

# Appendices

## Online Appendix A: Model Description

This model represent a continuum between hierarchical and neutral model in the same spirit as in [1]. The model is a stochastic cellular automata (CA) or also called interactive particle system [2]. In these kind of models space is discretized into a grid and only one individual can occupy a particular position. Each position represents an area fixed by the investigator to mimic the real system. Time is continuous so the update of the model is asynchronous. I update one randomly chosen site at a time and to perform one complete time interval  $cJ$  sites have to be updated, where  $c$  is a constant that describes the overall rate at which transitions are occurring and  $J$  is the size of the grid [2].

The model use periodic boundary conditions, which makes the landscape a torus. It means that sites on the top edge of the grid are neighbors of those on the bottom edge, and sites on the right edge are neighbors of those on the left. With this choice I can avoid edge effects and is equivalent to thinking that the grid is embedded in a large community.

The size of the community is given by  $J = \dim X \times \dim Y$ , where  $\dim X$  and  $\dim Y$  are the dimension of the grid. Thus  $J$  is the maximum number of individuals in the simulated area.

In this model all individuals have almost the same parameters, besides they should belong to different species [3], and each species is assigned with a number. There are only two possible differences between species:

- They may have a different frequency in the metacommunity and also different abundances in the local community.
- Hierarchical competition: species with lower numbers have a probability to replace species with higher numbers as in [5]. Thus a species with number 1 have a probability to replace species with number 2 and greater. The species with number 2 can replace species starting from 3. The probability of replacement is a parameter, when it is 0 replacement occurs only when a species dies.

The colonization-competition and other trade-off are not explicitly included in the model. But a colonization-competition trade-off can be established if species numbering is arranged in inverse order as it's abundance  $X_i$  in the metacommunity, the most competitive species (with number 1) will have the lowest migration rate and the less competitive will have the highest migration rate.

There are four processes included in the model: death, local dispersal, and migration, starting with an empty site the following events can happen:

- (1) With probability  $m$  an individual of a species  $i$  can migrate from the metacommunity at a rate proportional to its frequency  $X_i$  in the metacommunity.
- (2) When the grid is not full, individuals give birth with rate 1 to a new individual that disperse to the neighborhood with a dispersal kernel. Here I use an exponential kernel but an inverse power kernel [4] is also implemented in the software.
- (3) Individuals die a rate  $\mu$
- (4) When an individual dies it is replaced by a migrant from metacommunity with probability  $m$  and with probability  $1 - m$  by an individual from the neighborhood. The neighborhood is established using the dispersal kernel with average distance  $d$ . Once the grid is full it keeps full, because when an individual dies is immediately replaced by another. This is called the zero-sum assumption in neutral models.
- (5) If the individual does not die it can be replaced by an individual from the metacommunity or neighborhood as in (5), but an individual of species with number  $k$  can replace an individual of a species  $k + 1$  with probability  $\rho$ . Thus a hierarchical ordering of species is established. When this probability is zero the model behavior is neutral.

The possible dispersal kernels are:

1. Exponential  $p(x) = \lambda e^{-\lambda x}$  with  $mean = 1/\lambda$  where  $x \geq 0$ .
2. Power law as defined in [4],  $p(x) = \frac{\alpha-1}{x_{min}} \left( \frac{x}{x_{min}} \right)^{-\alpha}$  with  $mean = \frac{\alpha-1}{\alpha-2} x_{min}$  where  $\alpha > 1$  and  $x \geq x_{min}$ . In all cases I used  $x_{min} = 1$ .

Simulations started with an empty grid that is colonized by migrants. They run until  $T=500$ , at this time the grid is completely occupied and steady state in richness was reached (Appendix Figure 1). Analyzing the Shannon Index  $H$ , some stochastic oscillations with differing periods appear (Appendix Figure 2). This adds some variability to repeated simulations for each parameter combination, and make the comparison of these communities more realistic, because natural communities generally suffer from disturbances that keeps them far from a steady state.

The C++ source code of the model is available at <https://github.com/lisaravia/neutral>.

## Appendix A References

1. Gravel D, Canham CD, Beaudet M, Messier C (2006) Reconciling niche and neutrality: the continuum hypothesis. *Ecol Lett* 9: 399–409. doi:10.1111/j.1461-0248.2006.00884.x.
2. Durrett R, Levin SA (1994) Stochastic spatial models: a user’s guide to ecological applications. *Philosophical transactions of the Royal Society of London Series B* 343: 329–350.
3. Hubbell SP (2001) The unified neutral theory of biodiversity and biogeography. Princeton University Press. 375p.
4. Marco DE, Montemurro MA, Cannas SA (2011) Comparing short and long-distance dispersal: modelling and field case studies. *Ecography* 34: 671–682. doi:10.1111/j.1600-0587.2010.06477.x.
5. Tilman D (1994) Competition and biodiversity in spatially structured habitats. *Ecology* 75: 2–16.

## Appendix B Tables

$D_q$ Type	SAD	Spatial Pattern	Number of Species	Side	Freq > 60	Freq > 90
DqSAD	Logseries	Randomized	8	256	0.87	0.45
			8	512	0.75	0.46
			64	256	0.99	0.51
			64	512	0.91	0.50
			256	256	1.00	0.51
			256	512	1.00	0.50
		Regular	8	256	0.90	0.63
			8	512	0.82	0.54
			64	256	0.98	0.64
			64	512	0.97	0.57
			256	256	1.00	0.85
			256	512	1.00	0.72
	Uniform	Randomized	8	256	0.43	0.00
			8	512	0.34	0.00
			64	256	0.69	0.18
			64	512	0.57	0.00
			256	256	1.00	0.33
			256	512	1.00	0.23
		Regular	8	256	0.99	0.30
			8	512	0.53	0.38
			64	256	1.00	0.79
			64	512	1.00	0.76
			256	256	1.00	0.82
			256	512	1.00	0.77
DqSRS	Logseries	Randomized	8	256	1.00	1.00
			8	512	1.00	1.00
			64	256	1.00	1.00
			64	512	1.00	1.00
			256	256	1.00	1.00
			256	512	1.00	1.00
		Regular	8	256	1.00	1.00
			8	512	1.00	1.00
			64	256	1.00	1.00

Uniform	Randomized	64	512	1.00	1.00
		256	256	1.00	1.00
		256	512	1.00	1.00
		8	256	1.00	1.00
		8	512	1.00	1.00
		64	256	1.00	1.00
		64	512	1.00	1.00
		256	256	1.00	1.00
	Regular	256	512	1.00	1.00
		8	256	1.00	1.00
		8	512	1.00	1.00
		64	256	1.00	1.00
		64	512	1.00	1.00
		256	256	1.00	1.00
		256	512	1.00	1.00
		256	512	1.00	1.00

---

Table 1: Proportion of  $R^2$  that are greater than 0.6 and 0.9 for simple spatial patterns. The  $R^2$  are from linear regressions used to estimate generalized dimensions  $D_q$  in simulated spatial patterns. Two different spatial patterns were used: *Regular* & *Randomized*, with two species abundance distributions: *Logseries* & *Uniform*, and using two approaches: *DqSRS* and *DqSAD* (See main text). *Side* is the side of grid used for simulate the spatial pattern.



$D_q$ type	Side	Metacommunity Species	Mean Species	$\rho$	Freq>60	Freq>90
DqSAD	256	11	6.1	0.000	0.98	0.81
			6.1	0.001	1.00	0.81
			6.3	0.010	0.97	0.72
			6.8	0.100	0.97	0.64
			5.0	1.000	0.90	0.49
		86	41.9	0.000	1.00	1.00
			42.8	0.001	1.00	1.00
			45.7	0.010	1.00	0.72
			27.9	0.100	0.96	0.60
			10.7	1.000	0.89	0.49
		341	109.9	0.000	1.00	1.00
			112.2	0.001	1.00	0.99
			121.1	0.010	1.00	0.70
			51.0	0.100	0.98	0.65
			13.9	1.000	0.91	0.51
DqSAD	512	11	6.8	0.000	1.00	0.81
			7.0	0.001	0.99	0.81
			7.3	0.010	0.95	0.80
			7.9	0.100	0.87	0.53
			3.7	1.000	0.83	0.54
		86	44.5	0.000	1.00	1.00
			45.1	0.001	1.00	1.00
			46.6	0.010	1.00	0.64
			42.8	0.100	0.85	0.52
			17.8	1.000	0.83	0.49
		341	146.5	0.000	1.00	0.99
			147.8	0.001	1.00	0.99
			159.9	0.010	1.00	0.60
			99.5	0.100	0.87	0.52
			36.7	1.000	0.84	0.50
DqSRS	256	11	6.1	0.000	1.00	1.00
			6.1	0.001	1.00	1.00
			6.3	0.010	1.00	1.00
			6.8	0.100	1.00	1.00

DqSRS	512	86	5.0	1.000	1.00	1.00
			41.9	0.000	1.00	1.00
			42.8	0.001	1.00	1.00
			45.7	0.010	1.00	1.00
			27.9	0.100	1.00	1.00
		341	10.7	1.000	1.00	1.00
			109.9	0.000	1.00	1.00
			112.2	0.001	1.00	1.00
			121.1	0.010	1.00	1.00
			51.0	0.100	1.00	1.00
		11	13.9	1.000	1.00	1.00
			6.8	0.000	1.00	1.00
			7.0	0.001	1.00	1.00
			7.3	0.010	1.00	1.00
			7.9	0.100	1.00	1.00
		86	3.7	1.000	1.00	1.00
			44.5	0.000	1.00	1.00
			45.1	0.001	1.00	1.00
			46.6	0.010	1.00	1.00
			42.8	0.100	1.00	1.00
		341	17.8	1.000	1.00	1.00
			146.5	0.000	1.00	1.00
			147.8	0.001	1.00	1.00
			159.9	0.010	1.00	1.00
			99.5	0.100	1.00	1.00
			36.7	1.000	1.00	1.00

Table 2: Proportion of  $R^2$  that are greater than 0.6 and 0.9 for simulated communities with different degree of neutrality given by the parameter  $\rho$ . The  $R^2$  are from linear regressions used to estimate generalized dimensions  $D_q$ .  $D_q$  was estimated using two approaches:  $DqSRS$  and  $DqSAD$  (See main text). I used a Logseries metacommunity with 11, 86 and 341 species and  $Side$  is the side of the simulation grid.

