Applications of Dynamical Systems in Ecology

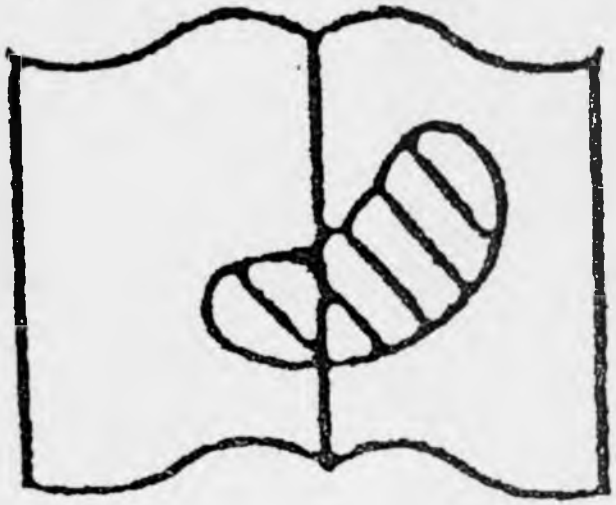
by

Howard B. Wilson

Thesis submitted for the Degree of Doctor of Philosophy

at the University of Warwick, November, 1993.

VARIABLE PRINT QUALITY



Declaration

The material in this thesis is original except for chapters 1, 3 and 5 which are background review chapters. Chapters 2,4 and 7 have been published as three separate papers in the Proceedings of the Royal Society **B** and chapter 6 will be published in Transactions of the Royal Society **B**.

All the original work in this thesis is the result of joint efforts between myself and my supervisor David Rand.

Summary

This thesis consists of five original pieces of work contained in chapters 2, 4, 6, 7 and 8. These cover four topics within the subject area of theoretical ecology: epidemiology, chaos in ecology, evolution and spatially extended ecological systems.

Chapter 2 puts forward a new mechanism for producing chaos in ecology. We show that near extinctions in the SEIR model stabilise a chaotic repeller. This mechanism works for a wide-range of parameter values and so resolves the debate about which dynamic regime is associated with realistic values. It also highlights the problem of treating fluctuations as being either deterministically or stochastically produced.

Chapter 4 describes a new technique for identifying chaos based on measuring the divergence of trajectories over a range of spatial scales. It correctly identifies noise scales and chaos in model systems and is also applied to some real ecological data sets.

In chapters 4 and 5 we set evolutionary game theory in a nonlinear dynamical framework. We introduce a powerful new tool, the selective pressure, for analysing ecological models and identifying evolutionary stable states. It allows analysis of systems where complex attractors exist. We also study the evolution of phenotypic distributions and provide a new mechanism for evolutionary discontinuities.

In chapter 6 we look at an individually-based spatially extended system. This model is spatially heterogeneous and stochastic. However we show that the dynamics on a certain scale are deterministic and low-dimensional. We show how to identify the most efficient spatial scale at which to monitor the system.

Contents

Acknowledgements

Firstly I wish to thank my supervisor David Rand. A remarkable mathematician and personality, I have had great pleasure in his company. He never fails to come up with interesting ideas and without him this thesis would be nothing like what it is. I consider myself truly lucky to have had such a good supervisor who provided such a stimulating work environment. I also wish to thank my friend Margaret who helped and supported me so much in Germany. I will stay in touch. Thanks also to Hubbie and Martin.

I thank all my friends who have helped me in the last three years, particularly Chris, Sue, Bally, Gee and Zeena.

At work Adam and Friedhelm helped me in Germany; at Warwick Mark Muldoon, Greg King and Pete Ashwin helped me out with various problems and Jacqui McGlade, who went to strenous efforts to help fund me, could get a grant out of BCCI.

The work was supported by the Forschungszentrum Jülich, Warwick University and the Science and Engineering Research Council.

1 Dynamical systems in epidemiology.

1.1 Introduction.

A central theme of the study of epidemiology is to explain the pattern and persistence of diseases. Childhood diseases have received a great deal of interest in this respect for a number of reasons. There are striking patterns in the incidence of infections, the transmission mechanism seems to be relatively simple and well-understood, there are relatively long notification series for these diseases and more recently, because they seem to be the best ecological candidate for an example of low-dimensional deterministic chaos. As such there has been a great deal of work trying to understand and explain their dynamics. With the availability of vaccines and the possibility of control this work has become additionally important. The greater our understanding the better able we are to implement control strategies that are effective and optimal.

1.2 Seasonality and periodicity.

A number of features of the incidence rates of childhood diseases are apparent. Firstly there seems to be an annual rise and decline in the incidence of infection. Large number of cases occur in the winter months and in the summer months the number of infections are consistently low. Several hypotheses could explain this pattern. The pathogen's virulence could depend on climate, for example. However detailed analysis points to the school year as being the driving force of these seasonal patterns. During the summer months children are out of doors and generally not coming into close contact with a lot of children. Conversely, during the school term large numbers of children come into close contact in classrooms. Detailed analysis of measles in England and Wales [31] indicates that the transmission of the disease is much higher during school terms and is particularly high at the beginning of the term when there are large numbers of susceptibles. The transmission rate then drops during the school holidays at Christmas

and Easter and particularly during the long summer break. This relationship had been put forward by a number of authors previously [100] and [54], but was clearly demostrated by the analysis of *weekly* incidence data, as performed by Fine and Clarkson [31].

On top of this annual seasonality in infection there are recurrent epidemics, with the inter-epidemic period varying depending on the disease. For some diseases the inter-epidemic period is fairly constant while for others it is more variable. For example the number of measles cases in England and Wales seems to follow a biennial pattern, i.e. a small epidemic occurs one year followed by a large epidemic the year after. New York City post-1945 also seems to have a biennial pattern in measles while pre-1945 it is more variable. Other diseases such as mumps, rubella and whooping cough have longer inter-epidemic periods. Work by Hamer [41] and Soper [100] using simple "mass action" models, was able to explain this phenomenon. The explanation derives from the constant influx of new susceptibles, i.e. births, into the population. After one epidemic had occurred another epidemic could not happen until a sufficient number of new susceptibles had built up for the disease then to spread through. The time between epidemics is a function of the biological parameters, such as the infectious and latency period, of the particular disease [2]. This early work served to illustrate the importance of dynamic models for explaining observed patterns.

1.3 Compartmental models.

One particularly fruitful approach to modelling the dynamics of diseases has been the use of compartmental models [3]. The population is separated into different categories and then equations describing the dynamics between and within categories are formulated. These equations can be stochastic, which assign probabilities to transformations between states, or deterministic. Clearly the real process by which a disease spreads through a population is a discrete stochastic phenomenon. However, when populations are

large, the changes occurring can be deterministically modelled by the expected changes. The deterministic system is thus an approximation to the mean behaviour of the real system. When populations are small this approximation is no longer valid and this point will taken up in more detail in section 2.5. However the deterministic models are far easier to manipulate and analyse and are hoped to be realistic for large cities such as New York City (population $\approx$ 10 million). Various factors such as age-structure, [92], and spatial distribution, [69], at varying levels of complexity can be included within a compartmental model. However, many of the features of childhood diseases can be reproduced with very simple models.

1.3.1 SEIR models.

An appropriate model for many diseases is the $SEIR$ model. The population is separated into four categories; susceptible (S), exposed (E), infective (I) and recovered (R). Individuals are born into the population as susceptibles. Through contact with infectives susceptibles move into the exposed category. After a latency period these individuals become infective. After an infectious period infectives become recovered and they are then assumed immune to further infections for life. For some diseases immunity is not permanent and there will be transitions from the recovered state to the susceptible category. For other diseases it might not be applicable to have an exposed class, in which case an SIR model would be more appropriate. The simplest implementation of this approach is to assume a fixed population size and homogeneous mixing so that four ordinary differential equations can be written down:

$$\begin{aligned} dS/dt &= m(N-S) - bSI \\ dE/dt &= bSI - (m+a)E \\ dI/dt &= aE - (m+g)I \\ dR/dt &= gI - mR. \end{aligned} \tag{1}$$

Since the population is fixed then the population of R is completely determined by the other three through the relationship $N = S + E + I + R$ and therefore doesn't need to be included. The population size N is assumed constant and normalised to 1. The quantity $1/m$ is the average life expectancy of an individual and can be obtained from census data, $1/a$ is the mean latency period and $1/g$ the mean infectious period both of which can be obtained from the medical literature. The most important parameter is the contact rate, b, between susceptibles and infectives since it is these contacts which drive the disease. This parameter has to be estimated indirectly from age-specific serological profiles.

For estimated parameters for common diseases such as measles the solution of these equations is a damped oscillation. This is obviously inconsistent with the recurrent epidemics seen in reality. It has been noted, [6] and [2], that stochastic fluctuations are enough to stabilise these oscillations. However, as noted in section 1.2, there is strong evidence for a seasonality in transmission. It is reasonable, therefore, to assume some periodicity in the contact parameter reflecting the periodic nature of the school year. London and Yorke [54] (also [100]) analysing monthly data, show that the contact rate appears smooth and periodic with period one year. Hence we can model b as a function of time with period one year. It can be simply given by $b(t) = b_0(1 + b_1 \cos 2\pi t)$. The parameter, $b_1 > 0$, reflects the degree of seasonal forcing. Other more complicated forms for b [50] can be extrapolated from weekly data. With this modification the $SEIR$ equations can now sustain oscillations. Aron and Schwartz, [5], show that there is a period-doubling sequence starting from an annual cycle and leading to chaos as the degree of seasonal forcing, b_1, increases. They also show that there is a functional relationship between between the degree of seasonal forcing, b_1, the contact rate, b_0, and the propensity for the model to produce biennial patterns. For low values of the contact rate it takes larger values of seasonal forcing to produce biennial oscillations and *vice versa*. Hence large populations, with large contact rates, will be more susceptible to

biennial outbreaks and also to chaos. However, the regular nature of these oscillations is also inconsistent with the variability seen in real epidemics. Some authors (e.g. [73]) have, therefore, suggested using high values of b_1 ($b_1 > 0.272$) which then produce chaotic solutions to the model. Such high values of b_1 are considered by some to be biologically unreasonable, [20] and [93]. To a large extent this argument is resolved in chapter 2. However, this has also led to more realistic models being formulated.

1.3.2 Age-structured models.

The above $SEIR$ model relies on homogeneous mixing of the population. This model and related ones explain many of the patterns we observe in real epidemics. However, as mentioned above, not all. One explanation of this is that, in reality, the diseases are driven by children coming together at school and infecting one another. This is the justification behind seasonally varying contact rates, as this reflects the seasonal nature of the school year and infection transmission. A more appropriate model might be one where this age-dependence in contact rates is explicitly taken into account. Schenzle [92] introduced such a model. This model keeps the compartmental nature of the $SEIR$ model and the simple biology. The population is then divided up into 21 cohorts ages 0-1, 1-2, .., 20-21, 21+. Each cohort comprises those born during the same school year. New-born susceptibles go into cohort 1. Individuals move into the next cohort simultaneously at the end of each school year. The mortality rate is zero in the first 20 cohorts and constant in the last cohort. The cohorts fall into 4 age classes reflecting: pre-school(1-5), primary school (6-10), adolescents (11-20) and adults (21). The pattern of contact rates is reflected in the discretised contact matrix. These contact parameters are estimated from age-structured case reports or from serological data. The specific pattern of the school calendar acts as a seasonal input. Thus the model carefully includes the most important heterogeneities neglected by the standard $SEIR$ model. This is reflected in the model's quantitative performance, which is better

than the global mass-action models. In particular it yields the biennial measles cycle, the characteristics of which compare favourably to those observed in England prior to vaccination.

1.4 Fade outs.

In order for the deterministic $SEIR$ model to demonstrate irregular chaotic epidemics comparatively high amplitude seasonal forcing is needed. One problem with this is that it generates low levels of infectives in the inter-epidemic periods (the "troughs"). At very low numbers of infectives the $SEIR$ model is no longer a reasonable approximation to reality (but see chapter 2). An approach more in keeping with the actual structure of an epidemic is to reformulate the deterministic $SEIR$ model as a stochastic Monte Carlo model. Here the equations are separated out into transitions between different states with an associated transition probability. For instance the probability that an individual moves from a susceptible state to an exposed state is proportional to $b(t)(S(t)/N)(I(t)/N)$, where N is the population size. This approach deals with individuals and is thus a discrete process. The introduction of a finite population size introduces some important real effects. Using stochastic models and observed data, Bartlett [6] established the important concept of *threshold community size* for the persistence of an infection. He noted that larger cities such as Manchester and Birmingham never show fade-out of infection while small towns have very irregular epidemics with many fade-outs. He identified the relationship of population size with fade-outs and showed that his coupled stochastic models also possessed this property. For measles the urban critical community size is of the order of $250,000$. Olsen et al [74] performed Monte Carlo simulations with population sizes such as Copenhagen ($\approx$ 1 million). However, they still observed fade-out of infection and so introduced a constant immigration probability. This population size is well above Bartlett's' threshold community size for the persistence of infection. One resolution of this problem is that the models don't incorpo-

rate important heterogeneities of a population which might enhance persistence. Since infectious diseases are driven by people coming into contact with each other, spatial heterogeneity might be an important factor. Bartlett used a spatial grid with migration of infectives between neighbouring cells to model this process. He demonstrated the relation between population size and fade-out of infection. Recently, Grenfell [40] has looked at the importance of seasonality and spatial coupling on the probability of fade-out of infection. However he found there was still a significant probability of fade out which is not observed in real epidemics.

1.5 Noise.

Such compartmental modelling, as explained above, has been successful at explaining the gross features of childhood diseases. However, the size of the epidemics varies in a seemingly irregular way. One question, therefore, is whether these annual fluctuations in the size of epidemics is a stochastic phenomenon or whether it is a case of low-dimensional chaos. There have been various mechanisms put forward whereby stochasticity, or noise, could be present in the system and thus explain the observed fluctuations. One source of noise is the reporting rate, or measurement noise. For New York, Baltimore and Copenhagen the reporting rate has been estimated at 12.5-50% for measles and 8-30% for chickenpox [54], [89]. In England and Wales, reporting rates for measles seems to be higher at 66%, [31]. If reporting rates vary this could bring observational errors into the data. For instance in 1941 in New York City there was an exceptionally large number of measles cases reported. This was probably due to the publicity given to the epidemic in that year. Measles data from England and Wales shows drops in the number of cases reported at the end of the calendar year. This has been attributable to postal delays caused at Christmas time. Additionally there was a postal strike in the first quarter of 1971 which led to a drop and then a surge in the number of cases documented. This type of variability may be important but is probably

not sufficient to explain all the observed fluctuations. Another source of noise could be variability in important parameters such as the birth rate or contact rate, so called dynamical noise (e.g. caused by a particularly severe winter keeping people inside more often then usual). Alternatively random movement of infected individuals into or out of the population, inherent stochasticity, could be a source of noise. It is possible that such stochasticity is enough to explain the observed variance and undoubtedly there is such noise present in the data. However the question is to what extent the observed fluctuations are due to this noise and what could be due to deterministic chaos. This is important as understanding structure within the data and the underlying dynamics means we will be better placed to control the disease through more effective vaccination policies. However, the question is a debatable one. Various analyses of the data and models formulated to mimic the dynamics of the disease have been put forward in an attempt to resolve it.

1.6 Traditional methods of analysis.

1.6.1 Power spectra.

An alternative method of representing the incidence rates of diseases (or any time series) is to study the data in the *frequency* rather than the *time* domain.

A process can be represented either in the time domain where some measurement is a function of time, $h(t)$, or the frequency domain where a process is represented as an amplitude as a function of frequency, $H(f)$. Often the amplitude is a complex number representing phase also. One can go back and forth between these representations by means of the *Fourier transform* equations:

$$\begin{aligned} H(f) &= \int_{-\infty}^{\infty} h(t) e^{2\pi i f t} dt \\ h(t) &= \int_{-\infty}^{\infty} H(f) e^{-2\pi i f t} df \end{aligned} \tag{2}$$

Hence we can represent a time series, $h(t)$, as a summation, or integral, of sin waves

of different frequencies and amplitudes. The relative amount a certain frequency is represented in $h(t)$ is reflected by the amplitude at that frequency, $H(f)$. Therefore dominant peaks, frequencies with a relatively large amplitude, in a Fourier transform represent periodic waves which are dominant in the original time series. In the analysis of epidemics we are only interested in the total power at a certain frequency. Since $H(f)$ is generally a complex number, $x + iy$, representing phase also, we plot $x^2 + y^2$ versus f. This is known as the *power spectrum.*

Several authors have studied the power spectra of epidemics including Schaffer and Kot [89], Schaffer et. al. [90] and Olsen et. al. [74]. Schaffer et. al. looked at incidences of chickenpox, measles, mumps and rubella in several North American cities and in Copenhagen, Denmark. All had dominant peaks in their power spectra at one cycle per year. Hence there is a strong annual forcing in the incidence of infection. While this was the only peak in the power spectrum for chickenpox, other diseases had additional peaks. Measles had significant peaks corresponding to cycles of 2-3 years, for mumps there were cycles of 3-4 years and rubella 5-7 years. Power spectra are thus able to pick out the periodic features of the time series . However, they have yielded little further information except that the spectra are fairly flat which is indicative of noise or aperiodic motion (but see [34] and [18] for more information that could possibly be extracted).

1.7 Nonlinear analysis.

1.7.1 Dimension.

Estimating the dimension of a system is attempting to establish that the underlying system generating the observed behaviour is low-dimensional. Stochastic systems are of infinite dimension and so should not converge to a low dimension. A first step in gauging the dimension of the system is a qualitative inspection of the geometry of reconstructed attractors in the phase space. When data is plotted as incidence versus time, then there

may appear a periodicity in the timing of outbreaks but no regular pattern in their magnitudes. Hence one might conclude that the fluctuations are random. However, attractor reconstruction may demonstrate that the fluctuations are deterministic.

Various authors, e.g. [89] and [74], have used phase space portraits to illustrate the low-dimensional structure of the dynamics of measles. Using embedding techniques (see section 3.2.2) it is possible to reconstruct the attractor of a scalar time series. Although Takens' theorem states that an embedding is achieved for $E \geq 2d + 1$, in practice it is often possible to choose an embedding dimension much smaller than this. Using this technique Schaffer and Kot [89] studied the phase space portraits of measles in New York City. The trajectory in the phase space clearly does not wander at random, but instead traces out a low-dimensional object that appears to lie within a 2-dimensional surface.

As outlined in section 3.3.4 the correlation dimension introduced by Grassberger and Proccacia (G-P algorithm) provides a convenient method for quantifying dimension analysis. Work by Schaffer et. al. [90] indicates no convergence to a low dimension as the embedding dimension is increased. This they attribute to the fact that the data is 'spiky', i.e. is nonuniform and hence needs much larger amounts of data to resolve. The modified G-P algorithm they suggest needs the calculation of first return maps as outlined in section 1.7.2. As such it has the same problems associated with it.

1.7.2 Poincaré sections.

In order to reduce the complexity of a system it is useful to study its *Poincare map*. Given a D-dimensional flow one can transversely intersect that flow with a hyperplane. The intersection, the Poincaré section, of the flow with the plane will then have dimension $D-1$. Hence we are able to abstract from our original continuous system a discrete representation of lower order. Successive points $\mathbf{x}_n, \mathbf{x}_{n+1}, \ldots$ on the Poincaré section are

related by the Poincaré map, $\mathbf{P}$, defined by:

$$\mathbf{x}_{n+1} = \mathbf{P}(\mathbf{x}_n).$$

The dynamics of the original system are thus captured by the Poincaré map. A theory of Oseledec [75] guarantees that the properties of this map will be independent of the choice of slicing plane. For a 2-dimensional flow the Poincaré section will be 1-dimensional, either a line or a curve. Thus the mapping, $\mathbf{P}$, might be relatively simple. Schaffer exploited this in studying epidemiological data.

After concluding that the motion in the phase space was 2-dimensional Schaffer and Kot looked at the Poincaré section. The points in this section all lay on a thin band, indicating a 1-dimensional intersection between the attractor and the Poincaré section. They then assumed a functional form for the Poincaré map:

$$y = axe^{bx}. \tag{3}$$

By fitting the points on the section to this map the parameters a and b are estimated. The importance of finding such a simple map is that it reveals an underlying deterministic rule for the dynamics. Once such a map has been estimated it is straightforward to extract various properties. In particular, Schaffer and Kot estimated the Liapunov exponents. The exponents calculated for measles data from New York and Baltimore were both positive. Interestingly, in this same paper, they go on to explore the effects of adding noise into the 1-D maps and point out the importance of repellers; a subject which will be taken up in some detail in chapter 2.

However, as Schaffer notes elsewhere, [90], this analysis is probably in error. A more reasonable assertion is that the measles dynamics correspond more closely to the $SEIR$ model, see section 1.3.1 , with a periodically varying contact rate. Poincaré maps are particularly useful for these forced systems. If we consider the forcing, $Acos\omega t$, as one of the variables, then the Poincaré section consists of points at times $t_n = 2\pi n/\omega$. For the $SEIR$ model this section is 2-dimensional (see figure 2). Hence an attempt to model

the dynamics with a 1-dimensional map must be inaccurate. The Poincaré sections of the original data appear 1-dimensional for the following reason. The data possesses large epidemics of approximately 10,000 infectives followed by troughs of less then 100 infected persons. The Poincaré sections are thus dominated by high-low oscillations. The fluctuations are effectively obscured due to the delay time used, 3 months, in the reconstruction. One coordinate of the the point of intersection on the Poincaré section is roughly proportional to the epidemic size. The other coordinate sits (due to the delay time used) in the trough between epidemics. As a result this coordinate shows relatively little variation and all points on the intersection lie near a 1-dimensional curve, [27]. To minimise this effect one should take logarithms of the data. When this is done the Poincaré sections lose, to some extent, their one-dimensional appearance.

Additionally, as noted by Schaffer and Kot, the results can be sensitive to the functional form assumed for the Poincaré maps. The form assumed here in equation (3) is arbitrary and does not reflect the underlying epidemiological mechanisms. We might expect other forms for **P** to give different results.

1.7.3 Sensitive dependence on initial conditions.

The most important property of a chaotic system is *sensitive dependence on initial conditions* (see section 3.2.1).This is equivalent to the presence of a positive Liapunov exponent. A great deal of work has focussed on directly measuring Liapunov exponents as well as indirectly. Details of these methodologies are given in section 3.3.

Schaffer attempted to calculate Liapunov exponents by two methods [90]. Firstly by directly estimating the divergence of nearby trajectories on the attractor using the Wolf method [115] or a modified Wolf method (which Schaffer calls the "Coyote" method). These methods both gave a positive exponent indicative of chaos. This is looked at in some detail in chapter 4 and as such there is considerable doubt over the figures he gives. The second method [89] was to take a Poincaré section through the attractor

and, if this is one-dimensional, calculate a one-dimensional first return map for the Poincaré section. From this it is straightforward to extract invariant measures such as the Liapunov exponent. However there is concern over this method as outlined in chapter 1.7.2 as the Poincaré sections are not one-dimensional.

Sugihara and May [106] used a simplex prediction method (see section 3.3). They looked at two epidemiological data sets: pre-vaccination measles and chickenpox cases from New York. They concluded that measles showed evidence of chaos while chickenpox was a noisy limit cycle.

In chapter 2 I show that the $SEIR$ model is chaotic for both chickenpox and measles parameter values. Additionally the model is chaotic for a wide range of values of the seasonal forcing not just for the restricted range considered too high by some authors. In chapter 4 I present evidence to show that the data from New York City for measles is chaotic, while chickenpox seems to be more periodic.

2 Chaotic stochasticity: A ubiquitous source of unpredictability in epidemics.

2.1 Introduction.

A topical and central question of epidemiology is whether or not epidemics such as those of chickenpox and measles have chaotic dynamics. The present controversy centres around the apparently contradictory theory and data [78]. While many of the observed time-series possess all the signatures of chaos, for realistic parameter values the associated models often do not [73]. We reconcile these observations and present a new mechanism whereby complex stochastic processes like epidemics can behave chaotically even though their mean-field approximations do not.

In this process the fluctuations from the mean field do not decrease with system size but are greatly amplified by the stabilisation of chaotic repellers of the mean field equations. Such chaotic stochastic processes are likely to be ubiquitous, especially in biological and chemical dynamics, because the ingredients of the underlying mechanism are very common. It occurs for realistic parameter values for both the chickenpox and measles models.

For deterministic systems, the most important characterisation of chaos is sensitive dependence upon initial conditions, i.e. the existence of a positive characteristic exponent χ (see section 3.2.1). Nearby orbits of such a system diverge exponentially fast at a mean rate $\exp(t\chi)$. This puts obvious limits upon the predictability of the system and determines the characteristic time, $t_p = \chi^{-1}$, giving the prediction horizon. A number of real-world epidemiological time-series show such exponentially decaying predictability [73], [91]. However, for deterministic systems the long-term behaviour of the system is usually determined by its attractors and in several cases the most realistic models for these epidemics only have periodic attractors (limit cycles). These have negative characteristic exponents.

An example is chickenpox. If parameter values appropriate to chickenpox are chosen for the commonly used SEIR equations (4) below, then the only attractor is a simple stable annual cycle. On the other hand, real-world chickenpox epidemics appear to have characteristic exponents which range from $\chi = 0.12$ to $\chi = 0.32$ bits per year [73], [91]. (In section 1.7.2 some reservations about these results were discussed). Moreover, although for measles the $SEIR$ equations are chaotic for some acceptable parameter values, several have argued that the contact rate b_1 used is too high [78] and have noted the importance of noise [78], [106], [89], [88] and section 1.5. .The alternative contact rate proposed for measles gives a periodic attractor corresponding to a biennial cycle rather than the chaotic signatures of the observed time-series. We treat both of these examples below.

Epidemics are complex stochastic processes at the microscopic level. The deterministic models are mean field equations which are derived by considering the behaviour of averaged quantities and assuming homogeneous mixing of the population. Their justification is the assumption that the stochastic fluctuations around these averaged quantities tends to zero as the system size grows. This is not valid in systems that possess near extinctions, where a class of the population becomes so small that the statistical effects become important and the mean field assumptions fail. Moreover, such near extinctions are also often associated with instabilities in the deterministic mean field equations. There are thus two contrasting intrinsic sources of unpredictability: (i) the sensitive dependence upon initial conditions which characterises deterministic chaos and (ii) the stochastic fluctuations arising from the probabilistic structure of the underlying epidemiological process. In addition, there are extrinsic sources such as stochastic fluctuations of the environment. We show that these factors can strongly interact with each other so that the stochastic process gives rise to large-scale chaotic behaviour with exponentially growing unpredictability even though the attractor of the mean field equations is periodic and completely predictable.

We divide the forms of stochastic fluctuations into the following three types.

(i) Fluctuations arising from the stochastic structure of the epidemic. These are particularly important when the number of infectives becomes so small that the assumptions of the mean field model are no longer valid.

(ii) Intrinsic noise. This includes randomness in the environment and parameters affecting the epidemic, and fluctuations in the population other than those covered by (i).

(iii) Measurement errors. In principle, these are easier to handle as they are not involved in the dynamics. Indeed, for long data sets there are possible techniques to remove such noise based on the chaotic structure of the dynamics.

Based on numerical work several authors have postulated that the relevant dynamics for chickenpox is a stable cycle in the presence of noise. However, neither multiplicative nor additive noise can produce a positive exponent without the existence of complex dynamics close to the limit cycle. Therefore such an explanation by itself is inadequate. It is also not difficult to see that realistic measurement error cannot contribute toward a positive characteristic exponent, thus we shall concentrate mainly upon the other two forms of noise.

2.2 Epidemiological models.

We firstly consider the following form of the SEIR equations (introduced in section 1.3.1):

$$\begin{aligned} dS/dt &= m(N-S) - bSI \\ dE/dt &= bSI - (m+a)E \\ dI/dt &= aE - (m+g)I. \end{aligned} \tag{4}$$

Individuals enter the population as susceptibles (S). Such a person becomes exposed (E) by contact with individuals (called infectives (I)) capable of transmitting the disease. After a latency period these exposed individuals become infectives and later immune or recovered (R). It is not necessary to include an equation for R because it does not enter the equations for dS/dt, dE/dt or dI/dt except through the relation $R+S+E+I = N$.

The population size N is assumed constant and normalised to 1. The quantity $1/m$ is the average life expectancy of an individual: $1/a$ is the mean latency period and $1/g$ the mean infectious period. The most important parameter, b, represents the effective contact rate, the average fraction of susceptibles contacted by a single infective, which themselves catch the infection. It is a periodic function of time with period one year which we will take to be given by $b(t) = b_0(1 + b_1 \cos 2\pi t)$.

We shall use the values of m, a and g shown in table 1. These values are taken from Olsen and Schaffer [73] and have been obtained from census data (in the case of m) and the medical literature (in the case of a and g). The parameters involved in b are

	Measles	*Chickenpox*
m	0.02 year^{-1}	0.02 year^{-1}
a	35.84 year^{-1}	36.0 year^{-1}
g	100 year^{-1}	34.3 year^{-1}
b_0	1800 year^{-1}	537 year^{-1}
b_1	0.28 year^{-1}	0.3 year^{-1}

Table 1: Parameter values for measles and chickenpox.

calculated indirectly by using age-specific serological profiles to obtain an average age of infection. We note that although the $SEIR$ equations qualitatively capture the main dynamical features of epidemics, it does not give a good quantitative representation of observed measles dynamics [92]. For this it is necessary to incorporate the age structure of both the population and the contact rate, b [92].

As the equation depends periodically upon time (with period one year) it has associated with it a Poincaré map P defined as follows:

$$P(S(0), E(0), I(0)) = (S(1), E(1), I(1)).$$

All of the long-term dynamical phenomena that we discuss can be read off from this map.

To get some general idea of the structure of the attractors for the above system we have plotted bifurcation diagrams for the Poincaré map P. In these we solve equation (4) to obtain $P^n(S(0), E(0), I(0)) = (S(n), E(n), I(n))$, remove transients and then, for each parameter value considered, project the resulting attractor onto the I axis. Figure 1(a) shows the diagram obtained when we interpolate between the chickenpox and measles parameter values of table 1. It shows the stability of the annual limit cycle for chickenpox (which also persists for lower values) as well as the fact that, near the measles values, there are multiple attractors. The latter follows from the observation that the attractor suddenly jumps from being a biennial cycle to a much larger chaotic attractor. We know that the biennial cycle in fact continues further and period doubles and becomes chaotic in a continuous fashion so coexisting with the large chaotic attractor and finally merging with it. An inspection of phase space confirms this. Note also that this diagram shows that chaotic attractors only occur in the immediate neighbourhood of the measles parameters of table 1. By contrast, our mechanism will work at all the parameter values of this diagram as well as values below them.

The interpolation involved here mainly involves the contact parameter b_1. In figure 1(b) we have plotted a similar bifurcation diagram but without removing all the transients and for the chickenpox parameter values with varying b_1. For the noise-free system the projection of the attractor is a single point. But when b_0 is modulated by a small amount of Gaussian noise then we see the dramatic effect shown in figure 1(b). Without noise, the effective chickenpox attractor is a single point (representing a limit cycle) but with only 3% noise it has exploded to a size which is two orders of magnitude

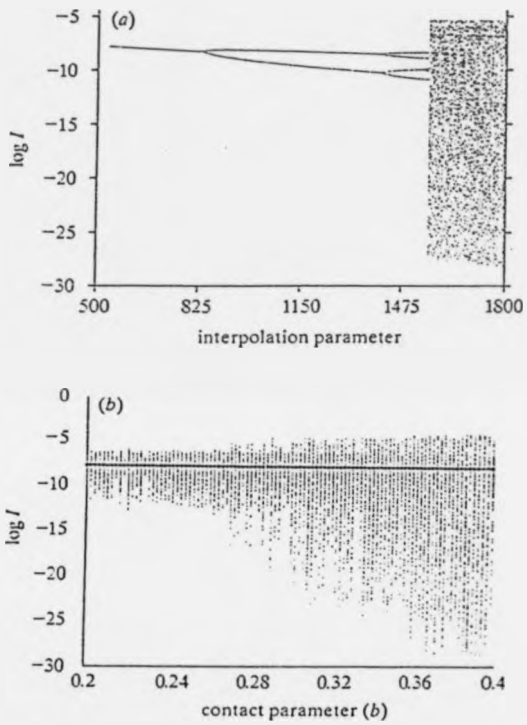


Figure 1: (a) Bifurcation diagram for the SEIR equations (4) where all the parameters are interpolated linearly between the values for chickenpox and measles given in table 1 and the abscissa is labelled with the corresponding value of b_0. Transients are removed by ignoring the first 300 years. (b) Bifurcation diagram for the chickenpox values, but where b_1 is varied. The diamonds plot the position of the attractor without noise. The other points are what is observed when b_0 is modulated by 3% Gaussian noise so that $b = b_0(1 + \xi(t))(1 + b_1 \cos 2\pi t)$.

greater than the noise amplitude. This occurs because of the existence of a chaotic repeller [97].

2.3 A simple chaotic repeller.

Consider a mapping $f : \Re \rightarrow \Re$ on the closed unit interval I defined by:

$$f(x) = ax(1-x), \quad a > 2+\sqrt{5} \tag{5}$$

Since $a > 4$, after one iteration of the mapping f some points will be mapped outside the unit interval and some will remain inside. Those points leaving the interval will tend towards $-\infty$ under the action of the map f. Those points that remain in I after the second iterate of f will be a subset of those remaining after one iteration. The action of the map f continues to subdivide the unit interval. Almost all orbits $f^i(x), 0 < i < \infty$, $x \in I$, eventually escape from I and tend towards $-\infty$. But there is a set of points Λ whose iterates remain in I. This is an invariant set in that if $x \in \Lambda$ then $f^n(x) \in \Lambda$, for all $n > 0$. This invariant set: $\Lambda = \{x : f^i(x) \in I, i > 0\}$ is a cantor set as it is a closed, totally disconnected and perfect subset of I. The requirement $a > 2+\sqrt{5}$ ensures that $\mid f'(x) \mid > 1$ for all $x \in \Lambda$.

Definition 1 *A set Λ is a repelling hyperbolic set for f if Λ is closed, bounded and invariant under f and there exists an $N > 0$ such that $\mid (f^n)'(x) \mid > 1$ for all $n \geq N$.*

2.4 Chaotic repellers and the SEIR equations.

Consider the Poincaré map P of equation (1) with the chickenpox parameter values given in table 1. For these parameters P has a single attracting fixed point $x_* = (S_*, E_*, I_*)$ corresponding to the simple annual cycle. For almost all initial conditions $x_0 = (S_0, E_0, I_0)$, if $x_n = P^n(x_0)$, then x_n converges to the fixed point x_* as $n \rightarrow \infty$. However, this is not true for all initial conditions. There is a set Λ in the phase space such that, if x_n does not converge to the fixed point x_* as $n \rightarrow \infty$ then instead it converges

to Λ. Moreover, this set Λ is chaotic in the sense that it has a fractal structure (it is a Cantor set) and the orbits in it have a positive characteristic exponent [97]. This set is qualitatively similar, but more complex, to the one introduced in section 2.3. On the other hand, the set of initial conditions that converge to it only has zero volume, so, for deterministic systems, it is effectively unobservable and only gives rise to metastable transient chaos [109]. Almost all orbits will eventually escape from Λ and be captured by the periodic attractor. However, orbits starting close to Λ take a long time to converge to the attracting periodic orbit, and while they stay close to Λ, like the orbits in Λ, they behave in a chaotic fashion. Moreover, Λ acts like a chaotic pinball machine. Its stable and unstable manifolds partition the phase space of the system into boxes of all scales through which orbits must pass in a specified sequence. Almost all orbits will eventually find their way through it, but small deviations will lead to radically different routes. If noise or fluctuations are constantly moving the state from one box to another then the orbit may never find its way out. Chaotic repellers are created after a period doubling sequence, but in multi-dimensional systems they can also occur before the main period doubling sequence. Experience shows and theory indicates that they are far more ubiquitous than chaotic attractors [109], [97].

In figure 2 we plot this chaotic repeller in the following way. Firstly, we note that for large times, E is a function of S and I, and is very close to I. Thus we need only plot S and I as these determine the system. We start a large number of initial conditions of the form $(S, E = I, I)$ along the top edge of the rectangle Γ of figure 2. We only consider that part of the orbit which stays inside Γ and discard the initial and final 25% of this, plotting the remainder. This captures the non-trivial structure of the repeller.

