An Introduction to Mathematical Modelling

Glenn Marion, Bioinformatics and Statistics Scotland
Given 2008 by Daniel Lawson and Glenn Marion

2008

Contents

List of Figures

1 Introduction

1.1 What is mathematical modelling?

Models describe our beliefs about how the world functions. In mathematical modelling, we translate those beliefs into the language of mathematics. This has many advantages

1. Mathematics is a very precise language. This helps us to formulate ideas and identify underlying assumptions.

2. Mathematics is a concise language, with well-defined rules for manipulations.

3. All the results that mathematicians have proved over hundreds of years are at our disposal.

4. Computers can be used to perform numerical calculations.

There is a large element of compromise in mathematical modelling. The majority of interacting systems in the real world are far too complicated to model in their entirety. Hence the first level of compromise is to identify the most important parts of the system. These will be included in the model, the rest will be excluded. The second level of compromise concerns the amount of mathematical manipulation which is worthwhile. Although mathematics has the potential to prove general results, these results depend critically on the form of equations used. Small changes in the structure of equations may require enormous changes in the mathematical methods. Using computers to handle the model equations may never lead to elegant results, but it is much more robust against alterations.

1.2 What objectives can modelling achieve?

Mathematical modelling can be used for a number of different reasons. How well any particular objective is achieved depends on both the state of knowledge about a system and how well the modelling is done. Examples of the range of objectives are:

1. Developing scientific understanding

   - through quantitative expression of current knowledge of a system (as well as displaying what we know, this may also show up what we do not know);

2. test the effect of changes in a system;

3. aid decision making, including

   (i) tactical decisions by managers;
   (ii) strategic decisions by planners.

1.3 Classifications of models

When studying models, it is helpful to identify broad categories of models. Classification of individual models into these categories tells us immediately some of the essentials of their structure.

One division between models is based on the type of outcome they predict. Deterministic models ignore random variation, and so always predict the same outcome from a given starting point. On the other hand, the model may be more statistical in nature and so may predict the distribution of possible outcomes. Such models are said to be stochastic.

A second method of distinguishing between types of models is to consider the level of understanding on which the model is based. The simplest explanation is to consider the hierarchy of organisational structures within the system being modelled. For animals, one such hierarchy is:

High	*herd*
	individual
	organs
	cells
Low	*molecules*
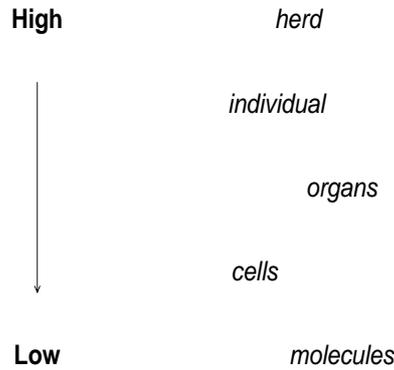
A model which uses a large amount of theoretical information generally describes what happens at one level in the hierarchy by considering processes at lower levels these are called *mechanistic models*, because they take account of the mechanisms through which changes occur. In *empirical models*, no account is taken of the mechanism by which changes to the system occur. Instead, it is merely noted that they do occur, and the model trys to account quantitatively for changes associated with different conditions.

The two divisions above, namely deterministic/stochastic and mechanistic/empirical, represent extremes of a range of model types. In between lie a whole spectrum of model types. Also, the two methods of classification are complementary. For example, a deterministic model may be either mechanistic or empirical (but not stochastic). Examples of the four broad categories of models implied by the above method of classification are:

	Empirical	Mechanistic
Deterministic	Predicting cattle growth from a **regression relationship** with feed intake	Planetary motion, based on Newtonian mechanics (**differential equations**)
Stochastic	Analysis of **variance** of variety yields over sites and years	Genetics of small populations based on Mendelian inheritance (**probabalistic equations**)

One further type of model, the system model, is worthy of mention. This is built from a series of sub-models, each of which describes the essence of some interacting components. The above method of classification then refers more properly to the sub-models: different types of sub-models may be used in any one system model.

Much of the modelling literature refers to 'simulation models'. Why are they not included in the classification? The reason for this apparent omission is that 'simulation' refers to the way the model calculations are done - i.e. by computer simulation. The actual model of the system is not changed by the way in which the necessary mathematics is performed, although our interpretation of the model may depend on the numerical accuracy of any approximations.

1.4   Stages of modelling

It is helpful to divide up the process of modelling into four broad categories of activity, namely *building*, *studying*, *testing* and *use*. Although it might be nice to think that modelling projects progress smoothly

from building through to use, this is hardly ever the case. In general, defects found at the studying and testing stages are corrected by returning to the building stage. Note that if any changes are made to the model, then the studying and testing stages must be repeated.

A pictorial representation of potential routes through the stages of modelling is:

Building

Studying

Testing

Use
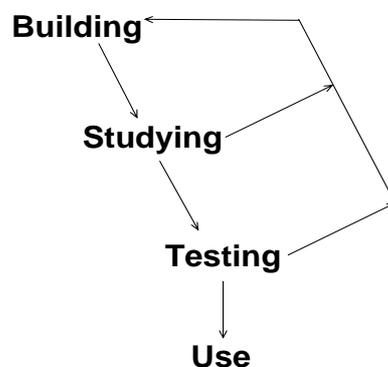
This process of repeated iteration is typical of modelling projects, and is one of the most useful aspects of modelling in terms of improving our understanding about how the system works.

We shall use this division of modelling activities to provide a structure for the rest of this course.

2 Building models

2.1 Getting started

Before embarking on a modelling project, we need to be clear about our objectives. These determine the future direction of the project in two ways.

Firstly, the level of detail included in the model depends on the purpose for which the model will be used. For example, in modelling animal growth to act as an aid for agricultural advisers, an empirical model containing terms for the most important determinants of growth may be quite adequate. The model can be regarded as a summary of current understanding. Such a model is clearly of very limited use as a research tool for designing experiments to investigate the process of ruminant nutrition.

Secondly, we must make a division between the system to be modelled and its environment. This division is well made if the environment affects the behaviour of the system, but the system does not affect the environment. For example, in modelling the growth of a small conifer plantation to predict timber yields, it is advisable to treat weather as part of the environment. Its effect on growth can be incorporated by using summary statistics of climate at similar locations in recent years. However, any model for the growth of the world's forests would almost certainly have to contain terms for the effect of growth on the weather. Tree cover is known to have a substantial effect on the weather via carbon dioxide levels in the atmosphere.

