provides that

$$
\frac{p(r, \theta|b, \mu, \kappa)}{r} = p(x, y|b, \mu, \kappa) = d.
\tag{11}
$$

Since we know d, we can solve for θ to get $\theta_{\min} = \mu - \alpha$, and $\theta_{\max} = \mu + \alpha$, where

$$
\alpha = \arccos \left[\log \left(\frac{2d\pi \left(b^2 + r^2\right)^{3/2} \text{BesselI}[0, \kappa]}{b} \right) / \kappa \right].
\tag{12}
$$

Thus, $P(x, y|b, \mu, \kappa)$ can be calculated by substituting $\theta_{\min}$ and $\theta_{\max}$ into Equation 10. Because the integration cannot be performed analytically, numerical integration techniques must be employed to calculate the cumulative density. Because the density is more easily calculated than the cumulative density, the $q-th$ bivariate quantile can most easily be found by finding the point (x, y) on the quantile with angle equal to the mean angle μ and the corresponding value of the density $p(x, y)$ can then be used to find all points with equal density. This equivalently finds the q-th bivariate quantile, but reduces the number of times that numerical integration need be performed.

The methods presented in this Section have been applied using *Mathematica* Version 6.0.2. The code is available from the author.
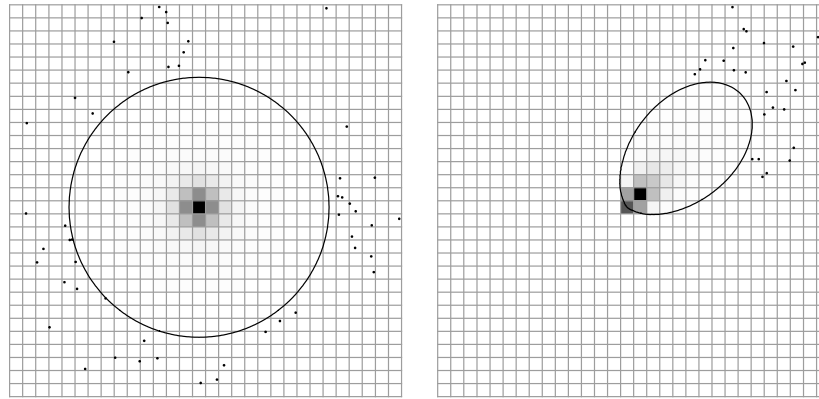
(a) Isotropic Cauchy dispersal kernel, with $b = 1$. (b) Cauchy von Mises product kernel, with $b = 1, \mu = \pi/4$ and $\kappa = 10$.

Figure 2. Illustration of the efficient algorithm for spread simulation with the dispersal origin at the centre of the centre cell. The *0.9-th* quantile is shown. Within it, counts of individuals dispersing to each cell are shown in greyscale with values increasing from white to black. Beyond the *0.9-th* quantile, individual random samples are generated and sshown as black points.

3 AN ALGORITHM FOR EFFICIENT STOCHASTIC SIMULATION OF SPATIAL SPREAD

The algorithm is as follows. For any given simulation, first

1. Choose a value $0 \leq q \leq 1$ and find the q-th bivariate quantile.
2. For grid cells fully contained within or intersecting the q-th bivariate quantile, integrate the PDF over the cell to calculate the probability that an individual will spread to that cell. This need be done only once for any given q.

Then, each time the simulation requires the dispersal of N individuals,

3. Multiply the probabilities from Step 2 by N to get predicted counts for each cell within the q-th quantile.
4. Calculate the total number M of individuals spread to cells within the q-th quantile.
5. Generate $M - N$ individual samples from the tail of the distribution.
6. For each grid cell completely outside the q-th bivariate quantile, count the number of samples that lie in that cell.

Figure 2 illustrates the efficient spread simulation algorithm using two dispersal kernels: the isotropic Cauchy with $b = 1$ and the Cauchy von Mises product with $b = 1, \mu = \pi/4$ and $\kappa = 10$.

