PHYSICAL REVIEW E VOLUME 55, NUMBER 3 MARCH 1997

Speciation and extinction in a simple model of evolution

D. A. Head* and G. J. Rodgers†
Department of Physics, Brunel University, Uxbridge, Middlesex UB8 3PH, United Kingdom

(Received 6 November 1996)

We introduce a simple model of macro-coevolution that allows the branching and termination of species lines and also incorporates external influences to the ecosystem. The strength of the external influences and the likelihood of speciation and extinction are defined from the fitness landscapes by two parameters, δg and δs. Results from numerical simulations show that the total number of species fluctuates about a natural system size N_∞. We present a mean-field theory that predicts $N_\infty \propto (K-1)\delta s/\delta g^2$, where $K-1$ is the system connectivity and δs is small. This result compares well with the numerical simulations. For large δs, we demonstrate why this expression changes to $N_\infty \propto (K-1)\delta s^2/\delta g^2$. We compare the model to the fossil record, and comment on the role of ecological niches in models of evolution. [S1063-651X(97)15403-4]

PACS number(s): 87.10.+e, 05.40.+j

I. INTRODUCTION

Attempts to model the evolution of the Earth's ecosystem have been hampered by its sheer size and complexity. Exact analysis of even small subsystems is often impractical, so tractable models of the large-scale dynamics can only be constructed after considerable simplifications have been made. To this end, Bak and Sneppen [1] have introduced a model that captures the essential ingredients of a coevolving system without trying to account for the complex internal dynamics of each species. Each species is represented by a single number, which is related to the average time it takes to evolve to a new form. This simplification allowed extensive numerical and analytical studies to be performed, which nonetheless revealed similarities with the fossil record, most notably the existence of punctuated equilibria and a power-law distribution of extinction sizes.

In the original Bak-Sneppen model, extinction is modeled by a species changing form in response to the change in form of an interacting species. This is a rather narrow definition, so recently a variety of models have been introduced that incorporate more sophisticated mechanisms for extinction. One approach [2] has been to account for influences external to the ecosystem, such as meteor impacts, volcano eruptions, etc., by assigning a fitness value to each species. In addition to the usual dynamics, all species with a fitness less than some global random noise value are also made extinct. A number of models have allowed for the permanent termination of species lines, caused either by the population reducing to zero [3], viability dropping below a threshold value [4], or competition from similar species in neighboring regions [5]. Generally, these models exhibit features that agree with observed paleobiological findings.

All these models [1–5] have tightly controlled the total number of species in the system, usually having a fixed system size where every species made extinct is immediately replaced by a new one. This has been justified by the concept of *ecological niches*, where it is assumed there are a constant number of niches in the system, each of which contains exactly one species. This is in direct contradiction to the fossil record [6], which clearly shows a wide variation in lifeform diversity, including periods of both apparent stasis and rapid growth. One previous attempt to tackle this problem allowed each species the possibility of speciation (splitting into two subspecies) or extinction whenever it evolved to a new form [7]. Depending upon how this probability is defined, the number of species either tends to zero or increases without limit as the system evolves.

In this paper, we present a modified Bak-Sneppen model in which speciation and extinction are defined purely from the internal dynamics of each species. The system evolves to a state in which the total number of species fluctuates around a steady state value. We refer to this value as the *natural system size*, because there is nothing that directly controls its value or even its existence (for instance, there is no reference to niches). The model is defined and justified in Sec. II. In Sec. III, results obtained from numerical simulations are compared to the fossil record. A mean-field theory for the model has been derived that predicts how the natural system size depends upon the system parameters, where there are now three such parameters. This analysis is explained in full and compared to numerical results in Sec. IV. The main results are summarized and some possible extensions to the model are suggested in Sec. V.

II. DEFINITION OF THE MODEL

A key component of Darwinian evolution is that variations in an organism's form (that is, its *phenotype*) can alter the probability of that organism's survival, which, when combined with a suitable hereditary mechanism, gives rise to natural selection. Rather than try to reduce the evolution of a species to its component individuals or genes, however, models based on the Bak-Sneppen approach take the whole species as the fundamental unit of selection. The relationship between phenotype and survival probability is quantified by a *rugged fitness landscape* [8], which is a function of the species itself, the species directly connected to it in the food chain, and also of environmental factors. A schematic example is given in Fig. 1, where for simplicity we have com-

