

Cheng, J., Mi, X., Nadrowski, K., Ren, H., Zhang, J. and Ma, K. 2012. Separating the effect of mechanisms shaping species-abundance distributions at multiple scales in a subtropical forest. – *Oikos* 121: 236-244.

## Appendix A1

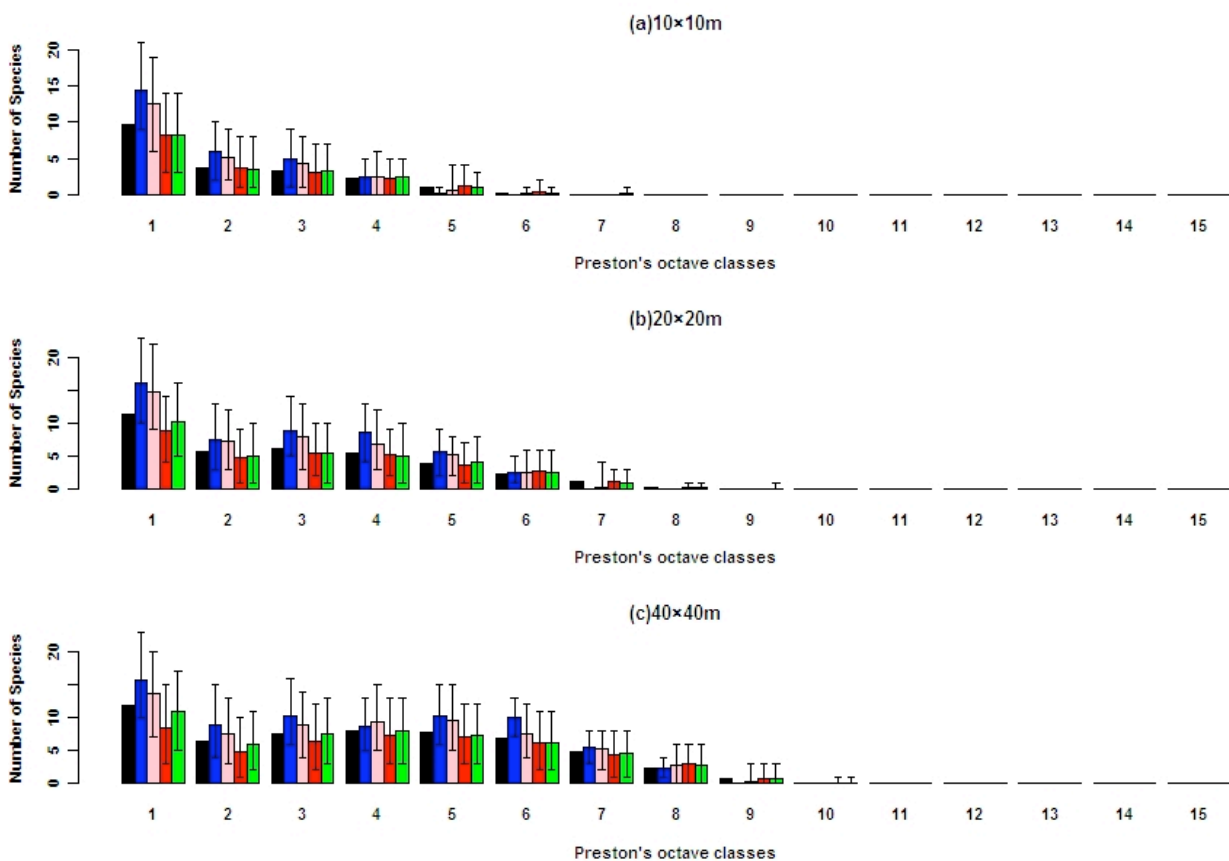
Loadings of each environment factors to the first four principal components