Side	Metacommunity No. Species	Replacement $\rho$	Mean No. Species
256	11	0.000	6.1
256	11	0.001	6.1
256	11	0.010	6.3
256	11	0.100	6.8
256	11	1.000	5.0
256	86	0.000	41.9
256	86	0.001	42.8
256	86	0.010	45.7
256	86	0.100	27.9
256	86	1.000	10.7
256	341	0.000	109.9
256	341	0.001	112.2
256	341	0.010	121.1
256	341	0.100	51.0
256	341	1.000	13.9
512	11	0.000	6.8
512	11	0.001	7.0
512	11	0.010	7.3
512	11	0.100	7.9
512	11	1.000	3.7
512	86	0.000	44.5
512	86	0.001	45.1
512	86	0.010	46.6
512	86	0.100	42.8
512	86	1.000	17.8
512	341	0.000	146.5
512	341	0.001	147.8
512	341	0.010	159.9
512	341	0.100	99.5
512	341	1.000	36.7

Table 3: Mean number of species for simulated communities with different degree of neutrality given by the parameter  $\rho$ . I used a Logseries metacommunity with 11, 86 and 341 species and *Side* is the side of the simulation grid.

Side	Metacommunity No. Species	Mean No. Species	Type	Power	n(Power)	Type I Error	n(Type I)
512	11	5.54	SAD	0.239	1000	0	225
512	11	5.54	DqSRS	0.676	1000	0.058	225
512	11	5.54	DqSAD	0.964	1000	0.44	225
512	86	34	SAD	0.687	1000	0.013	225
512	86	34	DqSRS	0.7	1000	0.089	225
512	86	34	DqSAD	0.761	1000	0.218	225
512	341	108	SAD	0.756	1000	0.009	225
512	341	108	DqSRS	0.7	1000	0.089	225
512	341	108	DqSAD	0.885	1000	0.191	225
256	11	5.32	SAD	0.012	1000	0	225
256	11	5.32	DqSRS	0.638	1000	0.116	225
256	11	5.32	DqSAD	0.865	1000	0.431	225
256	86	31.84	SAD	0.662	1000	0.018	225
256	86	31.84	DqSRS	0.702	1000	0.116	225
256	86	31.84	DqSAD	0.914	1000	0.360	225
256	341	78.3	SAD	0.793	1000	0.044	225
256	341	78.3	DqSRS	0.703	1000	0.191	225
256	341	78.3	DqSAD	0.909	1000	0.289	225

Table 4: Power and Type I error rate for SAD, DqSAD and DqSRS. The number of points used for SAD is the number of species found in the compared communities, and the number of points for multifractal spectra are the  $q$  used for the estimation ranging -24 to 24 ( $n=35$ ). The complete set used is  $q=\{-24,-22,-20,-18,-16,-14,-12,-10,-8,-6,-4,-3,-2.5,-2,-1.5,-1,-0.5,0,0.5,1,1.5,2,2.5,3,4,6,8,10,12,14,16,18,20,22,24\}$ .

## Appendix figures

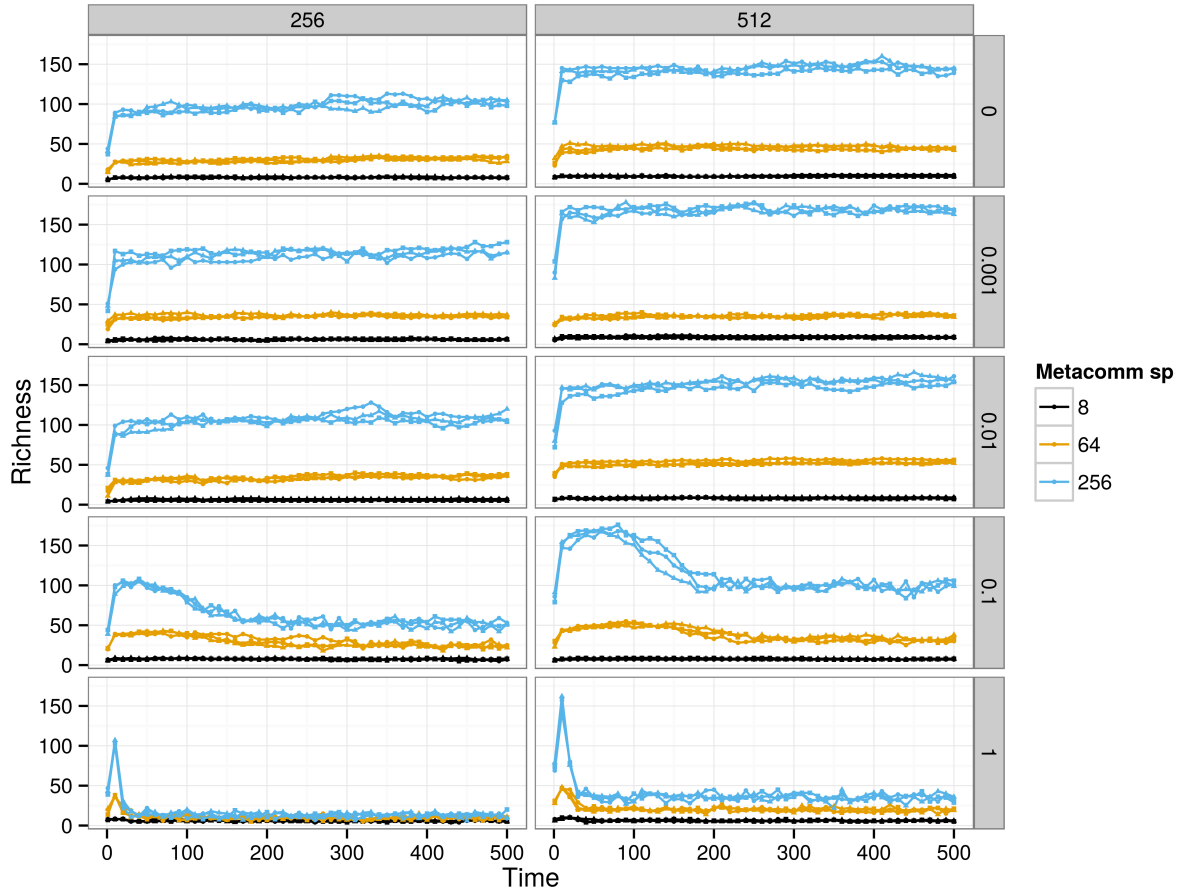


Figure 1: Time series of richness of the neutral/hierarchical model for all parameters used in power calculations (see Table 1 in the main text). Three simulations for each parameter combination are showed . The columns represent different sizes of the model grid (size) and rows different values of the parameter  $\rho$