---

*Electronic address: David.Head@brunel.ac.uk

†Electronic address: G.J.Rodgers@brunel.ac.uk

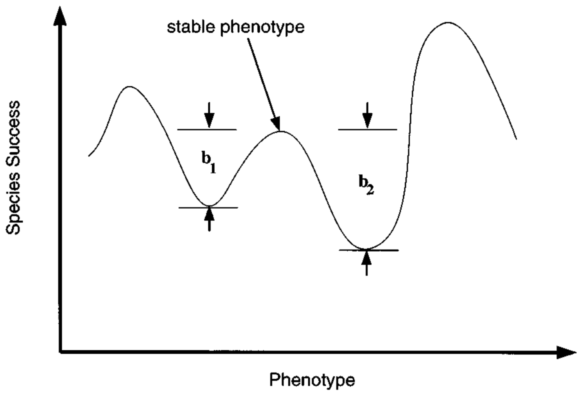

FIG. 1. A schematic diagram of a fitness landscape. Local maxima correspond to stable phenotypes. The two smallest barriers against mutation are given by b_1 and b_2.

pressed all possible phenotypic variations onto the single horizontal axis. The vertical axis corresponds to some measure of the species survival, which could be any one of a number of definitions for biological fitness [9]. Local maxima are the attractors of these landscapes, in much the same way that local minima are the attractors of potential energy landscapes. A species passes between adjacent maxima with a probability exponentially small in, and hence on a time scale exponentially large in, the barrier height. Previous models have considered only the smallest barrier against mutation, citing the exponential time differences as justification for rejecting the larger ones. Since we eventually wish to include branching into two different species, we reject all but the two smallest barriers, b_1 and b_2.

Our model is defined as follows. The system consists of N species, each of which is completely defined by two real numbers, $\{b_{1i},b_{2i}\}$, $i=1,\ldots,N$, corresponding in no particular order to the two smallest barriers against mutation. Initially each barrier is uniformly distributed on the interval $[0,1]$. At every time step, the lowest of the $2N$ barriers is found. This is the b_{ki} that obeys $b_{ki}<b_{mn}$ for all $m\neq k$ and $n\neq i$, where $k=1$ or 2. The species i is then *mutated* — that is, both b_{1i} and b_{2i} are assigned new values from the uniform distribution, corresponding to a new fitness landscape. Note that both barriers are changed irrespective of the value of k. To incorporate coevolution, those species linked to i in a manner yet to be defined are similarly mutated.

As the model has been defined so far, the barriers are static between mutations and so the larger barrier for each species is not involved in the dynamics. Therefore if we were to remove the larger of the two barriers upon every mutation, we would have the same system with the same dynamics but now with just one barrier per species. This process of choosing the smallest of two uniformly distributed barriers can be mimicked by drawing a single barrier from the different distribution $P(x)=2(1-x)$, $x\in[0,1]$. Thus we can map the two-barrier model onto the original Bak-Sneppen model with barriers drawn from a modified probability distribution. This change is unimportant since the model is robust — that is, its essential behavior is insensitive to arbitrary details such as the particular choice of probability distribution. With the inclusion of speciation and extinction, however, both barriers become dynamically involved and so this mapping is no longer valid.

Before coming to these new features, it should be noted that the fitness landscape is also dependent on events unrelated to other species, including geographical, climatic, and astronomical changes. The influence of these factors on each species' fitness landscape, and hence their barriers, is likely to be very complex. We assume that their total effect is random and incorporate this into the model as noise of $O(\delta g)$, where δg is a new parameter. Specifically, we say that at every time step, each barrier in the system is transformed according to

$$b_{ki}\rightarrow b_{ki}+\delta g_{ki}, \quad i=1,\ldots,N, \quad k=1 \text{ or } 2 \tag{1}$$

where the $\delta g_{ki}(t)$ are uniformly distributed on $[-\delta g/2,\delta g/2]$ and uncorrelated in time. We shall call this δg noise.

We now consider extinction. Although a species that undergoes a burst of mutations may appear in the fossil record as a separate extinction and speciation, we prefer to think of this as a psuedoextinction and reserve the term extinction for the permanent termination of a species line. There are many possible mechanisms for extinction given in the literature [10,11]. Within the context of this model, we define extinction to be the inability for a species to adapt to new conditions or competitors. In terms of the barriers, a species becomes extinct if, when it is chosen for mutation, it has both $b_{1i}>1$ and $b_{2i}>1$. The justification for this is that a species with only very large barriers against mutation has become so inflexible that it is no longer able to adapt and dies out. Fixing the threshold value to be 1 may seem to be somewhat arbitrary, but the fitness scale is entirely arbitrary anyway and fixing the value like this removes the need for extra parameters. The model has proved to be robust under a variety of extinction thresholds, including ones defined by probability distributions. The value of 1 was chosen so that there would be no extinction when $\delta g=0$ and the original model might be recovered. Since extinction and speciation are to be treated independently, the vacancy left by an extinct species is not immediately refilled and the system size is reduced by 1, $N\rightarrow N-1$.

