Species Abundance: Natural Regulation of Insular Variation

Terrell H. Hamilton •
Ira Rubinoff
Robert H. Barth, Jr.
Guy L. Bush

Reprinted from Science, December 20, 1963, Vol. 142, No. 3599, pages 1575-1577

Species Abundance: Natural Regulation of Insular Variation

Abstract. *Variation in numbers of land plant species on islands in the Galapagos Archipelago can be predicted on the basis of elevation, area of the adjacent island, distance from the nearest island, and distance from the center of the archipelago, but not on the basis of the area of the host island. Multiple linear regression* ($y = bx_1 + bx_2 \ldots$) *gives better "goodness of fit" than curvilinear analysis* ($y = bx^z$). *The variation in number of species on large islands can be predicted more accurately than the variation on small ones. Ecologic diversity and isolation are the natural regulators of species abundance.*

The factors regulating numbers of species on member islands of an archipelago can be divided into those that impede or promote inter-island dispersal of individuals (for example, degree of isolation as measured by distance between islands) and those relating to successful establishment of natural populations on adjacent islands (for example, ecologic diversity as estimated by area, elevation, and other factors) (*1*). A major goal of evolutionary biology is to measure these factors and thus to determine whether they can be used independently or interdependently to predict population variables such as species numbers for naturally isolated areas (*2*).

Of several models available for predicting species numbers from observed variation in environmental factors, two are

$$y = bx \qquad (1)$$

where $x = x^z$ with $z = 1$, and

$$y = bx^z \qquad (2)$$

where $z \neq 1$. Model 1 predicts the dependent variable y as a function (b = regression coefficient) of the independent variable x, and assumes that unit changes in y associated with unit change in x are the same regardless of the magnitude of x. This is linear regression by the formula

$$\hat{Y} = a + bX$$

where a is the intercept value. Model 2 assumes that y changes progressively with change in x, and is the familiar curvilinear regression equation, where

$$\log \hat{Y} = \log b + z \log X$$

Both models may be readily adapted for multiple regression analysis to test for variation in Y associated with that of an X, independent of other X's. As to which model is applicable to a given set of data, it seems statistically sound to use the one having the better "goodness of fit"—that is, the model whose X or X's account for the greatest component of the variance of Y is the better predictor (*3*).

Preston (*4*) has recently discussed group variations in values for coefficient z resulting from application of model 2 to variation in isolated or sample numbers of animals and plant species as a function of area. He finds for land plants (*5*) of the Galapagos Archipelago a species-area relation (*6*) with a z value (*4*) of 0.33. This is somewhat larger than the theoretical value of 0.27. Preston notes an appreciable spread in species-area points and comments rightly that area is not the only factor determining richness of faunas and floras. In the present report we extend his analysis by (i) determining whether model 1 or model 2, in single or multiple factorial analysis, is a better predictor for insular variation in reported (*5*) numbers of land plant species in the archipelago and by (ii) examining the influences on insular floral richness of environmental variants other than area in the context of the colonization barrier for small islands.

In Table 1 are listed, for 17 islands of the Galapagos, data for the following independent variables: X_1, area (*4*); X_2, elevation (*5*); X_3, distance to nearest island (*1*); X_4, distance from center of the archipelago (*5*); and X_5, area of the adjacent island. X_1 and X_2 are positive indices to ecologic opportunity. X_3 and X_4 are positive indices to isolation, and X_4 is also a measure of the position effect (*1*). X_5, selected a priori, is used as a possible inverse index to isolation since a given island with a small neighboring one might be more isolated than it would be if its neighbor were larger and hence possessed greater numbers of potential dispersers.

For model 1, least squares estimates by computer analysis (*7*) give the following multiple linear regression equation:

$$Y = 16.4917 - 0.0030\,(X_1) + 0.0723^{**}\,(X_2) + 0.3982^{*}\,(X_3) - 1.0734^{**}\,(X_4) - 0.0096^{*}\,(X_5) \qquad (3)$$

Variance (R^2) for species number is 0.8398, receiving these contributions from the X's: 0.0067, X_1; 0.4457, X_2; 0.1367, X_3; 0.0657, X_4; 0.1850, X_5. Area of an island (X_1) is, itself, of little or no importance in predicting species numbers as such. In contrast, elevation (X_2) appears to be the major factor influencing species numbers on islands. It is followed in decreasing order of importance by X_5 (area of the adjacent island), X_3 (distance to the nearest island), and X_4 (distance from the center of the archipelago).

The estimate that, within the archipelago, the number of land plant species for a given island increases (72.3 species per 1000 feet) with elevation,

Table 1. Insular number of land plant species and some other environmental factors for the Galapagos Archipelago. The numbers of the islands correspond to those in Fig. 1.