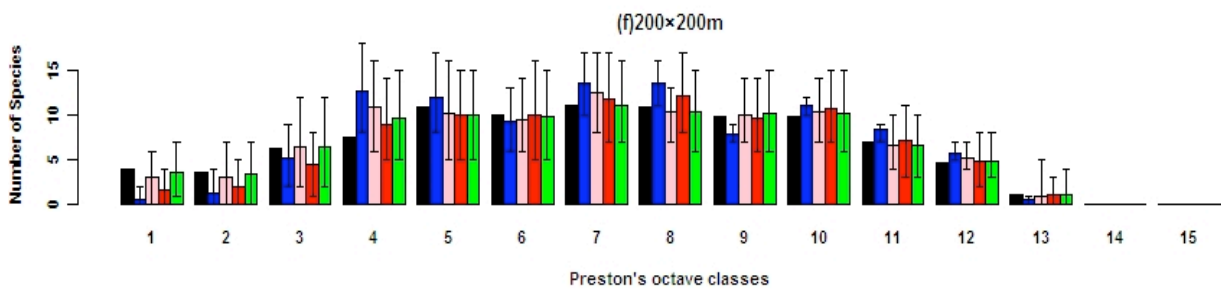
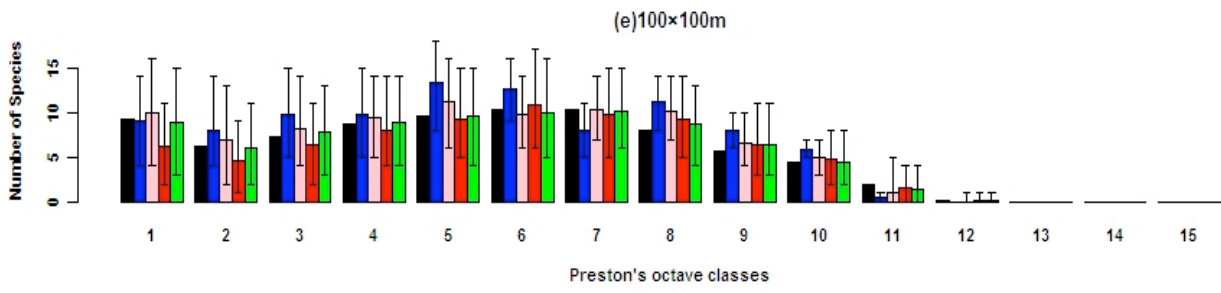
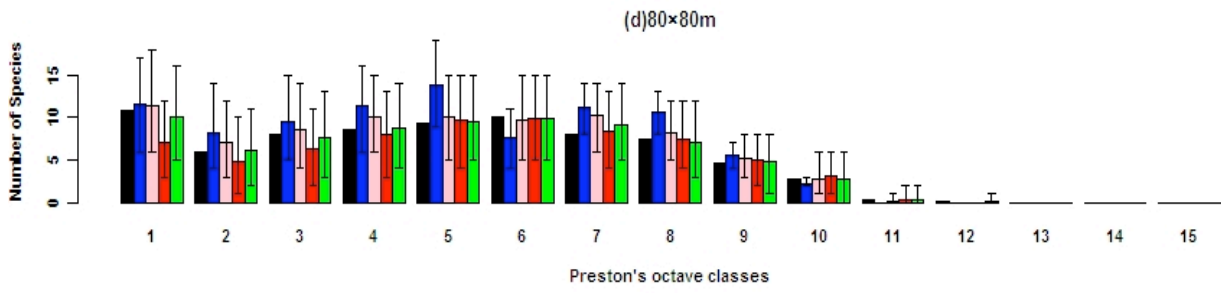
Habitat variables	Principal components 1	Principal components 2	Principal components 3	Principal components 4
Mean elevation	0.9978	0.0624	-0.0048	0.0198
Convexity	0.0326	-0.2204	0.5009	-0.8339
Slope	-0.0157	0.3674	-0.7508	-0.5486
cos(aspect)	0.0008	0.0001	0.0034	-0.0077
sin(aspect)	-0.0002	-0.0034	-0.0074	0.0014
Fe	0.0002	-0.0001	0.0051	-0.0060
Mn	-0.0001	-0.0059	-0.0200	0.0252
Zn	-0.0001	0.0066	0.0000	0.0076
Cu	0.0003	0.0004	-0.0014	0.0039
K	-0.0009	0.0052	-0.0071	0.0097
P	-0.0045	0.0232	-0.0083	0.0254
Ca	-0.0001	0.0093	-0.0124	0.0297
Mg	-0.0005	0.0082	-0.0086	0.0201
Na	0.0012	-0.0004	-0.0051	0.0044
B	0.0005	-0.0008	0.0018	-0.0016
Si	0.0016	-0.0123	-0.0026	-0.0082
Al	0.0000	0.0034	-0.0007	-0.0021
N	-0.0006	0.0042	-0.0032	0.0062
pH	0.0000	0.0008	-0.0001	0.0012
N mineralization rate	-0.0017	0.0114	0.0020	0.0073
Bulk density	0.0545	-0.9006	-0.4294	-0.0169
Soil moisture	-0.0014	0.0161	0.0042	0.0088
Standard deviation	56.5427	7.9814	6.4891	5.8465
Proportion of variance	0.9575	0.0191	0.0126	0.0102
Cumulative proportion	0.9575	0.9766	0.9892	0.9994

## Appendix A2

The bar plot of observed and predicted species–abundance distribution (SAD) at different scales ( $10 \times 10 - 200 \times 200$  m) in Gutian plot. The vertical bars are 95% confidential intervals.

