3 **Appendix S1. Description of an algorithm generating simulated community**

4 **data along two environmental gradients (2D simulated community data set)**

5 The model for generating simulated community data was inspired by the COMPAS model

6 introduced by Minchin (1987). The current implementation builds upon the original one-

7 gradient version introduced by Fridley et al (2007) and uses also parts of the original script

8 written by Jason Fridley (published as Appendix S1 in Fridley et al. 2007; for details, see the

9 original description).

10   Generally, the model is based on two orthogonal 'environmental' gradients of equal or

11 different length (*gradient 1* and *gradient 2*). Along each of these gradients, S species response

12 curves with randomly generated optima and niche-widths were generated, representing

13 potential abundances of species in a given location of the gradient. The species response

14 curves were modelled using the Beta function, allowing for an asymmetric unimodal shape

15 (Minchin 1987). In the next step, N_p sampling locations were randomly selected in the space

16 defined by the two orthogonal gradients. In each sampling location, N_{ind} individuals were

17 randomly assigned to different species. The probability that a particular individual will be

18 assigned to a certain species was proportional to this species' potential abundance in a given

19 sampling location, which equals to square-rooted multiplication of species potential

20 abundances along the first and the second environmental gradient (as defined by relevant

21 species response curves).

22   For the purpose of this study, I created 10 community data sets of increasing beta

23 diversity, while fixing some of the parameters to be constant for all data sets. *Gradient 1* has

24 the length fixed to 1000 units in all data sets; increasing beta diversity of individual data sets

25 is achieved by increasing the length of *gradient 2*, which imitates the situation when data set

26   expands over more communities either in ecological or geographical space. The length of

27   *gradient 2* ranged from 1000 units for "one community" to 10 000 units for "ten

28   communities", where *number of communities* is an arbitrary measure with each community

29   spanning 1000 units along *gradient 2* (see Appendix S1: Fig. S1 for examples of community

30   data sets composed of one, five and ten communities). With the increasing length of *gradient*

31   *2*, the number of species in the simulated data set also increased ($S = 100 \times$ *number of*

32   *communities*), while the number of individuals remained fixed in all data sets to 500

33   individuals per sample. Niche widths for all species along both gradients were random

34   numbers generated from the uniform distribution in the range of 500 to 1000 units, and the

35   number of samples created for each community remained constant ($N_p = 50$; see Appendix

36   S1: Table S1 for the setting of all parameters). Position of samples along *gradient 1* are used

37   as the sample attributes, while positions of species optima along *gradient 1* are used as the

38   species attributes.

39   The increasing number of communities is reflected by an increase in the beta diversity

40   of the data set, which is quantified as a variation in the matrix of compositional distances

41   calculated using squared Whittaker distance (Legendre & De Cáceres 2013). Whittaker

42   distance is one complement of Whittaker's index of association (Whittaker 1952), and in

43   Legendre & Legendre (2012) is denoted as D_9. It is Manhattan type distance based on species

44   profiles, comparing species composition of samples by differences in their relative species

45   abundances. Since CWM calculation is also based on relative species abundances (species

46   profiles), this measure is suitable for calculation of beta diversity in the context of CWM

47   analysis.

48   The R code for generating 2D simulated community data is implemented in the R

49   package *weimea* (Zelený, unpubl., available here as Appendix S5 and also from

50   https://github.com/zdealveindy/weimea) and consists of two functions: *simul.comm.2*, which

51 creates parameters of simulated community with species in 2D space, and *sample.comm.2*,

52 which samples such simulated community at random or selected positions along both

53 gradients. Additionally, there is the function *draw.ecospace*, which visualizes the distribution

54 of species and samples in simulated ecospace (see Appendix S1: Fig. S1), and the function

55 *betadiv.ia* calculating the beta diversity of the species composition matrix as a variation in the

56 matrix of squared distances among samples calculated by Whittaker distance.

57

58 **References**

59 Fridley, J., Vandermast, D.B., Kuppinger, D.M., Manthey, M. & Peet R.K. 2007. Co-

60 occurrence based assessment of habitat generalists and specialists: a new approach for the

61 measurement of niche width. *Journal of Ecology* 95: 707–722.

62 Legendre, P. & Legendre, L. 2012. *Numerical ecology, Third English Edition*. Elsevier

63 Science, Amsterdam, The Netherlands.

64 Legendre, P. & De Cáceres, M. 2013. Beta diversity as the variance of community data:

65 dissimilarity coefficients and partitioning. *Ecology Letters* 16: 951–963.

66 Minchin, P.R. 1987. Simulation of multidimensional community patterns: towards a

67 comprehensive model. *Vegetatio* 71: 145–156.

68 Whittaker, R.H. 1952. A study of summer foliage insect communities in the Great Smoky

69 Mountains. *Ecological Monographs* 22:1–44.

70

71 **Appendix S1: Table S1**

72 The parameter setting of ten 2D simulated community data sets with increasing beta diversity.
73 Parameters for *simulated community data* reflect the setting of the model, while those of
74 *sampled data set* indicate average data set parameters calculated from sampling 1000
75 simulated community data sets. *Gamma diversity* is the number of species in the data set;
76 *mean alpha diversity* is the mean number of species in individual samples; *beta diversity* is
77 the variation in the dissimilarity matrix with squared values of Whittaker distance.

