Oikos

OIK-02838

Blanchet, F. G., Legendre, P. and He, F. 2015. A new cost-effective approach to survey ecological communities. – Oikos doi: 10.1111/oik.02838

Appendix 1

Simulation results of the multivariate correlation between increasingly precise partial-abundance community data and the complete-abundance community data using raw data

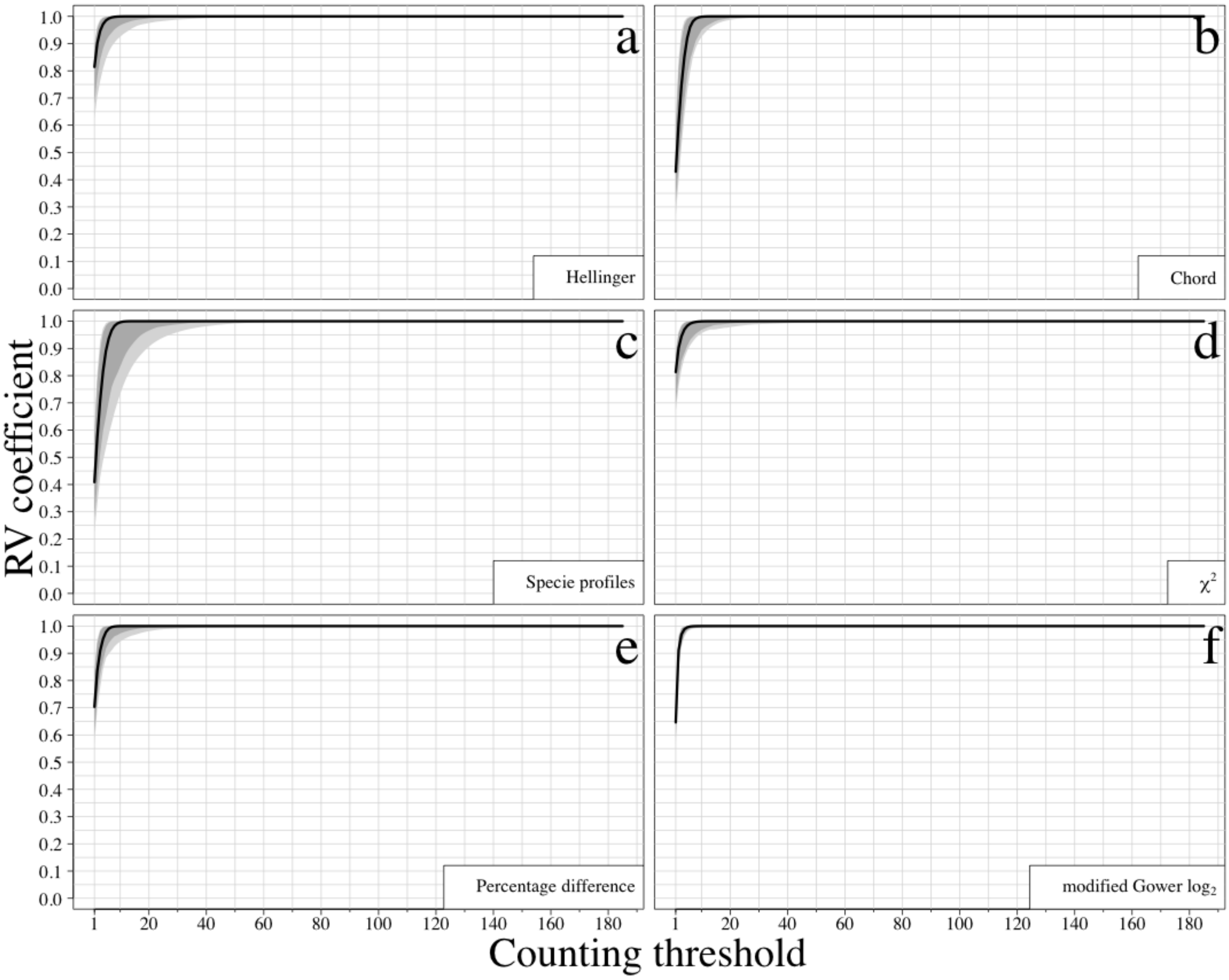

Figure A1.1. Simulation results of the multivariate correlation (using RV coefficient) between increasingly precise partial-abundance community data and the complete-abundance community data using raw data. The counting threshold (abscissa) is the maximum number of individuals counted for a species within a sampling unit. The ordinate represents the RV coefficients. The range spatial distribution of the simulated species was broad and the abundance of each species was sampled from a lognormal distribution. Each panel presents the results for a specific distance. Light grey areas represent the 99% empirical confidence intervals of the simulation results (constructed using the 5th and 995th largest RV coefficients associated to each increasingly precise partial-abundance data, over 1000 simulations), the dark grey areas the 95% empirical confidence intervals (constructed using the 25th and 975th largest RV coefficients associated to each increasingly precise partial-abundance data), and the black lines are the medians per count threshold value.