We now consider how to detect the existence of a positive characteristic exponent for the repeller, as this method provides a particularly useful signature which we shall apply to our model epidemiological data. Let $x(t) = (S(t), E(t), I(t))$ be a solution of (1). For $i = 0, 1, 2, \ldots$ let $y_i(t)$ be a solution of the equation for $i\tau \leq t \leq (i+1)\tau$ such

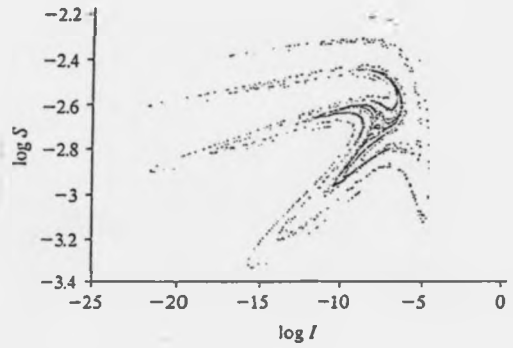


Figure 2: An approximation of the chaotic repeller of the SEIR equation (4) with the chickenpox parameter values of table 1. The square marks the position of the fixed point corresponding to the annual cycle.

that $y_i(t)$ is very close to (say within δ) of $x(t)$ for this period of time, and such that $x(i\tau) - y_i(i\tau)$ is parallel to $x(i\tau) - y_{i-1}(i\tau)$. Let

$$d_i = ||x(i\tau) - y_i(i\tau)||/||x((i-1)\tau) - y_i((i-1)\tau)||$$

and $D_n = d_1 \cdots d_n$. For almost all choices of the y_i the largest characteristic exponent χ_0 of the attractor is estimated by the long time limit of $n^{-1} \log D_n$. We emphasise that we are not so much interested here in the precise value of χ_0, but in detecting when it is positive.

We plot $\log D_n$ against n in figure 3(b) and note that it grows linearly for a while at rate χ_0 approximately equal to 0.38 bits per year, but then the slope, χ_0, suddenly plunges to a negative value of approximately -0.17. We interpret this as the signature of the existence of a chaotic repeller. The increasing D_n correspond to periods when the solution shadows the chaotic repeller (the chaotic transient) and the plunge corresponds to the solution finally entering the immediate basin of the periodic attractor. This behaviour is to be contrasted with figure 3(a) which is instead for a solution of the

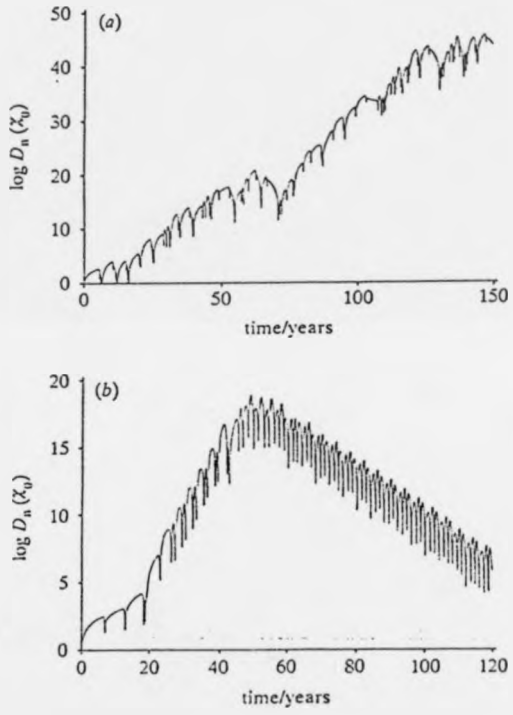


Figure 3: (a) The graph of D_n against n for the SEIR equation (4) with the measles parameter values of table 1. (b) As (a) but for the chickenpox parameters of table 1. Note the non-uniform growth in $\log D_n$ which is due to the fact that chaotic attractors and repellers contain a mixture of regions of high and low instability.

measles parameters of table 1. In this case we observe perpetual linear growth with a slope, χ_0, approximately equal to 0.32 bits per year. This signals the presence of a chaotic attractor.

Several authors have argued that this value of b_1 is too high for measles [78]. But, if b_1 is less than 0.272 then the only attractor is a limit cycle and the situation is similar to that of chickenpox above. In particular, for b_1 not too small there exists a chaotic repeller for measles.

2.5 Fluctuations and noise.

Let us consider the contribution μ to dE/dt which models the interaction, contact and infectious transfer between infectives and susceptibles. In equation (4) we have used $\mu = bSI$. Such a term is reasonable if the size of both I and S does not become too small. However, solutions of (4) for the parameters in table 1 at times have extremely small values of I. Although the value of I for the annual cycle for chickenpox is approximately 10^{-4}, orbits in the repeller contain common events where I becomes as small as 10^{-6} and rare events where it becomes as small as 10^{-10}. Transient events are observed with values as low as 10^{-14}. Values less than 10^{-6} correspond to only a few infected individuals in the population or the extinction of the infection. The existence of these near extinctions is thought to be realistic and the presence of a repeller clearly has consequences for the theory of extinctions. But with so few individuals the approximation of μ is not reasonable.

The real process by which susceptible individuals are brought into contact with and infected by infectious ones is clearly very complex. The simplest model for such a situation is to assume (i) that each susceptible and infected individual interacts with other members of the population and this other person is chosen randomly and independently of the other such choices (ii) that such interactions occur at a constant rate τ and (iii) that the infectives and susceptibles are distributed uniformly and randomly throughout

the population.

Consider a time interval $(t, t+\delta t)$. The probability of an infective making an infectious contact with another individual during this period is $p = \tau\delta t$. Let $s = NS(t)$ and $i = NI(t)$ denote the number of susceptibles and infectives at time t. Then the probability that a susceptible makes at least one infectious contact is $1-(1-\tau\delta t)^i \approx i\tau\delta t$, where we ignore terms that are $O(\delta t^2)$. Therefore, the probability that j individuals are infected during this period is

$$p(j) = {}^sC_j(1-\tau i\delta t)^{s-j}(\tau i\delta t)^j.$$

If $i\delta t \to 0$ this is approximated by the Poisson distribution

$$p(j) = \frac{(\tau si\delta t)^j}{j!}e^{-\tau si\delta t}. \tag{6}$$

Thus we replace the first equation (4) by

$$\begin{aligned} dS/dt &= m(N-S) - d\mu \\ dE/dt &= d\mu - (m+a)E \\ dI/dt &= aE - (m+g)I + I_0. \end{aligned} \tag{7}$$

where the term $d\mu(t)$ is such that its integral on an interval $(t, t+\delta t)$ has the distribution $p(j)$ given in (6). We do not attempt to stochastically model the other terms in the equation which correspond to immigration, birth, death and the transition from being exposed to being infective since these are far better controlled and accurately represented by the mean field approximation.

We have numerically simulated the stochastic solutions of this equation with N in the range $10^7 - 10^{10}$. With these values of N the minimum value of $i = NI$ in the repeller is of the order of $10-1000$ for the common events. The rare events, where it is even smaller, give extinctions if, as in (4), the term I_0, which represents a very small constant import rate of infectives into the population is zero. We prevent these extinctions by taking I_0 positive but very small, usually of the order of 10^{-6} to correspond to the import of a few

individuals per year. We now observe that the solutions for the chickenpox parameter values of table 1 corresponding to those computed above now have values of D_n which grow linearly indefinitely with time (see figure 4(a)) and have a positive characteristic exponent χ_0 approximately equal to 0.77. In figure 4(b) we have plotted a typical trajectory to verify that its density is a smoothed out version of that for the repeller. Thus we see that the stochastic fluctuations corresponding to low I can stabilise the chaotic repeller.

We obtain similar results if we assume stochastic fluctuations of the contact rate. For example, in figure 5 we show the growth in the D_n when we take $b = b_0(1 + \xi(t))(1 + b_1 \cos 2\pi t)$ where ξ is white noise with mean 0, amplitude $a = 0.03$ and variance $\sigma = a^2$.

Finally, we note that when there is a coexisting periodic attractor and chaotic repeller, the existence of near extinctions can lead to large fluctuations in the size of the basin of the attractor with time of the year and hence to the critical size of perturbations. They are moreover associated with instabilities which amplify fluctuations by an order of magnitude or more between periods of low and high I. This follows because (i) I varies over several orders of magnitude and (ii) the equations corresponding to (4) for the quantities $\log S$, $\log E$ and $\log I$ are well-balanced and for these the basin is roughly constant with phase and the expansion is roughly of order unity. In the case of the chickenpox parameters, this amplification is by a factor of more than 20. This phenomenon greatly aids the mechanism we have described since the stochastic fluctuations at small I are greatly magnified.

2.6 Conclusion.

The deterministic mean field models for chickenpox and measles epidemics have a very rich dynamical structure with complex bifurcations, multiple co-existing attractors, fractal basin boundaries and chaotic repellers. In particular, they possess a chaotic repeller for a very large range of parameter values. They also contain near extinctions where the

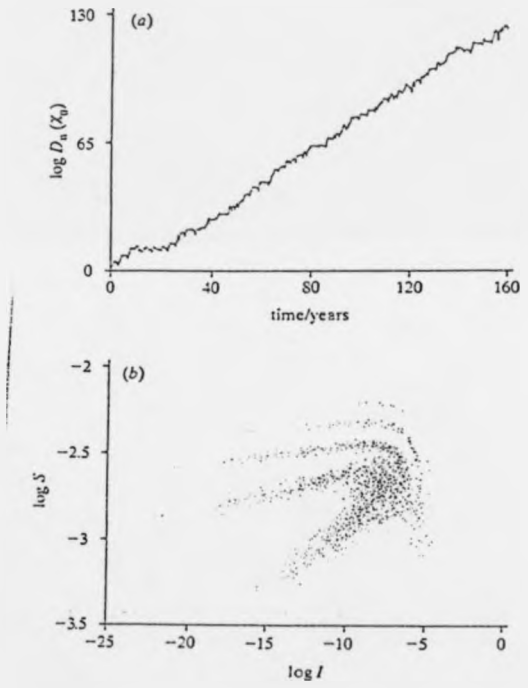


Figure 4: (a) The graph of D_n against n for the SEIR equation (3) with the chickenpox parameter values of table 1, $I_0 = 10^{-6}$ and with the contacts determined by the Poisson distribution. (b) The distribution of a typical orbit. Note that it is a smoothed out version of the repeller shown in figure 2

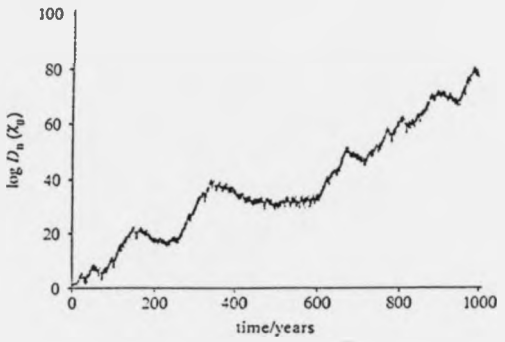


Figure 5: The graph of D_n against n for the SEIR equation (4) with the chickenpox parameter values of table 1 and $b = b_0(1 + \xi(t))(1 + b_1 \cos 2\pi t)$ where ξ is white noise with mean 0, amplitude $a = 0.03$ and variance $\sigma = a^2$.

number of infectives becomes very small and the deviation from the mean behaviour is significant. We have shown that the corresponding stochastic fluctuations stabilise the chaotic repellers so that the resulting time-series are practically indistinguishable from those of a chaotic attractor. This is a potential explanation for the present discrepancy between theory and experiment. Because of its simplicity and the ubiquity and stability of its ingredients, this mechanism is likely to be very common in biological, chemical and some physical dynamical systems. For example, since they appear to fit into our scheme of things (including the existence of periods of very low numbers), it is tempting to use these ideas to model the dynamics of Nicholson's blowflies [70] and other biological populations to explain the conjectured chaos as discussed by May [61].

3 Detecting chaos.

3.1 Introduction.

Until recently most biological and ecological systems have been presumed to posses simple dynamics: either an equilibrium or limit cycle with stochastic perturbations superimposed. The discovery of low-dimensional deterministic chaos in the early sixties [55] and its development and application in the physical sciences in the early seventies [86] has stimulated great advances in the nonlinear analysis of systems. May [60] brought this body of work to the attention of ecologists by demonstrating that even the simplest nonlinear model for population dynamics had a rich dynamical structure, including chaos. Since then, the potential for biological and ecological systems to display chaos has been demonstrated in numerous different areas including epidemiology, physiological systems and population dynamics. The importance of this new phenomenon is that it may enhance our understanding, control and prediction of previously badly understood, or presumed random, systems.

Initial techniques for identifying chaotic systems were developed in the physical and mathematical sciences. Deterministic chaos is usually characterised by the existence of an attractor with the following three properties: (i) sensitive dependence upon initial conditions, (ii) a fractal geometric structure and (iii) positive Sinai-Kolmogorov entropy. Of these (i) is the most fundamental for the following reasons: (a) it is the cause of the exponentially decaying predictability characteristic of chaos; (b) it implies (iii); (c) not all chaotic attractors are fractal and the Hausdorff dimension is severely constrained by the sizes of the Liapunov characteristic exponents.

Sensitive dependence is equivalent to the fact that, in a chaotic system, typical nearby orbits diverge exponentially fast from each other, at least on average. The Liapunov characteristic exponents (LCEs) are the different long-time exponential rates at which nearby orbits diverge or converge. Thus we have sensitive dependence provided the

largest LCE, χ, is positive. Hence, methodologies for detecting chaos have focussed on measuring the Liapunov exponents (although measuring the dimension has also been a very active area).

3.2 Dynamical systems.

Consider a dynamical system of the form

$$\frac{d\mathbf{y}}{dt} = \mathbf{F}(\mathbf{y}), \tag{8}$$

where y is a D-dimensional vector $(y_1, y_2, ..., y_D)$. Under certain conditions on F this has a unique solution at a time t given an initial value $\mathbf{y}_0$. Thus we write

$$\mathbf{y}_t = f^t(\mathbf{y}_0).$$

For *dissipative* systems the flow will contract onto a set of dimension d where $d < D$. These sets are called attractors.

Definition 2 *An attractor is an invariant set Λ with neighbourhood $\mathcal{N}$ such that if $\mathbf{y} \in \mathcal{N}$ then the distance $(f^t(\mathbf{y}), \Lambda) \to 0$ as $t \to \infty$.*

The study of dynamical systems is the study of the geometry of motion in the *phase space*. The phase space of a dynamical system is a coordinate system which uniquely specifies the state of the system at a particular time, t. Usually the state variables are used as the coordinates. A qualitative description of the dynamic behaviour of F would begin with the geometric phase-space identification of all possible attractors. These could include fixed point, periodic, quasi-periodic or chaotic attractors. For certain parameter values it is also possible for multiple attractors to exist in different parts of the phase space. Once the attractors are known one can associate all the starting conditions that settle to it. This point set is the *basin of attraction*. All the basins of attraction should constitute the entire phase space.

3.2.1 Liapunov exponents.

In order to illustrate the properties of chaos and Liapunov exponents consider a 1-dimensional mapping $f : \Re \rightarrow \Re$. Assume there exists an attractor, Λ. Let $x_t = f^t(x_0)$ and $y_t = f^t(y_0)$ where x_0 and y_0 are both points on Λ and are a small distance, ε, apart. Hence:

$$| y_0 - x_0 | = \varepsilon.$$

Then:

$$\begin{aligned} y_t - x_t &= f^t(y_0) - f^t(x_0), \\ &= (f^t)'(x_0) \cdot (y_0 - x_0) + O(\varepsilon^2), \end{aligned}$$

since $f(y_0) = f(x_0) + \varepsilon f'(x_0) + O(\varepsilon^2)$, by Taylor's Theorem. By the chain rule:

$$| y_t - x_t | = | (y_0 - x_0) | \cdot \prod_{i=0}^{t-1} | f'(x_i) | + O(\varepsilon^2). \tag{9}$$

If the uncertainty grows exponentially fast then (9) must be consistent with:

$$| y_t - x_t | = | (y_0 - x_0) | \cdot 2^{\lambda t},$$

where λ is the exponential rate of growth. The Liapunov exponent, λ, is thus defined as:

$$\lambda = \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{i=0}^{t-1} log_2 | f'(x_i) | . \tag{10}$$

The limit as $t \rightarrow \infty$ is needed for two reasons: (i) the rate of divergence or convergence will vary at different parts of the attractor and (ii) we are interested in the long-term behaviour which is independent of the initial condition, x_0, chosen. The quantity λ thus becomes a measure on the attractor, Λ.

For systems of higher dimension the Liapunov exponent becomes a generalised eigenvalue averaged over the attractor as the derivative is replaced by the Jacobian at that point. There will be an exponent for each phase-space dimension describing orbital contraction or convergence in a particular direction. This direction will change over

the attractor so it becomes impossible to associate a specific direction to a Liapunov exponent over long periods of time. The set of all the Liapunov exponents is called the Liapunov spectrum. Chaos is then defined as the presence of at least one positive exponent in the Liapunov spectrum. For periodic attractors, or point attractors, all the exponents will be negative or zero.

In general, if **F** is known then it is relatively straight forward to calculate the Liapunov spectrum directly from **F**, [95] and [7] . This can be done by calculating the long-time product of local Jacobians along an orbit over all the attractor Λ. This matrix is thus the long-time average convergence or divergence. Normally, however, we do not know **F** *a priori*. In fact often we have only an observed scalar time series.

3.2.2 Embedding.

Although we may be studying a D-dimensional system typically, in data from real systems, we have only an observed 1-dimensional time series $\{x_i\}_{i=1}^{N'}$ where i is an integer, α is the sampling time between successive measurements and N' the number of points in the time series. However we can understand and analyse the system by making use of embedding theorems [114] as first applied to time series by Packard et. al. [76] and put on a firm mathematical basis by Takens [107]. The basic idea behind this methodology is that the past, and future, of a time series contain information about the present, unobserved state variables. This information can be utilised in the form of a *delay vector*. In fact there are many choices of possible coordinate system, e.g. successive time derivatives, but the easiest and most widely used is delay coordinates. Takens studied the delay reconstruction map Φ:

$$\Phi(\mathbf{y}) = (\upsilon(\mathbf{y}), \upsilon(f^1(\mathbf{y})), ..., \upsilon(f^{2m}(\mathbf{y}))),$$

where $\upsilon(\mathbf{y})$ corresponds to a value of a measurement made on the system in state $\mathbf{y}$. He proved that Φ is an embedding when $m \geq 2d+1$. An embedding is a smooth, one-to-one coordinate transformation with a smooth inverse (i.e. Φ is a *diffeomorphism*). Hence

the reconstructed attractor is topologically equivalent to the attractor for the whole system and thus preserves important geometrical invariants such as dimension of the attractor and Liapunov exponents of a trajectory.

For the observed scalar time series, $x_i = v(\mathbf{y}_t)$, the function v corresponds to measuring one state variable of the system at a particular time, $t = \alpha i$. Hence the delay coordinates are:

$$\mathbf{x}_i = (x_i, x_{i-\tau}, ..., x_{i-(E-1)\tau}), \tag{11}$$

where E is known as the *embedding dimension* and τ the *delay time*, which is conventionally taken to be an integer here. Thus the time between elements of the embedded vector is $\tau\alpha$. Unfortunately, Takens' theorem gives us no information as to how to choose the delay time τ. Additionally E is often hard to choose as the dimension, d, of the attractor of F, is not known *a priori*. The number of embedded vectors, N, is $N = N' - E + 1$. The notation outlined above will be used throughout.

3.3 Basic methodologies.

Using the method of delay coordinates, various techniques and algorithms for distinguishing chaotic and stochastic fluctuations have been proposed. These techniques can be separated into four categories: directly measuring the Liapunov exponents either through a) the divergence of nearby trajectories or b) estimating tangent maps; c) investigating the decay in predictability and d) measuring the number of degrees of freedom, or state space variables, needed to describe the dynamics. The first three categories are based on the property of a positive Liapunov exponent. The fourth category, the dimension of the system, is attempting to establish that the underlying system generating the observed behaviour is low-dimensional. Stochastic systems are of infinite dimension so should not converge to a low dimension.

Methods (a) and (b) can, in principle, measure the complete Liapunov spectrum; negative as well as positive exponents. In practice, with finite, noisy data sets, the

calculation of negative exponents is very difficult. However, it is sufficient to measure the largest exponent to identify whether a system is chaotic or not.

3.3.1 Divergence of trajectories.

Although the reconstructed attractor is defined by a single trajectory from a discrete series of measurements it can provide points that may, because of the topological mixing, be considered to lie on different orbits provided our trajectory is long enough. Thus, Wolf et al [115] proposed directly measuring the divergence of nearby orbits. Given a time series $\{x_i\}_{i=1}^{N'}$ we form delay vectors, $\mathbf{x}_i$, as stated above. The nearest neighbour (in the Euclidean sense) is located to the initial point $\mathbf{x}_1$ and the initial distance between these points is denoted $L(0)$. For a chaotic system this distance will grow exponentially fast. However we wish only to measure the local stretching on the attractor and so this distance should only be allowed to grow for a small amount of time, T. If the time is too long the trajectories might pass through a folding region on the attractor. After a fixed time, $T = m\alpha$, our initial point will have evolved a number of steps, $m = T/\alpha$, to $\mathbf{x}_{1+m}$ and the distance apart is now $L'(1)$. We then look for a new data point that is a near neighbour to $\mathbf{x}_{1+m}$ but approximately preserves the original orientation of $\mathbf{x}_{1+m}$ and our original evolved neighbour. The new distance apart is $L(1)$ and the points are again evolved a time T. This is then repeated until the whole time series has been traversed. The largest Liapunov exponent is thus estimated as:

$$\lambda_1 = \frac{1}{MT}\sum_{k=1}^{M} log_2 \frac{L'(k)}{L(k-1)},$$

where M is the total number of replacement steps.

In the limit of an infinite amount of noise-free data the largest exponent is obtained by definition. Depending on the nature of the attractor, estimates of within a few percent of the true exponent value can be obtained for noise-free time series of a few thousand points or less. For noisy time series it is important to ignore length scales of the order of the noise but still operate at small scales at which only the stretching nature of the

attractor is probed. These problems are explored in some detail in chapter 4.

3.3.2 Tangent space.

Eckmann and Ruelle [23] and Sano and Sawada [87] suggested obtaining tangent maps to the reconstructed dynamics. The evolution of a tangent vector, ξ, in a tangent space at $\mathbf{y}$ is represented by linearising Eq. 8,

$$\dot{\xi} = \mathbf{T}(\mathbf{y}) \cdot \xi, \tag{12}$$

where $\mathbf{T} = \mathbf{DF} = \frac{\partial \mathbf{F}}{\partial \mathbf{y}}$ is the Jacobian matrix of $\mathbf{F}$. The solution of the linear equation (12) can be obtained as

$$\xi(t) = \mathbf{A}^t \xi(0), \tag{13}$$

where $\mathbf{A}^t$ is the linear operator which maps tangent vector $\xi(0)$ to $\xi(t)$. We can then calculate $\mathbf{A}$ at a time j, $\mathbf{A}_j$, over all the attractor Λ of $\mathbf{F}$. The Liapunov exponents are then given by the eigenvalues of the product of the matrices $\mathbf{A}_0 \cdot \ldots \cdot \mathbf{A}_n$.

From the observed data $\{x_i\}_{i=1}^{N}$ it is possible to estimate the linear map $\mathbf{A}_j$. Consider a small ball of radius ε centered at the embedded vector $\mathbf{x}_j$. Find any points $\mathbf{x}_{k_i}$ within this ball, i.e $\| \mathbf{x}_{k_i} - \mathbf{x}_j \| \leq \varepsilon$, where $\|$ denotes the Euclidean norm. The displacement vector $\mathbf{w}^i = \mathbf{x}_{k_i} - \mathbf{x}_j$ is thus a good approximation to the tangent vector, ξ, at the point j if the radius, ε, is small enough. After the evolution of a time interval $u = m\alpha$, the point $\mathbf{x}_j$ will evolve to $\mathbf{x}_{j+m}$ and the neighbouring points $\mathbf{x}_{k_i}$ to $\mathbf{x}_{k_i+m}$. The displacement vector $\mathbf{w}^i$ is thus mapped to $\mathbf{z}^i = \mathbf{x}_{k_i+m} - \mathbf{x}_{j+m}$. The matrix $\mathbf{A}_j$ given by

$$\mathbf{z}^i = \mathbf{A}_j \mathbf{w}^i$$

is an approximation of the flow map $\mathbf{A}^t$ at $\mathbf{x}_j$ given in Eq. (13). The map can now be estimated by a least-squares fit procedure which minimises the average of the squared error norm, S, between z^i and $\mathbf{A}_j \mathbf{w}^i$ with respect to the elements of the matrix $\mathbf{A}_j$, i.e

$$S = \min_{\mathbf{A}_j} \frac{1}{n} \sum_{i=1}^{n} \| \mathbf{z}^i - \mathbf{A}_j \mathbf{w}^i \|^2,$$

where n is the number of points in the ball. This procedure leads to $n * E$ simultaneous linear equations in $E * E$ unknowns. Hence it is well-defined if the number of points in the ball, n, is larger than the dimension, E, of the embedded vectors. In practice it is usual to take n somewhat greater than E for statistical reasons (but see [23]).

3.3.3 Predictability.

Farmer and Sidorowich [30] formulated a prediction technique also based on local linear maps. In a chaotic system nearby trajectories diverge apart exponentially fast at a rate given by the Liapunov exponent. This means that any initial error in estimating the state of the system will grow exponentially fast through time. Hence the accuracy of prediction of a system from an initial point into the future will decay exponentially as you try to predict the evolution of the point further ahead in time. For a periodic or steady state system this predictability should not decay. Hence an exponential decrease in predictability should be a signature of chaos. Farmer and Sidorowich exploited this property by measuring the predictability of a time series by the following method based on local linear maps.

A state, x_i, will evolve to some other state, x_{i+m}, a time $T = m\alpha$ in the future. Assume a map, $\mathbf{A}$, exists such that $x_{i+m} = \mathbf{A}(\mathbf{x}_i)$, where $\mathbf{x}_i$ is the delay vector. To approximate the map $\mathbf{A}$ we look for all nearest neighbours, $\mathbf{x}_k$, to $\mathbf{x}_i$, i.e. $\| \mathbf{x}_i - \mathbf{x}_k \| < \varepsilon$ and then fit the coefficients of $\mathbf{A}$ using a least-squares fitting procedure as described above. The point x_{i+m} is then predicted from $\mathbf{A}(\mathbf{x}_i)$. The accuracy of this prediction will decay exponentially with increasing T.

Further work in this area has been done by several authors; notably Casdagli [14] who suggested using radial-basis functions instead of linear maps and other authors who have explored the use of higher-order local maps. Neural networks have also been used to learn optimal local maps and seem to perform as well or better than Farmer and Sidorowich's method.

where n is the number of points in the ball. This procedure leads to $n * E$ simultaneous linear equations in $E * E$ unknowns. Hence it is well-defined if the number of points in the ball, n, is larger than the dimension, E, of the embedded vectors. In practice it is usual to take n somewhat greater than E for statistical reasons (but see [23]).

3.3.3 Predictability.

Farmer and Sidorowich [30] formulated a prediction technique also based on local linear maps. In a chaotic system nearby trajectories diverge apart exponentially fast at a rate given by the Liapunov exponent. This means that any initial error in estimating the state of the system will grow exponentially fast through time. Hence the accuracy of prediction of a system from an initial point into the future will decay exponentially as you try to predict the evolution of the point further ahead in time. For a periodic or steady state system this predictability should not decay. Hence an exponential decrease in predictability should be a signature of chaos. Farmer and Sidorowich exploited this property by measuring the predictability of a time series by the following method based on local linear maps.

A state, x_i, will evolve to some other state, x_{i+m}, a time $T = m\alpha$ in the future. Assume a map, $\mathbf{A}$, exists such that $x_{i+m} = \mathbf{A}(\mathbf{x}_i)$, where $\mathbf{x}_i$ is the delay vector. To approximate the map $\mathbf{A}$ we look for all nearest neighbours, $\mathbf{x}_k$, to $\mathbf{x}_i$, i.e. $\| \mathbf{x}_i - \mathbf{x}_k \| < \varepsilon$ and then fit the coefficients of $\mathbf{A}$ using a least-squares fitting procedure as described above. The point x_{i+m} is then predicted from $\mathbf{A}(\mathbf{x}_i)$. The accuracy of this prediction will decay exponentially with increasing T.

Further work in this area has been done by several authors; notably Casdagli [14] who suggested using radial-basis functions instead of linear maps and other authors who have explored the use of higher-order local maps. Neural networks have also been used to learn optimal local maps and seem to perform as well or better than Farmer and Sidorowich's method.

Sugihara and May [106] proposed a simplex prediction method. To predict the evolution of a point, $\mathbf{x}_i$, you look back in the time series for points that will enclose $\mathbf{x}_i$ within a simplex. They then look at the evolution of the simplex points a certain time $T = m\alpha$ into the future and use this to estimate a linear mapping from $\mathbf{x}_i$ to $\mathbf{x}_{i+m}$. Since the decay in prediction accuracy is due to a positive Liapunov exponent, λ, predictions will decay as exp(-tλ). However none of the decays Sugihara and May observe are exponential. This is important because non-chaotic systems can produce non-exponential decay. For example (due to R.Law and J.Reed (pers. comm.)) a power law decay is obtained from $x_{t+1} = f(x_t) + \sigma_n$ where $\sigma_n = \sum_{i=1}^{5} 2^i \sigma_{n-i}$ and $f(x) = xe^{1.5(1-x)}$. Additionally the dynamical instability associated with $\lambda > 0$ is essentially a local linear phenomenon. It is therefore important to able to work at small scales in the phase space where the nonlinearity of the dynamics is small. A consequence of this is that for prediction decay to indicate $\lambda > 0$, it is necessary for short term predictions to be accurate. Sugihara and May observe order 1 errors after only one month.

3.3.4 Dimension.

The dimension of a system is, in some way, the amount of information needed to specify the position of a point on an attractor to within some degree of accuracy. It is also a lower bound on the number of essential variables needed to model the dynamics. As such it has received a great deal of attention. There are a host of different dimension measures including for example the fractal dimension, D, information dimension, Hausdorff dimension and Liapunov dimension. For low-dimensional systems these are often, but not always, similar in value. The most widely used is probably the Grassberger-Proccacia [38] and [39] correlation dimension. Here we calculate the number of pairs of points within a certain distance, r, apart. The correlation integral $C(r)$ is defined as:

$$\begin{aligned} C(r) &= \frac{\text{number of pairs of points separated by less than r}}{\text{total number of points}} \\ &= \text{probability}(\| \mathbf{x}_i - \mathbf{x}_j \| < r), \end{aligned}$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ are randomly chosen points from the attractor Λ. This is approximated by:

$$C_E(r) = \lim_{N \to \infty} \frac{1}{N^2} \sum_{i,j=1}^{N} \Theta\{r > \| \mathbf{x}_i - \mathbf{x}_j \|\},$$

where Θ is the Heaviside step function which is equal to zero for the negative argument and one otherwise, N is the number of embedded vectors and $i \neq j$. For small r we expect $C_E(r) \propto r^{\nu_E}$. Hence a log-log plot will have gradient ν_E. Care must be taken to investigate $C_E(r)$ with different values of E. As E is increased then ν_E should converge to the correlation dimension, ν. Non-convergence indicates a high-dimensional system.

In reference [38] Grassberger and Proccacia compare their correlation dimension with the fractal dimension. Consider a coverage of an attractor by hypercubes of length length l. If the attractor is fractal then the number $M(l)$ of cubes that contain a piece of attractor is

$$M(l) \propto l^{-D},$$

where D is the fractal dimension [58]. This measure is insensitive to to the density of points in different areas of the attractor. It is only related to the geometrical structure. The correlation dimension is thus a superior measure as it is sensitive to the dynamical process, i.e. the fact that different areas of the attractor are visited at different rates. It also has a far greater convergence rate than D and so is more practically applicable.

3.4 Data requirements.

A lot of the techniques for identifying chaos need large amounts of data for reliable results. Indeed spurious results can be obtained when using small data sets. Eckmann and Ruelle [25] proposed the following minimum data requirements.

Any method to calculate Liapunov exponents requires that near a point one finds other points so that the rate of divergence of orbits can be estimated. The number of points in a ball of radius r around a point x is:

$$N(r) \propto r^d,$$

where d is the dimension of the attractor. Since $N(D) = N$, where D is the diameter of the reconstructed attractor and N the number of points:

$$N(r) = N\left(\frac{r}{D}\right)^d.$$

Now we need $N(r) \gg 1$ for statistical reasons and $\frac{r}{D} \ll 1$ in order to avoid nonlinearities. Hence we need:

$$logN > dlog(\frac{D}{r}).$$

For dimension calculations a similar argument holds except that the maximum number of pairs of points, assuming we calculate the correlation dimension, is $1/2(N - E)(N - E + 1) \approx 1/2N^2$. Hence we get:

$$2logN > dlog(\frac{D}{r}).$$

The quantity $\frac{r}{D}$ will depend on the extent of nonlinearity in the data, however we can estimate it to be of the order 0.1. Hence, for Liapunov calculations, assuming a low-dimensional attractor with $2 < d < 3$, we will need between 100 and 1000 points as a minimum. For larger dimensional systems, e.g. $d \approx 10$, then impractically large amounts of data are needed, $\approx 10^{10}$ points. Hence one must be very careful in the application of the above techniques to small amounts of data.

In ecology there are precious few long data sets. Hence there has been a need to formulate techniques that not only work in the presence of noise but also with small amounts of data. There are some, however, that may be within the range stated above. For instance, measles epidemics in New York City before vaccination began is a good candidate for analysis as the time series is ≈ 500 points and there is evidence to suggest it has a low-dimensional attractor of dimension $2 < d < 3$. Liapunov exponents calculations, from the above analysis, seem to need the square of the number of points used in dimension calculations. However, dimension calculations seem particularly ill-behaved with small amounts of data. In chapter 4 I will outline a technique for calculating Liapunov exponents which cannot assign them a particular value but can distinguish

whether an exponent is positive or negative even with small amounts of data. This is the central problem facing the detection of chaos in ecological systems.

3.5 Noise.

In physical systems it is often possible to extract large amounts of relatively clean data and it has thus been possible to demonstrate chaos in various well-controlled physical experiments in the laboratory [33]. Chaos in biological systems, however, has been much harder to identify. This is primarily due to (i) the far greater amount of 'noise' or stochasticity in biological systems and that (ii) data sets tend to be of a much shorter length than from physical experiments.

For infinite amounts of noise-free data the methods in sections 3.3.1, 3.3.2 and 3.3.3 achieve the Liapunov exponents by definition. The question, therefore, is how well the techniques work for finite data sets with varying amounts of noise? Noise has important effects on the particular methodologies.

Of importance to all the methodologies is the effect noise has on the embedding technique. Casdagli et. al., [16], studied the effects of, and how to achieve an optimal reconstruction in, the presence of noise. The delay vector technique is the most widely used, and easiest, but may not be the best. Singular value decomposition is an optimal local reconstruction but this is also complicated by the presence of noise.

In theory a stochastic system has infinite dimension as the noise tends to fill out the state space. For length scales smaller than the noise amplitude we would therefore expect the G-P correlation dimension to be near the embedding dimension, E, as the noise fills out a portion of E-dimensional space. For length scales far greater than the noise amplitude we would expect the dimension to behave as if noise was not present, [99]. The intermediate case is problematic and modifications have been proposed to attempt to address this problem, [99]. However, dimension estimates seem to be particularly sensitive to noise.

Liapunov exponents are defined at any level of noise, [49], and therefore provide a good criterion for identifying chaos when a stochastic component may be present, [27]. However, estimating the divergence of trajectories can be problematic in the presence of noise as the trajectories being compared will not have the same sequence of random events. Consequently the divergence between them might simply be due to the random component, [71]. This problem is resolved, at least for moderate amounts of noise, in a new method outlined in chapter 4.

Liapunov exponents are also the cause of the exponentially decaying predictability seen in chaotic systems. However when such techniques, e.g. [106], have been applied to noisy ecological data exponential decay is rare. This is important as stochastic systems can produce power law decays, particularly autocorrelated noise.

3.5.1 Autocorrelated noise.

The signatures of chaos identified in relatively noise-free signals are often masked by noise and are thus not apparent in a time-series with a significant stochastic component. Additionally it seems possible to tailor a noise-generating process that can reproduce many supposedly chaotic signatures [79]. Ellner, [27], in a review of detecting chaos in population dynamics, emphasises this point and uses a non-chaotic stochastic population model to generate many supposedly chaotic features. The point here is that *surrogate* non-chaotic data sets cannot be distinguished from truly chaotic data (see section 3.5.2). Of particular importance in ecological systems is the presence of *coloured* noise. By this is meant *autocorrelated* noise, $\varepsilon(t)$, which has the general form:

$$\varepsilon(t+1) = \beta\varepsilon(t) + Z(t),$$

with autocorrelation β and where $Z(t)$ is a sequence of independent random variables with mean zero and variance σ^2.

