The Theory of Range-Size (RS) Distributions

SYDNEY ANDERSON[1]

ABSTRACT

If the frequency distribution of sizes of geographic ranges (the range size or RS distribution) of the species comprising the continental fauna for a given taxonomic group of organisms is viewed as a system in dynamic equilibrium and if the observed "hollow curve" or logarithmic distribution is taken into consideration, it follows that the geographic range of a species, regardless of its size, is more likely to decrease than to increase. Some of the conditions of the system, including the complications of extinction and speciation, are examined by simulating the system with a Markov chain model on a digital computer.

Some general attributes of the system are: diversity and the RS distribution are maintained dynamically, any local change in number of species or transition probability (TP) ramifies throughout the system, and the system converges on a theoretical equilibrium (that keeps changing, in the real world). A simple Basic Logarithmic Model that assumes a logarithmic series of TP values accounts for most of the variation in RS distributions of North American vertebrates. The fit of the model is not improved by making TP values a function of border length rather than area. The Basic Logarithmic Model, regarded as a randomly determined Null Hypothesis, is rejectable. There remain residuals or deviations of data from this model that need further explanation.

Actual RS distribution curves for groups of North American vertebrates can be closely simulated by specifying six values: minimum and maximum possible ranges, a logarithmic series of TP values, an inflection point (somewhere between 5.0 and 6.3 km²), and TP ratios (of the probability for an increase, I, to that for a decrease, D) between 0.66 and 0.75 above the inflection point and from 0.85 to 0.89 below.

INTRODUCTION

A general theory for the frequency distribution of range sizes of species in a continental faunal system is developed here. The effects of various assumptions about increases and decreases in ranges, extinctions, speciation, and other processes on the model are examined. The range of conditions that produce specific models resembling faunal data for North American vertebrates are found by trial and error with a computer by using a Markov chain process.

The frequency distribution of the sizes of geographic ranges of species in each of the different groups of vertebrates in North

[1] Curator, Department of Mammalogy, American Museum of Natural History.

America is a "hollow curve" (Anderson, 1977, 1984a, 1984b). Hollow curves for range sizes were reported also for North American mammals and for African birds by Rapoport (1982). Fishes have relatively small ranges and birds have relatively large ranges among North American vertebrate groups (fig. 1). Some aspects of these distributions that are not otherwise evident may be revealed by changing the scales of graphical presentations (compare fig. 2 with fig. 1).

Species with small ranges are more common than those with larger ranges. The number of species decreases by about an order of magnitude for each increase in range size of an order of magnitude, except at the small end of the scale (fig. 3). Between values of 10^2 and 2×10^7 km², the relationship of numbers of species per 100 km² increment in range size (y) having ranges of different sizes (x) takes the form: $y = 100 \times x^{-1}$. All range measurements are here considered on a decimal scale and in km². Range sizes in a given group vary over several orders of magnitude (fish, for example, vary over at least six orders of magnitude).

The precise form of the frequency distribution varies from group to group. Sometimes the curve approximates a lognormal distribution (fig. 4), but what, if any, biological significance this particular distribution may have is not clear.

DEFINITIONS

A few terms and concepts need to be defined because they are not familiar or because they are used in one restricted sense here.

Diversity. The number of species in a fauna. This is the simplest of several generally used measures of diversity.

Fauna. The combination of species of a given larger taxonomic group that are present in a given geographic area. For example, we may refer to the mammalian fauna of North America.

Hollow curve. Any highly concave curve, such as those shown in figure 1. Such curves are produced by a great variety of circumstances both actual and mathematical (for a historical review and some examples see Anderson, 1974; for theoretical and mathematical background see Boswell and Patil, 1971).

Range size (RS). The geographic extent or area of the distribution of a species. In the studies of North American vertebrates referred to here, the entire area within a curving linear boundary drawn somewhat subjectively around known localities on a map was measured in km²; this measurement was then rounded to one significant figure.

RS distribution. The frequency distribution of sizes of geographic ranges of species in a given larger taxonomic group being considered in a given geographic area, such as turtles in North America.

Transition probability (TP). The probability that a species with a given size of geographic range will either increase or decrease its range by a given amount in a given time. D will refer to a TP for a decrease and I for an increase.

FACTORS PRODUCING A GIVEN DIVERSITY AND RS DISTRIBUTION

At any one time, diversity and RS distribution are the products of the following conditions and processes. Any general or comprehensive theory or model of diversity and RS distributions must embody at least these six factors.

Factor 1. Past diversity. The diversity at any one time is viewed as the net result of the additions to and removals from some earlier diversity. Therefore, that earlier diversity must be taken into account in modeling a system in which diversity changes with time.

Factor 2. Past RS distribution. If the RS distribution is viewed as changing from some prior condition under the influence of changed TP values (factor 3), then the prior condition (factor 2) is important in describing the entire process.

Factor 3. Transition probabilities. These are mathematical expressions inferred from the observation that ranges of certain sizes are more common than ranges of other sizes, whatever the causes may be.

Factor 4. Sizes of ranges of species added to the fauna. In modeling RS distributions over long (evolutionary) periods of time some assumptions must be made about the nature of the process of speciation and the consequent distribution of sizes of ranges of species

arising in situ or being added to the fauna from some outside source.

Factor 5. Additions. Additions to the world's biota (defined as the species thereupon) come only from the evolutionary process of speciation.

Additions to any geographical subset of the world's biota, such as the fauna of a continent, can also come from some other part of the world. This arrival may result from sufficient individuals dispersing across the intervening barrier to establish a self-maintaining population, or by the disappearance of a barrier as in the merging by tectonic movements of the earth's crust of two or more formerly separate land masses with different faunas.

There are intermediate conditions between the complete separation of land masses and their complete merger and there is much biogeographical literature dealing with these conditions, under topics such as island biogeography, land bridges, and great faunal interchanges (between North and South America, for example). And, chiefly in regard to smaller areas and shorter time intervals, there is also a large volume of ecological literature dealing with relative degrees of separation of habitat patches and their causes and effects.

Three different probabilities are relevant here.

A. The probability that an individual of a species not already part of the North American fauna, for example, will arrive in North America from another part of the world varies with the species and with the larger group of organisms. Extreme examples are: in some species of birds the probability for each year is near one since stragglers are observed nearly every year; and in some species of fish the probability is effectively zero, it being unlikely that any living individual has ever moved from one continent to another.

B. The probability that a resident self-perpetuating population of a species represented by these dispersing individuals will become established in North America is much lower than the preceding probability and generally higher than the probability described in the next paragraph.

C. The probability that a species will be added to the fauna of North America by merger of large land masses is much lower in most groups. This sort of event occurs every few tens of millions of years. However, when it does occur a relatively large readjustment of the fauna may follow.

Of the above probabilities, A and B are logically related and seem reasonable; C is more subtle and less certain. There may be groups of organisms having small size, small ranges, lack of dispersal abilities, or other attributes such that the probability of transferral to and establishment on another continent is actually lower for A or B than for C, even though C is very low.

Factor 6. Removals. In modeling RS distributions, some assumptions must be made about the processes that remove species from the fauna, namely extinction and continental fragmentation:

Extinction of the resident population of a species on a continent under consideration removes the species from that continent's fauna whether or not the species continues to survive elsewhere. Extinction occurs more frequently than continental fragmentation. Continental fragmentation is a relatively rare event.

The probable rates of these two removal processes are important to the theory and practice of biogeography. How does the RS distribution (in regard to both sizes of ranges and their placement on the continent) influence the probability that a given species will be removed by a given continental fragmentation? This process will not be included in the models developed here. Some related questions, concerning the merging of continents rather than their fragmentation, were discussed briefly by Anderson and Evensen, 1977.

MODELING RS DISTRIBUTIONS

The RS distribution is here simulated as a system in a series of models on a computer. The expectation in developing these models and comparing them with real data is not that

a model will be found that fits the data and therefore absolutely proves that the assumptions of the model are correct. Firstly, models are inherently simplified and, secondly, there may be other models that also fit. However, when the model does not fit, then probably one or more of the assumptions of the model are not valid for the real system being simulated. The process, as I view it, is one of testing alternatives in order to decrease or increase our confidence in them, rather than absolutely proving or disproving them. Computer models simulating natural systems can be useful in winnowing and testing hypotheses.

Let us make some initial assumptions about the system and about the model itself.

1. *Diversity is maintained dynamically.* The faunal diversity of any given geographic area can be viewed in the terms of speciation, colonization, extinction, and emigration (see Williams, 1964, for examples and discussion of "birth-death" models in general). Thus diversity is dynamic in the sense that active processes are involved.

2. *An RS distribution is dynamic.* An RS distribution also is dynamic because the sizes of individual ranges change frequently, on a time scale of a few years. The changes vary from relatively minor oscillations about a species boundary with the average position remaining about the same for hundreds of years to large-scale shifts in which the boundary retracts in some places and expands in other places. In extreme cases this may result in the total range moving from one place to another completely different place. There were major changes of this type in some North American species in the Pleistocene as ice sheets advanced and contracted.

3. *There are no additions or removals.* This assumption is made as an initial simplification (in models A through E) in order to examine more easily how the RS distribution relates to different TP values. Additions and removals are then considered in models F through I.