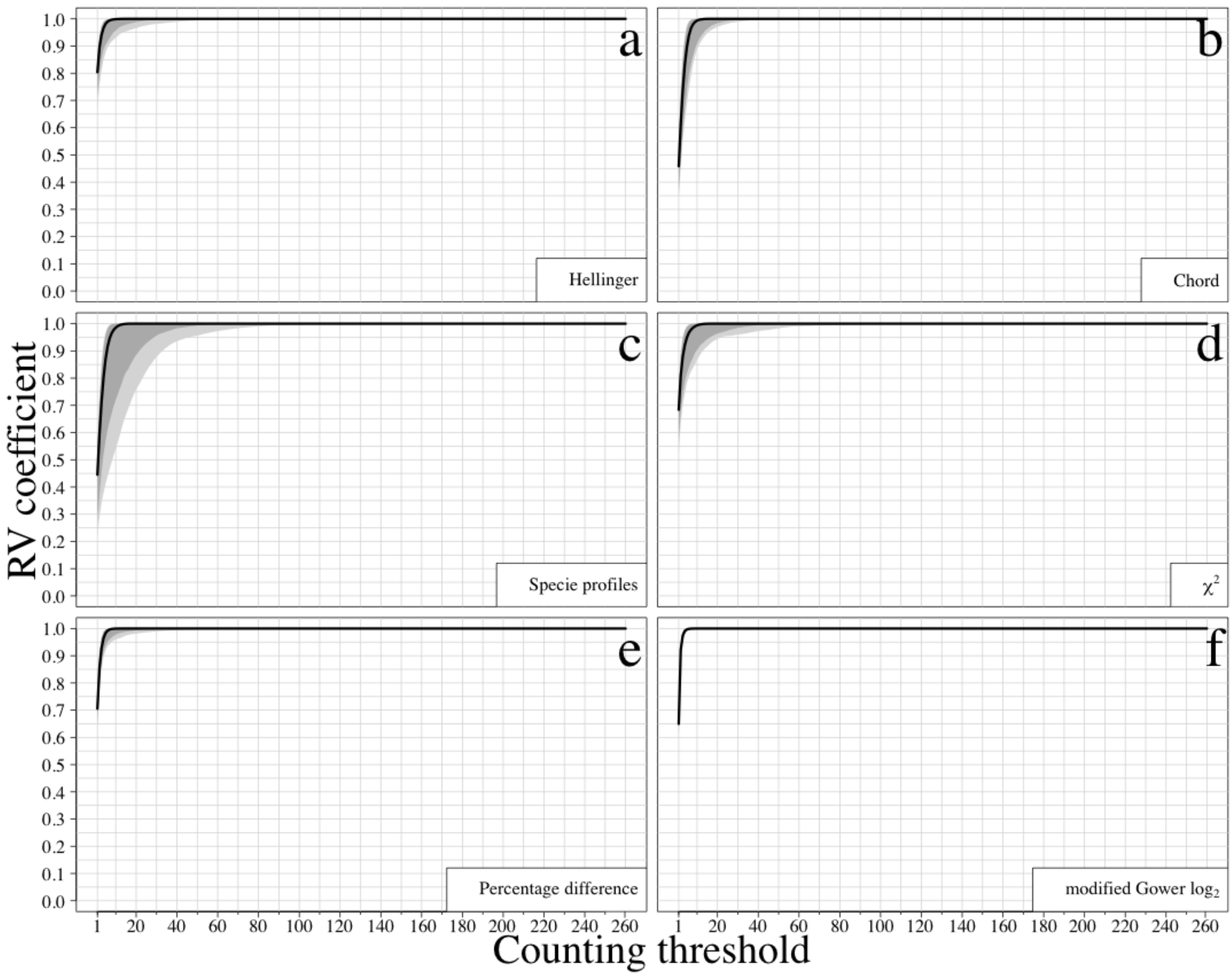

Figure A1.2. Simulation results of the multivariate correlation (using RV coefficient) between increasingly precise partial-abundance community data and the complete-abundance community data using raw data. The counting threshold (abscissa) is the maximum number of individuals counted for a species within a sampling unit. The ordinate represents the RV coefficients. The simulated species were highly aggregated and the abundance of each species was sampled from a lognormal distribution. Each panel presents the results for a specific distance. Light grey areas represent the 99% empirical confidence intervals of the simulation results (constructed using the 5th and 995th largest RV coefficients associated to each increasingly precise partial-abundance data, over 1000 simulations), the dark grey areas the 95% empirical confidence intervals (constructed using the 25th and 975th largest RV coefficients associated to each increasingly precise partial-abundance data), and the black lines are the medians per count threshold value.

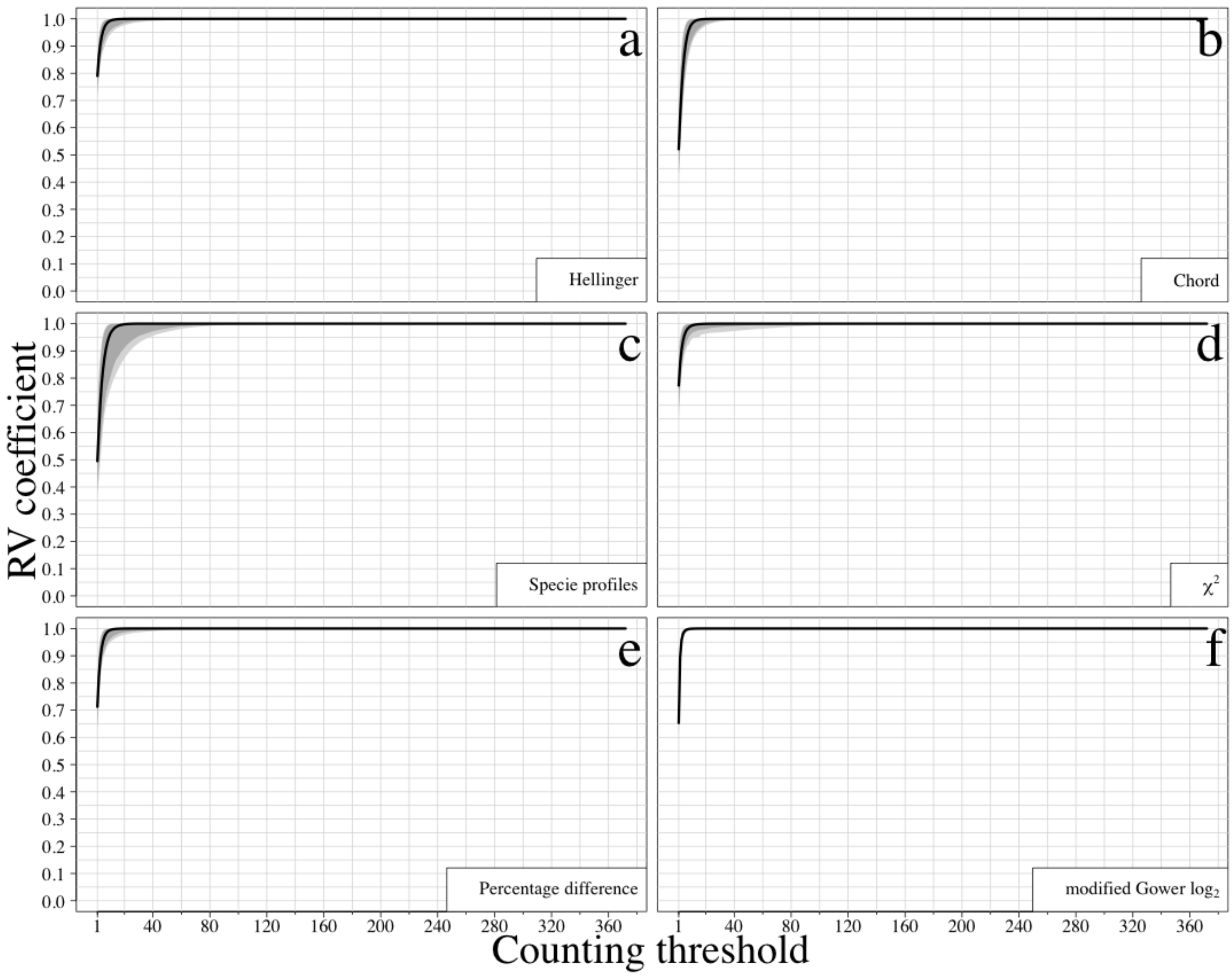

Figure A1.3. Simulation results of the multivariate correlation (using RV coefficient) between increasingly precise partial-abundance community data and the complete-abundance community data using raw data. The counting threshold (abscissa) is the maximum number of individuals counted for a species within a sampling unit. The ordinate represents the RV coefficients. The range spatial distribution of the simulated species was broad and the abundance of each species was sampled from a broken-stick model. Each panel presents the results for a specific distance. Light grey areas represent the 99% empirical confidence intervals of the simulation results (constructed using the 5th and 995th largest RV coefficients associated to each increasingly precise partial-abundance data, over 1000 simulations), the dark grey areas the 95% empirical confidence intervals (constructed using the 25th and 975th largest RV coefficients associated to each increasingly precise partial-abundance data), and the black lines are the medians per count threshold value.

