Anne Chao and Lou Jost. 2012. Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. *Ecology* 93:2533−2547. http://dx.doi.org/10.1890/11-1952.1

APPENDIX H. Some basic theoretical properties of coverage-based species accumulation curves

As we show in Fig. 1 of the main report, the theoretical coverage-based SAC plots the expected species richness with respect to expected coverage. That is, we plot $E(S_m)$ with respect to $E(C_m)$. Based on the formulas $E(S_m)$ and $E(C_m)$, which are given respectively in Eqs. 1 and 2 of the main report, we state some theoretical properties for coverage-based SACs below. Then we use examples to demonstrate the general shapes of the coverage-based SACs.

We first define the coefficient of variation (CV) of a set of relative abundances as the ratio of their standard deviation to their mean. The magnitude of CV is used to characterize the degree of unevenness (or heterogeneity) among the relative abundances. The shape of a coverage-based SAC generally depends on the value of CV. For a completely even community, the CV value is 0. The larger the CV, the greater the degree of unevenness.

Some basic properties for coverage-based SACs (Fig. H1):

(1) As we proved in Appendix B, the curve is a non-decreasing function of expected coverage; it starts at the base point (0, 0) and ends at the point (1, S).

(2) Given there are S species, when all species have the same abundances (CV= 0), the curve is a straight line connecting (0, 0) and (1, S). The straight line shows that x% of species cover x% of community's individuals. This is like the "perfect equality line" in a Lorenz curve. (The basic concept of our coverage-based SAC is different from a Lorenz curve, which is a curve to measure inequality; see Gastwirth (1972) for a review. This is because we evaluate expected values in both axes whereas in a Lorenz curve both axes represent the cumulative percentages.)

(3) Given there are S finite species, when one species dominates and all the others have vanishingly small abundances (CV tends to infinity), then the curve gives values 0 when the expected coverage < 1 and jumps to a value of unity when the expected coverage is one. That is, the SAC includes one horizontal line and one vertical line. This is like the "perfect inequality line" in a Lorenz curve.

(4) For other cases, the curve is between the perfect equality line and the perfect inequality line, as will be shown in the following examples.

(5) Any point $(x^*, y^*) = (E(C_m), E(S_m))$ on the coveraged-based SAC (Fig. H1) can be interpreted as follows: when there are y^* species in a sample, on average, this sample will cover x^* of community individuals. It can also be interpreted as: x^* individuals, on average, contain y^* species. Therefore, coverage-based SACs can reveal community structure (evenness) in an average sense. For the special case of $y^* = 1$, the corresponding X-coordinate is $E(C_1) = \sum_{i=1}^{S} p_i^2$, which is the Simpson concentration. This implies that the coverage-based curve can also readily reveal the Simpson concentration.

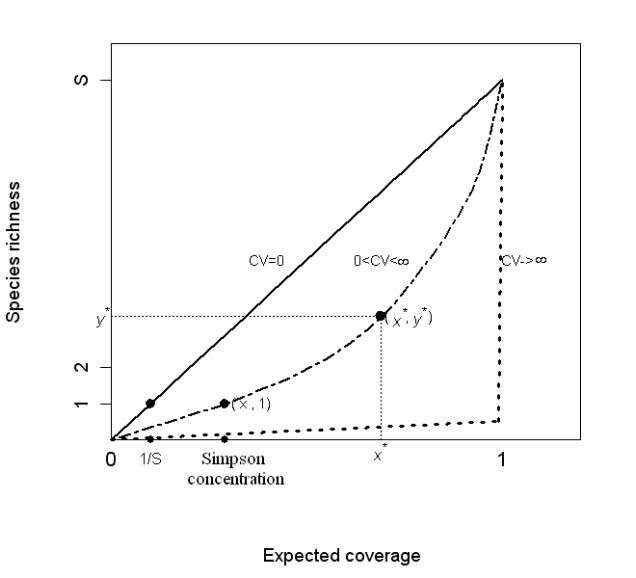

Fig. H1: Coverage-based SAC when species richness *S* is fixed. The straight solid line (perfect equality line) corresponds to a community with equal relative abundances (CV = 0). The dotted line (perfect inequality line) corresponds to a community with one dominant relative abundance and the others vanishingly small ($CV \to \infty$). The SAC of any other community is between the two extreme cases. See text.

