Ecography

Supplementary material

Appendix 1.

Model Overview, Design Concepts, Details and Algorithm

To overcome the difficulty of describing and replicating the results originated from an Individual Based Model (IBM), here we described the IBM developed in our study following the ODD protocol (model overview, design concepts and details, Grimm et al. 2010).

Overview

Purpose

In this study, we developed a neutral simulation individual-based model. In our model, individuals are ecologically equivalents and empirically estimated environmental productivity affects the size of communities (total number of individuals). The model can be used to understand the underlying mechanism behind the statistical association between diversity and productivity in macroecological scales in the absence of niche-based processes. Thus, we employ the model for a mechanistic evaluation of an ecological question (hypothesis) that has been long verbally discussed: does productivity affect the accumulation of diversity over space and time? (Hutchinson 1959, Brown 1981).

State Variables and Scale

Our model consists in a two-dimensional geographical map space. We used the Australian continent, divided into 1° x 1° equal-area grid cells (664 grid cells). Each cell on the grid is characterized by its geographical coordinates, its community size (number of individuals, J) and the observed Net Primary Productivity (NPP).

The temporal resolution of our model is represented by 'generations'. One generation is represented by the minimum amount of time (i.e., time steps) required for the replacement (by death, recolonization or speciation) of all individuals in the geographical domain, roughly represented by total number of individuals in the geographical domain (Boucher et al. 2014).

Process Overview and Scheduling

Within each model time step, one individual is randomly chosen and may die, disperse, speciate or reproduce according to probabilities implemented as free parameters (see *Algorithm*). The model runs for a predefined number of generations. For example, if the model runs for 6,000 generations and the carrying capacity of all the geographical cells ($n=664$) is defined by 30 individuals ($J=30$), the model will run for 119,520,000 (664*30*6,000) individual replacements. A single simulation is finished when the predefined number of generations is reached. The simulation is repeated with the same parameter sets to produce independent model replicates. Variations between model replicates are caused by stochasticity.

Design Concepts
Emergence

A presence absence matrix (PAM) and a phylogeny emerge from the neutral simulated events, as a consequence of random events of speciation, dispersal, birth and death. A unique PAM and a unique phylogeny are produced for each model replicate.

Interaction

Implicit competition occurs among individuals within the same geographical cell because the number of individuals cannot increase beyond the predefined number of individuals that each cell presents. Thus, it is assumed that resources are limited in each geographical cell, and that communities are saturated with individuals competing for space (Hubbel 2001).

Stochasticity

All events in the model are stochastic. Individuals are randomly chosen from the geographical domain and undertake neutral events that are defined according to probabilities.

Model Results

The PAM and phylogeny (outputs) of the IBM generated for each model replicate is used to compare the simulated patterns of diversity against empirical patterns.

Model Details

Initialization

The model starts with all cells populated by an ancestral species. At each time step one individual is randomly chosen to undergo neutral stochastic events (see *Algorithm*).

Input

We studied the effect of productivity on the emergence of spatial patterns of mammal's diversity in the Australian continent. Thus, the inputs for the model are (i) a

gridded map of the geographical domain and (ii) the Mean Net Primary Productivity extracted for each geographical cell.

Algorithm

1. The pairwise distances among all cells of the Australian gridded map is used to represent the geographical arrangement of the continent. In the simulation, each cell is assumed to be connected with all its adjacent neighboring cells (maximum of eight neighboring cells, for non-coastal cells, i.e., Moore neighborhood).

2. All communities (i.e. geographical cells) are saturated and populated by *J* individuals that compete only for space. At the beginning of the simulation all individuals belong to the same ancestral species.

3. A single individual is randomly selected across the geographical domain to undergo random demographic events.

4. The randomly chosen individual may speciate (υ), or die, based on death probability (1- υ).

5. If an event of death is the outcome of the stochastic dynamics, a descendent of a randomly chosen individual (parent) will colonize the empty spot. The parent may be from one of the neighboring cells in an event of dispersal (with probability *m*), or from the same cell in which the death event occurred (1-*m*).

6. The model is run for a predefined number of generations, enough to reach stability of the patterns in the metacommunity.