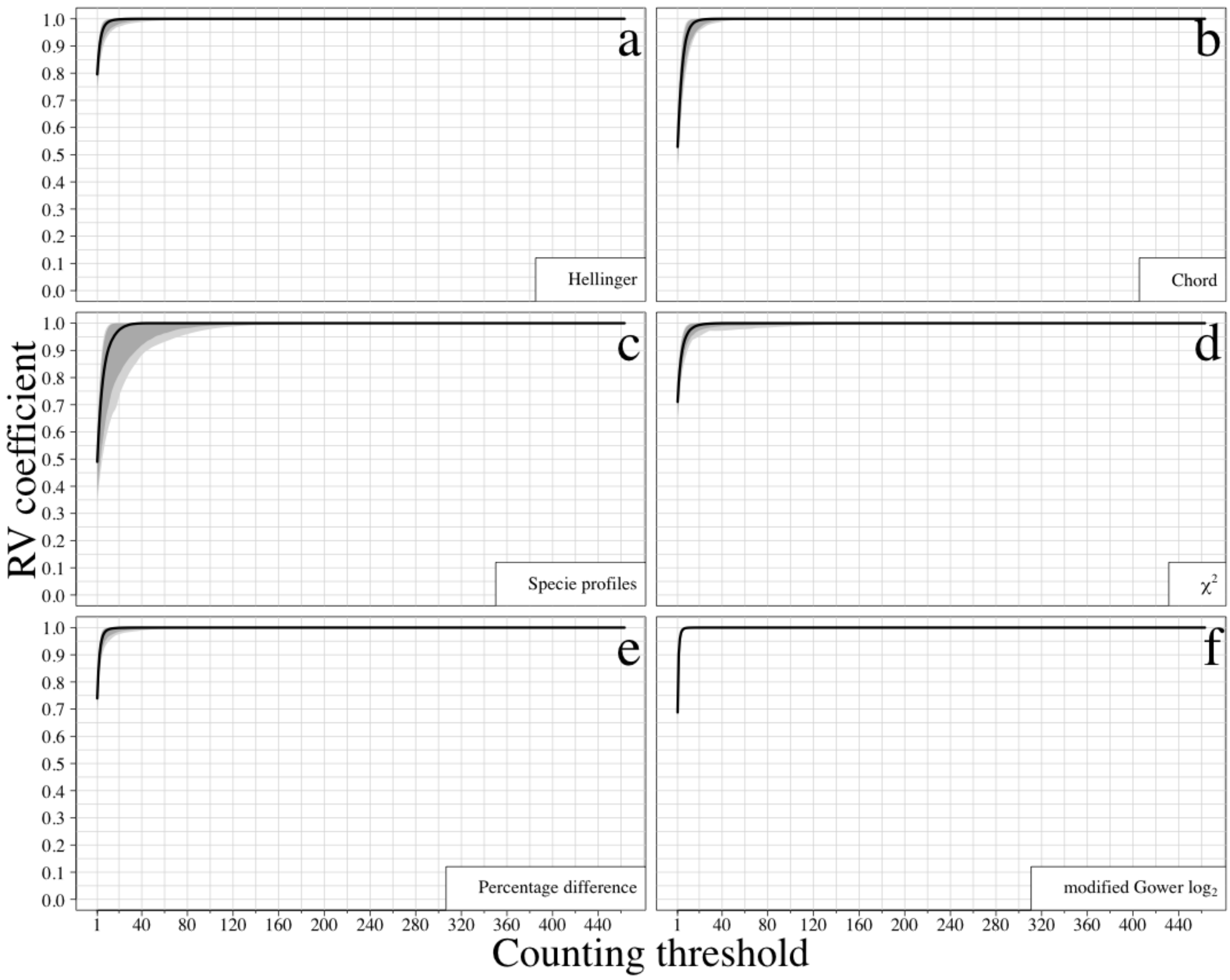

Figure A1.4. Simulation results of the multivariate correlation (using RV coefficient) between increasingly precise partial-abundance community data and the complete-abundance community data using raw data. The counting threshold (abscissa) is the maximum number of individuals counted for a species within a sampling unit. The ordinate represents the RV coefficients. The simulated species were highly aggregated and the abundance of each species was sampled from a broken-stick model. Each panel presents the results for a specific distance. Light grey areas represent the 99% empirical confidence intervals of the simulation results (constructed using the 5th and 995th largest RV coefficients associated to each increasingly precise partial-abundance data, over 1000 simulations), the dark grey areas the 95% empirical confidence intervals (constructed using the 25th and 975th largest RV coefficients associated to each increasingly precise partial-abundance data), and the black lines are the medians per count threshold value.

