

# Estudo função de dispersão na simulação neutra coalescente

Mori, Danilo Pereira; Coutinho, Renato Mendes; Prado, Paulo Inácio K

3 de abril de 2018

## Introdução

Os modelos neutros de campo médio apresentam dois parâmetros para descrever a distribuição de abundância de espécies (SAD): theta e m (Hubbell 2001, Etienne 2005). Já o modelo neutro de espaço explícito coalescente depende de dois parâmetros: U (taxa de especiação na paisagem) e d (distância média de dispersão dos indivíduos na metacomunidade) (Rosindell et al 2008).

Theta é o chamado número fundamental da biodiversidade, número adimensional que pode ser descrita como a relação entre o número de indivíduos( $J_M$ ) e da taxa de especiação da metacomunidade (U):  $\theta = 2J_M U$  (Hubbell 2001). Assim, se  $\theta = 50$ , então, “o produto de duas vezes o tamanho da metacomunidade e a taxa de especiação (novas espécies por evento de nascimento na metacomunidade) é igual 50 novas espécies por geração” (Ricklefs 2012). Considerando a metacomunidade como uma paisagem finita composta de habitat e não habitat e saturada (todas as unidades de habitat estão colonizadas antes do evento de morte - *zero-sum assumption*), podemos escrever  $\theta$  como:

$$eqn1 : \theta = 2ApDA_{media}U$$

Onde  $DA_{media}$  = densidade media de indivíduos na paisagem (ind./ha), p = porcentagem de cobertura vegetal da paisagem (# unidades de habitat/ # total de unidades na paisagem [REESCREVER]), A = área total da paisagem (500 ha) e U = taxa de especiação na paisagem. Essa igualdade pressupõem que a densidade de unidades de habitat (ou de não habitat) é constante na paisagem. Considerando  $DA_{media} = DA_{obs}$ , podemos utilizar essa igualdade para obter U a partir do modelo neutro de campo médio (EI) e  $\theta$  a partir do modelo neutro de espaço explícito (EE).

$m$  é a probabilidade de uma morte na ‘comunidade local’ ser sucedida de um evento de imigração (de movimentação sucedida de colonização e amadurecimento até a idade reprodutiva) de um indivíduos da metacomunidade. Assim, se  $m = 0.1$ , então 10% dos indivíduos na amostra (i.e. a ‘comunidade local’) são imigrantes que chegaram ao acaso da paisagem ao redor (adaptado de Ricklefs 2012). Esse parâmetro descreve a limitação à dispersão no nível da amostra, imaginando uma comunidade local que se relaciona por dispersão com um o conjunto de comunidades locais onde todas podem potencialmente trocar indivíduos (a metacomunidade/paisagem) (Etienne et al. 2007). Chisholm & Lichstein (2009) estabeleceram uma aproximação entre  $m$  e  $d$ :

$$m = P_{amostra}d/\pi A_{amostra}$$

Onde  $P_{amostra}$  e  $A_{amostra}$  são respectivamente o perímetro e área da amostra. Essa aproximação independe da forma da amostra ou da distribuição de probabilidade subjacente a função de dispersão (probabilidade de um indivíduos colonizar um sítio como função da distância do sítio e da planta progenitora).

## Função de dispersão na simulação coalescente

Utilizando apenas amostras contíguas, aproximamos-as como um quadrado na simulação coalescente, assim, podemos simplificar a formula acima considerando que  $P_{amostra} = 4l_{amostra}$  e  $A_{amostra} = l_{amostra}^2$ , onde  $l_{amostra}$  = lado da amostra; e como  $A_{amostra} = J/DA_{obs}$ :

$$m = d(4/100\pi\sqrt{J/DA_{obs}})$$

A variável  $d$  está descrita na unidade de espaço utilizada para definir  $DA_{obs}$ . A unidade de espaço utilizada em  $DA_{obs} = \text{individuos}/ha$  é o hectare ( $ha$ ) =  $100^2$  metros, assim, para escrever  $d$  em metros dividimos por 100, uma vez que  $\sqrt{ha} = 100$  metros.

A função de dispersão descreve a probabilidade de um indivíduo colonizar um sítio como função da distância do sítio e da planta progenitora. Em geral essa função utiliza uma distribuição de probabilidade para aproximar a função de dispersão (Rosindell et al. 2008). Na simulação coalescente o espaço não é contínuo, ele é descrito como unidade de habitat discreto - onde apenas um indivíduo pode se estabelecer, que podemos aproximar como um quadrado de lado ( $l_{cel}$ ) =  $100/\sqrt{DA_{obs}}$  (unidade de espaço = metros/ $\sqrt{\text{individuos}}$ ). Assim, ao utilizar distribuições de probabilidade para aproximar a função de dispersão na simulação coalescente temos que converter a escala da distribuição de probabilidade da escala padrão para a escala da simulação.