We now consider three simple communities, whose size- and coverage-based SACs are shown in Figs. H2 and H3 respectively.

Community 1 (5 species, completely even, CV = 0): species relative abundance $p_i = 0.2$ for $i = 1, 2,\ldots, 5$. The coverage-based curve for an even community is a straight line as shown in the plot below (Fig. H3, solid line). The Simpson concentration is 0.20 (i.e., the X-coordinate when $y = 1$).

Community 2 (5 species, moderately uneven, CV = 0.77): species relative abundance {0.5, 0.2, 0.1×3} (i.e., there is one species with relative abundance 0.5, one species with relative abundance 0.2, and three species with relative abundance 0.1.) The coverage-based curve reveals that, on average, one species covers approximately 32% of community's individuals (Fig. H3, dashed line); two species cover approximately 59% of community's individuals. A similar interpretation can be given for any point in the curve. The Simpson concentration is 0.32.

Community 3 (5 species, highly uneven, CV = 1.90): species relative abundance {0.96, 0.01×4}. The coverage-based curve reveals that, on average, one species covers approximately 92% of community's individuals (Fig. H3, dotted line); two species cover approximately 97% of community's individuals; three species cover approximately 98% of community's

individuals. A similar interpretation can be given for any point in the curve. The Simpson concentration is 0.92.

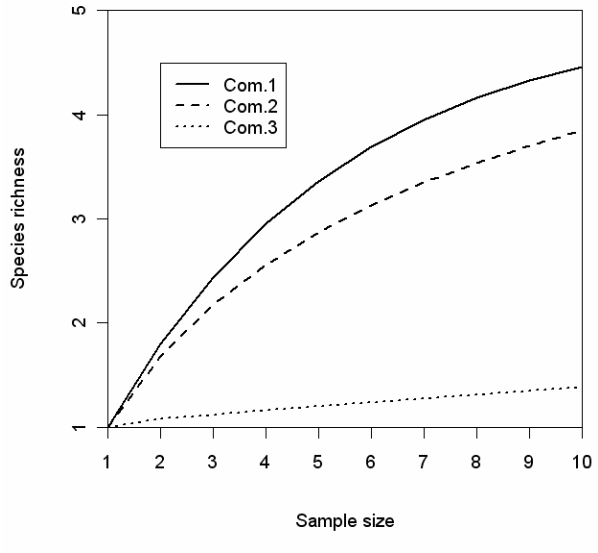

Fig. H2: Size-based SAC

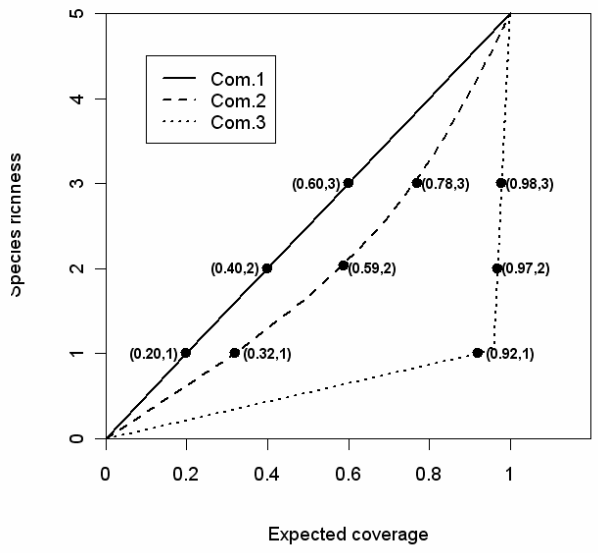

Fig. H3: Coverage-based SAC

We next compare four communities with 100 species, whose size- and coverage-based SACs are shown in Figs. H4 and H5 respectively.

Community 1 (100 species, completely even, CV = 0): species relative abundance $p_i = 0.01$ for i = 1, 2,…, 100.

Community 2 (100 species, moderately uneven, CV = 0.57): species relative abundance {0.02×22, 0.01×28, 0.006×40, 0.004×10} (i.e., there are 22 species with relative abundance 0.02; 28 species with relative abundance 0.01, …etc.)

Community 3 (100 species, highly uneven, CV = 1.95): species relative abundance

{0.1×2, 0.08×2, 0.06×2, 0.04×2, 0.004×90}.

Community 4 (100 species, very highly uneven, CV = 3.30): species relative abundance

{0.24, 0.16, 0.14, 0.1, 0.04×3, 0.02×3, 0.002×90}.

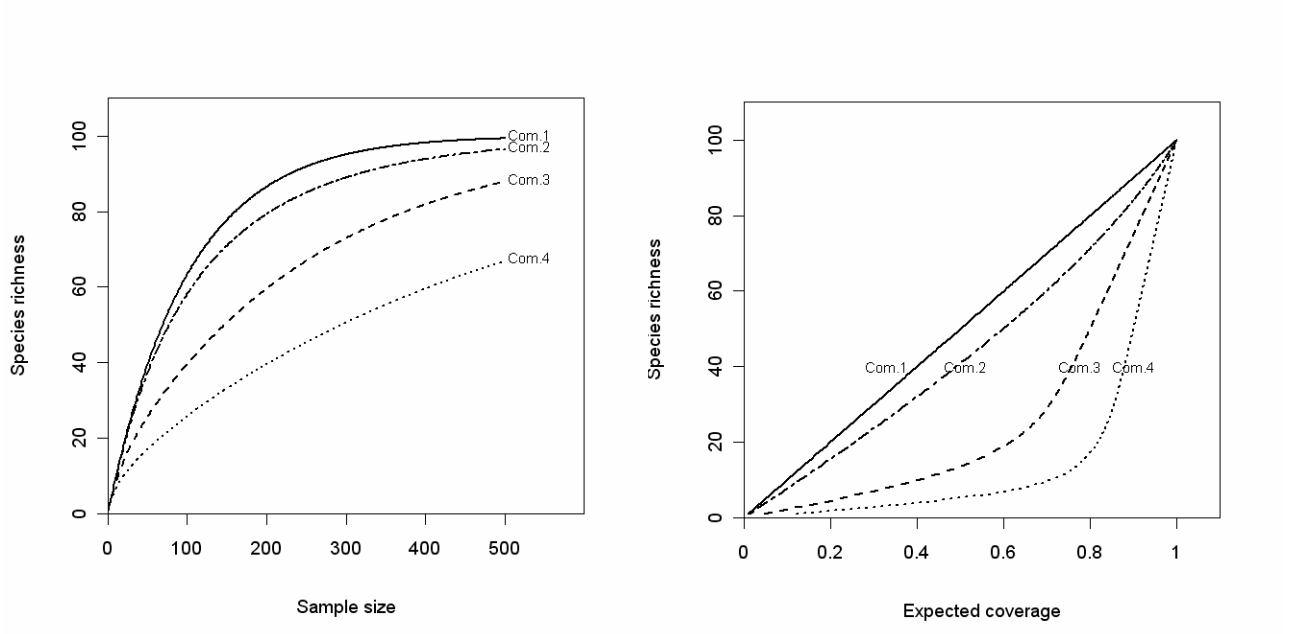

Fig. H4: Size-based SAC

Fig. H5: Coverage-based SAC

The size-based and coverage-based SACs provide different information. The size-based SACs in Figs. H2 and H4 provide sampling information about how species richness increases with sampling efforts in each community; a steep slope at any size indicates a higher probability of discovering new species whereas a flat slope implies low chance to find new species. The initial slope of each curve quantifies the Gini-Simpson index of the corresponding community (Lande et al. 2000).

When we compare size-based SACs among multiple communities, we standardize sample size, which is determined by the investigator, not the communities. In contrast, in coverage-based

SACs, we standardize sample coverage, a community-level characteristic. As shown in Figs. H3 and H5, coverage-based SACs reveal community structure including dominance (or evenness) information and the Simpson concentration. Also, the curve is bounded by two simple reference lines (perfect equality and inequality lines). If the curve is close to the perfect equality line, implying the degree of heterogeneity among species relative abundances is low. If the curve is close to the perfect inequality line, implying the degree of heterogeneity among species relative abundances is high. Thus, it provides more insights and robust comparison among multiple communities. (The size-based SACs lack two simple reference lines.)

LITERATURE CITED

Gastwirth, J. L. 1972. The estimation of the Lorenz curve and Gini index. The review of Economics and Statistics 54:306-316.

Lande, R., P. J. DeVries, and T. R. Walla. 2000. When species accumulation curves intersect: implications for ranking diversity using small samples. Oikos 89:601-605.