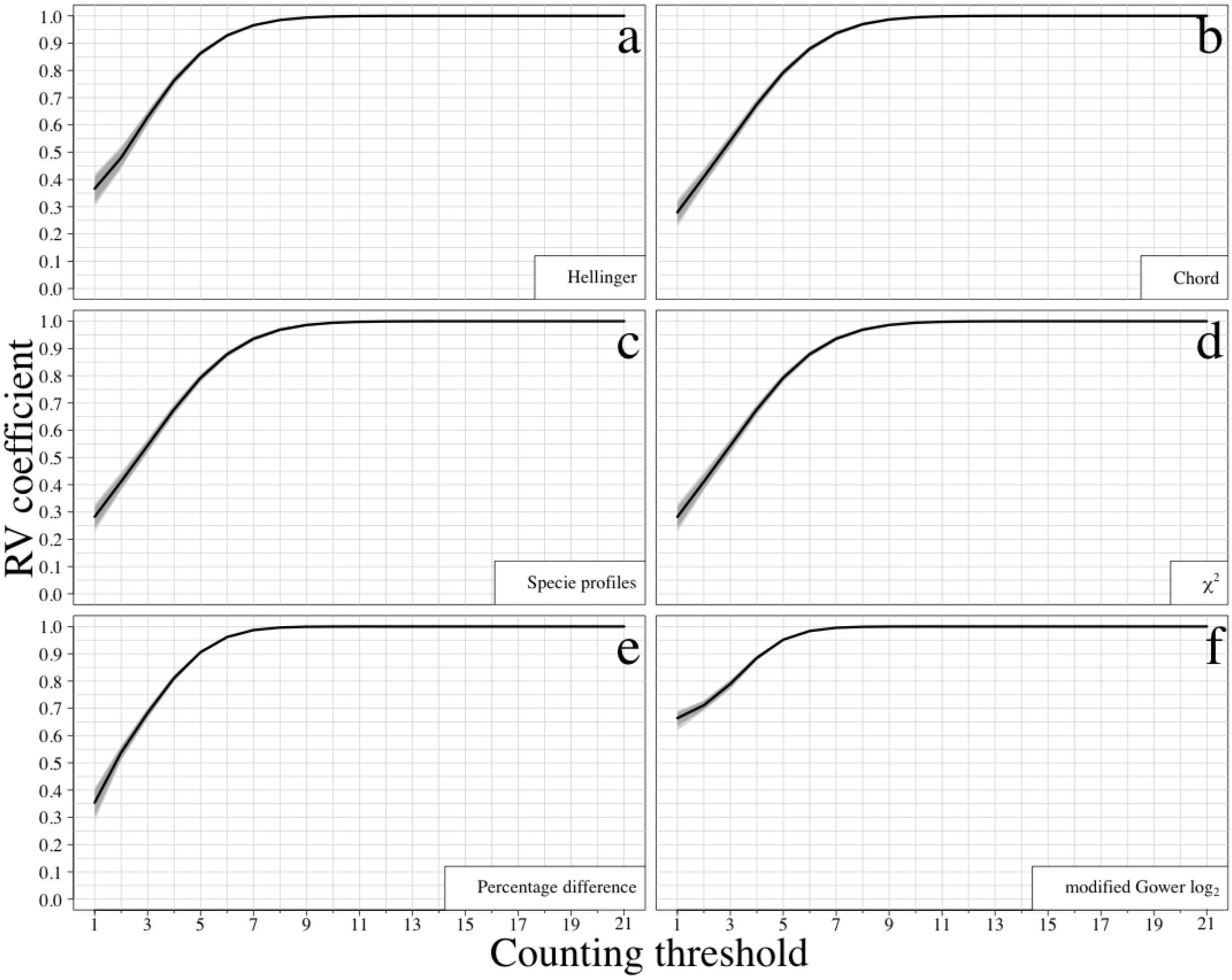

Figure A1.5. Simulation results of the multivariate correlation (using RV coefficient) between increasingly precise partial-abundance community data and the complete-abundance community data using raw data. The counting threshold (abscissa) is the maximum number of individuals counted for a species within a sampling unit. The ordinate represents the RV coefficients. Five hundred individuals of each species were randomly distributed across the sampling area. Each panel presents the results for a specific distance. Light grey areas represent the 99% empirical confidence intervals of the simulation results (constructed using the 5th and 995th largest RV coefficients associated to each increasingly precise partial-abundance data, over 1000 simulations), the dark grey areas the 95% empirical confidence intervals (constructed using the 25th and 975th largest RV coefficients associated to each increasingly precise partial-abundance data), and the black lines are the medians per count threshold value.

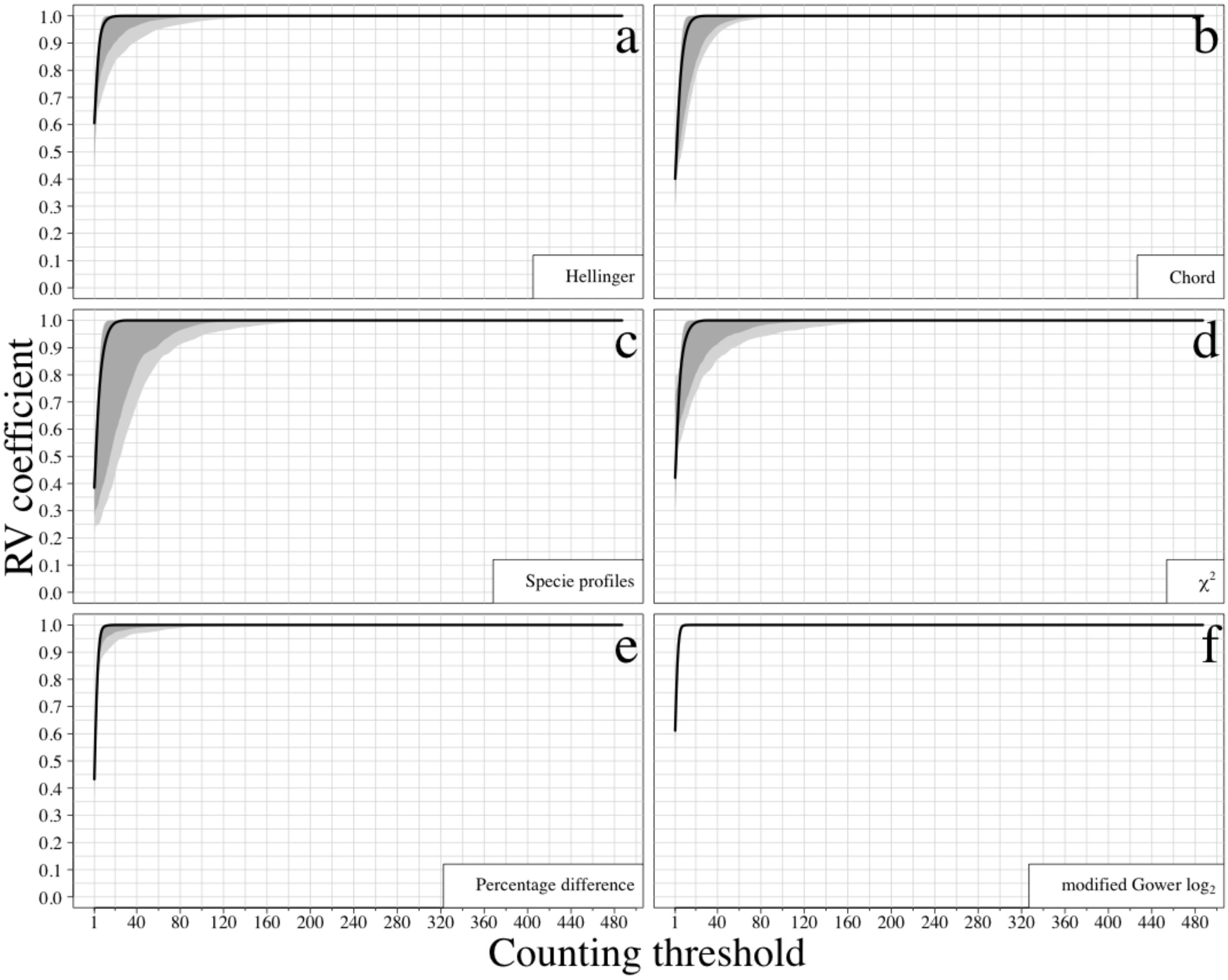

Figure A1.6. Simulation results of the multivariate correlation (using RV coefficient) between increasingly precise partial-abundance community data and the complete-abundance community data using raw data. The counting threshold (abscissa) is the maximum number of individuals counted for a species within a sampling unit. The ordinate represents the RV coefficients. The range spatial distribution of the simulated species was broad and the abundance of each species was roughly 500. Each panel presents the results for a specific distance. Light grey areas represent the 99% empirical confidence intervals of the simulation results (constructed using the 5th and 995th largest RV coefficients associated to each increasingly precise partial-abundance data, over 1000 simulations), the dark grey areas the 95% empirical confidence intervals (constructed using the 25th and 975th largest RV coefficients associated to each increasingly precise partial-abundance data), and the black lines are the medians per count threshold value.