7. Each model replicate (i.e., simulation repeated with the same parameter sets) exports a PAM and a phylogeny.

Submodels

Speciation Mode

Many realistic speciation modes have been proposed for neutral models (Rosindell et al.2010, Desjardins-Proulx & Gravel, 2012). However, we modeled speciation as an instantaneous event, in compatibility with the broad spatial and temporal scales used in our study, as well as for the sake of simplicity and computational tractability (see Davies et al. 2011 and Boucher et al. 2014 for similar assumptions). In this study, speciation events are modeled by *point mutation* (Hubbell 2001), by randomly selecting one individual among all the individuals of the geographical domain, and altering its species identity based on speciation probability (υ). In *point mutation* mode, only the randomly selected individual becomes one of the new daughter species (Incipient species abundance = 1). Phylogenies are thus reconstructed based on the history of speciation events. Thus, in our model, time and phylogeny are measured in units of simulation steps. Consequently, the calculated phylogenetic diversity in the model differs in scale from empirical observations. Anyway, differences in scale and units of variables do not affect Pearson's correlation coefficient, as it measures only the association (standardized covariance) between the variables. However, to calculate and map the residuals (difference between observed and predicted PD values), we standardized observed and predicted PD to the same scale (0 to 1). The same procedure was followed for species richness. The neutral model is not designed to replicate the absolute number of species in nature, as the number of individuals of real-world species is unknow and computationally intractable. Thus, to calculate the map of residual richness (difference between observed and predicted richness) we standardized observed and predicted richness to the same scale (0 to 1). When comparing the

simulated and empirical richness (or PD), we want to reproduce the relative variation of richness across space, not the exact absolute number of real-world species in each grid cell.

The same procedure was followed for species richness because the neutral model is not designed to replicate the absolute number of observed species in nature, as the number of individuals of real-world species is unknow and computationally intractable.

Parameter Estimation

In this study, we tested whether our neutral model, with spatially varying community size, is capable of replicating empirical heterogeneous pattern of mammalian diversity in Australia. The multi-dimensional "parameter landscape" is defined by the parameters of the logistic function (α and β), dispersal limitation (m) and speciation rate (υ), which we explored using a Gibbs sampling Monte Carlo Markov Chain (MCMC) simulation. This method can be used to evaluate the effect of parameters on the predictive power of the model (e.g. model goodness-of-fit, f). By design, the sampler explores in greater detail (i.e. more frequently) the areas of parameter space of higher goodness-of fit. It also visits areas of the parameter landscape with lower goodness-of-fit, but with lower frequency. Thus, instead of maximizing the model predictive power, the sampler provides a complete picture of the parameter landscape, which can be described by the density of sampled parameters within a given parameter region. This density of sampled parameter values can be used to estimate the uncertainty and sensitivity of the analysis. Here, the summary statistic used in the Gibbs sampler (f) is a pseudo-likelihood measure (i.e. probability of the data, given the model and its parameters) approximated by the goodness-of-fit measure. f is the estimated

Pearson's coefficient of linear correlation (r) between simulated and empirical richness across the Australian continent.

Here we summarize how the Gibbs sampler operates: (i) the sampler defines a particular parameter combination and launches the simulation model using those parameters; (ii) once the simulation is finished, the predictions of the model, given the parameter set, are extracted; (iii) the goodness-of-fit between (Pearson's r) between observed (real-world) and predicted (simulated) data is calculated; (iv) the simulation is replicated 100 times, using the same parameter combination; (v) the pseudo-likelihood of the model and parameter combination (f) is calculated as the average r among all replicates of the parameter combination; (vi) the sampler generates a disturbance (variation) in the model parameters and returns to step i; (vii) once the sampler explored the parameter space in great detail, it produces the distribution of all the explored parameters; (viii) the distribution of parameters values are averaged and a parameter set defined with the average values is considered as the estimated parameter set. Note that the parameter set that maximized the model prediction is not used, but the parameter set is calculated based on the average of all parameters sets explored by the model (i.e. those with low, medium or high f to empirical data). We ran the sampler during preliminary evaluations of the model and used previous runs as burn-ins. The initial parameters were based on our previous knowledge of model performance under different values of dispersal, speciation, α and β (see Results section). In our analysis, each model run is defined as a unique combination of parameter values, which is replicated 100 times. We ran ~ $6*10^5$ replicates to explore ~6,000 parameter combinations (MCMC iterations – total chain length).