Nearly all techniques for identifying chaos work less well in the presence of coloured noise as opposed to Gaussian, or "white" noise. For example autocorrelated noise can

cause prediction decay, see section 3.3.3. A considerable number of ecological data sets display "coloured" spectra (e.g. [77] and [102]).

One conclusion to be drawn from the problems associated with noise is that it it important to use a battery of distinguishing characteristics rather than relying on one particular method.

3.5.2 Null hypothesis and surrogate data.

An approach taken by Theiler et. al. [110] and others, eg [98], [104] and [27], is the use of *surrogate data*. This method specifies some process, qualitatively different to the process we are attempting to identify, as a *null hypothesis*. Data sets are then generated consistent with this hypothesis. The discriminating method, or statistic, with which we are attempting to establish the existence of some property, e.g. chaos, is then applied to the generated surrogate data and to the original data. If the results are significantly different for the ensemble of values from the surrogate data than from the original data then the null hypothesis is rejected and the existence of the property is verified.

Some care needs to be taken in deciding which null hypothesis to test and therefore which process is used to generate surrogate data. Normally various properties of the original data are preserved in the null hypothesis in order to isolate some specific property to test.

In testing for nonlinearity in data Theiler et. al. [110] suggested various algorithms. One method, used in section 4.3, is to hypothesise the data comes from a linear Gaussian process while preserving the Fourier spectrum of the original data. One way to do this is to compute the Fourier transform of the original data, randomise the phases at each frequency by multiplying by $e^{i\theta}$, where θ is chosen randomly and independently from the interval $[0, 2\pi]$. The inverse transform is then the surrogate data.

Using such methods Theiler et. al. achieved good results for some model systems with small numbers of data points and large amounts of noise. Additionally they were

able to identify nonlinearity in some real data such as Rayleigh-Benard convection data.

Other authors propose surrogates based on stitching together Gaussian waveforms, [104], with random parameters estimated from the data or sine and cosine waves with added noise [27]. The particular method proposed is dependent on the problem and data being addressed.

4 Detecting chaos in a noisy time series.

4.1 Introduction.

A number of authors have addressed the problem of distinguishing chaos from noise in complicated noise-contaminated nonlinear systems (e.g. [30] and [14]). These systems are commonly found in biology, ecology and epidemiology (see Sugihara and May [106] and [78] and [27]). In this note we propose a new approach based on a method outlined by Wolf et. al. [115], see section 3.3.1.

Sensitive dependence is equivalent to the fact that, in a chaotic system, typical nearby orbits diverge exponentially fast from each other (see section 3.2.1), at least on average. The Liapunov characteristic exponents (LCEs) are the different long-time exponential rates at which nearby orbits diverge or converge. Thus we have sensitive dependence provided the largest LCE, χ, is positive. The exponential separation only occurs while the orbits are so close that their separation is described by the linearisation of the dynamical system. It ceases when nonlinear effects come into play. In the repeated stretching and folding that produces chaos, the stretching is essentially a local linear phenomenon that determines χ, and the folding a larger scale property that should not enter into exponent calculations. Hence, in calculating χ, it is important to work at scales d in the phase space smaller than the scale d_{nonlin} above which the nonlinearity is significant. However, real data is always contaminated with some degree of noise and it is precisely at small scales $d < d_{\text{noise}}$ that this noise has the greatest effect. We therefore should require that $d_{\text{noise}} < d < d_{\text{nonlin}}$. The usual mathematical definition of χ involves taking the limit $d \to 0$ which is clearly injudicious in the presence of noise.

We will show that in some noisy systems it is not possible to give a value for χ. On the other hand, an essential point about our approach is that, even in such a system, it can indicate the sign of χ and hence signal the presence of chaos. It also indicates when it is reasonable to assign a value to χ.

4.2 Methodology.

For most real world ecological or biological systems it is often difficult to identify all of the state variables since they are generally embedded in a more complex, higher-dimensional system. Very often we only have a time-series $\{x_i\}_{i=1}^{N'}$ of discrete measurements of a single observable. The well-known technique of phase space reconstruction with delay coordinates (see section 3.2.2) makes it possible to obtain from such a time-series an attractor whose Liapunov exponents are identical with those of the original attractor.

Firstly, we choose an embedding dimension, E, and a time-delay, τ, and then use delayed coordinates to represent each lagged sequence of data points

$$\mathbf{x}_i = (x_i, x_{i-\tau}, ..., x_{i-(E-1)\tau}),$$

as points in E-dimensional space, where E is known as the *embedding dimension* and τ the *delay time*, which is conventionally taken to be an integer here. The time-series is discretely sampled $t_i = i\alpha$ where α is the sample time. In all our examples we take $\tau = 1$. Let Λ denote the trajectory $\{\mathbf{x}_i\}_{i\geq 0}$. Because of the topological mixing, our reconstructed attractor, though defined by a single trajectory, can provide points that may be considered to lie on different trajectories provided our trajectory is long enough.

Rather than try to estimate a single value for χ we instead approximate the function $\chi(d)$ at different spatial scales d. We fix, assuming it exists, a spatial scale $d < d_{\text{nonlin}}$. Suppose that $\mathbf{x}_s$ and $\mathbf{x}_t$ are points of the trajectory whose initial distance apart $d(0)$ is approximately d. Let $d(u) = ||\mathbf{x}_{s+u} - \mathbf{x}_{t+u}||$. If the motion is chaotic, we expect $d(u)$ to grow exponentially in time provided it is less than kd, where k is some small fixed integer. When $d(u)$ is greater than kd we replace $\mathbf{x}_{s+u}$ by a point $\mathbf{x}_{s'}$ in our trajectory which is within d of $\mathbf{x}_{t+u}$ and, with respect to $\mathbf{x}_{t+u}$, the angular separation between $\mathbf{x}_{s'}$ and $\mathbf{x}_{s+u}$ is small. Continuing in this way, we can measure the exponent χ. To make this precise we now give a slightly more formal definition of this approach.

Given the time series x_i above, we define vectors z_i and y_i in Λ and a sequence of steps m_i and a_i as follows:

(i) Let $z_0 = x_0$ and $a_0 = 0$.

(ii) Let $L(m) = ||x_{a_m} - y_m||$, where $y_m = x_s$ is the vector in Λ closest to $x_{a_m} + (1/k)(z_m - x_{a_m})$. Let $t > 0$ be such that for $\alpha u < t$, $d/k < ||x_{a_m+u} - y_{m+u}|| = L'(m) < kd$ while for $\alpha u \geq t$ this is not the case. Then $a_{m+1} = a_m + u$ and $z_{m+1} = y_{m+u}$.

Let $D_0 = 0$ and define

$$D_{m+1} = D_m + \frac{1}{\alpha u} \log_2 \frac{L'(m)}{L(m)}$$

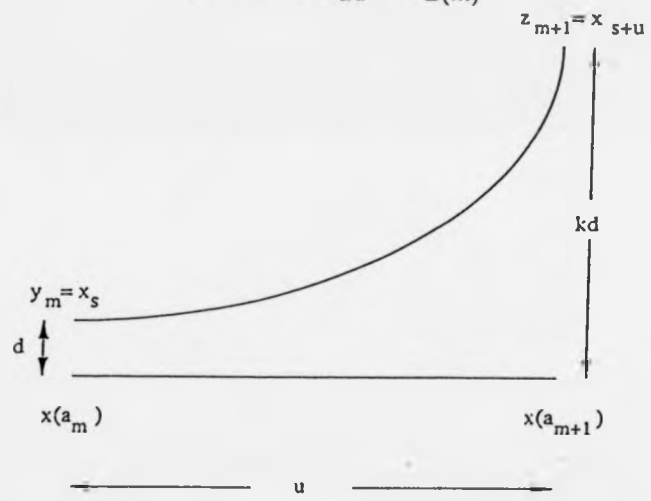


Figure 6: A schematic diagram illustrating the algorithm used in calculating $\chi(d)$ for a specific scale d. For a negative exponent the vector z_m will be a distance $k^{-1}d$ from x_{a_m+1}.

Then $\chi(d)$ is approximated by by the slope of the best fitting straight line through the set of points $\{(m, D_m)\}$.

In a noise-free chaotic system, $\chi(d)$ will be approximately constant for small d and should converge to the largest characteristic exponent χ as $d \to 0$. On the other hand,

if the noise amplitude a is large in magnitude compared with d then $\chi(d)$ will be larger than χ. This is due to the effect of noise pushing initially close points apart (the typical behaviour in multi-dimensional space).

To see the effect of this we can crudely argue as follows. With noise, our nearby trajectories start d apart and, after one time step $t = \alpha$ in our time series, end, not $2^{\chi\alpha}d$ apart but $2^{\chi\alpha}d+a$. Thus our estimate for the local divergence of the trajectories is given by $\alpha^{-1}\log_2(2^{\chi\alpha}d+a/d)$. When a is large compared to $2^{\chi\alpha}d$, which is of the same order as d, this is approximated by $\alpha^{-1}(\log_2 a - \log_2 d)$. As N is fixed, this will tend to decrease with gradient α^{-1} with increasing $\ell = \log_2 d$. On the other hand, when d is large compared to a, $\chi(d)$ is approximately the true exponent value χ and we expect to see no change in $\chi(d)$ with slightly increasing d. If d is increased further, $\chi(d)$ decreases as nonlinear effects become important and eventually converges to 0 when d is of the order of the width of the attractor.

In order to measure χ, we need that $d_{\text{noise}} < d_{\text{nonlin}}$. The scale d_{nonlin} is related to the smallest length scales at which the nonlinearity becomes important. For some systems this scale will be clear from the graph, while for others an accurate estimate will be needed (e.g. see [15]). As a first estimate we arbitrarily, but conservatively, take this scale to be 5% of the transverse attractor width. In order to estimate d_{noise} we extend a technique due to Broomhead and King [10] based on ideas from singular systems theory [8]. We do not intend to explain all the background but instead refer the reader to the references above. Given an embedded time series, $\mathbf{x}_i$, as above ($\tau = 1$ and E is suitably chosen) we let $\mathbf{Y}$ be the matrix whose i'th row is the vector $\mathbf{x}_i$ for $i = 0, 1, \ldots\ldots, N$ and let the trajectory matrix, $\mathbf{X} = N^{-\frac{1}{2}}\mathbf{Y}$. In order to centre the mass at the origin we subtract the mean of each column from each element of the column. If all data points are used $N = N_T - (E-1)$ where N_T is the number of data points in our original time series. The singular value decomposition of this matrix is given by $\mathbf{X} = \mathbf{S}\mathbf{\Sigma}\mathbf{C}^{\mathbf{T}}$ where $\mathbf{S}$ and $\mathbf{C}^{\mathbf{T}}$ contain the left and right singular vectors and the

entries of the diagonal matrix Σ are the singular values $\sigma_1 > \sigma_2 > ..... > \sigma_E$. As in Broomhead and King we expect all the singular values to be non-zero due to noise. However, we can distinguish between singular values associated with a deterministic component in the time series and those dominated by noise due to the presence of a non-zero noise floor in the singular spectrum. Given that we can separate the singular values which are noise dominated, we expect the noise amplitude at each time step, d_α, to equal the total noise power in the singular spectrum. Hence

$$d_\alpha = \left(\frac{1}{(E-n)} \sum_{i=n}^{E} \sigma_i^2 \right)^{\frac{1}{2}}$$

where n is the largest eigenvalue which is noise dominated. Note also that the variance of the time series is equal to the total power in the spectrum:

$$\frac{1}{N_T} \sum_{i=0}^{N_T-1} (\nu_i - \bar{\nu})^2 = \frac{1}{E} \sum_{i=1}^{E} \sigma_i^2.$$

4.3 Results.

Firstly, we consider model systems to illustrate these ideas. Unless otherwise stated noise is Gaussian white noise with mean 0, amplitude a and variance a^2. This is added to one dimension of the system dynamically at each time step. For differential equations the noise is scaled by the square root of the time step. The variance is thus kept constant over one period. The exponent $\chi(d)$ has units bits per unit time for the Lorenz equations, bits per year for the SEIR equations and the real epidemic data and bits per mean orbital period for the coupled oscillator data. The results in figure 7 are for the Lorenz equations with chaotic parameter values chosen. We have used the known equations of motion to generate two trajectories separated initially by d with different noise realisations added to each. We then follow the procedure as explained in the methodology except we generate new vectors rather than searching for them through a time series. We see the expected dependence of $\chi(d)$ upon $\ell = \log_2 d$ most clearly. For small ℓ the slope is α^{-1}. Then the graph shows a plateau giving the approximate true

value of χ. For larger ℓ, the nonlinear effects become apparent. From this graph one can also deduce the scales d_{noise} and d_{nonlin} at which the noise and nonlinearity are effective.

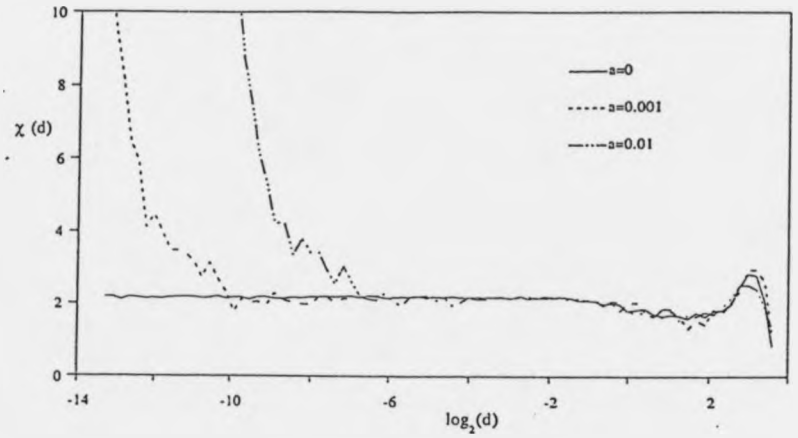


Figure 7: The graph of $\chi(d)$ versus $log_2(d)$ using the Lorenz equations with $\sigma = 16.0$, $\rho = 45.92$, $b = 4.0$ and sample time 0.005. The equations have been used to generate trajectories a known distance apart. Hence no embedding is needed as the exact state of the system is known. We calculate the largest exponent χ to be 2.17 bits per unit time for $a = 0$.

The other graphs are all calculated from a time-series treated as though it was data with no prior knowledge of the system that produced it. The deviation of these other figures from the nearly ideal picture is due to the relative scarcity of data points.

In figure 8 we have used the SEIR equations for modelling epidemics and the measles parameter values taken from Olsen and Schaffer [73]. It should be noted that the time-series are only of 500 points, yet the difference between the chaotic measles dynamics and the periodic orbit of the Lorenz equations is clear. Additionally when there is no noise in the system it is possible to obtain a reasonable estimate of χ for the chaotic measles dynamics. On the other hand, in the presence of significant noise it is not possible to associate a value to χ even though we can clearly distinguish its sign, i.e.

the presence or absence of chaos.

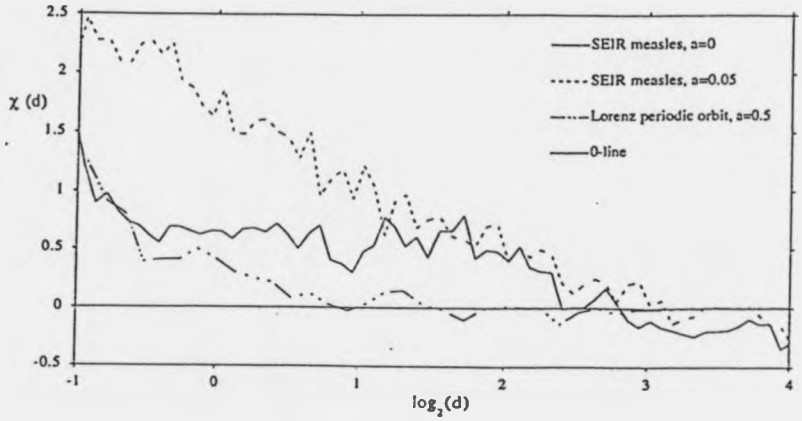


Figure 8: The graph of $\chi(d)$ versus $log_2(d)$ using time-series of the SEIR equations with measles parameter values, $m = 0.02$, $a = 35.84$, $g = 100.0$, $b0 = 1800.0$ and $b1 = 0.28$, and for a periodic orbit of the Lorenz equations, $\sigma = 10.0$, $\rho = 350.0$ and $b = 8/3$. In all our examples τ, the delay time for the embedding equals one. The time series length is 500 points, $\alpha = 0.1$ years for the SEIR equations and 0.1 for the Lorenz equations and the embedding dimension, E, equals 6. Note the reasonable approximation to the true exponent value, $\chi = 0.46$ bits per year, in the time-series with no noise. The Lorenz time-series has ten times the noise amplitude in order that the signal to noise ratio is comparable to that in the SEIR time-series.

For the Lorenz periodic orbit, when ℓ is slightly larger than the noise level $\chi(d)$ is negative reflecting convergence onto the attractor due to a negative characteristic exponent. However when ℓ is much greater than the noise level we cannot find points displaced off the attractor by this distance. We can only find points further away on the deterministic orbit and so cannot measure any convergence. Hence $\chi(d)$ is approximately zero at these scales. This shown more clearly in figure 9 using 1000 points from a periodic orbit of the the Lorenz equations.

In figure 10(a) we plot the normalised singular spectrum for a time-series from the

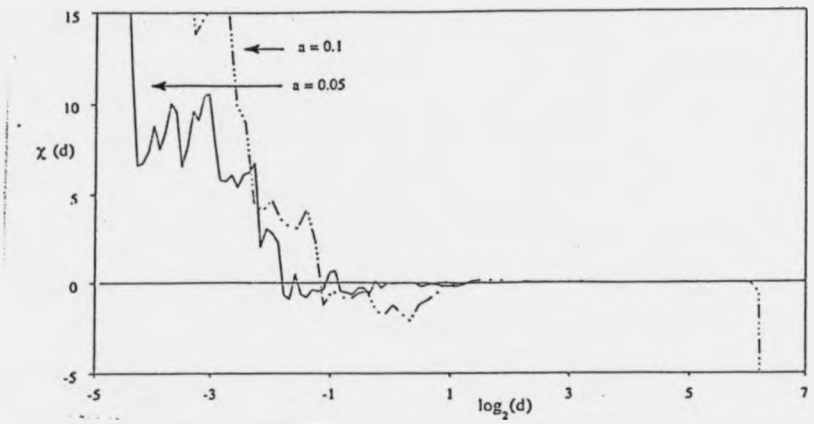


Figure 9: The graph of $\chi(d)$ versus $log_2(d)$ using time-series from the Lorenz equations, $\sigma = 10.0$, $\rho = 350.0$ and $b = 8/3$. The time series length is 1,000 points, $\alpha = 0.005$ and the embedding dimension, E, equals 6.

Lorenz equations with measurement noise, $a = 0.01$, added. After the ninth eigenvalue we see the constant noise floor and so we use this eigenvalue as our cutoff. In figure 10(b) we plot the log of the estimated noise amplitude $d_n = d_\alpha \alpha^{-\frac{1}{2}}$, against the cutoff used. For measurement noise we see an almost perfect prediction of the noise in the system. Figures 10(c) and (d) show the same except for dynamical noise. The quantity $log_2(d_{\text{noise}})$ is then calculated by $log_2(E^{\frac{1}{2}} d_n)$ where E is the embedding dimension used in calculating $\chi(d)$. We obtain estimates for the Lorenz equations, with $E = 3$, of $log_2(d_{\text{noise}}) = -9.5$ for $a = 0.001$ and $log_2(d_{\text{noise}}) = -6.3$ for $a = 0.01$. There is a very good comparison of these estimates of $log_2(d_{\text{noise}})$ with the nearly ideal graph in figure 7. We have used the quantity $log_{10}(d_n)$ for figure 10 to illustrate the very good estimate of the noise amplitudes. In following noise amplitude graphs we will plot the quantity $log_2(d_{\text{noise}})$ in order to facilitate direct comparison with the graphs of $\chi(d)$.

Finally, we consider some real data sets. The first is a time-series from an electronic circuit model of the chaotic Van der Pol-Duffing oscillator, with no external forcing and three degrees of freedom. In figure 11 we see that there is an approximate plateau and

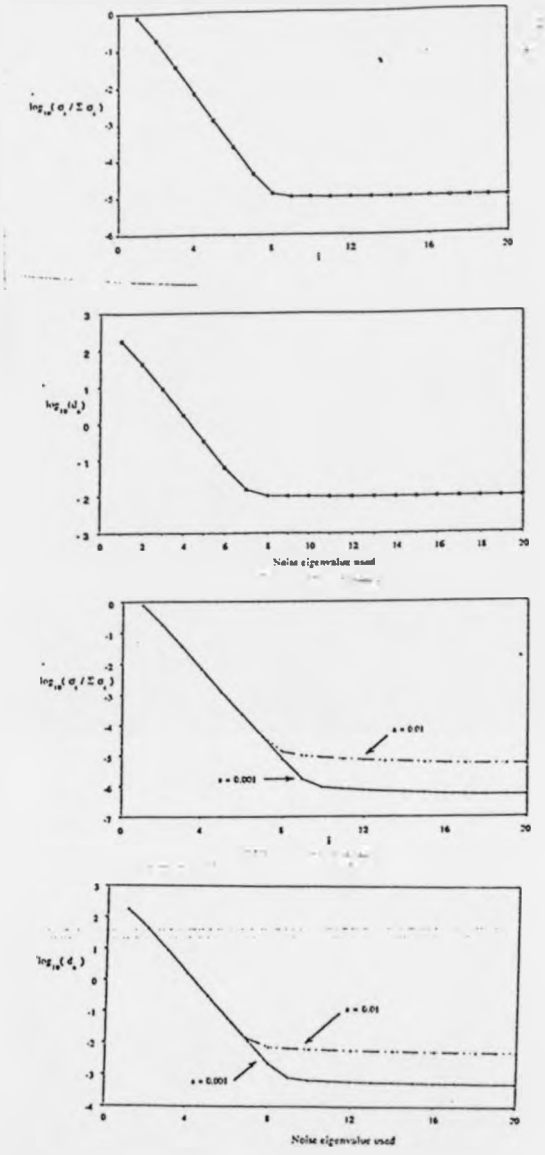


Figure 10: The singular spectrum and estimated noise for the Lorenz equations, $\sigma = 16.0$, $\rho = 45.92$ and $b = 4.0$, using 10,000 points, window width = 20 and $\alpha = 0.005$. (a) The normalised singular spectrum for measurement noise $a = 0.01$. (b) The estimated noise level for measurement noise $a = 0.01$. (c) The normalised singular spectrum for the Lorenz equations and dynamical noise $a = 0.001$ and $a = 0.01$. (d) The estimated noise level for dynamical noise $a = 0.001$ and $a = 0.01$.

so the time-series is chaotic with $\chi \approx 0.6$ bits per mean orbital period.

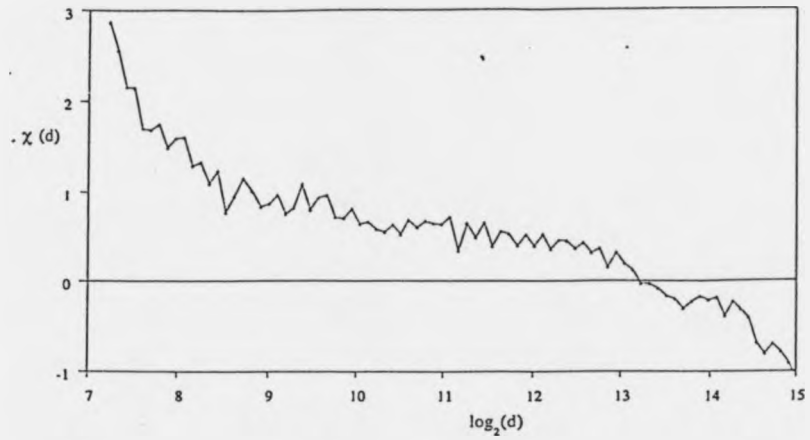


Figure 11: The graph of $\chi(d)$ versus $log_2(d)$ using a time series from an electronic circuit model of the chaotic Van der Pol-Duffing oscillator, with no external forcing and three degrees of freedom. The time series length is 5,000 points, $\alpha = 1/30$ of the mean orbital period and the embedding dimension, $E = 7$.

Figure 12(a) is calculated from a time-series of monthly reported measles cases from 1928-1966 and monthly reported chickenpox cases from 1928-1972 in New York City. The difference between the two is apparent with the chickenpox case possessing some of the features of a periodic orbit and the measles case looking chaotic. Further work is needed to precisely distinguish these cases as the lengths of these two time-series are on the boundary for this method. Figure 12(b) is the estimated noise in each time series. This gives estimates for $log_2(d_{\text{noise}})$ in figure 12(a) as -2.9 for measles and -3.4 for chickenpox. The graphs appear to have much more noise than the figures above suggest. This is due to scarcity of data since this defines a scale below which we cannot find sufficient points to form a realistic embedded trajcctory.

Figure 13(a) is from a time-series of measles cases reported weekly in England and Wales from 1948 to 1966. Here we need to know precisely the scale d_{nonlin} in order to interpret this graph. If $log_2(d_{\text{nonlin}}) < 0$ then the exponent is positive, otherwise the

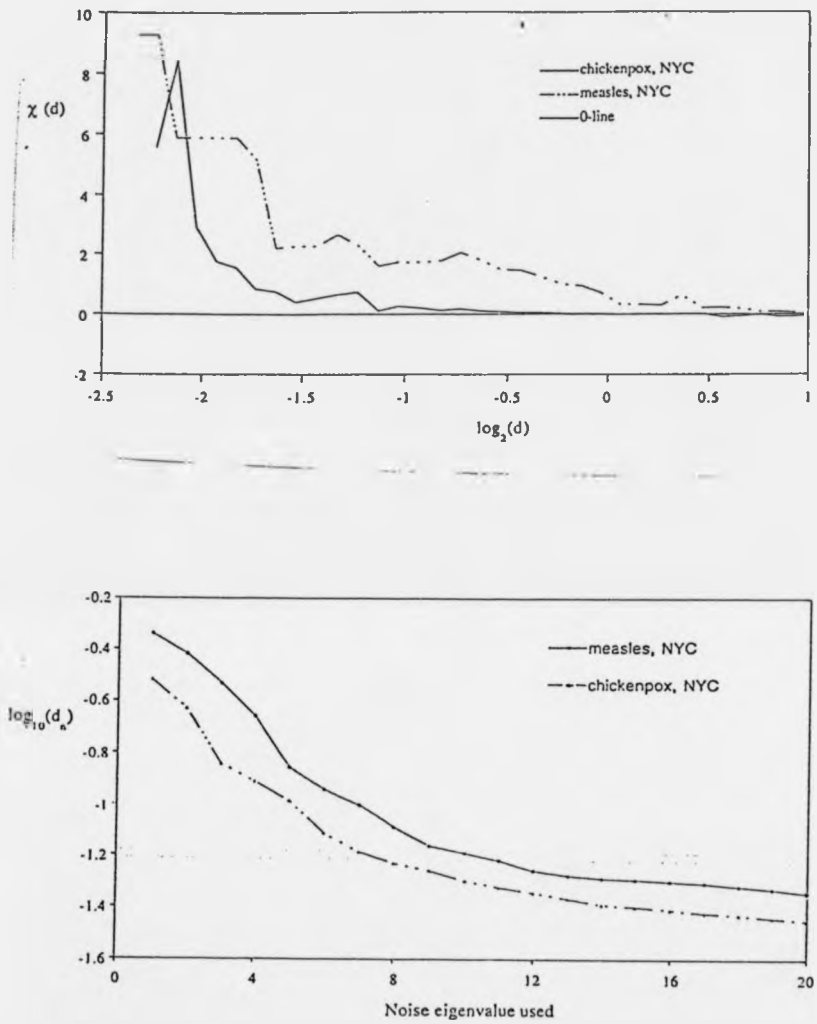


Figure 12: (a) The graph of $\chi(d)$ versus $log_2(d)$ using time-series of monthly reported measles and chickenpox cases as reported by doctors in New York City from 1928 to 1966 and from 1928 to 1972 respectively. The time-series lengths are 456 and 534 points, α equals one month with embedding dimension, $E = 6$ used. (b) The estimated noise level for each time-series, window width = 20.

exponent is negative. Preliminary results, based on the width of the attractor, suggest $log_2(d_{\text{nonlin}}) \approx 1.0$. Further work directly measuring the degree of linearity of the data at various spatial scales is needed to estimate this more precisely (see [15]). Figure 13(b) is the estimated noise.

It has been suggested to us that we generate surrogate time series by taking the Fourier transform of a data set then inverting the transform using the observed amplitudes but with random phases (as in Theiler et. al. [110], see section 3.5.2). If our method is able to distinguish these stochastic surrogates from the data then it is a very reliable one. This transformation will effectively increase the noise in the system as well as destroying any dynamic structure. We illustrate this in figures 14(a) and (b) for the Lorenz equations showing the difference between the original data, which has a clear plateau, and the surrogate data. When there is not a clear plateau present, though, this will not be a good comparative method as it is important to compare time-series of similar signal-to-noise ratios. Hence the surrogate data in figure 14 is not comparable to the SEIR data in figure 8.

In general it seems possible to mimic any signature of chaos by tailoring some noise generating function. Hence we feel it is important to use a series of tests in order to distinguish chaos and not to rely solely on one method.

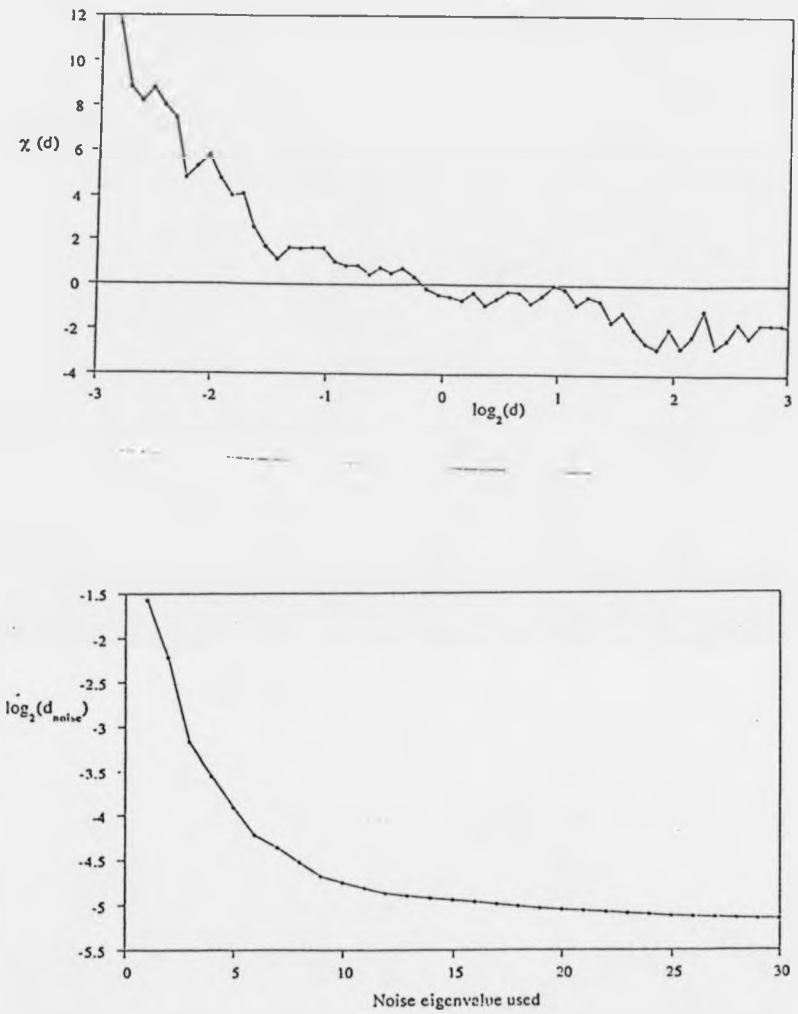


Figure 13: (a) The graph of $\chi(d)$ versus $log_2(d)$ using time-series of weekly reported measles cases in England and Wales from 1948 to 1966. The time-series length is 991 points, α equals one week with embedding dimension $E = 6$ used. Being weekly, this time-series has more points than the two from New York City. However, the overall time period (18 years versus 38 and 42 years) is much shorter. This is probably the reason why this graph is less clear than the previous two. (b) The estimated noise level for the time-series, window width = 30.

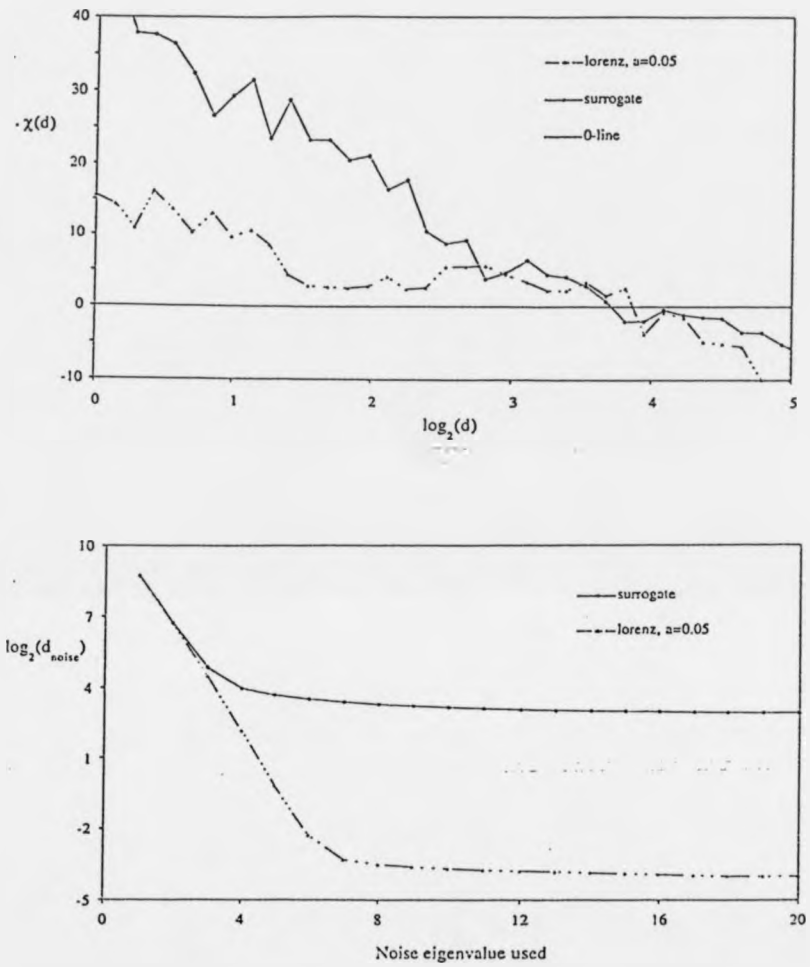


Figure 14: (a) A comparison of a Lorenz data set and a surrogate data set generated from it. The parameters used for the Lorenz equations are $\sigma = 16.0$, $\rho = 45.92$ and $b = 4.0$. The surrogate data was generated by taking the Fourier transform of the Lorenz data, multiplying each complex amplitude by $e^{i\theta}$ where θ is independently chosen for each frequency from $[0, 2\pi]$, and then taking the inverse transform. The sample time = $\alpha = 0.005$, time series length was 1,000 points and the embedding dimension was $E = 6$. (b) The estimated noise amplitude for each data set in (a). This clearly shows the large increase in noise in the surrogate data. Hence for very noisy time-series generating surrogate data will not be a good comparative tool. The window width used in the SVD procedure was 20.

5 Dynamics and evolution.

5.1 Density-dependent selection.

On an ecological time-scale there exist processes that regulate population densities. These processes also have evolutionary effects and as such the two need to be analysed together. Early work looking at the evolutionary effects of density-dependent population regulation was done by MacArthur [56] and MacArthur and Wilson [57]. They identified two evolutionary processes arising from such regulation: "r-selection" and "K-selection". Populations that are expanding and have negligible crowding effects undergo r-selection where selection acts to maximise the reproductive rate or intrinsic rate of increase, r. Crowded populations undergo K-selection where selection acts to maximise the carrying capacity, K. Roughgarden [85] used Sewall Wright's notion of selective value and showed that in both cases evolution favoured the genes that produced the phenotype with the highest selective value. This was a 2-allele genetic model with one homozygote having high r/low K and the other having low r/high K. The density-dependence was introduced as a linear drop off in the selective value with increasing population size. As pointed out in this paper the selective value can be interpreted as a measure of the fitness of a phenotype and also as a determination of the growth characteristics of a population of that phenotype. The two are intrinsically related.