Appendix B

Percentage of sites required in a pilot study to accurately estimate the number of individuals that need to be considered when sampling partial abundances

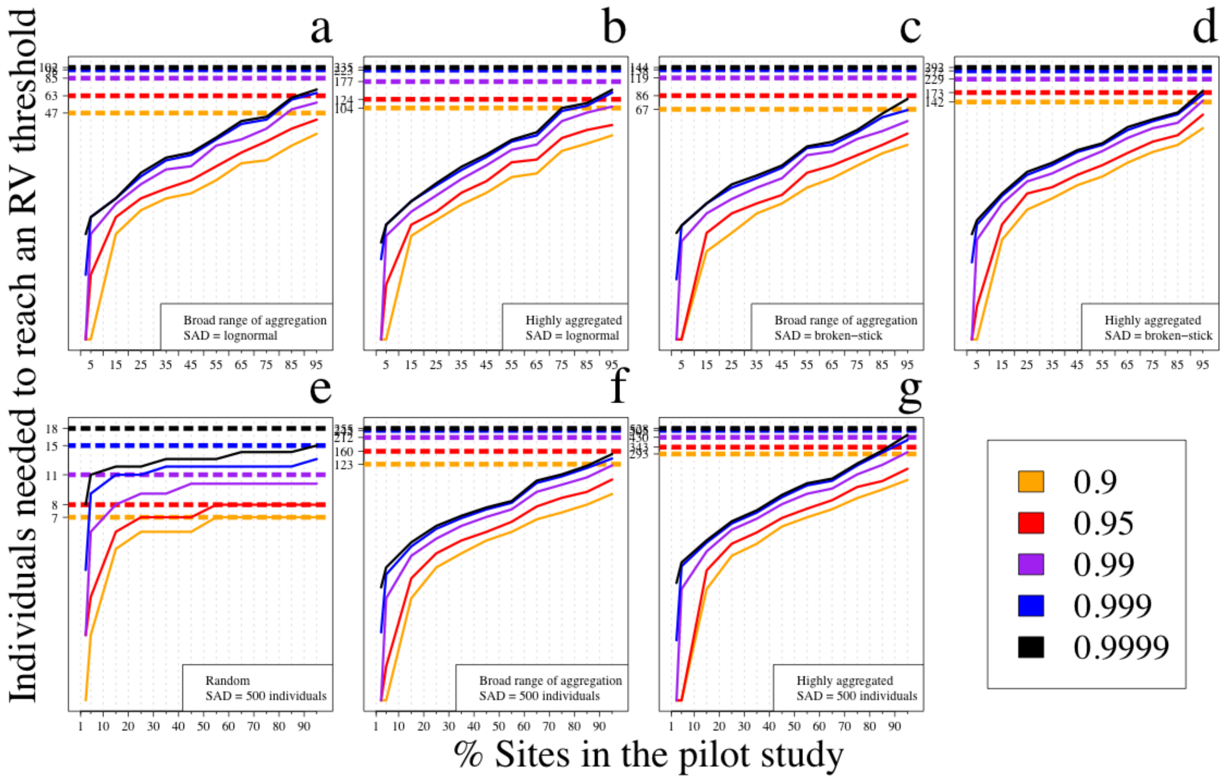

Figure A2.1. Percentage of sites required in a pilot study to accurately estimate the number of individuals that need to be considered when sampling partial abundances. In this figure, we focus on the Euclidean distance and on the number of individuals (ordinate) needed to meet 0.9, 0.95, 0.99, 0.999 and 0.9999 RV coefficients calculated between partial and complete-abundance data. Each panel represents one set of simulated communities. The survey-wide RV coefficients are represented by dotted lines. They are the lower bounds of the 99% confidence intervals of the simulations results presented in Fig. 2. The full lines represent the RV coefficient between partial and complete-abundance calculated using pilot study data. To obtain the pilot study RV coefficient, the sampling units were ordered to form an abundance gradient, from the ones that contained the lowest maximum counts of individuals for any one species to the ones that presented the largest counts. Following this order, the sites were sequentially included in the pilot study. The ordinates were log-transformed for visual clarity. The counting threshold is the maximum number of individuals counted for a species within a sampling unit.

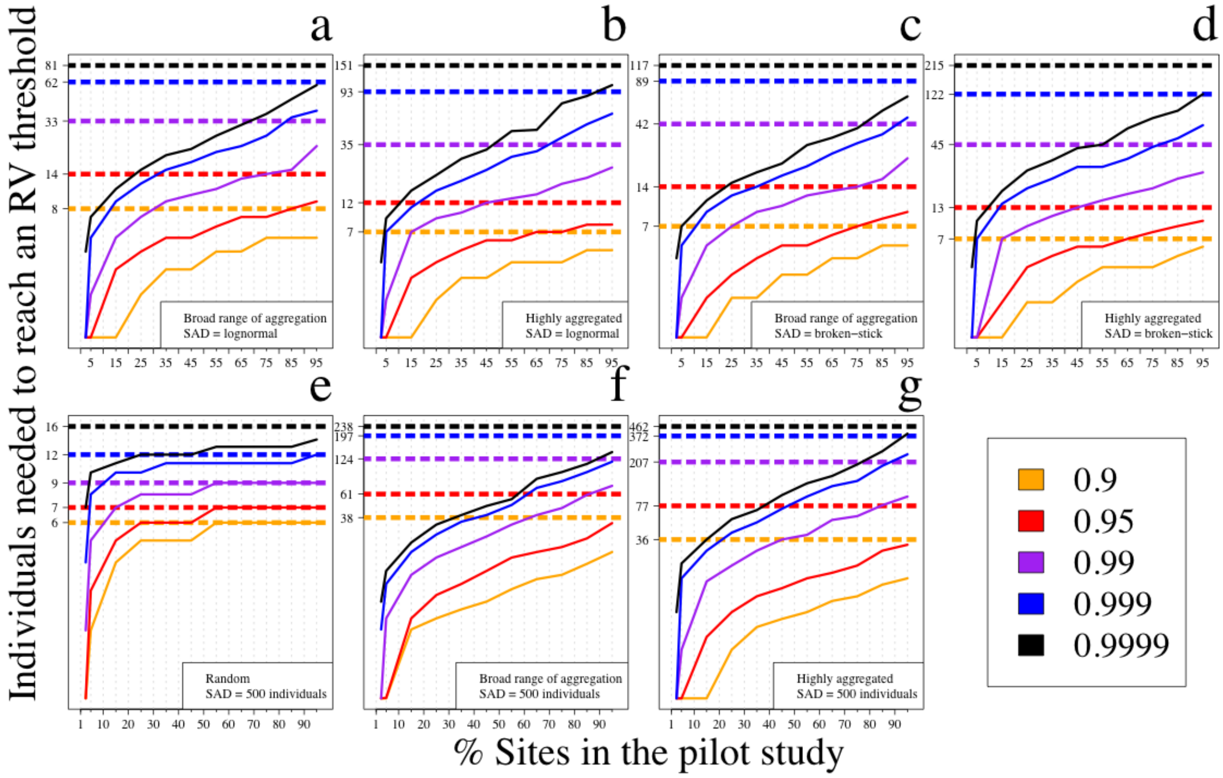

Figure A2.2. Percentage of sites required in a pilot study to accurately estimate the number of individuals that need to be considered when sampling partial abundances. In this figure, we focus on the Hellinger distance and on the number of individuals (ordinate) needed to meet 0.9, 0.95, 0.99, 0.999 and 0.9999 RV coefficients calculated between partial and complete-abundance data. Each panel represents one set of simulated communities. The survey-wide RV coefficients are represented by dotted lines. They are the lower bounds of the 99% confidence intervals of the simulations results presented in Fig. 2. The full lines represent the RV coefficient between partial and complete-abundance calculated using pilot study data. To obtain the pilot study RV coefficient, the sampling units were ordered to form an abundance gradient, from the ones that contained the lowest maximum counts of individuals for any one species to the ones that presented the largest counts. Following this order, the sites were sequentially included in the pilot study. The ordinates were log-transformed for visual clarity. The counting threshold is the maximum number of individuals counted for a species within a sampling unit.