A distribuição de probabilidade que utilizamos na função de dispersão foi a distribuição Laplace. A distribuição Laplace, que pode ser entendida como uma distribuição exponencial negativa espelhada, é descrita por dois parâmetros:  $\mu$ , parâmetro de localização; e  $b$ , parâmetro escalar, que determina a dispersão estatística da distribuição. Considerando que todos os indivíduos na simulação possuem a mesma função de dispersão e na simulação coalescente voltamos no tempo afim de reconstruir a genealogia da comunidade, podemos inverter a função de dispersão e centra-la no sítio vago afim de estimar a localização do indivíduo que será o progenitor daquele indivíduo que ocupou o sítio vago (Rosindell et al. 2008), assim  $\mu = 0$ .  $b$  é estimado como a média do desvio absoluto de  $\mu$  e portanto pode ser interpretado como a distância média de dispersão ( $d$  da aproximação de Chisholm & Lichstein 2009). A variância da distribuição Laplace pode ser descrita como:  $var = 2b^2$ , logo,  $b = \sqrt{var}/\sqrt{2}$ ; podemos parametrizar a distribuição pelo desvio padrão ( $sd = \sqrt{var}$ ) e temos que:

$$b = sd/\sqrt{2}$$

A unidade de  $b$  é a dimensão espacial padrão, metros. Como  $b = d$  então podemos reescrever a aproximação de Chisholm & Lichstein (2009) em termos do desvio padrão da função de dispersão considerando a distribuição Laplace:

$$\text{eqn2 : } m = 4sd/100\pi\sqrt{2J/DA}$$

O benefício de parametrizar a função de dispersão em termos de desvio padrão é que fica mais fácil comparar funções de dispersão utilizando outras distribuições de probabilidade na simulação coalescente, como a distribuição normal (Rosindell & Etienne 2011) e uniforme (Campos et al. 2013) [REESCREVER].

### Função de dispersão e chuva de propagulos

A chuva de propágulos pode ser entendida como o produto da fecundidade e função de dispersão (Clark et al. 1999). Por conta do pressuposto da equivalência funcional todos os indivíduos produzem o mesmo número de propágulos por unidade de tempo (Hubbell 2001), assim, podemos simular cenários de limitação à dispersão em função da porcentagem de propágulos que permanece até determinada distância da planta progenitora. Dessa maneira, não precisamos definir a dispersão em termos de distância *per se* mas em termos de porcentagem de indivíduos que permanecem na área imediata da planta progenitora. Para isso é necessário estabelecer uma distância padrão da planta progenitora e estimar ou definir a porcentagem de indivíduos que se mantêm até esta distância padrão. Como distância padronizamos  $l_{cel}$ , assim, cada paisagem possui uma distância padrão que depende da densidade observada de indivíduos naquela paisagem. Podemos estimar qual a porcentagem de propágulos até a distância padrão que um determinado  $sd$  gera, partindo de um  $m$  (eqn 1); ou podemos informar *a priori* quais as porcentagens de interesse e estimar o  $sd$  necessário para gerar tais porcentagens. Na simulação coalescente, utilizamos 12 valores de porcentagem para simular os cenários de limitação à dispersão: 99%, seq(95,50,by=-5)% e 25%. Apesar da simulação coalescente ser bem

eficiente e permitir simular paisagens infinitas, funções de dispersão que apresentam dispersão muito elevadas são computacionalmente muito onerosas (Rosindell et al. 2008) e apresentariam pouco realismo biológico (Nathan 2006), logo, não utilizamos porcentagens muito baixas (e.g. 1%).