Speciation occurs when two subpopulations reach a state of reproductive isolation and hence should be considered as separate subspecies. This could occur, for instance, by spatial isolation, such as the introduction of a geographical barrier, or by some form of genetic variation resulting in a hybrid zone with low fitness. In fact, the possible causes of speciation [12] are too numerous to account for independently. Within the framework of the fitness landscapes, however, there is a way of defining speciation that neatly encapsulates many of the possible biological mechanisms. Let P_1 denote the probability of species i to cross over the barrier b_{1i}, and P_2 the probability to cross over b_{2i}. Then P_1/P_2 decays exponentially in $(b_{1i}-b_{2i})$ and is, respectively, 0 or ∞ for b_{1i} significantly larger or smaller than b_{2i}. However, for $b_{1i}\approx b_{2i}$ there is a finite probability of subpopulations crossing over different barriers. For simplicity, we say that species i will speciate if, when chosen for mutation, it has $|b_{1i}-b_{2i}|$ smaller than some new parameter δs. For further realism, we could incorporate the possibility of one species splitting into three or more subspecies simultaneously, by including further barriers b_{3i}, b_{4i}, and so on. This would occur so rarely as to barely contribute to the dynamics, however.

To complete the description of this model, we must define the system connectivity. Difficulties are immediately apparent in any lattice based approach. For instance, branching would not be possible for any species that has all its adjacent lattice sites occupied. Also, the extinction of a single species in a one-dimensional food chain would split the chain into two. An alternative approach, which has already been studied with simpler definitions for extinction and speciation [7], is to arrange all species past and present on a tree, with the ends of the branches corresponding to those currently active. In this paper, however, we have chosen to simplify the dynamics by assigning to every species $K-1$ neighbors, selected at random from the remaining $N-1$ in the system [13]. These neighbors are reselected upon every mutation, hence the disorder is annealed. Since the system has no spatial definition it cannot be regarded as self-organized critical [14].

In summary, this model is defined as follows. The system consists of N species, each of which has two barriers, b_{1i} and b_{2i}, $i=1,\dots,N$. All barriers are initially drawn from the uniform probability distribution on the interval [0,1]. The algorithm for every time step then proceeds as follows. (i) The smallest barrier in the system, b_{ki}, is found, where $k=1$ or 2. Species i is then mutated, which means that both b_{1i} and b_{2i} are given new random values from the uniform probability distribution. (ii) $K-1$ other species are chosen at random and similarly mutated. (iii) If a species selected for mutation has $b_{1i}>1$ and $b_{2i}>1$, it is removed from the system. $N\rightarrow N-1$. (iv) If a species selected for mutation has $|b_{1i}-b_{2i}|<\delta s$ but is not made extinct, it branches into two new species with random barriers. $N\rightarrow N+1$. (v) Every barrier b_{ki} is transformed to $b_{ki}+\delta g_{ki}$, where the δg_{ki} are uniformly distributed on $[-\delta g/2,\delta g/2]$ and are reselected at every time step. The order of steps (iii) and (iv) can be reversed if either δg or δs is small. This robustness is lost when both parameters are large, when reversing the order will result in almost no extinction.

The original Bak-Sneppen model is recovered in the limit $\delta g\rightarrow 0$ and $\delta s\rightarrow 0$, with one crucial difference. In [1,2] the time step is set at $\delta t\sim\exp(b_{ki}/b_0)$, since this is the expected time it takes to pass over the minimum barrier b_{ki}, where b_0 fixes the scale. Our model, however, also includes δg noise corresponding to influences unrelated to the barriers and consequently on an independent time scale. Rather than include two separate time scales for the coevolutionary and external processes, we have chosen to ignore such considerations and simply set δt to be constant. The physical ramifications of this choice will be discussed when the model is compared to the fossil data in Sec. III.

With such specific definitions of general processes, it is obviously important to check for robustness. We have tried various definitions for speciation and extinction based on uniform, exponential, and normal distributions, and similarly for the distribution of the δg noise. In all these cases, the essential behavior of the system remained unchanged, the only differences being purely quantitative.

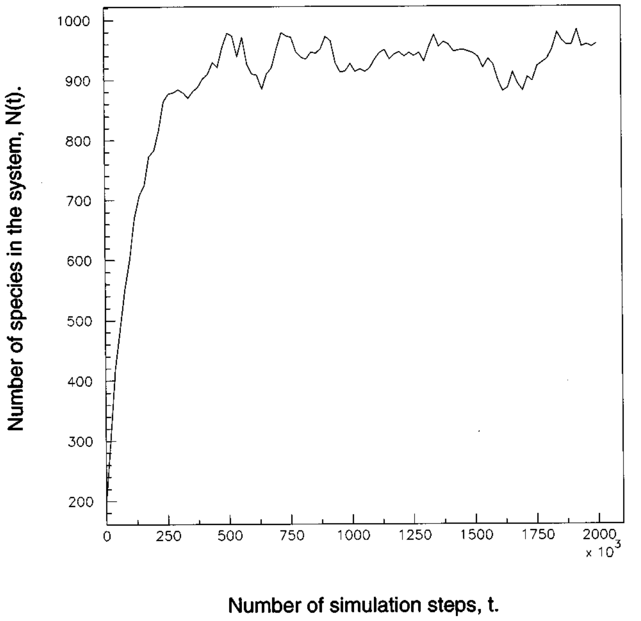

FIG. 2. Plot of the system size $N(t)$ against t, with $K=4$, $\delta g=0.02$, and $\delta s=0.001$, starting from a system with $N_0=200$ species.