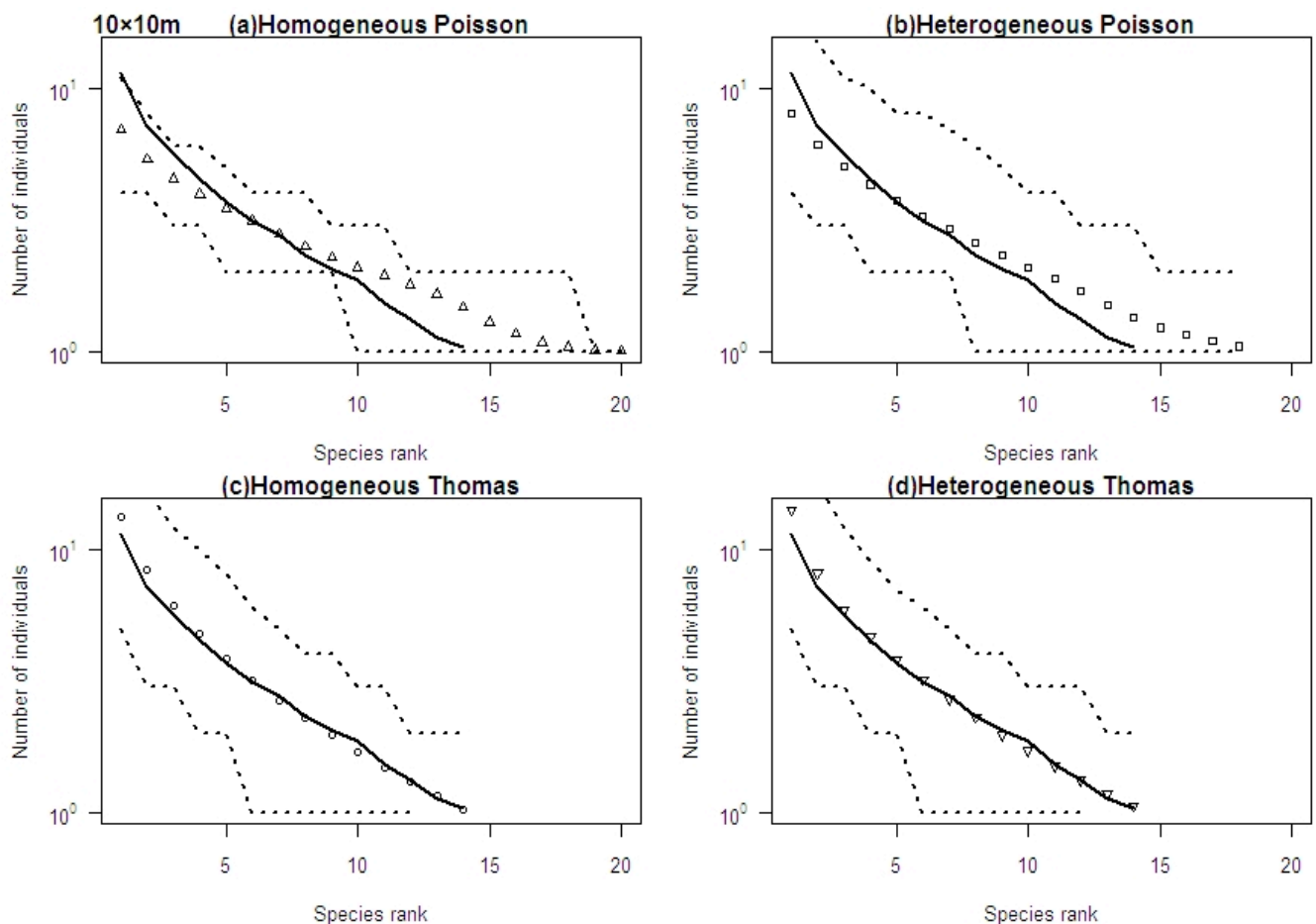
Different colors represent different SADs predicted by different processes: black histogram is the mean value of observed SADs from 100 samples in Gutian plot at corresponding scale, blue histogram by homogeneous Poisson process, pink histogram by heterogeneous Poisson process at corresponding scale, red histogram by homogeneous Thomas process, green histogram by heterogeneous Thomas process at corresponding scale. The SADs are plotted using Preston's binning method. The numbers on the  $x$  axis represent Preston's octave classes, and octave 1 represents 1, 2, 3–4, 5–8, 9–16, and so on (Hubbell 2001).