5.2 Frequency-dependent selection and linear game theory.

5.2.1 Introduction.

Maynard-Smith and Price [66] introduced the concept of game theory to biology. This deals with the complications of *frequency-dependent* selection within a strategy model. Although game theory was developed in economics it has turned out to be more easily applicable in biology. There are two reasons for this: (i) a common currency is used for payoffs for different outcomes from different strategies and in biology there is a genuine

currency with which to compare all outcomes, *Darwinian fitness.* (ii) The solution of the game is sought in terms of *evolutionary stability* as opposed to rationality in economics. This is defined as:

Definition 3 *An evolutionary stable strategy, an ESS, is a strategy such that if all members of a population adopt it, then no mutant strategy could invade it under the influence of natural selection, [64].*

Stability is an important concept as non-stable strategies are not expected to be observed in populations since, by definition, they will be invaded and taken over by other strategies.

Evolutionary game theory was first applied by Maynard-Smith and Price to deal with animal conflicts. As such it was couched in terms of strategies. However it can equally apply to any phenotypic variation, e.g. dispersal strategy. Thus, evolutionary game theory refers to the evolution of phenotypes in a population when the fitness of a particular phenotype depends on the frequencies of other phenotypes in the population. An example is mimicry in the passion flower butterfly, *Heliconius melpomeme* and *H. erato*. *H. erato* is a distasteful butterfly to predators which learn to avoid butterflies with the warning pattern (a striking colouration). *H. melpomeme* has evolved to mimic the colouration of *H. erato* so it too avoids predation. This will only work while the relative frequency of *H. melpomeme* remains small, otherwise predators will re-learn that there is a good chance of a fair-tasting meal from the strikingly coloured butterflies.

An essential first step of a game theoretic analysis is to specify the set of possible phenotypes that could be taken up. In the case above it is possible that the set is nearly continuous from all one colour to a fantastic mixture of colour and patterns. In specifying this set any *constraints* must be identified. For instance it may not be possible to display a certain colour. Constraints such as these and *trade-offs*, where the development of one characteristic inhibits or constrains another, turn out to be crucially important and game theory highlights them by ensuring they must be specified

beforehand.

5.2.2 The Hawk-Dove game.

The hawk-Dove game, introduced by Maynard Smith and Price [66], is the basic model in evolutionary game theory. Two animals are competing for a resource which will increase their fitness, if obtained, by a value V. The resource may be a particularly good territory for breeding, for example. Two strategies are available to the contestants; to be a Hawk or a Dove. The Hawk strategy is to escalate the contest, i.e to fight. The Dove strategy is only to display and if the opponent escalates to run away. Thus, if a Hawk meets a Dove the Hawk escalates, the Dove runs away and the Hawk gains all the resource V. If a Hawk meets another Hawk then they both escalate and fight and there is a chance of an injury at a cost C. The uninjured Hawk then gains the resource. On average a Hawk will win the resource half the time and be injured half the time. Thus the average payoff to a Hawk in a Hawk-Hawk contest will be $(V-C)/2$. When a Dove meets another Dove they both display and they share the resource equally, so the payoff will be $V/2$. We can thus write down the *payoff matrix* for the different strategies. In

	H	D
H	$(V-C)/2$	V
D	0	$V/2$

Table 2: Payoff matrix for the Hawk-Dove game.

this matrix the entries, or payoffs, show the increase in fitness to the strategies as the outcome of the particular contest. Let W_0 be the fitness of an individual prior to the contest. Let the fitness of a Hawk strategy be $W(H)$ and the fitness of a Dove be $W(D)$. Let $E(D,H)$ be the payoff from strategy D playing strategy H and p be the proportion

of the population playing the Hawk strategy. Then:

$$\begin{aligned} W(H) &= W_0 + pE(H,H) + (1-p)E(H,D) \\ W(D) &= W_0 + pE(D,H) + (1-p)E(D,D). \end{aligned} \tag{14}$$

The dynamics of this game and the frequency of Hawks in the next generation, p', are given by:

$$p' = pW(H)/\overline{W}, \tag{15}$$

where $\overline{W}$ is the average fitness of the population and is given by $\overline{W} = pW(H) + (1-p)W(D)$. The dynamics of the Doves can be completely determined from the dynamics of the Hawks since the total proportion of Hawks and Doves equals one. The dynamics are 1-dimensional and take place on a line. Equation 15 is analytically complex to solve but we can understand the dynamics by looking for stable strategies, or ESS's. The dynamics of a mutant strategy J in a population of strategy I are given by:

$$\frac{q'}{q} = \frac{W(J)}{qW(J) + (1-q)W(I)},$$

where $q \ll 1$ is the proportion of mutants, J. Since $q \ll 1$ the number of mutants will only decrease if $W(I) > W(J)$. The fitnesses of the strategies are:

$$\begin{aligned} W(I) &= W_0 + (1-q)E(I,I) + qE(I,J) \\ W(J) &= W_0 + (1-q)E(J,I) + qE(J,J). \end{aligned} \tag{16}$$

Since $q \ll 1$ this requires for all $J \neq I$,

$$\textit{either} \quad E(I,I) > E(J,I) \tag{17}$$

$$\textit{or} \quad E(I,I) = E(J,I) \quad \textit{and} \quad E(I,J) > E(J,J). \tag{18}$$

These conditions for an ESS were given in Maynard-Smith and Price [66].

5.2.3 Pure strategies, mixed strategies and polymorphic populations.

For the Hawk-Dove game it is clear that Dove can never be an ESS as $E(H,D) > E(D,D)$ and so a Hawk strategy could always invade. Hawk will be an ESS if $(V - C)/2 > 0$, i.e. $V > C$. This is intuitive as this means the reward for gaining the resource outweighs the cost of an injury. Hence the only possible pure strategy is Hawk. Given a payoff matrix, such as in table 2, it is possible to immediately identify ESS's as the diagonal payoff needs to be greater than any of the other payoffs in that column for a strategy to be an ESS. If $V < C$ there is no pure ESS but there is the possibility of a mixed strategy ESS.

A mixed strategy is one where an individual plays different strategies for set proportions of time. This mixed strategy is then asexually passed onto the next generation. To identify ESS's in the mixed strategy case a theorem proved by Bishop and Cannings [9] is needed: If I is a mixed ESS which includes, with non-zero probability, the pure strategies A,B,C..., then

$$E(A,I) = E(B,I) = E(C,I)... = E(I,I). \tag{19}$$

To find the mixed ESS in the Hawk-Dove game we need to solve $E(H,I) = E(D,I)$. Let P be the mixed ESS and the time spent playing the Hawk strategy:

$$\begin{aligned} E(H,P) &= PE(H,H) + (1-P)E(H,D), \\ E(D,P) &= PE(D,H) + (1-P)E(D,D). \end{aligned} \tag{20}$$

Thus:

$$\begin{aligned} P\frac{(V-C)}{2} + (1-P)V &= (1-P)\frac{V}{2}, \\ \Rightarrow (1-P)V &= (C-V)P, \\ \Rightarrow P &= \frac{V}{C}. \end{aligned} \tag{21}$$

Hence, if there is a mixed ESS then it will be given by equation (21). We need to check that this satisfies equation (18) since $E(P,P) = E(H,P) = E(D,P)$. Therefore we

need $E(P,H) > E(H,H)$ and $E(P,D) > E(D,D)$.

$$\begin{aligned} E(P,H) &= PE(H,H), \\ &= \frac{V}{C}.\frac{(V-C)}{2}, \\ &> E(H,H) \quad ,as \quad C > V. \\ E(P,D) &= PE(H,D) + (1-P)E(D,D), \\ &> E(D,D) \quad ,as \quad E(H,D) > E(D,D). \end{aligned}$$

Thus a mixed strategy given by $P = V/C$ is an ESS when $V < C$. When $V > C$ then the pure hawk strategy is an ESS.

The third situation that is possible is that there could be a population made up of pure strategies of Hawk and Dove which could be evolutionary stable. If $V > C$ then a pure Hawk population would be an ESS. If this is not the case, is there a stable genetic polymorphism made up of a mixture of pure breeding Hawks and Doves? At equilibrium the fitnesses $W(H)$ and $W(D)$ must be equal. That is

$$pE(H,H) + (1-p)E(H,D) = pE(D,H) + (1-p)E(D,D)$$

where p is the frequency of Hawks. This equation is the same as equation (20). Thus the condition for a mixed ESS is the same as for genetic polymorphism. This is also true when there is more than two strategies. However is this polymorphism stable? When there are only two strategies then when the mixed ESS is stable then so is the polymorphism. When there are more than two strategies then the situation is more complex. There are examples when the mixed ESS is unstable and the polymorphism is stable and *vice versa*. However a polymorphic population in which only pure strategies are possible will always be stable if it satisfies conditions (17) and (18) against invasion by any pure *or mixed strategy*. It must satisfy these conditions for any mixed strategy as it is possible for a population to be stable against invasion by all pure strategies while still being unstable to a mixed strategy invasion. For example the payoff matrix in table 3, due to [64], has strategy A stable to invasions by pure B or C but not against the

mixed strategy $(0, \frac{1}{2}, \frac{1}{2})$.

	A	B	C
A	1	1	1
B	1	0	10
C	1	10	0

Table 3: A game in which strategy A is stable to any pure invasion but not if B and C mutants invade simultaneously.

5.2.4 Playing the field.

In some cases the fitness of a certain strategy will depend not on the outcome of individual contests but on some average property, or strategy, of the population as a whole. An example of this type is the evolution of the sex ratio. The success of a particular strategy, i.e. the proportion of male or female offspring an adult has, depends on the ratio of the number of males to females in the whole population. In this case let the fitness of a single J strategist in a population of I strategists be $W(J, I)$. Clearly I will be an ESS if for all $J \neq I$, $W(J, I) < W(I, I)$. If $W(J, I) = W(I, I)$ then we need to think of a population P consisting of a small proportion q of J strategists and $(1 - q)$ I strategists. Then the fitness of the J strategy is $W(J, P)$ and if this is less than $W(I, P)$ then I is again an ESS.

In a 2-strategy game then I is an ESS if $W(I, I) > W(J, I)$ and J is an ESS if $W(J, J) > W(I, J)$. If neither condition is true then the ESS will be a mixture of J and I. In this case the frequencies of I and J can be found, as in the Hawk-Dove game, using the condition $W(I, P) = W(J, P)$ where P is the equilibrium population. For a continuously varying strategy, s, finding the ESS is more difficult. We seek a strategy s^* which is uninvadable, i.e. $W(s, s^*) < W(s^*, s^*)$ where $s \neq s^*$ and is a small invading

population. If the function $W(s, s^*)$ is differentiable then we can find s^* by locating the function's maximum, i.e.

$$\frac{\partial W(s, s^*)}{\partial s} \mid_{s=s^*} = 0, \tag{22}$$

and

$$\frac{\partial^2 W(s, s^*)}{\partial s^2} \mid_{s=s^*} < 0.$$

If (22) has a unique solution then the the strategy s^* will be an ESS. If there is more than one possible solution then the situation is more complex as the strategy s^* is a local maximum. If it is also a global maximum then it is an ESS. Otherwise we need to redefine the concept of stability to include invasion only by those mutants close to s^*. The strategy s^* will then be locally evolutionary stable.

5.2.5 Assumptions, complications and extensions of the model.

Various assumptions are inherent in the simple Hawk-Dove game analysed above. Some have used these assumptions and simplifications to suggest the limited and artificial nature of the game theoretic approach. On the contrary though, it seems that the basic concept is very robust and able to deal with a wide variety of different circumstances. The following assumptions have been relaxed in order to extend the model into new and interesting areas.

(i) The population reproduces asexually. Most populations that are of interest are sexual and so extending the approach to diploid populations with genetics is clearly a natural progression.

If the ESS is a phenotype that can be produced by a homozygous diploid genotype then the ESS will be stable. If the ESS is a mixed one which can only be achieved by a genetically variable population then complications can arise. Work by Maynard Smith [63], Maynard Smith and Parker [65] and others has demonstrated that sexual populations evolve "as close as possible" to the ESS. Hines [46] showed that for the single-locus multi-allele case with additive inheritance the population mean will evolve to

the ESS when possible. Treisman [111] considered the cases of both additive inheritance and strict dominance for the two-allele model and reached a similar conclusion. The ESS approach is in fact relevant for fairly general inheritance patterns. However, it is possible that natural genetic constraints can adversely affect the stability of an ESS or can prevent the population from evolving to it. This is due to the non-convex nature of the set of possible mean strategies.

It is often reasonable to assume that there is some additive genetic variance underlying phenotypic variability because artificial selection experiments have almost always revealed such variance. Thus we can assume that 'like begets like'. Then in general as the number of loci, or number of alleles per locus increases, it becomes more likely that a population will reach the ESS, [96].

(ii) The population is randomly mixed. If dispersal of individuals is not random then it is more likely that one mixes with like individuals. Models looking into effects of spatial dispersion have used discrete lattices with discrete updating (cellular automata). Analysis of the Hawk-Dove game in this context has shown that for some payoffs that the ESS formulation would predict a pure Hawk strategy ESS the spatial model shows a dynamic coexistence of Hawks and Doves. This is consistent with findings that spatial heterogeneity increases species diversity [43].

In genetic models non-randomly dispersing populations may mean there is a greater chance of meeting relatives than otherwise. This can lead to altruistic behaviour or cooperation. In the standard ESS formulation any behaviour that helped an opponent with no gain for oneself is selected against. In the sexual population a concept similar to that of Hamilton's inclusive fitness needs to be incorporated. This has been approximated in work by Grafen [37], Aoki [4] and others [48]. Some of the results are analogous to that suggested by the concept of inclusive fitness while others seem not to be.

(iii) Infinite population size. This ensures that the probability of meeting a particular

strategy is proportional to the density of that strategy and that mutants are rare. The use of deterministic models of infinite population is justified on the basis that for large populations deviations from expected values are negligible in size and effect. This has yet to be fully explored for its appropriateness in ESS population model formulations. Some theory suggests that finite population size in deterministic models will result in a slightly greater tendency for pure as opposed to mixed strategies while some stochastic models suggests a centralizing tendency towards the ESS.

(iv) No variability. Some randomness or noise can be introduced by assuming a variable environment, mistakes in strategy expression or inexact strategy transmission. If environmental variability is modelled by small random changes in the payoff matrix then it results in greater average fitness for those strategies near the ESS and hence will tend to decrease diversity [47]. Inexact strategy transmission will result in greater diversity [117].

(v) Games are symmetrical, i.e. the two individuals are in exactly the same position with the same choice of strategies and payoffs. This is clearly invalid for biological populations since they display a rich variety of structure. Animal conflicts are often between a large animal and a smaller one or between a male and a female for example. An asymmetry which alters payoffs will often alter the choice of action. For example Riechert's study [83] of the funnel web spider showed that it was the size of spider (and the quality of the web site) which influenced the course of action and the result in contests over ownership of webs. However, it is also the case that an asymmetry which doesn't alter payoffs or the chance of success in escalated contests can still determine the choice of action. An example is the Hawk-Dove-Bourgeois game where there is a contest between an owner of a resource and an intruder. Hawk and Dove are the same as before whilst the third strategy Bourgeois is "if owner play Hawk; if intruder play Dove". The payoff matrix for this game is shown in table 4. From the payoff matrix it is clear that B is an ESS. Thus an asymmetry of ownership is used to settle the contest

	H	D	B
H	-1	2	0.5
D	0	1	0.5
B	-0.5	1.5	1.0

Table 4: The Hawk-Dove-Bourgeois game.

even though it doesn't affect the payoffs or success in fighting.

(vi) The basic model can be extended to study the evolution of learning processes. One can look for evolutionary stable learning rules, i.e. a learning rule such that a population adopting it cannot be invaded by mutants adopting a different learning rule. Under certain assumptions Harley [42] proved that an evolutionary stable learning rule is one which will take an initial population to the ESS within one generation (such a rule is called "a rule for ESS's"). Harley proposed the following as a realistic learning rule:

$$\begin{aligned} f_i(1) &= r_i / \sum_{i=1}^{n} r_i, \\ f_i(t) &= \frac{r_i + \sum_{\tau=1}^{t-1} m^{t-\tau-1} P_i(\tau)}{\sum_{i=1}^{n} (r_i + \sum_{\tau=1}^{t-1} m^{t-\tau-1} P_i(\tau))} \end{aligned} \tag{23}$$

where $f_i(t)$ is the probability of playing action i on trial t, $P_i(t)$ is the payoff received on trial t for action i and there are n possible actions or strategies. There are two learning rule parameters r_i the residual preference for action i and m the discount memory factor, $0 < m < 1$. The closer m is to one then the more attention is payed to earlier payoffs. This learning rule will be returned to in section 6.10.

5.3 Coevolution.

5.3.1 The Red Queen's Hypothesis.

In an influential paper in the early seventies Van Valen [112] discovered an important phenomenon in the fossil record. This was that the extinction rate of groups is approximately constant. This was derived from empirical data on thousands of extinct and living species from a very wide-ranging variety of different genera and families. A log-log plot of the proportion of the original sample that survives for a given time will have a linear relationship for a constant probability of extinction. Considering sampling error, which may be large when considering the fossil record, there is a clear linear relationship for many taxa. Van Valen recast this discovery in an ecological form: "the effective environment of the members of any homogeneous group of organisms deteriorates at a stochastically constant rate". Environment here means the physical and biotic environment, i.e. its physical surrounding and surrounding species. A homogeneous group is defined in terms of ecology and factors impinging on the regulation of population density (as such it is somewhat ambiguous and will depend on the circumstances).

That the probability of extinction of a taxon is independent of age suggests a randomly acting process. However the probability of extinction must be deterministically linked to the ecology of the taxon in question. Van Valen suggested the following link.

Consider a fixed amount of a limiting resource. If one species evolves an adaptation which increases its resource holding power, or resource gaining power, then this must involve a decrease to another species. In this sense the resource is like a fixed-volume mobile landscape with peaks and troughs. The height of the landscape at a particular point indicates the relative success of the species at that point compared to others and the fitness of a species is assumed to be related to this height. Adaptations in one species thus adversely effect another and cause a decrease in their fitness, ϕ. On average a species will generate a response which equals the deterioration caused by

other species. This reciprocity can mean that evolution will continue indefinitely. This gives it its name, "The Red Queen's Hypothesis". So called because the Red Queen in Through the Looking Glass [13] says to Alice that you have to keep running just to keep still, i.e. it is necessary to continually evolve yourself just in order to keep up with other species. However, at any particular moment this response to other species' adaptations will vary. Species will differ in the threshold of ϕ, or its components, to which they cannot respond since response to these stresses will often mean paying a cost elsewhere. If they cannot respond they might then go extinct. Since the extinction rate is constant this implies that the stresses on species are sufficiently varied that they affect most species similarly over long periods of time. The conclusion from the model is thus that species will continually evolve at a constant rate even when there are no physical changes in the environment.

While the Red Queen's Hypothesis explains the empirical data it is not the only explanation. The hypothesis needs to be tested before we can be sure it is the correct one. Nevertheless it has generated considerable theoretical interest and authors have tried to explore its consequences. The two main consequences of the theory are the fact that evolution is perpetual and constant even when there is no change in the physical environment and that ecological dynamics are crucial in determining the path of evolution.

5.3.2 The Red Queen and arms races.

The Red Queen describes how adaptations in one species might change the selection pressure on another species. This might then induce counter adaptations in the other species. If this happens reciprocally, Red Queen evolution ensues which was described by Dawkins and Krebs [19] as analogous to an "arms race" An example of such a phenomenon might be the adaptations for predator avoidance in a prey species and adaptations for catching prey in predators. These adaptations have an assumed cost

which is that less energy is spent on other adaptations such as reproduction. Hence, in the absence of a changing environment, a species will evolve to an optimum which balances competing selection pressures. However, the continual evolution of a predator (prey) ensures the continual evolution of anti-predator (prey-catching) adaptations. One such example may possibly be the trend for larger brain sizes in carnivorous and herbivorous mammals from the early Tertiary [35] to the present. As the brains of herbivorous mammals grew larger, presumably under intense selection pressure from carnivores, so the brains of predators increased [35]. Another example may be the coevolution of sabre-toothed tigers and their prey.

Arms races may be unequal which could end in one side winning. This might occur if there is unequal selection pressures, Dawkins' "life-dinner" principle, on the two species. The "life-dinner" principle comes from a predator-prey situation where the pressure on the prey to develop adaptations is more intense than on the predator. This is because a prey that fails to avoid predation will be eaten, while a predator that fails to catch a prey will only lose a dinner. If the selection pressure is very unequal then it is possible that the prey develops adaptations that will enable it to completely avoid predation. This would then result in specialised predators only hunting a few species of prey. Generalist predators, however, could evolve to exploit the "rare-predator" principle. This is when the prey only meets a predator rarely so that the selective pressure to evolve anti-predator adaptations will be small.

Species competing for similar resources will be unlikely to enter prolonged arms races as it is more probable that selection will lead to avoidance of competition through niche separation. This has been analysed by various authors, e.g. [52]. An example of an intra-specific arms races may be the competition among males for females. Larger males may tend to win contests so there is an evolutionary pressure for larger and larger males. Races of this kind may help to explain evolutionary trends for larger size (Cope's Rule).

This theory has been disputed by some authors (e.g. [1]) . They argue that the

selective pressure on species to evolve certain characteristics is a function of the detailed interactions determining population dynamics and fitness.

5.3.3 Coevolution of lag-loads.

In order to investigate evolutionary rates and in particular the efficacy, or not, of Red Queen evolution Maynard Smith [62] proposed what he called the lag-load model. This is a multi-species model which explicitly addresses the fact that the evolution of one species will be to the detriment of another.

For each species i there is a perfect phenotype whose fitness is defined to be $\hat{W}_i$. This phenotype is the fittest possible in the *contemporary* environment, i.e. it incorporates all possible favourable mutations, whether they have occurred or not, while all the other species do not change. Hence it is the fitness of species i when all other species are prevented from changing and species i has stopped evolving. Let the present species fitness be $\bar{W}_i$. The evolutionary lag, or lag-load, is then defined as :

$$L_i = \frac{\hat{W}_i - \bar{W}_i}{\hat{W}_i}. \tag{24}$$

L_i is thus the relative fitness deficiency from the optimum in species i. Maynard Smith then assumed that the rate of evolution of the lag-load is linearly dependent on L_i. Hence the further a species is from its optimum the greater the selective pressure on it to change. However a species' evolution is dependent on not only its own lag-load but also the evolution of species within the same ecology. Let the coefficient β_{ij} be the effect of one unit change in the lag-load of species j on the lag-load of species i. For s species we now have s simultaneous differential equations given by:

$$\frac{\partial L_i}{\partial t} = k_i(\sum_j \beta_{ij} L_j - L_i), \tag{25}$$

where k_i is the rate of change of L_i per unit L_i and $\beta_{ii} = 0$. Normally $\beta_{ij} > 0$ as a decrease in the lag-load of species j will lead to an increase of the lag-load in species i.

The Red Queen's hypothesis can now be restated in terms of lag-load: In the s-dimensional lag-load space of the s species, there will be a stable, internal equilibrium point. This means that (i) all lag-loads, L_i, are positive, (ii) all the values $\partial L_i/\partial t$ will be zero and (iii) any perturbations from this point will result in dynamics bringing the system back to it, [84].

As the basic model stands Maynard Smith showed that it cannot predict the Red Queen hypothesis, only stasis. However in a further paper, [103], speciation and extinction rates are incorporated into the model which led to a variety of conclusions: stationarity, Red Queen evolution or extinction depending on parameter values. This is looked at further in section 5.3.5 where the results of this model are compared to others.

There are various criticisms of the lag-load model. (i) Firstly lag-load is something that could never be measured in real systems, indeed it is doubtful that it could even be measured in the laboratory. (ii) Lag load is a scalar quantity whereas evolutionary pressures have a direction to them. Lag-load should be a vector indicating improvement against a particular species. This can bring out inconsistencies, such as the fact that lag-loads do not unambiguously determine the sign of the coefficients β_{ij}, i.e. a similar change in the lag-load of two species could have opposite effects on the lag-load of another species. (iii) The model pays little attention to phenotypic properties which are the actual characteristics that determine success and selection pressures. Thus crucially important factors such as constraints are ignored in the model. Strategies are effectively unbounded.

5.3.4 Coevolutionary games.

The modelling of coevolution as a frequency-dependent evolutionary game is appropriate as it is the strategies of one's coevolutionary partners which are the driving force for one's own evolutionary change. This is further improved upon by the inclusion of density-

dependent effects since evolution must take place in an ecology where population size has an effect on the dynamics.

In the following formulation, due to Brown and Vincent [11], gene frequencies are not modelled specifically but instead the focus of interest is the phenotype, referred to as the strategy. Not including genetics means various evolutionary constraints imposed by specific genetic mechanisms, such as sexual reproduction in diploid populations, are not modelled. However, it has been shown that despite these constraints the population will usually evolve to as close to the ESS as possible, [65]. The fitness of an individual, G, using a scalar strategy u is written as:

$$G(u, \mathbf{u}, \mathbf{p}, N), \tag{26}$$

where $\mathbf{u} = (u_1, u_2, .., u_r)$ is the vector of all current strategies in the population and $\mathbf{p} = (p_1, p_2, .., p_r)$ is the vector of frequencies of each strategy and N is the population size. The fitness of u is frequency-dependent through $\mathbf{p}$ and density-dependent through N. The mean fitness of the population as a whole, $\bar{G}$, is given by:

$$\bar{G} = \sum_i p_i G(u_i, \mathbf{u}, \mathbf{p}, N), \tag{27}$$

and so the dynamics are:

$$N_{t+1} = N_t \bar{G}. \tag{28}$$

In order for a single strategy, say u_1, to be an ESS then it must be the strategy that maximises individual fitness as given by equation (26). If not, then the strategy could be invaded. If the strategy set is unconstrained then a necessary condition for u_1 to be an ESS is:

$$\frac{\partial G(u, \mathbf{u}^*, \mathbf{p}^*, N^*)}{\partial u} \Big|_{u=u_1} = 0, \tag{29}$$

where the vector $\mathbf{u}^*$ has every component u_1, $\mathbf{p}^*$ has $p_1 = 1$ and all other entries zero and N^* is the equilibrium population size of a population composed solely of individuals u_1. From equations (27) and (28) we have the additional condition:

$$G(u_1, \mathbf{u}^*, \mathbf{p}^*, N^*) = 1. \tag{30}$$

The two equations (29) and (30) are used to solve for the two unknowns, u_1 and N^*. The stability of the solution also needs to be checked, if possible by looking at the second derivative of G with respect to u and a linear stability analysis of equation (30). Condition (29) can be thought of as the evolutionary process governing the change in strategy frequencies and condition (30) the ecological process governing the population dynamics. But these two processes are not independent and cannot be analysed separately.

For a coalition of strategies, $\mathbf{u^0} = (u_1, u_2, ..., u_s)$ where $s \leq r$, the conditions for the ESS are very similar. If the strategy set is unconstrained then equation (29) and equation (30) must be true evaluated for each element of the vector $\mathbf{u^0}$. Hence there are $2s + 1$ equations (since there is also $u_1 + u_2 + .. + u_s = 1$) to solve for s ESS coalition strategies, s coalition strategy frequencies and the equilibrium population size N^*. Hence an ESS can be composed of more than one strategy and coevolution can thus occur.

This formulation for finding the ESS is very general as long as the fitness generating function, G, can actually be written down. However it suffers the drawback of having to rely on equilibrium population dynamics at the ESS. This will not always be true and there are examples given in chapter 6 where an ESS has chaotic dynamics. A formulation to deal with complex dynamics is outlined in chapter 6. Vincent and Brown [113] also propose additions to analyse non-equilibrium dynamics.

With the function G in mind evolution can be visualised as occurring on a frequency-dependent adaptive landscape. This landscape is constructed by considering the individual's fitness as a function of its strategy. The rate of evolution is then proportional to the slope of the landscape. As evolution occurs the landscape will change due to the density- and frequency-dependence. A species will then always evolve with respect to the current landscape. The partial derivative of the function G with respect to an individual's strategy is effectively the selection pressure acting on that strategy. Hence

assuming the rate of evolution to be proportional to the selective pressure is reasonable. However, the rate of evolution is also dependent on the arrival of new mutations which is essentially a stochastic process. Therefore, assuming this relationship between selective pressure and evolutionary rate implies that the arrival of new mutations is independent of the phenotypic trait being analysed. This might not always be the case.

Metz, Nisbet and Geritz [67] (in work done parallel to the work outlined in chapter 6) propose a measure for defining fitness similar to the invasion exponent introduced in section 6.2. This is defined in a similar way to a Liapunov exponent and as such is applicable to any dynamical situation, not just the equilibrium case. This measure is the long-term growth rate of the population. In this sense it is analogous to fitness. Any bounded population will thus have zero fitness as ,over the long term, the population is neither decreasing or increasing.

To analyse evolution in ecologies they look at the initial growth rate of a small invading population in the environment set by the resident genotype. This growth rate can be positive or negative. A positive growth rate indicates an invasion can occur and evolution will tend to move the population in the direction of the mutant invaders. When any possible mutation has a negative growth rate then the resident genotype is an ESS. This was applied to a well-known single species population model. They indicated the direction of evolution and noted that the presence of constraints is crucial in determining where evolution will stop. This work looks ahead to the work outlined in chapter 6.

5.3.5 Constraints.

A main result from the work by Rosenzweig, Brown and Vincent [84] (which uses the formulation outlined in section 5.3.4) is that ESS's will dominate when constraints or interdependencies are put on phenotypic characters. Red Queen evolution depends on the existence of limitless, unbounded traits the extreme of which will make the

best phenotype (e.g. the biggest or most quickly growing). Since it is probable that limitless traits will be the exception (if possible at all) then ESS's and stasis will be the predominant feature of evolution. This will be interspersed with periods of fast evolution when interdependencies (or constraints) are broken and traits become more independent. However, Maynard Smith's original lag-load model predicted stasis even though the strategies, or phenotypic traits, are unbounded. This was resolved by Rosenzweig, Brown and Vincent [84] who assumed that given the lag-load of $s-1$ species then the lag-load of the last species is completely determined. Thus the dimension of the system is reduced by one. This is equivalent to assuming there is a fixed amount of resource available and perfection would be to have it all. The original formulation of the Red Queen by Van Valen also assumed this. With this addition the lag-load model again predicts the Red Queen.

Stenseth and Maynard Smith [103] made additions to the early lag-load model by adding speciation and extinction rates. This then predicted a variety of results including stasis and continual evolution depending on parameter values. However, once Rosenzweig et. al.'s addition is included then this model always predicts a Red Queen. Thus this earlier work on lag-loads is consistent with the conclusion that it is the presence or absence of constraints that is crucial in determining how evolution will occur.

In [84] the dynamics in the phenotype space are given by the partial derivative of the fitness function G with respect to the phenotypic trait of interest. An alternative approach is to introduce new equations describing the population dynamics of mutant strategies. This new mutant will compete with the original strategy through *interaction terms*, see section 6.1.2. Depending on how these interaction terms are specified then density and frequency-dependent selection as well as interactions between species can all be analysed together. As new mutations arise the number of equations can get impossibly unwieldy. Often though it is only necessary to deal with one mutation at a time. For Lotka-Volterra predator-prey systems it is possible to show that if a rare

mutation can spread then it will always take over and displace the original strategy. Thus evolution has occurred in the phenotype space and now mutants of the new species can be looked at. If a rare mutant cannot spread then it dies out and no evolution will occur. Hence for two coevolving species (e.g. predator and prey) the maximum dimension of the system is four.

Marrow et. al. [59] studied such a predator-prey Lotka-Volterra system. They introduced constraints by specifying how the phenotype strategies, in this case the body size of the predator and prey, influence the ecological dynamics. They assumed that the the predator-prey interaction term in the predator's equation is a bell shaped function of the body size of prey and predator. Hence an intermediate size of predator and prey will maximise the benefit of the prey to the predator. For the prey this is the opposite. This effective constraint set produces interesting results. For different parameter values ESSs and evolutionary saddle points are observed. Coevolution of the predator and prey can tend to an ESS or perpetual evolution in a Red Queen fashion.

It is the nature of the constraint set in this case that produces very interesting results showing that evolution can continue indefinitely or tend to an ESS. However, one drawback of the work was the continued assumption of equilibrium dynamics. In chapter 6 a new formulation is proposed which relaxes this assumption.

6 Evolutionarily stable attractors, the invasion exponent and phenotype dynamics.

6.1 Preliminary framework.

We start by considering a simple example with the aim of illuminating the general mathematical discussion that follows.

The example is a resource-predator-prey system with the following dynamics

$$\begin{aligned} \frac{x_1'}{x_1} &= (1+b_1)\exp\left(-\alpha\frac{x_1}{x_3} - c_1\frac{x_2}{1+d_1x_1}\right) \\ \frac{x_2'}{x_2} &= (1-d_2)\exp\left(c_2x_1\right) \\ \frac{x_3'}{x_3} &= (1+b_3)\exp\left(-\frac{x_3}{k} - c_3x_1\right) \end{aligned} \tag{31}$$

Here x_1, x_2 and x_3 denote respectively the population size of prey, predators and resource and the primed variables represent the corresponding numbers in the next period. The biological interpretation of the various terms in the equation is outlined in table 5. We do not attempt to defend this model on grounds of realism. It is just used here

Term in equation	Interpretation
$1+b_1$	prey's unconstrained birth rate
$-\alpha\frac{x_1}{x_3}$	decrease in fitness of prey due to resource limitation
$-c_1\frac{x_2}{1+d_1x_1}$	decrease in fitness of prey due to predation
$1-d_2$	predator's unconstrained death rate
c_2x_1	predator's fitness increase due to feeding
$1+b_3$	resource's unconstrained reproductive rate
$-\frac{x_3}{k}$	resource's carrying capacity limitation
$-c_3x_1$	resource's fitness decrease due to feeding by prey

Table 5: An explanation of the construction of equation 31.

as an illustrative example. As the parameters are changed, it displays a wide range of dynamical behaviour including periodic, quasi-periodic and chaotic attractors. For example, it is chaotic when the parameters are as in table 6. The attractor for the first set of these parameter values is shown in figure 15.

α	b_1	b_3	c_1	c_2	c_3	d_1	d_2	k
0.3	1.1	1.8	0.0025	0.0015	0.005	0.01	0.05	1000
0.5	1.1	2.8	0.001	0.0005	0.005	0.01	0.05	1000

Table 6: Parameter values for equation (31) giving chaotic dynamics.

In this example we distinguish three groups of species: resource, predator and prey. In the pure system there is one species in each group. However, we want to consider the effect of adding mutant resources, predators and prey. Therefore, we will allow for mutations within each group and thus consider systems with more than one species in each group.

6.1.1 Phenotypes

At this point it is important to discuss the relation between phenotype and dynamics. We regard the phenotype of a species group i as being described by a vector p_i of real numbers. Thus each aspect of the phenotype is capable of continuous variation. We let P_i denote the set of possible values that p_i can take. By p we denote the vector $(p_1, \ldots, p_s)$ which encodes all the phenotypes of our pure system, $s = 3$ for equation (31). One can think of p as representing the ecosystem's phenotype. The set of values which p can take is denoted by P.

The dynamics and evolution of a system such as that described by equation (31) are determined by the parameters that occur in the equations. Within our framework it is useful to consider a subset of the parameters that occur in the dynamical equations as being a part of a phenotype. When the dimension of the phenotype is higher than

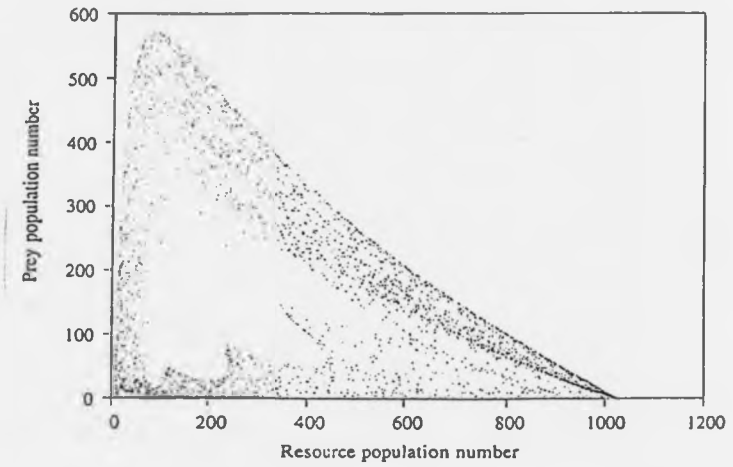


Figure 15: A 2-dimensional projection of the attractor for equation (31) with the first set of parameters as in table 6.

this subset there is some redundancy as other aspects of the phenotype cannot express themselves. Suppose that some character π of the phenotype has no influence upon the parameters of the equation. Then the behaviour of the system is independent of π. Thus, we can get evolutionary drift with respect to π and the system cannot be evolutionarily stable for this trivial reason. A similar remark holds if the dimension of p is greater than the number of parameters. Then we expect subsets of positive dimension in P to give rise to the same parameter values. Evolutionary drift can occur along these. These trivial obstructions to evolutionary stability should be removed by restricting p to only those characters that play a definite role in setting the parameters. One way to do this is to use the parameters and interactions to define the phenotype. Hence one takes the parameters that depend on p_i to be the vector p_i. In this case, there is an exact balance in the equation $s(p) = 0$ for evolutionary stability (see section 6.5). Thus, generically, we will obtain isolated points p as solutions and hence as ESA values.