The system as a general conceptual model was examined as a Markov chain process on a digital computer. It might also be manipulated on an analog computer or as a Monte Carlo model on a digital computer by tracking individual species ranges whose changes are determined by the use of random numbers in such a way that their average shifts conform to specified TP values. The Markovian simulation was selected because it is relatively easy to create and manipulate and is an adequate approximation to actual operation of such a system. The computer model is also simplified by following changing probabilities (which usually have fractional values) of cell contents through sequences of probability cycles. A different and more complex model would be needed to maintain only integral values to represent species in each cell. In reality, of course, species are not fractional and the number of species in any given cell at any one time is an integer value.

The basic methodology of the model is that values are grouped into cells. A cell is an arbitrarily defined range of measurements (in this case of sizes of geographic ranges). For example, we might consider range sizes as grouped into cells of 1000 km^2 intervals and place all range sizes (regardless of where the species occur) of from zero to 1000 (all discussion of range sizes is in terms of km^2) in the first cell and ranges from 1000 to 2000 in the second cell, etc.

4. *Movements occur only between adjacent cells.* We assume that a species cannot move from one cell to a given larger or smaller cell without passing through any intermediate cell or cells in the linear series.

There is a special case in which this is not true, namely the vicariant division of a species by evolution. In this case some barrier arises which divides a species range into two parts. The population of one or of both of the parts then changes until finally they are different species. By this process the range of the original species is divided directly into two parts without either of the parts passing through the intermediate range sizes between the original range and the range of the daughter species. This process is included in one of the later models (H) developed here. It is negligible over short time intervals and therefore in most ecological modeling, but it is important over long periods and in modeling evolutionary events.

Suppose that a species on its way to extinction undergoes a contraction of its RS (which passes through successive cells) until it reaches the minimal cell and eventually passes from that cell to extinction. This is probably the usual course of extinction. In

contrast, however, it is theoretically possible to suppose that the individuals comprising the total population of a species might become fewer and fewer, until a widely scattered and very low-density population suddenly ends with the simultaneous death of the last widely scattered individuals, in which case the RS would not pass through successively smaller range sizes as it approached extinction. This scenario seems much less probable than the former and therefore will not be considered further in the modeling here. These considerations are relative to the scales of space and of time being considered; for example, the few years required for the elimination of the passenger pigeon are practically an instant, from the perspective of a geological time scale. Evolutionary and ecological effects of time scales and of randomness were discussed by Levandowsky and White (1977). Another interesting presentation of effects of scale on form, from a more mathematical and theoretical perspective, is that of Mandelbrot (1982).

5. *Transition probabilities are defined.* These probabilities express changes from one cell to another. The terminal cells, the ones for the smallest and largest possible range sizes, must have different TP functions than the intermediate cells, because the definitions of maximum and minimum exclude the possibility of transition to a larger or smaller size, respectively.

Considering only two cells, if the TP (in any given interval of time) for an increase in RS were 0.2 and for a decrease were 0.1, then at equilibrium the number of species in cell 1 would be half that in cell 2. If the number of species in cell 1 is $N(1)$ and in cell 2 is $N(2)$, then equilibrium is reached when:

$$N(1) \times 0.2 = N(2) \times 0.1$$

To generalize the above example, if the TP for an increase from cell 1 to cell 2 = $I(1)$, from cell 2 to cell 3 = $I(2)$, ... $I(n)$, and if the TP for a decrease from cell 2 to cell 1 = $D(2)$, from cell 3 to cell 2 = $D(3)$, ... $D(n)$, then the content of any cell n will be:

$$N(n) = N(n - 1) \times I(n - 1) + N(n) \\ \times (1 - D(n) - I(n)) \\ + N(n + 1) \times D(n + 1)$$

At the equilibrium point the net rate of outflow from each cell would equal the net rate of inflow.

6. *The system converges on equilibrium.* Given the transition probabilities between all cells in the entire linear sequence of cells, an equilibrium will be approximated eventually which will be the same regardless of the initial RS distribution of species. This can be demonstrated by constructing a model via a computer program that will accept any initial RS distribution and then run through repeated cycles in which the contents of each cell are multiplied by the specified TP values and the contents of the cells are adjusted accordingly. In strictly mathematical terms, the above system converges on equilibrium rather than absolutely reaches it. In reality, conditions external to the species themselves may never remain constant long enough for a close approach of conditions to equilibrium. If this is true, then the system tracks a changing theoretical equilibrium.

There may be another equilibrium at a higher level of abstraction in which some faunal phenomenon is approaching a theoretical equilibrium condition at the same rate that the external conditions are changing. This assumes that environmental conditions are changing in such a way that the theoretical equilibrium state is moving away from the actual current state of the faunal phenomenon rather than toward it. The faunal phenomenon referred to might be the RS distribution, as in the present context, or it might be the morphological adaptations of species or taxonomic diversity in some other evolutionary context. The distance from the current state of the RS distribution to the current theoretical equilibrium might then remain constant. The RS distribution would then be tracking the changing environment at the same rate that it is changing.

7. *Changes ramify through the entire system.* An interesting property of such a system is that any change of the TP between any two cells in the entire system or any change in the content of a cell anywhere in the system will result in a subsequent readjustment of cell contents throughout the system.

GRAPHICAL REPRESENTATIONS

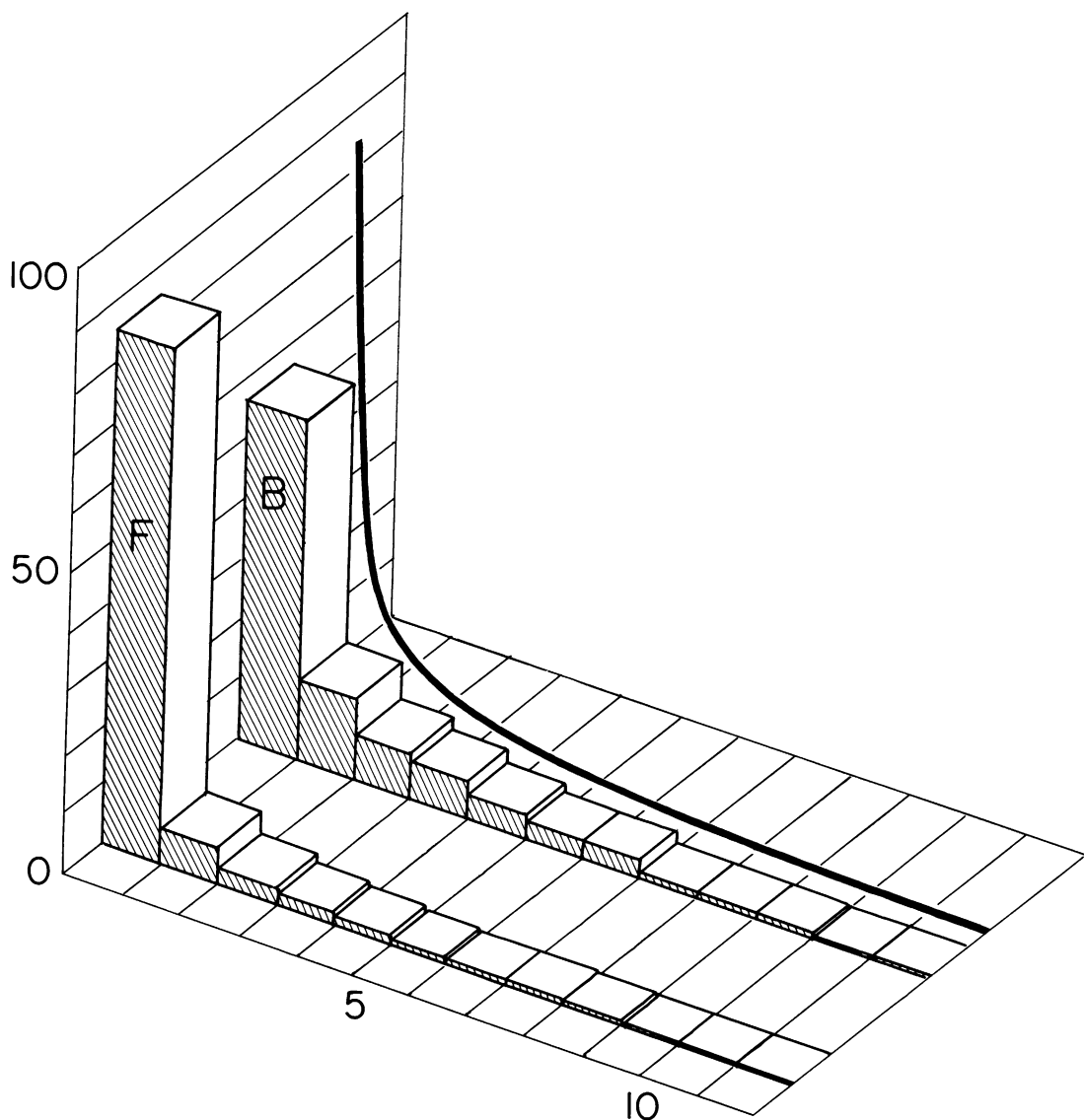There are many graphical means for summarizing frequency distributions such as the

FIG. 1. Hollow curve distributions illustrated by frequency histograms of range sizes of North American fishes (F) and birds (B). The percentage of species in each 1 million km² size-class is shown on an arithmetic scale.

RS distribution. Most people are more or less comfortable with simple frequency histograms using arithmetic scales (like that shown in fig. 1). Those who are uncomfortable with anything else will be uncomfortable with most of what follows because a simple arithmetical frequency histogram is virtually useless in the present context. Various more abstract graphs are needed to illustrate what is happening.

The frequency histogram (as shown rather elaborately in fig. 1) for RS distributions has already undergone one conversion from the raw data. The ordinate has been converted from absolute values to percentages so that fish and birds may be compared more easily. Both axes are arithmetically scaled, however.