III. NUMERICAL RESULTS AND COMPARISON TO THE FOSSIL RECORD

When $\delta g=0$ there is no extinction and the system size $N(t)$ increases without limit. This is also the case when $K=1$. When $\delta s=0$ speciation becomes impossible and N tends to zero. With $\delta g>0$, $\delta s>0$, and $K>1$, N approaches a constant value N_∞, which is independent of the initial value N_0. It then fluctuates around this value but never settles down within the time frame considered, typically up to around $10^4\times N_\infty$ time steps. An example is given in Fig. 2. For values of N_0 far from N_∞, $N(t)$ initially increases or decreases linearly. Note that we have implicitly assumed that δg and δs are small, since large values are physically unrealistic.

That such a natural system size should exist at all is by no means obvious, since speciation and extinction are both defined independently of N. It exists because the δg noise has a different effect on speciation than it does on extinction. The random noise is just as likely to push two barriers apart as to bring them together, so the rate of speciation remains roughly unchanged. However, the noise acts asymmetrically on barriers near the threshold for extinction, tending to push species over this threshold into the small tail corresponding to those species that will be made extinct when next mutated. The important point to remember is that the δg noise acts on all N species at every time step, hence the rate of extinction increases with N but the rate of speciation will remain constant. A steady state will be found when these two rates balance. This qualitative reasoning is confirmed by the mean-field analysis in Sec. IV.

For marine organisms, a graph of the number of families of species against time [6] initially increases, then levels out throughout much of the Palaeozoic era. After a sudden drop caused by the mass extinction at the end of the Permian period, the graph increases linearly beyond the previous plateau and is still increasing to the present day. We present the following possible explanation for this behavior based on our model. The marine record could be viewed as being split into two parts separated by the end-Permian mass extinction, where the more recent part has a much higher value for N_∞. This would certainly account for the shape of the fossil record, but we still need to justify the different values for N_∞. This comes from closer examination of the mass extinction itself [15], which caused the predominantly sedentary marine life to quickly evolve to much more mobile forms. Increased mobility would mean a lessened susceptibility to geographic fluctuations, hence δg would decrease and so N_∞ would indeed increase (see Sec. IV).

The data for continental organisms does not agree as well. In this case, the plot of diversity shows an increase faster than linear and no apparent plateaus. The model we have presented clearly fails to account for this behavior and we can only suppose that more realistic enhancements may improve its validity. One such modification would be to allow the tendency to branch or become extinct to itself be a hereditary characteristic, that is to replace the global parameters δs and δg with species-dependent δg_i and δs_i. Simulations allowing such quantities to vary randomly by a small amount upon mutation do indeed appear to remove the steady state and have N increase faster than linearly. This is to be expected, since branching now favors species with a larger δs_i. It would be interesting to see what a detailed study of such a model might reveal.

The rate of mutation is $K\delta t$, which is constant. This is not true for most other Bak-Sneppen models, where δt is a function of the minimum barrier as explained in Sec. II. It was this variation that allowed for the power-law distribution of extinction sizes, which was one of the central results of the original study. We cannot hope to reproduce such behavior without first separating the time scales for mutation and δg noise, something that would require the introduction of at least one extra parameter. The extinction of species lines in our model follows a Poisson distribution, as expected for a random selection process.

IV. MEAN-FIELD ANALYSIS

In this section we extend the mean-field analysis of Flyvbjerg *et al.* [16] to our model. Although a full solution was not obtained, unsurprising in light of the dynamical complexity, we have derived an expression for the dependence of N_∞ on the parameters K, δs, and δg, which agrees with the numerical results. Initially, however, we ignore extinction and speciation and just consider how the analysis can be expanded to cope with pairs of barriers.

We define $p(x,y)dxdy$ to be the probability that a randomly selected species i has $x<b_{1i}<x+dx$ and $y<b_{2i}<y+dy$. This can be related to $p_{\min}(x,y)$, the distribution for when x or y is the lowest barrier in the system, by

$$p_{\min}(x,y)=Np(x,y)Q^{N-1}(x,y), \tag{2}$$

where $Q(x,y)$ is the probability that a species has both barriers greater than the smaller of x and y,

$$Q(x,y)=\int_m^1\int_m^1 p(x',y')dx'dy', \tag{3}$$

where $m=\min(x,y)$. The justification of Eq. (2) is that the species with the minimum barrier could be any one of N in the system, and the other $N-1$ species must have both barriers larger than the minimum. Note that $p(x,y)=p(y,x)$ and the time dependence is implicit. At each time step, $p(x,y)$ will change by an amount $\Delta p(x,y)$ defined by

$$\Delta p(x,y)=-\frac{1}{N}p_{\min}(x,y)-\frac{K-1}{N-1}\left(p(x,y)-\frac{1}{N}p_{\min}(x,y)\right)+\frac{K}{N}, \tag{4}$$

where the first term on the right-hand side accounts for the mutation of the species with the lowest barrier, the second for the the mutation of the $K-1$ random nearest neighbors, and the third term handles the K new pairs of barriers.