There are several parameters in equation (31). Of these b_1 is a function of the

phenotype of species one, $b_1 = b_1(p_1)$ and $b_3 = b_3(p_3)$, $d_i = d_i(p_i)$ and $k = k(p_3)$. The other parameters are a function of more than one phenotype. For example c_1 relates the decrease in fitness in the prey due to predation. Evolution of the prey phenotype may act to decrease this while evolution of the predator phenotype will have the opposite effect. Since the parameters are functions of the phenotypes they are not independent and there may be constraints on the values that they can take. The phenotypic constraint manifold P is the set of all possible values of the vector of parameters. Because of the constraints the dimension of this is often less than the number of parameters.

These constraints on p are represented by the structure of P and this is why we call P the *phenotypic constraint manifold.*

There are two sorts of constraints. The first concerns constraints internal to P_i. These may occur because two parameters in the equation for the ith species group, e.g. $b_1 = b_1(p_1)$ and $c_1 = c_1(p_1)$, are functions of a single component of p_i. Then locally b_1 will be a function of c_1. We will see that the existence of these internal constraints is very important for the existence of ESAs[1] and is also an important natural property of real ecosystems.

The other constraints are cross-group constraints where a parameter in the equation for the ith species group depends upon a phenotype of another species group p_j with $j \neq i$. This is the case, for example, for the parameter c_1 in equation (31) which depends upon the phenotype of both the prey and the predator. However, it is usually the case that these parameters are part of the interaction terms that we are going to define below. It then follows that their dependence upon the p_j for which $j \neq i$ is not relevant to the question of evolutionary stability.

Our approach allows for the simultaneous coevolution of all the species phenotypes,

[1] This is relevant to the debate on the Red Queen Hypothesis. We would suggest that this hypothesis is essentially equivalent to the non-existence of ESAs. Therefore, in view of our results, its veracity is largely equivalent to the absence of constraints that enhance the possibility of an ESA. The debate should therefore concentrate upon understanding the structure of these constraints in real ecosystems.

and all our general ideas are developed in this context. However, in the examples that we treat, we consider mutations of only a single phenotypic attribute that is linked to parameters that are internally constrained.

We can write equation (31) in the abstract form

$$x_i' = X_i(x, p) \qquad (i = 1, 2, 3) \tag{32}$$

showing the dependence of the dynamics upon the phenotype parameter.

6.1.2 Interactions.

Let us now return to our example and suppose that a mutant prey species is present, consisting of y_1 individuals. Then to equation (31) we must add an equation for y_1 similar to that for x_1 but with a different value p_1' of the phenotypic parameter p_1.

The original and mutant prey will compete for the same available resource x_3. The resource-limiting term in equation (31) is $\exp(-\alpha x_1/x_3)$ which expresses the effect of the magnitude of the number of prey per unit of resource. When the mutant population y_1 is present, the number of prey per unit of resource is instead $(x_1 + y_1)/x_3$. Thus, the resource-limiting term should be replaced by $\exp(-\alpha(x_1 + y_1)/x_3)$. In fact, in general, it should be of the form $\exp(-e_{11}/e_{13})$, where e_{1j} is the total number of individuals belonging to a species in group j. When there are cross-group constraints these terms will also include a parameter reflecting the phenotype of species group j.

It is necessary for our development to single out such terms e_{ij} which represent those average quantities of species group j which enter the equations for group i. We call them interaction terms and incorporate them in our formulation. Thus, we rewrite equation (31) for the dynamics with no mutants as

$$\begin{aligned} \frac{x_1'}{x_1} &= (1 + b_1)\exp\left(-\alpha\frac{e_{11}}{e_{13}} - \frac{e_{12}}{1 + d_1 e_{11}}\right) \\ \frac{x_2'}{x_2} &= (1 - d_2)\exp(e_{21}) \\ \frac{x_3'}{x_3} &= (1 + b_3)\exp\left(-\frac{e_{33}}{k} - e_{31}\right) \end{aligned} \tag{33}$$

where we have only allowed the parameters c_i to have the possibility of cross-group constraints. It is these terms which include the parameters c_i that we shall focus on in order to clarify the nature of the of the interaction parameters, e_{ij}. When there are no mutants present and no constraints then $e_{ij} = c_i x_j$. When there is a single mutant species y_j present then $e_{ij} = c_i(x_j + y_j)$. In general

$$e_{ij} = c_i \int \xi_j(p_j) dp_j$$

where ξ_j is the phenotypic distribution of species group j. In the pure case the distribution ξ_j has only one phenotype present and is thus a delta function, δ_{p_j} centered at p_j. Thus $\xi_j = x_j \delta_{p_j}$ and and $e_{ij} = c_i \int x_j \delta_{p_j} dp_j = c_i x_j$. If there are cross-group constraints then the parameter c_i is a function of the phenotypes p_i and p_j. Hence

$$\begin{aligned} e_{ij} &= \int c_i(p_i, p_j) \xi_j(p_j) dp_j \\ &= \text{species group j weighted mean of } c_i \end{aligned}$$

The interaction terms are a function of the distribution of species group j, which reflects the phenotypic distribution of j, and the phenotype i. The matrix e has the entries e_{ij}, thus $e = (e_{ij}(\xi_j, p_i))$. We can now write the equations with no mutants in the abstract form

$$\begin{aligned} x_i' &= X_i(x, e_i, p) \qquad (i = 1, 2, 3) \\ e_i &= (e_{ij}(x_j \delta_{p_j}, p_i))_{j=1,2,3} \end{aligned} \tag{34}$$

We call this the *pure dynamics* of the system. This formulation has been derived because all our definitions and constructions follow from the form of the pure equations and the interactions.

Now introduce a small mutant population $y = (y_1, \ldots, y_s)$ with phenotype $p' = (p'_1, \ldots, p'_s)$. The new phenotypic distribution is given by

$$x\delta_p + y\delta_{p'} = (x_1 \delta_{p_1} + y_1 \delta_{p'_1}, \ldots, x_s \delta_{p_s} + y_s \delta_{p'_s})$$

and therefore the equations for the new system are given by

$$\begin{aligned} x_i' &= X_i(x, e_i, p) \quad (i = 1,2,3) \\ y_i' &= X_i(y, e_i', p') \quad (i \in M) \\ e_i &= (e_{ij}(x_j\delta_{p_j} + y_j\delta_{p_j'}, p_i))_{j=1,2,3} \\ e_i' &= (e_{ij}(x_j\delta_{p_j} + y_j\delta_{p_j'}, p_i'))_{j=1,2,3} \end{aligned} \tag{35}$$

where M is the set of i such that $p_i' \neq p_i$. We specify the set M because we do not want to include equations for mutants which do not differ from the original species.

We call this system the *p'-mutated equation* and note that it is completely determined by the pure equations and the interactions.

In the above example x_i is a scalar and so e_{ij} is just a number. In general, though, x_i and y_i can be vector quantities reflecting age-structure for example. In this case e_{ij} will also be a vector. The species groups interact here through abundance but generally may interact through average properties such as mean strategy or biomass etc..

6.2 The invasion exponent.

This section contains an important characterisation of evolutionary stability which provides the mathematical and computational tool for analysing examples. The full-blown definition for general attractors requires the use of less well-known mathematical tools (such as invariant measures), and is given in [80]. Here we illustrate the case where the attractor Λ of our pure system is either a fixed point or a periodic orbit.

We consider when an attractor Λ of the pure dynamics with phenotype p is stable to mutation from phenotype p to p'. The mutated dynamics are given by equation (35).

Let us denote by y the vector (y_i) where the index i runs over those i such that $p_i' \neq p_i$. Equation (34) for the pure dynamics defines a mapping $x \mapsto x' = f(x)$ and equation (35) for the p'-mutated dynamics defines a mapping $(x, y) \mapsto (x', y') = g(x, y) = (g_1(x, y), g_2(x, y))$ where $g_1(x, 0) \equiv f(x)$ and $g_2(x, 0) \equiv 0$.

We are interested in determining whether or not there are small invading populations y which will grow under iteration of the mapping g. In [80] it is shown that in the generic case such an invasion will occur if it occurs in the system where the equation for y is replaced by its linearisation. This defines a mapping, V:

$$(x, y) \xmapsto{V} (f(x), T_x \cdot y) \tag{36}$$

where $T_x = d_y g(x, 0)$ is the partial derivative of g with respect to y evaluated at $(x, 0)$. Below we show how to calculate this for some important examples.

Now let us consider the simplest case where the attractor of f is just an attracting fixed point, $\Lambda = \{x_0\}$. Then the action of V on the x component is trivial and the map reduces to

$$y \mapsto T_{x_0} \cdot y$$

where T_{x_0} is the linear map $d_y g(x_0, 0)$. Thus we can apply standard theory to deduce that the behaviour is determined by the eigenvalues of T_{x_0}. If all the eigenvalues of T_{x_0} are inside the unit circle in the complex numbers $\mathbf{C}$ then the system is stable. We express this in the following way. Let λ denote the eigenvalue with largest modulus and let $\chi = \log|\lambda|$. Then the system is stable to invasion by a small population y with phenotype p' if $\chi < 0$ or equivalently, $|\lambda| < 1$.

A similar result holds if our attractor is a periodic orbit $x_0, \ldots, x_{q-1}$ of period q. Then we let λ be the eigenvalue with largest modulus of the product $T_{x_{q-1}} \cdots T_{x_0}$ corresponding to iterating once around the orbit. Again, the system is stable to small invasions if $\chi = \log|\lambda| < 0$.

To the attractor Λ of f and a mutated phenotype p' we are going to associate a number $\vartheta_p(\Lambda, p')$ which characterises the stability of Λ with respect to the mutation p'. We call this number the *invasion exponent*. This measures the rate of growth of a small invading population with phenotype p'. A positive growth rate ϑ means that a small population with phenotype p' will be able to invade and either take over or coexist

with the original population. In our examples the principle of mutual exclusion holds which means that, after successful invasion, the invading population actually takes over and replaces the original population. Thus we observe evolution from p to p'. The magnitude of ϑ is related to the selective pressure and determines the speed at which the invasion initially takes place and hence the speed of the evolution. A negative ϑ implies that invasion by a small population with phenotype p' is impossible. In the simple cases above we would set

$$\vartheta_p(\Lambda, p') = \chi.$$

Then Λ is evolutionarily stable if for all $p' \neq p$ near p, $\vartheta_p(\Lambda, p') < 0$. Moreover, if $\vartheta_p(\Lambda, p') > 0$ for p' arbitrarily near p then it is evolutionarily unstable.

The action of the mapping V on y is linear. Thus, we may expect that on average, under iteration, the length $||y||$ of y grows or decays exponentially fast. The exponential growth rate associated with a generic choice of y and x in the attractor Λ thus provides us with our invasion exponent, i.e. χ is the growth rate of a typical y vector. This means that if $(x_n, y_n) = V^n(x, y)$, then $||y_n||$ grows like $\exp(n\chi)$ or more precisely that

$$\chi = \lim_{n\to\infty} \frac{1}{n} \log ||y_n||$$

This gives us a numerical method of evaluating the invasion exponent. When P is 1-dimensional it is useful to consider the functions $f_+(p) = \vartheta_p(p+\varepsilon)$ and $f_-(p) = \vartheta_p(p-\varepsilon)$. These respectively measure the selective advantage of $p+\varepsilon$ and $p-\varepsilon$ over p. If $f_+(p) > 0$ then a small population with phenotype $p' = p+\varepsilon$ will be able to invade a p population. If mutual exclusion holds, then we get evolution to increasing values of p. If $f_+(p) < 0$, the invading population will die out. If $f_-(p) > 0$ then a small $p' = p - \varepsilon$ population will be able to invade the p population and if $f_-(p) < 0$, it will not. For small $\varepsilon > 0$, the zeros of these two functions approximate an *evolutionary stable attractor*, an ESA, (see section 6.3) in the following way: if for $p < p_*$, $f_+(p) > 0$ and for $p > p_*$, $f_-(p) > 0$ then this indicates that Λ_* is an ESA. If the inequalities are the other way round, then we

may regard Λ_* as an *evolutionary repeller*. If P is an interval then the right-hand (resp. left-hand) end-point is a boundary ESA if $f_+ > 0$ (resp. $f_+ < 0$) near the end-point.

Since to detect an ESA it is only necessary to consider one of f_+ or f_- in later sections we will concentrate our attention upon f_+.

6.3 Evolutionarily stable attractors.

Now we suppose that equation (34) has an attractor Λ. It is well-known that the equations of population dynamics have a rich variety of attractors including stationary, periodic, quasi-periodic and chaotic ones. One advantage of our approach is that it applies to all these cases.

We now make the standing assumption that if a species has a zero population then equation (34) implies that it remains zero for all time. The only way it can become non-zero is by mutation. Then the set

$$\Lambda_0 = \{(x, y) : x \in \Lambda, y = 0\}$$

is an invariant set for the dynamics of the p'-mutated equation (35).

Definition 4 *The attractor Λ of the pure system (34) is said to be strongly evolutionarily stable if for all p' in P near p, Λ_0 is an attractor for the p'-mutated system (35).*

This means that for all p' in P near p, a small invading mutant population y will die out and the system will relax back to its pure state. For deep mathematical reasons, see [80], this turns out to be too strong a condition for chaotic attractors because generically there are often ways in which chaotic attractors can be invaded. However, the possible invasions have measure zero in some sense and are therefore not observed and irrelevant. Therefore, we use the slightly weaker definition of evolutionary stability given below. For non-chaotic attractors the strong evolutionary stability and the weaker form are equivalent. Experience shows that for practical purposes this is as effective a condition

as strong evolutionary stability. Moreover, we conjecture that in systems with a small amount of stochastic noise the two notions are equivalent.

The weaker notion of evolutionary stability means that the probability of very small invasions succeeding is very small and goes to zero with the size of the invasion. It is precisely defined as follows. Let U be a neighbourhood of Λ in the x-space. Let U^ϵ denote the set of all points (x, y) such that $x \in U$ and $||y|| < \epsilon$. Let $\Lambda^\epsilon_{p'}$ denote the set of all points $(x, y) \in U^\epsilon$ such that, if $(x_n, y_n) = g^n(x, y)$, $||y_n|| \to 0$ as $n \to \infty$.

Definition 5 *We say that Λ is evolutionarily stable to p' if the Lebesgue measure of the set of points in U^ϵ but not in $\Lambda^\epsilon_{p'}$ tends to 0 as $\epsilon \to 0$. We say that Λ is evolutionarily stable if it is evolutionarily stable to p' for all p' near p*

We call such an attractor an ESA. The associated phenotype p is called an ESA value. We say that Λ is *globally evolutionarily stable* if this stability holds not only for small perturbations p' of p in P, but also for all p' in P.

The precise form of the above conjecture on noise is as follows. If Λ is evolutionarily stable and there is some small amount of noise perturbing the system, then the probability that $\Lambda^\epsilon_{p'} = U^\epsilon$ tends to one as $\epsilon \to 0$. Thus, with noise, strong stability and this weaker notion are equivalent.

We distinguish interior ESAs from boundary ESAs. If p is an ESA value and the phenotypic constraint manifold is a smooth manifold near p then we say that the ESA is *interior*. If not then we say that it is a *boundary ESA*. We use this terminology because in our examples this occurs when P is an interval and one of the end-points of P gives an ESA.

The dynamics of these invasions is, of course, given by the p'-mutated dynamics. The correct mathematical definition of ϑ is given in reference [80]. The relation of $\vartheta_p(\Lambda, p')$ to ESAs is given by the following theorem.

Theorem 1 *1. If $\vartheta_p(\Lambda, p') < 0$ then Λ is evolutionarily stable to p'.*

2. *If $\vartheta_p(\Lambda, p') > 0$ then Λ is evolutionarily unstable to p'.*

3. *If $\vartheta_p(\Lambda, p') < 0$ for all $p' \neq p$ near p then Λ is evolutionarily stable. If there exists a sequence $p_i \to p$ such that $\vartheta_p(\Lambda, p_i) > 0$ then Λ is not evolutionarily stable.*

We illustrate these ideas by returning to our resource-predator-prey example, equation (33). Suppose that our phenotype parameter is $p = b_1$. Then only the prey are allowed to mutate. Let us assume that there are no constraints on p. Then:

Proposition 1 *There are no ESAs for the unconstrained system.*

The reason is obvious but non-trivial to prove because of the complexity of the attractors of (33) and the proof is given in section 6.6. A prey with a given phenotype b_1 can always out-compete one with a slightly lower b_1. If we plot f_+ then we find that it is always positive. Clearly, to get evolutionary stability we need to introduce constraints.

Indeed, if the phenotype parameters are constrained then (33) does have an evolutionarily stable attractor. For example, it is natural to suggest a trade-off between the prey growth parameter b_1 and its contact parameter c_1 with the predator. It is sufficient that c_1 increases with b_1. For definiteness, we assume that c_1 is proportional to b_1. At the same time, because the contact rate of prey with predators and predators with prey must be equal, let us keep the ratio of c_2 to c_1 fixed. We regard the other parameters α, c_3, d_i as fixed. Thus our phenotype space P is given by

$$P = \{(b_1, c_1, c_2) : b_1 \geq 0, b_1/c_1 = l, c_1/c_2 = m, \} \tag{37}$$

where $l = 440$ and $m = 1.7$ for the first set of parameter values and $l = 1100$ and $m = 2.0$ for the second set. The phenotype space is 1-dimensional and parametrised by $b_1 \geq 0$.

Proposition 2 *If the parameter values (except b_1 which is variable) are as in table 6, this constrained system has at least one evolutionarily stable attractor. This is chaotic.*

A plot of f_+ for equation (33) for parameter values corresponding to the two cases given in table 6 is shown in figure 16. In the first case there is a single ESA value $p_* \approx 0.59$ and the corresponding ESA is chaotic. In the second there are two ESA values separated by an evolutionary repeller. One is a fixed point and corresponds to $p_\dagger \approx 0.75$. The other, which corresponds to $p_\ddagger \approx 1.1$, is chaotic. Thus we see that even in relatively simple systems we should expect multiple ESAs. Moreover, this and our other examples show that evolutionary stability and chaotic dynamics are perfectly compatible, as is any other common form of dynamical behaviour.

6.4 Game theory in a dynamic environment.

Our approach allows us to consider many extensions of the usual game-theoretic formalism. We illustrate this by considering one example based upon the usual Hawk-Dove prototype. We will suppose that the species playing a mixture of these two strategies is a predator and that the resource gained in each contest is a proportion V_0 of the prey population x_1. Thus the payoff matrix is shown in table 7. The phenotype, p, is the

	H	D
H	$(V-C)/2$	V
D	0	$V/2$

Table 7: Payoff matrix for the Hawk-Dove predator game with prey where $V = V_0 x_1$.

amount of time spent playing the Hawk strategy. The main difference from the usual theory is that $V = V_0 x_1$ depends upon the number of prey x_1 and this is changing dynamically. If we denote the payoff matrix by $E = (E_{ij})$ and use the same notation

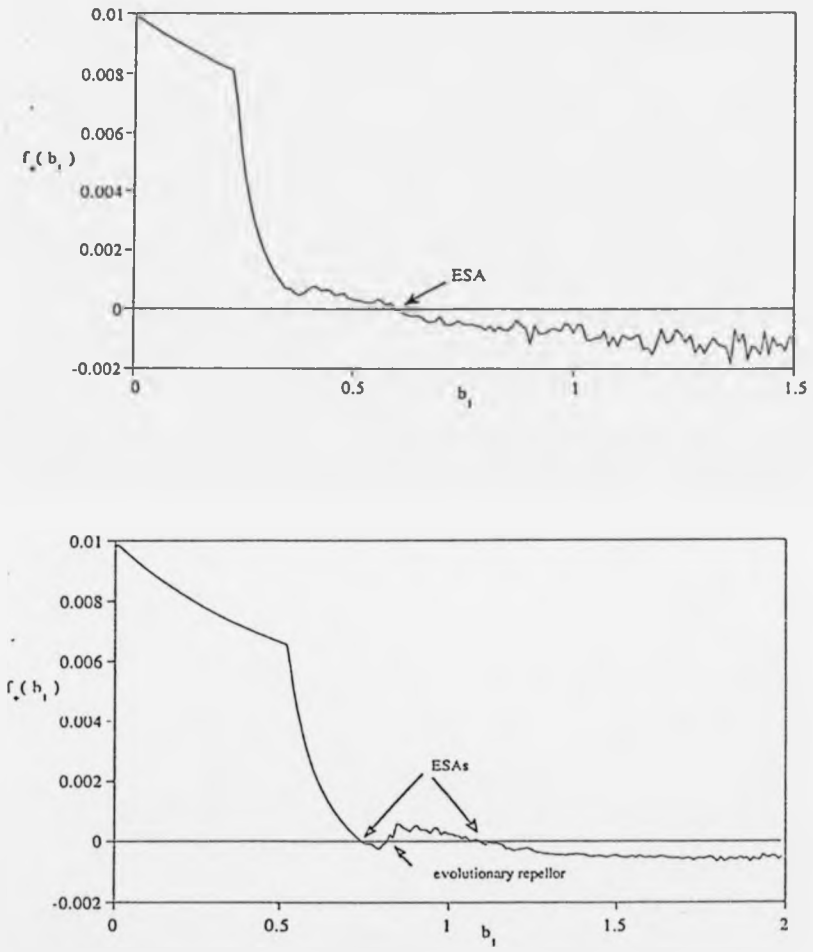


Figure 16: The graph of f_+ for equation (33) for parameter values corresponding to the two cases given in table 6. The jagged parts of the graph correspond to parameter values where the attractor is chaotic. The jaggedness is caused by the nonuniform convergence of time-series for the ergodic measures of chaotic attractors.

as above, then our pure equations are[2]

$$\begin{aligned} x_1' &= x_1(1+b)\exp\left(-\alpha\frac{x_1}{k}-\frac{c_1e_{12}}{1+d_1e_{11}}\right) \\ x_2' &= x_2e^{-\sigma}E(p,e_{22}) \end{aligned} \tag{39}$$

where $E(p,q)=\sum_{i,j}p_iq_j\exp(c_2E_{ij})$, $e_{11}(\xi_1)$ is the total prey population $\int\xi_1(p_1)\,dp_1$, $e_{12}(\xi_2)$ is the total predator population $\int\xi_2(p_2)\,dp_2$ and $e_{22}(\xi_2)$ is the mean predator population strategy

$$\int p_2\xi_2(p_2)\,dp_2/\int\xi_2(p_2)\,dp_2.$$

We have included the death rate $-\sigma$ in the equation for the predators so that, without the increase in fitness gained by winning prey, their numbers would naturally decline.

[2]One aspect of our treatment of linear evolutionary games is nonstandard. If the payoff to strategy i against strategy j is E_{ij}, then the contribution towards fitness is taken to be $\exp(\tau E_{ij})$ where τ (=c_1 in table (8)) is some constant largely set by the time-scale of the interactions. Then the mean fitness of an individual playing the strategy p in a population whose mean strategy is e is $E(p,e)=\sum_{i,j}p_ie_j\exp(\tau E_{ij})$. For discrete time dynamics this has a number of advantages while also giving the same ordinary differential equation as is usual for the continuous time dynamics. The advantages include (i) natural positivity, unlike E_{ij}, $\exp(E_{ij})$ cannot be negative; (ii) invariance of the dynamics under addition of a constant to a column of the payoff matrix E_{ij}; and (iii) sensible scaling with the unit of time τ. For example, in the classical Hawk-Dove game $E_{HH}=(V-C)/2$ which can be negative, and this cannot be overcome by adding a constant to the fitnesses as this changes the classical version of the discrete-time dynamics as formulated in [64]. Replacing the payoffs such as E_{HH} by terms of the form $\exp(cE_{HH})$ overcomes all these problems for mappings and preserves the continuous-time dynamics. The continuous-time dynamics are derived as follows: Assume that E_{ij} is the contribution to fitness over a small time period τ. Let $W(p,e)=\sum_{i,j}p_ie_jE_{ij}$. Then,

$$\begin{aligned} x(t+\tau) &= x(t)\sum_{i,j}p_ie_j\exp(\tau E_{ij}), \\ x(t+\tau) &= x(t)\sum_{i,j}p_ie_j+\tau x(t)\sum_{i,j}p_ie_jE_{ij}, \\ \frac{x(t+\tau)-x(t)}{\tau} &= x(t)\sum_{i,j}p_ie_jE_{ij}, \end{aligned} \tag{38}$$

and therefore in the limit $\tau\to 0$ we obtain $dx/dt=W(p,e)x(t)$. This is the same differential equation as that obtained in [108] and [118]. Finally, we stress that our preference for this formulation does not affect our conclusions in any important way.

It will be shown in section 6.7 that this system has an ESA at $p_* = \bar{V}/C$ where $\bar{V} = V_0\bar{x}_1$ and $\bar{x}_1$ is the time-averaged value of x_1. This fact is also clear from the graph of the function f_+ (see section 6.2) which was calculated numerically and is plotted in figure 17. This ESA corresponds to that for the classical linear game. However, it is also clear from this plot that there is a significant difference, because $p = 1$ is also a boundary ESA value.

The reason for this can be clearly seen from figure 18 where we plot the the observed value of $\bar{V}/C$ against the proportion of Hawk behaviour. Since this proportion is equivalent to p, we use p to denote it. Note that one of the values p_* where the graph crosses the diagonal corresponds to the ESA and the other p_r to the evolutionary repeller. From the graph one can easily see the following: if p is less than p_* then the proportion of Hawk behaviour is less than $\bar{V}/C$ and therefore there is a selective advantage to more hawkish behaviour. By similar reasoning, if p is greater than p_* and less than p_r then $p > \bar{V}/C$ and more dove-like behaviour is selected. But if p is greater than the evolutionary repeller then $p < \bar{V}/C$ and therefore more hawk-like behaviour is selected. This explains why $p = 1$ is a boundary ESA value.

b	k_1	c_1	c_2	d_1	V_0	C	σ
0.5	1000	0.01	0.001	0.005	1.0	75 *variable*	0.006

Table 8: Parameter values for for the Hawk-Dove system given by equation (39).

For low p the system has a fixed point as its attractor while for larger values it has a quasiperiodic attractor. In this example the ESA is a fixed point.

Later in section 6.9.3 we will see that this example has particularly interesting phenotype dynamics for both the corresponding mixed strategy and polymorphic population problems. In the first case it has a highly structured asymptotic distribution of strategies and, in the second, it displays complicated oscillatory behaviour.

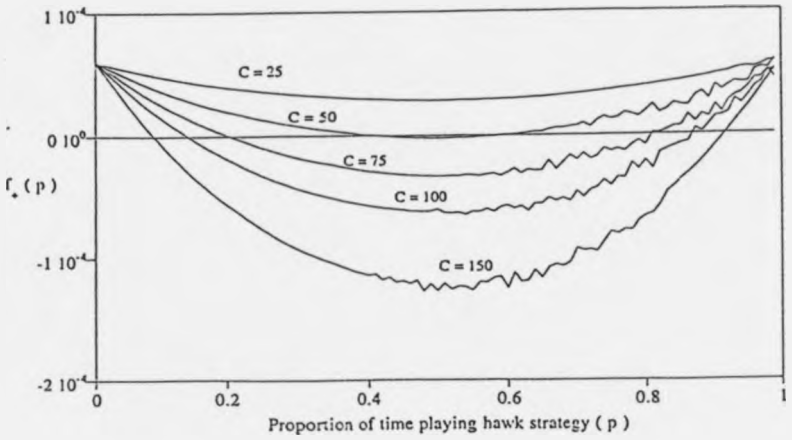


Figure 17: The graph of f_+ for the Hawk-Dove predator-prey system given by equation (39) for parameter values as given in in table 8 and for variable injury cost C.

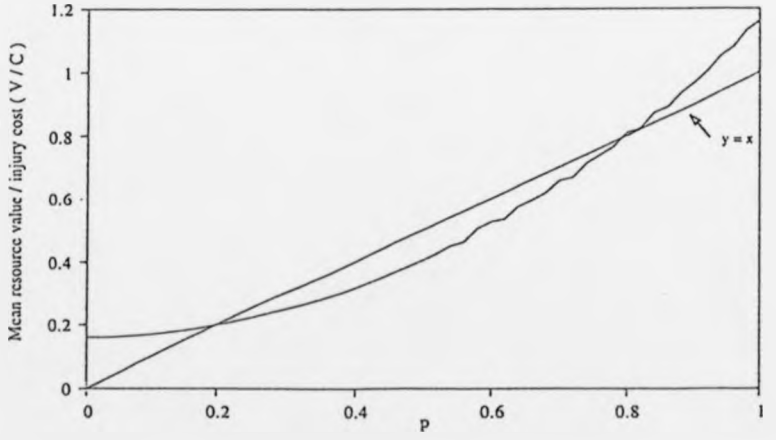


Figure 18: The graph of $\bar{V}/C$ against the proportion of Hawk behaviour, where the injury cost, C, is 75.

6.5 Differential selective pressure.

The equations for the mutant species are given by $y_i'/y_i = F_i(y, e, p')$ where $p_i' \neq p_i$. If there is only one mutant species then the invasion exponent, $\vartheta_p(p')$, is given by averaging $log F_i(0, e_i, p')$ over the attractor using the so-called natural measure, ν_*. In reference [80] it is shown that when there is more than mutant species then the growth rate of the ith component, $\vartheta_{p,i}(p')$, is given by looking at the partial derivative with respect to the ith component (using the assumption that if $y_i = 0$ then $y_i' = 0$). Furthermore this partial derivative only depends on p_i' and not on p' as all cross group terms enter through the interaction terms which are evaluated at $y = 0$.

Let $\theta_i(x, p_i') = log F_i(0, e_i, p_i')$. Then

$$\vartheta_{p,i}(p_i') = \int \theta_i(x, p_i')\nu(dx). \tag{40}$$

Expanding $\theta_i(x, p_i')$ using Taylor's Theorem:

$$\theta_i(x, p_i') = \theta_i(x, p_i) + d_{p_i'}\theta_i(x, p_i)\cdot(p_i' - p_i) + \frac{1}{2}d^2_{p_i'}\theta_i(x, p_i)\cdot(p_i' - p_i)^2 + \mathcal{O}((p_i' - p_i)^3). \tag{41}$$

Thus,

$$\begin{aligned}\int \theta_i(x, p_i')\,\nu(dx) \;=\; & \int \theta_i(x, p_i)\,\nu(dx) + \left(\int d_{p_i'}\theta_i(x, p_i)\,\nu(dx)\right)\cdot(p_i' - p_i) \\ & + \left(\frac{1}{2}\int d^2_{p_i'}\theta_i(x, p_i)\,\nu(dx)\right)\cdot(p_i' - p_i)^2 + \mathcal{O}((p_i' - p_i)^3). \end{aligned}\tag{42}$$

If each y_i is either 1-dimensional (and under certain conditions if y_i is a vector, [80]) then the first term in equation (42), $\int \theta_i(x, p_i)\,\nu(dx)$, is zero. This is because x is bounded away from 0 and infinity in the sense that there is a $k > 0$ such that for all $x \in \Lambda$, $k^{-1} < x_i < k$.

Definition 6 *We call $s_i(p) = \int d_{p_i'}\theta_i(x, p_i)\,\nu_*(dx)$ the differential selective pressure of the ith species group. The differential selective pressure of the ecology is the function $s(p) = (s_0(p), \ldots, s_s(p))$.*

Theorem 2 *If p is an interior point of P, p is an ESA value if, for all i, $s_i(p) = 0$ and $s_i'(p)$ is negative definite (or just negative in the 1-dimensional case). Conversely, if Λ is an ESA then $s_i(p) = 0$ for all i.*

If $s_i(p) \neq 0$ for some i then $\vartheta_{p,i}(p_i')$ would be positive for some i and so mutants could invade.

In the 1-dimensional case we can observe that the function $f_+(p)$, introduced in section 6.2, and $s(p)$ effectively determine each other. This is because, from equation (40) and equation (42):

$$\vartheta_p(p_i') = s(p) \cdot (p' - p) + \mathcal{O}((p_i' - p_i)^2).$$

Therefore, since $f_+(p) = \theta(x, p + \varepsilon)$,

$$f_+(p) = s(p) \cdot \varepsilon + \mathcal{O}(\varepsilon^2).$$

We now use the differential selection pressure to find ESAs analytically.

6.6 ESAs in the resource-predator-prey model.

We consider why propositions 1 and 2 are true. The equation for the mutant prey y is

$$y'/y = (1 + b_1') \exp\left(-\alpha \frac{x_1 + y}{x_3} - c_1 \frac{x_2}{1 + d_1(x_1 + y)}\right)$$

where b_1' is a function of p'. Thus,

$$\theta(x, p') = \theta_1(x, p') = \log(1 + b_1') - \alpha \frac{x_1}{x_3} - c_1 \frac{x_2}{1 + d_1 x_1}$$

In the unconstrained case,

$$d_{b_1'}\theta(x, p') = (1 + b_1')^{-1} > 0.$$

Thus, if p' differs from p only in having a larger b value, then the system is unstable to p'.

On the other hand, for the constrained system of proposition 1,

$$\theta(x,p') = \log(1+b_1') - \alpha\frac{x_1}{x_3} - \frac{b_1'}{l}\frac{x_2}{1+d_1x_1}.$$

where $l = 440$ for the first set of parameter values of table 6 and $l = 1100$ for the second set. Therefore,

$$d_{b_1'}\theta(x,p') = (1+b_1')^{-1} - \frac{x_2}{l(1+d_1x_1)}.$$

Let $\mu = l^{-1}\int x_2(1+d_1x_1)^{-1}\,\nu_*(dx)$. Then

$$s(b) = \int d_{b_1'}\theta(x,p')\,\nu_*(dx) = (1+b_1')^{-1} - \mu$$

and

$$s'(b) = d^2_{b_1'}\theta(x,p') = -(1+b_1')^{-2} < 0$$

hence $s(b) = 0$ if $b = \mu^{-1} - 1$. Since $0 < \mu \leq 1$, there is such a value. Moreover $s'(p) < 0$. Consequently, this value corresponds to an interior ESA for the constrained system. Note that this argument does not exclude the possibility of more than one ESA.

In figure 19 we show the graphs of ϑ_p for the two ESA values p for the second set of parameter values.

6.7 ESAs in the Hawk-Dove predator-prey system.

We now consider the Hawk-Dove predator game with dynamic prey as discussed above in section 6.4. For this example, $y'/y = \varphi(x,y,p')$ where $\varphi(x,p') = \varphi(x,0,p')$ is given by

$$\begin{aligned}\varphi(x,p') &= e^{-\sigma}\sum_{i,j} p_i'p_j\exp(cE_{ij}) \\ &= \exp\left(-\sigma + c\sum_{i,j} p_i'p_j\,E_{ij}\right) + \mathcal{O}(c^2).\end{aligned}$$

since $\sum_{i,j} p_i'p_j = 1$. Thus, ignoring terms which are of order c^2,

$$\theta(x,p') = -\sigma + c\sum_{i,j} p_i'p_j\,E_{ij}$$

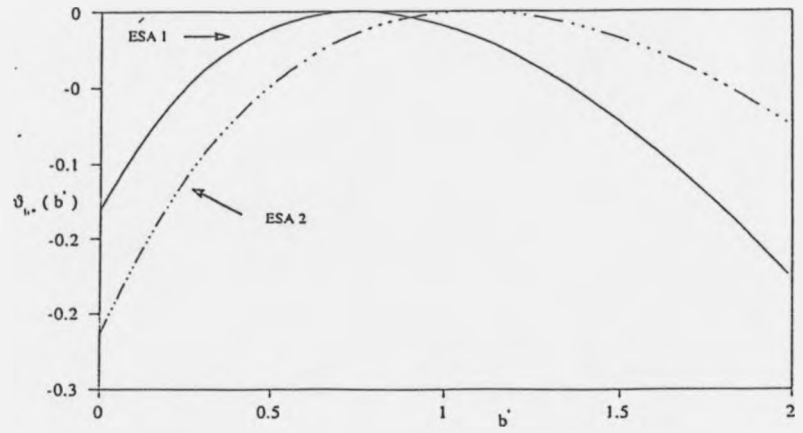


Figure 19: A plot of the function $\vartheta_{b_*}(b_1')$ for the two ESA values of the resource-predator-prey model corresponding to the second set of parameter values in table 6.

