

Mathematical modelling of species formation

KIM CHRISTENSEN AND HENRIK JELDTOFT JENSEN

Introduction

Charles Darwin's theory of biological evolution is well over a century old. Biologists have collected an enormous amount of empirical data concerned with almost every detail of the mechanism involved in the evolution of species by natural selection¹. Nevertheless, almost all aspects of evolution theory are still subject to lively debate². Among the issues studied are the specific ways selection has brought about this or that feature of a given species. A famous example is the success of the melanic form of *Biston betularia* over the lighter form, in nineteenth century's England. The standard explanation is the change in colour of the bark of trees from light to dark, brought about by the pollution of the industrial revolution favoured the melanic form. It would be no surprise if specific case studies continue for ever. It is, maybe, more surprising that fundamental problems are still very much open to debate. Questions such as what is a species, why are species necessary? Or what do we really mean by evolution being driven by *survival of the fittest*. What is fitness? The ability to survive? Or a bit more specific; why do some organisms use sexual reproduction while others manage with asexual reproduction, and what determines the length of the number of genes in an organism.

Kim Christensen is a lecturer in Theoretical Condensed Matter Physics in the Blackett Laboratory, Imperial College, Prince Consort Road, London SW7 2BW, UK, URL: <http://www.sst.ph.ic.ac.uk/>, e-mail: k.christensen@ic.ac.uk. His research interest is in statistical mechanics and in particular in complex systems with an emphasis on far from equilibrium phenomena. The main objective has been to investigate the origin of complexity, manifested as fractals and other scale free phenomena, observed in nature. His work has dealt with models of earthquakes and avalanches in piles of granular media such as sand and rice.

Henrik Jeldtoft Jensen is a professor of Mathematical Physics in the Department of Mathematics, Imperial College, 180 Queen's Gate, London SW7 2BZ, UK, URL: http://www.ma.ic.ac.uk/mathematical_physics/, e-mail: h.jensen@ic.ac.uk. Together with Christensen, he has a long standing interest in the statistical mechanics of complex systems. His research has focuses on the conditions necessary for complex scale free behaviour to emerge in models of macroscopic systems. He is the author of the book: Self-organized criticality: Emergent Complex Behaviour in Physical and Biological Systems.

The answers to three types of questions we do not expect to be able to find solely from a refined empirical understanding of specific cases. We need a general theoretical framework. A framework that sees evolution as one specific realisation of general laws of dynamics of complex systems³. The viewpoint that general mathematical laws exist for systems as complex as biological evolution is in itself a controversial one. Some people feel that biology is so complex and complicated that the simplifications needed to make mathematical modelling possible prohibit a mathematical description like those successfully employed in physics.

We do not think this is correct. Just because a phenomenon is complex, and complicated by the simultaneous action of many causes, it does not necessarily exclude that simple laws are at work. An example of this is the development of statistical mechanics. Through the last hundred years, statistical mechanics has demonstrated that it *is* possible to formulate general mathematical laws for phenomena such as phase transitions, superconductivity and turbulence which all involve an enormous number (typical about 10^{23}) of interacting particles.

Statistical mechanics methods are being applied to the study of evolution by a large number of researchers. It is impossible for us here to give a comprehensive review of all these activities, but a very good introduction is given by Peliti⁴. Here we will limit the discussion to a few concrete models which attempt to illuminate different aspects of the mechanisms driving evolution. We have chosen our models by three criteria. The first, that we are aware of their existence. We do acknowledge that there may be other models more appropriate for our purpose. Second, we have selected models that lucidly describe the statistical mechanical approach. Finally we have picked models concerned with the aspects of evolution that are most likely to be governed by some (yet to be discovered) general 'laws'.

In this paper we discuss a truly simple model introduced by Bak and Sneppen⁵. The model addresses the question concerning the intermittent, rather than smooth, extinction rate observed in the fossil record. This behaviour has been discussed by Eldredge⁶ and Gould⁷ and is termed *punctuated equilibrium*. The model focuses on the role played by co-evolution of species. The model assumes the existence of species and ascribe a fitness parameter to each species in an *ad hoc* way. The model is entirely concerned with species and no attempt is made to describe the relationship between the dynamics of species and the individuals, that is, the model is concerned with *macro-evolution*.