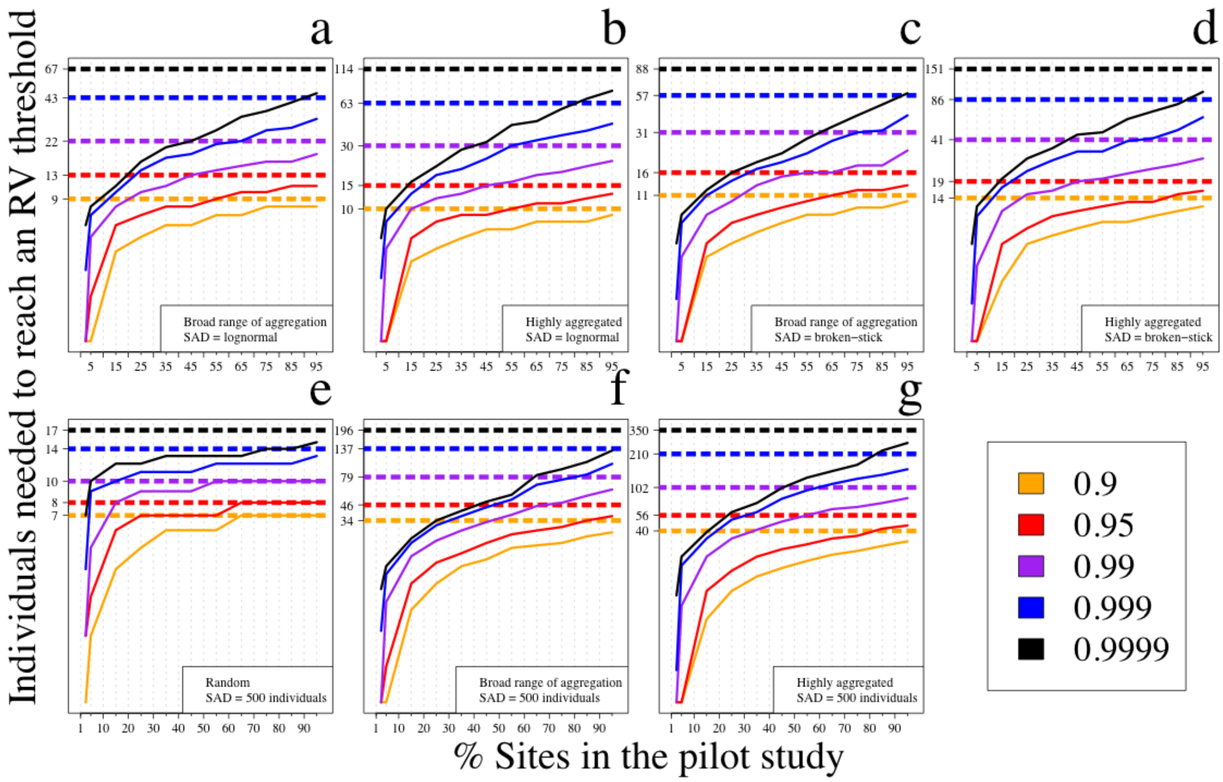

Figure A2.3. Percentage of sites required in a pilot study to accurately estimate the number of individuals that need to be considered when sampling partial abundances. In this figure, we focus on the chord distance and on the number of individuals (ordinate) needed to meet 0.9, 0.95, 0.99, 0.999 and 0.9999 RV coefficients calculated between partial and complete-abundance data. Each panel represents one set of simulated communities. The survey-wide RV coefficients are represented by dotted lines. They are the lower bounds of the 99% confidence intervals of the simulations results presented in Fig. 2. The full lines represent the RV coefficient between partial and complete-abundance calculated using pilot study data. To obtain the pilot study RV coefficient, the sampling units were ordered to form an abundance gradient, from the ones that contained the lowest maximum counts of individuals for any one species to the ones that presented the largest counts. Following this order, the sites were sequentially included in the pilot study. The ordinates were log-transformed for visual clarity. The counting threshold is the maximum number of individuals counted for a species within a sampling unit.