The system evolves to a state where $\Delta p=0$. Using Eqs. (2) and (4) we derive the steady state equation

$$\frac{K}{N}-\frac{K-1}{N-1}p-\frac{N-K}{N-1}pQ^{N-1}=0. \tag{5}$$

As in the original model [1], we must consider two separate regimes. For $Q<1-O(1/N)$, the third term on the left hand side of Eq. (5) vanishes in the large-N limit and we get

$$p(x,y)=\frac{K}{K-1}. \tag{6}$$

The other regime corresponds to $Q(x,y)=1+O(1/N)$ and hence, from the definition of $Q(x,y)$ Eq. (3), we can see that $p(x,y)=O(1/N)$. This means the second term in Eq. (5) will vanish, giving

$$pQ^{N-1}=\frac{K}{N}, \tag{7}$$

and hence from Eq. (2)

$$p_{\min}=K. \tag{8}$$

The boundary between these two regimes can be found by remembering that both p and $p_{\min}$ are probability distributions and so must normalize to one. The final solution is therefore

$$p(x,y)=\begin{cases}\dfrac{K}{K-1} & \text{for } x \text{ and } y>1-\sqrt{\dfrac{K-1}{K}}\\ O(1/N) & \text{otherwise,}\end{cases} \tag{9}$$

$$p_{\min}(x,y)=\begin{cases}O(1/N) & \text{for } x \text{ and } y>1-\sqrt{\dfrac{K-1}{K}}\\ K & \text{otherwise.}\end{cases} \tag{10}$$

The regime for large barriers corresponds to species that can only get mutated by selection as a random neighbor, whereas

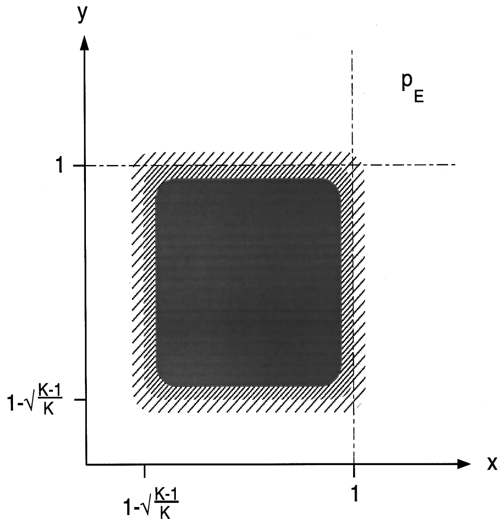

FIG. 3. Schematic of solution of $p(x,y)$ for $\delta g>0$. The shaded region is the density of $p(x,y)$, lighter gradations corresponding to lower densities. p_E refers to the region $x>1$ and $y>1$, where species are liable to extinction.

the asymptotic tail for small barriers describes the finite number of species that can be selected as having the minimum barrier. These two regimes are separated by the threshold value of $1-\sqrt{(K-1)/K}$, which for large K is approximately $1/2K$. This comes as no surprise, since splitting up each pair of barriers would give the original single-barrier model with $2K-1$ random nearest neighbors, which also has a threshold value of $1/2K$. Breaking up pairs in this way involves replacing just one fixed connection with a random one, a small difference when K is large.

We now turn to consider $\delta g>0$ and $\delta s>0$. To replace the constant N with a dynamical variable $N(t)$ will quickly make analysis impractical, so instead we fix $N=N_\infty$ and assume that speciation balances extinction. Our hope is that the fluctuations of N around N_∞ do not significantly contribute to the solution and can to a good approximation be ignored. Note that extinction has now become dynamically redundant, since making a species extinct and immediately replacing it is no different from a normal mutation. The effects of speciation can similarly be ignored, and we need only account for the δg noise.

With $\delta g>0$, $p(x,y)$ will become blurred at the edges and have a tail extending into the regions $x>1$ and $y>1$. A schematic diagram of the new solution is given in Fig. 3. Since the δg noise is independent of the mutation process, we can calculate its contribution to $\Delta p(x,y)$ separately and simply add it to the expression for $\Delta p(x,y)$ when $\delta g=0$ (4). The new steady-state equation is

$$\frac{K}{N}\theta(x)\theta(1-x)\theta(y)\theta(1-y)-\frac{K-1}{N-1}p-\frac{N-K}{N-1}pQ^{N-1}+\frac{\delta g^2}{24}\nabla^2 p=0, \tag{11}$$

$$\text{where}\quad \theta(x)=\begin{cases}1 & \text{if } x\geqslant 0,\\ 0 & \text{otherwise.}\end{cases} \tag{12}$$

The θ functions in the first term are needed since x and y can now take values outside $[0,1]$. The last term on the left-hand side of Eq. (11) accounts for the noise, where ∇^2 is the Laplacian operator. A full derivation of this term is given in the Appendix. Note that it is only valid for when δg is small. The system behavior for large δg is considered at the end of this section.