We also describe a model introduced by Higgs and Derrida⁹ in which the focus is on the dynamics of reproduction of the individuals in the population. The aim of the model is to show that biological

species can arise in a population when successful reproduction can happen only between individuals with sufficiently similar genetic sequence. This model puts no explicit selective pressure on the individuals. Furthermore it operates with a fixed number of individuals and a fixed length of the genome. In contrast to the Bak–Sneppen model, Higgs' and Derrida's model operates with individuals and studies how species are formed, that is, the model is concerned with *micro-evolution*.

The last model we present is our own construction and constitutes research still in progress. The model includes a number of ingredients from other models. Our intention is to formulate a model of great enough flexibility that within the same model we can address questions such as how do species form, what are the mechanisms determining the length of the genome, how does asexual and sexual reproduction compete. We will also use this model to study aspects of macro-evolution such as the lifetime of species and the temporal behaviour of the extinction rate. Moreover, we use the model to describe the mathematical properties of how the gene sequence of the individuals surviving in the model is related to each other. From this we hope to be able to address the questions: What does evolution do? Does evolution work towards a population of greater diversity, complexity and/or fitness?

Model of punctuated equilibrium

We first describe the observation that Bak and Sneppen address by their model. Does evolution progress in a steady incremental fashion? A steady flow of old species disappearing while new and fitter species appear. Or is the appearance and disappearance of species a jerky process consisting of long periods of quiet happy co-existence between species – punctuated by brief hectic periods of extinction and ensuing proliferation of new species. According to Gould⁷ the latter is the case. From studies of the fossil record Raup⁸ presented paleontological evidence for extinction to be, as he phrased it, episodic at all scales: species survive for long periods and then disappear over a relative short span of years. Moreover, one often observes that the extinction of one species occur simultaneously with the extinction of a number of other species. The number of extinct species during a certain burst of extinction varies greatly.

Let us accept that evolution is a process of long time spans of equilibrium separated by short periods of activity. The question is then what causes the jerkiness. One suggestion has been that the periods of extinction are induced by some external force. For

instance, a meteorite bombardment. Another suggestion ascribes certain sudden mass extinction to volcanic activity.

From a statistical mechanics point of view one is tempted to ask why we in the first place expected the dynamics of a system as complex as the entire ecology of the Earth to be smooth? There may be convincing evidence that some extinction events are caused by meteorites hitting our planet or major volcanic eruptions; however, are we at all entitled to consider intermittent dynamics in a complex system as something that begs an external cause?

In addition to presenting the evidence for this avalanche-like nature of extinction, Raup also suggested that extinction might be a rather arbitrary process. A matter of 'chance-susceptibility' to the surrounding conditions which a species is always subject to at a given moment in history. Raup mentions as an example that mammals and dinosaurs coexisted for more than 100 Myr. Only after the dinosaurs had disappeared did the mammals undergo an explosive evolutionary success aided possibly by the absence of the dinosaurs. The dinosaurs may not necessarily have become obsolete in the Darwinian sense. Their extinction might as well simply have been affected by an unusual fluctuation in the ever present environmental stress. Extinction, argues Raup, might not play the constructive drive behind evolution, as is assumed traditionally. Extinction might be more accidental and arbitrary than hitherto assumed.

Bak and Sneppen tried to extract the essence of these speculations in a very simple probabilistic model. Think of a fixed number of species N . The fitness of a specific species enumerated by x is characterised by a number $B(x)$ between 0 and 1. A value $B(x)$ close to zero signifies that in the given environment the species number x possesses a phenotype that is badly suited to compete with the other species that species number x is interacting with.