References

Boucher, F. C. et al. 2014. Neutral biogeography and the evolution of climatic niches. - Am. Nat. 183: 573–84.

Boucher, F. C. et al. 2014. Neutral biogeography and the evolution of climatic niches. - Am. Nat. 183: 573–84.

Brown, J. H. 1981. Two decades of homage to santa rosalia: Toward a general theory of diversity. - Integr. Comp. Biol. 21: 877–888.

Davies, T. J. et al. 2011. Neutral biodiversity theory can explain the imbalance of phylogenetic trees but not the tempo of their diversification. - Evolution (N. Y). 65: 1841–1850.

Desjardins-Proulx, P. and Gravel, D. 2012. How Likely Is Speciation in Neutral Ecology? - Am. Nat. 179: 137–144.

Grimm, V. et al. 2010. The ODD protocol: A review and first update. - Ecol. Modell. 221: 2760–2768.

Hubbell, S P. 2001. The Unified Neutral Theory of Biodiversity and Biogeography. Princeton University Press, Princeton.

Hutchinson, G. E. 1959. Homage to Santa Rosalia or Why Are There So Many Kinds of Animals? - Am. Nat. 93: 145–159.

Rosindell, J. et al. 2010. Protracted speciation revitalizes the neutral theory of biodiversity. - Ecol. Lett. 13: 716–727.

Table A1. The effect strength of primary productivity on the predictive power of the model (measured as the correlation between observed and predicted PD of mammals in Australia). The predictive power of the model increases as the effect of productivity in community size gets stronger. The same pattern is observed for different ratios between speciation and dispersal (υ/m). The parameters used for different υ/m ratios are presented on Table A2.

α	β	Mean Pearson's r for different ratios between speciation and dispersal rate (υ/m) based on 100 replicates								
		0.5	1	2	3	4	5	6	7	8
0	0	-0.15	-0.05	-0.19	-0.18	-0.14	-0.08	-0.01	-0.05	-0.02
-17.83	0.02	0.01	0.07	0.02	0.05	0.11	0.12	0.10	0.13	0.05
-17.31	0.020256	0.14	0.14	0.07	0.11	0.11	0.06	0.15	0.16	0.15
-16.80	0.020513	0.06	0.09	0.17	0.13	0.10	0.10	0.19	0.18	0.13
-16.29	0.020769	0.17	0.13	0.19	0.20	0.17	0.21	0.19	0.19	0.14
-15.77	0.021026	0.20	0.16	0.23	0.21	0.22	0.20	0.24	0.20	0.23
-15.26	0.021282	0.26	0.22	0.21	0.23	0.24	0.27	0.29	0.29	0.25
-14.75	0.021538	0.29	0.24	0.27	0.28	0.27	0.33	0.30	0.30	0.35
-14.24	0.021795	0.30	0.40	0.31	0.30	0.31	0.32	0.36	0.32	0.33
-13.72	0.022051	0.32	0.38	0.34	0.33	0.34	0.37	0.32	0.35	0.32
-13.21	0.022308	0.37	0.38	0.35	0.34	0.34	0.36	0.38	0.34	0.38
-12.70	0.022564	0.37	0.41	0.35	0.37	0.37	0.40	0.37	0.37	0.37
-12.18	0.022821	0.39	0.40	0.36	0.36	0.36	0.39	0.41	0.37	0.37
-11.67	0.023077	0.42	0.43	0.42	0.40	0.40	0.45	0.42	0.42	0.41
-11.16	0.023333	0.42	0.46	0.43	0.42	0.43	0.42	0.42	0.43	0.45
-10.65	0.02359	0.50	0.48	0.44	0.44	0.44	0.48	0.50	0.47	0.46
-10.13	0.023846	0.50	0.52	0.49	0.47	0.50	0.51	0.49	0.48	0.51
-9.62	0.024103	0.55	0.53	0.52	0.51	0.53	0.52	0.53	0.54	0.54
-9.11	0.024359	0.58	0.59	0.56	0.56	0.56	0.55	0.57	0.55	0.58
-8.59	0.024615	0.63	0.64	0.61	0.60	0.60	0.60	0.63	0.60	0.61