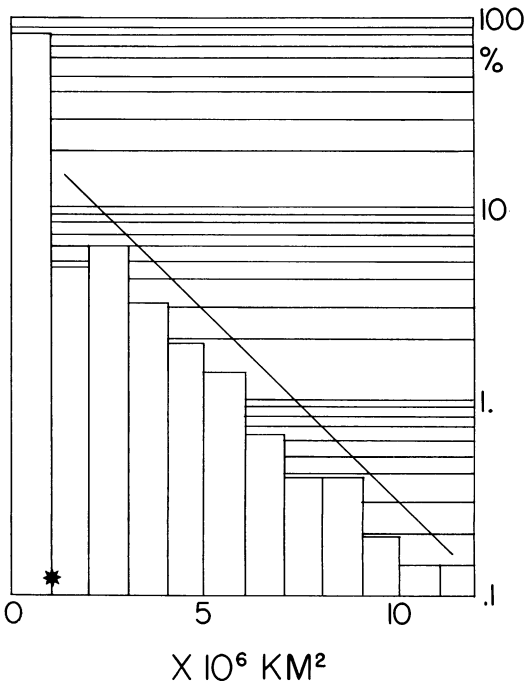In figure 2 (showing fish only), the vertical axis of figure 1 has been converted to a log-

FIG. 2. The effect of converting the vertical scale from arithmetic to logarithmic. Frequency histogram for fish, data as in figure 1.
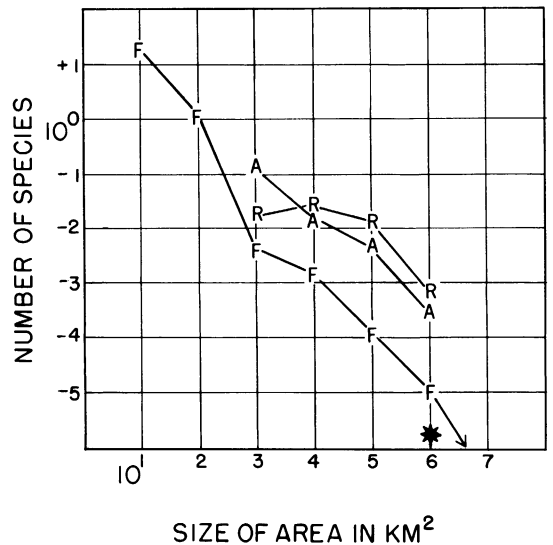

SIZE OF AREA IN KM^2

FIG. 3. The negative area-rarity correlation (from Anderson, 1984b). Graph for North American fish, amphibians, and reptiles showing numbers of species (averaged for each succeeding order of magnitude) having ranges of any given size. Counts of species are expressed in terms of the number per 100 km^2 increment.

arithmic scale. A logarithmic distribution is suggested by the approximation of the frequency values to a straight line. This relationship was not clearly evident in figure 1.

In figure 3 (from Anderson, 1984b; and in subsequent figures), the horizontal axis also has been converted to a logarithmic scale and it is now possible to see logarithmic relationships that were not so evident in figures 1 or 2. In comparing figures 2 and 3, note that the logarithmic relationship that is evident in figure 3 lies entirely within the leftmost cell of figure 2. The value of 1×10^6 is marked with an asterisk on both figures to emphasize this difference. Figure 2 reveals information about sizes of ranges larger than 1×10^6 that is hidden at the extreme right in figure 3 and figure 3 provides information about sizes of ranges smaller than 1×10^6 that is hidden at the extreme left in figure 2.

A graphical test for lognormality is provided in figure 4. The vertical axis is a probability scale and the horizontal axis is a logarithmic scale. A straight line for the plotted cumulative frequencies would indicate a log-

normal distribution. These RS distributions are logarithmic but not exactly lognormal, although the approach is close for the reptiles (R).

In figure 5, the values are plotted cumulatively, with the percentages of species scaled arithmetically at the left and logarithmically at the right. All data points except the right-hand one for each frequency class are omitted and a line is drawn connecting these terminal points in order to make the pattern easier to see. Figures 1 through 5 are different ways of looking at the same set of data.

Any cumulative curve of this type will begin at the upper left of the graph and end at the lower right. The left curve in figure 5 is the type used for data on various vertebrate groups in earlier studies (Anderson, 1977, 1984a, 1984b) and this type of curve is used here to compare the models simulated on computers.

SELECTED MODELS

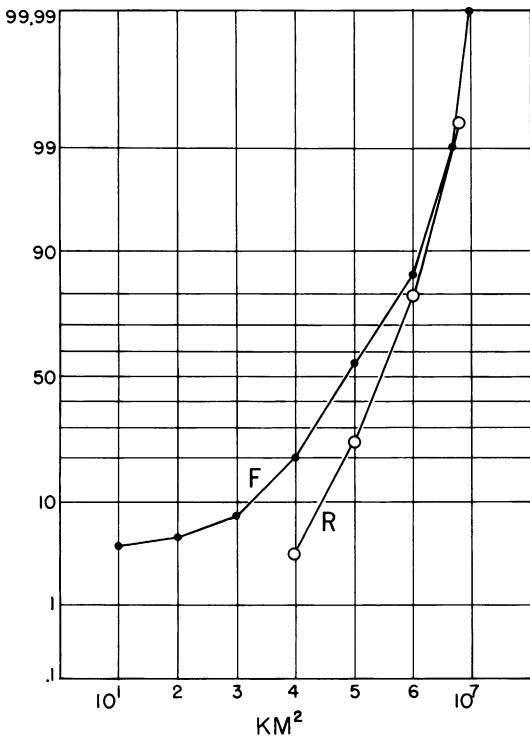*Model A.* This, the Basic Random Model, may be regarded as a Null Hypothesis in-

FIG. 4. A test for lognormality. Cumulative percentages of species in ×10 (1–10, 11–100, 101–1000, etc.) log classes of geographic ranges in km² (abscissa) are plotted against cumulative percentages on a probability scale (on the ordinate), to see whether distributions are lognormal. This would be the case if points fell on a straight line. Fish (F) and reptiles (R) are plotted. Neither is precisely lognormal although reptiles approach that condition (from Anderson, 1984b).

volving the simplest possible assumptions about transition probabilities. We will subsequently reject the hypothesis, but this model is theoretically possible, not unreasonable, and helpful in thinking about the system. The simple assumptions are:

(1) Cells are of equal size on an arithmetic scale.

(2) Transition probabilities are equal. All species have the same probabilities for change. The TP for an increase (I) of a given absolute amount (in km²) in range size is equal to the TP for a decrease (D) of like amount.

Given these conditions, the RS distribution will be uniform among ranges of all pos-

sible sizes. If the distribution is not uniform initially, then it will move toward uniformity, regardless of the initial distribution.

The upper boundary condition or maximum for sizes of ranges of terrestrial animals in North America is about 2×10^7 km² which is the area of the continent (as treated by Anderson, 1977, 1984a; south through Panama). The theoretical minimum might be perceived as the space occupied by the last individual of a species just before it dies (about 1×10^{-6} km²). Although there are a few species that actually have ranges smaller than 100 km² (such as several species of pupfish each known to inhabit only one tiny isolated spring in the deserts of the Great Basin), I have arbitrarily assumed, for practical purposes, a minimum RS of 100 km².

If the sizes of ranges for each group in figure 3 were distributed uniformly among all possible sizes, the line for each group would be horizontal. Since the lines are not horizontal, one or more of the assumptions of Model A is invalid. The minimum, maximum, and cell dimensions are axiomatic, so the assumption of equal TP values is not tenable. Clearly a logarithmic model would fit better than an arithmetic one. It seems reasonable to equate an increase of say 5 percent in one range size with 5 percent in another one, even though the absolute change is quite different.

Model B. The Basic Logarithmic Model. This might also be termed the Smaller is Better Model. A logarithmic relationship could be embodied in the Markovian model by specifying cell dimensions logarithmically, by specifying the TP logarithmically, or both. The assumptions of Model B, which involve both, are:

(1) Cells are of equal increments on a logarithmic scale. The possible range sizes are arbitrarily placed in 54 cells (with values of 2.0, 2.1, . . . , 7.3, to encompass actual ranges from 1×10^2 to 2×10^7 km²).

(2) Transition probabilities are equivalent between all cells, in the sense that all species have the same probabilities for an increase or decrease sufficient to move the species to an adjacent cell.