We notice here that the model focuses on co-evolution only. The fitness of a species is entirely determined by the accidental peculiarities of co-existing species. The physical environment whether harsh or friendly is not directly considered. This seems at first to be rather silly. But after a moment's thought, one realises that if one can show that a completely self-contained system of co-evolving interacting species can exhibit punctuated equilibrium then one has definitely demonstrated that the punctuation observed in the fossil record is not in itself a proof that the environment of the earth has suffered devastation at intermittent intervals. On the other hand, one reaches then the conclusion that the punctuated equilibrium is the expected dynamical mode of evolution simply because evolution is about the dynamics of a hugely complex interacting network of entities.

Let us return to the definition of the model. We need to decide how the species interact, how their fitness changes and how such a change influences other species. Bak and Sneppen followed a general principle in statistical mechanics; namely to choose the simplest possible representation of a phenomenon. This is always a good starting point if there is no striking evidence that the simplicity will leave out essential features of the phenomenon. We want to adopt the viewpoint of Raup mentioned above. Namely, fitness is a relative quantity and the fitness of one species can increase or decrease when another species disappears. Species with a low fitness are the most likely to become extinct. Hence, the dynamics of the Bak–Sneppen model consist of locating the species x_s with the *smallest* fitness parameter B . This species is removed from the model. In order to maintain the same total number N of species we immediately replace the extinct species number x_s by a new species. Since mutations are random we expect when a new species appears that its fitness relative to the surroundings is random. Accordingly the new species at x_s is given a new fitness $B'(x_s)$ between zero and one chosen at random. The replacement of the species number x_s by a new species of a new phenotype will influence all the species interacting with that species. This is represented in the following simple way. Species number x_s-1 and number x_s+1 are ascribed new randomly chosen fitnesses $B'(x_s-1)$ and $B'(x_s+1)$ respectively. This procedure is then repeated over and over again. Each step consists of locating the smallest B -value and the renewal of this B -value and of the two ‘neighbouring’ species. The model evolves through extremal dynamics in the sense that only the neighbourhood of the smallest B values is allowed to evolve at a given moment.

The model was introduced in an attempt to explain the intermittent or punctuated equilibrium dynamics apparently observed in the fossil record. So how do we identify extinction events in the model? We consider the temporal evolution of the smallest value of all the B -values in the model. Let $B_s(t)$ denote the value of the smallest fitness value at time step t . The signal $B_s(t)$ fluctuates up and down with time. An extinction event is defined as the set of consecutive updates for which $B_s(t)$ is below a chosen value B_0 . So an extinction event starts at time t_0 as $B_s(t)$ passes from above to below B_0 and this extinction event stops the first time $B_s(t)$ goes above B_0 again. We can think of such set of events as wiping out all insufficiently fit species of a fitness value B below the considered B_0 . The size of this extinction event will be given by the number of time steps B_s spent below B_0 .

The dynamics of the model creates a state in which nearly all species has a B value uniformly distributed between the value $B_c = 0.6670$

and 1, see Figure 1. As B_0 approaches B_c from below the distribution of sizes, s , of the extinction events follows a power law: $P(s) \propto s^{-\tau}$ with an exponent $\tau \approx 1.09$. The fossil record also indicates a power law for the size distribution of extinction, though the power observed is about 2 and not close to 1 as in the Bak–Sneppen model. However, the model and the fossil record is in qualitative agreement. There is no *typical* size of an extinction event. The power law form of the distribution of sizes of extinctions signifies this. To appreciate this point one should think of the Gaussian (or normal) bell-shaped distribution. For such a distribution the possible events are those centred close to the peak of the distribution. This is different from a power law distribution. Here events stretching over many orders of magnitude will occur with a significant probability.