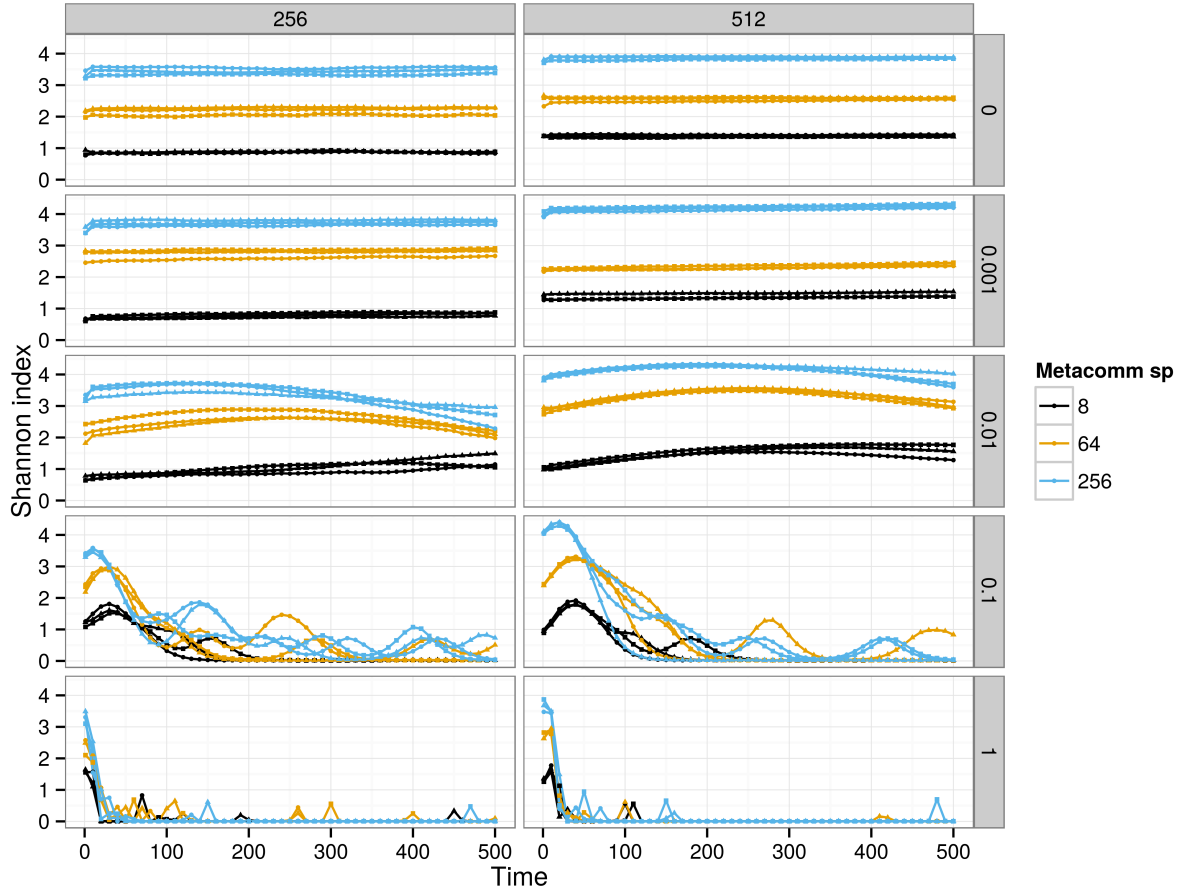


Figure 2: Time series of Shannon diversity index ( $H$ ) of the neutral/hierarchical model for all parameters used in power calculations (see Table 1 in the main text). Three simulations for each parameter combination are showed . The columns represent different sizes of the model grid (size) and rows different values of the parameter  $\rho$

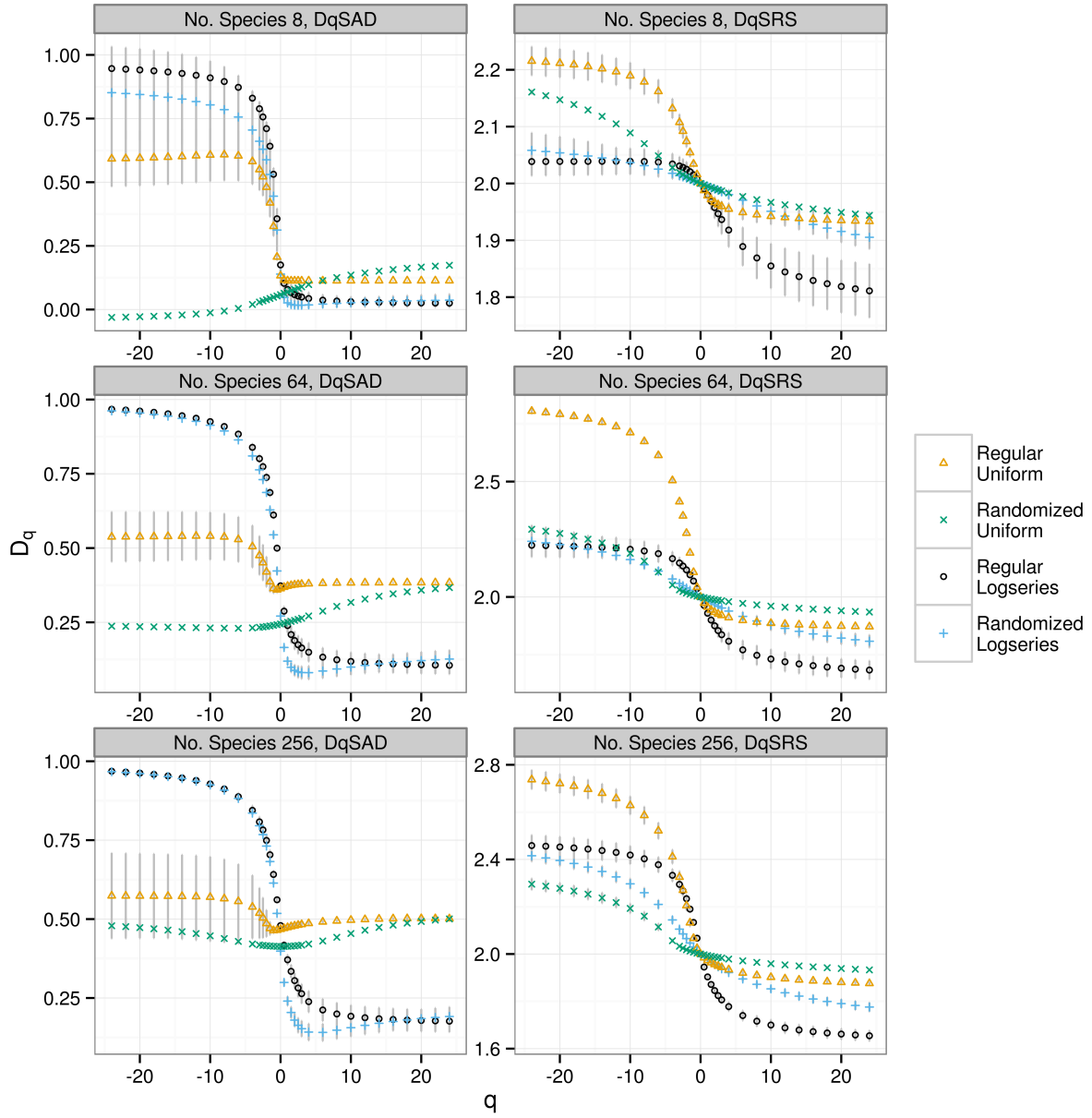


Figure 3: Generalized dimension spectra  $D_q$  of simulated species spatial patterns. The points are means of 10 repetitions of simulated patterns using a spatial grid of side=512. A Logseries or uniform species abundance distribution were used, with 8,64 and 256 species. Two forms of generalized dimensions were estimated: DqSRS, from species rank surface  $D_q^{SRS}$  and DqSAD, estimated from species abundance distribution  $D_q^{SAD}$ . Two different spatial patterns were used: Regular, the species are distributed in vertical bands, Randomized the spatial distribution of species was randomized.

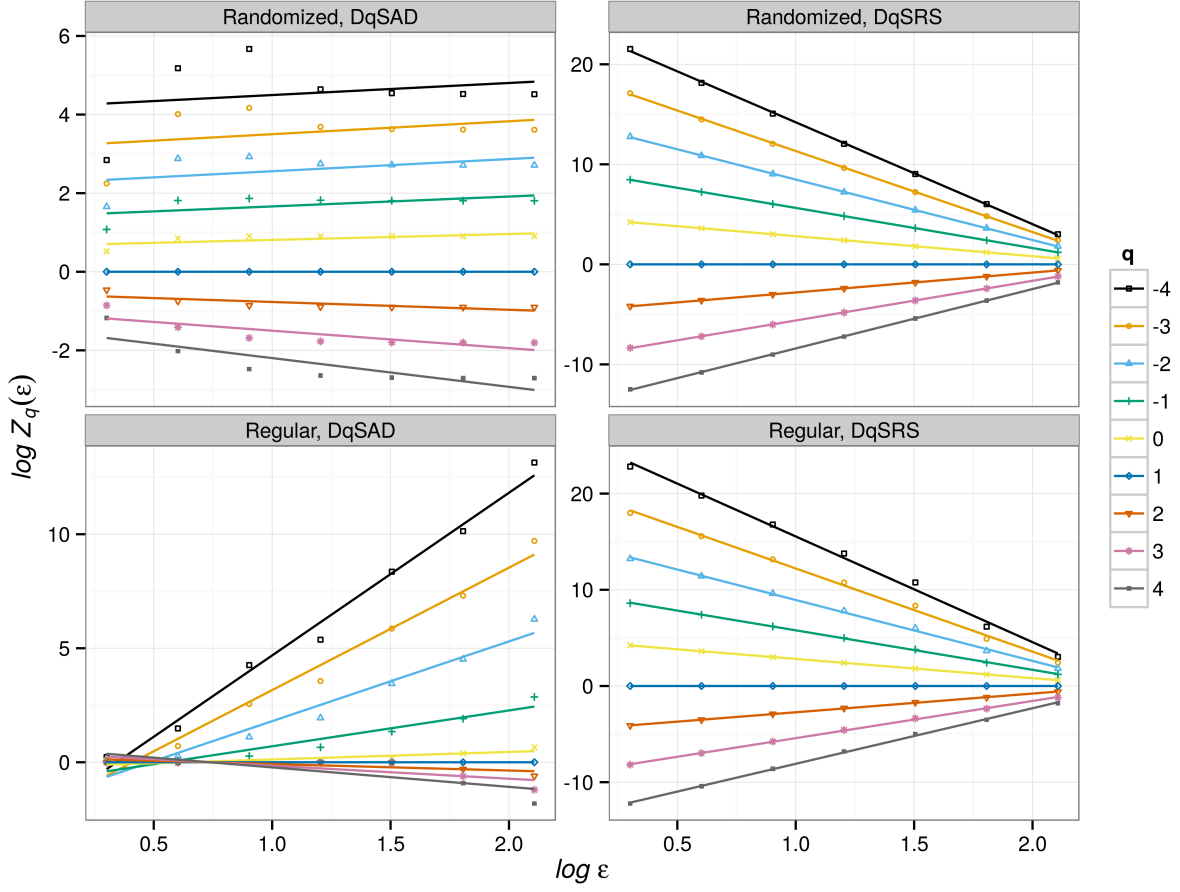


Figure 4: Linear fit from generalized dimension ( $D_q$ ) estimation from simple species spatial patterns with 64 species and a uniform abundance distribution. The spatial grid has a side=256 sites and two different spatial patterns: a) Regular, a regular spatial pattern with species distributed in vertical bands of equal width. b) Randomized, the positions of species in the grid are randomized. Two kinds of generalized dimension were estimated: DqSRS corresponds the fit of  $D_q^{SRS}$  (see text) and DqSAD is the fit from the estimation of  $D_q^{SAD}$  (see text).  $Z_q(\epsilon)$  corresponds to the partition function calculated for a box with side  $\epsilon$ .