The amount of change on an arithmetic scale needed to do this is not equal as a percentage, however. For example, the change from the midpoint of the cell with the value
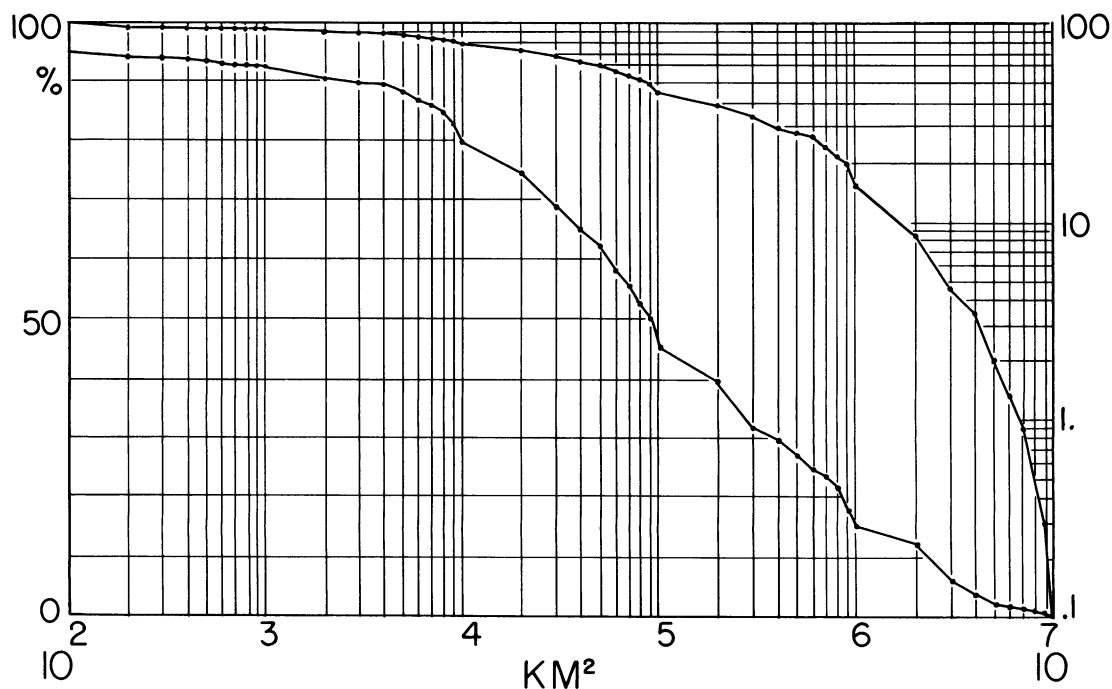
FIG. 5.   Graph showing the effect of plotting the same data for fish that were used in figures 1 and 2 as a cumulative frequency, and the effect of using arithmetic and logarithmic ordinate scales. Percentages are shown on the ordinate and are expressed on an arithmetic scale at the left and a logarithmic scale at the right.

of 2.0 on a log scale (=100) to the midpoint of the adjacent cell with the value of 2.1 (=125.89) is 25.89, which is 25.89 percent of 100 and 20.57 percent of 125.89. The same percentage relationships apply to every pair of adjacent cells. If we assume, as mentioned above, that equal percentage changes are equivalent for purposes of our Null Hypothesis, then a change of 25.89 percent is less likely to occur than one of 20.57 percent. The TP ratio of I/D is 0.79.

In figure 6 the results of two computer runs (p for changes between cells = 0.1 in both directions) beginning with different RS distributions, illustrate the convergence of the system on an equilibrium. A diagonal line represents the equilibrium in which there would be equal numbers of species in each cell (as in Model A, except that the cells are now of logarithmically graded sizes).

Model C. An Exaggerated Smaller is Better Model. The effect of making the I/D ratio less than that of the Basic Logarithmic Model

(0.79) is shown in figure 7. The ratio here is 0.26 [0.26 = (0.1/0.3) × 0.79, where 0.1 and 0.3 are the probabilities for changes from cell to cell].

The computer run began with species uniformly distributed among all cells. The RS distribution converges on a curve to the left of the curve for 400 cycles, in which each cell will have three times as many species as the next larger cell and the most species, of course, will be in the smallest possible cell.

Comparison of figures 6 and 7 with the left curve of figure 5 (representing an actual RS distribution drawn in a comparable way) indicates that Model B is closer than Model C to the real distributions. Departures of several real distributions from the condition of Model B are shown in figure 8.

The probability ratio (I/D) of 0.26 drives the curve of figure 7 far to the left. A ratio closer to 0.79 would drive any beginning curve to an equilibrium position more or less paralleling the intermediate curves of figure 7,
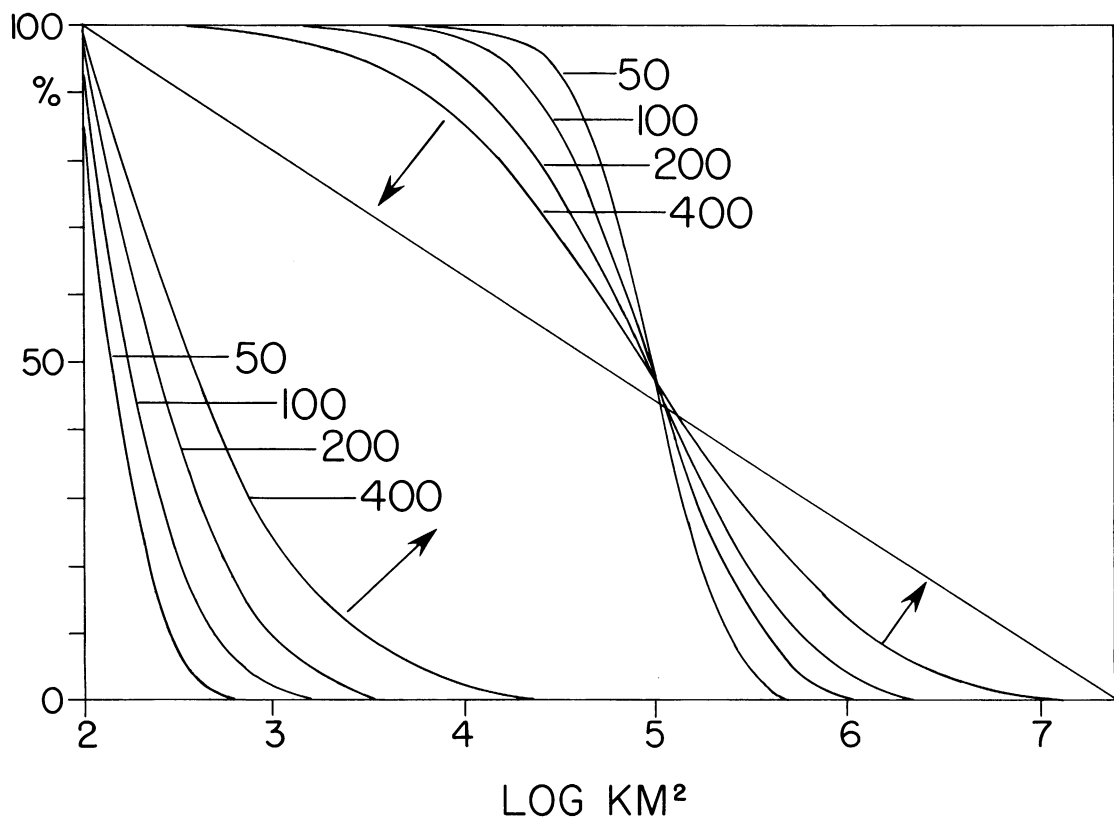
FIG. 6. The Basic Logarithmic Model, with I/D ratio of 0.79 (or with equal probabilities between log cells of increases and decreases). Sets of curves showing changes in cumulative percentage distributions of range sizes in computer simulations after various numbers (doubling values) of cycles. At the left are curves after 50, 100, etc. cycles, beginning with all species in cell 1 (the smallest possible value), and with transition probabilities of 0.1 in both directions for every cell. At the right are curves, beginning with all species in the cell for 1×10^5 km², and with the same TP values. In all such cases the distribution converges (arrows) on the diagonal line, which represents the system with equal numbers of species in each cell.

and the nearer the ratio is to 0.79, the nearer the equilibrium curve will be to the diagonal line. In all cases, however, the curve will lie below the diagonal.

Model D. A Reduced Ratio Model (less bias in favor of smaller ranges). The effect of making the I/D ratio more than that of the Basic Logarithmic Model (0.79) is shown in figure 9.

Since the curves representing actual RS distributions of North American vertebrate groups mostly lie above the diagonal (all are within the shaded area in fig. 9), the probability ratio (I/D) for the animals must be, in some significant part, more than 0.79 (al-

though still less than 1.0). The other curves are sample equilibrium curves when the ratios of probability of decrease to that of increase is 0.79 and four values more than 0.79. In contrast to the condition shown in figure 7, all curves lie above the diagonal.