Figure 1 also shows the probability distribution of the smallest B -value in the system. One observes that the value of the smallest fitness value B is likely to be close to zero. Hence, most of the time the evolution of the model consists of removing species which are very unfit. However, once in a while the species eliminated might have a rather high fitness (large value of B), the species can be eliminated

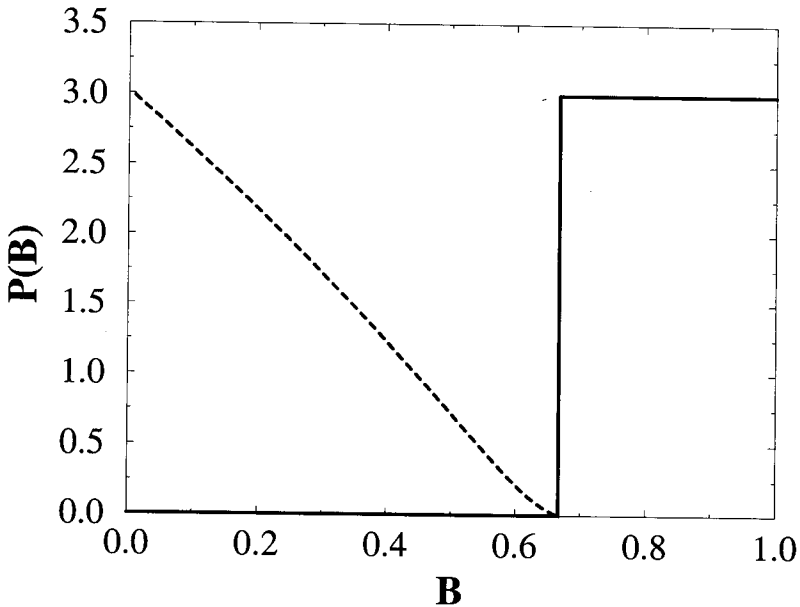


Fig. 1a The solid line shows the probability density of the fitness B in the Bak-Sneppen model after a transient period. The B -values are distributed uniformly between $B_c = 0.6670$ and 1. **b** The dashed line is the probability density of the smallest B -value in the system.

simply because all other species in existence at that moment happens to be even fitter. This aspect of the model stresses the point that fitness is a relative quality.

Species formation from genetic distance criterion

The building blocks of evolution consist of the individual organisms and not of species directly. Species is our attempt to structure the complexity of the biological world. Intuitively we group plants or animals in the same species if they look the same. It is the same degree of similarity that the palaeontologist uses when structuring the fossil record. Obviously it can be difficult to decide exactly how similar two individuals have to be in order to belong to the same species. Are brown and black dogs members of the same species? A more objective criterion consists of what is called the biological species concept. This definition consider two individuals to be of the same species if they can produce fertile offspring. There are, of course, many cases in practice where it is impossible to apply this criterion, not least in the case of fossils or in the case of organisms that reproduce asexually. The definition suggests, however, that what counts is some genetically similarity more than phenotypic look-alike. One is lead to the idea that individuals belong to the same species if their genetic content is sufficiently similar. Higgs and Derrida has constructed a model of a population of sexually reproducing individuals. In this model species are formed when only individuals with sufficiently similar genetic code are allowed to produce offspring.

Think of a population consisting of N individuals. The individual enumerated α is represented by a sequence of L numbers $\vec{S}^\alpha = (S_1^\alpha, S_2^\alpha, \dots, S_L^\alpha)$. One can think of \vec{S}^α as the genome of the individual α and the elements S_i^α as the genes. For simplicity only two forms or allele is allowed for each gene: $S_i^\alpha = -1$ or $S_i^\alpha = 1$. We can define the distance between the genomes of two different individuals α and β as

$$d_{\alpha\beta} = \frac{1}{2L} \sum_{i=1}^L |S_i^\alpha - S_i^\beta| \quad (1)$$

Since S_i^α and S_i^β assumes the two values -1 and 1 only, the term $|S_i^\alpha - S_i^\beta|$ is either 0, when $S_i^\alpha = S_i^\beta$, or 2, when $S_i^\alpha \neq S_i^\beta$. Therefore $Ld_{\alpha\beta}$ is equal to the number of locations in which the two genomes \vec{S}^α and \vec{S}^β differs.

The production of offspring is done as follows way. First two parents are selected. An individual α is selected at random among the N

individuals. Then another individual is selected at random among the individuals for which the distance $d_{\alpha\beta}$ to the first parent α is smaller than some specified distance d_{\max} . A genome \vec{S}^γ is now constructed for the offspring. The gene S_i^γ is chosen to be equal to S_i^α or S_i^β with equal probability. Mutations can occur with a certain probability p_{mut} . A mutation consists of changing the sign of the chosen gene. Say we first selected $S_i^\gamma = S_i^\beta$. The effect of the mutation is to flip the sign so that we end up with $S_i^\gamma = -S_i^\beta$.