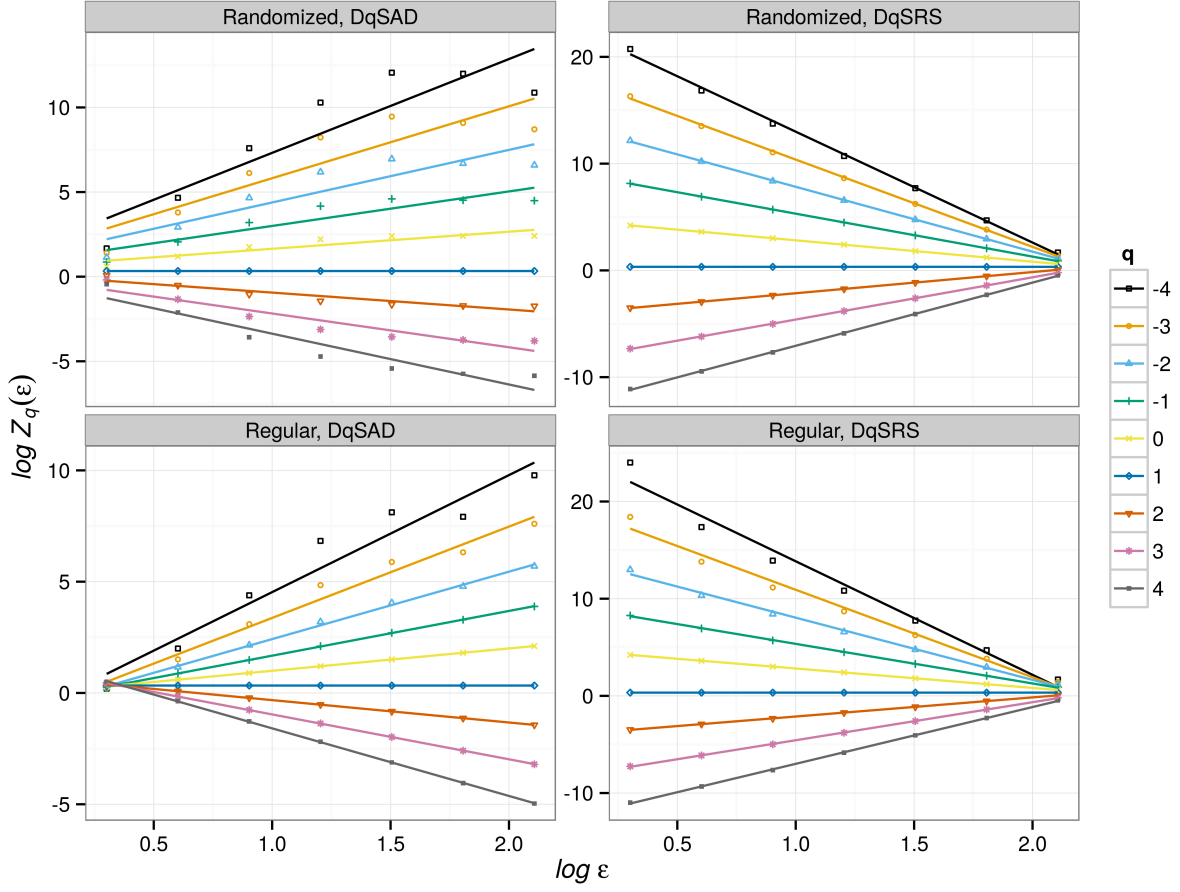


Figure 5: Linear fit from generalized dimension ( $D_q$ ) estimation from simple species spatial patterns with 256 species and a uniform abundance distribution. The spatial grid has a side=256 sites and two different spatial patterns: a) Regular, a regular spatial pattern with species distributed in vertical bands of equal width. b) Randomized, the positions of species in the grid are randomized. Two kinds of generalized dimension were estimated: DqSRS corresponds the fit of  $D_q^{SRS}$  (see text) and DqSAD is the fit from the estimation of  $D_q^{SAD}$  (see text).  $Z_q(\epsilon)$  corresponds to the partition function calculated for a box with side  $\epsilon$ .

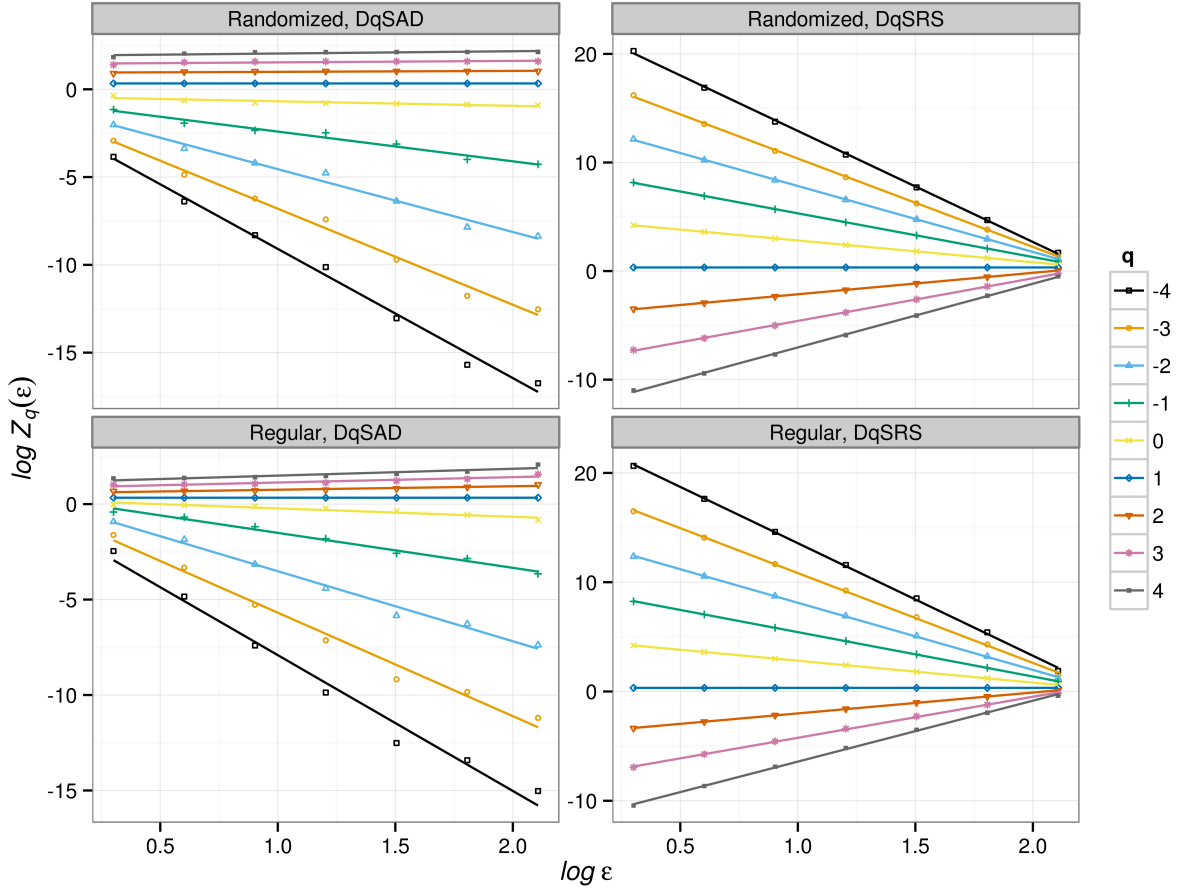


Figure 6: Linear fit from generalized dimension ( $D_q$ ) estimation from simple species spatial patterns with 8 species and a logseries abundance distribution. The spatial grid has a side=256 sites and two different spatial patterns: a) Regular, a regular spatial pattern with species distributed in vertical bands of equal width. b) Randomized, the positions of species in the grid are randomized. Two kinds of generalized dimension were estimated: DqSRS corresponds the fit of  $D_q^{SRS}$  (see text) and DqSAD is the fit from the estimation of  $D_q^{SAD}$  (see text).  $Z_q(\epsilon)$  corresponds to the partition function calculated for a box with side  $\epsilon$ .