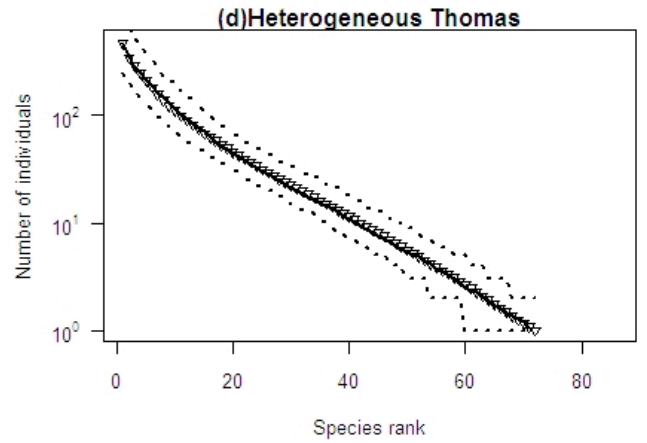
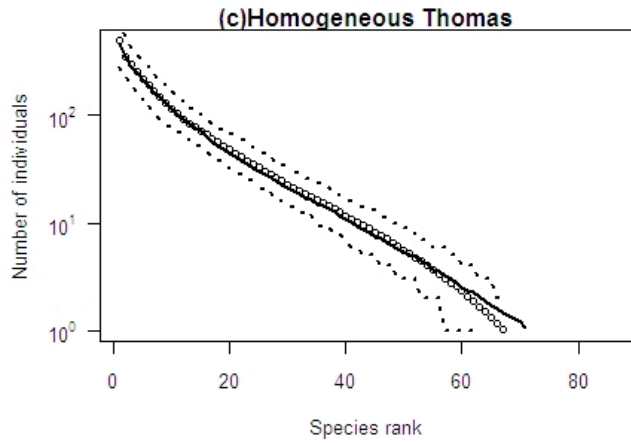
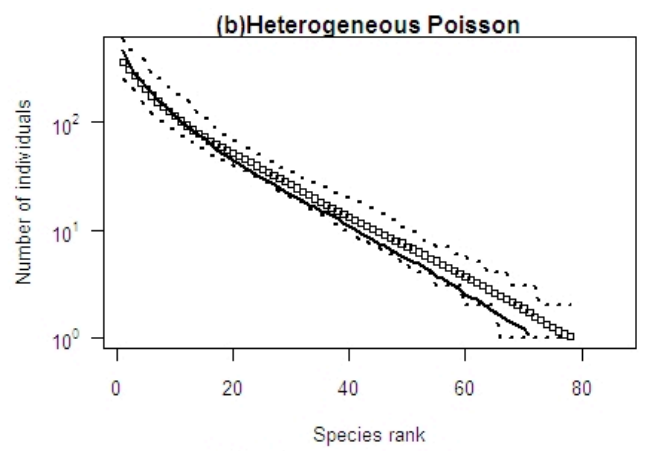
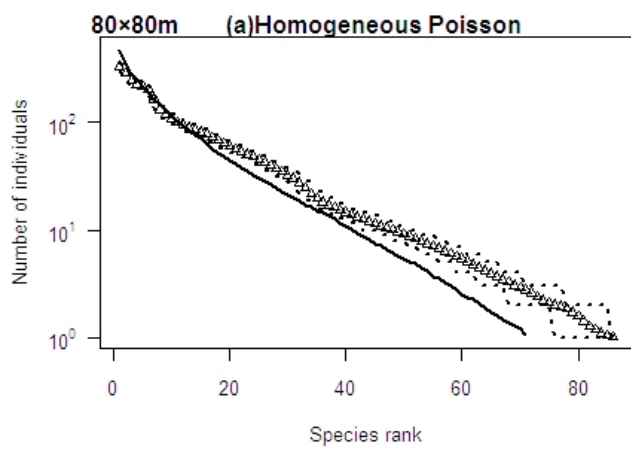
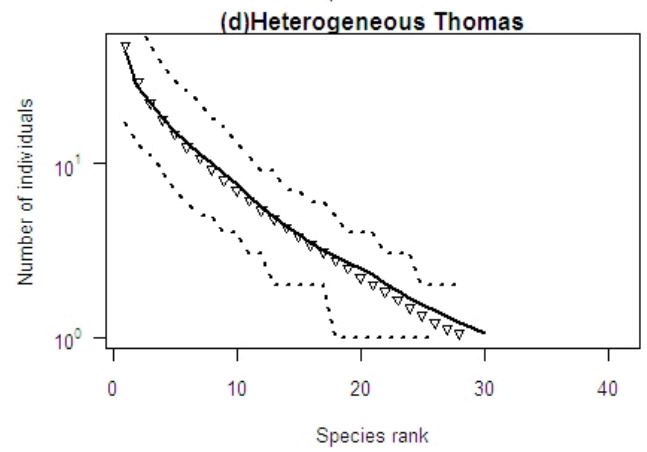
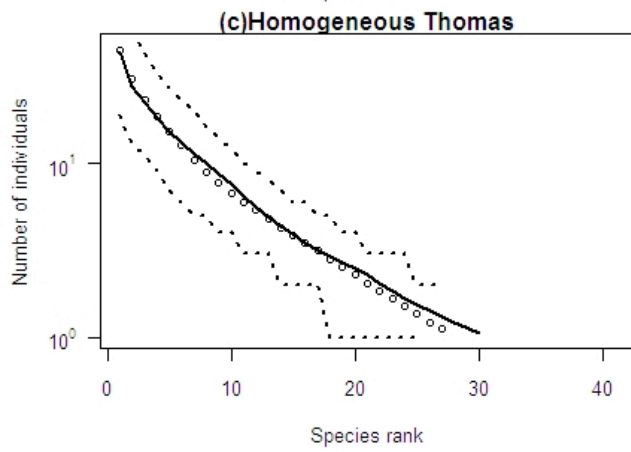
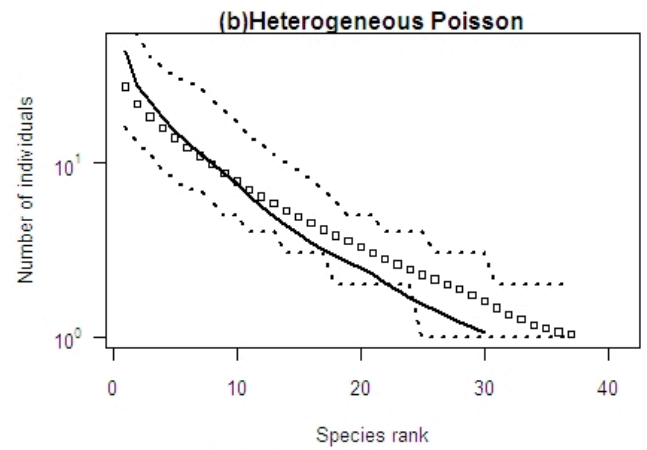
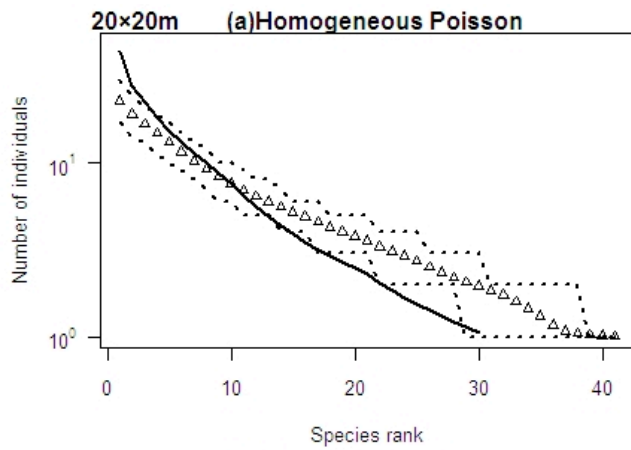