Model E. The Border Model. The previous models make the transition probability a function of the size of the range (a two-dimensional area). Since the only place that a range can change in size is at its border, it might be more reasonable to make the probability of change a function of the length of the border (which is measured in one dimension). This model tests the effect of making
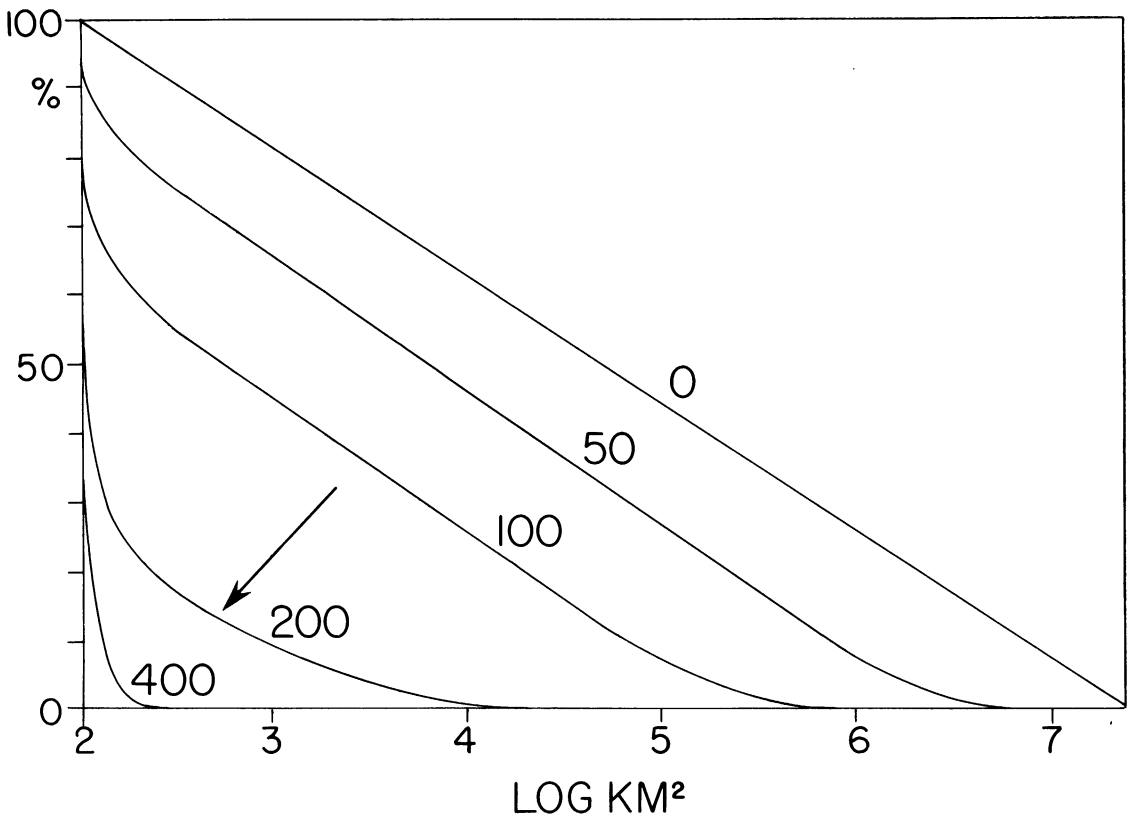
FIG. 7. The Exaggerated Smaller is Better Model. Intercellular transition probabilities of 0.3 and 0.1 are given. Computer simulation of system, beginning with equal numbers of species in all cells, showing cumulative curves after various numbers of cycles.

the probability a function of length of the border rather than size of area. Assumptions of the model are:

(1) Cells are logarithmically equal as in Models B to D.

(2) Transition probability functions are equal for increases and decreases for each cell, but are a function of the square root of the area represented by each cell.

The square root of the area expressed as a percentage of the size of the area ranges from 10 for the smallest cell (1×10^2) to 0.02 for the largest cell ($1 \times 10^{7.3}$). Each cell has a slightly lower TP for a decrease (D) than the smaller of the two adjacent cells has for an increase (I) and so at equilibrium will have more species than the smaller cell. The system at equilibrium was computed and displays a curve slightly above that labeled 0.87 in figure 9.

The relation between length of border and

area of range will depend on the shape of the range including its gross dimensions and the fine structure of the border (Rapoport, 1982). The smallest border for a given area occurs when the area is circular. An oval, rectangle, or any other shape for an area of the same size will have a longer border. There is no maximum border length in theory unless one postulates a limit on how finely folded a border can be (Mandelbrot, 1982).

It has been noted (Cain, 1944; Anderson, 1984a) that some species ranges tend to have greater longitudinal spans than latitudinal spans. Rapoport (1982, p. 151) plotted the long axes of mammalian ranges on a map of North America and of bird ranges on a map of South America and thus presented a more detailed picture of the orientation of long axes. Most of these axes parallel coastlines or mountain ranges rather than latitudinal lines.

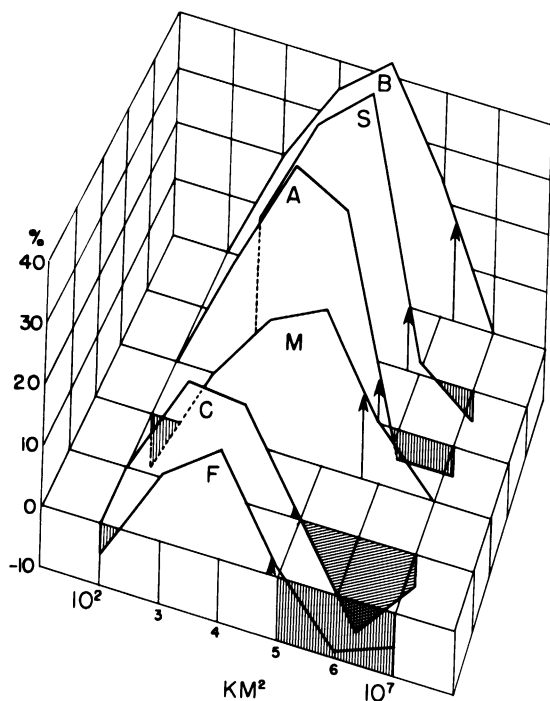If the average North American mammal's

FIG. 8. The residuals or percentage points by which cumulative distributions of range sizes of different groups depart from the Basic Logarithmic Model as explained in text. The range size with the greatest departure is the "most favored." The groups are fish (F), salamanders (C for Caudata), mammals (M), frogs and toads (A for Anura), snakes (S), and birds (B). The geometric means of range sizes (not of the departure values shown) are indicated by arrows.

range were a straight-sided rectangle, the length would need to be 25 times the width in order to provide a border of the measured length (based on Rapoport's measurements for 62 species). Because actual borders are irregularly curved rather than straight lines, the greatest length of a range is generally less than 25 times its gross width. In Rapoport's data the border is 10.34 times the square root of the area on the average, and the correlation of area and this multiple (values from Rapoport's table, p. 26, calculated by me) was $r = 0.54$. The range of a population (such as a colony of mold on an agar plate) expanding from a propagule in an isotropic and suitable medium is circular. Few vertebrates are in

that situation and a nearly circular range is unusual. These are all interesting aspects for consideration in developing more detailed models than those discussed here.

Model F. An Extinction Model. In models A to E, the number of species has been held constant and the effects of different assumptions about range sizes and probability functions upon the RS distribution have been examined. In this model extinction will be assumed to occur and its effects will be examined. Assumptions of this model are:

(1) Cells are equal increments on a logarithmic scale as in models B to E.

(2) Transition probabilities are equal as in Model B.

(3) Extinctions occur from cell 1 with the same probability as all other transitions.

In the first computer simulation under these assumptions (fig. 10), TP is set at 0.2 (for all transitions between any two adjacent cells and for the transition from cell 1 to extinction). Since no provision has been made for additions of species to replace those removed by extinction, this system will gradually decline in diversity and converge on total extinction of the fauna. What we are interested in examining is the effect of the extinctions through the one-way gate at the bottom of the RS distribution upon the shape of the cumulative curve for that distribution. This simulation begins with 1080 species uniformly distributed among cells (represented by the diagonal line in fig. 10). The original fauna of 1080 has been reduced to 887 after 400 cycles. If the data for the 400-cycle curve below the diagonal are plotted as cumulative percentages of these 887 species, the curve represented by the broken line above the other curves is produced.

Model G. Extinction with Replacement Model. The previous Extinction Model converges on an equilibrium only in oblivion, when all species are extinct. In order for the system to reach a different equilibrium it is necessary to postulate the addition of new species. Two sources of new species have been noted above. For purposes of this simplified model, the computer program was modified to add a new species whenever a species became extinct. Assumptions of this model are:

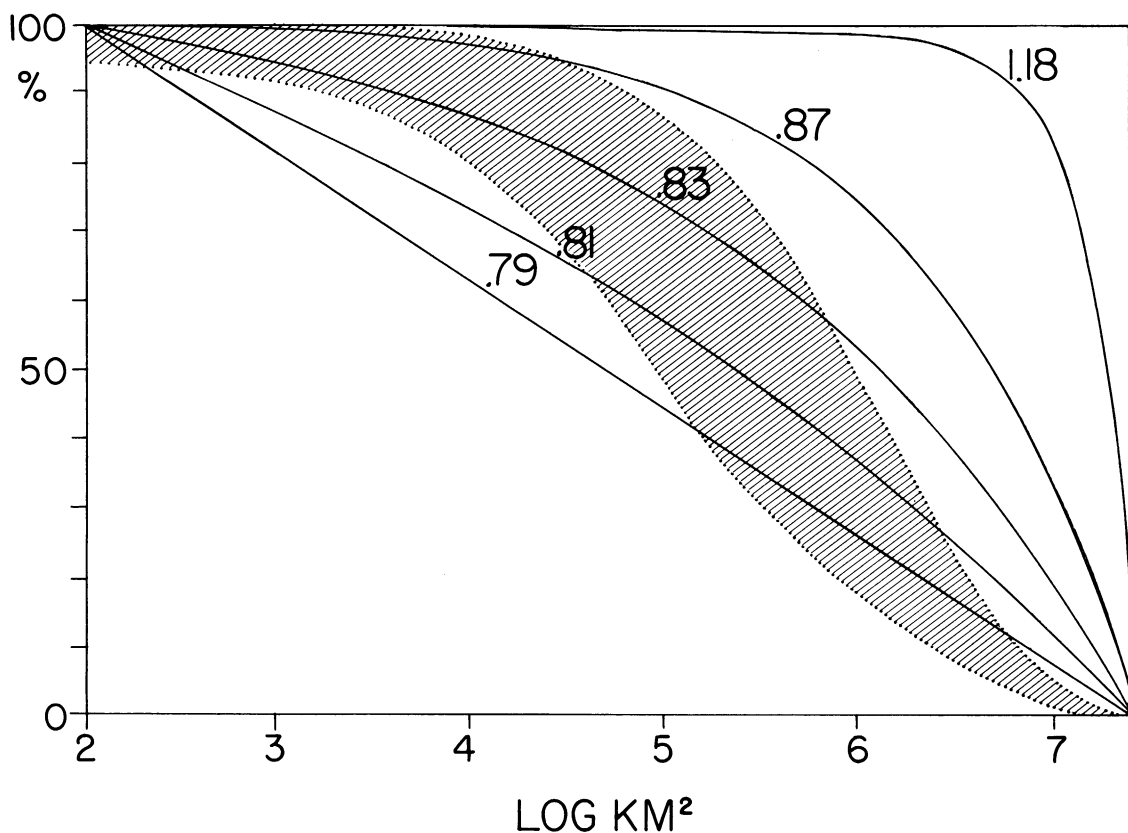(1) Cells are logarithmic as in models B to F.