-8.08	0.024872	0.64	0.63	0.61	0.63	0.64	0.64	0.63	0.63	0.64
-7.57	0.025128	0.65	0.67	0.65	0.66	0.67	0.66	0.68	0.65	0.66
-7.06	0.025385	0.67	0.69	0.68	0.70	0.68	0.67	0.69	0.67	0.68
-6.54	0.025641	0.71	0.70	0.72	0.70	0.71	0.72	0.71	0.70	0.72
-6.03	0.025897	0.72	0.72	0.70	0.70	0.72	0.72	0.70	0.70	0.71
-5.52	0.026154	0.74	0.72	0.73	0.73	0.72	0.70	0.68	0.72	0.71
-5.00	0.02641	0.74	0.74	0.73	0.74	0.73	0.69	0.68	0.72	0.72
-4.49	0.026667	0.74	0.72	0.74	0.73	0.71	0.71	0.71	0.70	0.72

Table A2. Different ratios between speciation and dispersal rate assumed in Table A1.

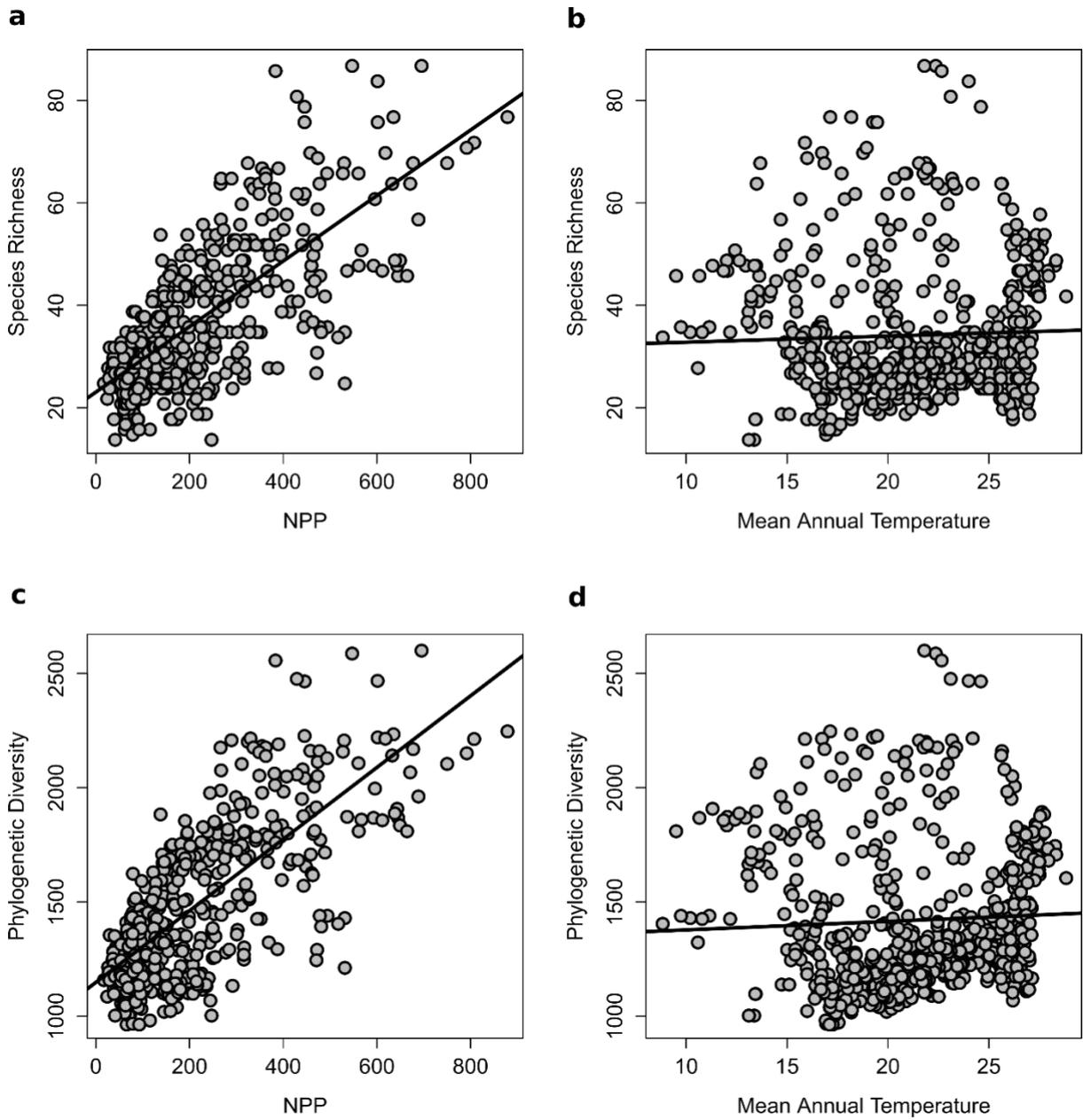| υ/m	Dispersal rate (m)	Speciation rate (υ)
0.5	0.1000	0.0500
1.0	0.0004	0.0004
2.0	0.0214	0.0422
3.0	0.0161	0.0474
4.0	0.0109	0.0448
5.0	0.0056	0.0291
6.0	0.0056	0.0343
7.0	0.0056	0.0396
8.0	0.0056	0.0448