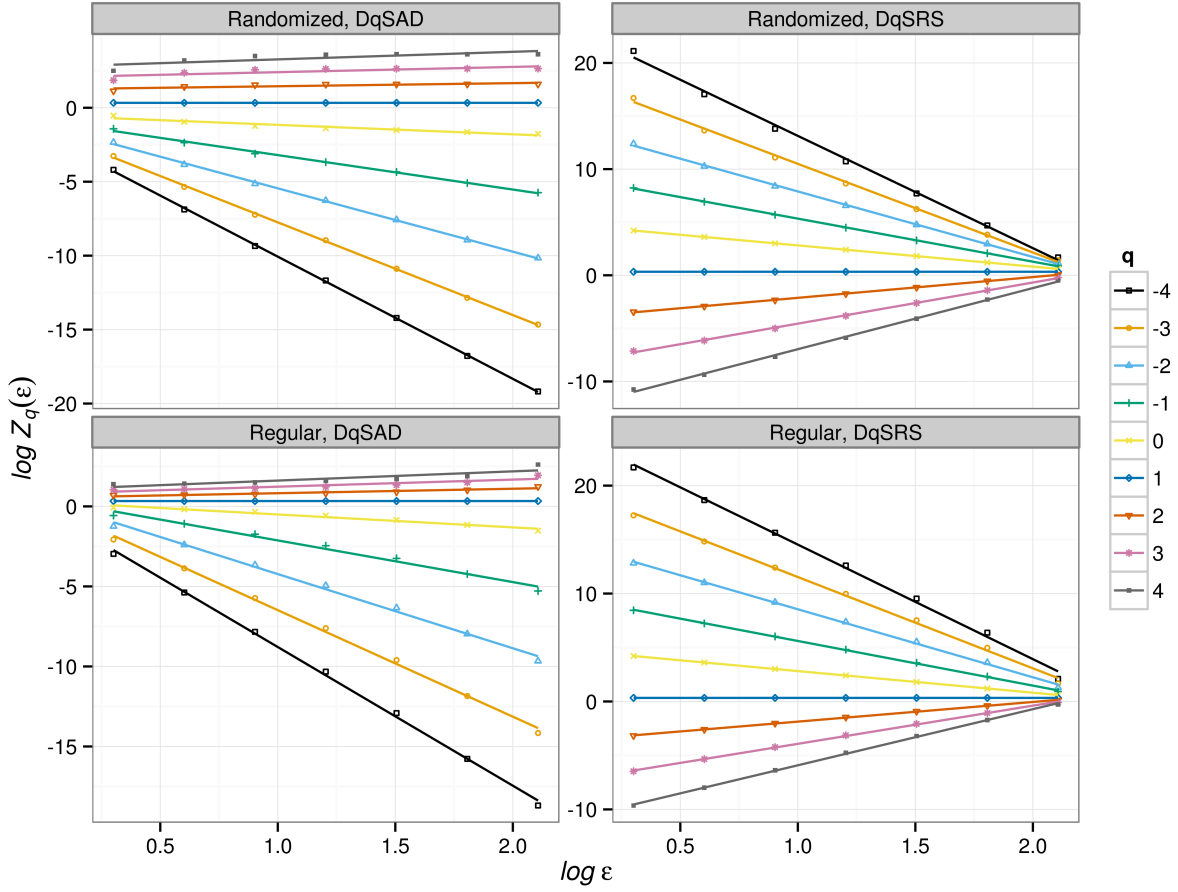


Figure 7: Linear fit from generalized dimension ( $D_q$ ) estimation from simple species spatial patterns with 64 species and a logseries abundance distribution. The spatial grid has a side=256 sites and two different spatial patterns: a) Regular, a regular spatial pattern with species distributed in vertical bands of equal width. b) Randomized, the positions of species in the grid are randomized. Two kinds of generalized dimension were estimated: DqSRS corresponds the fit of  $D_q^{SRS}$  (see text) and DqSAD is the fit from the estimation of  $D_q^{SAD}$  (see text).  $Z_q(\epsilon)$  corresponds to the partition function calculated for a box with side  $\epsilon$ .

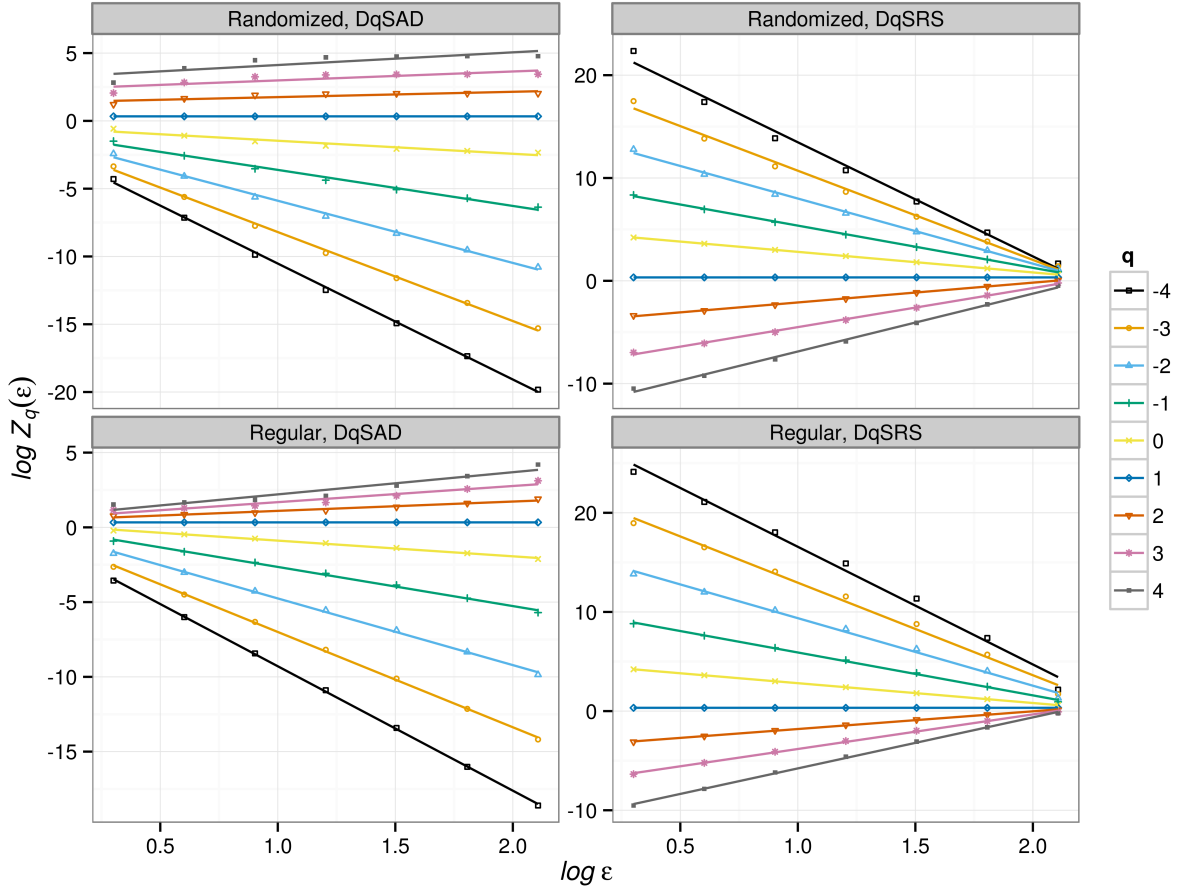


Figure 8: Linear fit from generalized dimension ( $D_q$ ) estimation from simple species spatial patterns with 256 species and a logseries abundance distribution. The spatial grid has a side=256 sites and two different spatial patterns: a) Regular, a regular spatial pattern with species distributed in vertical bands of equal width. b) Randomized, the positions of species in the grid are randomized. Two kinds of generalized dimension were estimated: DqSRS corresponds the fit of  $D_q^{SRS}$  (see text) and DqSAD is the fit from the estimation of  $D_q^{SAD}$  (see text).  $Z_q(\epsilon)$  corresponds to the partition function calculated for a box with side  $\epsilon$ .