The requirement that two individuals can produce an offspring only if their genetic distance satisfies the distance criterion $d_{\alpha\beta} < d_{\max}$ makes the population separate into different isolated groups. That is formation of species takes place. This corresponds to only some specific sequences \vec{S}^α being present in the population. Say the population splits up into two species A and B characterised by a genome \vec{S}_A and \vec{S}_B respectively. All individuals in the model will then have a genetic sequence which is either very similar to \vec{S}_A or to \vec{S}_B . The sequences do not have to be *identical* to \vec{S}_A or \vec{S}_B since a certain number of genes are allowed to differ according to our species criterion. One way to characterise mathematically the formation of species is to study all the possible distances $d_{\alpha\beta}$ between all the individuals. In the case where the population splits into two species only, we will find that $d_{\alpha\beta}$ is close to one of three values d_A , d_B or d_C . The values d_A is the typical distance between two individuals both belonging to species A and similar d_B is the typical distance between two individuals both belonging to species B. The distance d_C is found whenever we calculate the distance between an individual from species A and one from species B.

As time passes and reproduction takes place mutations will occur. These mutations will make the location \vec{S}_A and \vec{S}_B of species A and B drift. This is seen in Figure 2 where we show a sketch of how the distances in the model of Higgs and Derrida evolve with time. It is important to stress that the population in this model only splits up into distinct species when the reproduction is limited to individuals with sufficiently similar genetic sequence. If the distance condition d_{\max} is made too large everybody can mate with everybody and the population consists of individuals with genetic sequences randomly distributed among all possible sequences.

Let us note a few points of special interest. This model demonstrates one mechanism by which species formation can occur without a geographic isolation between populations. This form of speciation is called sympatric speciation and is observed in nature¹⁰. An other point is that the definition of species as individuals with a distance between their genetic sequence smaller than a certain value leads to the follow amusing phenomena. Consider three individuals,

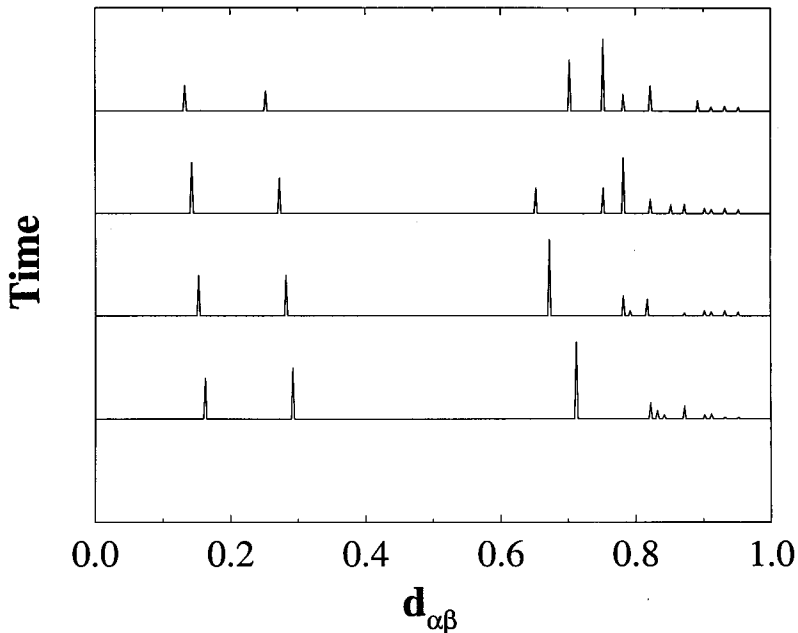


Fig. 2. The species formation illustrated by the genetic distance $d_{\alpha\beta}$ in the model by Higgs and Derrida.