Supplementary figures

Figure A1. Relationship between observed species richness and phylogenetic diversity with Net primary productivity and Mean Annual Temperature.
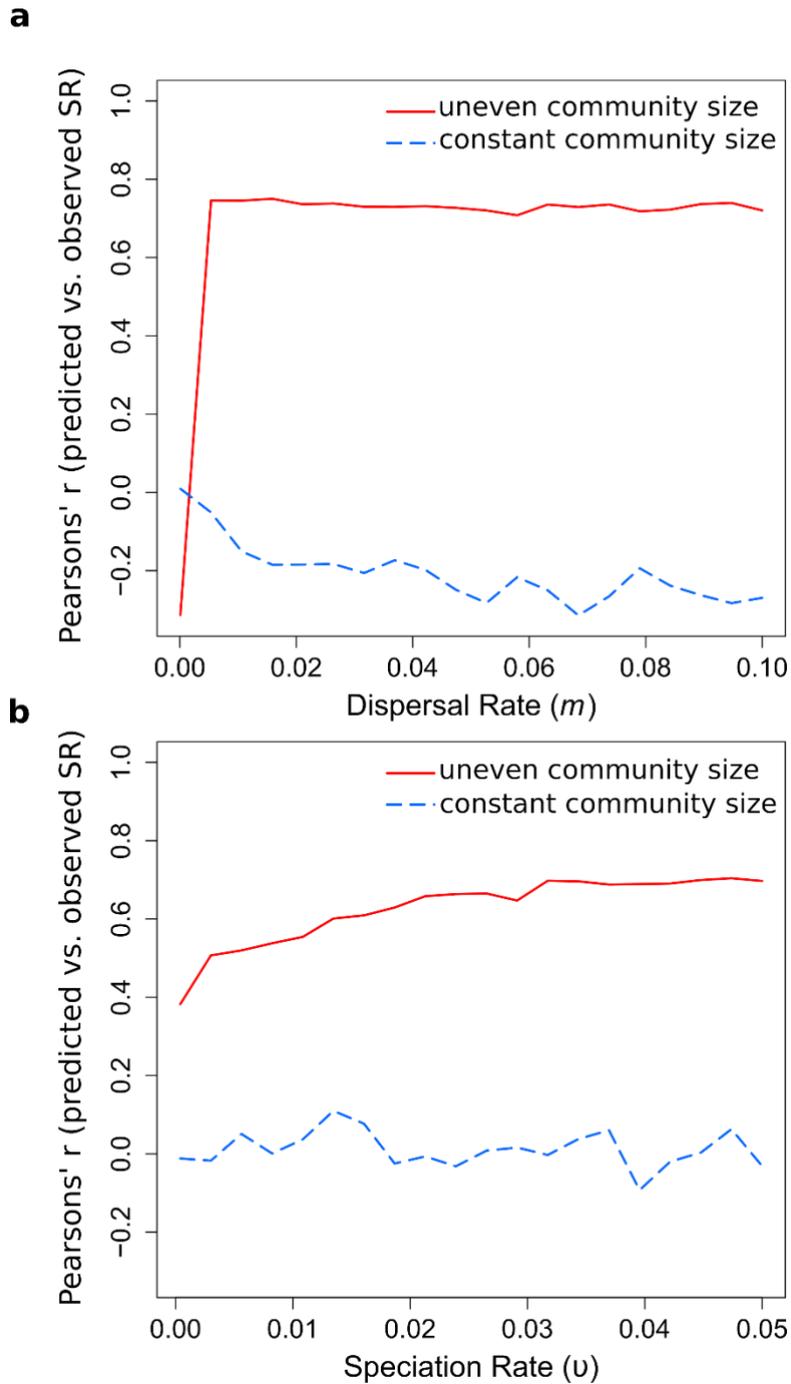
Figure A2. Effect of dispersal rate (a) and speciation rate (b) on the predictive power of neutral models. The continuous red line illustrates the neutral model that assumes the effect of primary productivity on community size. The