No. of comm.	Simulated community data			Sampled data set (average across 1000 simulations)		
	Length of *gradient 1*	Length of *gradient 2*	Number of species	Gamma diversity	Mean alpha diversity	Beta diversity
1	1000	1000	100	99.3±0.9	23.2±1.4	0.621±0.032
2	1000	2000	200	194.6±3.2	25.5±1.3	0.786±0.018
3	1000	3000	300	283.7±6.2	26.2±1.2	0.853±0.013
4	1000	4000	400	364.7±9.3	26.6±1.1	0.888±0.009
5	1000	5000	500	436.4±13.5	26.8±1.1	0.909±0.008
6	1000	6000	600	500.8±18.0	26.9±1.1	0.924±0.007
7	1000	7000	700	557.8±21.2	27.0±1.0	0.934±0.007
8	1000	8000	800	610.4±24.5	27.2±1.0	0.942±0.006
9	1000	9000	900	653.8±27.8	27.2±1.0	0.948±0.006
10	1000	10000	1000	695.5±31.0	27.2±1.0	0.954±0.005

78

79

(a) No of communities = 1

(b) No of communities = 5

(c) No of communities = 10

80
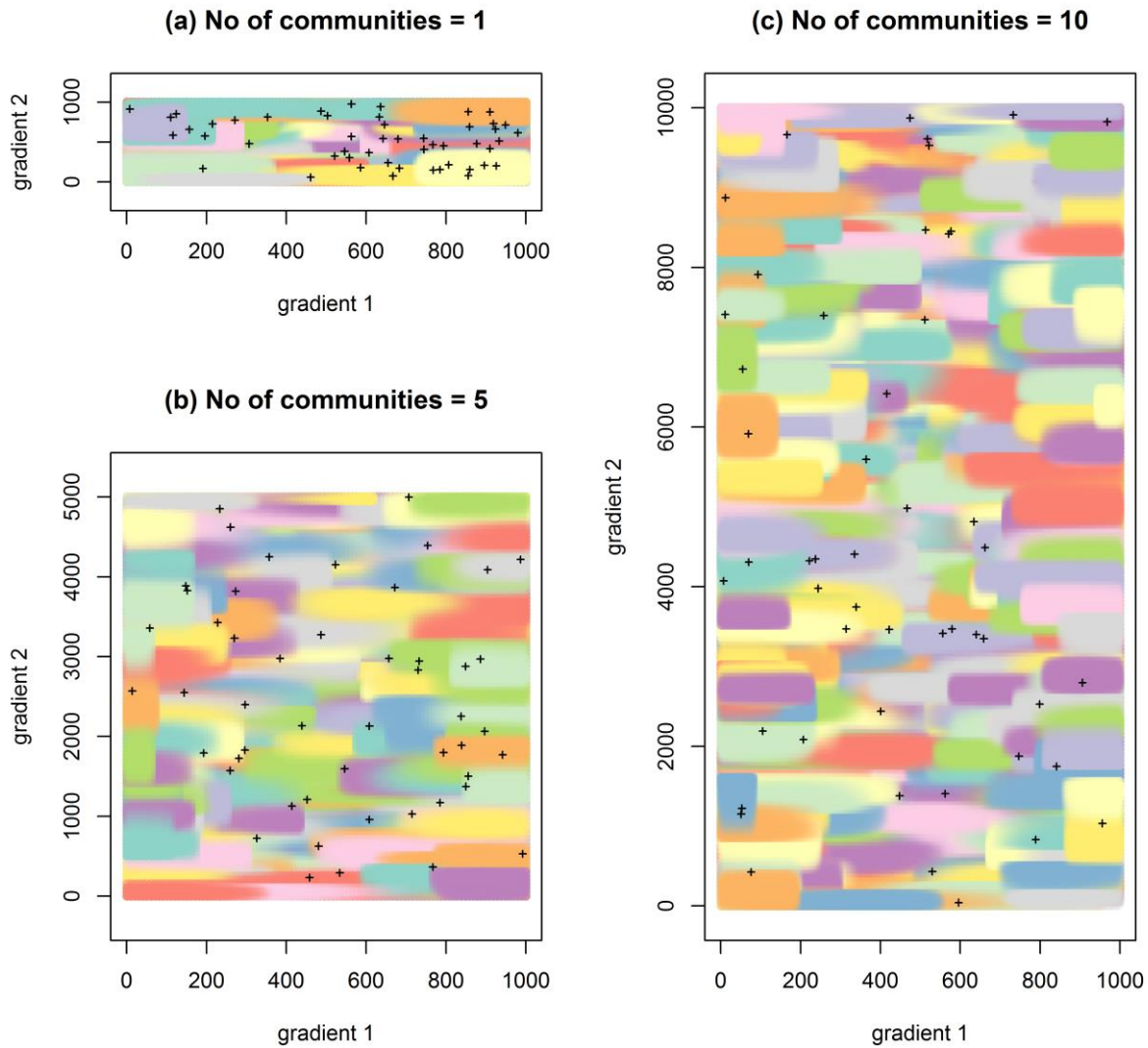
Appendix S1: Figure S1

Visualization of the model creating 2D simulated community data set, namely ecospace defined by two unevenly long virtual environmental gradients (*gradient 1*, always 1000 units long, and *gradient 2*, with range of lengths 1000–10 000 units). Each color represents a region with highest probability of occurrence of particular species, and each symbol (cross) indicates the sampling location in the ecospace (the realization of one permutation). The models of three community data sets with different beta diversity are shown, with numbers of communities equal to a) one, b) five and c) ten (with each community equal to 1000 units of *gradient 2*).

90