Island and No.	Area X_1 (mi²)	Elevation X_2 (ft)	Isolation (mi)		Area of adjacent island X_5 (mi²)	Species numbers		
			X_3	X_4		Observed Y	Predicted $\hat{Y}_1$	Predicted $\hat{Y}_2$
1. Culpepper	0.9	650	21.7	162	1.8	+7	−24	+28
2. Wenman	1.8	830	21.7	139	0.9	14	+14	35
3. Tower	4.4	210	31.1	58	45.0	22	89	47
4. Jervis	1.9	700	4.4	15	203.9	42	49	39
5. Bindloe	45.0	1125	14.3	54	20.0	47	94	101
6. Barrington	7.5	899	10.9	10	389.0	48	76	56
7. Gardiner	0.2	300	1.0	55	18.0	48	−19	17
8. Seymour	1.0	500	0.5	1	389.0	52	+16	29
9. Hood	18.0	650	30.1	55	0.2	79	124	74
10. Narborough	245.0	4902	3.0	59	2249.0	80	95	176
11. Duncan	7.1	1502	6.4	6	389.0	103	49	55
12. Abingdon	20.0	2500	14.3	75	45.0	119	169	77
13. Indefatigable	389	2835	0.5	0	1.0	193	212	206
14. James	203.0	2900	4.4	12	1.9	224	225	166
15. Chatham	195.0	2490	28.6	42	7.5	306	259	164
16. Charles	64.0	2100	31.1	31	389.0	319	220	113
17. Albemarle	2249.0	5600	3.0	17	245.0	325	325	367

$\hat{Y}_1$ = species number predicted by model 1, multiple regression; $\hat{Y}_2$ = species number predicted by model 2 using only area.

increases (0.4 species per mile) with increased distance from the nearest island, decreases (1.1 species per mile) with increased distance from the center of the archipelago, and increases (1.0 per 100 square miles) with reduction in size of the adjacent island, is in keeping with our presuppositions (*1*). Svenson (*8*) has suggested that wealth of plant species in the islands results from altitude and area. Our analysis provides a partial qualification of his statement. The first-order correlation ($r_{y1} = +0.57$) between insular numbers of land plant species and insular area thus appears to be more apparent than real ($r_{y1.2345} = -0.20$; insignificant) and may result from covariations with elevation and the isolation factors ($r_{12} = +0.73$; $r_{13} = -0.28$; $r_{14} = -0.23$; $r_{15} = +.05$).

For model 2, multiple curvilinear regression analysis with logarithms of the data gives the equation:

$$\log \hat{Y} = 1.0359 + 0.2255\ (\log X_1) + 0.2520\ (\log X_2) + 0.0659\ (\log X_3) - 0.2345\ (\log X_4) + 0.0262\ (\log X_5) \quad (4)$$

Contributions to the variance of log Y ($R^2 = 0.6726$) are: 0.5849, X_1; 0.0110, X_2; 0.0254, X_3; 0.0476, X_4; 0.0037, X_5. The best and perhaps only predictor of log Y is the logarithm of area, but the variance for log Y by model 2 is less than that for Y by model 1. This suggests for model 2 a poor "goodness of fit," an observation also found when area alone is used as a predictor by the formula whose unrectified values are:

$$\hat{Y} = 28.58\ X_1^{.33**} \quad (5)$$

The infidelity of the latter, in contrast to multiple regression analysis by model 1, in prediction of observed values for species numbers is shown in Fig. 1.

The preceding discussion demonstrates an approach to the study of natural control of species abundance which at best is limited. It is not that one model predicts or fails to predict, but that several will predict with varying degrees of accuracy (Fig. 1). Here we emphasize three points. First, the model which estimates most precisely the primary measurements of populations or species attributes is more likely to quantitatively represent organismic responses to environmental variants. Second, transforming original measurements by conversions to logarithms necessitates alteration of the varying relation among sample measurements; thus arithmetic-to-logarithmic analyses can curve linear relationships as well as straighten curvilinear ones. Third, natural regulation of species abundance, as well as of species characters, is undoubtedly multiple rather than single factorial, and as many factors as are intuitively of importance—ecologic, geographic, genetic, historic, accidental, behavioral, and others—need to be quantified and evaluated for predictive power (*9*).