α , β , and γ . Assume that $d_{\alpha\beta} < d_{\max}$ and that $d_{\beta\gamma} < d_{\max}$. There is nothing to prevent $d_{\alpha\gamma} > d_{\max}$ so let us assume this to happen. In this case we have that individuals α and β are in the same species, so is β and γ , but individual α and γ belong to different species. Situations like this do occur in nature. One example is the subspecies of the Rocky Mountain deer mouse¹¹, for which subspecies, say, A and B can interbreed, and so can subspecies B and C, but A and C cannot.

The Garden of Eden model

Our next and final model is mathematically related to the model of the previous section. The model is also in spirit related to the Bak–Sneppen model considered above since we pay special attention to the effect of co-evolution. The model operates with individuals which spontaneously form species under certain conditions. The hope is that the dynamics of species can be ‘derived’ as an emergent property of the model. We do observe indications of punctuated equilibrium in the model. It is not very easy to identify the punctuated equilibrium in the Bak–Sneppen model because the model operates

with a fixed number of species and because the number of species renewed is the same (namely three) in each time step. In our model, species are emergent structures in a population of a fluctuating size. The dynamics of the individual members of the population leads to species being created and going extinct.

The individuals are represented in the same way as in the Higgs–Derrida model by a sequence $\vec{S}^\alpha = (S_1^\alpha, \dots, S_L^\alpha)$ with $S_i^\alpha = -1$ or 1 . As in the Higgs–Derrida model two individuals α and β can only produce an offspring γ if the individuals are sufficiently similar. Since not all genetic information is relevant for mating to be possible, we define the distance between two individuals in a way slightly different from equation (1). Our definition is

$$\tilde{d}_{\alpha\beta} = \frac{1}{2} \frac{1}{L/2} \sum_{i=1}^{L/2} |S_i^\alpha - S_i^\beta| \quad (2)$$

That is, when determining how similar two genetic sequences are we include only the first half of the sequence. The remainder of the sequence is imagined as having no relevance for the ability to mate. An offspring can only be produced if $\tilde{d}_{\alpha\beta} < d_{max}$. When an offspring γ is produced the genetic sequence S_i^γ is as in the Higgs–Derrida model made up by randomly selecting between the two parent genes S_i^α and S_i^β . Mutations can occur with a probability p_{mut} and consists again of changing the sign of the selected gene.

The fitness of an individual is a function of the other individuals present at a given moment and of the physical conditions. This is modelled as follows. Below, we define a number, H^α , for each individual. This number will be used to control the success of the reproduction attempted between an individual α and individual β in the following way. We first choose at random an individual α . Next we choose at random another individual β amongst the individuals with $\tilde{d}_{\alpha\beta} < d_{max}$. The probability p_{offs} that an offspring is produced from α and β is determined by the value of the number H^α and H^β according to

$$\begin{aligned} p^\alpha &= \frac{\exp(H^\alpha)}{1 + \exp(H^\alpha)} \\ p^\beta &= \frac{\exp(H^\beta)}{1 + \exp(H^\beta)} \\ p_{offs} &= \frac{1}{2} (p^\alpha + p^\beta) \end{aligned} \quad (3)$$

We notice that if H^α and H^β assume large positive values p_{offs} will be close to one, and an offspring will be produced. If H^α and H^β assume large negative values p_{offs} is close to zero, and it is unlikely that the encounter between α and β will lead to the production of an offspring.

The number H^α is determined in the following way. Assume that there are $N(t)$ individuals alive at time t . We define a H^α by

$$H^\alpha = \frac{1}{N(t)} \sum_{\beta=1}^{N(t)} \left(\sum_{i=1}^L J_i(\vec{S}^\alpha, \vec{S}^\beta) S_i^\alpha S_i^\beta \right) - \mu N(t) \quad (4)$$

where $J_i(\vec{S}^\alpha, \vec{S}^\beta)$ is a function describing the coupling strength between the genes of the individual α and β . We choose these coupling strengths to be random, but once chosen they do not change again. This coupling is similar to the couplings in the NKC model introduced by Kaufman³. The idea is as in the Bak–Sneppen model that the fitness depends on the features of the other creatures around. One might think that it is more appropriate to include the effect of the surrounding individuals in some sort of killing probability p_{kill}^α related to each individual. This can certainly be done. What we do for simplicity is that all individuals experience the same probability $p_{kill}^\alpha = p_{kill}$ to be removed during a time step. All differences in fitness is included in the p_{offs} which differs from individual to individual.