2.2 Systems analysis

2.2.1 Making assumptions

Having determined the system to be modelled, we need to construct the basic framework of the model. This reflects our beliefs about how the system operates. These beliefs can be stated in the form of underlying assumptions. Future analysis of the system treats these assumptions as being true, but the results of such an analysis are only as valid as the assumptions.

Thus Newton assumed that mass is a universal constant, whereas Einstein considered mass as being variable. This is one of the fundamental differences between classical mechanics and relatively theory. Application of the results of classical mechanics to objects traveling close to the speed of light leads to inconsistencies between theory and observation.

If the assumptions are sufficiently precise, they may lead directly to the mathematical equations governing the system.

In population studies, a common assumption is that, in the absence of limiting factors, a population will grow at a rate which is proportional to its size. A deterministic model which describes such a population in continuous time is the differential equation.

$$\frac{dp}{dt} = ap$$

where $p(t)$ is population size at time t, and a is a constant. Solution of this equation by integration gives

$$p(t) = p(0)e^{at}$$

where $p(0)$ is population size at time zero. According to this solution, populations grow in size at an exponential rate.

Clearly, not all populations grow exponentially fast. Since the differential equation arose from an interpretation of the assumption, we must look to the assumption for an explanation for this discrepancy. In this case, the explanation is the qualifier "in the absence of limiting factors". Most

natural populations are subject to constraints such as food supply or habitat which restrict the range of sustainable population sizes.

It is important that all assumptions are stated clearly and concisely. This allows us to return to them later to assess their appropriateness.

Another assumption which we made to obtain the differential equation was that growth takes place continuously. If the population consisted of discrete generations, we would have used the difference equation

$$d_{i+1} = bd_i \,,$$

where d_i is the size of the ith generation. This has solution

$$d_i = d_0 b^i$$

where d_0 is the initial population. Note that the solutions of the differential and difference equations can coincide at time $t = i$ if $d_0 = p(0)$ and $b = e^a$

Yet another assumption we have made is that the population behaves according to a deterministic law. Typically when populations are large we expect randomness or variation to be limited, although as we will see in Practical 2.1, this can be misleading. However, when populations are small we intuitively expect stochastic variability to be important. This forces us to consider births separately from deaths. A model for population change between times t and $t + \delta t$ is:

Event	Effect on population, p	Probability of Event
Birth	$p(t + \delta t) = p(t) + 1$	$cp(t)\delta t$
Death	$p(t + \delta t) = p(t) - 1$	$fp(t)\delta t$
No change	$p(t + \delta t) = p(t)$	$1 - cp(t)\delta t - fp(t)\delta t$

This description of the model shows how to (approximately) simulate the stochastic model: choose a sufficiently small time step δt (such that all the probabilities are less than one); and then choose one of the possible events with the probability shown. This last step is achieved by drawing a random number y uniformly on $[0, 1]$ (many packages have such a random number generator). A birth occurs if $y < cp(t)\delta t$, a death if $y < cp(t)\delta t + fp(t)\delta t$, otherwise no event occurs. Clearly this can be extended to any number of event types (see e.g. Renshaw 1991 for more details). This stochastic model has the same expected value as the differential equation model if $(c - f) = a$. In general, even this degree of correspondence between deterministic and stochastic models is hard to achieve.

Note: In general the correspondence between deterministic and stochastic population models is rather simple. The model itself is define by the population size p and the rates, for example birth rate, b(p), and death rate, d(p). In the deterministic case these define the rate of change of the population size n thus,

$$\frac{dp(t)}{dt} = b(p) - d(p) \,,$$

In the stochastic case these define (for suitably small δt) the probabilities of birth and death *events*, namely

Event	Effect on population, p	Probability of Event
Birth	$p(t + \delta t) = p(t) + 1$	$b(p)\delta t$
Death	$p(t + \delta t) = p(t) - 1$	$d(p)\delta t$
No change	$p(t + \delta t) = p(t)$	$1 - b(p)\delta t - d(p)\delta t$

A third choice we have made in formulating our simple population growth model is to assume the population is uniformly distributed in space, and has no contact with other populations. If we were

interested in the spatial distribution of our population then we would need to introduce a spatially explicit model with rates describing the movement of individuals from one area to another. A common approach to this problem is illustrated in Figure 1. Here our nonspatial model is taken to be one of a (large) number of *colonies* and then interactions (here movements of individuals) between them are specified. As we shall see later spatial heterogeneity can often qualitatively change model behaviour.
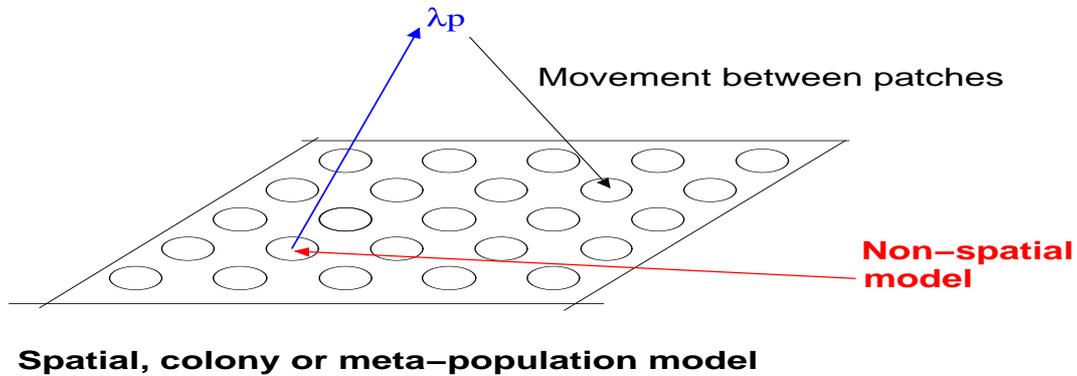
Figure 1: A schematic description of a spatial model

Having seen how many different assumptions have to be made to arrive at a simple model of population growth, we must proceed with extreme caution when formulating models of complicated systems. It is often wise to examine several different versions (e.g deterministic, stochastic, spatial ...) based on the same basic model as this improves confidence in any results derived from the model. Moreover as we shall see later in the exercises, relaxing certain assumptions, for example by adding stochasticity, can reveal flaws in model structure.