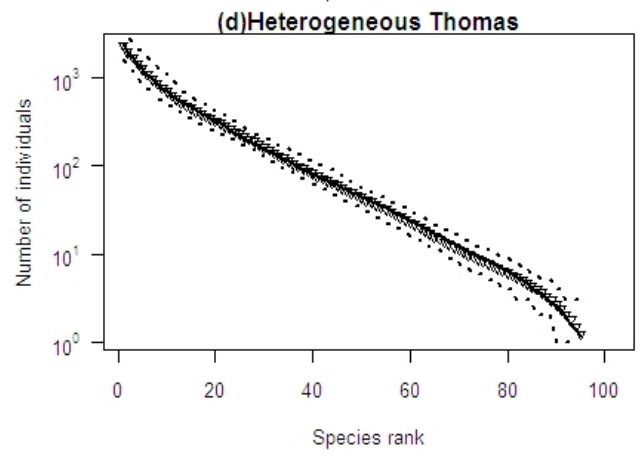
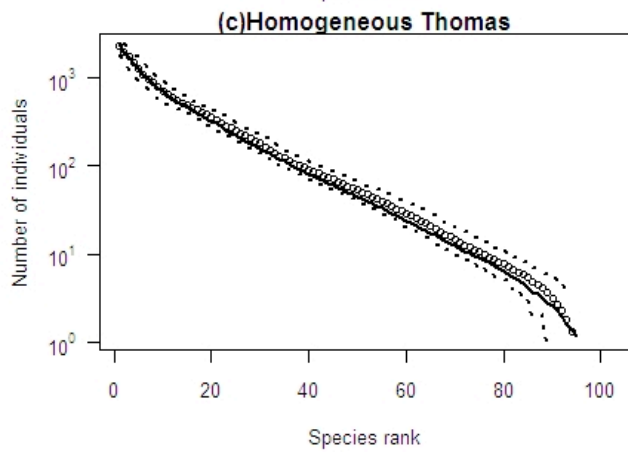
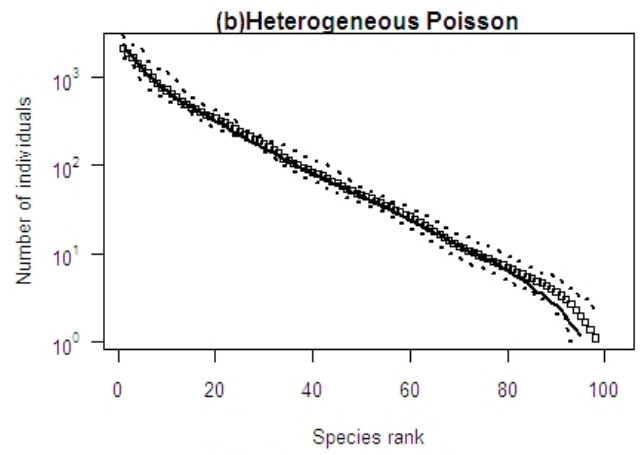
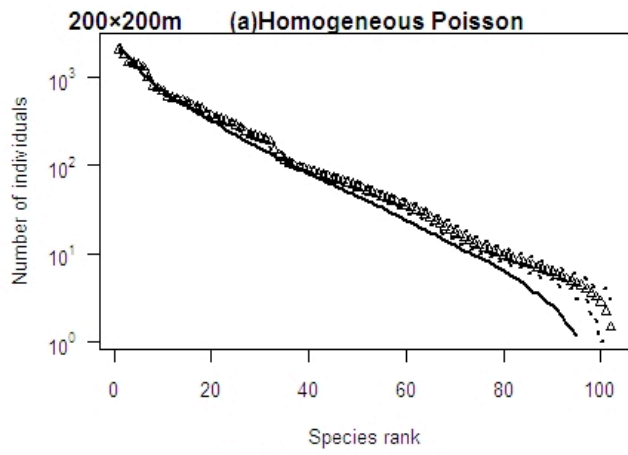


## Appendix A3

95% confidence intervals of the four processes at scales of  $10 \times 10 - 200 \times 200$  m. Solid line represents mean value of observed species abundance distribution (SADs) of 100 samples in Gutian plot at corresponding scale; Points are the mean values of simulated SADs from 1000 (10 of each simulated communities) samples at corresponding scale. Different symbols represent different SADs predicted by different processes:  $\square$ : homogeneous Poisson process,  $\square$ : heterogeneous Poisson process,  $\circ$ : homogeneous Thomas process,  $\square$ : heterogeneous Thomas process.