Let p be the proportion of time spent playing Hawk, then the proportion of time playing dove is $1-p$ and:

$$\theta(x,p') = -\sigma - \frac{C}{2}pp' + \frac{V}{2}(p'-p) + \frac{V}{2}$$

where $V = V_0 x_1$. Thus,

$$d_{p'}\theta(x,p') = -Cp/2 + V/c$$

and

$$\int d_{p'}\theta(x,p') = -Cp/2 + V_0\bar{x}_1/2$$

where $\bar{x}_1 = \int x_1\, \nu_*(dx)$. It follows that

$$s(p) = \frac{1}{2}(V_0\bar{x}_1 - Cp).$$

Thus $s(p) = 0$ if for the natural measure ν_* for the phenotype p,

$$p = \frac{V_0\bar{x}_1}{C}.$$

However, in this case, $d^2_{p'}\theta \equiv 0$ and therefore for this value of p

$$s'(p) = 0. \tag{43}$$

6.8 The generic structure of $\vartheta_p(p')$.

The degeneracy expressed by equation (43) is a result of the linear dependence of the payoff upon p'. For this linear game theory, it is the reason why it is necessary to use the second-order condition in the ESS criterion ([66]). For general nonlinear games or dynamical systems such second-order conditions are not needed. This is illustrated by the previous example. For such generic cases, we observe that that if p is an ESA value then the invasion exponent function $\vartheta_p(p')$ has a quadratic maximum at p where it takes the value 0. This is the situation shown in figure 19.

6.9 Phenotype dynamics.

We now consider the dynamics on phenotype space implied by the system consisting of both the pure dynamics and the interactions. The difference between the phenotype dynamics and before is that a distribution of phenotypes is present in each species group and we study the way in which this distribution changes in time. However, in all the examples that follow we study only a distribution of phenotypes in one species group. All other species groups have only one phenotype present. We will also introduce a mutation process which will allow for mutations between phenotypes within a species group. We let $x_i(p_i)$ denote the phenotypic density of individuals in species group i, i.e. $x_i(p_i)\,dp_i$ is the number of individuals in species group i whose phenotype lies in a volume dp_i based at p_i.

Mutation-free dynamics.

The (mutation-free) dynamics are then given by

$$x_i'(p_i) \;=\; X_i(x(p), e_i, p_i) \tag{44}$$

$$e_i = e_i(x,p)$$

where X_i is given by the pure dynamics in equation (34), $x(p) = (x_1(p_1), \ldots, x_s(p_s))$, $p = (p_1, \ldots, p_s)$ and $e(x,p)$ is the value of the interactions corresponding to the distribution given by x.

This defines a dynamical system. Let us denote the mapping given by this by L i.e. $L(x) = x'$.

We can restore the pure dynamics, equation (34), from equation (44) by taking for x_j the distribution $x_j^0 \delta_{p_j}$ for each j where δ_{p_j} is the delta function on P_j concentrated at p_j (i.e. each species group j is represented by a single population of x_j^0 individuals each with the phenotype p_j). The p'-mutated dynamics (35) are then also given by (44) by taking for x_j the distribution $x_j^0 \delta_{p_j}$ if $p'_j = p_j$ and $x_j^0 \delta_{p_j} + y_j^0 \delta_{p'_j}$ when $p'_j \neq p_j$.

Mutations.

Now, we consider how to add mutations to this process. We discuss this in terms of distributions on the phenotype space. These represent the distribution of the phenotypes present in the system. In the case where only a single phenotype p_j is present in species group j, this is represented by a delta function δ_{p_j}. We assume that under mutation in group j such a pure situation changes to the probability distribution $M_j(\delta_{p_j})$. This means that after one time step, mutation causes the distribution of phenotypes to change from δ_{p_j} to $M_j(\delta_{p_j})$. Usually, $M_j(\delta_{p_j})$ will be a smooth distribution close to the delta function δ_{p_j}.

Now if $x = x(p)$ then the phenotype dynamics are given by $F(x) = M(L(x))$ i.e.

$$\begin{array}{ccccc} & & L & & M \\ x & \rightarrow & L(x) & \rightarrow & M(L(x)). \\ & & \text{local dynamics} & & \text{mutation process} \end{array}$$

6.9.1 Computational methods.

The implementation of the phenotype dynamics involves the discretisation of the phenotypic constraint manifolds P_i. If they are 1-dimensional then we approximate them by considering an equally spaced lattice of phenotypes p_i^j, $j = 1, \ldots, N$. We denote the spacing by $\alpha = \alpha(P)$ and always check that our results are not dependent upon the particular choice of α. The phenotypic population distribution $x_i(p_j)\,dp_j$ is then approximated by the vector $(x_i(p_i^1), \ldots, x_i(p_i^N))$. The interactions e_{ij} which are all integrals over dp_j in the examples that we consider then become sums.

For the 1-dimensional discretised case the mutation process is taken to be of the following form

$$M_j(\delta_{p_j^k}) = (1-\lambda)\delta_{p_j^k} + \frac{\lambda}{2}(\delta_{p_j^{k+1}} + \delta_{p_j^{k-1}}) \tag{45}$$

where the parameter λ is the *mutation rate* and $0 < \lambda < 1$. In this discretised situation, the delta function $\delta_{p_j^\ell}$ is just represented by the vector in $\mathbb{R}^N$ all of whose entries are zero except that corresponding to the ℓth entry. In all the examples we set $\lambda = 0.001$.

6.9.2 Phenotypic attractor for the constrained resource-predator-prey model.

Recall that in this constrained problem we only allow the phenotype of the prey to vary. Thus only b_1, c_1 and c_2 vary and these are related so that the phenotype $p = p_1$ is determined by b_1 and hence identified with it. Therefore, the discretised phenotype dynamics for the resource-predator-prey system whose pure dynamics and interactions are given by equation (33) are, denoting $x_1(p^\ell)$ by x_1^ℓ and $c_1(p_1^\ell)$ by c_1^ℓ:

$$\begin{aligned}
\frac{(x_1^\ell)'}{x_1^\ell} &= (1+b_1)\exp\left(-\alpha\frac{e_{11}}{e_{13}} - \frac{e_{12}}{1+d_1e_{11}}\right) \\
\frac{x_2'}{x_2} &= (1-d_2)\exp(e_{21}) \\
\frac{x_3'}{x_3} &= (1+b_3)\exp\left(-\frac{e_{33}}{k} - e_{31}\right)
\end{aligned} \tag{46}$$

where $e_{11}^\ell = \sum_\ell x_1^\ell$, $e_{12}^\ell = c_1^\ell x_2$, $e_{13}^\ell = e_{33} = x_3$, $e_{21} = \sum_\ell c_2^\ell x_1^\ell$ and $e_{31} = c_3 \sum_\ell x_1^\ell$. The parameter values are as in table 6. Recall that for the first set of values, there is a single

ESS at $p = p_*$. There is a distribution $\xi_1(p_1')$ such that, under the phenotype dynamics (with the mutation process defined by equation (45)), any initial distribution converges to $\xi_1(p_1')$. This asymptotic distribution is a "smoothed out" delta function and is shown in figure 20. Note that it is very much wider than the mutated delta function $M_1(\delta_{p_*})$. In figure 21 we show the time dependence of the mean phenotype. It shows convergence to the ESA value.

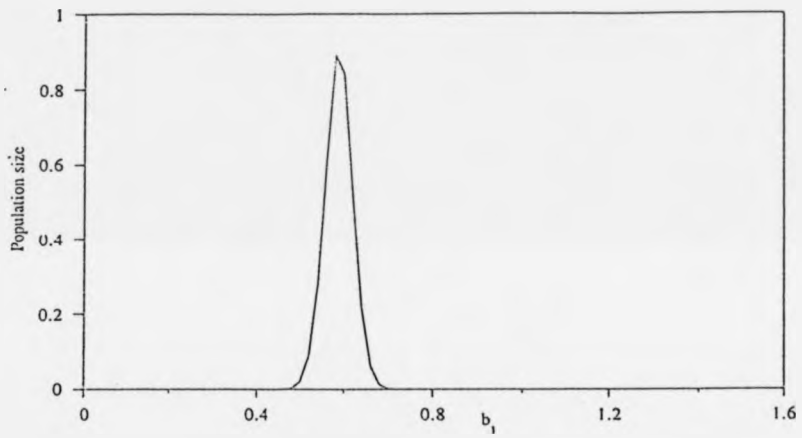


Figure 20: The asymptotic distribution for the phenotype dynamics of the resource-predator-prey system whose pure dynamics are given by equation (33) with parameter values from the first set in table 6.

Not surprisingly in the second case, where there are two ESA values separated by an evolutionary repeller, the behaviour is more complicated. Any initial condition converges to one of two asymptotic distributions, but to which depends upon the initial condition and the mutation rate. If the latter is relatively large, then the asymptotic state is always a "smoothed out" delta-function close to the larger ESA value as in figure 22. If the mutation rate is smaller and the support of the initial condition is very close to the smaller ESA value or to the left of this, then the asymptotic state is a "smoothed

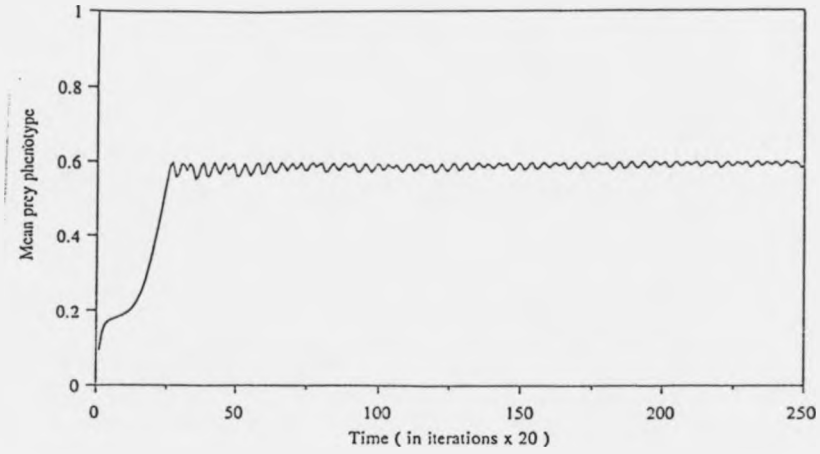


Figure 21: Time dependence of the mean phenotype showing convergence to the ESA value.

out" delta function close to this ESA value. This is also shown in figure 22. Otherwise, it converges to the other asymptotic state.

In figure 23 we show the time dependence of the mean phenotype for an initial condition which, depending on the size of the mutation rate, converges to one of the two asymptotic distributions. The phenotypes present in the initial condition are all to the left of the smaller ESA value.

6.9.3 Phenotype dynamics for the Hawk-Dove system of section 6.4.

Mixed strategies.

Recall that we studied this in section 6.4 for the case where each predator played a mixed strategy of Hawk or Dove. We therefore consider the phenotype dynamics for this case first before proceeding to the case of a polymorphic population.

It follows from the definition of the pure dynamics and the interactions that the

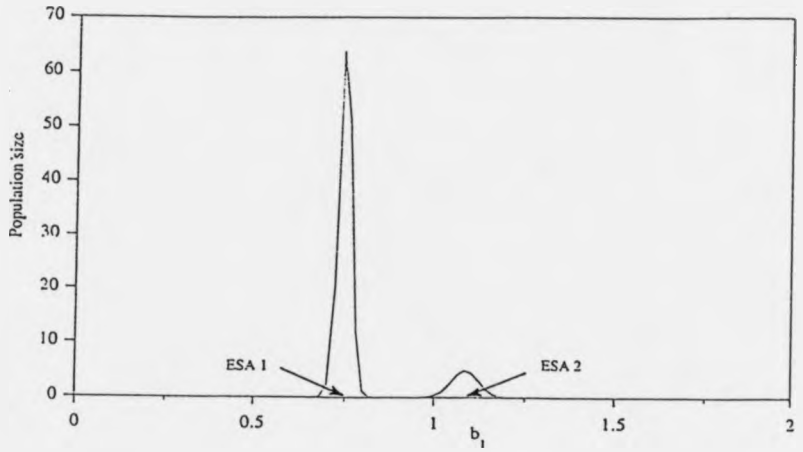


Figure 22: The two asymptotic distributions for the second set of parameter values in table 6 for the resource-predator-prey system.

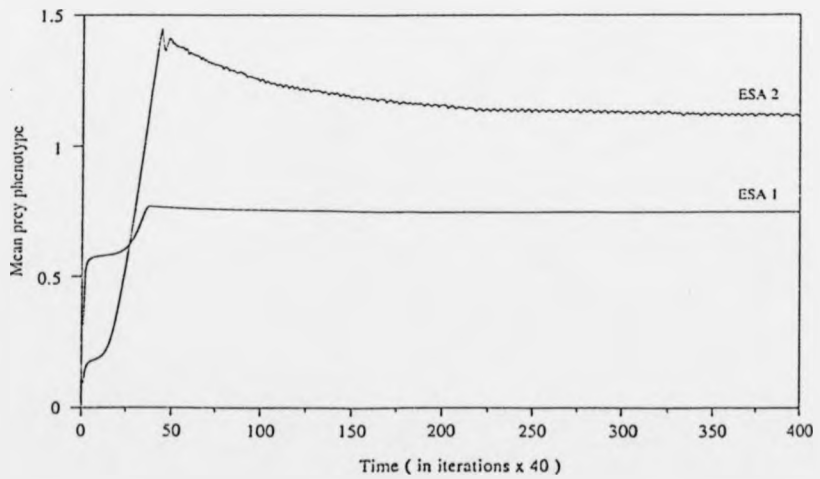


Figure 23: Time dependence of the mean phenotype showing convergence to the two ESA values.

phenotype dynamics are given by

$$\begin{aligned} x_1' &= x_1(1+b)\exp\left(-\alpha\frac{x_1}{k}-\frac{e_{12}}{(1+d_1x_1)}\right) \\ x_2' &= x_2e^{-\sigma}E(p,e_{22}) \\ e &= e(x) \end{aligned} \tag{47}$$

where $E(p,q)=\sum_{i,j}p_iq_j\exp(c_2E_{ij})$, $e_{12}(\xi_2)$ is the total predator population weighted mean of c_1, $\int c_1(p_1,p_2)\xi_2(p_2)\,dp_2$, and $e_{22}(\xi_2)$ is the mean predator population strategy $\int p_2\xi_2(p_2)\,dp_2/\int\xi_2(p_2)\,dp_2$. We discretise these as follows, denoting $x_2(p^\ell)$ by x_2^ℓ and $c_1(p^\ell)$ by c_1^ℓ,

$$\begin{aligned} x_1' &= x_1(1+b)\exp\left(-\alpha\frac{x_1}{k}-\frac{e_{12}}{(1+d_1x_1)}\right) \\ (x_2^\ell)' &= x_2^\ell e^{-\sigma}E(p^\ell,e_{22}) \\ e_{12} &= \sum_\ell c_1^\ell x_2^\ell \\ e_{22} &= \sum_\ell p^\ell x_2^\ell/\sum_\ell x_2^\ell. \end{aligned}$$

Note that $c_1^\ell=c_1$ does not depend upon ℓ since we assume that c_1 is the same for both Hawks and Doves. Therefore e_{12} is c_1 times the total predator population size.

Recall that this system has two ESA values separated by an evolutionary repeller at $p_r\approx 0.8$ (with quasi-periodic dynamics): the first at $p_*=V_0\bar{x}_1/C\approx 0.2<p_r$ and the other a boundary ESA at $p=1$. The phenotype dynamics reflect this. If the initial condition is a distribution with only phenotypes p with values greater than the evolutionary repeller p_r then, under the dynamics without mutation, it converges to a delta function at the boundary ESA. With mutation this asymptotic distribution is smoothed out. Other initial conditions converge to the distribution shown in figure 24 whose mean is at the interior ESA. This is not close to the delta function at p_* that might be expected, and the addition of mutation makes virtually no difference. The mean of this distribution is the ESA value, $p_*=V_0\bar{x}_1/C$. Thus, the population plays the same mean strategy as the pure ESA, but there is very great phenotypic variation in the population.

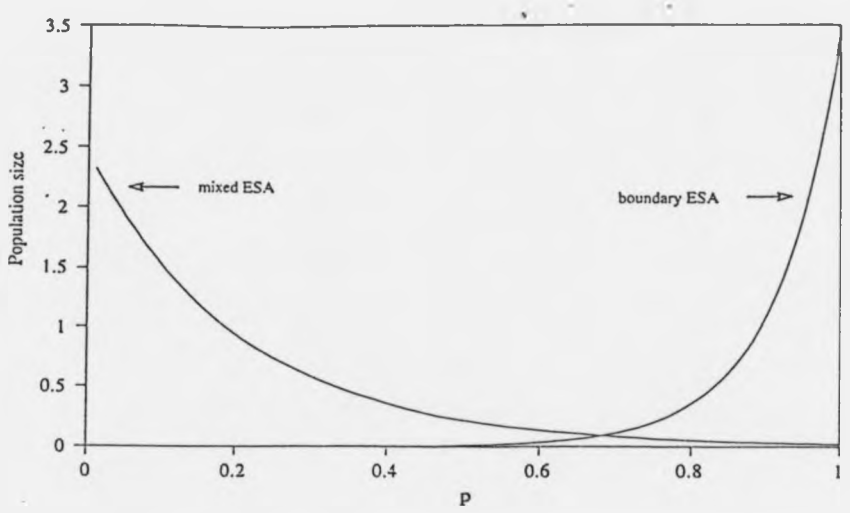


Figure 24: The asymptotic distributions for the phenotype dynamics for the Hawk-Dove predator-prey system. The value of the parameter C is 75.

In figure 25 we show the time dependence of the mean strategy for two initial conditions which converge to each of the two asymptotic distributions.

Polymorphic populations.

The dynamical behaviour of polymorphic populations is very different. In a polymorphic population each individual plays one of the pure strategies and not a mixed strategy as in the previous case. Firstly, let us consider the pure equations for this system. The phenotype p of the predators belongs to the discrete set $P = \{H, D\}$. Those individuals with $p = H$ play the pure Hawk strategy and those with $p = D$ play the pure Dove strategy. The pure equations are given by

$$\begin{aligned} x_1' &= x_1(1+b)\exp\left(-\alpha\frac{x_1}{k} - \frac{e_{12}}{(1+d_1x_1)}\right) \\ x_2' &= x_2 e^{-\sigma} E(p, e_{22}) \qquad (48) \\ e_{12} &= \text{predator population mean of } c_1 = c_1x_2 \\ e_{22} &= \text{mean population strategy} = p. \end{aligned}$$

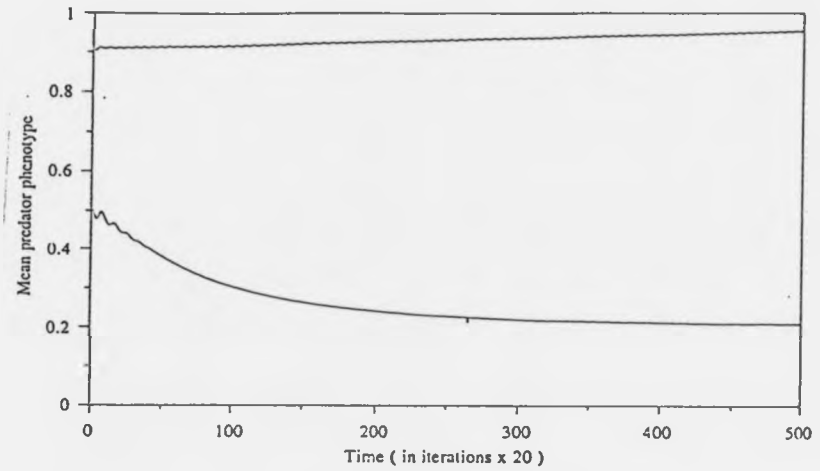


Figure 25: Time dependence of the mean strategy showing convergence to the two ESA values.

The notion of an ESA for the pure dynamics is not so interesting for this system because of the discreteness of P. Trivially the only possible ESAs correspond to systems that consist only of Hawks or only of Doves and small phenotypic mutations are not possible.

However, the phenotype dynamics which are obtained from this system are very interesting and are a generalisation of the Jonker-Taylor-Zeeman dynamics [108], [118]. The equations for this are

$$\begin{aligned} x_1' &= x_1(1+b)\exp\left(-\frac{x_1}{k} - \frac{e_{12}}{(1+dx_1)}\right) + 1 \\ \frac{x(H)'}{x(H)} &= e^{-\sigma}E(H, e_{22}) \qquad (49) \\ \frac{x(D)'}{x(D)} &= e^{-\sigma}E(D, e_{22}) \end{aligned}$$

(50)

where e_{22} is the mean predator population strategy and e_{12} is the predator population mean $c_1(D)x(D) + c_1(H)x(H)$ of c_1. The latter equals $c_1(x(D) + x(H))$ since we take $c_1(D) = c_1(H) = c_1$. The term 1 which is added to the right-hand side of the first

equation represents a very small import of prey into the system. It is added to control numerical instability associated with near-extinctions. It does not affect our conclusions.

When the prey are not present and the value of the resource is set at a constant V, then, as is well-known, any initial condition $(x(D), x(H))$ with $x(H) \neq 0$ and $x(D) \neq 0$ converges to $x(H) = V/C$ if $V/C < 1$ and to $x(H) = 1$ if $V/C > 1$. The behaviour is very different when the prey are present. For the parameter values in table 9, we observe complex oscillatory behaviour as shown in figure 26.

b	c_1	c_2	d	k	V_0	C	σ
6.5	0.01	0.001	0.005	2000	1	450	0.006

Table 9: Parameter values for the Hawk-Dove predator-prey equation (49).

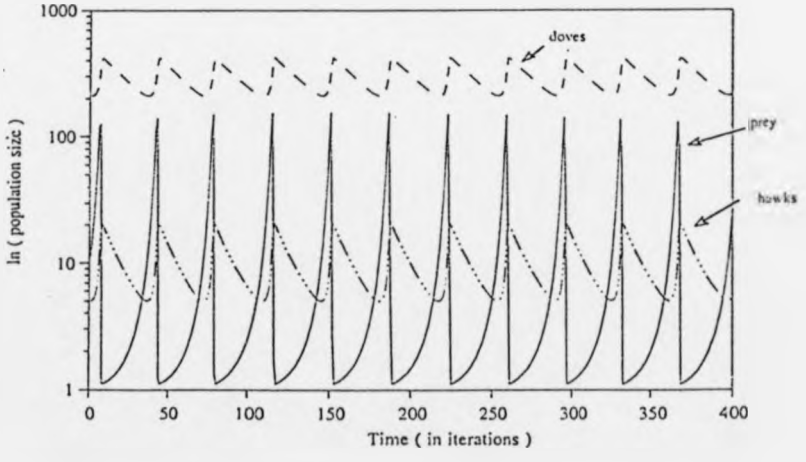


Figure 26: Time dependence of the numbers of Hawk and Dove predators and prey.

6.9.4 Phenotype dynamics for the predator-prey system with strategic prey.

We now consider another type of interesting phenotypic dynamics and introduce a new system. This is again a predator-prey system, but now we assume that the prey are strategic. Their strategy is represented by a number p in the unit interval $P = [0,1]$ which represents the weight given to searching for food. Those with $p = 0$ put a large effort into actively searching for food with a consequently higher risk of predation. Those with $p = 1$ do not actively search for food and are therefore immune to predation but are likely to go hungry. Intermediate strategies are represented by intermediate values of p. For the pure equations and interactions we take the following:

$$\begin{aligned} (x_1^\ell)' &= x_1^\ell(1+r)\exp\left(-\frac{p^\ell e_{11}}{k} - (1-p^\ell)\frac{c_1 x_2}{(1+d_1 e_{11})}\right) + \frac{x_1^\ell}{e_{11}} \qquad (51) \\ x_2' &= x_2(1-d_2)\exp(e_{21}) + 1 \end{aligned}$$

where e_{11} is the total prey number and e_{21} is the prey population weighted mean of c_2. For the pure case $e_{11} = x_1$. As in the previous example, the terms x_1^ℓ/e_{11}^ℓ and 1 are introduced to control numerical instabilities that arise because of near extinctions in the dynamics. The first term is chosen so that it goes to zero with x_1 to allow competition with successfully invading mutants to wipe the original population out. Normally, we would introduce a further interaction term, e_{12}, corresponding to the total number of predators. However, since we are not going to consider mutations of the predator, this is not necessary.

In figure 27 we plot the graph of f_+ for this equation when the parameter values are as in in table 10. We observe that there are two ESA values separated by an evolutionary repeller. There is a boundary ESA at $p = p_b = 0$ and evolutionary repeller at $p_r \approx 0.07$ and an ESA at $p_* \approx 0.23$. In the phenotype dynamics these ESAs compete with the result that the mean strategy $\bar{p}$ oscillates in an irregular fashion between $p = 0$ and $p = 1$. When the system is near the state where $\bar{p}$ is close to 0 or 1, then the population's

phenotypic distribution is bunched up close to this end-point. As the system moves to the opposite extreme all the individuals move to the other end of phenotype space. For this reason, and because the strategy is akin to confidence, we regard this as a *biological business cycle.*

r	c_1	c_2	d_2	k	d_1
1.8	0.01	0.001	0.1	130	0.001

Table 10: Parameter values for the strategic prey equation (51).

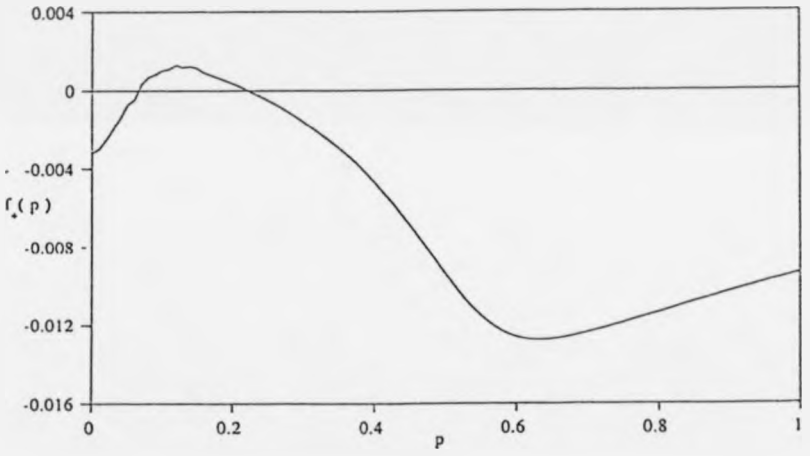


Figure 27: The graph of f_+ for the risk trade-off model (51) for parameter values given in table 10.

6.10 Learning dynamics.

In this section we are interested in the following learning problem. We consider a general dynamical system of the form

$$\xi' = \Xi(\xi, x, c)$$

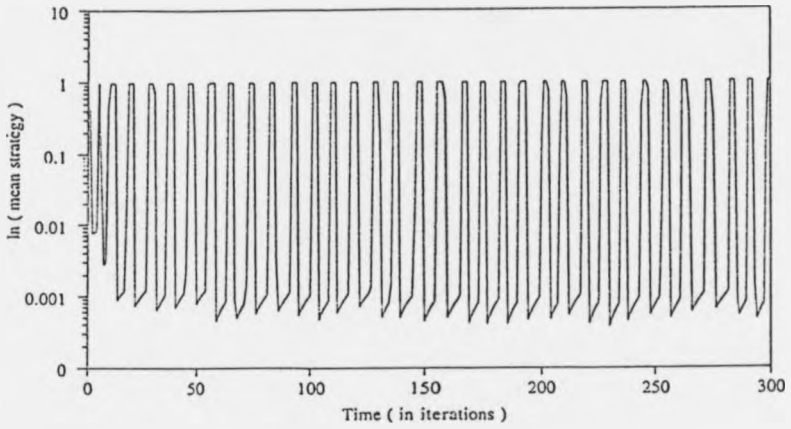


Figure 28: Time dependence of the mean prey strategy caused by the oscillating evolution between the two almost-boundary strategies.

$$x' = \lambda(\xi, x, c).$$

Here ξ is some dynamically varying background environmental variable and x describes a distribution of strategies or behaviour amongst a population of agents. The variable c represents external control variables. Unlike our previous systems, the strategy dynamics are not given by reproduction but rather by learning. This is represented by the function λ which gives the learning rule. This can represent a sort of phenotype dynamics because x can contain distributions of strategies.

The function λ will be our phenotype and we will allow mutations in this learning rule.

As an example we will consider the following simple learning system. We assume a discrete set of strategies $S = \{1, \ldots, m\}$ and assume a population Π consisting of individuals π. The state of the system is given by the configuration $\underline{\sigma} = (\sigma_{i,\pi})$ where $\sigma_{i,\pi}$ is the probability that $\pi \in \Pi$ plays strategy i.

We denote by $P_i(\underline{\sigma})$ the payoff to strategy i if the population is playing the configuration $\underline{\sigma} = (\sigma_{i,\pi})$. Then we define $\underline{\sigma}' = \lambda(\underline{\sigma})$ by

$$\sigma'_{i,\pi} = r_i + \frac{1}{Z}\left((\sigma_{i,\pi} - r_i)^m \exp(P_i(\underline{\sigma}) - \bar{P}(\underline{\sigma}))\right) \tag{52}$$

where $\bar{P}(\underline{\sigma})$ is the population's mean payoff and Z is chosen so that $\sum_i \sigma_{i,\pi} = 1$. Note that this generates the sequence

$$\sigma_{i,f}(t+1) = \frac{1}{Z}\left(r_i + \exp(m\tilde{P}_i(t) + m^2\tilde{P}_i(t-1) + \cdots)\right)$$

where $\tilde{P}_i(t) = P_i(t) - \bar{P}(t)$ is the excess payoff at the tth time step. This is analogous to the Harley learning rule [42] (see section 5.2.5). The parameter m is a discount factor. It is also important to note that the initial condition $\sigma_{i,\pi}$ must satisfy $\sigma_{i,\pi} > r_i$. Otherwise the probabilities will become negative.

To make things simple let us further suppose that the payoffs are given in the usual linear way by a payoff matrix $E = (E_{ij})$. Then the payoff is given by

$$P_i(\underline{\sigma}) = \sum_j q_j E_{ij}$$

where $q_j = \sum_{\pi'} \sigma_{j,\pi'} / \sum_i \sum_{\pi'} \sigma_{i,\pi'}$ is the population's mean strategy.

The simplest case to consider is where $\Pi = \{1\}$. One can regard this as the situation where everyone is forced to play the same strategy. Then the state of the system is given by $\underline{\sigma} = (\sigma_i)$. It follows that $q_j = \sigma_j$.

We use the payoff matrix of our previous Hawk-Dove predator-prey system discussed in section 6.9.3. Instead of being prey we regard the agents corresponding to the population $\xi = x_1$ as consumers and the agents corresponding to the predators as agents who either adopt a high-risk aggressive selling strategy (Hawks) or else are risk-averting (Doves). The strategic learning process λ is as described above.

Thus, if we let ξ denote the number of consumers, $x(D)$ and $x(H)$ denote the numbers of Hawk and Dove agents, $\tau = x(H)+x(D)$, $h = x(H)/\tau$, $d = x(D)/\tau$ and $\sigma = x(H)/\tau$,

then our pure dynamics are given explicitly by

$$\begin{aligned} \xi' &= \xi(1+b)\exp\left(-\frac{\xi}{k} - c_1\frac{\tau}{1+d_1\xi}\right) \\ h' &= r_H + \frac{1}{Z}(h-r_H)^m \exp(E(H,p) - E(p,p)) \\ d' &= r_D + \frac{1}{Z}(d-r_D)^m \exp(E(D,p) - E(p,p)) \\ \tau' &= \exp(c_3 E(p,p)). \end{aligned} \tag{53}$$

The parameters are given in table 11. For these parameter values, the attractor of this system is a fixed point.

b	c_1	c_2	d	k	V_0	C	σ	c_3	m
0.5	0.01	0.001	0.001	1000	1	180	0.02	0.05	0.8

Table 11: Parameter values for the learning system given by equation (53).

Now recall that we think of λ as the phenotype p and consider evolutionary stability to mutated learning processes. Here we only consider evolution within the class of learning processes λ of the form described above. These are parametrised by the discount factor m and the residual biases or preferences r_H and r_D. The evolution of m is simple. Under evolution, the value of m increases. Here we fix its value at 0.8.

Thus we consider $p = (r_H, r_D)$. For simplicity we consider these separately. Recall that we must have $r_i < 1$ since, as was pointed out above the initial condition $\sigma_{i,\pi}$ must be greater than r_i. For the Hawk's residual preference r_H there are two boundary ESAs at $r_H = 0$ and $r_H = 1$ separated by an evolutionary repeller. The value $r_H = 1$ is not a boundary ESA value in the usual sense. This is because in practise it cannot be realised because of the constraint $\sigma_{H,\pi} > r_H$. However, the system will evolve to values arbitrarily close to $r_H = 1$ if the initial strategy has a greater value than the evolutionary repeller.

For the Dove's residual preference r_D there is an interesting ESA at $r_D = r_D^* \approx 0.34$. The graph of f_+ for r_D is shown in figure 29. Thus we see that the learning evolves

to give us a specific learning structure with a specific residual preference. We observed that the ESA value and the evolutionary repeller are given by a condition analogous to that for the classical linear Hawk-Dove game. They are precisely the values for which the proportion of time playing Hawk equals $\bar{V}/C$ where $\bar{V}$ is the mean resource value $V_0\bar{\xi}$.

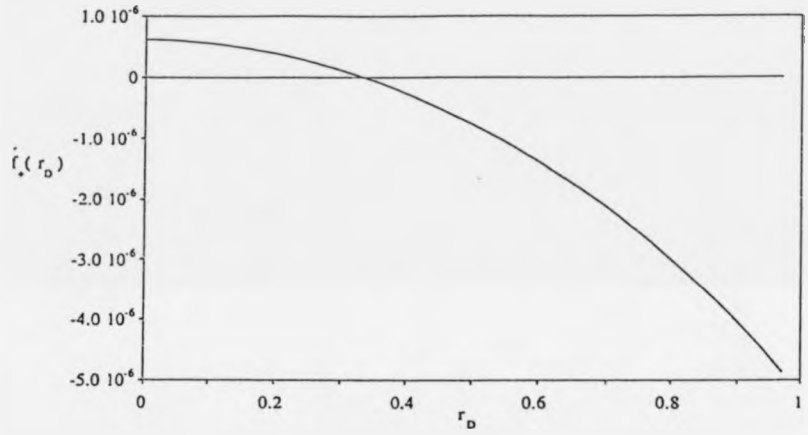


Figure 29: The graph of f_+ for the learning system given by equation (53) when $p = r_D$, for the parameter values given in table 11. The parameter r_H is set to 0.

We then consider the phenotype dynamics for this system where we separately take the phenotype p to be one of r_H and r_D. We take the usual mutation process given by equation (45) in our phenotype dynamics. For $p = r_H$ initial distributions converge under the phenotype dynamics to one of two distributions depending upon initial conditions. These are shown in figure 30. They show great phenotypic variation. However, we believe that this is due to the fact that the selective pressures are small and therefore the effect of the mutation process is amplified. Without mutations, the system will converge to a delta function at one of the ESAs. For $p = r_D$ they converge to the interesting distribution shown in figure 31. This gives a wide distribution of phenotypes with mean

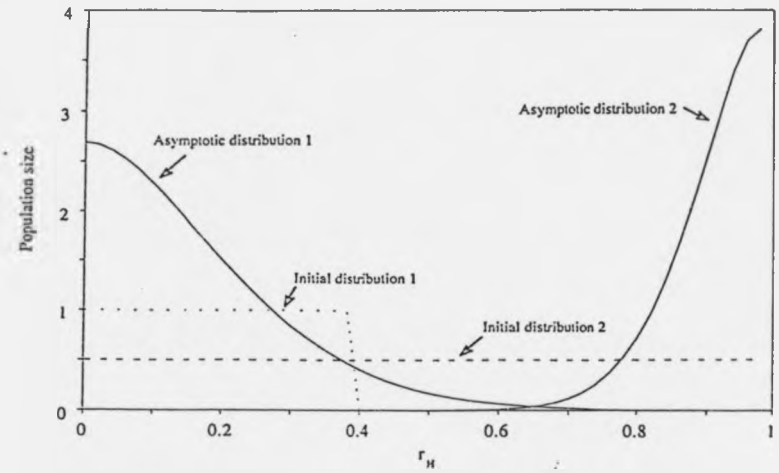


Figure 30: The asymptotic phenotype distributions for $p = r_H$ for the learning system given by equation (53). The parameter r_D is set to 0. With the distributions we show two initial conditions that converge to them.

value at the ESA value r_D^*. Thus we conclude that in such a population with phenotypic variation, the learning rule will evolve (under non-reproductive learning evolution) to one with the r_H having a very low or high value and the r_D having great phenotypic diversity around the mean given by the ESA value.