Each of the $K-1$ random neighbors selected every time step will be made extinct if they have $x>1$ and $y>1$. Thus the rate of extinction k_E is given by

$$k_E=(K-1)\int_1^\infty\int_1^\infty p(x,y)\,dx\,dy. \tag{13}$$

Strictly speaking, the distribution in this equation should be $p-(1/N)p_{\min}$, but this distinction can be ignored for large N. Since both barriers are large, $Q^{N-1}\sim 0$ and Eq. (11) can now be simplified by the transformation

$$x\rightarrow x'=\alpha(1-x), \tag{14}$$

$$y\rightarrow y'=\alpha(1-y), \tag{15}$$

$$p\rightarrow p'=\frac{K-1}{K}p, \tag{16}$$

$$\alpha^2=\frac{48(K-1)}{\delta g^2 N}, \tag{17}$$

to give

$$2\nabla'^2 p'(x',y')=p'(x',y')-\theta(x')\theta(y'). \tag{18}$$

For either x' or y' negative, corresponding to $x>1$ or $y>1$, the second term on the right-hand side of Eq. (18) vanishes and the equation can be solved by separation of variables. Coupled with the boundary condition $p'(x',y')\rightarrow 0$ for $x'\rightarrow-\infty$ or $y'\rightarrow-\infty$, the solution is

$$p'(x',y')=ce^{(1/2)(x'+y')}, \tag{19}$$

where c is an arbitrary constant. However, it is clear from the numerical results that although the assumption of separable variables is good for this regime, it is not valid when $x'>0$ and $y'>0$. Without matching solutions for these two different regimes, it is impossible to fix the value of c. That c is independent of K and α follows from the absence of these parameters in Eq. (18). Therefore transforming back into the original variables gives the explicit parameter dependence,

$$p(x,y)=c\frac{K}{K-1}e^{-(\alpha/2)(x+y-2)}\qquad\text{for } x>1 \text{ and } y>1. \tag{20}$$

Substituting this into Eq. (13) gives

$$k_E\propto\delta g^2 N\frac{K}{K-1}. \tag{21}$$

When $N=N_\infty$ this will be balanced by the rate of speciation, k_S. The lack of a solution of Eq. (11) for $x<1$ and

TABLE I. Observed values of the natural system size N_∞ for varying parameter values, averaged over 10^7 time steps. The numbers in brackets refer to the uncertainty in the last digit(s). Note that the line for $K=4$, $\delta g=0.02$, and $\delta s=0.001$ appears three times, to allow for easier comparison.

K	δg	δs	Observed N_∞	$N_\infty \delta g^2/[\delta s(K-1)]$
2	0.02	0.001	320(5)	128(2)
4	0.02	0.001	955(5)	127(1)
8	0.02	0.001	2236(10)	128(1)
4	0.02	0.001	955(5)	127(1)
4	0.04	0.001	241(2)	128(1)
4	0.08	0.001	61(1)	130(2)
4	0.02	0.001	955(5)	127(1)
4	0.02	0.002	1770(20)	118(2)
4	0.02	0.004	3250(50)	109(2)

$y<1$ means we have been unable to derive an explicit expression for speciation when $\delta g>0$. Instead, we use the $\delta g=0$ solution as a first approximation. k_S will be proportional to the density of species with $|x-y|<\delta s$. Furthermore, since the species with the minimum barrier can also branch, we must calculate the contributions from p and $p_{\min}$ separately and weight them accordingly, i.e.,

$$k_S=\int\int \theta(\delta s-|x-y|)\{p_{\min}(x,y)+(K-1)p(x,y)\}dx\,dy. \tag{22}$$

Substituting the explicit expressions for p and $p_{\min}$ from Eqs. (9) and (10) gives

$$k_S\approx 2K\delta s \qquad \text{for small } \delta s, \tag{23}$$

which is explicitly independent of N. With $\delta g>0$, $p(x,y)$ broadens and so the real rate of speciation will be slightly lower.

Setting $k_E=k_S$ gives an expression for the parameter dependence of N_∞,

$$N_\infty \propto \frac{\delta s}{\delta g^2}(K-1). \tag{24}$$

If Eq. (24) were exact the quantity $N_\infty \delta g^2/[\delta s(K-1)]$ would be invariant. This should be compared to the numerical results given in Table I. The agreement with the mean-field theory is very good for varying δg and K, but there is a noticeable divergence as δs increases, corresponding to the approximate derivation of k_S [Eq. (23)].