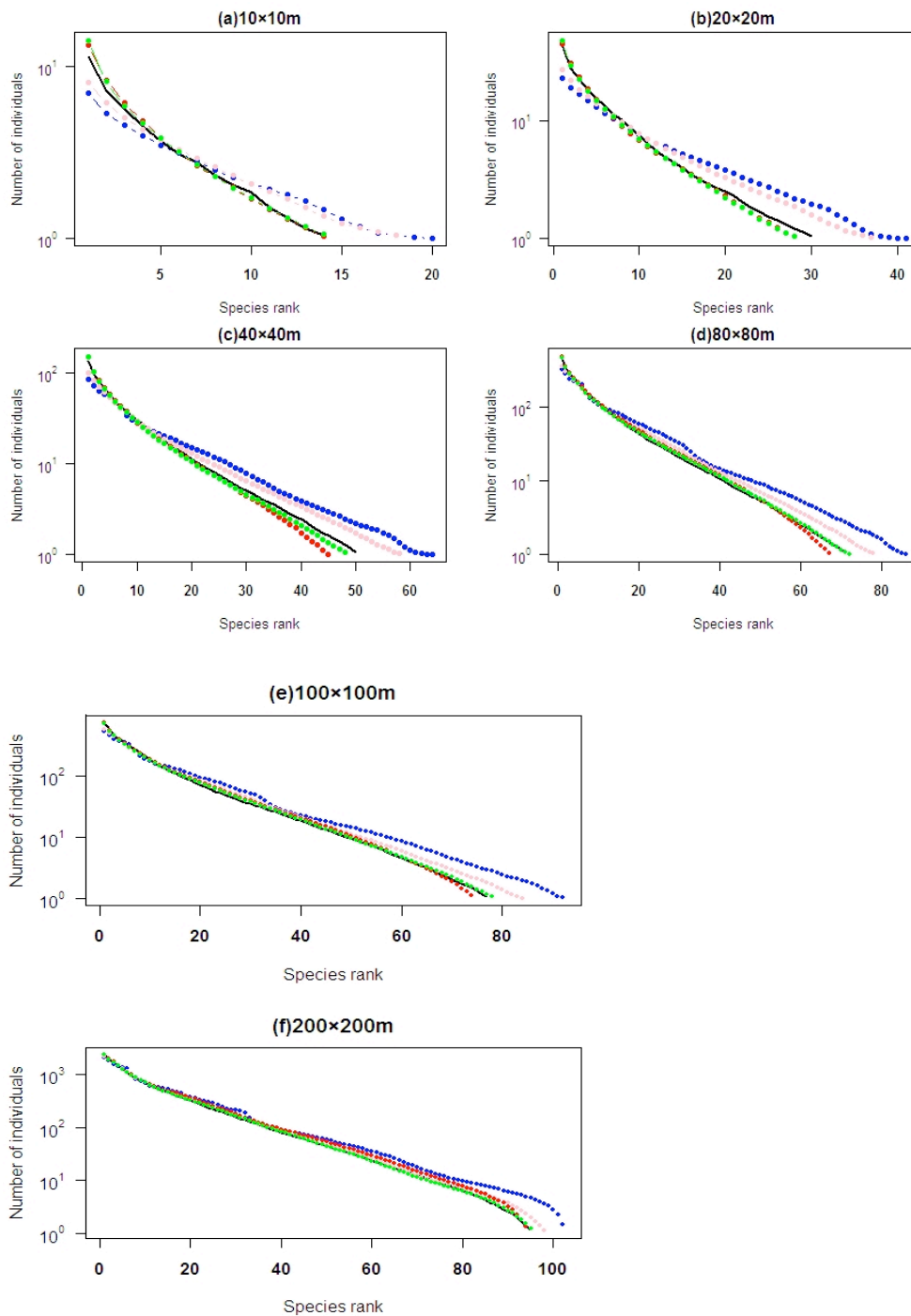






## Appendix A4

Rank–abundance curves simulated from the four models at different scales: Observed curves of species–abundance distribution (SADs) are shown as black line. Blue dotted-lines are the mean value of SADs from 1000 (10 of each simulated communities) samples simulated by homogeneous Poisson process, pink dotted-lines by heterogeneous Poisson process, red dotted lines by homogeneous Thomas process, and green dotted-lines by heterogeneous Thomas process.



## Appendix A5

Algorithms for the four spatial point models and model parameterization used in inference mechanisms structuring species abundance distributions (SADs)

Calculations of AIC are also detailed. Møller and Waagepetersen (Møller and Waagepetersen 2004) and Shen et al. (2009) carefully described the algorithms of four models used in this study, and Waagepetersen and Guan (2009) and Waagepetersen (2007) clearly exemplified the steps of model parameterization. Here we outline the framework of four models and steps of parameterization. For more details, please refer to three above-mentioned literatures.