4 ALGORITHMIC EFFICIENCY

We compare the speed of our efficient algorithm with two approaches that also allow the simulation of rare events: (1) generation of N random samples from a bivariate dispersal kernel and (2) integration of the PDF to get the probability P_{ij} of dispersal to each cell in the grid and generation of a random sample from a Binomial(N, P_{ij}) distribution to simulate the number of individuals dispersing to the cell. We use two dispersal kernels: an isotropic Cauchy distribution with $b = 1$ and a Cauchy von Mises product distribution with $b = 1, \mu = \pi/4$ and $\kappa = 10$; and three different grid sizes: 10×10, 100×100 and 1000×1000. Each grid cell has dimension 1×1. The tests were conducted on an IBM Thinkpad Z60m with 2MB of RAM.

When using the Binomial approach described in Section 1, the probabilities P_{ij} need only be performed once for any simulation. As shown in Table 4, the time required for this increases linearly with the number of cells in the grid. Similarly, for our efficient algorithm, Steps 1 and 2 need only be performed once. However, the time taken to perform these steps remains constant for different grid sizes.

|	Binomial by grid size			Efficient algorithm
	10×10	100×100	1000×1000	
Isotropic Cauchy	1.016	97.359	9732.766	38.36
Cauchy von Mises	12.750	246.968	23674.953	4532.766

Table 1. Comparison of performance times for the"one off" part of a simulation.

Figure 3 shows, for three grid sizes, the times required by each algorithm to simulate dispersal events of different sizes using an isotropic Cauchy dispersal kernel. The time required to generate individual random samples increases linearly with the number N of individuals to be dispersed, and is not affected by the grid size. The same is true for the efficient algorithm. However, the time taken to simulate dispersal by the Binomial approach increases as the size of the grid increases. While this approach is faster for small grid sizes and $N \leq 10^4$, it is very slow compared to either generating individual random samples or the efficient algorithm. For larger N, the normal approximation to the Binomial distribution can be used and this approach becomes faster than generating individual random samples but it is sill slower than the efficient algorithm.

Figure 4 shows the times required by each algorithm to simulate dispersal using a Cauchy von Mises product kernel. The efficient algorithm is faster than generating individual random samples, but is slower than the Binomial approach for smaller grid sizes; becoming faster when the grid size is 1000×1000. The efficient algorithm is slower when using the Cauchy von Mises product kernel because random samples from a von Mises distribution are generated using an accept-reject method (Best and Fisher, 1979). For dispersal kernels that allow the use of the inverse CDF method for generating random samples, such as the isotropic Cauchy, our algorithm will be more efficient than the Binomial approach for moderate to large grid sizes, with speeds comparable to those reported for the isotropic Cauchy dispersal kernel.

(a) 10×10 grid. (b) 100×100 grid. (c) 1000×1000 grid.
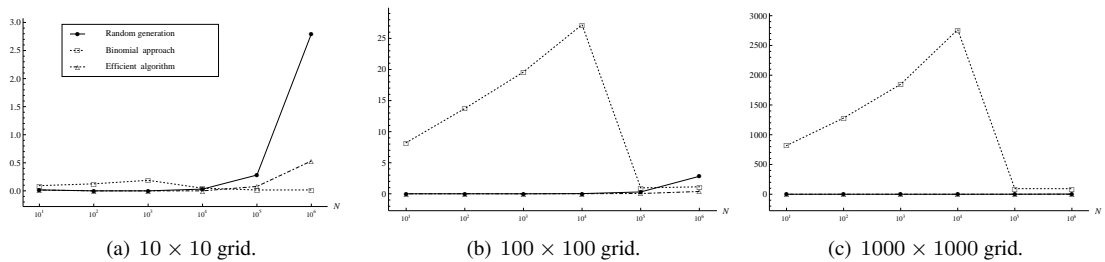
Figure 3. Time taken to simulate a dispersal of N individuals using an Isotropic Cauchy dispersal kernel over different grid sizes.

(a) 10×10 grid. (b) 100×100 grid. (c) 1000×1000 grid.
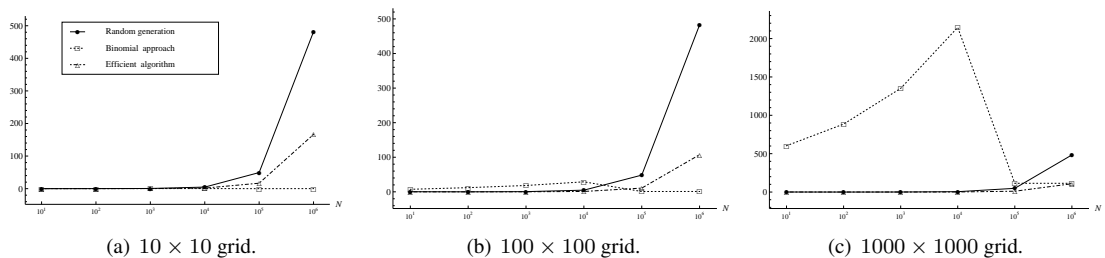
Figure 4. Time taken to simulate a dispersal of N individuals using a Cauchy von Mises dispersal kernel over different grid sizes.