2.2.2 Flow diagrams

Where the system being modelled is more complex, we cannot simply jump from an assumption to an equation. We must be much more methodical, both when describing the system and when stating assumptions. Flow diagrams are a visual aid to this end. In their most basic form, they consist of a series of boxes linked by a network of arrows. The boxes represent physical entities which are present in the system, whilst the arrows represent the way these entities inter-relate. Commonly used symbols in flow diagrams are:

□ state or level variables

◯ source or sink

→ channel of material flow

--→ channel of information flow

⋈ control on rate of flow

There are many ways of conducting a logical analysis of a system as a prelude of drawing a flow diagram. Often, the most profitable method is to consider the transfer of a resource throughout the system. This works best when the transfer is subject to a conservation law. Properties which may be conserved include: energy; mass; atomic elements. Which is best for any application depends entirely on the nature of interactions taking place in the system. In population models one often considers compartments representing different sections of the population with flows between them. For example in the model of births and deaths considered above the flow into the population would correspond
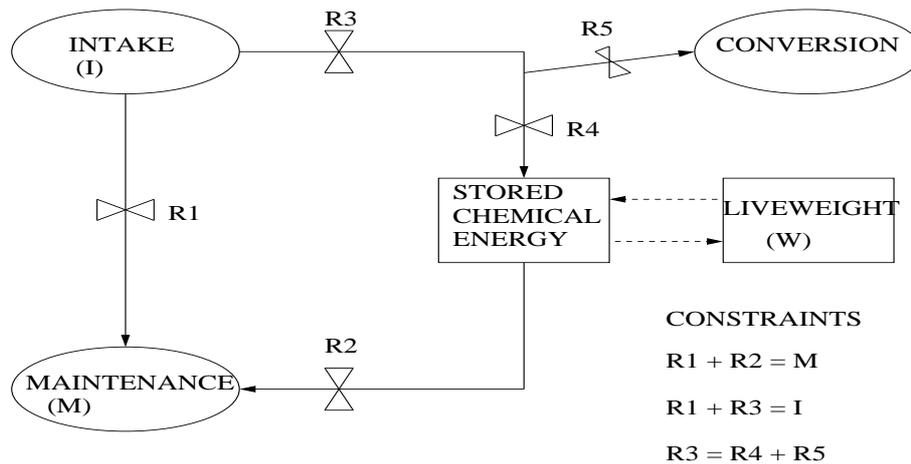
Figure 2: A flow diagram of an Energy Model for Cattle Growth

to the birth rate and a flow into a compartment representing the number of deaths would equal the death rate.

Figure 2 shows a flow diagram for an energy conservation mode of cattle growth. The model assumes that the only relevant information about an animal is its live-weight, W, the only source of energy is food intake, I, and that this energy is lost either as a result of maintenance (M) or due to converting food to body tissue. In terms of our flow diagram definitions W is a state variable, I is a source and M is a sink. $R1$ to $R5$ are model parameters.

2.3 Choosing mathematical equations

Once the structure of a model has been determined, mathematical equations must be chosen to describe the system. It is worth choosing these equations carefully - they may have unforeseen effects on the behaviour of the model.

2.3.1 Equations from the literature

It may be that somebody else has published an equation relating to the quantities you are interested in. This provides a good starting point, but it is necessary to proceed with caution. Problems encountered may include;

- equations derived from data with a range of explanatory variables which does not contain the range required for model application;

- experimental conditions (the environment) differ substantially from conditions to be encountered during model application;

- equations describe behaviour of the bulk of the data with no attempt to account for known departures at the end of range, or no account taken of variability.

Some areas of science are sufficiently well studied that appropriate forms of analysis have become standard. it is then relatively safe to assume that a similar analysis (and hence equation structure) carries over to similar problems.

Often, equations in the literature will not be expressed in exactly the form required for the model. The dependent and explanatory variables in a regression may be transposed. Or an equation may describe the change in weight of an animal over time when the model requires knowledge about the
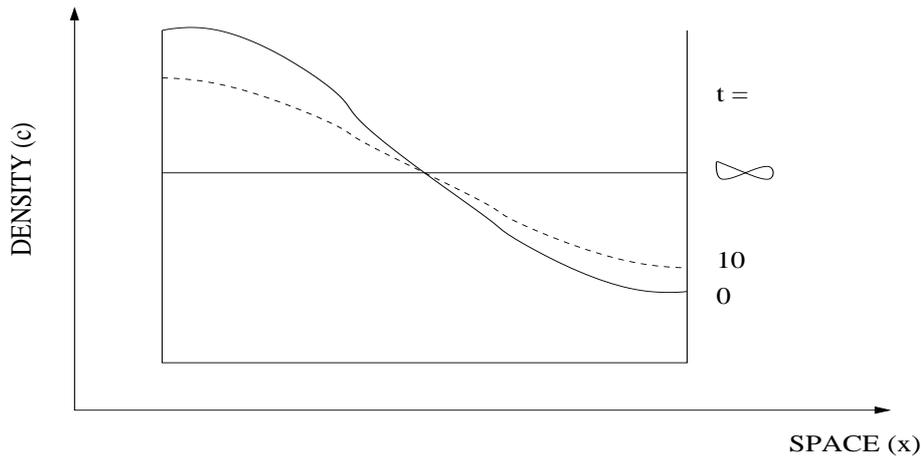
Figure 3: Diffusion of a population in which no births or deaths occur.

rate of change. In either case, we may have to accept that the parameter estimates given are not necessarily the best ones for our purposes.

2.3.2 Analogies from physics

Physicists have built mathematical models to describe a wide range of systems. Often, the systems can be specified precisely, making the application of mathematical equations relatively simple.

There are many cases where we can draw on this body of knowledge to suggest equations for biological systems. For example, local migration of organisms is often considered to be equivalent to a diffusion process. Diffusion of large numbers of tiny particles has been well studied, and is described by the equation.

$$\frac{\partial C(x,t)}{\partial t} = D\frac{\partial^2 C(x,t)}{\partial x^2} \, ,$$

where C(x,t) is the concentration of particles at location x at time t. This equation gives no hint as to the future motion of any individual particle. It is simply a description of collective behaviour. Figure 3 illustrates how a spatially distributed population, obeying the diffusion equation changes over time.

2.3.3 Data exploration

Where no information exists about the form of a relationship, the only way forwards is to acquire a body of data and fit equations to it. This has the advantage that we are in control of the analysis. This topic is worth a course of its own. The BioSS course "Regression and Curve Fitting using Genstat" is recommended. Here, we restrict comment to exploration of growth processes described by differential equations.