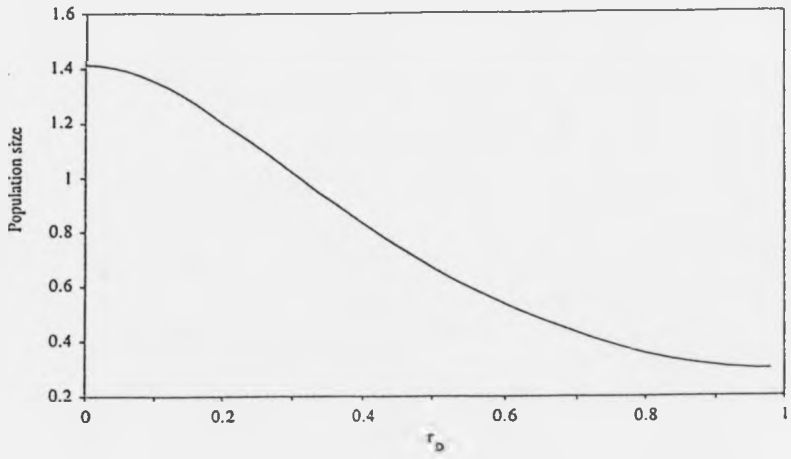


Figure 31: The asymptotic phenotype distribution for $p = r_D$ for the learning system given by equation (53). The parameter r_H is set to 0.

7 Evolutionary catastrophes, punctuated equilibria and gradualism in ecosystem evolution.

7.1 Introduction

In chapter 6 and in [80] a general mathematical theory of Darwinian evolution in ecosystems is developed. In this chapter we wish to address the important issue of gradualism *v.* punctualism in evolutionary theory, [26], [36] and [101]. We discuss this in terms of a simple illustrative example, but emphasise that it follows from reference [80] that our results apply quite generally and are ubiquitous and wide-ranging. We show that our model displays stasis, gradualism and punctuated equilibria and we use it to present a new general mechanism for such phenonema involving what we call *evolutionary catastrophes*. We also consider the Red Queen Hypothesis, [112] and [103], and discuss a mechanism for speciation. We argue that the possibility of Red Queen evolution is a

function of the system's phenotypic constraints. Thus, we propose that rather than ask about its general occurrence one should switch attention to the structure of phenotypic and genotypic constraints in ecosystems.

7.2 Evolutionary stability of ecologies

7.2.1 The model and evolutionarily stable attractors

One of the basic concepts of chapter 6 is that of an *evolutionarily stable attractor*. This is formulated along lines which generalise Maynard Smith's evolutionarily stable strategies, see section 5.2, to a much wider class of dynamical systems including coevolving ecologies with complex dynamics. Here, rather than give a general mathematical description, we conduct the discussion in terms of the following illustrative example. It is the predator-prey-resource ecosystem introduced in section 6.1. The dynamics are given by a mapping which displays a wide range of dynamical behaviour including chaos. This mapping is of the form:

$$\frac{x_i'}{x_i} = F_i(x, e, p) \quad (i = 1, 2, 3). \tag{54}$$

Here x_1, x_2 and x_3 denote respectively the population size of a prey species, a predator species and a resource species with respective phenotypes p_1, p_2 and p_3. Each aspect of the phenotype is capable of continuous variation. We let P denote the set of possible values that the phenotype vector $p = (p_1, p_2, p_3)$ can take. The primed variables represent the corresponding numbers in the next period. The variable $e = (e_{ij}(p_i))_{i,j=1,2,3}$ describes the interaction of these species with any mutants that are present. The term $e_{ij}(p_i)$ represents the effect of species group j upon those individuals in species group i with phenotype p_i. In fact, in equation (54), e_{ij} is either the total population size of species group i or else represents the jth species group population weighted mean of some contact parameter $c_i(p_i, p_j)$. For example, e_{21} measures the aggregated effect on the predator individuals with phenotype p_2 of the total prey population.

In the pure case, $F_1 = (1+b_1)\exp\left(-\alpha x_1/x_3 - c_1 x_2(1+d_1x_1)^{-1}\right)$, $F_2 = (1-d_2)\exp(c_2x_1)$ and $F_3 = (1+b_3)\exp(-x_3/k - c_3x_1)$. The biological meaning of the various terms is explained in table 12. However, we want to write these so that the equations are still correct even when mutants are present. Thus we will write them in terms of the variables $e = (e_{ij})(p_i)_{j=1,\ldots,s}$. Then $F_1 = (1+b_1)\exp\left(-\alpha e_{11}/e_{13} - e_{12}(1+d_1e_{11})^{-1}\right)$, $F_2 = (1-d_2)\exp(e_{21})$ and $F_3 = (1+b_3)\exp(-e_{33}/k - e_{31})$.

Term in equation	Interpretation
$1+b_1$	prey's unconstrained birth rate
$-\alpha\frac{x_1}{x_3}$	decrease in fitness of prey due to resource limitation
$-c_1\frac{x_2}{1+d_1x_1}$	decrease in fitness of prey due to predation
$1-d_2$	predator's unconstrained death rate
c_2x_1	predator's fitness increase due to feeding
$1+b_3$	resource's unconstrained reproductive rate
$-\frac{x_3}{k}$	resource's carrying capacity limitation
$-c_3x_1$	resource's fitness decrease due to feeding by prey

Table 12: An explanation of the construction of equation 54.

If there are no mutant species then $e_{ij} = e_{ij}^{\text{pure}} = x_j$ if $ij = 11$, 13 or 33 and $e_{ij} = e_{ij}^{\text{pure}}(p_i) = c_i(p_i,p_j)x_j$ if $ij = 12$, 21 and 31. Then we call equation (54) the *pure system*. Let us consider now the situation where there are mutants in each species group i with phenotype p_i' and population size y_i. Then $e_{ij} = x_i + y_i$ if $ij = 11$, 13 or 33 and $e_{ij} = e_{ij}^{\text{pure}}(p_i) = c_i(p_i,p_j)x_j + c_i(p_i,p_j')y_j$ if $ij = 12$, 21 and 31 and the dynamical equation is

$$\frac{x_i'}{x_i} = F_i(x,e,p), \quad \frac{y_i'}{y_i} = F_i(y,e,p') \quad (i=1,2,3) \tag{55}$$

where $y = (y_1,y_2,y_3)$.

We call this the *p'-mutated dynamics*. We ask if there is a negligible probability

that any such a small population with these mutant phenotypes can invade. If so, then we say that p is *p'-stable*. Using this notion, we say that an attractor Λ of the pure system (54) is *evolutionarily stable* if for all p' near p, p is p'-stable. Then we call p an ESA value. In chapter 6 we illustrate the wide range of this concept. The way in which we have formulated our definition of p-stability is dictated by the subtlety of chaotic attractors. However, from a practical point of view the subtle theoretical definition which is given in [80] can be regarded as equivalent to the obvious one since the differences are practically unobservable.

7.2.2 The invasion exponent

The *invasion exponent*, $\vartheta_p(p') = \vartheta(\Lambda_p, p')$ is a mathematical tool which enables analytical and numerical analysis of relatively complex situations, such as the coevolution of multiple species with chaotic population dynamics. It characterises the evolutionary stability of an attractor $\Lambda = \Lambda_p$, corresponding to a phenotype p, to a small mutant population with phenotype p'. It measures the rate of growth of the invading mutant population. It can be calculated for all dynamical situations including those with periodic, quasi-periodic and chaotic attractors. For equilibria, its definition corresponds to notions that have already appeared in the literature (see for example [17] and [64]). However, for general attractors its definition involves less well-known mathematical ideas such as the ergodic invariant measures of the attractor Λ. It is given by averaging $\log F_i(0, (e_{ij}^{\text{pure}}(p_i'), p')$ over the attractor Λ_p using the so-called natural measure.

The important point is this:

Λ_p is an ESA if $\vartheta_p(p')$ is negative for all p' in P near but not equal to p

In fact, if $\vartheta_p(p') < 0$ then p is p'-stable and conversely, if $\vartheta_p(p') > 0$, then p can be invaded by p'-mutants. For ESA values p in the interior of P, the generic situation is that $\vartheta_p(p')$ is of the form $-\kappa(p'-p)^2 + \mathcal{O}((p'-p)^3)$ for p' near p with $\kappa > 0$. However, if p is an ESS of a linear evolutionary game, then $\kappa = 0$. This degeneracy is reflected in

the fact that Maynard Smith's condition for an ESS (equations (17) and (18) and 2.4 of [64]) mixes a first order condition with one that is second order. Generic nonlinear systems only require a first order condition.

7.2.3 ESAs in the resource-predator-prey system.

If there are no constraints upon the parameters of equation (54) then there are no ESAs. This is easily proved using the results of chapter 6. However, the absence of such constraints is biologically unrealistic. For example, it is natural to suggest a trade-off between the prey growth parameter b_1 and its contact parameter c_1. A change in behaviour which causes an increase in b_1 may involve more exposure to predators and hence an increase in c_1. Such a relationship implies the existence of an ESA. For the purposes of exposition we will assume a linear relationship ($c_1/b_1 = k$). At the same time, since the contact rate of prey with predators and predators with prey must be equal, we keep the ratio of c_2 to c_1 fixed ($c_2/c_1 = l$). The other parameters α, c_3, d_i are fixed. Thus our phenotype space P is given by

$$P = \{(b_1, c_1, c_2) : b_1 \geq 0, c_1/b_1 = k, c_2/c_1 = l\}. \tag{56}$$

It is 1-dimensional and parametrised by $b_1 \geq 0$.

In figure 32 we plot the *differential selective pressure*

$$s(p) = \lim_{\varepsilon \to 0} \frac{\vartheta_p(p+\varepsilon)}{\varepsilon}.$$

If $s(p) > 0$ (resp. < 0) then higher (resp. lower) values of p can invade. If p is an ESA value, then $s(p) = 0$. In our examples the principle of mutual exclusion holds which means that, after successful invasion, the invading population actually takes over and replaces the original population. Thus, by inspecting the graph of s we can see which of its zeros are ESA values and which initial phenotypes evolve to which ESA.

In figure 32 we show the graph of s for three values of the parameter b_3 and with the parameter values given in table 13. We observe that for some values of b_3, there is

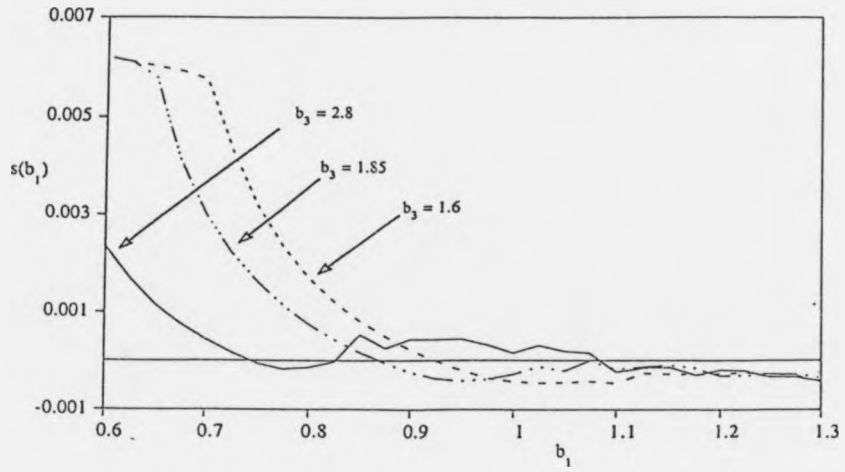


Figure 32: The differential selective pressure s as a function of b_1 for the constrained system with parameter values given by table 13 and three values of b_3. The zeros of these graphs are either ESA values or evolutionary repellers. Which, can be determined from the sign of s on either side of the zero. The jaggedness of the graph in places is caused by the nonuniform convergence of time-series for the ergodic measures of chaotic attractors.

α	b_1	b_3	c_1	c_2	c_3	d_1	d_2	k
0.5	1.1	*variable*	0.001	0.0005	0.005	0.01	0.05	1000

Table 13: Parameter values for equation (54) giving chaotic dynamics.

one ESA value while, for others, there are two. When there is more than one they are separated by an *evolutionary repeller*. Moreover, the associated ESAs in this example are often chaotic. Thus we see that even in relatively simple examples we should expect to see multiple ESA values. In such a situation there may be some indeterminacy in the evolution.

7.3 Evolutionary catastrophes and punctuated equilibria

This evolution to the ESA takes place on an ecological time-scale. We now want to consider longer time-scales. We postulate some environmental change taking place on a time-scale which is slow compared to the prey species' adaptation. As the ecosystem's environment changes we find that the ESA values change in ways which correspond to stasis, gradualism and punctuated equilibria.

However, not only do we consider the behaviour of the ESAs and the evolution of pure strategies, but we also consider the evolution of a distribution of phenotypes in the population. We study the detailed dynamics and evolution of the phenotypes using the phenotypic dynamical system introduced in section 6.9. This incorporates a small mutation rate between neighbouring phenotypes.

We assume that the effect of this slow environmental change is to decrease the parameter $\varepsilon = b_3$. Changes to other combinations of parameters can produce similar behaviour.

For all the values of the environmental variable ε, the attractor of the phenotype dynamics is a distribution with very small standard deviation and whose mean is an ESA value.

When ε is large there are two ESA values $p_*^1 < p_*^2$. The ESA value p_*^2 dominates the other in the following sense. From arbitrary initial conditions, the population evolves under the phenotype dynamics to one with all its phenotypes close to this ESA value.

At $\varepsilon = \varepsilon_{\text{crit}} \approx 1.85$ the dominant ESA value collides with the evolutionary repeller

and, for smaller values, only the other ESA value persists. Thus as ε passes through ε_{crit} we see a very fast evolution in the population's phenotype distribution. We term such a phenomenon an *evolutionary catastrophe.* The variation in the population's mean phenotype as ε changes is shown in figure 33. In this we see *gradualism* ($\varepsilon < \varepsilon_{crit}$), *punctuated equilibrium* ($\varepsilon = \varepsilon_{crit}$) and *stasis* ($\varepsilon > \varepsilon_{crit}$).

There is a resemblance to the jumps observed in Catastrophe Theory. In Catastrophe Theory, the jump occurs in phase space when the attractor of a dynamical system disappears. The jump is along a trajectory of the dynamical system. In an evolutionary catastrophe, an ESA value disappears and the resulting jump in phenotype space is due to mutation and natural selection. A possible explanation of punctuated equilibria in terms of Catastrophe Theory was proposed in [21] . However, this relied upon the assumed existence of a fitness function which evolution sought to maximise.

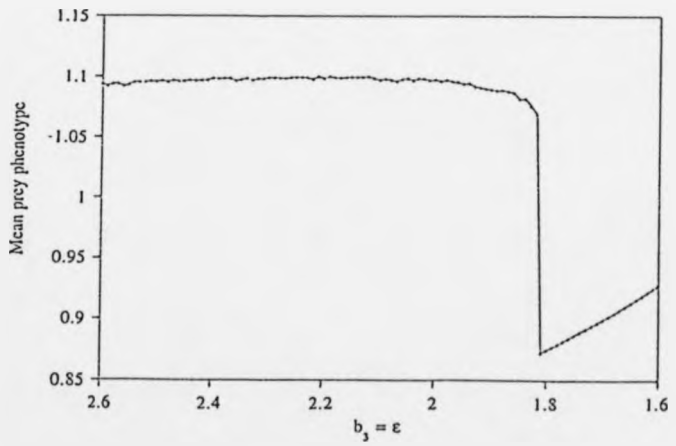


Figure 33: The graph of the mean prey phenotype as the parameter ε is slowly decreased.

7.4 Discussion

We argue that both evolutionary gradualism and evolutionary discontinuity akin to punctuated equilibria are natural and ubiquitous phenomena of this nonlinear dynamical theory of evolution. They should be viewed as different sides of the same coin rather than as opposing theories. Which will be observed will depend upon the detailed dynamical, ecological and environmental structure. Over long times of environmental change both should occur, side by side. Our discussion presented above also gives insights into the nature of Red Queen evolution and suggests a mechanism for speciation.

7.4.1 Constraints and the Red Queen Hypothesis

According to the Red Queen Hypothesis [112], even if the surrounding physical environment is constant, the majority of organisms are most of the time unlikely to be completely well-adapted since their ecological environment changes as a result of the evolution of coexisting species. On the other hand, an ecology at an ESA clearly violates this. Thus, Red Queen evolution can only occur if the system does not evolve to an ESA. We therefore interpret the absence of ESAs as being equivalent to the perpetual coevolution implied by the Red Queen Hypothesis and are led to the problem of determining when ESAs exist. This point has also been made by Marrow et. al. [59]. Constraints in their model come through the functions assumed for the interaction between prey and predator (see section 5.3.5).

We have already observed that if there are no constraints upon the parameters of equation (54) then there are no ESAs. The proof of this is in section 5.3.5, which clearly applies to a much wider range of problems. Moreover, the general results given there, strongly suggest that this is the general rule. Furthermore, these results can be used to determine which constraints are important in specific problems. Therefore, we suggest that, so far as this question is concerned, we should concentrate on understanding the nature of these constraints in real ecosystems. We note that ecological coevolution

certainly does not imply Red Queen evolution. This supports a similar conclusion in previous work on the equilibrium case, [84].

7.4.2 Speciation

It has been suggested that the episodes of rapid evolutionary change coincide with events of speciation, [26], [36] and [101]. This can be understood by applying an argument of Zeeman, [119], from a different context to our theory as follows.

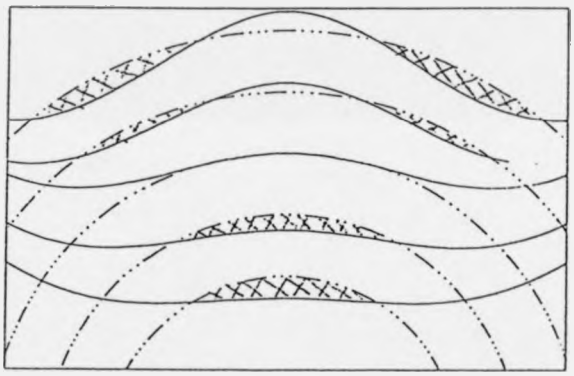


Figure 34: Evolution of the phenotypic support of the population. The phenotypic support at different times is shown hatched. The circular broken contours are the mutation fronts and the other contours are the selection fronts. The population's phenotypic support is caught between these. The phenotypic support is initially connected. However, after it has passed through the middle contour, it separates into two disconnected populations.

Consider a phenotypically variable population. We call the set of phenotypes in the population it's *phenotypic support*. It can be thought of as a subset of the phenotype space P above. Just after an evolutionary catastrophe, this phenotypic support is

rapidly changing. At a given time, mutation acts on it to increase it's extent and natural selection acts to reduce it. The most important part of this process for speciation takes place at the boundary of the phenotypic support.

The effect of natural selection can be measured by our differential selective pressure s introduced above. This is particularly informative when P is multi-dimensional. It defines a "pressure field" on P which tells us the pressure of selection between two nearby phenotypes in P. Using it we can determine the *selection front* of a population i.e. the set of phenotypes within the population which cannot be invaded by the rest of the population. During the rapid evolution following an evolutionary catastrophe, we expect this to be on the boundary of the phenotypic support.

Just after the evolutionary catastrophe, the phenotypic support will be connected and correspond to a single species. However, we can apply Zeeman's [119] argument to show that if the selection front is not aligned to the mutation front, then after a while the phenotypic support will break up into several components. These phenotypically disconnected components are the precursors of new species. They could, for example, now evolve to different ESAs or undergo Red Queen evolution with diverging evolutionary paths. The way in which the population splits and speciates is illustrated in figure 34.

8 Low-dimensional spatial dynamics in an artificial ecology.

8.1 Introduction.

Biological populations are composed of individuals whose movement is limited in space. Consequently ecological systems show heterogeneity and patchiness at a broad range of temporal and spatial scales and this heterogeneity is fundamental to population dynamics, [28] and [29], and to stability, [43] and [44]. Three basic types of models have been formulated to analyse spatial heterogeneity. (i) Assume homogeneous mixing within a patch and then link the patches by dispersal, [81], [82] and [45]. (ii) Reaction-diffusion equations, [94] and [72], which assume mass-action at every point in space and diffusion throughout the spatial domain. (iii) Individual-based spatially extended models called cellular automata, [32], [51], [68] and [22]. This third class of model is interesting as its basis is with the individual and dynamics and competition are built into the model through the use of simple transition rules and space.

We introduce an *artificial ecology* model of a resource-predator-prey system. This is a generalisation of a cellular automaton. The model is defined by the states a cell can be in and the rules governing the evolution of the cells. In this model the rules are stochastic, i.e. the state of the neighbourhood of a cell at time t determines the probability of various events in the neighbourhood occurring which determines the state at time $t + 1$.

We address the problem of spatial scale, i.e. the appropriate scale at which to make measurements of the population. This is a fundamental question in ecology and it is well-known (e.g. [105]) that the scale at which measurements are taken will influence the results and the conclusions gained. Too small a scale and the results may be dominated by noise. Too large a scale and dynamics may be masked out. We show how to extract a clearly defined spatial scale at which the system should be sampled to extract

the interesting dynamical effects. Similar work on this question has been noted and addressed by a number of authors (e.g. [53] and [12]);

Notwithstanding the probabilistic nature of the transition rules, at appropriate spatial scales the population dynamics are deterministic. Additionally these dynamics are low-dimensional. This allows us to characterise the complex spatial patterns observed in the ecology by a low-dimensional vector. We use these ideas to develop a general theory for data analysis of ecosystems to allow monitoring and management and to provide tools for detecting long-term structural change or drift.

8.2 The model.

Artificial ecologies (AEs) are a slight generalisation of probabilistic cellular automata (PCAs). Physical space is represented by a 2-dimensional $L \times L$ lattice Ω of cells. Each cell x can be in any one of a number of discrete states $s_1, \ldots, s_d$ and time t proceeds discretely. Therefore, the state of the system is given by a configuration $S = \{S_x\}_{x \in \Omega}$. The state $S(t)$ at time t determines a probability distribution on the potential future states and $S(t+1)$ is chosen randomly from this.

In a PCA each site is updated independently. This is not appropriate for many ecological applications because an event at one site can determine a specific change at another neighbouring site. Thus, in an AE the state of each neighbourhood determines a probability distribution on a finite set of admissible events. Each of these events is a transformation of the state on the neighbourhood. Thus the state $S(t)$ at time t determines a probability distribution of the states at time $t+1$. In a PCA all events only change the central site.

In this paper we consider a 2-dimensional resource-predator-prey system. Each cell can be in one of five states; resource, predator, predator in a resource, prey or empty. A predator in a resource is effectively both and its evolution can be described by what happens to each separately. Hence we will ignore this as a separate state. The boundary

conditions are wrap-around which can be interpreted as giving a torus as physical space.

To define the transition rules, it is simplest to explain our model by focussing on how the individuals within it act. This completely defines the event set and the event probabilities for a given neighbourhood state. In this description by an adjacent site to x we mean one of the sites $x+(1,0)$, $x-(1,0)$, $x+(0,1)$ and $x-(0,1)$.

- A resource site will grow into an adjacent empty site with probability g.
- A prey will move into and eat an adjacent resource site. Thus the adjacent resource site will become a prey and the original site will become empty. If there are no resource in adjacent sites a prey will move randomly into one of its adjacent sites.

  There is a certain probability of giving birth P_b into an adjacent empty site. This probability is zero if a prey individual hasn't eaten for a certain amount of time, t_{P_b}. A prey will die with probability one if it hasn't eaten for a certain amount of time t_{P_d} (where $t_{P_d} > t_{P_b}$).
- An empty site does nothing.
- A predator has a more complex neighbourhood to that of the other species. Its immediate neighbourhood is an 8-cell neighbourhood: $\mathcal{N}(x) = \{x+z : |z_1| \leq 1, |z_2| \leq 1, z \in \mathbf{Z}^2, z \neq (0,0)\}$. A predator will move into and eat a neighbouring prey site.

  As for the prey there is a certain probability of giving birth Z_b into a neighbouring empty site. This probability is zero if a predator individual hasn't eaten for a certain amount of time, t_{Z_b}. A predator will die with probability one if it hasn't eaten for a certain amount of time, t_{Z_d} ($t_{Z_d} > t_{Z_b}$), .

  If there are no neighbouring prey then a predator has a hunting ability in that it can sense prey much further away than it can actually move. A predator can "see" a prey anywhere in a 48-cell neighbourhood, i.e. the neighbourhood is $\mathcal{N}(x) = \{x+z : |z_1| \leq 3, |z_2| \leq 3, z \in \mathbf{Z}^2, z \neq (0,0)\}$. A predator, as the prey,

can only move one cell in one time step and will thus move to the cell in its 8-cell neighbourhood which is nearest (in the Euclidean sense) to the located prey.

All the simulations in this paper have the parameter values as given in table 14.

g	P_b	t_{P_b}	t_{P_d}	Z_b	t_{Z_b}	t_{Z_d}
0.4	0.2	2	8	0.1	5	8

Table 14: Parameter values for the resource-predator-prey model. All times are in time steps of the model.

The model is meant to simulate a resource-predator-prey ecology in order to probe biological and mathematical mechanisms that might emerge from a spatially extended system. It is not meant to be biologically exact. For instance all the species are moving and reproducing on similar spatial and time scales. This would not always be true and would be an important, and interesting, refinement to the model to if this were not the case.

8.3 Spatial scale.

The importance of spatial scale in ecology is fundamental. The results one obtains both from models and field studies is influenced by the scale at which we observe and and make our measurements, (e.g. Sugihara, Grenfell and May 1991). For many ecological systems, if they are sampled on too small a scale, then the data is noisy and dominated by stochastic effects. On the other hand, if they are sampled on too large a scale then the interesting effects are averaged out. We wish to address the question of how to choose the scale at which to monitor the system. We show that in our system there is a clearly-defined intermediate scale on which to observe the dynamics which gives us the maximum amount of ecological information about the system and its dynamics relative to the amount of data and work necessary to describe it. This scale is an inherent

characteristic of the ecosystem and is also the appropriate scale of measurement.

We investigate the effects of scale in our model by simply increasing or decreasing a window of size V (where $V = l^2$ is the number of sites in the observation window and $l \leq L$) through which we are observing and making measurements of the model. Without loss of generality we can fix the top corner of our window as the top corner of the lattice. A typical observable in biological systems is the population size of a species. Studying how this evolves is the study of the population dynamics. After iterating the model for a number of time steps so that the system has settled onto its attractor, we count the number of cells in our window of each particular species to give the species' population size. We then output the numbers at each time-step; thus generating a time-series, x_t. In the specific discussion of our model this species is usually the resource.

When l is very small (say $l < l_1$) we are observing on a scale smaller than the natural scale of the dynamics. The dynamics are thus dominated by stochastic fluctuations. These take place at a scale of the order of a few pixels as it is on this scale that randomness enters the model. We expect that as l increases, the decrease in the relative size of the stochastic fluctuations scales as $V^{-\frac{1}{2}} = l^{-1}$.

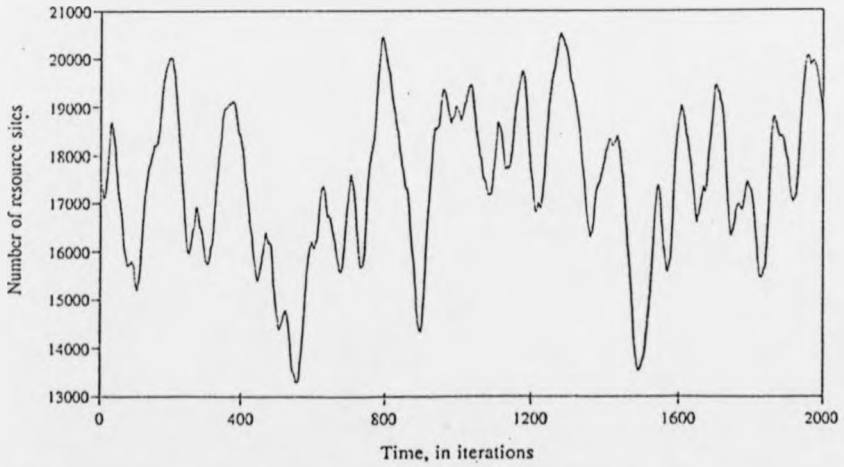


Figure 35: A time series of the number of resource sites. The lattice size is 150x150.

On the other hand, at intermediate scales $l_1 < l < l_2$ the motion we see is dominated by deterministic oscillations in the species numbers x_i. Figure 35 is a typical time-series of the numbers of resource. However, when l is very large (say $l > l_2$), distant parts of the lattice begin to act independently or become uncoupled. The result of this is a decoupling of the phases of the oscillations which results in an averaging effect. Thus as l is increased an averaging effect occurs so that the size of these oscillations decreases at a rate proportional to $V^{-\frac{1}{2}}$ and the dynamics will tend to a fixed point. We can use this fact in order to identify a spatial scale L_s above which distant parts of the lattice are acting independently.

For two variables X_i and Y_i the variance of their sum is equal to: $Var(X_i + Y_i) = Var(X_i) + Var(Y_i) + 2/N \sum_{i=1}^{N}(X_i - \bar{X}) * (Y_i - \bar{Y})$: the sum of the variances plus the correlation between the two variables. When X and Y are independent then this correlation is zero and $Var(X_i + Y_i) = Var(X_i) + Var(Y_i)$. When the window size is very large we expect that it is a conglomeration of d independent patches, $(X_i, ..., X_d)$, each of size V_s. For a size V we expect $d = V/V_s$ patches each with variance $Var(V_s)$. Thus the variance of the whole window, V, is:

$$Var(V) = Var(X_i, ..., X_d) = \sum_{i=1}^{d} Var(X_i) + \sum_{i,j=1}^{d} Cor(X_i, X_j),$$

where $Cor(X_i, X_j)$ is the correlation between the patches i and j. The patches are acting independently and so are uncorrelated except for the fact that they are all acting out dynamics on the same attractor. Hence there will be some small correlation between the patches related to the attractor dynamics within a patch. Hence:

$$Var(V) = \frac{V}{V_s} Var(V_s) + \frac{V^2}{V_s^2} Cor(V_s),$$

where $Cor(V_s)$ is the correlation between patches of size V_s due their acting out similar dynamics. The important point being that above the scale V_s the variance scales very differently with the window size to when $V < V_s$. The scaling above V_s should only come from the correlations due to patches being on the same attractor. In figure 36 we plot

$Var(V)/V$ against V. We see a clear change in the scaling behaviour at $V = 10,000$. This is the scale V_s.

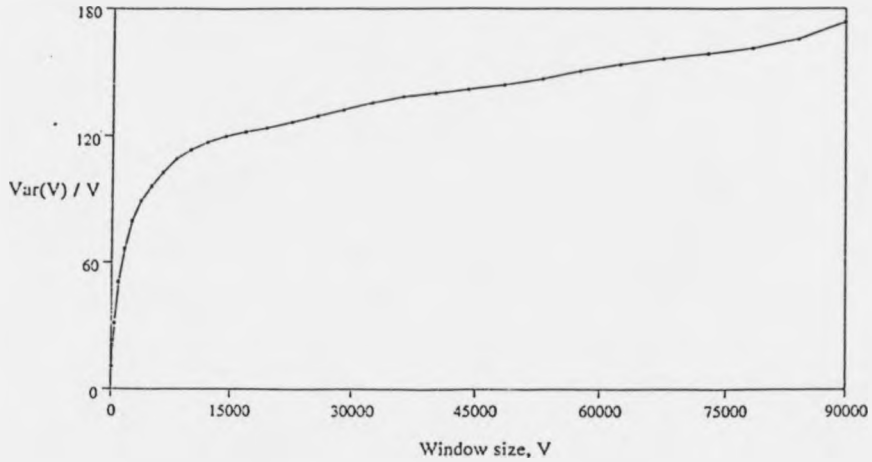


Figure 36: The variance of the resource time series as viewed through a window of size V. The lattice size is 300x300.

8.4 Determinism, dynamics and dimensionality.

After letting initial transients die away the model appears to settle down onto an attractor, Λ. This is a set of states such that:

(i) as $t \rightarrow \infty, S_t \rightarrow \Lambda$

(ii) once the system is on the attractor it doesn't leave it, i.e. if $S_t \in \Lambda$ then $S_{t'} \in \Lambda, \quad \forall \quad t' > t.$

The fact that attractors exist in cellular automata is well-known (e.g. [116]). It is important because only the attractor of the model needs to be studied since the automaton will always settle onto it. The spatial structure of the states in the attractor is quite complex, non-homogeneous and time-dependent. Resource is patchily distributed with prey associated with the edges of patches and predators with prey.

Locally the model is stochastic as the rules governing the transition from one state to another are probabilistic. We wish to show that, notwithstanding this, the population dynamics are governed by a deterministic function. By deterministic we mean that there is some time-independent functional relationship between the present state, x_t and the past states $x_{t-1}, x_{t-2}...$:

$$x_t = F(x_{t-E\tau}, x_{t-(E-1)\tau}, ..., x_{t-\tau}), \tag{57}$$

where E is an integer known as the *embedding dimension*, and τ the *delay time*. We make no attempt to determine the function F, only to demonstrate that such a function must exist. As well as the deterministic component to the dynamics there is also present a small noise component, ξ_t, where ξ_t is a random variable:

$$x_t = F(x_{t-E\tau}, x_{t-(E-1)\tau}, ..., x_{t-\tau}) + \xi_{t-1}. \tag{58}$$

However at intermediate length scales this noise is small and the dynamics are dominated by the deterministic component.

By definition determinism means that given some state the system will evolve in a fixed and predictable manner. If the system returns to this state then it will evolve in the same manner as before. To test this in our time-series we first form time delayed vectors:

$$\mathbf{x}_t = (x_{t-E\tau}, x_{t-(E-1)\tau}, ..., x_{t-\tau}).$$

Choosing some state at time t, $\mathbf{x}_t$, we find a previous state, $\mathbf{x}_{t-t'}$, very close (in the Euclidean sense) to our original state, i.e $||\mathbf{x}_t - \mathbf{x}_{t-t'}|| < \varepsilon$. We check that, as in a deterministic system, the evolution of these two states will be very similar.

In our case the deterministic part of the population dynamics are chaotic. This means that nearby states diverge apart from each other exponentially fast at a rate determined by the Liapunov exponent. In figure 37 we show the typical divergence of two nearby states. The exponential divergence is clear. We interpret this as further evidence of

determinism since a stochastic system is unlikely to produce such a pattern. In figure 38 we show the average of the log of the divergence calculated over all vectors in the time series. This is a straight line indicating exponential divergence. Note that the best-fit straight line has a correlation coefficient close to one. We only expect to see exponential divergence at small scales in the phase space, below the scale at which nonlinearities are affecting divergences. This system is highly nonlinear and so this scale, although it will vary at different points on the attractor, is relatively small. Hence we only plot the divergence for a short time.

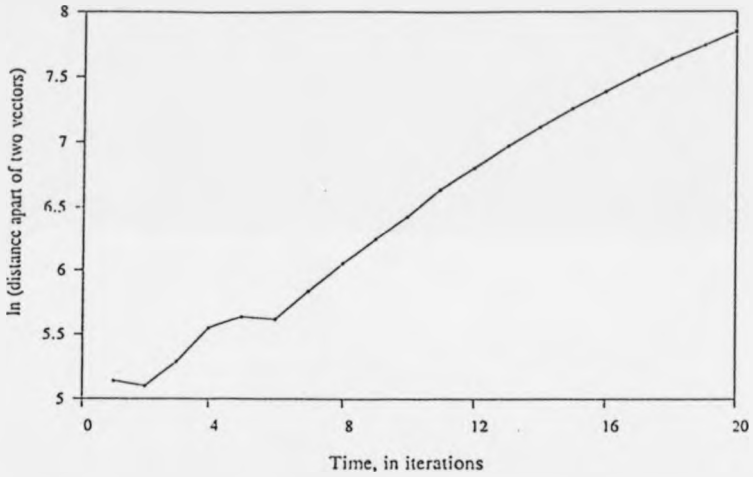


Figure 37: The typical divergence of two nearby states in the resource time series. The embedding dimension, $E = 4$ and the time delay $\tau = 5$. The lattice size is 100x100.

