

Estudo da relação entre os parâmetros de um modelo neutro de campo médio e um modelo neutro de espaço explícito

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 metacommunidade 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 θ 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 θ 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 simplicar 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 $DA_{obs} = \text{individuos}/\text{ha}$ é o héctare (ha) = 100^2 metros, assim, para escrever d em metros dividimos por 100, uma vez que $\sqrt{\text{ha}} = 100$ 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: μ , 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 μ 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)

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))
}

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 o 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 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).

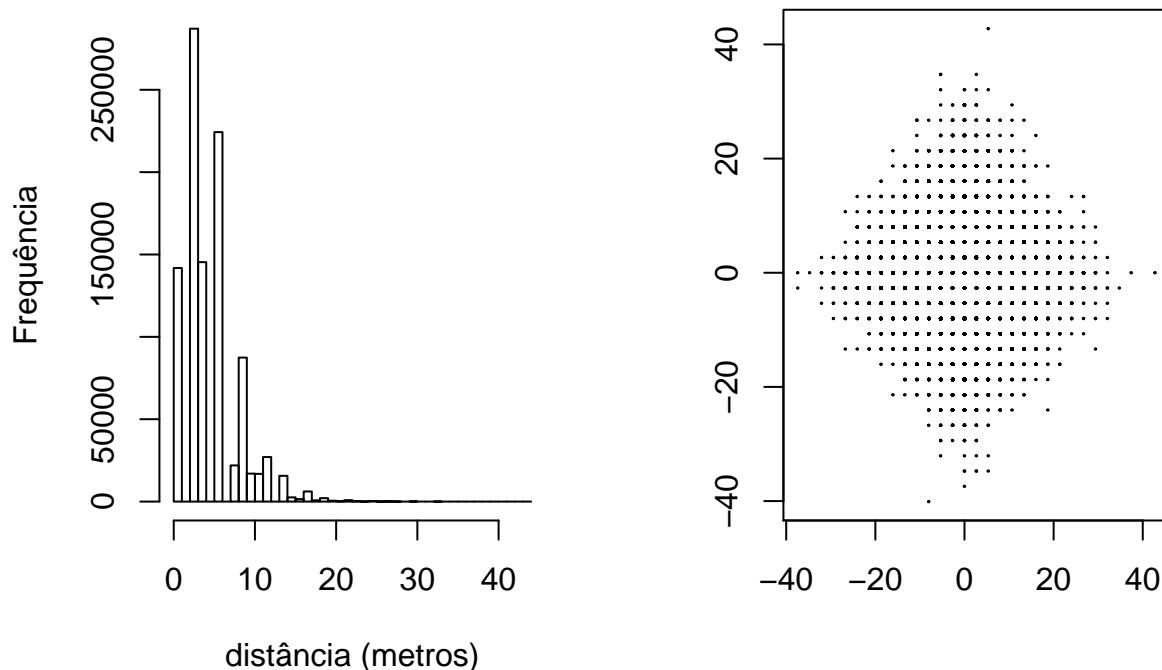


Figura 1. Simulação do processo de dispersão usado na simulação coalescente (Rosindell et al. 2008), consideramos DA = 1400 indivíduos/ha, desvio-padrão da função de dispersão = 4 metros e 1e6 pontos. No primeiro gráfico temos a distribuição de distâncias à planta progenitora (m). No segundo gráfico 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 aproximaria de um círculo), a maneira como a simulação coalesceente (Rosindell et al. 2008) aproxima a dispersão faz com que a forma da chuva de propagulos se aproxime de um losango.

Podemos calcular qual a correspondente chuva de propagulo (%k) de um m conhecido, considerando a densidade observada e o tamanho da amostra. Considerando $m=0.11$ (valor estimado para BCI REFERÊNCIA), $J=20852$ (número de indivíduos em BCI), $A=50\text{ha}$ (área amostrada em BCI) e $DA=417.04$, podemos calcular sd correspondente considerando uma função de dispersão com distribuição de Laplace:

```
J = 20852
DA = 20852/50
(sd = 0.11 * 100 * pi * sqrt((2 * J)/DA)/4) #eqn 2

## [1] 86.3938
```

O sd estimado é bem elevado, acredito que o principal motivo para isso é que a área amostrada é grande e a densidade dos indivíduos é baixa, gerando um baixo valor de indivíduos (J) para uma respectiva área. Assim, a função de dispersão descreve eventos de colonização em unidades de habitat que possuem arrestra grande.