When our data consist of observations y_i at times t_i, $i = 1...n$, we can estimate the derivative between times t_i and t_{i+1} by $d_{i+1} = (y_{i+1} - y_i)/(t_{i+1} - t_i)$. By plotting d_{i+1} against $(y_{i+1} + y_i)/2$, an estimate of the average value of y, we can investigate the relationship between dy/dt and y. In the case of exponential growth, we would expect the graph to show a straight line. Where growth is limited, we would expect to see some departure from linearity. For example, logistic growth, described by the equation

8
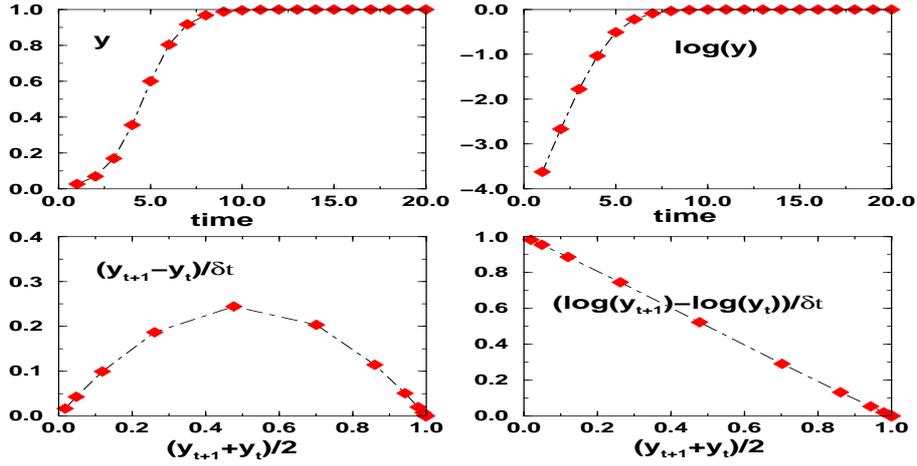
Figure 4: The relationship between logistic growth a population data

$$\frac{dy}{dt} = ry(a - y)\,,$$

should lead to a quadratic curve. Since it is generally rather difficult to determine visually whether or not a curve is quadratic, we would like to find a plot which gives a straight line when growth follows the logistic equation. To motivate this plot, we write the logistic equation as

$$\frac{1}{y}\frac{dy}{dt} = ry(a - y)\,,$$

If we could calculate the left-hand side, sometimes called the proportional growth rate, from the data, then plotting against y would indeed yield a straight line. There is a mathematical result which says that

$$\frac{1}{y}\frac{dy}{dt} = \frac{d}{dt}(\log y)\,,$$

Hence if we calculate $(d_i = \log y_{i+1} - \log y_i)/(t_{i+1} - t_i)$ we can proceed as above. This is illustrated in Figure 4. Departure from linearity in the fourth plot not only indicates that the logistic is not the correct model for the data, but it may also suggest the kind of amendment, which is appropriate.

2.4 Solving equations

2.4.1 Analytically

There is much to be gained from obtaining an analytical solution to a model. This will allow us to perform all of the manipulations implied by the model with the minimum of fuss. Note that a full analytical solution for a stochastic model involves finding the distribution of outcomes, but we may feel satisfied if we can solve the equations for the mean and standard deviation. In general, obtaining an analytical solution is rarely a simple matter.

In certain special cases, it is possible to obtain a mathematical solution to a system of equations. For example, in compartment analysis we often encounter linked differential equations of the form:

9

$$\frac{dx}{dt} = ax + by + cz$$

$$\frac{dy}{dt} = dx + ey + fz$$

$$\frac{dz}{dt} = gx + hy + kz$$

These equations are linear since there are not product terms of the form xx or xy on the right hand side. Because of the properties of linear systems, these differential equations can always be solved by a standard method. However, the method of solution is restricted to linear systems. If we were to consider a similar system, which contained just one non-linear term, then we could not apply this method. The analytical solution, if it exists, must be sought in other ways.

If the model consists of just one differential equation, then there is a good chance that it has already been studied. Texts such as Abromowitz and Stegun (1968) contain a large number of standard integrals, and it is worth looking to see if the equation of interest is included. Similarly analytic results can also often be obtained for linear stochastic models. However, when models contain nonlinearity, and most interesting models do, analytic results are typically harder to obtain than for the corresponding deterministic system. If the model is more complicated, and especially if the structure of the model is likely to be changed, then it is hardly worth even trying to find an analytic solution.

2.4.2 Numerically

When analytical methods are unproductive we can use numerical methods to obtain approximate solutions. Although they can never have the same generality as analytical solutions, they can be just as good in any particular instance.

Numerical solution of model equations generally mimics the processes described in the model. For difference equations, numerical solution is exact since we can use the rules laid down in the equations to follow the evolution of the system. With a stochastic model, we can repeatedly simulate outcomes using a random number generator as described earlier, and combine a large number of simulations to approximate the distribution of outcomes.

Differential equations provide a rather more difficult problem. The basic method is to divide continuous time into discrete intervals, and to estimate the state of the system at the start of each interval. Thus the approximate solution changes through a series of steps. The crudest method for calculating the steps is to multiply the step length by the derivative at the start of the interval. This is called Euler's method. More sophisticated techniques are used in performing the Runge-Kutte types of integration. Fourth order Runge-Kutte is both commonly used and sufficiently accurate for most applications. It is always worth treating numerical solutions to differential equations with caution. Errors in calculating them may accumulate, as can be seen by considering the following example.