Deviations from regression, indicating error or variation unexplained by the X's were plotted against such, and new information is evident only for X_1 and X_3 plottings. Model 1 predicts floral richness for larger islands more accurately than it does for smaller islands (Fig. 2). Ecologic diversity tends to decrease with reduction of insular area and, concomitantly, increased opportunities are expected to occur for habitat- or niche-preemption by initial colonizers, for extinctions and losses of genetic variability associated with reduced or fluctuating population size, and for depauperate biotas due to sampling error in interisland dispersal. Further evidence for operation of this colonization barrier for the smaller is-

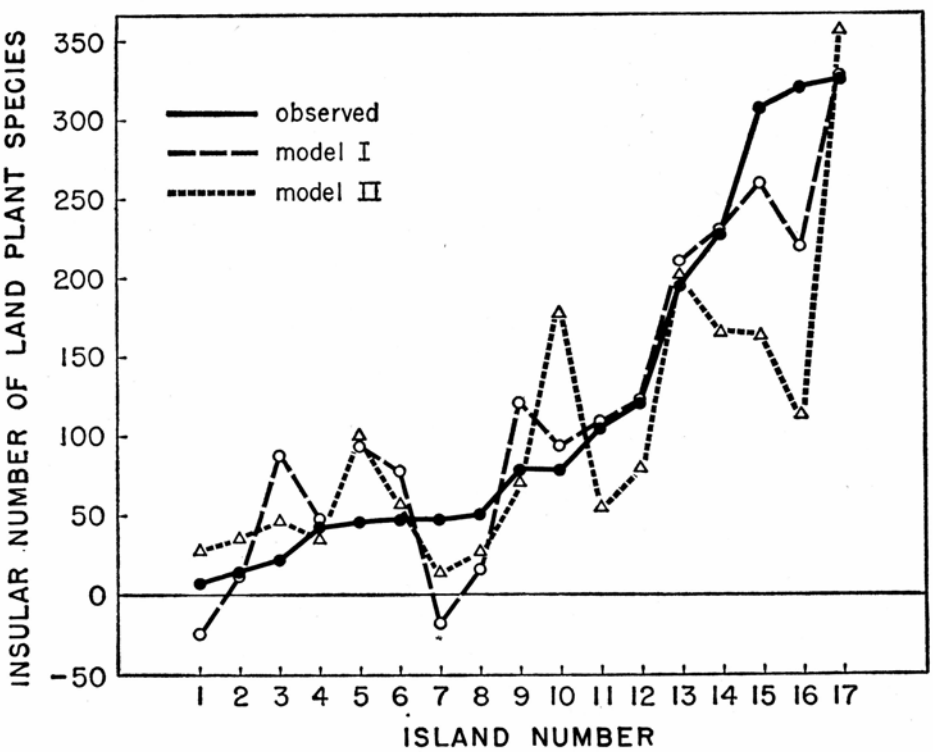

Fig. 1. Land plant species abundance predictions for 17 islands (Table 1) of the Galapagos Archipelago. Model 1, Eq. 3; model 2, Eq. 5 or $\log \hat{Y} = 1.4560 + 0.33^{**} \log X_1$. Multiple curvilinear regression by model 2 gives predictions similar to, but more divergent than, those given by the preceding formula.

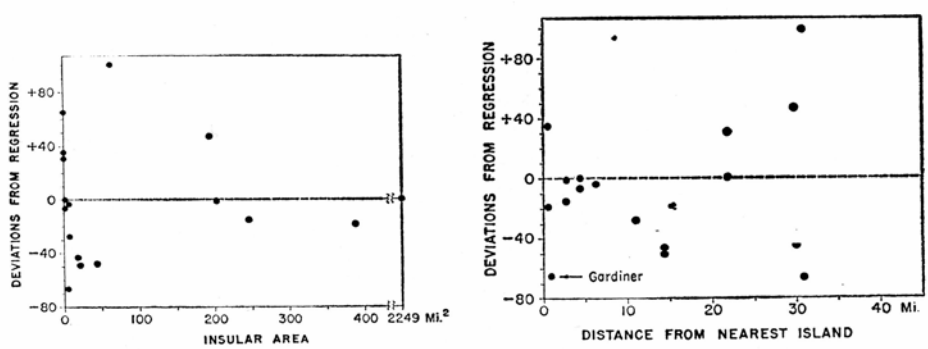

Fig. 2. Residual variation versus Galapagos environmental variants (model 1, Fig. 1). (Left) Species prediction error plotted against insular area (X_1) to show decreased prediction accuracy associated with decrease in insular area. (Right) Prediction error plotted against distance from nearest island (X_3) to show, with exception of Gardiner Island, the apparent increase in prediction error with increase in distance from nearest island. Together, the diagrams illustrate the inaccuracy of model 1 for small, isolated islands.

lands of the Galapagos comes from the observation (Fig. 2) that prediction error is greater for the more isolated islands. That correlation between area and isolation ($r_{13} = -0.28$) is slight hints that the distance effect may contribute to floral species variation independently of factors associated with reduced insular area.