Para estimar o  $sd$  necessário para gerar uma determinada porcentagem de propágulos até a distância padronizada, desenvolvemos uma função no ambiente de programação R (R language team). A seguir o código utilizado nessa função, note que a função permite utilizar 3 distribuições de probabilidade (uniforme, normal e Laplace), contudo utilizamos apenas a distribuição Laplace.

```
library(rmutil)

#' Kernel quantile
#'
#' Returns the quantile value that accumulates a given percentile of
#' the density of three types of dispersion kernels in a simulation of
#' a spatially explicit neutral dynamics (Rosindell et al. 2008).
#'
#' @param sigma real positive, the dispersion parameter of the kernel. For
#'   uniform it is the total width of the kernel (max - min), and for
#'   gaussian and laplacian kernels sigma is the standard deviation.
#' @param kernel character, either 'normal', 'uniform' or 'laplacian';
#'   the kernel type. See Rosindell et al. (2008) for details.
#' @param p real 0 < p > 1; the accumulated probability from the kernel
#'   center.
#' @param density real positive, the density of individual in the
#'   simulation grid, in individuals/hectare
#' @return kernel quantile, which is the distance in meters from the kernel
#'   center that encompasses a proportion p of total kernel
#'   density. For example, p = 0.5 returns the median dispersal
#'   distance.
#' @param npoints non-negative integer, number of distance points to
#'   simulate.
#' @details This function simulates the sampling of dispersion
#'   distances from the kernel as outlined by Rosindell et al
#'   (2008), which is not the same as a bivariate version of each
#'   distribution (e.g. a bivariate normal, laplacian,
#'   uniform). Rather, X and Y coordinates are sampled independently
#'   and the used to find the dispersal distance. Though an
#'   analytical solution might be feasible, this function uses the
#'   random sample functions of the univariate distributions and
#'   empirical quantile functions and so provides approximate values.
#' @references Rosindell J, Wong Y, Etienne RS, 2008, A coalescence
#'   approach to spatial neutral ecology, Ecological Informatics, Vol: 3,
#'   Pages: 259-271
qkernel <- function(sigma, kernel, p, density = 20852/50, npoints = 1e+05) {
  kernel <- match.arg(kernel, choices = c("normal", "gaussian", "laplace",
                                           "uniform"))
  d_ind_MA <- 100/sqrt(density)
  if (kernel == "laplace") {
    b_laplace <- sigma/sqrt(2)
    X_laplace <- d_ind_MA * round(rlaplace(npoints, s = b_laplace)/d_ind_MA)
    Y_laplace <- d_ind_MA * round(rlaplace(npoints, s = b_laplace)/d_ind_MA)
    dist_laplace <- sqrt(X_laplace^2 + Y_laplace^2)
    result <- quantile(dist_laplace, p)
  }
}
```

```

if (kernel == "normal" | kernel == "gaussian") {
  b_norm <- sigma
  X_norm <- d_ind_MA * round(rnorm(npoints, sd = b_norm)/d_ind_MA)
  Y_norm <- d_ind_MA * round(rnorm(npoints, sd = b_norm)/d_ind_MA)
  dist_norm <- sqrt(X_norm^2 + Y_norm^2)
  result <- quantile(dist_norm, p)
}
if (kernel == "uniform") {
  b_unif <- sigma/2
  X_unif <- d_ind_MA * round(runif(npoints, min = -b_unif, max = b_unif)/d_ind_MA)
  Y_unif <- d_ind_MA * round(runif(npoints, min = -b_unif, max = b_unif)/d_ind_MA)
  dist_unif <- sqrt(X_unif^2 + Y_unif^2)
  result <- quantile(dist_unif, p)
}
return(unname(result))
}

#' Finds kernel dispersal parameter
#'
#' Returns dispersal parameter of three types of kernel that
#' accumulates a proportion of the kernel density up to a given distance.
#'
#' @param kernel character, either 'normal', 'uniform' or 'laplacian';
#'       the kernel type. See Rosindell et al. (2008) for details.
#' @param p real 0 < p > 1; the accumulated probability from the kernel
#'       center.
#' @param distance real positive, the quantile that should accumulate
#'       p (for instance, if p=0.5 distance is the median distance)
#' @param density real positive, the density of individual in the
#'       simulation grid, in individuals/hectare
#' @param npoints non-negative integer, number of distance points to
#'       simulate.
#' @param sigma.min real positive, the minimum value of the dispersal
#'       value to try in the one-dimensional optmisation.
#' @param sigma.max real positive, the maximum value of the dispersal
#'       value to try in the one-dimensional optmisation.
#' @return the output of function uniroot, which is a list. The
#'       element 'root' of the list is the dispersal parameter necessary for the kernel having
#'       dist as the quantile for p. For example, p = 0.5 and dist =10
#'       returns the dispersal parameter such that the median dispersal
#'       distance is 10 meters. The solution is not exact because of
#'       rounding values to cell sizes of the simulation grid.
sigkernel <- function(kernel, p, distance, density = 20852/50, npoints = 1e+05,
  sigma.min = 1, sigma.max = 100) {
  f1 <- function(x) distance - qkernel(x, kernel, p, density, npoints)
  uniroot(f1, lower = sigma.min, upper = sigma.max)
}

```

Nossa função para estimar  $sd$  necessário para gerar determinada porcentagem até a distância padrão consiste de duas funções:  $qkernel$  e  $sigkernel$ .  $sigkernel$  retorna o parâmetro de dispersão necessário para que a função de dispersão tenha a distância padrão como quantil da porcentagem de interesse. Essa função depende da função  $qkernel$ , que por sua vez retorna o valor quantil que acumula determinado percentil de propágulos. A função  $qkernel$  pressupõem o processo de dispersão aleatório como está originalmente implementado na simulação coalescente descrito por Rosindell et al. (2008), que não é o mesmo que a versão bivariada da

distribuição de probabilidade subjacente à função de dispersão; ao invés a dispersão é simulada como duas amostras independentes da distribuição de probabilidade em eixos ortogonais, a distância de dispersão é descrito como o módulo do vetor resultado da amostra nos eixos ortogonais (figura 1, detalhes no código).