The cosine function y = cos(x), is the solution of the differential equation

$$\frac{d^2y}{dx^2} = -y \, ,$$

when y(0) = 1 and $dy/dx = 0$ when $x = 0$ i.e.,

$$\frac{dy}{dx} \mid_{x=0} = 0 \, ,$$

This differential equation is equivalent to the linear system

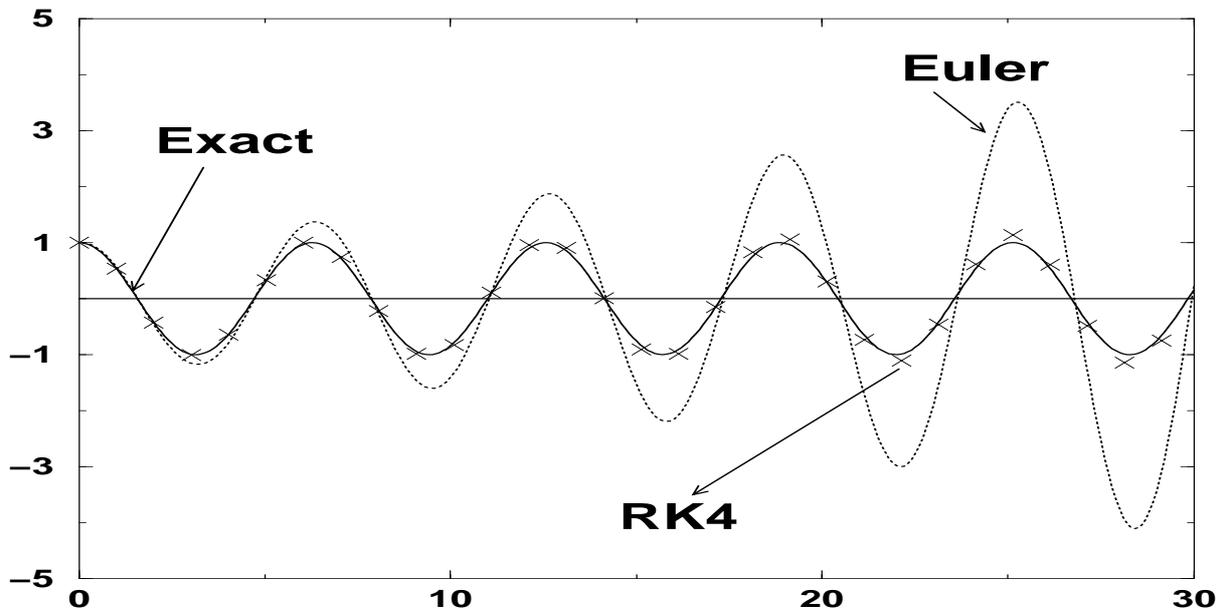$$\frac{dy}{dx} = z \quad \text{and} \quad \frac{dz}{dx} = -y$$

Figure 5: Numerical estimation of the cosine function

when $y(0) = 1$ and $z(0) = 0$.

Although the errors inherent in numerical integration of this system using Euler's method (Figure 2.4.2) can be reduced by shortening the step length, this increases the computational burden. Fourth order Runge-Kutte, on the other hand, can be seen to be much more accurate.

3 Studying models

It is important to realise that the behaviour of a model can be described in two ways. Qualitative description provides an answer to questions about "how", whereas quantitative description answers questions about "how much". In general, qualitative behaviour will be the same for whole families of models and hence is amenable to general results. This contrasts markedly with quantitative behaviour, which is often only relevant to an individual circumstance.

The qualitative behaviour of stochastic models is likely to show more diversity than the corresponding deterministic models. For example, different realisations of a stochastic population model may exhibit exponential growth and extinction. With stochastic models, therefore, it is important not only to describe the average behaviour but also to describe the range of types of behaviour.

3.1 Dimensionless form

One way of breaking away from the quantitative features of a particular model is to re-write the equation in terms of dimensionless quantities. This reduces the number of parameters in the equations, making qualitative studies easier and allowing direct comparisons of model types. Having dimensionless quantities also makes direct sense of statements about "small" or "large" values.

For example, the differential equation for logistic growth can be written in a dimensionless form which contains no parameters. The transformations which lead to this dimensionless form are simply rescalings of the measurement axes as shown in Figure 3.2. Note that there is an additional parameter in the logistic model corresponding to population size at some specific time. Changing this parameter corresponds to sliding the whole curve, left or right, along the time axis.

3.2 Asymptotic behaviour

A qualitative feature of the (scaled) logistic model is that the long-term behaviour of the population is independent of the initial conditions. Hence the population will always approach the value 1 as time increases. This value is called an asymptote. A population of size 1 is in equilibrium, since according to the model it will never change. Furthermore, this equilibrium is stable, since if the population size is changed slightly it will always return to 1.

Many other types of long-term behaviour are possible. Some, such as oscillations (also called limit cycles), recur at regular intervals: others continue to display irregular behaviour. Although there are analytical methods available for studying long-term behaviour of models, we will concentrate only on the graphical methods.

An alternative way of representing the logistic equation is to plot the derivative dy/dt against y. From Figure 3.2, we can see that dy/dt is positive for $0 < t < 1$ and negative for $1 < y$. Hence we can deduce that, in the long run, y will approach a value of 1, although the graph contains no explicit information about how fast this will happen.

If the population is not only self-regulating, but is also subjected to a loss of individuals due to some external influence, then the logistic model must be amended. One classic example of such a system is fish stocks being depleted by trawlers. If the trawlers operate with constant effort, then it is reasonable to assume that the rate of depletion is cy. As the value of the c increases, the stable population size decreases until it reaches zero (Figure 4).