Vamos avaliar qual é a correspondente chuva de propágulos considerando uma distribuição Laplace:

```
library(rmutil)
# distância entre os indivíduos e a distância padrão
d_ind_MA <- 100/sqrt(DA) # tanto a metrica da paisagem quanto a distancia de referencia
b_laplace <- sd/sqrt(2) ## relação entre sigma e b, a variância da laplaciana
## gerando valores de uma distribuição laplaciana e convertendo para a
## distância em metros
X_laplace <- d_ind_MA * round(rlaplace(1e+05, s = b_laplace)/d_ind_MA)
Y_laplace <- d_ind_MA * round(rlaplace(1e+05, s = b_laplace)/d_ind_MA) # idem para Y
dist_laplace <- sqrt(X_laplace^2 + Y_laplace^2) #distâncias dos pontos até a origem
(percentil <- length(dist_laplace[dist_laplace <= d_ind_MA])/length(dist_laplace)) # percentil

## [1] 0.00696

Uma maneira de avaliar se o percentil está corretamente é utilizar esse valor na função quantil e ver se o valor é próximo de 'd_ind_MA'. O quantil correspondente à esse percentil é igual 6.9109844, valor próximo em escala de 4.8967862. Uma segunda avaliação pode ser utilizar o valor do quantil para calcular o percentil

(percentil2 <- length(dist_laplace[dist_laplace <= unname(quantile(dist_laplace,
percentil))])/length(dist_laplace)) # percentil do quantil

## [1] 0.00696
```

O valor de percentil calculado pela distância referência, a distância entre indivíduos, e o quantil correspondente ao percentil possuem o mesmo valor. O erro da simulação deve ocorrer pelo uso da função 'round' na simulação de valores de uma distribuição Laplace em dois eixos ortogonais. Portanto, existe um problema de aproximação da função que provavelmente está relacionado com a necessidade de tornar discreta uma distribuição de probabilidade que é contínua, assim, esse erro independente da DA_{obs} .

Isso quer dizer que apenas 0.696% dos propágulos se mantêm até cerca de 4.8967862 da planta progenitora. Considerando um modelo coalescente, então isso representaria que menos de 1% dos propagulos permanece até l_{cel} da planta progenitora, ou seja, até duas vezes a área da sua copa (assumindo uma árvore que ocupa toda a área da unidade de habitat). Esse é um cenário de ausência de limitação à dispersão no nível da comunidade local. Acredito que esse valor ocorre pois no modelo neutro, o evento de dispersão é resultado do processo de dispersão *per se* e do de colonização seguido do amadurecimento até a idade reprodutiva. Os dados de BCI são de DBH ≥ 10 cm, considerando que existe considerável correlação entre o diâmetro da árvore e sua idade, podemos considerar que estamos amostrando em BCI indivíduos que já existem há muitas gerações. E como a produção de propagulos é muitas vezes mais rápido que o processo de dispersão (*sensu* teoria neutra) podemos considerar que já ocorreram diversos eventos de dispersão *per se* de modo que todas as classes de distância já foram apresentados, permitindo o surgimento de indivíduos em longas distâncias de suas respectivas árvores progenitoras. Esse efeito é forte o suficiente para produzir populações com índices de agregamento que tendem a zero. Uma vez que o modelo de campo médio, desconsidera o efeito da limitação à dispersão no nível de suas escalas (comunidade local e metacommunidade) e considera esse efeito apenas entre escalas (Etienne 2005), podemos considerar razoável que o percentil tenha sido tão baixo, isso mostra coerência interna do modelo.

Uma possível pergunta nesse contexto é qual a equivalente chuva de propágulo para que $m = 1$, ou seja, sem qualquer tipo de limitação à dispersão. Como esse valor provavelmente é muito baixo, vou aumentar o número de pontos utilizado para estimar o percentil:

```
# sd máximo teórico
sd_max <- 100 * pi * sqrt((2 * J/DA))/4
b_laplace <- sd_max/sqrt(2)
# aumentei o número de pontos para maximizar a resolução do percentil
X_laplace <- d_ind_MA * round(rlaplace(1e+06, s = b_laplace)/d_ind_MA)
Y_laplace <- d_ind_MA * round(rlaplace(1e+06, s = b_laplace)/d_ind_MA)
dist_laplace <- sqrt(X_laplace^2 + Y_laplace^2)
```

```
(percentil_max <- length(dist_laplace[dist_laplace <= d_ind_MA])/length(dist_laplace)) # percentil
## [1] 0.000105
```

Essa chuva de propágulos é o esperado para quando não existe limitação à dispersão entre as escalas comunidade local e a metacomunidade.

Alternativamente podemos estimar sd que gera um determinado percentil de propágulos até a distância referência. Essa abordagem é principalmente importante na simulação coalescente, onde não possuímos ferramentário estatístico para estimar os parâmetros do modelo U e a chuva de propagulos (%k) a partir do ajuste da SAD_{obs} por verossimilhança . Na simulação coalescente podemos estimar U a partir da riqueza observada utilizando uma modificação da formula semi-analítica apresenta no apêndice A de Rosindel et al. (2008). A formula estabelece relação entre a probabilidade de colonização de um singleton por evento de morte (ou seja, a probabilidade de uma nova espécie aparecer na paisagem - U) e a riqueza observada partindo de uma árvore genealogica da comunidade quando $U \rightarrow 0$, ou seja, quando há monodominância. No modelo coalescente, reconstruímos a arvore genealógica da comunidade no espaço, assim a função de dispersão altera a arvore genealógica da comunidade e por consequência a relação entre U e a riqueza observada. Dessa maneira, para trabalhar com o modelo neutro espacialmente explícito informamos *a priori* qual a chuva de propágulos de interesse e estimamos sd correspondentes. Utilizamos 12 valores de porcentagem de propágulos que permanecem até l_{cel} metros da planta progenitora: 99%, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 50%, 25%. Vamos calcular os respectivos valores de sd que produzem os percentis de interesse utilizando a função sigkernel:

```
df_ <- data.frame(l_cel = 100/sqrt(DA), percentil = c(99, seq(95, 50, by = -5),
25)/100, kernel_type = rep("laplace", 12), DA_obs = DA, J_obs = J, stringsAsFactors = FALSE)
# essa função é bem mais útil quando existem outros locais para se comparar
f_sapply <- function(i) {
  sigkernel(kernel = df_[i, "kernel_type"], p = df_[i, "percentil"], density = df_[i,
  "DA_obs"], distance = df_[i, "l_cel"])$root
}
df_$sd <- sapply(1:nrow(df_), f_sapply)
```