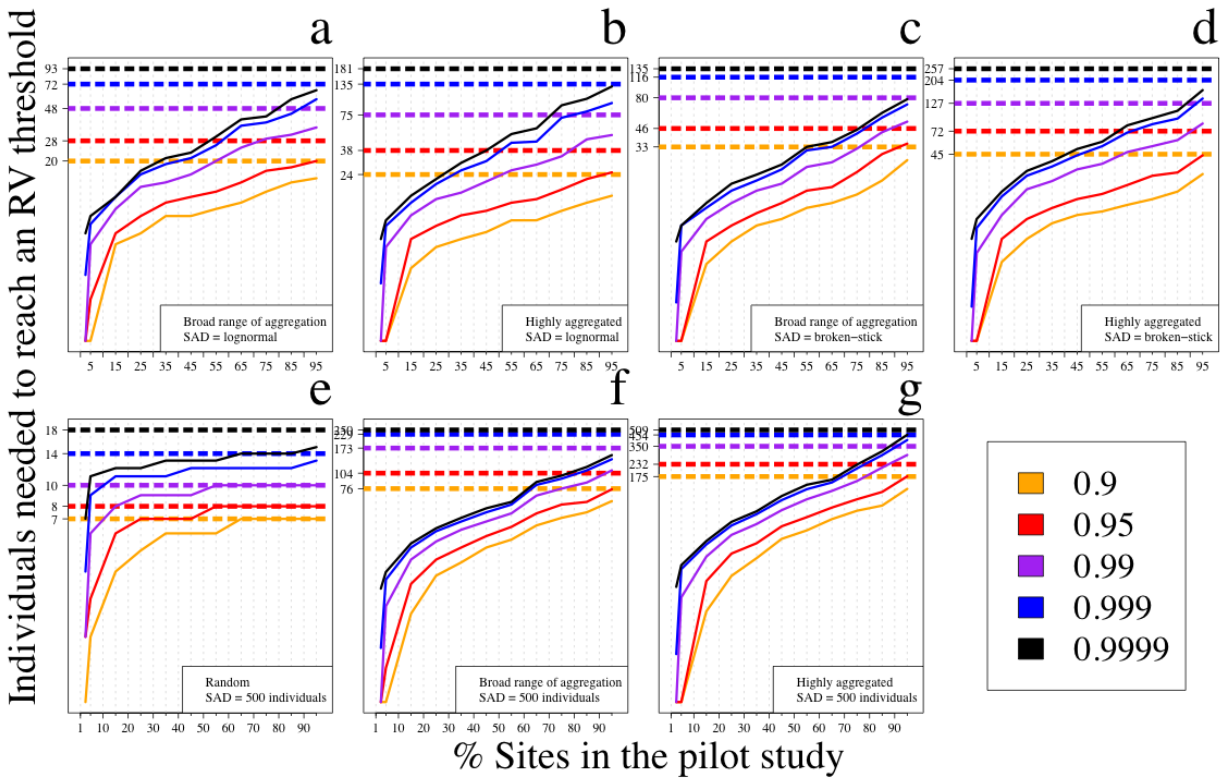

Figure A2.4. Percentage of sites required in a pilot study to accurately estimate the number of individuals that need to be considered when sampling partial abundances. In this figure, we focus on the distance between species profiles and on the number of individuals (ordinate) needed to meet 0.9, 0.95, 0.99, 0.999 and 0.9999 RV coefficients calculated between partial and complete-abundance data. Each panel represents one set of simulated communities. The survey-wide RV coefficients are represented by dotted lines. They are the lower bounds of the 99% confidence intervals of the simulations results presented in Fig. 2. The full lines represent the RV coefficient between partial and complete-abundance calculated using pilot study data. To obtain the pilot study RV coefficient, the sampling units were ordered to form an abundance gradient, from the ones that contained the lowest maximum counts of individuals for any one species to the ones that presented the largest counts. Following this order, the sites were sequentially included in the pilot study. The ordinates were log-transformed for visual clarity. The counting threshold is the maximum number of individuals counted for a species within a sampling unit.

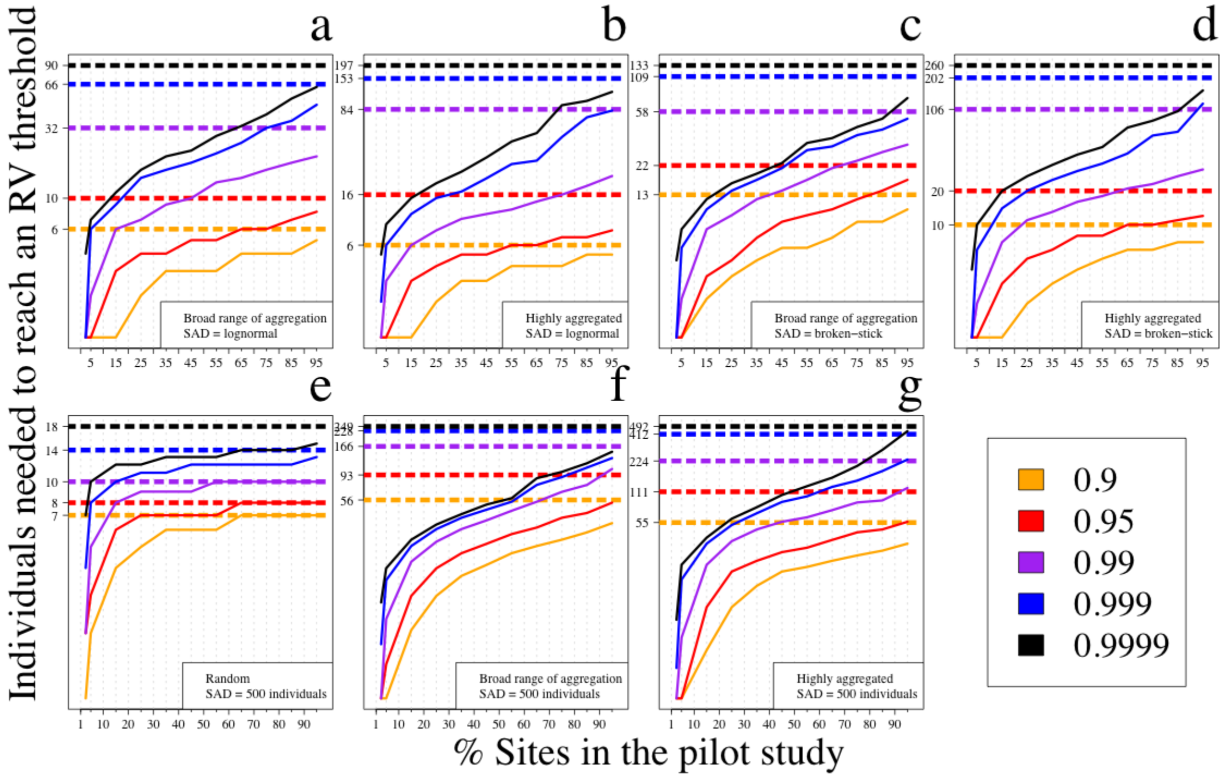

Figure A2.5. Percentage of sites required in a pilot study to accurately estimate the number of individuals that need to be considered when sampling partial abundances. In this figure, we focus on the χ^2-distance and on the number of individuals (ordinate) needed to meet 0.9, 0.95, 0.99, 0.999 and 0.9999 RV coefficients calculated between partial and complete-abundance data. Each panel represents one set of simulated communities. The survey-wide RV coefficients are represented by dotted lines. They are the lower bounds of the 99% confidence intervals of the simulations results presented in Fig. 2. The full lines represent the RV coefficient between partial and complete-abundance calculated using pilot study data. To obtain the pilot study RV coefficient, the sampling units were ordered to form an abundance gradient, from the ones that contained the lowest maximum counts of individuals for any one species to the ones that presented the largest counts. Following this order, the sites were sequentially included in the pilot study. The ordinates were log-transformed for visual clarity. The counting threshold is the maximum number of individuals counted for a species within a sampling unit.