Alternatively, if the trawlers take a catch of constant size, the rate of depletion is simply c. As c increases, the stable equilibrium value decreases and also the unstable equilibrium value increases. If the population ever falls below the unstable equilibrium, it will carry on decreasing until it becomes extinct. Thus the long-term future of the population depends on the way in which trawling depletes the population.
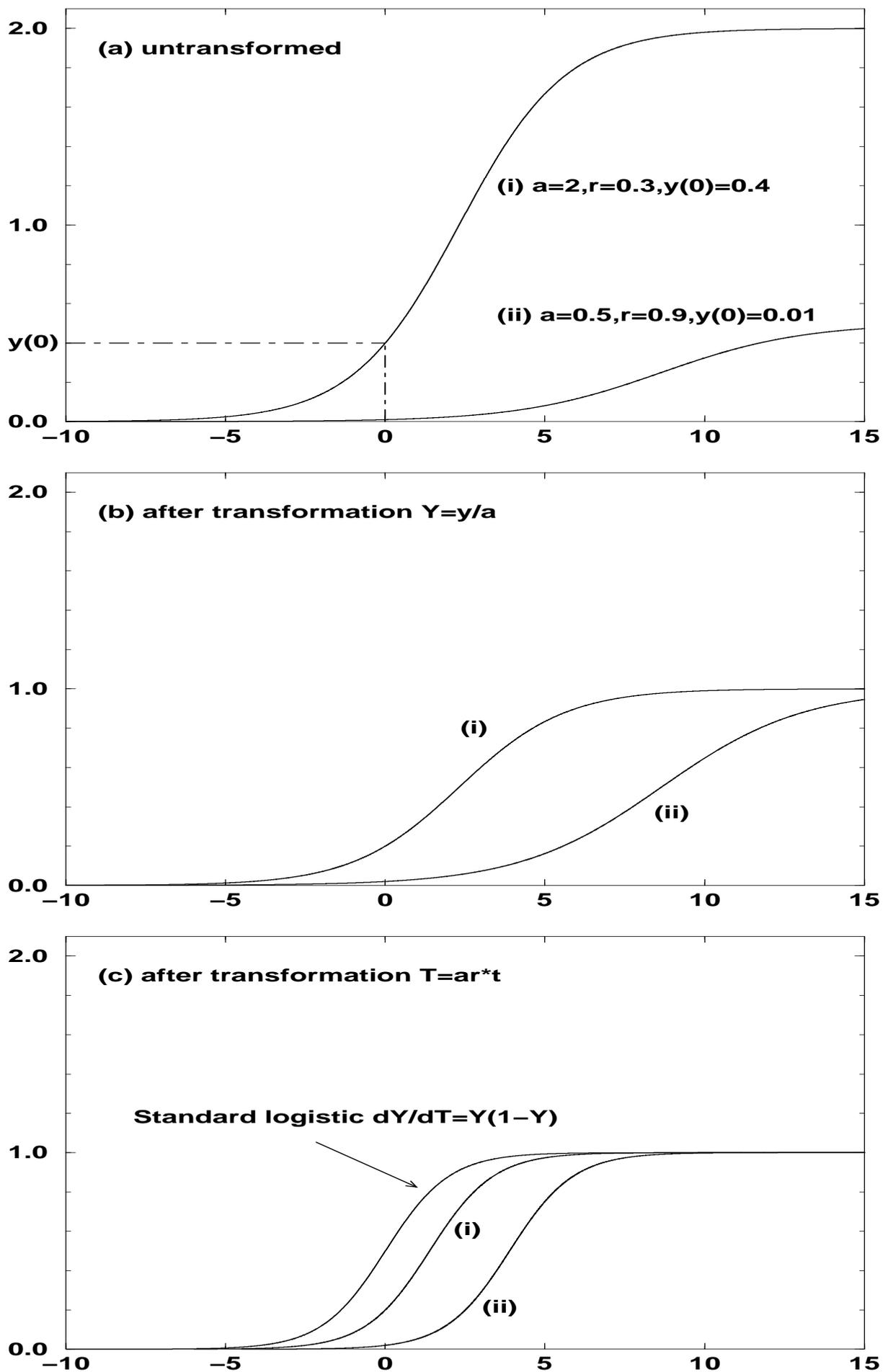
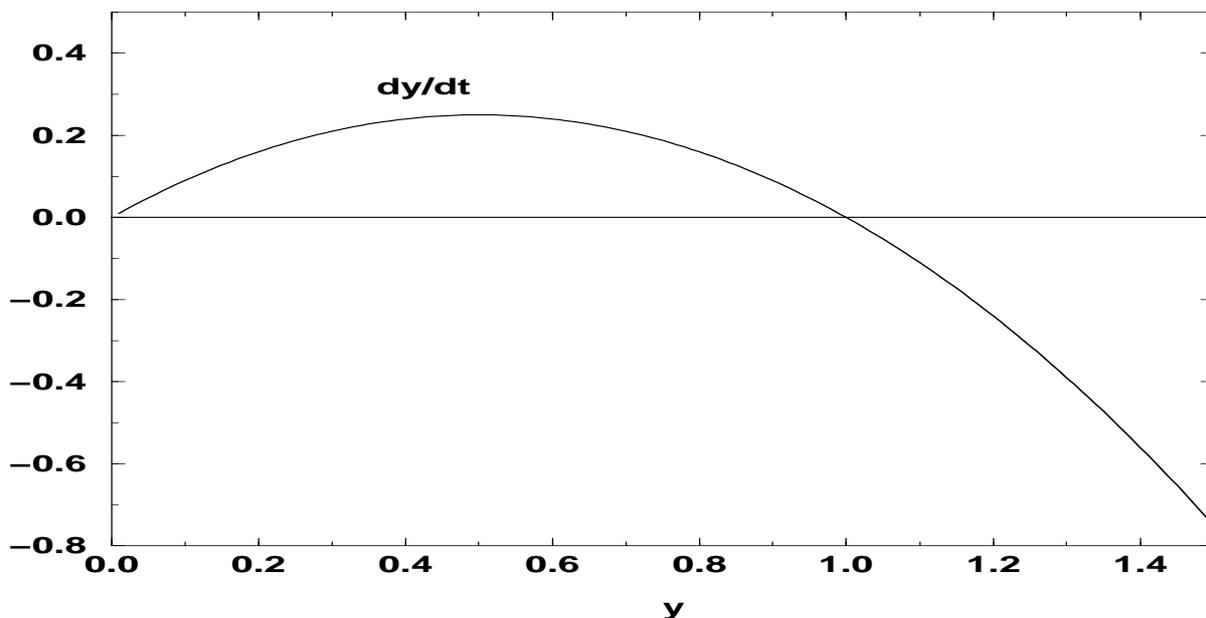Figure 6: Scaling of two logistic equations $dy/dt = ry(a - y)$ to dimensionless form.

Figure 7: Graph of dy/dt against y for the logistic curve given by $dy/dt = ry(a - y)$.

Where two populations are involved, we use a graph called the phase-plane. Figure 9 shows such a graph of the size of one population on the x-axis and the other on the y-axis. Each moment in time is represented by one point, and the entire history of the system is represented by a curve. Interacting populations can show a wide variety of behaviour. The qualitative features of this behaviour can be studied without calculating an analytical solution to the equations. Plot the lines for which $dy/dt = 0$ and $dx/dt = 0$ on the graph. These lines are called the null-clines. Where they cross, the populations have an equilibrium value. Each equilibrium may be either stable or unstable, depending on the coefficients in the model.

Not all systems settle down to regular long-term behaviour. Until recently, this was believed to be due to continual, irregular disturbances from the environment. However, it is now known that even simple deterministic mathematical models may show apparently erratic behaviour which does not die away with time (Figure 10). Such models are called chaotic. Models which exhibit chaotic behaviour are very sensitive to the choice of initial conditions. From two similar initial states, the difference between resulting states increases exponentially with time. We will see an example of deterministic chaos in the practical exercises.