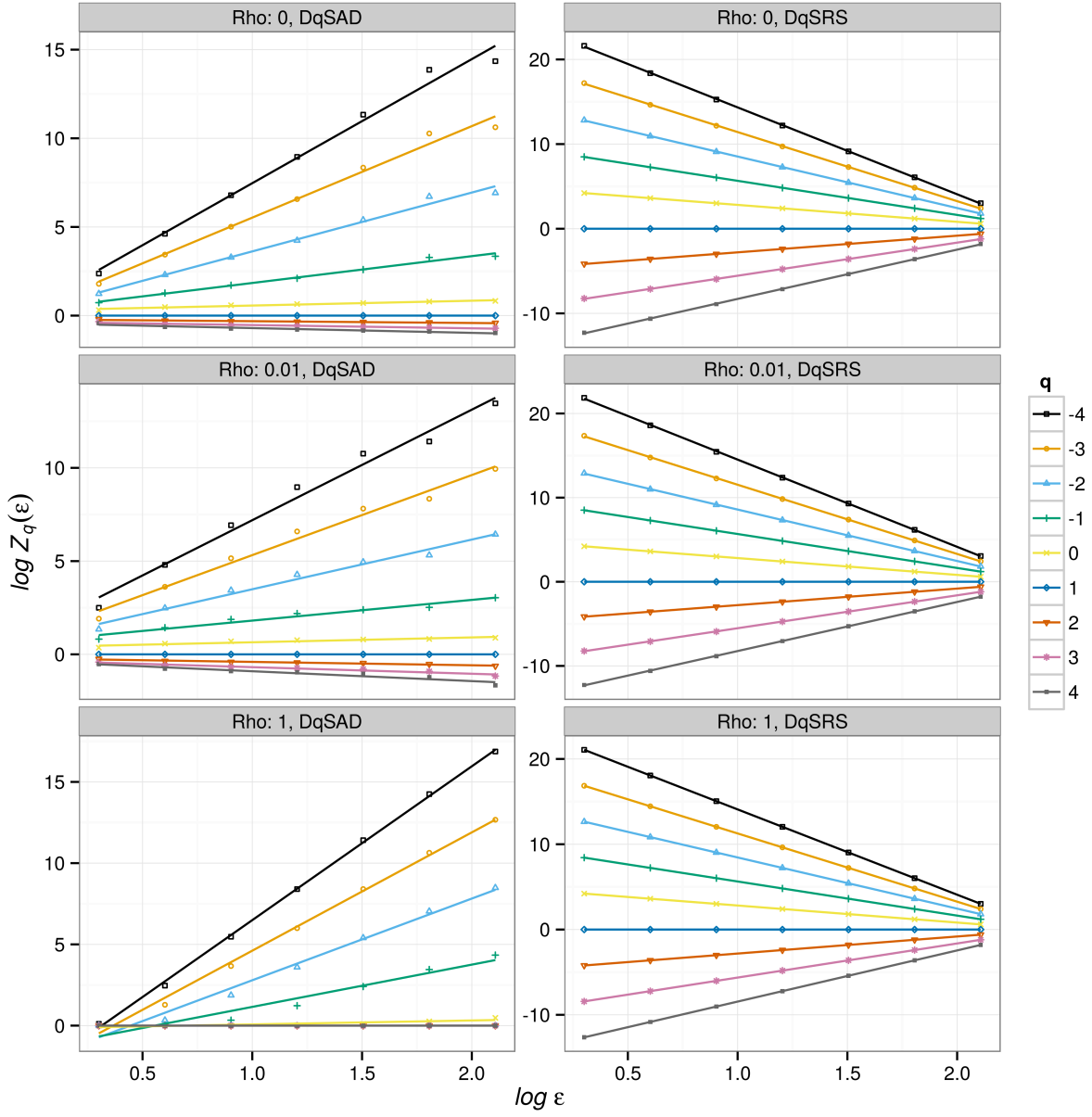


Figure 9: Linear fit of generalized dimension ( $D_q$ ) estimation from spatial patterns generated with a spatial neutral/hierarchical model with a 11 species logseries metacommunity and a simulation grid side=256 sites.  $Rho$  is the parameter that determines the degree of neutrality. When this parameter is 0 the model is completely neutral and there is no competitive replacement of species. When  $\rho$  is 1 competitive superior species always replaces inferior ones and the model is completely hierarchical.  $Species$  is the number of species actually present plot.

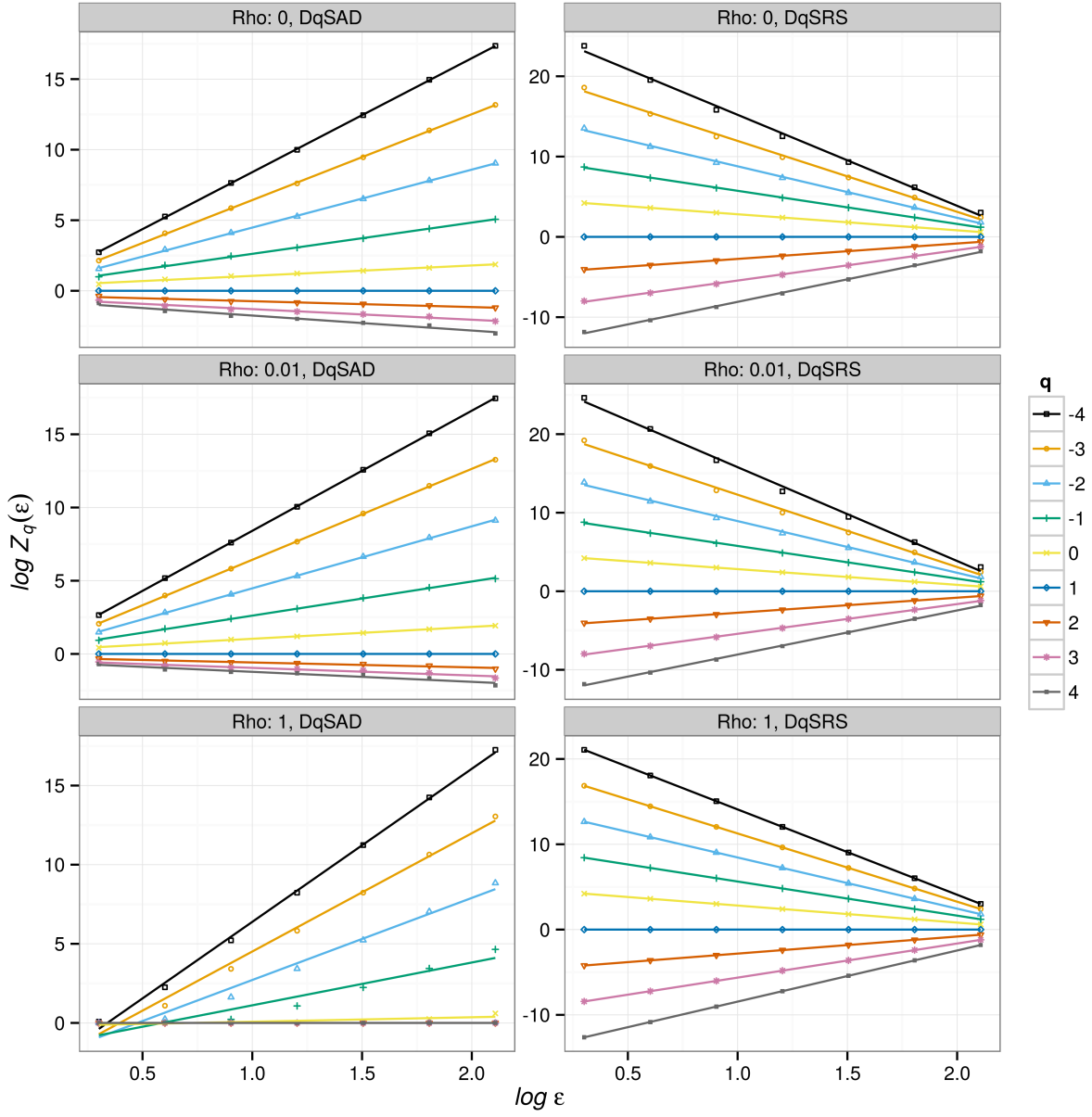


Figure 10: Linear fit of generalized dimension ( $D_q$ ) estimation from spatial patterns generated with a spatial neutral/hierarchical model with a 341 species logseries metacommunity and a simulation grid side=256 sites.  $Rho$  is the parameter that determines the degree of neutrality. When this parameter is 0 the model is completely neutral and there is no competitive replacement of species. When  $\rho$  is 1 competitive superior species always replaces inferior ones and the model is completely hierarchical.  $Species$  is the number of species actually present plot.

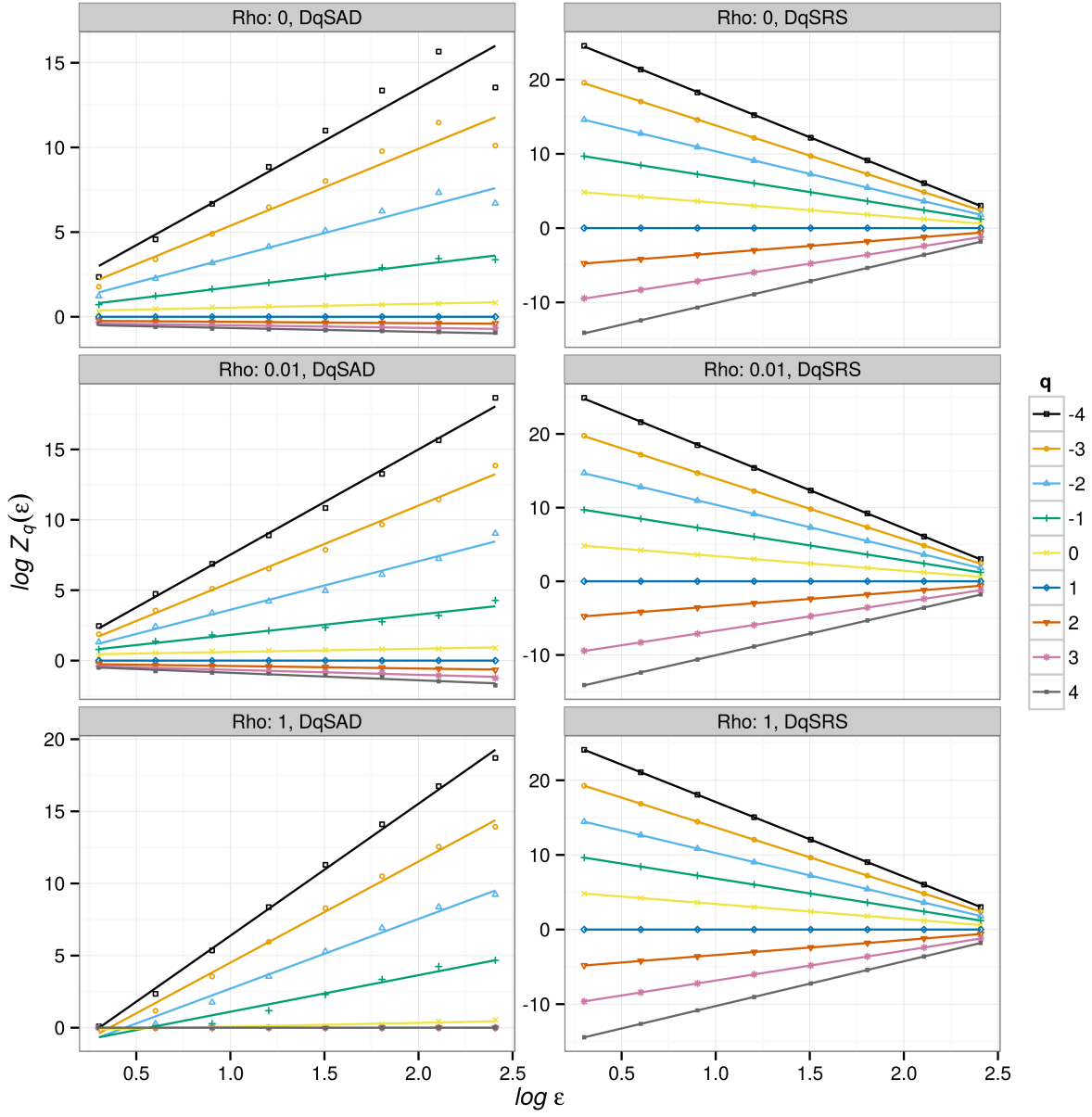


Figure 11: Linear fit of generalized dimension ( $D_q$ ) estimation from spatial patterns generated with a spatial neutral/hierarchical model with a 11 species logseries metacommunity and a simulation grid side=512 sites.  $\rho$  is the parameter that determines the degree of neutrality. When this parameter is 0 the model is completely neutral and there is no competitive replacement of species. When  $\rho$  is 1 competitive superior species always replaces inferior ones and the model is completely hierarchical.  $Species$  is the number of species actually present plot.