The physical conditions of the environment is included via the constant M in the term $-\mu N(t)$. This makes the value of H^α depend on the total number of individuals present. If we let μ assume a large positive value, even a moderate number of individuals $N(t)$ will tend to make H^α large and negative and, therefore, hinder the production of offspring. To model an environment able to support a large population we can choose μ to be a small positive number.

One time step consists of the following procedure

- (1) Choose at random an individual α .
- (2) Calculate H^α from equation (4).
- (3) Among the individuals within distance d_{max} from \vec{S}^α choose at random an individual β .
- (4) With probability p_{offs} (calculated from equation (3)) produce an offspring \vec{S}^γ .
- (5) Choose at random an individual and remove it from the population with probability p_{kill} .

The initial population consists for example of $N(0)$ individuals all with the same genetic sequence \vec{s}_0 . The dynamics of the model is generated by repeating item (1) to (5) over and over again.

In the three dimensional plot in Figure 3 we try to describe the evolution of the model. The genetic sequences are enumerated by a number k along the x -axis. Here k is between 0 and 1,024. Time, located along the y -axis, is measured in a number of time steps and assumes values from 0 to 10^7 . The z -axis signifies the relative number of individuals of a certain genetic sequence. For example, the height of a peak at $k = 250$ at time $t = 8 \times 10^6$ is equal to 0.25 if 25% of all

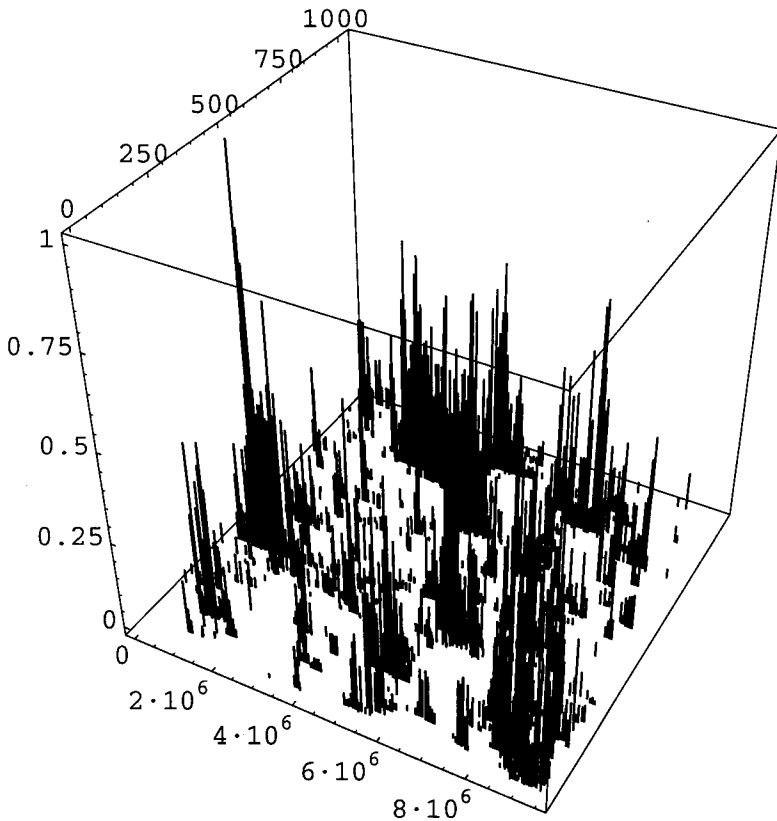


Fig. 3. The evolution of species in the Garden of Eden model. A point on the x -axis represent a specific gene sequence. Here, the total amount of points are $2^{10} = 1,024$, as $L/2 = 10$. The time along the y -axis measures the number of iterations in the model. The shown data covers the range $t \in [0, 10^7]$. Along the z -axis is the relative amount of individuals with that specific gene sequence.

the individuals present at time step 8×10^6 have the genetic sequence enumerated by $k = 250$.