dashed blue line illustrates the classical neutral model in which community size is constant over the geographical domain.
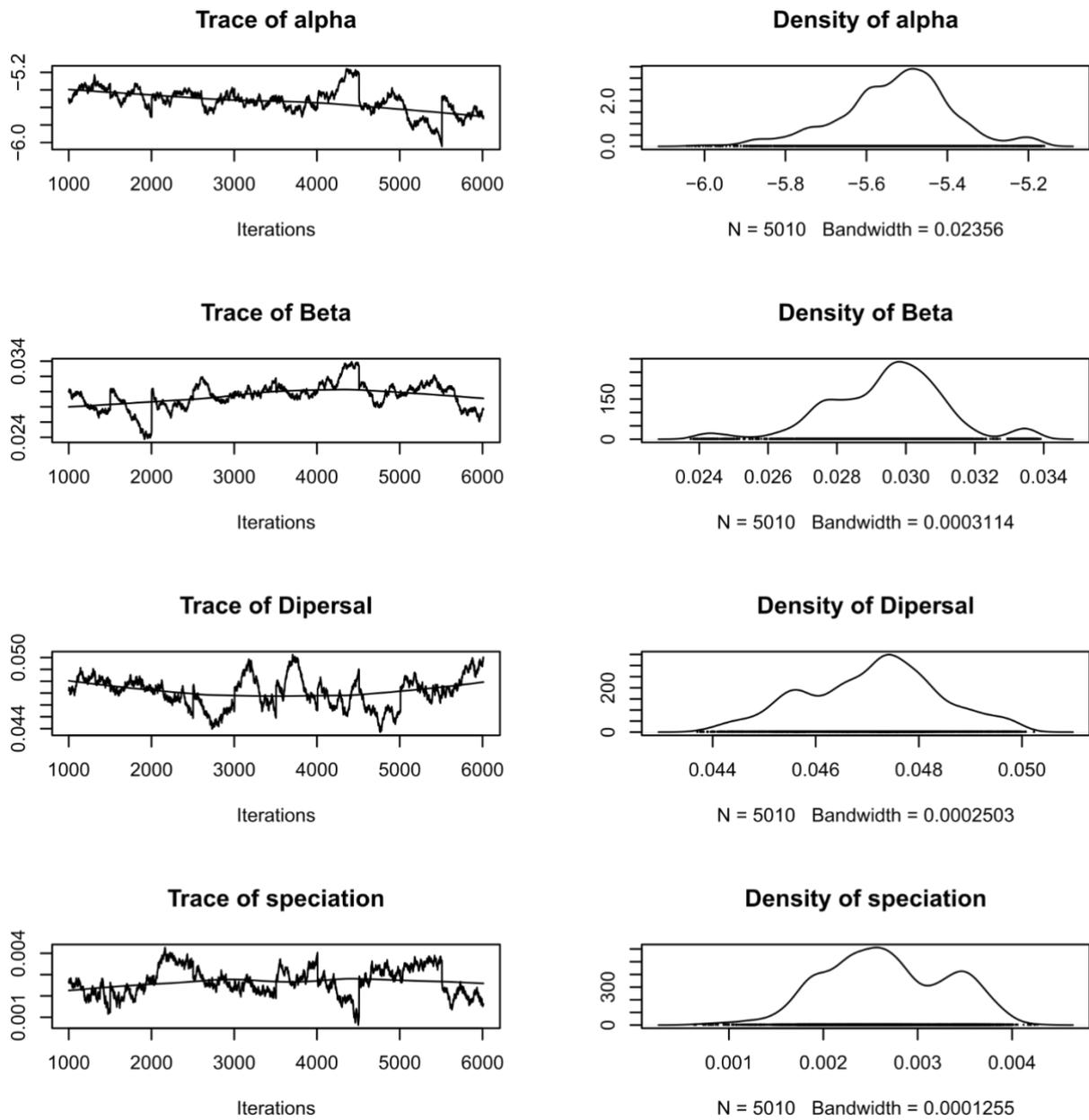
Figure A3. Gibb's MCMC search based on almost 6000 iterations. Parameters were estimated through a mean of their posterior distribution (α =-5.52774; β = 0.02942; m = 0.04709; υ = 0.00267).
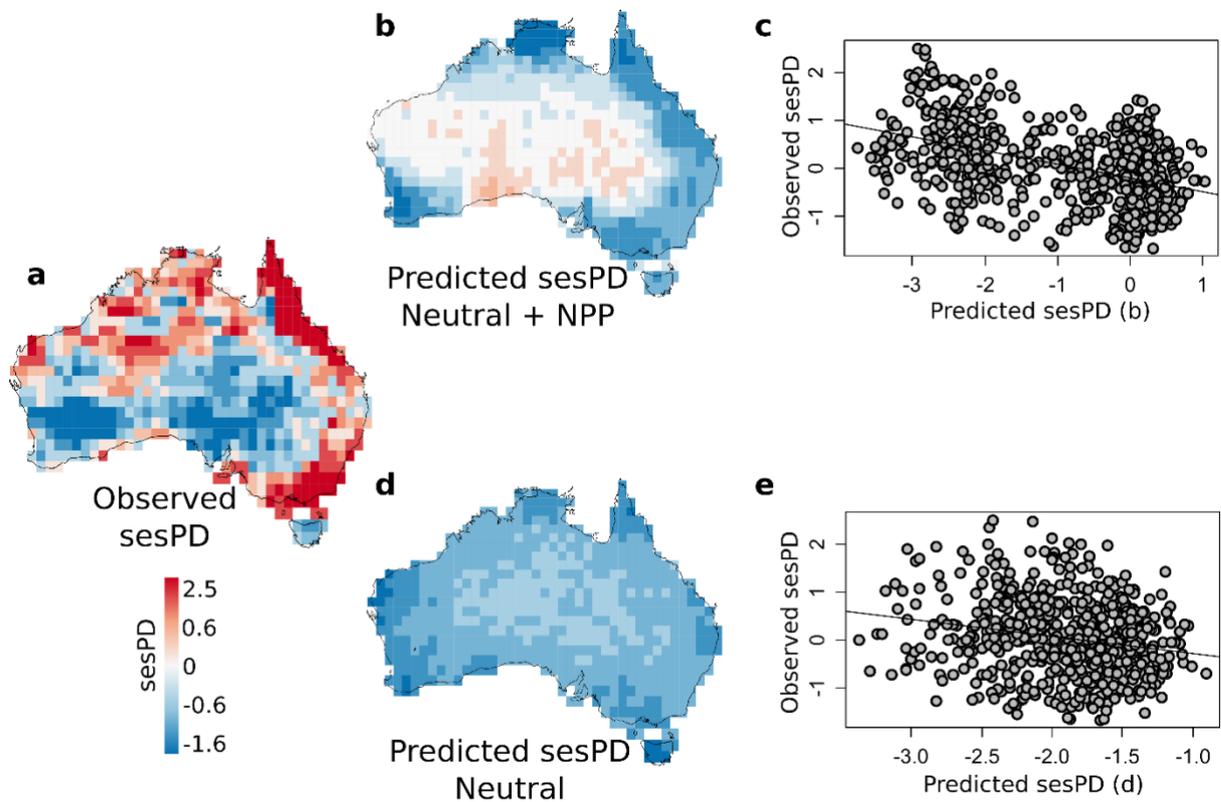
Figure A4. Standardized effect size of PD (sesPD) for observed and predicted data. *a*, observed sesPD for terrestrial mammals in Australia, *excluding monotremes*; *b*, the sesPD predicted by a neutral model that incorporates the effect of productivity on community size; *c*, relationship between observed *a* and predicted *b*; *d*, the sesPD predicted by a neutral model that assumes no effect of productivity; *e*, the relationship between observed *a* and predicted *d*.