Com os valores de sd podemos estimar o correspondente 'm' utilizando a eqn2:

```
library(car)
df_ %>% mutate(m = sd * 4/(100 * pi * sqrt((2 * J)/DA)))
scatterplotMatrix(~percentil + sd + m, reg.line = "", smoother = "", data = df_,
main = "")
```

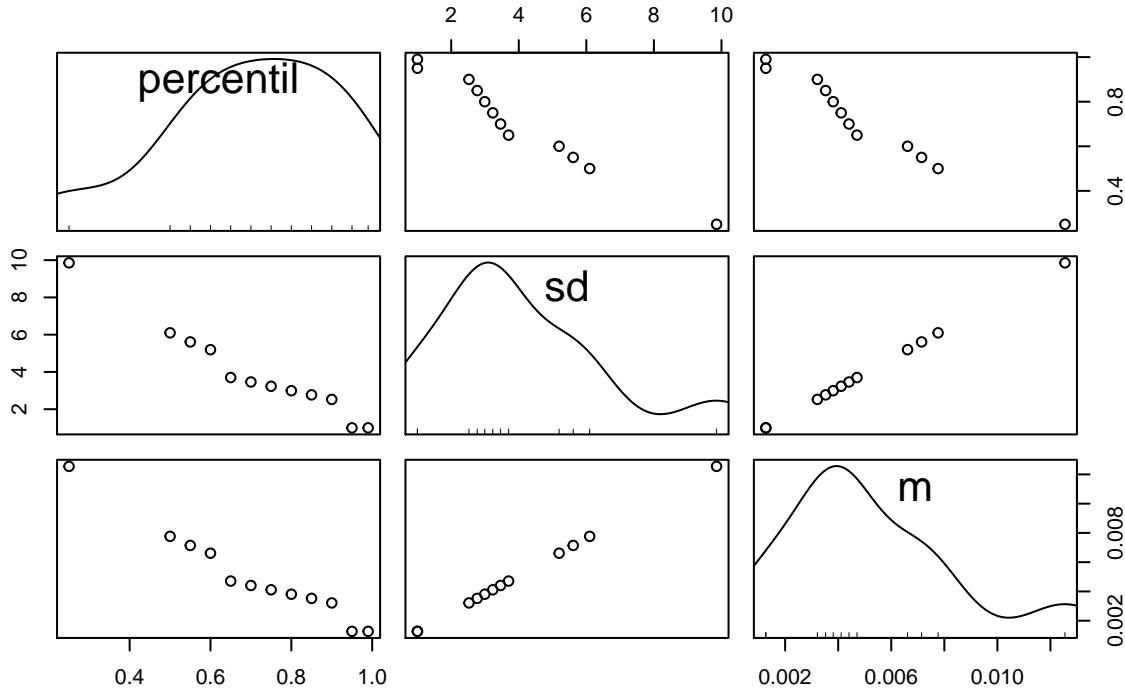


figura 2. Relação entre o percentil de propágulos que permanecem até a distância padrão de $100/\sqrt{DA}$ que é a distância dos indivíduos na simulação coalescente (painel[1,1]), o desvio padrão correspondente sd (painel[2,2]) e a taxa de imigração calculada a partir de sd e da equação 2. A relação entre m e sd é proporcional para valores de J e DA iguais; já a relação do percentil e sd é descrita por uma distribuição Laplace.

Os valores utilizados até aqui são observados em BCI, as características da amostra de BCI difere daquelas presentes no TreeCo em: i) paisagem aproximada como homogênea, ou seja, existem apenas unidades de habitat na paisagem (não há fragmentação); ii) a área amostrada de BCI é cerca de 50 vezes maior que a moda das áreas amostradas em nosso filtro de trabalhos do TreeCo; iii) BCI conta com o censu de árvores com $DBH \geq 10$ cm, os sítios selecionados no TreeCo utilizaram $DBH \geq 5$ cm, espera-se que isso tenha efeito em DA_{obs} . Assim, na próxima sessão vou fazer a conversão de parâmetros do modelo EI e EE das paisagens selecionadas no TreeCo

Convertendo Parâmetros de modelos neutros para as paisagens de TreeCo

Sdt: i) carregar df_resultados; ii) converter os parâmetros segundo sua natureza (EI \leftrightarrow EE); iii) plotar o obtido: