

# Comments

*Ecology*, 94(3), 2013, pp. 761–763  
© 2013 by the Ecological Society of America

## Species–area relationships always overestimate extinction rates from habitat loss: comment

JACOB BOCK AXELSEN,<sup>1</sup> URI ROLL,<sup>1</sup> LEWI STONE,<sup>1</sup> AND ANDREW SOLOW<sup>2,3</sup>

The species–area relationship summarizes the relationship between the average number of species in a region and its area. This relationship provides a basis for predicting the loss of species associated with loss of habitat (e.g., Pimm and Raven 2000). The approach involves two steps. First, as discussed in more detail below, the species–area relationship is used to predict the number of species that are endemic to the habitat at risk based on its area. Second, these endemic species are assumed to become extinct should this habitat be lost. In a controversial paper, He and Hubbell (2011) argued that the way in which the species–area relationship is used to predict the number of endemic species is incorrect when individual organisms are aggregated in space and argued that this explains a discrepancy between predicted and observed extinction rates associated with habitat loss. The controversy surrounding the paper focused primarily on the second part of their argument (Brooks 2011, Evans et al. 2011, He and Hubbell 2012, Pereira et al. 2012, Thomas and Williamson 2012). Here, we focus on the details underlying the first part.

Consider an overall region of area  $A$  that contains a total of  $s$  species. Let the random variable  $S(a)$  be the number of species contained in a subregion of area  $a$  within this overall region. The nature of the randomness in  $S(a)$  is discussed below. The species–area relationship (SAR) is defined as the expected value of  $S(a)$ :

$$\text{SAR}(a) = E[S(a)]. \quad (1)$$

It is straightforward to show that

$$\text{SAR}(a) = \sum_{j=1}^s p_j(a) \quad (2)$$

Manuscript received 11 January 2012; revised 26 April 2012; accepted 20 August 2012. Corresponding Editor: B. D. Inouye.

<sup>1</sup> Biomathematics Unit, Department of Zoology, Tel-Aviv University, Tel-Aviv 69978 Israel.

<sup>2</sup> Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543 USA.

<sup>3</sup> Corresponding author. E-mail: asolow@whoi.edu

where  $p_j(a)$  is the probability that species  $j$  is contained in a subregion of area  $a$ . Similarly, let the random variable  $N(a)$  be the number of species that are endemic to a subregion of area  $a$ . That is, these species are contained in the subregion but not in its complement, which has area  $A - a$ . These species are only assumed to be endemic with respect to the overall region and may occur outside it. The endemic–area relationship (EAR) is defined as the expected value of  $N(a)$ :

$$\text{EAR}(a) = E[N(a)]. \quad (3)$$

It is again straightforward to show that

$$\text{EAR}(a) = \sum_{j=1}^s q_j(a) \quad (4)$$

where  $q_j(a)$  is the probability that species  $j$  is endemic to a subregion of area  $a$ .

The use of the species–area relationship to predict the number of endemic species in a subregion is based on the following equation:

$$\text{EAR}(a) = s - \text{SAR}(A - a). \quad (5)$$

The central claim of He and Hubbell (2011:368) is that “[o]nly in a very special and biologically unrealistic case, when all species are randomly and independently distributed in space, is it possible to derive the EAR from the SAR.” This claim seems counterintuitive. By simple bookkeeping, for *any* fixed spatial distribution of individuals, the number of species endemic to *any* fixed subregion of area  $a$  must be *exactly* equal to the difference between  $s$  and the number of species contained in the complementary subregion of area  $A - a$ . It is hard to see how averaging these quantities over a stochastic process can disturb this equality. In fact, the following general argument shows that the claim of He and Hubbell (2011) is incorrect. The event that species  $j$  is endemic to a subregion of area  $a$  is equivalent to the event that it is not contained in a subregion of area  $A - a$ . The probability of the former event is  $q_j(a)$  and of the latter is  $1 - p_j(A - a)$ . These probabilities must be equal, ensuring that the equality in Eq. 5 is true. This argument is completely general and, in particular, is not based on an assumption of how individuals are distributed in space.

Before turning to the analytical and empirical results that He and Hubbell (2011) used to support their claim, it is necessary to distinguish between two ways of modeling randomness in  $N(a)$  and  $S(A - a)$ . One way is through the spatial distribution of individuals in the overall region. Here, the subregion of area  $a$  is regarded as fixed and expectation is taken over realizations of a stochastic process generating this spatial distribution.

For example, as He and Hubbell (2011) note, if this process consists of distributing  $m_j$  individuals of species  $j$  independently at random in the overall region, then  $p_j(A - a) = 1 - (a/A)^{m_j}$ ,  $q_j(a) = (a/A)^{m_j}$ , and the equality in Eq. 5 can be verified by substitution into Eq. 2 and Eq. 4.

He and Hubbell (2011) presented some analytical results showing that this result does not hold for a stochastic process in which, conditional on their total  $m_j$ , the number of individuals of species  $j$  within the subregion follows what they called a finite negative binomial distribution. This distribution is identical to the conditional negative binomial distribution described by Conlisk et al. (2007), which itself is identical to the Pólya-Eggenberger distribution (Johnson et al. 1997). This distribution can be used as a (conditional) model for the number of individuals within a subregion that are aggregated in space. Let the random variable  $M_j(a)$  be the number of individuals of species  $j$  contained within a subregion of area  $a$ . The conditional probability mass function of  $M_j(a)$  for this distribution is

$$\text{prob}[M_j(a) = m | m_j] = \frac{F(k, m)F([(A - a)/a]k, m_j - m)}{F([A/a]k, m_j)} \quad (6)$$

where, for any positive real number  $x$  and any nonnegative integer  $y$ ,  $F(x, y) = \Gamma(x + y)/\Gamma(x)y!$  where  $\Gamma$  is the gamma function. Here,  $k$  is a parameter that controls the degree of spatial aggregation of individuals for a subregion of area  $a$ . The essential error of He and Hubbell (2011) was to use the *same* value of the clustering parameter in the conditional distribution of the number of individuals contained in the complementary subregion of area  $A - a$  instead of the correct value  $[(A - a)/a]k$ . Thus, the conditional probability mass function of  $M_j(A - a)$  is

$$\begin{aligned} \text{prob}[M_j(A - a) = m | m_j] \\ = \frac{F([(A - a)/a]k, m)F(k, m_j - m)}{F([A/a]k, m_j)}. \end{aligned} \quad (7)$$

The scaling of the aggregation parameter with area was emphasized by Conlisk et al. (2007).

It follows from Eq. 6 that

$$q_j(a) = \text{prob}[M_j(a) = m_j | m_j] = F(k, m_j)/F([A/a]k, m_j) \quad (8)$$

and from Eq. 7 that

$$\begin{aligned} p_j(A - a) &= 1 - \text{prob}[M_j(A - a) = 0] \\ p_j &= 1 - F(k, m_j)/F([A/a]k, m_j) \end{aligned} \quad (9)$$

ensuring that the equality in Eq. 5 holds.

The second way to model randomness underlying the SAR and the EAR is through randomness in the location of the subregion of area  $a$ . Here, the locations of individuals are regarded as fixed, and expectation is over realizations of the random location of the subregion. This is the basis for traditional nonparametric estimation of the SAR (e.g., Smith et al. 1979) and also for the empirical results of He and Hubbell (2011). It is worth emphasizing that the results that follow are true for any fixed distribution of individuals.

Suppose that the overall region is divided into  $A$  nonoverlapping cells each of unit area. Let  $k_j$  be the number of cells that contain species  $j$ . In the standard model of a random subregion of area  $a$ , all of the  $\binom{A}{a}$  subsets of  $a$  cells (comprising a not necessarily contiguous subregion of area  $a$ ) are equally likely. Under this model,

$$p_j(A - a) = 1 - \binom{A - k_j}{A - a} / \binom{A}{A - a} \quad (10)$$

where  $\binom{x}{y} = 0$  if  $x < y$ . The quantity  $\binom{A - k_j}{A - a}$  is the number of ways to select  $A - a$  cells out of the  $A - k_j$  that do not contain species  $j$ . Turning to the probability  $q_j(a)$  that species  $j$  is endemic to the subregion consisting of  $a$  cells, this subregion must include all  $k_j$  cells that contain this species, leaving  $\binom{A - k_j}{a - k_j}$  ways to select the remaining cells. Thus:

$$q_j(a) = \binom{A - k_j}{a - k_j} / \binom{A}{a}. \quad (11)$$

This can be shown by the combinatorial identity  $\binom{x}{y} = \binom{x}{x - y}$  to be equal to  $1 - p_j(A - a)$ , ensuring that  $\text{EAR}(a) = s - \text{SAR}(A - a)$

An *essential* feature of these calculations is that the random process generating subregions of area  $A - a$  in forming  $\text{SAR}(A - a)$  is consistent with the one generating subregions of area  $a$  in forming  $\text{EAR}(a)$ . Taking all subregions of area  $a$  as equally likely is consistent with taking all subregions of area  $A - a$  as equally likely. A different process can be used to generate subregions of area  $a$  to form  $\text{EAR}(a)$ , but for the comparison with  $s - \text{SAR}(A - a)$  to make sense, it must be consistent with the process used to generate the complementary subregions in forming  $\text{SAR}(A - a)$ .

In the empirical analyses of He and Hubbell (2011),  $\text{EAR}(a)$  was formed by averaging the number of endemic species within square cells of area  $a$  and  $\text{SAR}(A - a)$  was formed by averaging the number of species within square cells of area  $A - a$ . However, the random process underlying the calculation of  $\text{EAR}(a)$  is not the same one underlying the calculation of  $\text{SAR}(A - a)$ . That is, realizations of the complement of a random square subregion of area  $a$  are *not* square subregions of area  $A - a$ . It is this inconsistency in the experiments of He and Hubbell (2011) that causes  $\text{EAR}(a)$  to differ

from  $s - SAR(A - a)$  and not aggregation in the locations of individuals.

Predicting the loss of species associated with loss of habitat based on the species–area relationship is clearly difficult. Among other things, such predictions take no account of the dynamics of extinction and therefore are difficult to evaluate from short-term observations. This and other complications suggest that predictions of species loss based solely on the species–area relationship should be taken only as a rough guide. The paper by He and Hubbell (2011) brought into question even this limited use of the species–area relationship in predicting species loss. We have shown here that the analytical basis of their claim was in error and that its empirical basis reflected an inconsistency in spatial sampling. We conclude that, properly understood, the statistical basis for this use of the species–area relationship remains intact.

#### *Acknowledgments*

The helpful comments of Fangliang He and two anonymous reviewers are acknowledged with gratitude. U. Roll is supported by the Adams Fellowship Program of the Israel Academy of Sciences and Humanities. L. Stone is supported by the Israeli Science Foundation.

#### *Literature cited*

- Brooks, T. M. 2011. Extinctions: consider all species. *Nature* 474:284.
- Conlisk, E., J. Conlisk, and J. Harte. 2007. The impossibility of estimating a negative binomial clustering parameter from presence-absence data: a comment on He and Gaston. *American Naturalist* 170:651–654.
- Evans, M., H. Possingham, and K. Wilson. 2011. Extinctions: conserve not collate. *Nature* 474:284.
- He, F. L., and S. P. Hubbell. 2011. Species–area relationships always overestimate extinction rates from habitat loss. *Nature* 473:368–371.
- He, F. L., and S. P. Hubbell. 2012. He and Hubbell reply. *Nature* 482:E5–E6.
- Johnson, N. L., S. Kotz, and N. Balakrishnan. 1997. *Discrete multivariate distributions*. Wiley-Interscience, New York, New York, USA.
- Pereira, H. M., L. Borda-de-Água, and I. S. Martins. 2012. Geometry and scale in species–area relationships. *Nature* 482:E3–E4.
- Pimm, S. L., and P. Raven. 2000. Biodiversity: extinction by numbers. *Nature* 403:843–845.
- Smith, W. K., J. F. Grassle, and D. Kravitz. 1979. Measures of diversity with unbiased estimates. Pages 177–191 in J. F. Grassle, editor. *Ecological diversity in theory and practice*. International Cooperative Publishing House, Fairland, Maryland, USA.
- Thomas, C. D., and M. Williamson. 2012. Extinction and climate change. *Nature* 482:E4–E5.