We will now move on to consider two quantitative methods for studying models. The first of these, sensitivity analysis, is used to find out how dependent outcomes are on the particular values of chosen parameters. The second method can be used to obtain an approximate understanding of the model in a few specific cases.

3.3 Sensitivity analysis

The aim of sensitivity analysis is to vary model parameters and assess the associated changes in model outcomes. This method is particularly useful for identifying weak points of the model. These can then be strengthened by experimentation, or simply noted and caution taken in any application.

If the model is particularly simple, it may be possible to differentiate the outcome with respect to each parameter in turn. The derivatives give the exact rate of change of predictions with respect to the parameters. With more complicated models, differentiation is best avoided and numerical methods applied. Figures 4 and 5

How de we decide on the size of changes to make to each parameter? This choice should depend on

14

(a) Constant effort model: dy/dt=y(1−y)−cy
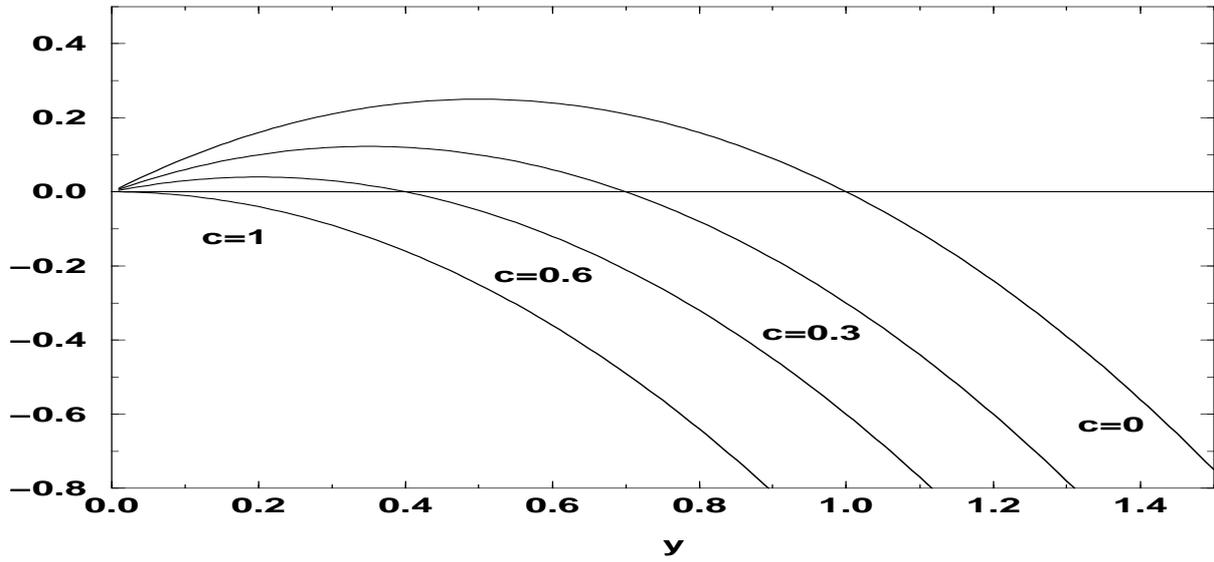
(b) Constant catch model: dy/dt=y(1−y)−c
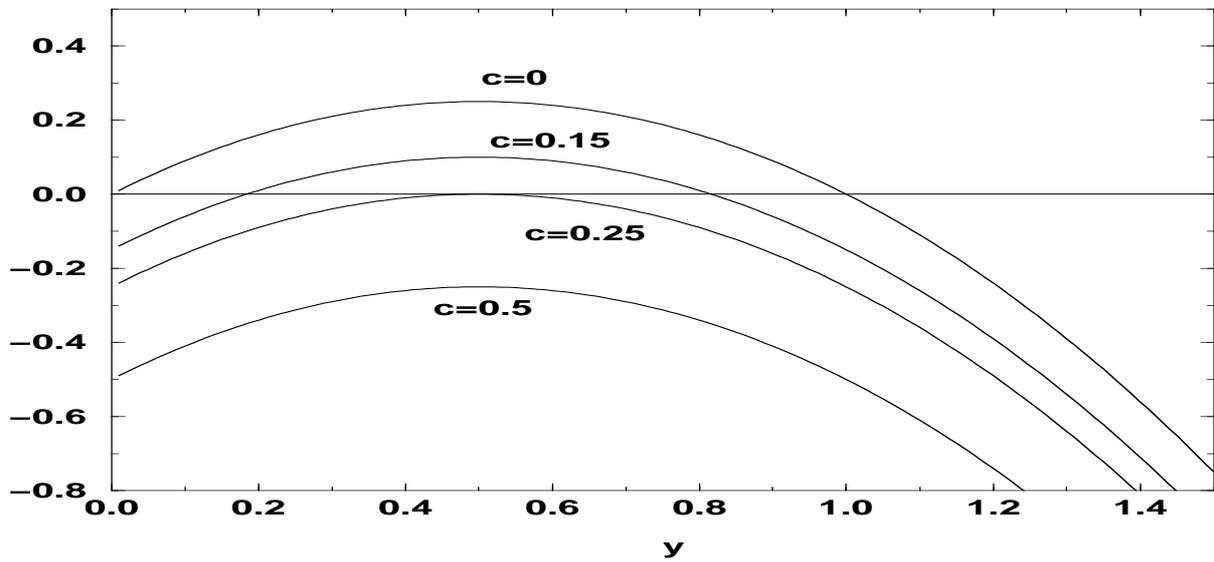
Figure 8: Plots of dy/dt against y for modified logistic equations
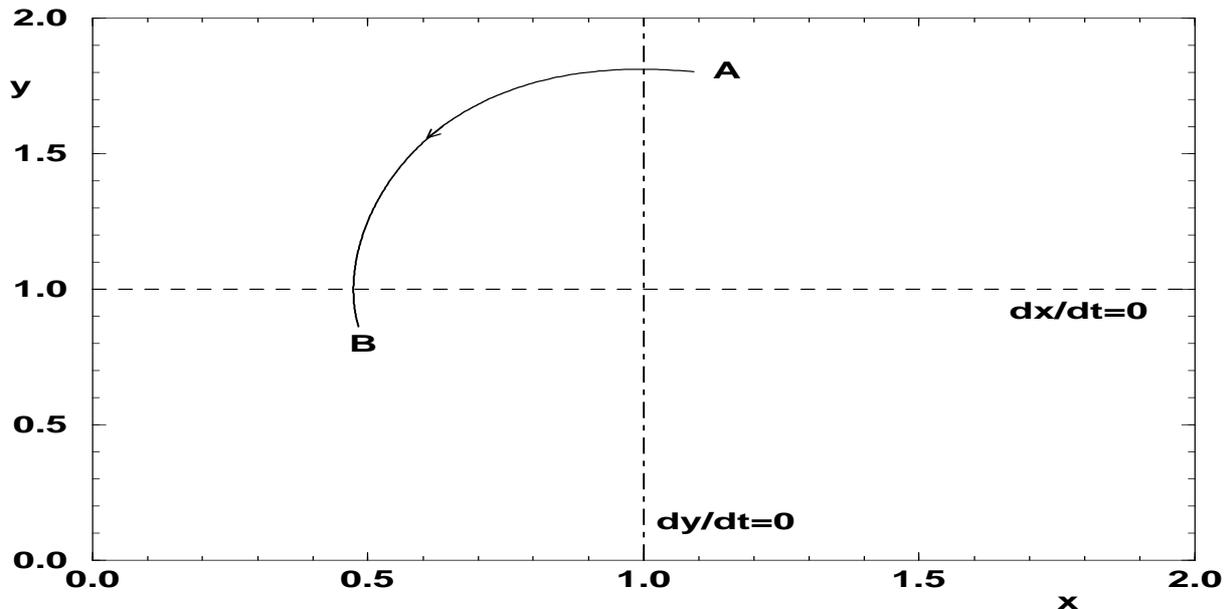
Figure 9: Phase plane diagram for the predator-prey system: $\frac{dx}{dt} = x(1-y)$ (prey) & $\frac{dy}{dt} = -y(1-x)$ (predator), showing the states passed through between times t_1 (state A) and time t_2 (state B).