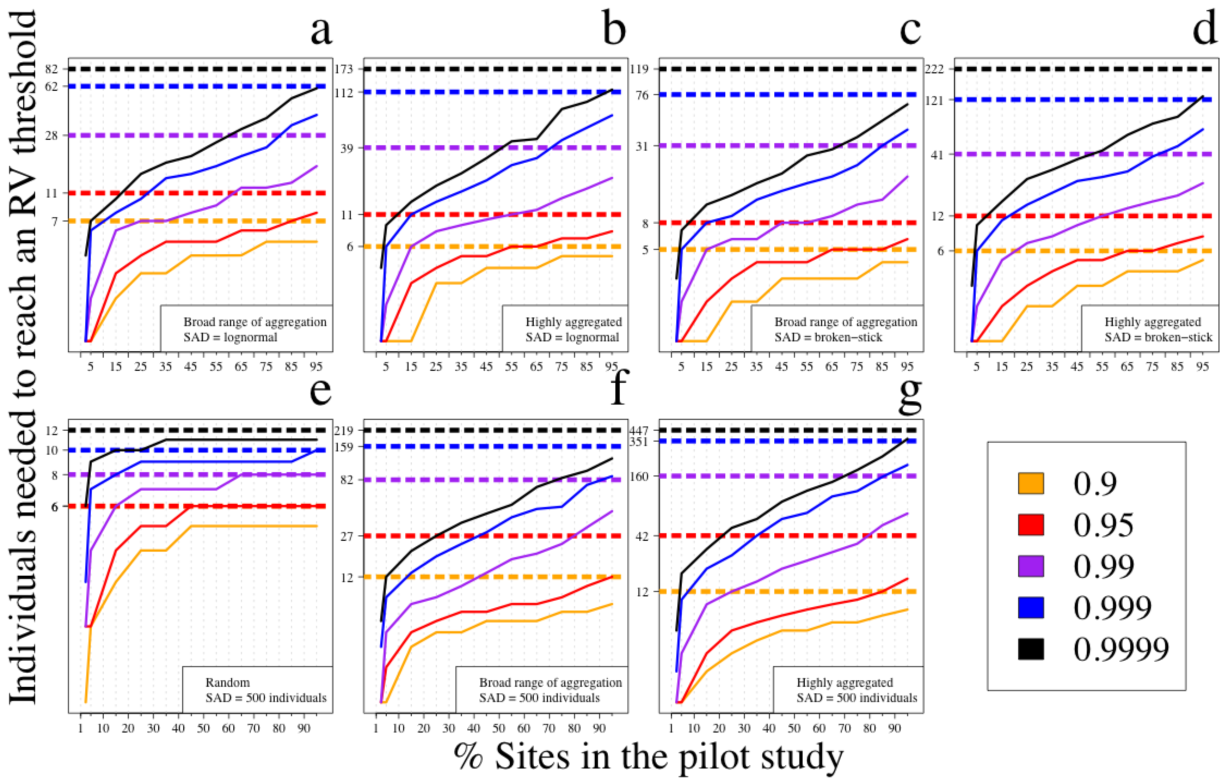

Figure A2.6. Percentage of sites required in a pilot study to accurately estimate the number of individuals that need to be considered when sampling partial abundances. In this figure, we focus on the percentage difference distance and on the number of individuals (ordinate) needed to meet 0.9, 0.95, 0.99, 0.999 and 0.9999 RV coefficients calculated between partial and complete-abundance data. Each panel represents one set of simulated communities. The survey-wide RV coefficients are represented by dotted lines. They are the lower bounds of the 99% confidence intervals of the simulations results presented in Fig. 2. The full lines represent the RV coefficient between partial and complete-abundance calculated using pilot study data. To obtain the pilot study RV coefficient, the sampling units were ordered to form an abundance gradient, from the ones that contained the lowest maximum counts of individuals for any one species to the ones that presented the largest counts. Following this order, the sites were sequentially included in the pilot study. The ordinates were log-transformed for visual clarity. The counting threshold is the maximum number of individuals counted for a species within a sampling unit.

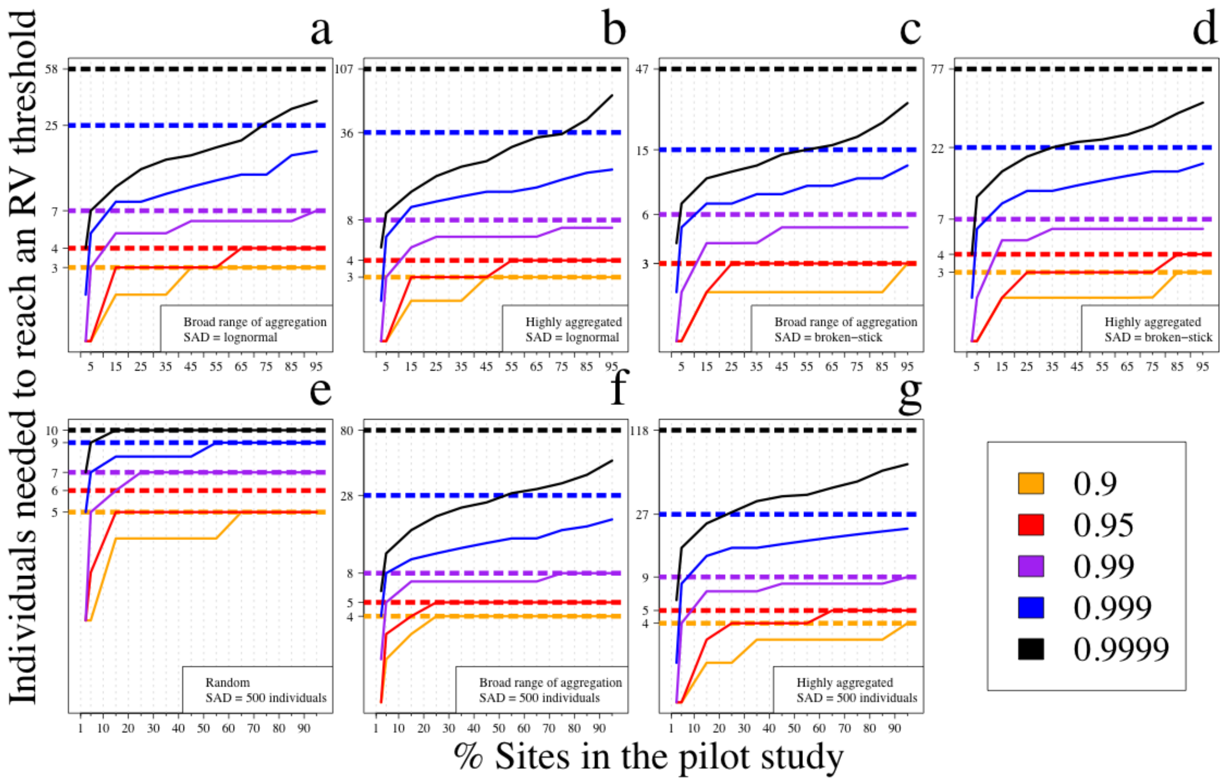

Figure A2.7. Percentage of sites required in a pilot study to accurately estimate the number of individuals that need to be considered when sampling partial abundances. In this figure, we focus on the modified Gower distance (using a based 2 log) and on the number of individuals (ordinate) needed to meet 0.9, 0.95, 0.99, 0.999 and 0.9999 RV coefficients calculated between partial and complete-abundance data. Each panel represents one set of simulated communities. The survey-wide RV coefficients are represented by dotted lines. They are the lower bounds of the 99% confidence intervals of the simulations results presented in Fig. 2. The full lines represent the RV coefficient between partial and complete-abundance calculated using pilot study data. To obtain the pilot study RV coefficient, the sampling units were ordered to form an abundance gradient, from the ones that contained the lowest maximum counts of individuals for any one species to the ones that presented the largest counts. Following this order, the sites were sequentially included in the pilot study. The ordinates were log-transformed for visual clarity. The counting threshold is the maximum number of individuals counted for a species within a sampling unit.