### 1) Description and model parameterization of Poisson process

A Poisson process  $X$  in a two-dimensional region  $S \subset \mathfrak{R}^2$  is characterized by two fundamental properties (1) the number of point  $N$  in any boundary subregion  $B$  has a Poisson distribution with mean intensity measure  $\mu(B)$  ( $\mu > 0$ ); (2) given that  $N(X \cap B) = n$ , the  $n$  points are independent and identically distributed, with intensity function  $\rho(u)$ :

$$\rho(u) = \alpha \exp(Z_{1:k}(u) \beta_{1:k}^T) \quad (1)$$

where  $u \in B$ , and  $\alpha > 0$ ,  $Z_{1:k}(u)$  is the  $1 \times k$  vector of non-constant environmental variables,  $\beta_{1:k}$  is the  $1 \times k$  vector of corresponding regression parameters.

If  $\rho(u)$  is constant for all  $u \in S$ , the Poisson process is homogeneous or stationary, that is, a completely spatial random process. If  $\rho(u)$  is a function of environmental variables  $Z_{1:k}(u)$  on location  $u \in S$ , the Poisson process is heterogeneous. The points of heterogeneous Poisson process have no interaction, but intensity changes with environmental variables. Waagepetersen (2007) suggested to extract the regression parameters  $\beta = (\beta_0, \beta_{1:k})$ , where  $\beta_0 = \log(\alpha)$  using maximum likelihood method based on the following equation to obtain  $\hat{\beta}$ :



$$l(\beta) = \sum_{u \in X \cap S} Z(u) \beta^T - \int_S (\exp(Z(u) \beta^T) du \quad (2)$$

where  $Z(u) = (1, Z_{1:k}(u))$ , and  $l(\beta)$  is the likelihood function.

## (2) Description and model parameterization of Thomas process

In Thomas process  $X$ , the clusters  $X_c$  of the offspring points are symmetrically and normally distributed around the parent points  $c$  in a homogeneous Poisson process of intensity  $\kappa$ . Given  $c$ , the clusters  $X_c$  are independent Poisson process with intensity functions:

$$\rho_c(u) = \exp(Z_{1:k}(u) \beta_{1:k}^T) \alpha k(u - c, \delta) \quad (3)$$

where  $\alpha > 0$ ,  $k(u - c, \delta)$  is a probability density depending on  $\delta > 0$  determining the spread of offspring points around  $c$ . Similar to Eq. 1,  $\exp(Z_{1:k}(u) \beta_{1:k}^T)$  still represents the covariance between point density and environmental variables  $Z_{1:k}(u)$  at location  $u$ .  $X$  is a homogeneous Thomas process when  $\exp(Z_{1:k}(u) \beta_{1:k}^T) = 1$ . Otherwise  $X$  is a heterogeneous Thomas process.

Assume that  $\exp(Z_{1:k}(u) \beta_{1:k}^T)$  is bounded by some constant  $M$ , and a cluster  $X_c$  may then be considered as independent thinning of a cluster  $X_c$  with intensity function  $M \alpha k(\cdot - c, \delta)$  where the spatially varying thinning probability is  $\exp(Z_{1:k}(u) \beta_{1:k}^T) / M$ . Using this thinning perspective, the intensity function of Thomas process  $X$  is:

$$\rho(u) = \alpha \kappa \exp(Z_{1:k}(u) \beta_{1:k}^T) \quad (4)$$

The parameters of heterogeneous Thomas process  $\hat{\beta}$  can be estimated using maximum likelihood method as Eq. 2. Other parameters  $\hat{\kappa}$  and  $\hat{\delta}$  can be estimated by minimum contrast methods:

$$m(\kappa, \delta) = \int_0^r (\hat{K}(t)^{1/4} - K(t; \kappa, \delta)^{1/4})^2 dt$$

where  $r$  is user-specific constants, and  $K$  is heterogeneous K-function of  $X$  which is defined as:

$$\hat{K} = \sum_{u, \eta \in X \cap S} \frac{1[0 < \|u - \eta\| < t]}{\exp((Z(u) - Z(\eta))\hat{\beta}^T)} e_{u, \eta}$$

where  $e_{u, \eta}$  is edge correction. Considering the bias of  $K$  increases with  $r$ , we set  $r = 100$  m. The expected number of offspring points around each parent point can be calculated as:  $\hat{\mu} = M / \hat{K}$ , and  $\hat{\alpha} = \exp(\beta_0) / \hat{K}$  for Thomas process.

## 2. Akaike's information criterion

The Akaike information criterion (AIC) is a way of selecting a model that has a good fit to the truth but few parameters from a set of models. In general case, AIC is defined as:

$$AIC = -2\ln(L) + 2k$$

where likelihood is the maximized value of the likelihood function for the estimated model and  $K$  is the number of free parameters in the model. A problem in the application of this criterion in our study is that estimation of our model parameters is not totally based on maximum likelihood method. However, we can use the following estimation  $\hat{A}$  in our current modeling framework (Webster and McBratney 1989):

$$\hat{A} = \left\{ n \ln \left[ \frac{2\pi}{n} \right] + n + 2 \right\} + n \ln R + 2k$$

where  $n$  is the number of ranks or octaves classes,  $k$  is the number of parameters in a model and  $R$  is the sum of squared deviations in species abundance at each species rank or octave class. The quantity in the curly brackets is constant for a given set of data and so models can be compared by computing:  $AIC \sim n \ln R + 2k$  and  $BIC \sim n \ln R + k \ln n$  (Webster and McBratney 1989). However, one of implicit assumption of this approximation is normally distributed deviations. We transformed the abundance-rank type SADs using  $\ln(A_i + c)$ , where  $A_i$  is the abundance of the  $i$ th species and  $c$  is a constant around 0.5. The deviations of transformed abundance-rank type SADs and Preston-like SADs are close to normal distribution. Thus, it makes sense to assume that the

deviations of SADs are normally distribution (Hilborn and Mangel, 1997), and we use  $n\ln R + 2k$  and  $n\ln R + k\ln n$  to approximate AIC and BIC.