FIG. 9. The Reduced Ratio Model. Graph of a sample of five cumulative curves showing equilibrium states when the I/D ratio is greater than 0.79. These equilibrium curves all lie above the diagonal. The cumulative curves for actual RS distributions of North American vertebrates all lie within the shaded area. The lower border is the curve for freshwater fish and the upper border is that for birds.

(2) Transition probabilities (set as 0.2) are equal as in Model F.

(3) Extinctions occur from cell 1 as in Model F.

(4) A new species with range size of 4.0 on a log scale, as explained below, is added every time that an extinction occurs.

The results of a computer simulation of Model G are shown in figure 11 (numbered solid curves above the diagonal line). Model G does provide a better fit to actual RS distribution curves (such as the left curve in fig. 5) than any of the other models, however there are the following noticeable differences:

(1) Conspicuous peaks in the residuals (as shown in fig. 9). If the base level were adjusted to reflect the equilibrium curve being approached by the Extinction with Replace-

ment Model rather than the diagonal line (of fig. 11, which is also the baseline for fig. 9), these peaks would be lower but would remain.

(2) Most peaks are at larger range sizes than in the model.

(3) Species in the largest range sizes in most groups are relatively few.

The above selection of 4.0 as the range size of species to be introduced was arbitrary but not capricious. The rationale for selection is as follows. If a range of any given size is assumed to divide at random (as a broken stick in this example, for reasons discussed by Anderson and Evensen, 1977) into two parts (for the fission products or species present after an allopatric or parapatric vicariant event of speciation), the larger of the two parts on the
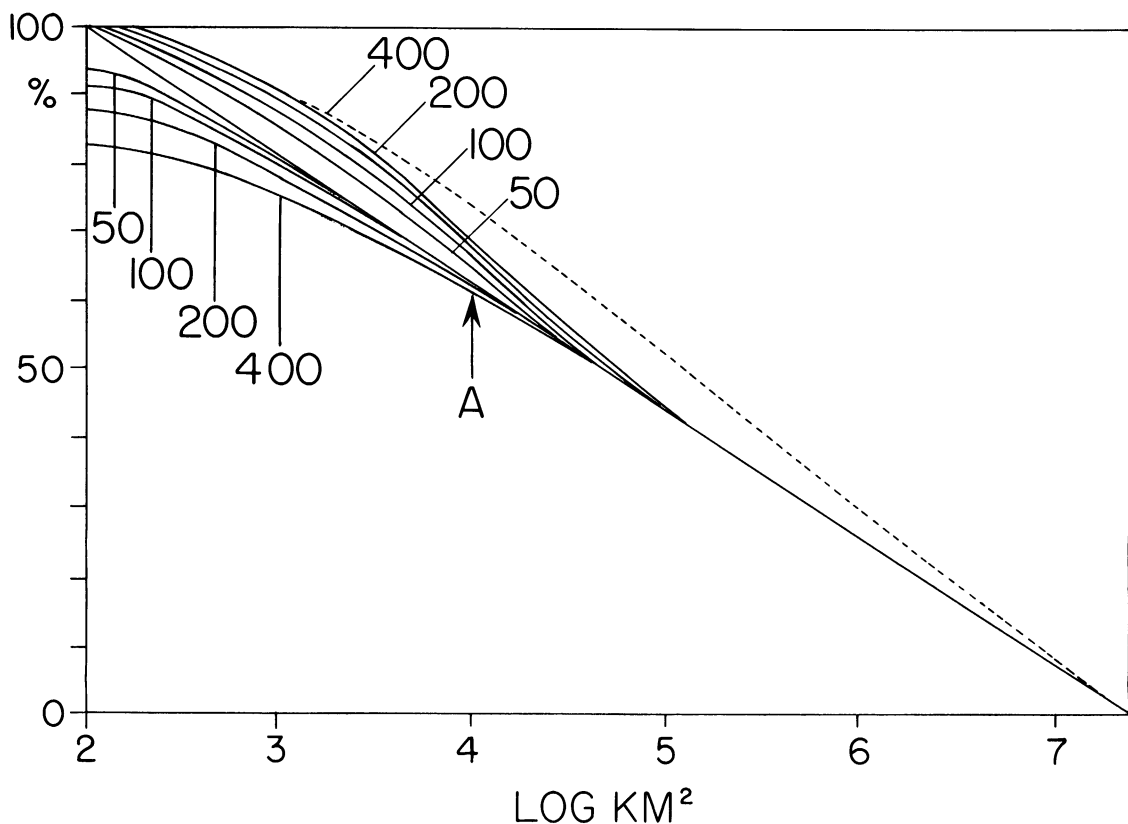
FIG. 10. The Extinction Model (shown below the diagonal line, beginning with 20 species in each cell and with all transition probabilities set at 0.2) after various numbers of cycles. The percentage scale for these curves is of the initial fauna of 1080 species, which declines over time with extinctions, there being no replacements to compensate for these losses. Above the diagonal is the "Extinction-Replacement Model," which has conditions as in the Extinction Model except that a new species having a range size (A) of 4.0 is added for each species that becomes extinct. The broken line is explained in text.

average will equal 0.75 of the original range and the smaller will equal 0.25. Because of the logarithmic scale of cell sizes, the two resulting ranges will, on the average, fall in the first and the sixth cells below the original cell. For example, a range of 100,000 would divide into ranges of 75,000 and 25,000. The logs of these three values are roughly 5.0, 4.9, and 4.4. Since the median cell has a size of 4.7, its fission products would have sizes of about 4.6 and 4.1. The latter value was rounded to 4.0 for the size of the added species and the originating value was not changed. This simplified the calculations and was deemed acceptable since the TP values tend to reset the RS values toward equilibrium relatively rapidly anyway.

In the following model this splitting of species to yield additional species is expressed in more detail. The presumed size or range of the average species being split is increased by an order of magnitude to see if the discrepancy (2, just above) of larger range sizes would be thereby reduced, and a better fit of model to data achieved.

Model H. A Second Extinction-Replacement Model, assumes that species with smaller ranges are less likely to split than species with larger ranges and will examine how this affects the RS distribution curve. The assumptions of this model are:

(1 to 3) As in model G and earlier models.

(4) The average species undergoing fission is assumed to have an RS value of 5.7, or ten times the median size of 4.7 in Model G.

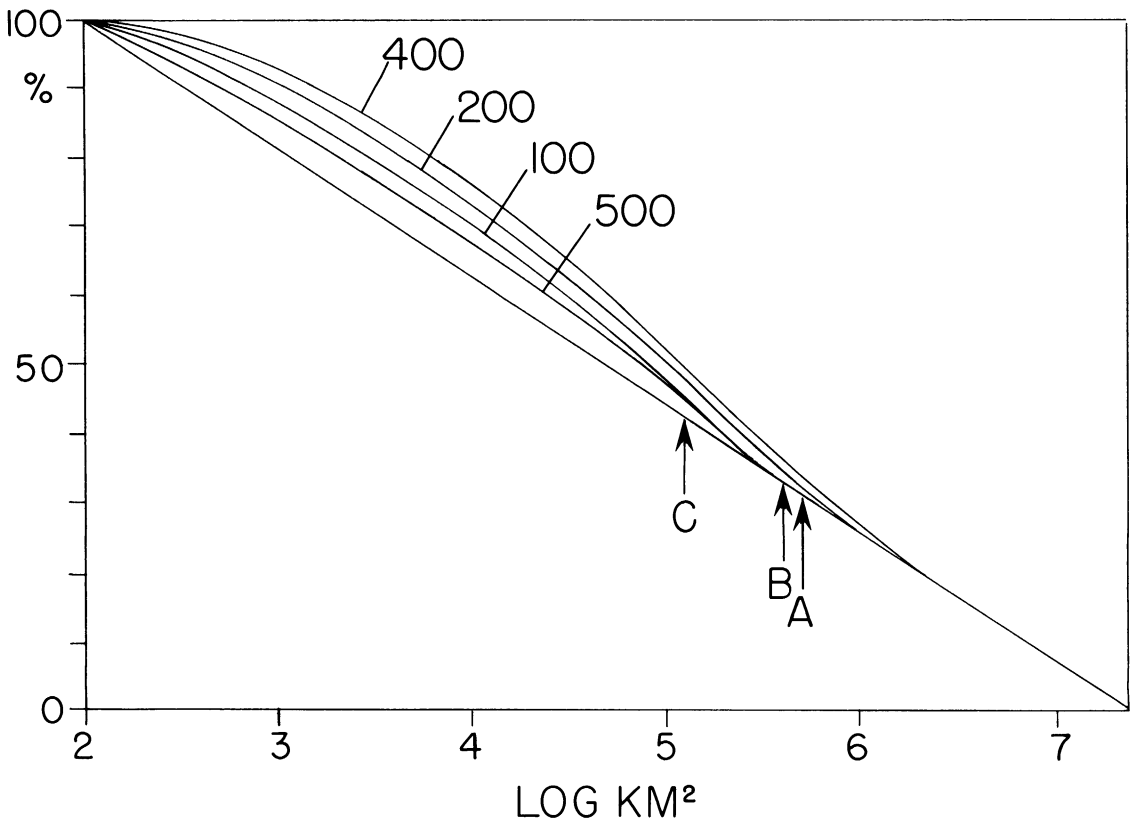The computer program was modified to

FIG. 11. The Extinction-Replacement Model as in figure 10, but with a species of size A divided into two species of sizes B and C each time a species becomes extinct. The simulation began with equal numbers of species in each cell.

reduce the RS of one species (the one being split) from 5.7 to 5.6 and to add a species at 5.1 each time an extinction occurs from cell 1.