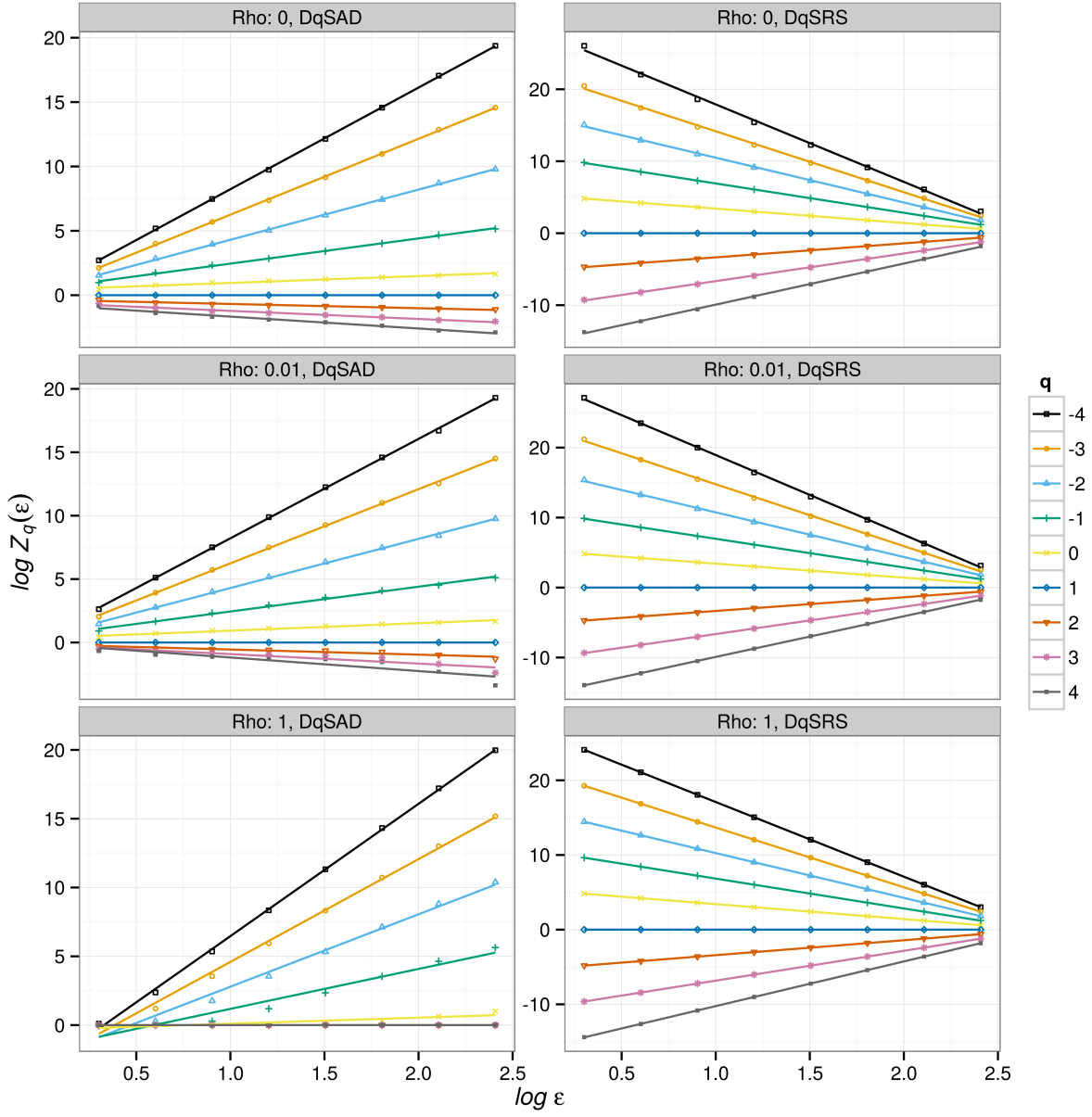


Figure 12: Linear fit of generalized dimension ( $D_q$ ) estimation from spatial patterns generated with a spatial neutral/hierarchical model with a 86 species logseries metacommunity and a simulation grid side=512 sites.  $Rho$  is the parameter that determines the degree of neutrality. When this parameter is 0 the model is completely neutral and there is no competitive replacement of species. When  $\rho$  is 1 competitive superior species always replaces inferior ones and the model is completely hierarchical.  $Species$  is the number of species actually present plot.



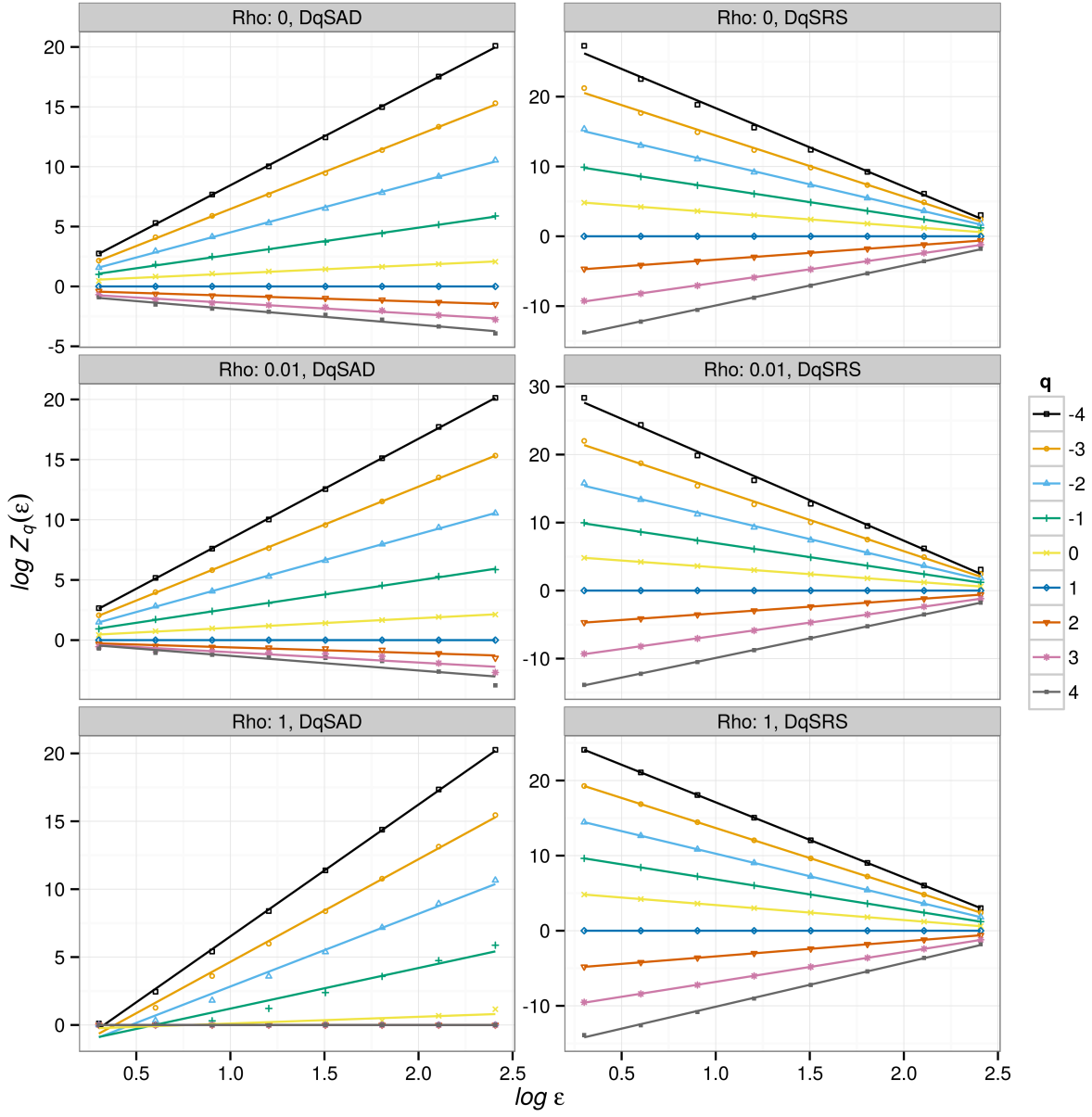


Figure 13: Linear fit of generalized dimension ( $D_q$ ) estimation from spatial patterns generated with a spatial neutral/hierarchical model with a 341 species logseries metacommunity and a simulation grid side=512 sites.  $Rho$  is the parameter that determines the degree of neutrality. When this parameter is 0 the model is completely neutral and there is no competitive replacement of species. When  $\rho$  is 1 competitive superior species always replaces inferior ones and the model is completely hierarchical.  $Species$  is the number of species actually present plot.