To test the robustness of approximated AIC and BIC, we also use adjusted mean sum of squared residual ( $MR_a$ ) to compare the performance of the four models.  $MR_a$  can be calculated as following (Efron and Tibshirani 1993, Hilborn and Mangel 1997):  $MR_a = R/(n - 2k)$  for abundance-rank type SAD, and  $MR_a = R/(n - k)$  for Preston-like SAD, where  $n$ ,  $R$  and  $k$  is as above-described.  $MR_a (R/(n - 2k))$  is approximate the same as Mallows  $C_p$  which is special case of AIC for general models (Efron and Tibshirani 1993, Hilborn and Mangel 1997), and  $MR_a$  does not assume normally distributed deviations. Using  $MR_a$  as criterion of model selection, we find the similar result as approximated AIC and BIC (except at scales of  $20 \times 20$  and  $20 \times 20$  m for homogeneous Thomas process and heterogeneous Thomas process model, Table 1, Appendix A6 Table A1, A2). These similar results from AIC, BIC and  $MR_a$  indicate that approximated AIC and BIC are robust for our data.

## References

- Efron, B. and Tibshirani, R. J. 1993. An introduction to the bootstrap. – Chapman and Hall.
- Hilborn, R. and Mangel, M. 1997. The ecological detective: confronting models with data. – Princeton Univ. Press.
- Møller J. and Waagepetersen R. P. 2004. Statistical inference and simulation for spatial point processes. – Chapman and Hall/CRC Press.
- Shen G. C. et al. 2009. Species-area relationships explained by the joint effects of dispersal limitation and habitat heterogeneity. – *Ecology* 90: 3033–3041.
- Waagepetersen, R. 2007. An estimation function approach to inference for inhomogeneous Neyman-Scott processes. – *Biometrics* 63: 252–258.
- Waagepetersen, R. and Guan Y. 2009. Two-step estimation for inhomogeneous spatial point

processes and a simulation study. – J. R. Stat. Soc. Ser. B 71: 685–702.

Webster, R. and McBratney, A. B. 1989. On the Akaike information criterion for choosing models for variograms of soil properties. – Eur. J. Soil Sci. 40: 493–496.

## Appendix A6

Adjusted mean sum of squared residual ( $MR_a$ ) and approximated Akaike's information criterion (AIC) values of four spatial point pattern models

Table A1. Adjusted mean sum of squared residual ( $MR_a$ ) of four spatial point pattern models for species abundance distributions (SADs) of Preston-like and rank-abundance type

SAD type	Scale ( $m^2$ )	Homogeneous Poisson (purely random)	Heterogeneous Poisson (niche)	Homogenous Thomas (dispersal)	Heterogeneous Thomas (niche + dispersal)
Preston-like	10×10	4.6901	3.1390	<b>0.3823</b>	0.5387
	20×20	6.0563	4.2647	1.2469	<b>0.5807</b>
	40×40	4.8618	1.9616	1.9869	<b>0.3348</b>
	80×80	9.6176	4.5474	4.4849	<b>3.3380</b>
	100×100	6.4288	3.4251	4.0809	<b>2.9873</b>
	200×200	10.9165	5.4832	6.3714	<b>4.8850</b>
Rank-abundance type	10×10	0.8247	0.5743	<b>0.2162</b>	0.7291
	20×20	10.3879	6.7116	<b>0.4141</b>	0.8034
	40×40	50.97835	26.3659	4.8685	<b>4.7146</b>
	80×80	289.3106	154.8750	19.0198	<b>5.7320</b>
	100×100	823.8921	461.0679	<b>36.4542</b>	62.7459
	200×200	2833.289	662.8333	275.7352	<b>89.7644</b>

The lowest  $MR_a$  in the four spatial process models in bold.

Table A2. Approximated Akaike's information criterion (AIC) values of four spatial point pattern models for species abundance distributions (SADs) of rank-abundance type with original abundance data.

SAD type	Scale ( $m^2$ )	Homogeneous Poisson (purely random)	Heterogeneous Poisson (niche)	Homogenous Thomas (dispersal)	Heterogeneous Thomas (niche + dispersal)
Rank-abundance type	10×10	130.8519	111.4253	<b>58.31158</b>	76.73214
	20×20	378.638	351.944	<b>147.0958</b>	176.6004
	40×40	674.3688	611.6455	421.6162	<b>420.0009</b>
	80×80	1027.247	868.7732	636.1655	<b>531.5328</b>
	100×100	1144.225	1038.147	<b>728.4186</b>	795.224
	200×200	1282.545	1133.868	1033.505	<b>929.0566</b>

The lowest AIC value in the four spatial process models in bold.