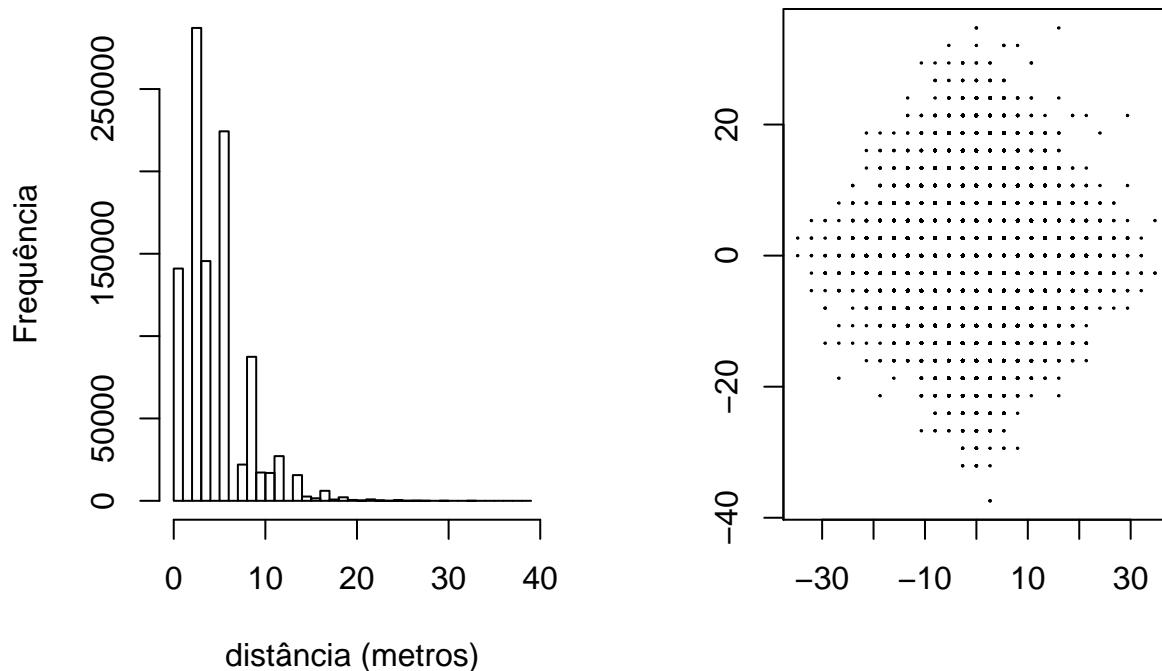


Figura 1. Simulação do processo de dispersão usado na simulação coalescente (Rosindell et al. 2008), consideramos  $DA = 1400$  indivíduos/ha, desvio-padrão da função de dispersão = 4 metros e foram simulados  $1e6$  pontos. No primeiro gráfico temos a distribuição de distâncias à planta progenitora (m). No segundo gráfico temos a distribuição da chuva de propagulos ao redor da planta progenitora que se encontra na origem do gráfico (ponto  $(0,0)$ ). Notem que diferente de uma dispersão que assume que a distribuição de probabilidade é bivariada (que seria mais próximo de um círculo), a maneira como a simulação coalescente (Rosindell et al. 2008) aproxima a dispersão faz com que a forma da chuva de propagulos se aproxime de um losango.

### Estudo de caso: BCI

Os census de Barro Colorado Island tem sido a base de dados padrão para avaliar modelos neutros. Aqui utilizamos a SAD observada em sua amostra de 50 ha para estimar  $\theta$  e  $m$  a partir de um modelo de campo médio (Etienne 2005) e estimamos  $U$  (a taxa de especiação) utilizando o modelo o modelo coalescente dado diferentes cenários de dispersão (Rosindell et al. 2008). Diferente de Rosindell et al. (2008) que considera uma paisagem infinita, vamos considerar um recorte de paisagem de 500 ha (ou  $5km^2$ ) e a dispersão como função da chuva de propágulos que é definida como a porcentagem de propágulos que permanecem até uma determinada distância padrão da planta progenitora. A distância padrão utilizada é a distância entre os organismos na simulação (ou o lado da célula de simulação -  $l_{cel}$ ) =  $100/\sqrt{DA}$  e as porcentagens de interesse são 99%, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 50%, 25%.