The simulation shown in Figure 3 began from $N(0) = 500$ individuals all of the same genetic sequence. The parameter of the simulation is $p_{mut} = 0.01$, $p_{kill} = 0.1$, $L = 20$ and $\mu = 0.005$. We see in Figure 3 that the initial single species population breaks up into multiple species. Lines of peaks correspond to a certain genetic sequence being present in the population with a certain frequency. When a line of peaks starts the species is born and when the line stops the species has gone extinct. The lifetime of a given species is given by the length of the corresponding line of peaks. The average number of individuals in this simulation was 425. Since the killing probability was $p_{kill} = 0.1$ on average the entire population is renewed every 4,250 time step. This implies that the total simulation corresponds to 23,500 generations. The lifetime distribution of the individuals is found to be exponential. From this we know that the lifetime of species is typically much, much longer than the lifetime of the individuals making up the species.

The model also exhibit punctuated equilibrium. This is seen from the fact that many species lines are typically created or extinguished simultaneously and that for long stretches along the y -axis the species survives in friendly co-existence.

The model can be generalised in a number of ways. We can for instance let individuals with a certain genetic sequence reproduce asexually. This will allow us to study the competition between asexual and sexual reproduction. The first is very efficient in transmitting beneficial genetic codes, the second is good at jumbling up genetic content. We can also study the length of the genome change, say as an effect of mutations. This will enable us to negotiate the mechanisms involved in determining the length of the genome. The physical environment can be included in a more sophisticated way by replacing the term $-\mu N(t)$ in H^α by a more complicated expression. One can imagine having an expression that depends on the genetic content \vec{S}^α of the individual considered. The environment term can also be made time dependent in order to study how a population responds to a temporary degradation of the resources, as has probably happened on a number of occasions when the Earth has been hit by meteorites or exposed to volcanic eruptions. All this has yet to be done.

Discussion

We have presented a few examples of attempts to model mathematically certain fundamental aspects of evolution. Mathematics has long been used successully to study and describe biological evolution

by people like J.B.S. Haldane, R.A. Fisher and J. Maynard Smith. The new approach to the research that we have tried briefly to sketch in this article consists of employing the methodology developed in statistical mechanics combined with computer simulations. This endeavour is still in its infancy. However, in our minds there is no reason why it should not be possible to develop a deeper understanding of the laws behind biological evolution if the approach, and methods indicated above are properly applied in close collaboration with the expertise of evolutionary biologists.

Acknowledgements

It is a great pleasure to express our thanks to Arijeet Datta, Simone Collobiano and Matt Hall. HJJ is supported by the EPSRC.

References

1. Smith, J. Maynard (1958) *The Theory of Evolution*. Cambridge University Press (Reprinted 1995).
2. Ridley, M. (ed.) (1997) *Evolution*, Oxford Readers
3. Kauffman, S. (1995) *At Home in the Universe. The search for the laws of complexity*. Viking, London
4. Peliti, L. *Introduction to the statistical theory of Darwinian evolution*, {<http://xxx.soton.ac.uk/archive/cond-mat/9712>}, paper no. 027.
5. Bak, P. and Sneppen, K. (1993) *Phys. Rev. Let.*, **74**, 4083
6. Eldredge, N. (1971) *Evolution* **25**, 156
7. Gould, S. J. (1977) *Palaeobiology* **3**, 135
8. Raup, M.D. (1986) *Science* **231**, 1528.
9. Higgs, P.G. and Derrida, B. (1992) *J. Molec. Evol.* **35**, 454
10. Kondrashov, A.S. and Kondrashov, F.A. (1999) *Nature*, **400**, 351
11. Cambell, N. A. (1996) *Biology*, Chap. 22, 4th. edn, The Benjamin/Cummings Publishing Company, Location