Further evidence comes from the dimension of the time-series. A stochastic system has infinite dimension as the noise tends to fill out the state space. Thus low-dimensionality is evidence for a deterministic system.

Given our embedded vectors, $\mathbf{x}_t$, we wish to calculate the dimensionality of this embedded space. To do this we need to know the number of linearly combined orthogonal vectors needed to fully describe the dynamics in this space. In order to investigate this we make use of a technique introduced into dynamical systems theory by Broomhead and King [10] based on ideas from singular systems theory [8].

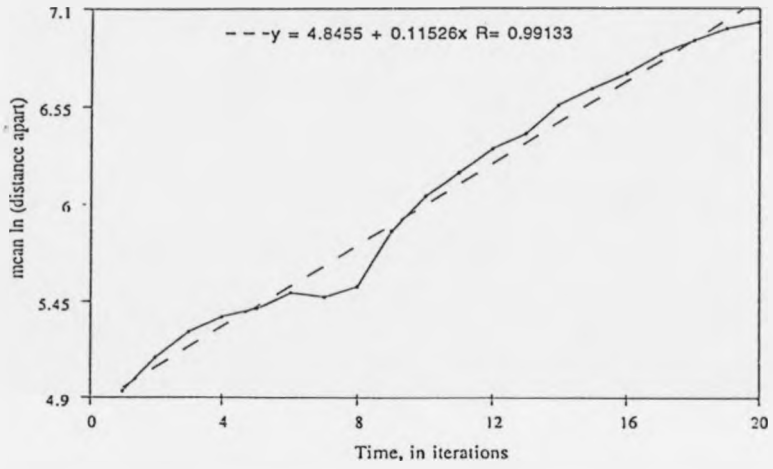


Figure 38: The logarithm of the average divergence of vectors in the resource time series. The embedding dimension, E, used is 4 and the time delay $\tau = 5$. The length of the time series is 10,000 points. The lattice size is 100x100. The figure was calculated by taking the first vector in the time series and locating the nearest neighbour, in the Euclidean sense. Their distance apart at time t is recorded. We then take the next vector in the time series and so on until the end. The average of the log of the distances apart at various times t is then plotted. A geometric average such as this is appropriate when trying to estimate Liapunov exponents.

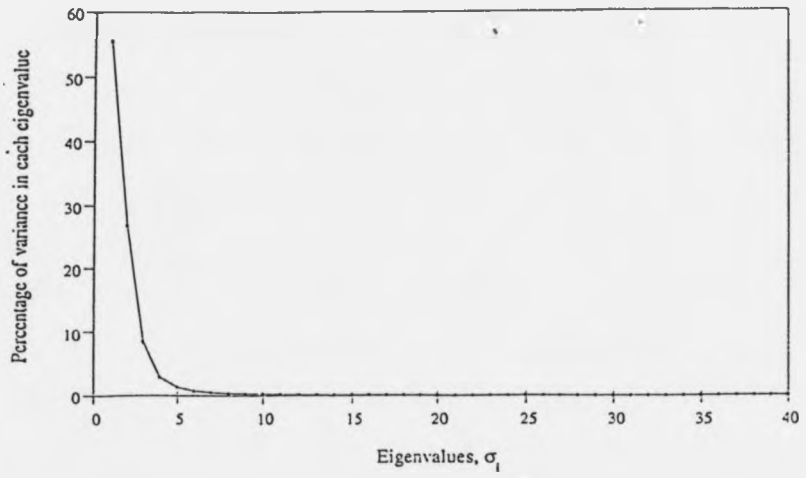


Figure 39: The normalised singular spectrum of the resource time series from a lattice size 150x150. The window size used in the SVD procedure was 40, the number of points was 10,000 and the time delay, $\tau = 1$.

If $\mathbf{x}_t$ is an embedded time series as above (where τ and E are suitably chosen) we let $\mathbf{Y}$ be the matrix whose i'th row is the vector $\mathbf{x}_i$ for $i = 0, 1, ..., N$ and let the trajectory matrix, $\mathbf{X} = N^{-\frac{1}{2}}\mathbf{Y}$. If all data points are used $N = N_T(E-1)$ where N_T is the number of data points in our original time series. In order to centre the mass at the origin we subtract the mean of each column from each element of the column. The singular value decomposition of this matrix is then given by $\mathbf{X} = \mathbf{S\Sigma C^T}$ where S and $\mathbf{C^T}$ contain the left and right singular vectors and the entries of the diagonal matrix $\mathbf{\Sigma}$ are the singular values $\sigma_1 > \sigma_2 > ..... > \sigma_E$. The singular vectors $e_1, e_2, ..., e_E$ form an orthogonal basis set that spans the embedding space. Each singular value is the mean square projection of the trajectory onto the corresponding singular vector. Hence the dimension of the embedding space is given by the number of non-zero singular values.

In the presence of noise, however, all singular values will be non-zero as the noise, as mentioned above, tends to fill out the state space. By inspecting the singular spectrum we can distinguish the singular values associated with a deterministic component and those associated soley with noise. Those associated with noise will tend to show up as

a constant noise floor. Figure 39 is a typical singular spectrum of a time-series of the resource. We see that 94% of the variance of the time-series is captured in the first four eigenvalues. This means that virtually all the dynamics in the original signal are occurring in a four-dimensional embedded space. Such a result is a strong indicator of determinism. It is also surprising considering the high-dimensionality of the original spatial system. We can project the time-series onto the first four eigenvectors in order to reconstruct the time-series minus some noise. Recall the matrix of embedded vectors and its decomposition, $\mathbf{X} = \mathbf{S\Sigma C^T}$. We can decompose the matrix of singular vectors, $\mathbf{C}$ into two subspaces:

$$\mathbf{C} = \mathbf{C}^D + \mathbf{C}^\epsilon, \qquad where \qquad \mathbf{C}^D = [c_1, ..., c_d, 0, ..., 0],$$
$$and \qquad \mathbf{C}^\epsilon = [0, ..., 0, c_{d+1}, ..., c_E].$$

Then:

$$\mathbf{XC}(= \mathbf{S\Sigma}) = \mathbf{XC}^D + \mathbf{XC}^\epsilon.$$
$$\Rightarrow \quad \mathbf{X} = \mathbf{XC}^D\mathbf{C}^T + \mathbf{XC}^\epsilon\mathbf{C}^T.$$

Thus d is the number of nonzero eigenvalues associated with the deterministic component of the signal and $\mathbf{X}^D = \mathbf{XC}^D\mathbf{C}^T$ is the projection onto the subspace associated with the deterministic dynamics. Each row, $\mathbf{p}_i$, of $\mathbf{X}^D$ is a d-dimensional vector which gives us a characterisation of the state of the system at time $t = i$. We show the reconstructed signal (the first column of $\mathbf{X}^D$) on top of the original in figure 40 and note its similarity to the original.

8.5 Detecting change.

An essential factor in the management of ecosystems is the detection of change. We outline a technique to detect long-term change or drift in the parameters governing the dynamics in an ecosystem. Drift can be detected at orders of 0.1% per period. The

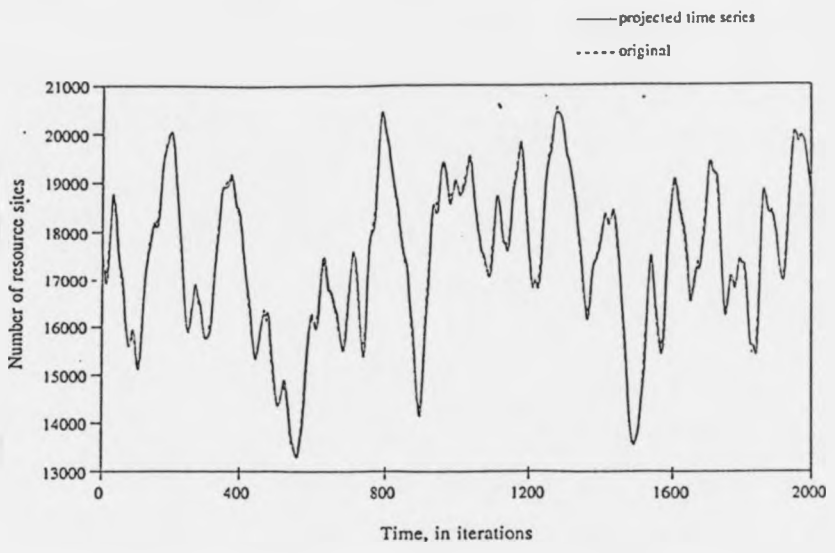


Figure 40: The resource time series projected onto the four eigenvectors with the highest eigenvalues (the first column of $\mathbf{X}^D$). The parameters in the SVD procedure are as in figure 39. The original time series is shown for comparison.

detection of change is also important in the analysis of ecological data. Nearly all data analysis techniques assume stationarity in the time series being analysed.

We make use of a technique introduced by Eckmann, Kamphosrt and Ruelle [24] for detecting non-stationarity in data sets. This is based on *recurrence plots*. Given a time series we form embedded vectors, $\mathbf{x}_t$, as in section 8.4. We then look for nearby vectors, $\| \mathbf{x}_t - \mathbf{x}_{t'} \| < \varepsilon$. This is done for all vectors and gives us a measure, $\nu(\varepsilon)$, on the recurrences in the time-series:

$$\nu(\varepsilon) = \frac{1}{N^2} \sum_{i,j=1}^{N} \Theta\{\varepsilon > \| \mathbf{x}_i - \mathbf{x}_j \|\},$$

where Θ is the Heaviside step function which is equal to zero for the negative argument and one otherwise. N is the number of embedded vectors and $|i - j| > 10$. Vectors on the same orbit very close in time will tend to be close together in distance. However these are not true recurrences and should not be included. Hence the requirement that the time difference between two vectors is at least 10, $|i - j| > 10$.

We consider two different problems. The first is when we have a continuous time

series from an ecology and wish to detect change occurring within it. In this case we look for nearby vectors, as above, but this time record the time difference $|t - t'|$ between the two. This is done for all vectors and we then plot the number of vectors found, $N_v(s)$, a time $s = |t - t'|$ apart. We plot this for the Lorenz equations in figure 41(a). We see that the time of recurrences is spread evenly. Compare this to figure 41(b) which is the plot when there is a 0.1% drift per unit time in the parameter ρ. We see that there are far fewer recurrences and that most of the recurrences are a short time apart. We emphasise the distribution of recurrences by plotting $D(u) = \sum_{s=u}^{N} N_v(s)$. For the stationary time series we see a linear decrease in $D(u)$ reflecting the even spread of $N_v(s)$. For the time series with drift we see an exponential decrease in $D(u)$. We plot this in figure 42.

The second problem is when we have two different time series and we wish to identify if there is a significant difference between the two. Using the measure $\nu(\varepsilon)$ the problem is reduced to comparing two measures and looking for statistical significance between the two. An improvement on this is to take cross-recurrences. We first calculate the measure $\nu(\varepsilon)$ for one of the time series. We then calculate $\nu(\varepsilon)$ for recurrences that occur between the two, i.e. $\mathbf{x}_i$ is taken from the first time series but recurrences $\mathbf{x}_j$ are taken from the second time series.

We show two graphs for the resource-predator-prey artificial ecology. Figure 43(a) is the internal recurrence plot plotting $N_v(s)$ against $s = |t - t'|$ for a time series using the parameter values as in table 14. Figure 43(b) is the cross-recurrence plot between the first time series and one where $g = 0.6$. This figure clearly indicates substantial change between the two.

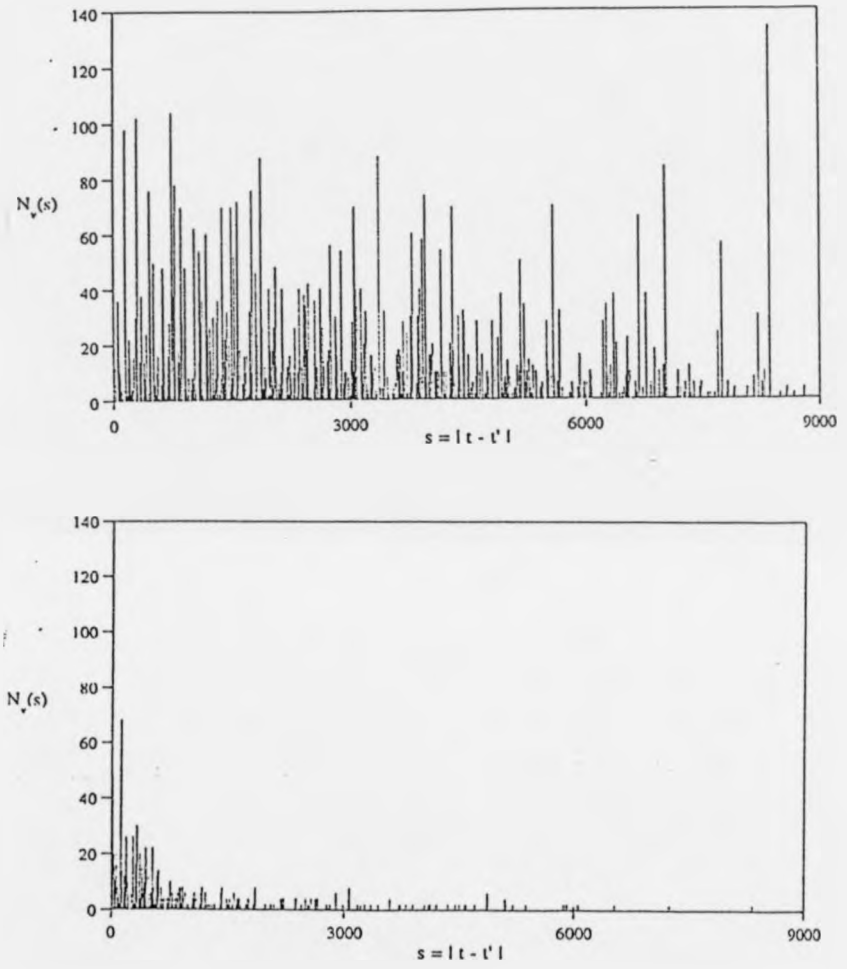


Figure 41: (a) The recurrence plot for the Lorenz time series using $\sigma = 16.0$, $\rho = 45.92$, $b = 4.0$ and sample time $\alpha = 0.05$. The number of points is 20,000. The embedding dimension used is 5, $\tau = 1$ and $\epsilon = 0.2$. (b) The recurrence plot for the Lorenz time series with drift of 0.1% in the parameter ρ, i.e. $0.001\rho\alpha$ was added to $\rho(t)$ each time step. The other parameters are as in (a).

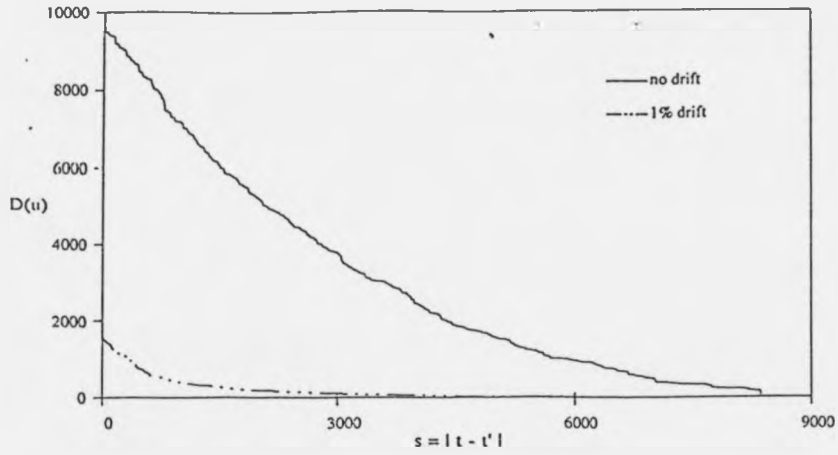


Figure 42: The distribution of the Lorenz recurrences, $D(u)$, with time. The parameter values are as in figures 41(a) and (b).

References

[1] Abrams, P.A. 1986. Adaptive responses os predators to prey and prey to predators: the failure of the arms-race analogy. *Evolution* **40** (6) 1229-1247.

[2] Anderson, R.M. and May, R.M. 1982. Directly transmitted infectious diseases: control by vaccination. *Science.* **215** 1053-1060.

[3] Anderson, R.M. and May, R.M. 1990 Infectious diseases of humans. Oxford University Press.

[4] Aoki, K. 1984. A quantitative genetic model of two-policy games between relatives. *J. Theor. Biol.* **109** 111-126.

[5] Aron, J.L. and Schwartz, I.B. 1984. Seasonality and period-doubling bifurcations in an epidemic model. *J. Theor. Biol.* **110** 665-679.

[6] Bartlett, M.S. 1957. Measles periodicity and community size. *J. Roy. Statist. Soc. A* **120** 48-60.

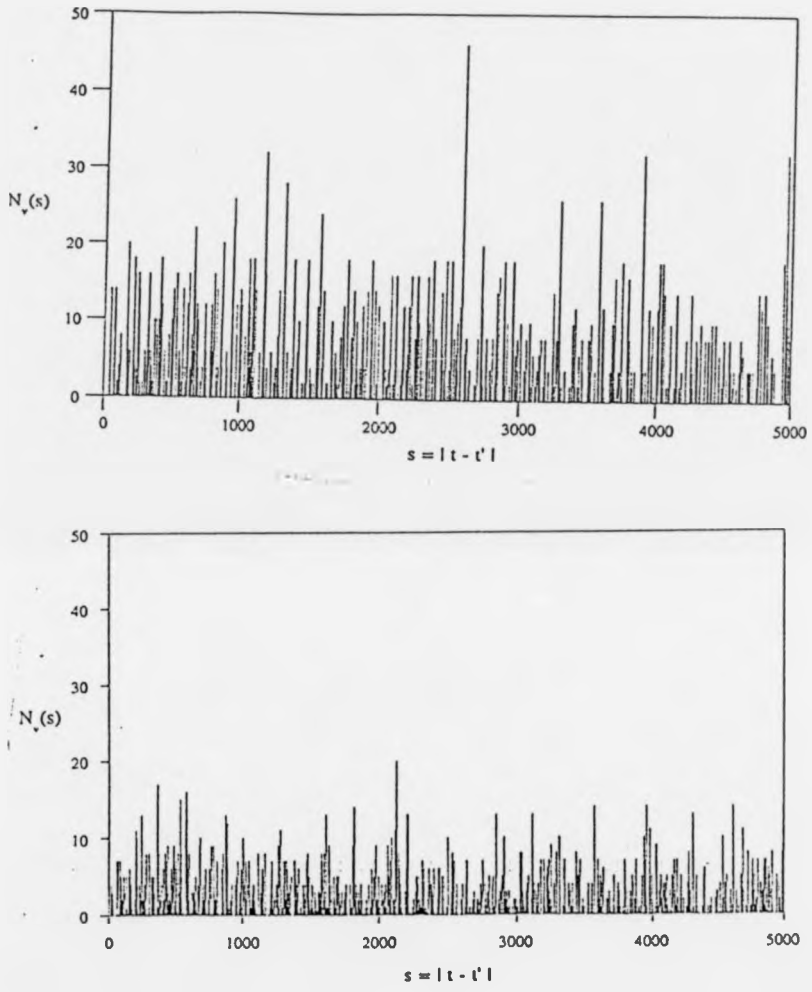


Figure 43: (a) The recurrence plot for the resource time series using the parameter values from table 14. The lattice size is 150x150, the time series length is 10,000 points, $E = 4$, $\tau = 5$ and $\varepsilon = 80$. (b) The cross-recurrence plot for the resource time series using $g = 0.4$ and $g = 0.6$ and the rest of the parameter values from table 14 and figure 43(a).

[7] Benettin, G., Galgani, L., Giorgilli, A. and Strelcyn, J.M. 1978. Tous les nombres de Liapounov sont effectivement calculables. *C. R. Acad. Sci. Paris* **286A** 431.

[8] Bertero, M. and Pike, E.R. 1982. Resolution in diffraction-limited imaging, a singular value analysis I: The case of coherent illumination. *Opt. Acta. 29* p.727

[9] Bishop, D.T. and Cannings, C. 1978. A generalised war of attrition. *J. Theor. Biol.* **70** 85-124.

[10] Broomhead, D.S. and King, G.P. 1986 Extracting qualitative dynamics from experimental data. *Physica* **20D** 217-236.

[11] Brown, J.S. and Vincent, T.L. 1987. Coevolution as an evolutionary game. *Evolution* **41** No. 1 66-79.

[12] Carlile, D.W., Skalski, J.R., Batker, J.E., Thomas, J.M. and Cullinan, V.I. 1989. Determination of ecological scale. *Landscape Ecology* **2** (4) 203-213.

[13] Carol, L. Through the looking glass.

[14] Casdagli, M. 1989. Nonlinear prediction of chaotic time series. *Physica D* **35** 335-356.

[15] Casdagli, M. 1992. Chaos and deterministic *versus* stochastic non-linear modelling. *J. Roy. Stat. Soc. B* **54** 303.

[16] Casdagli, M., Eubank, S., Farmer, J.D. and Gibson, J. 1991. State space reconstruction in the presence of noise. *Physica D* **51** 52-98.

[17] Cressman, R. 1990. Strong Stability and Density-dependent Evolutionarily Stable Strategies. *J. theor. Biol.* **145** 319-330.

[18] Crutchfield, J., Farmer, D., Packard, N., Shaw, R. and Jones,G. and Donnelly, R.J. 1980. Power spectral analysis of a dynamical system. *Phys. Lett.* **76A** No. 1 1-4.

[19] Dawkins, R. and Krebbs 1979. Arms races between and within species. *Proc. Roy. Soc. B* **205** 489-511.

[20] Dietz and Schenzle, D. 1990. In em discussion on Chance or Chaos? (by Bartlett, M.S.) *J. Roy. Statist. Soc. A* **153** 338.

[21] Dodson, M.M. 1975. Quantum evolution and the fold catastrophe. *Evolutionary theory* **1** 107-118.

[22] Durrett, R. 1988. Crabgrass, measles annd gypsy moths: an introduction to interacting particle systems. *Mathematical Intelligencer* **10** 37-47.

[23] Eckmann, J.-P. and Ruelle, D. 1985. Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.* **57** 617.

[24] Eckamnn, J.-P., Kamphorst, S.O. and Ruelle, D. 1987. Recurrence plots of dynamical systems. Preprint: Universite de Geneve.

[25] Eckmann, J.-P. and Ruelle, D. 1992. Fundamental limitations for estimating dimensions and Liapunov exponents in dynamical systems. *Physica D.* **56** 185-187.

[26] Eldridge, N. and Gould, S.J. 1972. Punctuated equilibria: an alternative to phyletic gradualism. In: *Models in Paleobiology* (ed. T. J. M. Schopf), Freeman, Cooper, San Francisco.

[27] Ellner, S. 1991. Detecting low-dimensional chaos in population dynamics: a critical review. In *Chaos and Insect Ecology* eds. Logan, J.A. and Hain, F.P. p63-90.

[28] Fahrig, L. and Paloheimo, J. 1988. Determinants of local population size in patchy habitat. *Theor. Pop. Biol.* **34** 194-213.

[29] Fahrig, L. and Paloheimo, J. 1988. Effect of spatial arrangement of habitat patches on local population size. *Ecology* **69** 468-475.

[30] Farmer, J.D. and Sidorowich, J.J. 1987. Predicting chaotic time series. *Phys. Rev. Lett.* **59** No. 8 845-848.

[31] Fine, P.E.M. and Clarkson, J.A. 1982. Measles in England and Wales-I: an analysis of factors underlying seasonal patterns. *Int. J. Epidem.* **11** No. 1 5-14.

[32] Gardener, M. 1970. *Scientific American* **223** (4) 120-124.

[33] Gollub, J.P. and Swinney, H.L. 1975. *Phys. Rev. Lett.* **35** 927.

[34] Gorman, M. and el-Hamdi, M. 1991. Power spectra as a guide for identification of deterministic chaos. *preprint, University of Houston*

[35] Gould, S.J. 1987. Ever since Darwin. London: Burnett-Deutsch.

[36] Gould, S.J. and Eldridge, N. 1977. Punctuated equilibria: the tempo and mode of evolution reconsidered. *Paleobiology* **3** 115-151.

[37] Grafen, A. 1979. The hawk-dove game played between relatives. *Anim. Behav.* **27** 905-907.

[38] Grassberger, P. and Procaccia, I. 1983. *Phys. Rev. Let.* **50** 5 346.

[39] Grassberger, P. and Procaccia, I. 1983. Measuring the strangeness of strange attractors. *Physica* **D9** 189.

[40] Grenfell, B.T. 1992. Chance and chaos in measles dynamics. *J. Roy. Statist. Soc. B* **54** No. 2 1-16.

[41] Hamer, W.H. 1906. Epidemic diseases in England. *Lancet, i* 733-39.

[42] C. B. Harley, Learning the evolutionarily stable strategy. *J. theor. Biol.* **89**, (1981) 611-633

[43] Hassell, M.P., Comins, H.N. and May, R.M. 1991. Spatial structure and chaos in insect population dynamics. *Nature* **353** 255-258.

[44] Hastings, A. and Wolin, C.L. 1989. Within-patch dynamics in a metapopulation model. *Ecology* **70** 1261-1266.

[45] Hastings, A. 1990. Spatial heterogeneity and ecological models. *Ecology* **71** (2) 426-428.

[46] Hines, W.G.S. 1980. An evolutionary stable strategy model for randomly mating sexual diploid populations. *J. Theor. Biol.* **87** 379-384.

[47] Hines, W.G.S. 1982. Mutations, perturbations and evolutionary stable strategies. *J. Appl. Probab.* **19** 204-209

[48] Hines, W.G.S. and Maynard Smith J. 1979. Games between relatives. *J. Theor. Biol.* **79** 19-30.

[49] Kifer, Y. 1986. Ergodic theory of random transformations. Boston:Birkhauser.

[50] Kot, M., Graser, D.J., Truty, G.L., Schaffer, W.M. and Olsen, L.F. 1990. Changing criteria for imposing order. *Ecol. Model.* **43** 75-110.

[51] Langton, C.G. 1986. Studying artificial life with cellular automata. *Physica* **22D** 120-149.

[52] Lawlor, L.R. and Maynard Smith, J. 1976. The coevolution and stability of competing species. *Am. Nat.* **110** 79-99.

[53] Levin, S.A. 1992. The problem of pattern and scale in ecology. *Ecology* **73** (6) 1943-1967.

[54] London, W.P. and Yorke, J.A. 1973. Recurrent outbreaks of measles, chickenpox, and mumps. I: Seasonal variation in contact rates. *Am. J. Epidem.* **98**, 453-68.

[55] Lorenz, E.N. 1963. Deterministic non-periodic flow. *J. Atmospheric Sciences* **20** No. 2 130-141.

[56] MacArthur, R.H. 1962. Some generalized theorems of natural selection. *Proc. Nat. Acad. Sci.* **48** 1893-1897.

[57] MacArthur, R.H., and Wilson, E.O. 1967. The theory of island biogeography. Princeton University Press.

[58] Mandelbrot, B.B. 1977. *Fractals-Form, Chance and Dimension* Freeman, San Francisco.

[59] Marrow, P., Law, R. and Cannings, C. 1992. The coevolution of predator-prey interactions: ESSs and Red Queen dynamics. *Proc. R. Soc. Lond. B* **250** 133-141.

[60] May, R.M. 1976. Simple mathematical models with very complicated dynamics. *Nature* **261** 459-467.

[61] May, R.M. 1986. When two and two do not make four: nonlinear phenomena in ecology. *Proc. R. Lond. B* **228** 241-266.

[62] Maynard Smith, J. 1976. What determines the rate of evolution? *American Naturalist.*

[63] Maynard Smith, J. 1981. Will a sexual population evolve to an ESS? *American Naturalist* **117** 1015-18.

[64] Maynard Smith, J. 1982. Evolution and the theory of games. Cambridge University Press.

[65] Maynard Smith, J. and Parker, G.A. 1976. The logic of asymmetric contests. *Animal Behaviour* **24** 159-175.

[66] Maynard Smith, J. and Price, G.R. 1973. The logic of animal conflict. *Nature* **246** 15-18.

[67] Metz, J.A.J., Nisbet, R.M. and Geritz, S.A.H. 1992. How should we define 'fitness' for general ecological scenarios? *Tree* **7** (6) 198-202.

[68] Miramontes, O., Sole, R.V. and Goodwin, B.C. 1992. Collective behaviour on random-activated mobile cellular automata. *Physica* **D**.

[69] Mollison, D. 1977. Spatial contact models for ecological and epidemic spread. *J. Roy. Statist. Soc.* **B 39** 283-326.

[70] Nicholson, A.J. 1954. An outline of the dynamics of animal populations. *Aust. J. Zoo.* **2** 9-65.

[71] Nychka, D., Ellner,S., Gallant, A.R. and McCaffrey, D. 1992. Finding chaos in noisy systems. *J. Roy. Statist. Soc. B* **54** No. 2.

[72] Okubo, A. 1980. Diffusion and ecological problems: mathematical models. Springer, New York.

[73] Olsen, L.F. and Schaffer, W.M. 1990. Chaos versus noisy periodicity: alternative hypotheses for childhood epidemics. *Science* **249** 499-504.

[74] Olsen, L.F., Truty, G.L. and Schaffer, W.M. 1988. Oscillations and chaos in epidemics: a nonlinear dynamic study of six childhood diseases in Copenhagen, Denmark. *Theor. Pop. Biol.* **33** 344-370.

[75] Oseledec, V.I. 1968. A multiplicative ergodic theorem. Lyapunov characteristic numbers for dynamical systems. *Trans. Mosc. Math. Soc.* **19** 197.

[76] Packard, N.H., Crutchfield, J.P., Farmer, J.D. and Shaw, R.S. 1980. Geometry from a time series. *Physical Review Letters* **45** 712.

[77] Pimm, S.L. and Redfearn, A. 1988. The variability of population densities. *Nature* **334** 613-614.

[78] Pool, R. 1989. Is it chaos or is it just noise? *Science* **243** 25-28.

[79] Provenzale, A., Smith, L.A., Vio, R. and Murante, G. 1992. Distinguishing between low-dimensional dynamics and randomness in measured time series. *Physica D* **58** 31-49.

[80] Rand, D.A., Wilson, H.B. and McGlade, J.M. 1993. Dynamics and evolution: Evolutionarily stable attractors, invasion exponents and phenotype dynamics. To appear in Trans. Roy. Soc. Lon. **B**.

[81] Reeve, J.D. 1988. Environmental variability, migration and persistence in host-parasitoid systems. *American Naturalist* **132** 810-836.

[82] Reeve, J.D. 1990. Stability, variability, and persistence in host- parasitoid systems. *Ecology* **71** (2) 422-426.

[83] Reichert, S.E. 1978. Games spiders play: behavioural variability in territorial disputes. *Behav. Ecol. Sociobiol* **3** 135-162.

[84] Rosenzweig, M.L., Brown, J.S. and Vincent, T.L. 1987. Red Queens and ESS: the coevolution of evolutionary rates. *Evolutionary Ecology* **1** 59-94.

[85] Roughgarden, J. 1970. Density-dependent natural selection. *Ecology* **52** No. 3 453-468.

[86] Ruelle, D. and Takens, F. 1971. On the nature of turbulence. *Commun. Math. Phys.* **21** 21.

[87] Sano, M. and Sawada, Y. 1985. Measurement of the Liapunov Spectrum from a Chaotic Time Series. *Phys. Rev. Let.* **55** 10 1082.

[88] Schaffer, W.M. 1985. Non-Linear Dynamics in Epidemiological Systems. *IMA J. Math. Appl. Med. Biol.*

[89] Schaffer, W.M. and Kot, M. 1985. Nearly one dimensional dynamics in an epidemic. *J. Theor. Biol.* **112** 403-427.

[90] Schaffer, W.M., Olsen, L.F., Truty, G.L., Fulmer, S.L. and Graser, D.J. 1987. Periodic and chaotic dynamics in childhood infections. In *From chemical to biological organization* Eds. Marksis, M. et. al.

[91] Schaffer, W.M., Olsen, L.F., Truty, G.L. and Fulmer, S.L. 1990. The case for chaos in childhood epidemics. In *The ubiquity of chaos.* (ed. S Krasner) pp.136-166, AAAS, Washington.

[92] Schenzle, D. 1984. An age-structured model of pre- and post-vaccination measles transmission. *IMA J. Math. App. Med. Biol.* **1** 169-191.

[93] Schwartz, I.B. 1989. In *Biomedical modelling and simulation.* Eds. Eisenfeld, J. and Levine, D.S. p201.

[94] Segal, L.A. and Jackson, J.L. 1972. Dissipative structure:An explanation and an ecological example. *J. Theor. Biol.* **37** 545-559.

[95] Shimada, I. and Nagashine, T. 1979. A numerical approach to ergodic problems of dissipative dynamical systems. *Prog. Theor. Phys.* **61** No.(6) 1605.

[96] Slatkin, M. 1979. The evolutionary response to frequency- and density-independent interactions. *American Naturalist* **114** 384-98.

[97] Smale, S. 1967. Differentiable dynamical systems. *Bull. Amer. Math. Soc.* **73** 747.

[98] Smith, L.A. 1992. Identification and prediction of low dimensional dynamics. *Physica D* **58** 50-76.

[99] Smith, R.L. 1992. Estimating dimension in noisy chaotic time series. *J. Roy. Statist. Soc. B* **54**.

[100] Soper, H.E. 1929. The interpretation of periodicity in disease prevalence. *J. Roy. Statist. Soc.* **92** 34-61.

[101] Stanley, S.M. 1979. *Macroevolution; Pattern and Process.* Freeman, San Francisco.

[102] Steele, J.H. 1985. Comparison of marine and terrestrial ecological systems. *Nature* **313** 355-358.

[103] Stenseth, N.C. and Maynard Smith, J. 1984. Coevolution in ecosystems: Red Queen evolution or stasis. *Evolution* **38** No.4 870-880.

[104] Stone, L. 1992. Coloured noise or low-dimensional chaos? correlations between the patches equals the correlation within a patch. *Proc. R. Soc. Lond. B* **250** 77-81.

[105] Sugihara, G., Grenfell, B.T. and May, R.M. 1990. Distinguishing error from chaos in ecological time series. *Phil. Trans. R. Soc. Lond.***B 330** 235-251.

[106] Sugihara, G. and May, R.M. 1990. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature,* **344** 734-741.

[107] Takens, F. 1981. Detecting strange attractors in turbulence, *In Dynamical Systems and Turbulence, Warwick 1980, ed. D.Rand and L.-S. Young. Lecture notes in Mathematics: Springer-Verlag* p 366-381.

[108] Taylor, J. and Jonker, L. 1987. Evolutionarily Stable Strategies and Game Dynamics. *Math. Biosci.* **40** 145-156.

[109] Tel, T. 1990. Transient chaos. Preprint Forschungszentrum Jülich.

[110] Theiler, J., Eubank, S., Longtin, A., Galdrikian, B. and Farmer, J.D. 1992. Testing for nonlinearity in time series: the method of surrogate data. *Physica D* **58** 77-94.

[111] Treisman, M. 1981. Evolutionary limits to the frequency of aggression between related or unrelated conspecifics in diploid species with simple mendelian inheritance. *J. Theor. Biol.* **93** 97-124.

[112] Van Valen, L. 1973. A new evolutionary law. *Evolutionary Theory* **1** 1-30.

[113] Vincent, T.L. and Brown, J.S. 1986. Evolution under nonequilibrium dynamics. In Proceedings of the 5th International conference on Mathematical modelling. Pergamon, N.Y.

[114] Whitney, H. 1936. Differentiable manifolds. *Ann. Math.* **37** 645.

[115] Wolf, A., Swift, J.B., Swinney, H.L. and Vastano, J.A. 1985. Determining Lyapunov exponents from a time series. *Physica D* **16** 285-317.

[116] Wolfram, S. 1984. Cellular automata as models of complexity. *Nature* **311** 419-424.

[117] Zeeman, E.C. 1981. Dynamics of the evolution of animal conflicts. *J. Theor. Biol.* **89** 249-270.

[118] Zeeman, E.C. 1979. Population dynamics from game theory. In *Proc. Conf. on Global Theory of Dynamical Systems.* Northwestern, Evanston. pp. 471-497.

[119] Zeeman, E.C. 1992. Evolution and catastrophe theory. In *Understanding Catastrophe.* (ed. J. Bourrian), CUP, pp. 83-101.

April 11, 1994