5 CONCLUSIONS

We presented a stochastic algorithm for simulating spread over a two-dimensional grid that achieves computational efficiency while allowing the the simulation of rare events. It is considerably faster than generating individual random samples from a bivariate dispersal kernel. It also performs better for larger grid sizes, and when there are larger numbers of individuals to be spread, than an approach that generates samples from a Binomial distribution for each grid cell using the probability of dispersal to that cell. The degree of computational efficiency achieved by the algorithm compared to the Binomial approach depends upon the speed with which random samples can be generated from the tail of the bivariate dispersal kernel used. Our algorithm is therefore faster using the isotropic Cauchy distribution, for which random samples can be generated from the tail of the distribution using the inverse CDF method, than for the Cauchy von Mises distribution, which requires the adoption of an accept-reject method for generating samples from its tail.

REFERENCES

Achim, A. and E. E. Kuruoglu (2005). Image denoising using bivariate alpha-stable distributions in the complex wavelet domain. *IEEE Signal Procesing Letters 12*(1), 17–20.

Best, D. J. and N. I. Fisher (1979). Efficient simulation of the von Mises distribution. *Applied Statistics 28*(2), 152–157.

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.

Colbach, N. and I. Sache (2001). Blackgrass *(Alopecurus myosuroides Huds.)* seed dispersal from a single plant and its consequences on weed infestation. *Ecological Modelling 139*(2-3), 201–109.

Diggle, A. J., M. U. Salam, G. J. Thomas, H. A. Yang, M. O'Connell, and M. W. Sweetingham (2002). AnthracnoseTracer: a spatiotemporal model for simulating the spread of anthracnose in a lupin field. *Phytopathology 92*(10), 1110–1121.

Fisher, R. A. (1937). The wave of advance of advantageous genes. *Annals of Eugenics 37*, 355–369.

Hastings, A., K. Cuddington, K. F. Davies, C. J. Dugaw, S. Elmendorf, A. Freestone, S. Harrison, M. Holland, J. Lambrinos, U. Malvadkar, B. A. Melbourne, K. Moore, C. Taylor, and D. Thomson (2005). The spatial spread of invasions: New developments in theory and evidence. *Ecology Letters 8*(1), 91–101.

Klein, E. K., C. Lavigne, P. H, M. Renard, and P. Gouyon (2006). Pollen dispersal of oilseed rape: Estimation of the dispersal function and effects of field dimension. *Journal of Applied Ecology 43*, 141–151.

Kot, M., M. A. Lewis, and P. van den Driessche (1996). Dispersal data and the spread of invading organisms. *Ecology 77*(7), 2027–2042.

Mollison, D. (1977). Spatial contact models for ecological and epidemic spread. *Journal of the Royal Statistical Society. Series B 39*(3), 283–326.

Nathan, R. and H. C. Muller-Landau (2000). Spatial patterns of seed dispersal, their determinants and consequences for recruitment. *TREE 15*(7), 278–285.

Pitt, J. P. W., S. P. Worner, and A. V. Suarez (2009). Predicting Argentine ant spread over the heterogenous landcsape using a spatially explicit stochastic model. *Ecological Applications 19*(5), 1176–1186.

Reza, F. M. (1994). *An Introduction to Information Theory*. Dover Publications.

Skellam, J. G. (1951). Random dispersal in theoretical populations. *Biometrika 38*, 196–218.

Stirzaker, D. (1999). *Probability and Random Variables: A Beginner's Guide*. Cambridge University Press.

Xu, X. and M. S. Ridout (1988). Effects of initial conditions, sporulation rate, and spore dispersal gradient on the spatio-temporal dynamics of plant disease epidemics. *Phytopathology 88*, 1000–1012.