Results of the simulation under Model H are shown in figure 11. The position of the greatest departure of the curve from the Null Hypothesis lies farther to the right than in Model G, but otherwise the shape of the curve is not much closer to the data. In the data the peak is higher and the curve drops below expectation in the high range sizes.

The models tested thus far show that the introduction of extinctions and species divisions moves the theoretical RS distribution curve of the Basic Logarithmic Model toward the actual RS distribution curves, as does the division of species with larger than average ranges rather than species selected at random. The actual RS distribution curves differ from the best approximation under the conditions

of models A to H. In these models there are relatively uniform gradients of transition probabilities. In reality, the probabilities for changes of ranges vary in a slightly more subtle way.

Model I. A sigmoid Curve Model. The assumptions of this model are:

(1) Cells are logarithmic as in models B to G.

(2) There is an inflection point of geographic range, below which the ratio of I/D is more than the basic ratio of Model B, and above which the ratio is less.

By changing one or more of the three values, namely the inflection point of the sigmoid curve of the cumulative distribution, the ratio of I to D below the favored size, and the ratio above it, we can observe how the change affects the RS distribution curve in computer simulations.
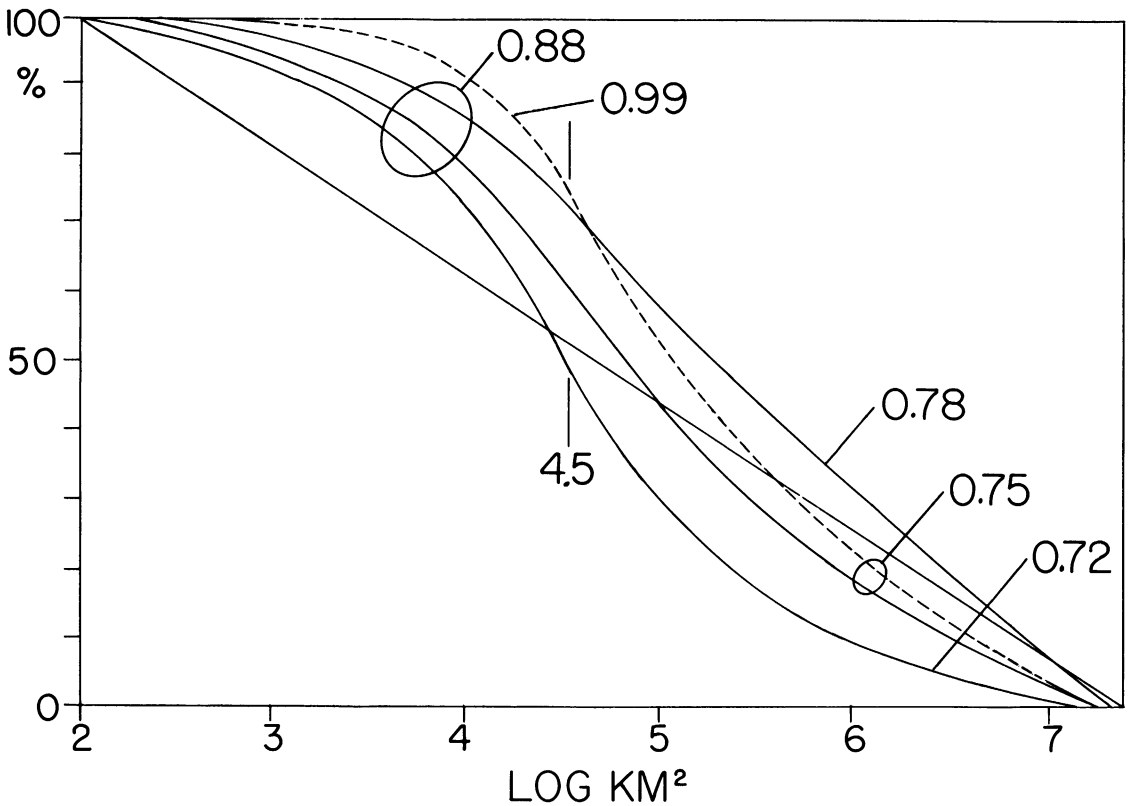
FIG. 12. The Sigmoid Curve Model. Computed cumulative equilibrium curves when the inflection point is assumed to be 4.5. Below this value the ratio of the probability for an increase (I) and that for a decrease (D) is less than the ratio (0.79) stipulated in the basic logarithmic model. Ratios of 0.9 and 0.8 (in the area below 4.5) are shown in sample curves, and (above 4.5) ratios of 1.01, 1.05, and 1.1 are illustrated. See discussion in text.

Although species with ranges smaller than the inflection point have a greater ratio (I/D) than occurs under the conditions of Model B, the basic probability for contraction, as measured in absolute arithmetic units, is still greater than for expansion.

In figure 12 are shown the computed equilibrium curves (cumulative percentages) when the I/D ratio is set at 0.88 for range sizes below 4.5 and at three different values less than 0.79 above 4.5. As seen in figure 7, a ratio of less than 0.79 produces a curve that bulges downward and (as seen in fig. 10) a ratio of more than 0.79 produces an upwardly bulging curve. For comparison, the broken line shows the equilibrium curve when the TP ratio below 4.5 is set at 0.99 instead of 0.88 and the ratio above 4.5 is set at 0.75.

The cumulative curve now approaches the sigmoid form of the real curves (within the shaded area of fig. 10). The inflection point (which is 4.5 in the examples of fig. 13) must be adjusted as shown in figure 14 to closely approximate the real curves. By trial and error it was established that the inflection point here is in the 5.0–6.3 range. The I/D ratio below the inflection point must be adjusted to the 0.85–0.89 range and above the inflection point it must be adjusted to the 0.66–0.75 range.

In summary, the real curves can be simulated rather closely by specifying (1) the minimum and maximum values, (2) a logarithmic series (of cell dimensions or of TP values), (3) the inflection point, and (4) the TP ratios, which are always more than 0.79 below and always less than 0.79 above the inflection point. Furthermore, the difference
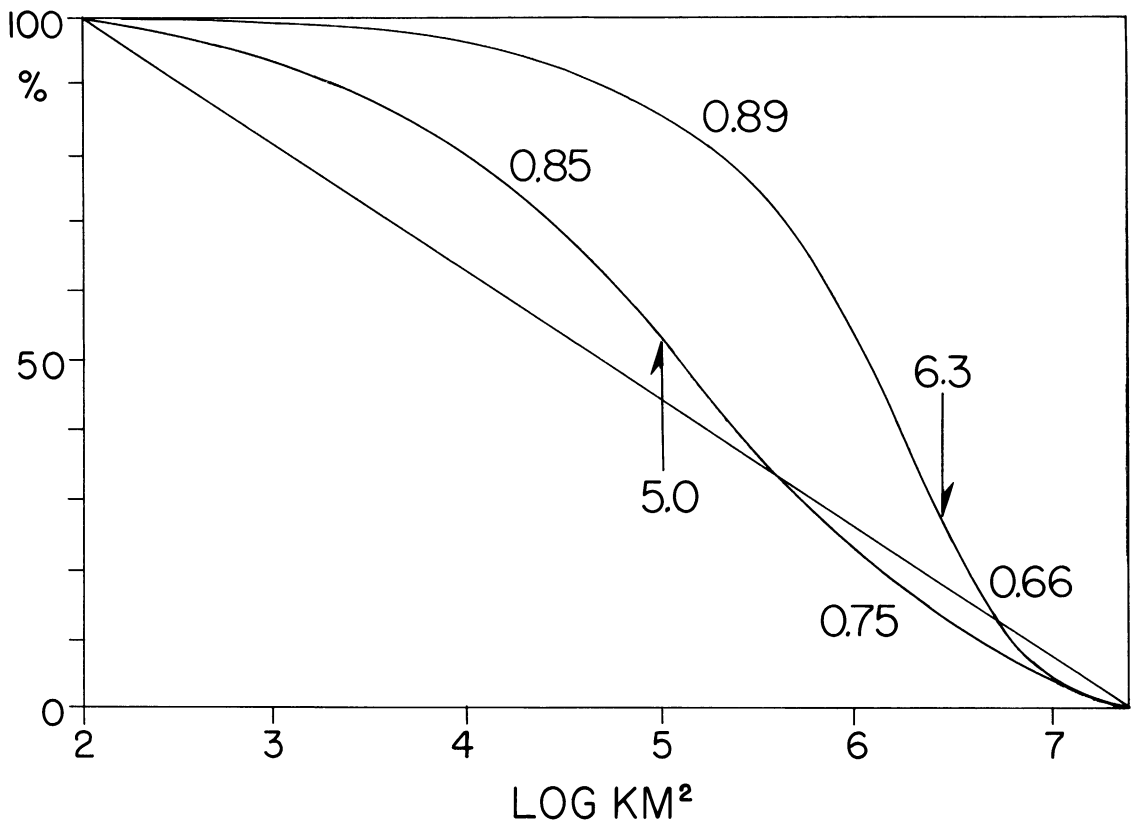
FIG. 13. Sigmoid Curve Model as in figure 12, but adjusted so that the curves approximate actual data curves such as the upper and lower curves of the shaded area in figure 9. See discussion in text.

from 0.79 of this ratio is always greater below than above the inflection point.