Although large values of δg and δs are physically unrealistic, to complete our study of this model we now describe its behavior for large parameter values. For large δg and small δs, N_∞ is so low that statistical fluctuations make every species in the system extinct. For small δg and large δs, the system size becomes very large and the average time between mutations increases accordingly. Hence the noise becomes the dominant driving force and the barrier distribution becomes very broad. Similarly, when both δg and δs are large the random noise dominates the system, so large δs corresponds with a noise-dominated regime. The derivation of the noise term in Eq. (11) is no longer valid and the rates of speciation and extinction must be found by alternative methods.

In the noise-dominated regime $p(x,y)$ is roughly symmetric about the x and y axes, since the noise is isotropic. Thus at most 1 in 4 species has both barriers greater than 1, and the rate of extinction will approach an upper bound value of

$$k_E\approx \frac{K-1}{4}. \tag{25}$$

The width of $p(x,y)$ can be found by the following argument based on a random walk approach. The average time between mutations as a random nearest neighbor is $\tau=(N-1)/(K-1)$. Each barrier is a one-dimensional random walker that starts from near the origin and moves until it is chosen for mutation, when it is sent back to near the origin. The average step size is $\delta g/4$, and so the average distance it covers in this time is $\sim\frac{1}{4}\delta g\,\tau^{1/2}\approx\frac{1}{4}\delta g\sqrt{N/(K-1)}$ for large N. This gives the width of the barrier distribution in both the x and y directions. The number of species in the infinite strip $|x-y|<\delta s$ is inversely proportional to the distribution width, and so the rate of speciation is now given by

$$k_S\propto \frac{K\delta s}{\delta g}\sqrt{\frac{K-1}{N}}. \tag{26}$$

As N increases, the rate of extinction will remain roughly constant but now the rate of speciation will decrease until a balance is found at $N=N_\infty$. From Eqs. (25) and (26), we find that

$$N_\infty\propto (K-1)\left(\frac{\delta s}{\delta g}\right)^2. \tag{27}$$

This expression replaces Eq. (24) when δs is large. It differs only in that it is now proportional to δs^2 rather than just δs. Figure 4 demonstrates the crossover between the two different forms of N_∞, where to keep the system sizes manageable we have fixed $\delta g=\delta s=\delta$. It is clear from this plot that $N_\infty\propto\delta^{-1}$ for small δ, but $N_\infty\propto\delta^0$ for large δ, as Eqs. (24) and (27) predict. We have also checked the validity of Eq. (27) for a range of K, δg and $\delta s\neq\delta g$.

V. DISCUSSION

In this paper, we have shown how the Bak-Sneppen model can be extended to include the branching and termination of species lines in a model that exhibits a natural system size. Thus there is no longer any need to assume a constant number of ecological niches to obtain sensible results, i.e., systems whose species number does not diverge to infinity or dwindle to zero. Indeed, rather than requiring niches, this model can be used to define what we mean by a niche and their number. Since we have a steady state, every extinction must be associated with exactly one speciation, although they can be in any order and have any time separation. In this sense, we can say that every species does indeed occupy a niche, but the number of niches now comes from within the system itself. This is an interesting corollary,

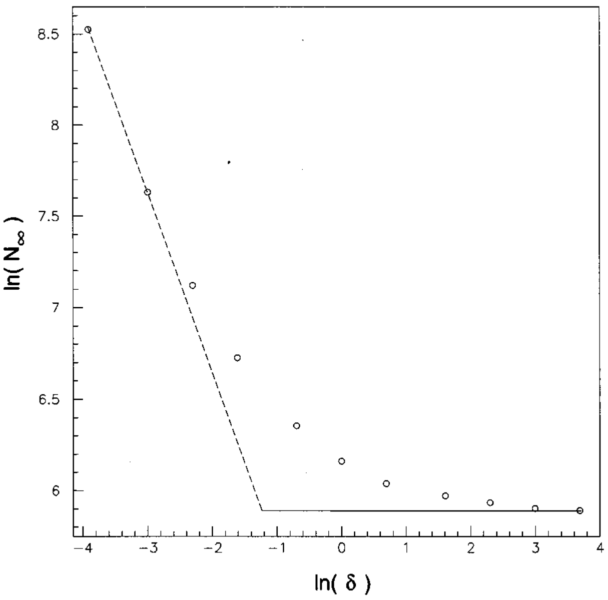

FIG. 4. Plot of the natural system size N_∞ against $\delta = \delta g = \delta s$. $K=2$ and the system size was averaged over 10^6 time steps. The straight lines are to guide the eye, where the solid line is horizontal and the dashed line has a slope of -1.

which should not be all that surprising, since any formal definition of an ecological niche would primarily involve all the species and hence depend on the whole ecosystem.