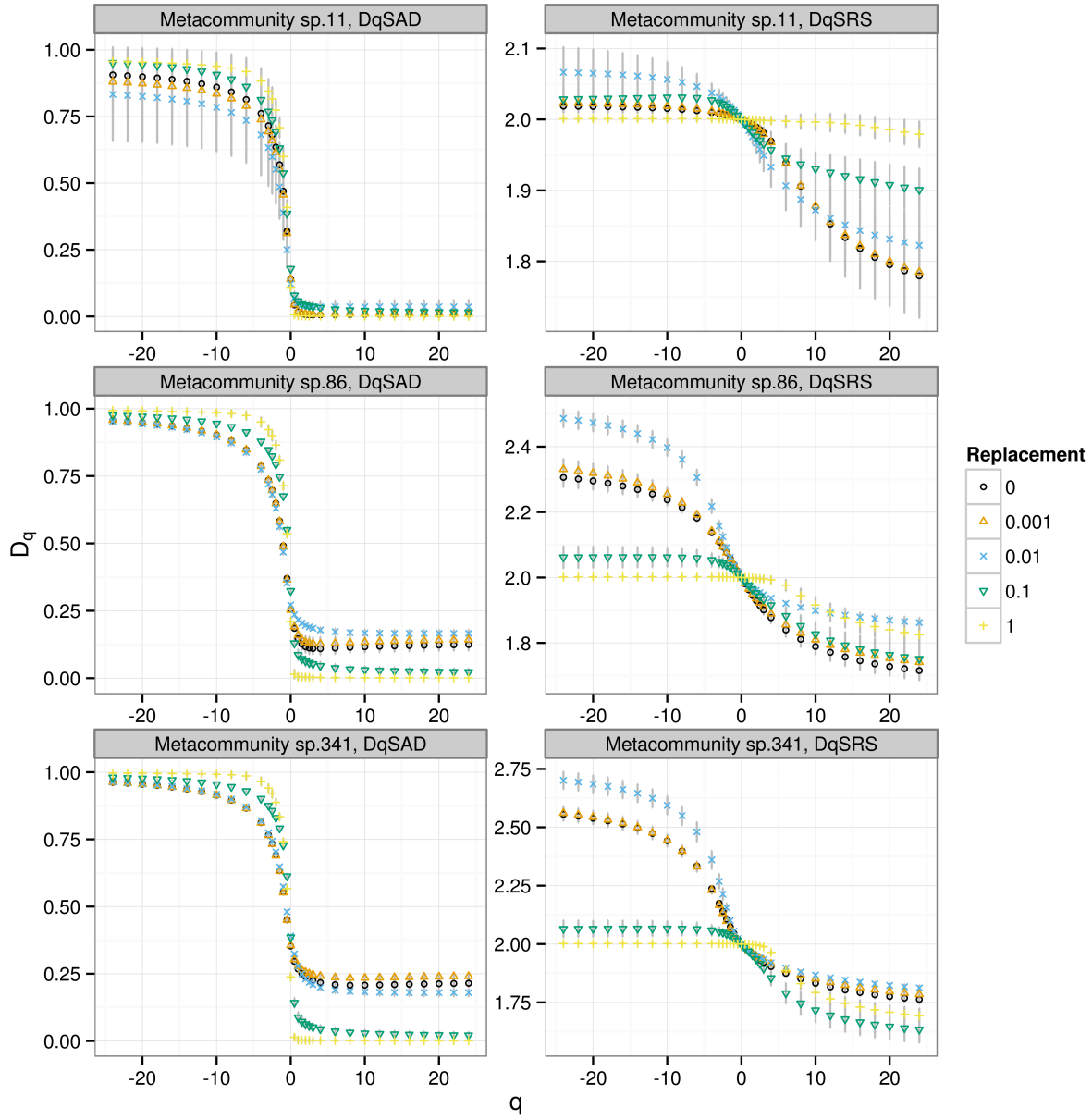


Figure 14: Generalized dimension spectra  $D_q$  of spatial patterns generated with a spatial neutral/hierarchical model. *Replacement* is the parameter that determines the degree of neutrality. When this parameter is 0 the model is completely neutral and there is no competitive replacement of species. When *Replacement* is 1 competitive superior species always replaces inferior ones and the model is completely hierarchical. The simulations use a metacommunity with a logseries abundance distribution with 11,86 and 341 species. The simulation grid side is 512, and the other parameters are: MortalityRate=0.2, DispersalDistance=0.4 (2.5 grid units), ColonizationRate=0.001.

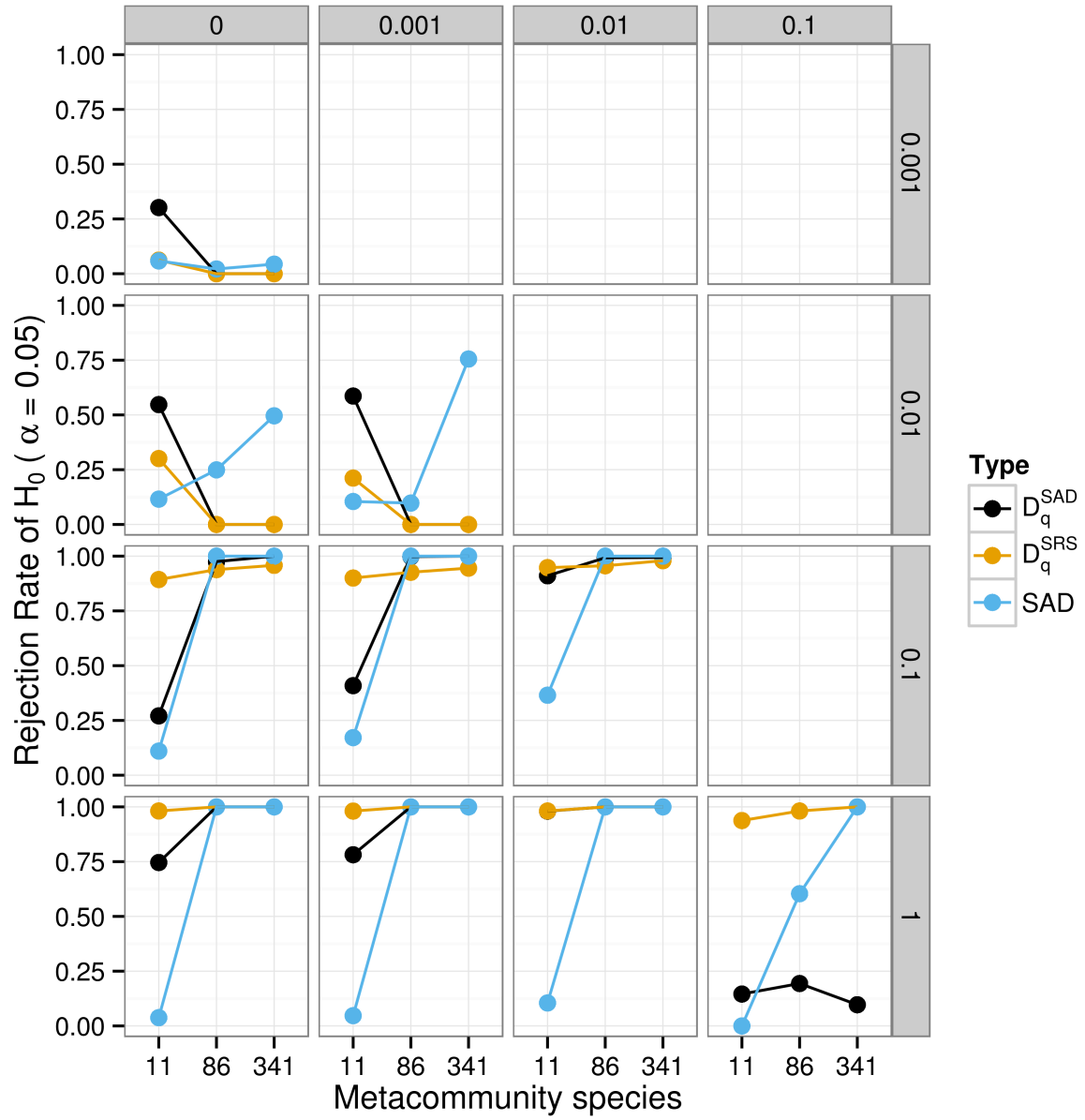


Figure 15: Power of the Anderson-Darling test for the hypothesis of differences between simulated neutral/hierarchical communities. The test uses generalized dimensions curves calculated from SAD ( $D_q^{SAD}$ ), generalized dimensions calculated from the species rank surfaces ( $D_q^{SRS}$ ) and the species abundance distributions (SAD). The compared communities differ only in parameter  $\rho$  (across panels) that determines the degree of neutrality/hierarchy. The number of comparisons to calculate the frequency is 2500 in all cases. Simulations use a metacommunity with a logseries abundance distribution with 11, 86 and 341 species; a grid side of 512 sites, the other parameters are given in the main text.