The three fitted variables, used to obtain the curves in figure 13, are the I/D ratio among range sizes of less than the inflection point, the ratio among values greater than the inflection point, and the inflection point itself. These values are shown in figure 13. The lower curve approximates the real curve for fish and the upper curve approximates the real curve for birds.

EXPLAINING THE RS DISTRIBUTION

The hollow curve or logarithmic frequency distribution is the most obvious general pattern when range sizes for a continental fauna are considered. It may well be asked what, if any, ecological implications the RS distribution has, or how it may be explained eco-

logically or in terms of the individual species of the fauna.

The "Basic Logarithmic Model," in which all species are equally likely to increase or decrease their ranges sufficiently to move from a given logarithmically dimensioned cell to an adjacent cell, is a (Null) hypothesis which accounts for most of the variation in range sizes of North American vertebrates. But there is residual deviation of data from this Null Hypothesis including differences between groups (fig. 8) and patterns within groups that need further explanation.

It might become possible to reason from knowledge or theory about the characteristics of individual species, other than the measurements of range sizes themselves, to the observed RS distributions. However, any such reasoning would need to consider the environment as well as the characteristics of the

species. It is not possible to reason from the observed generality about RS distributions to what the RS for any one species will be.

The main ecological implication here is that in many complex situations, ranging from the "community" ecology of microorganisms or mites in a cubic meter or soil to the biota of a continent, the best single predictor of certain major patterns seems to be luck.

The major patterns of taxonomic diversity described earlier (Anderson, 1974) and of RS distributions summarized here cannot now be derived from any reasonable set of deterministic assumptions known to me, nor from premises established by study of lower levels of ecological and evolutionary organization. Perhaps our knowledge will increase to the point that this will no longer be true.

"Philosophically, this approach does not refute a deterministic position, for an appearance of randomness can result from great ignorance as well as from real disorder. Practically speaking, it means that we are a long way from a 'causal' explanation for any specific large-scale example of faunal evolution or of ecological dynamics and that our predictions about these things are likely to be poor in most cases" (Anderson, 1974).

In the few cases where range sizes have been correlated with some other biological property of the same set of species or with some environmental variable, the analysis generally ends with the correlation. For example, it has been observed that larger fish (McAllister et al., in press) and larger mammals (Anderson, 1977; Brown, 1981; Rapoport, 1982; Van Valen, 1973) have larger geographic ranges, on the average. Such correlations may be in the neighborhood of $r = 0.4$. However, this is not a high correlation and it may also be true that among the species of any given size there are more species with smaller ranges than with larger ranges, and we are left with questions such as—why are logarithmic distributions so prevalent? How do body size, trophic level, heat tolerance, home range size, tooth size, competition, and other biological properties interact and relate to range size?

Pianka (1977), Kiester (1971), and McAllister et al. (in press) have used multivariate methods or multiple regression analysis to relate the species densities of different groups to each other and to various environmental factors, but none of these studies has directly included range sizes in the analysis. Smaller ranges have been associated in one context or another (summary in Anderson, 1984b, 1985) with more competition, greater species density, younger taxa, stenotopy, narrow ecological niches, reduced vagility, low population density, smaller individuals, lower trophic levels, less available space, and patchier distributions. Usually the association is suggestive rather than definitive. Rarely is a set of measurements of range size correlated with data on one of these factors, and I know of no analysis attempting to compare data on range size with data on two or more of these factors.

There has been some recent interest in the correlation of population density and range size. Bock (1984) summarized the few published works on the subject and used data on 70 species derived from Christmas Bird Censuses in the United States and southern Canada to derive correlations of from 0.346 to 0.476. J. H. Brown (personal commun.) has correlated ($r = 0.574$) the mean densities of individuals of 25 North American species of raptors and the areas of their winter ranges. The correlation of sizes of individuals and sizes of the geographic ranges of 144 species of North American fish has been calculated by McAllister et al. (in press), $r = 0.40$ for area versus length, and 0.38 for log area versus log length.

These values suggest the range of correlations that may be expected, but detailed comparisons of reported studies will be needed to evaluate the meaning of these and other differences.

The range of every species is determined by an interaction of the species and the environment on both short-term and long-term time scales. The physiological, morphological, and genetic properties of the populations of individuals comprising the species are all involved.

If the environment changes more rapidly than a species can adapt, then the species becomes extinct, and sooner or later this is what happens. Most species are extinct. Since most environmental changes are independent of organismic changes, and in significant measure are unpredictable, luck (or stochas-

tic processes in the jargon of science) plays a major role in determining the success of species. They gamble for existence as well as struggle for it, and in the long run most of them lose (Reddingius, 1971).

The existence of stochastic processes has been acknowledged often, but the relevance and importance of luck are not often addressed. Tilman (1982, p. 4), for example, wrote:

> Admittedly, numerous stochastic processes ... decrease the ability of any theory to predict accurately the dynamics of populations. However, this book is based on the assumption that the interactions among species and of species with their environment are sufficiently strong as to establish major patterns which are discernible over such stochastic noise.

The border of a geographic range of a species is where it is because of the dynamic interaction of organisms and environment. Beyond the border either conditions are not suitable, or insufficient time has elapsed since the species arrived or since the environment changed from unsuitable to suitable. The concept of suitable is not absolute but is relative to space and time. An environment that is unsuitable for a species at one time can become suitable by evolutionary adaptation by that species.

Our understanding of these complex interactions can be increased by more and better observational data, by better theory (including the formulation of more testable hypotheses), and by using conceptual models. With these models it is easier to ask complex "what if" questions. However, it may not be easier to get data to test the models; in fact, it may be quite laborious simply because more data are needed for complex multivariate analyses.

LITERATURE CITED

Anderson, S.
  1974. Patterns of faunal evolution. Quart. Rev. Biol., vol. 49, pp. 311–322, 15 figs.
  1977. Geographic ranges of North American terrestrial mammals. Amer. Mus. Novitates, no. 2629, pp. 1–15, 5 figs.
  1984a. Geographic ranges of North American birds. *Ibid.*, no. 2785, pp. 1–17, 9 figs.
  1984b. Areography of North American fishes, amphibians, and reptiles. *Ibid.*, no. 2802, pp. 1–16, 9 figs.
Anderson, S., and Charles S. Anderson
  1974. Three Monte Carlo models of faunal evolution. Amer. Mus. Novitates, no. 2563, pp. 1–6.
Anderson, S., and Mary K. Evensen
  1977. Randomness in allopatric speciation. Syst. Zool., vol. 27, pp. 421–430.
Bock, Carl E.
  1984. Geographic correlates of abundance vs. rarity in some North American winter land-birds. Auk, vol. 101, pp. 266–273.
Boswell, M. T., and G. P. Patil
  1971. Chance mechanisms generating the logarithmic series distribution used in the analysis of numbers of species and individuals. Pp. 99–127. *In* Patil, G. P., E. C. Pielou, and W. E. Waters (eds.), Statistical ecology, vol. 1, Spatial patterns and statistical distributions. Pennsylvania State Univ. Press, University Park, 582 pp.
Brown, J. H.
  1981. Two decades of homage to Santa Rosalia: toward a general theory of diversity. Amer. Zool., vol. 21, pp. 877–888.
Cain, S. A.
  1944. Foundations of plant geography. New York, Hafner, 556 pp.
Kiester, A. Ross
  1971. Species density of North American amphibians and reptiles. Syst. Zool., vol. 20, pp. 127–137.
Levandowsky, M., and B. S. White
  1977. Randomness, time scales, and the evolution of biological communities. Evol. Biol., vol. 10, pp. 68–161.
Mandelbrot, B. B.
  1982. The fractal geometry of nature. Freeman, New York, 468 pp. and 16 pls.
McAllister, D. E., S. P. Platania, F. W. Schueler, M. E. Baldwin, and D. S. Lee
  [in press] Ichthyofaunal patterns on a geographic grid. Chapter in Zoogeography of freshwater fishes of North America. Hocutt, C. H., and E. O. Wiley (eds.). New York, Wiley.
Pianka, E. R.
  1977. Reptilian species diversity. Pp. 1–34. *In* Gans, C., and D. L. Tinkle (eds.), Biology of the Reptilia, vol. 7, xvi + 720 pp.
Rapoport, Eduardo H.
  1982. Areography: geographic strategies of species. New York, Pergamon Press, xvi + 269 pp.

Reddingius, J.
    1971.   Gambling for existence. Bibliotheca
            Biotheoretica, vol. 12 (=Acta Biotheo-
            retica, vol. 20, suppl. 1), pp. 3–208.
Tilman, David
    1982.   Resource competition and community
            structure. Princeton Univ. Press,
            Monographs in Population Biology, no.
            17, pp. 1–288, 98 figs.

Van Valen, Leigh
    1973.   Body size and numbers of plants and
            animals. Evolution, vol. 27, pp. 27–35.
Williams, C. B.
    1964.   Patterns in the balance of nature and
            related problems in quantitative ecol-
            ogy. New York, Academic Press, vii +
            324 pp.