Despite the extra features, this model is qualitatively very similar to the original when δg and δs are small. This is confirmed by the mean-field analysis, which shows that the barrier distribution has only changed by $O(\delta g^2)$. This analysis has also produced an expression for the natural system size for when δs is small. Although it has no physical meaning, we have also demonstrated the existence of a noise-dominated regime, which occurs when δs is large and has a modified expression for the natural system size. It is interesting that in both cases the system size is proportional to its connectivity $K-1$. This occurs because, when δs is small, the rate of extinction depends on the noise but the rate of speciation does not. Conversely, when δs is large, it is the rate of extinction that is constant and now the rate of speciation decreases due to the noise. The amount of noise that each barrier is subjected to depends upon the average number of time steps since its last mutation, which in this random nearest neighbor model is $\sim N/(K-1)$. Thus, the two rates balance when $N \propto K-1$.

Comparison to the fossil record has demonstrated some agreement with the model, but also some qualitative differences. Although the model is already fairly complex, there are still many possible modifications that may improve its applicability. We have already mentioned introducing species-dependent δs_i and δg_i. Another enhancement would be to have some form of locally defined connectivity, to replace the random neighbor approach used here. Finally, having a barrier-dependent time scale separate from the random noise would allow for comparison with the distribution of extinction sizes from the fossil record. Many other variations could also be considered, but it seems likely that the study of such models would be confined to numerical work.

APPENDIX

Here we give the derivation of the δg-noise term used in the mean-field analysis (11). In the continuum limit, noise effects alone will result in $p(x,y)$ taking the mean value of the surrounding square with sides δg, that is

$$\Delta_{\text{noise}} p(x,y) = -p(x,y) + \frac{1}{\delta g^2} \int_{x-\delta g/2}^{x+\delta g/2} \int_{y-\delta g/2}^{y+\delta g/2} p(u,v)\,du\,dv. \tag{A1}$$

For small δg, $p(x,y)$ can be expanded according to Taylor's theorem as

$$p(x+\delta x, y+\delta y) = p(x,y) + \delta x \frac{\partial p}{\partial x} + \delta y \frac{\partial p}{\partial y} + \frac{\delta x^2}{2!}\frac{\partial^2 p}{\partial x^2} + \delta x\,\delta y \frac{\partial^2 p}{\partial x \partial y} + \frac{\delta y^2}{2!}\frac{\partial^2 p}{\partial y^2} + \cdots. \tag{A2}$$

On substituting this into Eq. (A1), the terms in δx, δy, and $\delta x\,\delta y$ integrate to zero, leaving the leading-order term

$$\Delta_{\text{noise}} p(x,y) = \frac{\delta g^2}{24} \nabla^2 p(x,y) + O(\delta g^3), \tag{A3}$$

where ∇^2 is the two-dimensional Laplacian operator,

$$\nabla^2 \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}. \tag{A4}$$

Equation (A3) is added to the expression for $\Delta p(x,y)$ when $\delta g = 0$ (4) to give the total change in $p(x,y)$ at every time step for $\delta g > 0$. Setting this to zero then gives the new steady state equation (11).

---

[1] P. Bak and K. Sneppen, Phys. Rev. Lett. **71**, 4083 (1993).
[2] B.W. Roberts and M.E.J. Newman, J. Theor. Biol. **180**, 39 (1996).
[3] G.Abramson, Phys. Rev. E **55**, 785 (1997).
[4] R.V. Solè and S.C. Manrubia, Phys. Rev. E **54**, R42 (1996).
[5] D. Aldous, J. Appl. Probab. **32**, 279 (1995).
[6] M.J. Benton, Science **268**, 52 (1995).
[7] M. Kramer, N. Vandewalle, and M. Ausloos, J. Phys. I (France) **6**, 599 (1996).
[8] S.A. Kauffman, *The Origins of Order* (Oxford University Press, Oxford, 1993).
[9] R. Dawkins, *The Extended Phenotype* (Oxford University Press, Oxford, 1982).
[10] M. Ridley, *Evolution* (Blackwell Scientific Publications, Boston, 1993).
[11] D.M. Raup, *Extinction: Bad Genes or Bad Luck?* (Norton, New York, 1991).
[12] J. Maynard-Smith, *The Theory of Evolution* (Cambridge Uni-

versity Press, Cambridge, 1993).

[13] J. de Boer, B. Derrida, H. Flyvberg, A.D. Jackson, and T. Wettig, Phys. Rev. Lett. **73**, 906 (1994).

[14] J. de Boer, A.D. Jackson, and T. Wettig, Phys. Rev. E **51**, 1059 (1995).

[15] D.H. Erwin, Sci. Am. **275** (1), 72 (1996).

[16] H. Flyvbjerg, K. Sneppen, and P. Bak, Phys. Rev. Lett. **71**, 4087 (1993).