That better predictions of Y should result from multiple rather than single regression analysis is not surprising. Of interest, however, is that linear, rather than curvilinear, multiple regression analysis gives more accurate predictions. In fact, multiple curvilinear analysis for these data gives predicted values less accurate than those determined by single curvilinear analysis with area only (Fig. 1). This raises the question of why area alone can be used to predict variation (in model 2) with an accuracy approaching that obtained by use of several factors (in model 1). Another question is why area by model 2 gives better predictions than it does by model 1? The answers may relate to the obvious: that the *number* of factors determining richness of insular floras or faunas increases progressively with increase in insular area. Thus use of logarithms and the model of $y = bx^z$, rather than of the actual numbers and $y = bx$, may give a prediction "curving in the right direction" for progressive, overlapping accumulation of elements of ecologic diversity associated with increased area. It is now clear that groups whose insular variation in species numbers have previously been studied by the Arrhenius approach (model 2) need to be examined by multiple regression analysis, utilizing linear, curvilinear, or mixed linear-nonlinear models.

Our study deals with insular variations in number of plant species for a cluster of small islands remote in the eastern Pacific, and two interrelated sets of problems are undoubtedly intermingled by the analysis: (i) insular production of endemic species versus insular increase of nonendemic species and (ii) whether insular number of species and number of individuals regulate, in part, each other. Preston's discussion (*4*) is the most recent one to approach the topic, and we will soon discuss it elsewhere (*10*). The present report, however, suggests that area itself exerts little control on insular species abundance in strong centers of endemic differentiation (*1*), with isolation and ecologic diversity being more important regulators. Area may be more important as a regulator in regions of larger land mass (large islands, continents) where barriers to dispersal are reduced and the degree of isolation is decreased. In this context the possibility arises that on small islands another aspect of the colonization barrier is the regulation of species numbers by numbers of individuals maintained or permitted by reduced ecologic diversity and competition, influenced in turn by vagaries of interisland dispersal (*11*).

TERRELL H. HAMILTON
Department of Zoology,
University of Texas, Austin
IRA RUBINOFF
Museum of Comparative Zoology,
Harvard University,
Cambridge, Massachusetts
ROBERT H. BARTH, JR.
GUY L. BUSH
Biological Laboratories,
Harvard University

References and Notes

1. T. H. Hamilton and I. Rubinoff, *Evolution*, in press.
2. This statement proceeds from comparable, but more inclusive, statements by L. B. Slobodkin, in *Growth and Regulation of Animal Populations* (Holt, Rinehart, and Winston, New York, 1961).
3. G. W. Snedecor, *Statistical Methods* (Iowa State College Press, Ames, 1956).
4. F. W. Preston, *Ecology* **43**, 185, 410 (1962).
5. A. L. Kroeber, *Floral Relations among the Galapagos Islands* (Univ. of California Press, Berkeley, 1916); R. I. Bowman, *Univ. Calif. Publ. Zool.* **58**, 1 (1961).
6. We use the expression "species-area relation," instead of the familiar "species-area curve," because the latter implies a curvilinear association between species numbers and area, an implication which may or may not be warranted.
7. We are indebted to Beverley Lee for programming assistance and computer feeding at the Harvard Computing Center. All regression coefficients were evaluated by Student's *t*-distribution, and may be considered insignificant unless accompanied by asterisk* (significant: $.05 \geq P > .01$) or asterisks** (highly significant: $P < .01$).
8. H. K. Svenson, *Am. J. Botany* **33**, 394, 427 (1946).
9. In addition to Slobodkin (*2*) and Preston (*4*), see S. Cain, *Am. Midland Naturalist* **19**, 573 (1938); L. C. Cole, *Ecology* **21**, 329 (1946); N. G. Hairston, *ibid.* **40**, 404 (1959); R. H. MacArthur, *Proc. Natl. Acad. Sci. U.S.* **43**, 293 (1957); C. B. Williams, *Nature* **152**, 264 (1943).
10. T. H. Hamilton, R. H. Barth, Jr., I. Rubinoff, in preparation. While the finding that small islands deviate from prediction more so than larger ones is not obvious and is not anticipated by our presuppositions resulting from a comparable study of the Darwin finches (*1*), we note Ernst Mayr's exact prediction of our finding in conversations about the biogeography of the Australopapuan biota prior to computer analysis of the present data.
11. Study was supported in part by a National Science Foundation research grant. We gratefully acknowledge the assistance and interest of W. F. Blair, E. Mayr, R. H. MacArthur, F. A. Pitelka, R. K. Selander, W. J. Smith, O. Solbrig, E. E. Williams, and E. O. Wilson.

29 July 1963