how well the parameter value can be determined. Hence the parameters which have been estimated from data, a multiple of the standard error is appropriate. Some parameters may have to be guessed, in which case a guess at percentage reliability is also required. It is worth remembering that the person who guessed the parameter value is quite likely to overestimate its reliability. Caution should be taken when parameter estimates are correlated, since if one parameter estimate is changed some of the others might have to be changed too.

3.4 Modelling model output

Evaluating complex models can often take a great deal of computer time. When the model has to be evaluated repeatedly, we may accept some loss of precision in the evaluation if it reduces the time taken per evaluation. Clearly, a reasonable approximation to the model is needed, but when we remember that the model itself is only an approximation we will realise that a small loss of detail should be no great worry if it speeds up the calculations enough.

There are two ways to proceed. The first is to develop some approximation by simplifying, or summarazing the model mathematically. An increasingly widespread example of this is the use of so called *closure equations* to summarize the statistics (typically the mean and variance) of a stochastic model. Since closure equations are differential equations the (approximate) results of many stochastic simulations are obtained for a cost roughly equal to one deterministic model run!

The alternative approach is both simple and more general, but may run against the grain. Firstly, forget all that you know about the interacting equations in the model. Secondly, conduct an experiment in which predictions are obtained from the model for carefully chosen values of the control variables. Thirdly, treat the outcomes as the results of a designed experiment and fit an empirical response surface to it. Fourthly, use the fitted surface to estimate model outcomes in the future.

The effectiveness of this procedure depends on a number of issues. Firstly, the model must give outcomes which very smoothly with the control variables in the region under consideration, so that some simple empirical approximation is possible. Secondly, the supposed experiment must cover the full range of control variables over which the empirical model will be used. Thirdly, since the problem is that the model outcomes take a long time to calculate, the experiment should be designed to

16
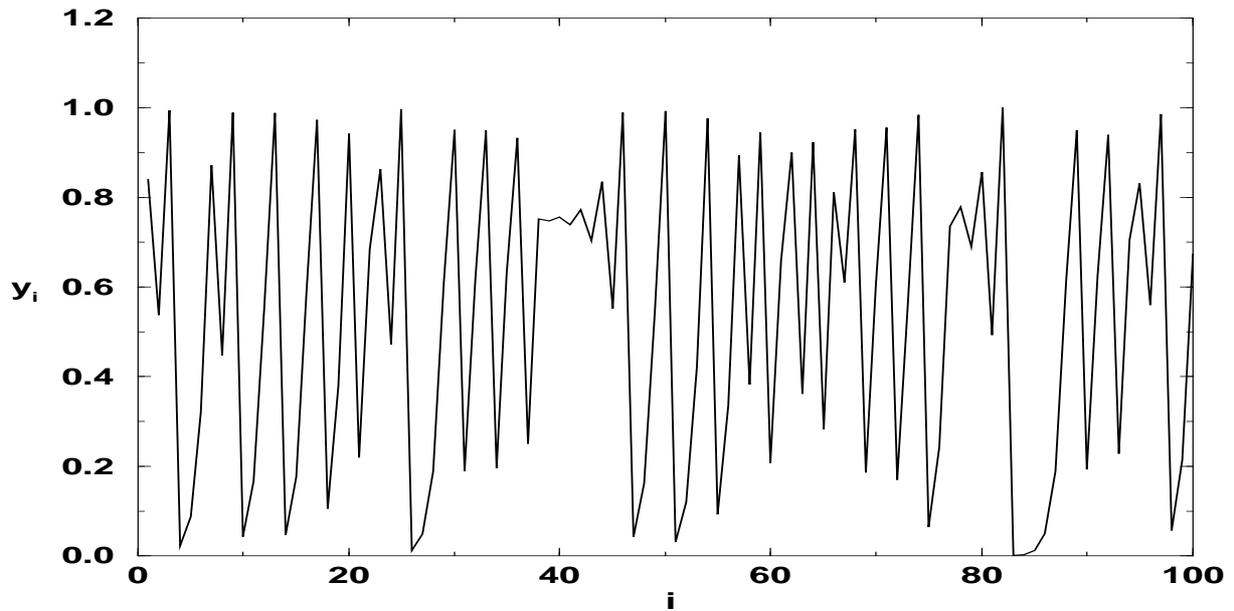
Figure 10: Graph of y_i against i for the chaotic difference equation $y_{i+1} = 4y_i(1 - y_i)$.

economise on the number of model evaluations required.

Fortunately, this procedure has been thoroughly covered in the statistical literature. The resulting experimental designs are called fractional factorial designs. The word fractional refers to the fact that we do not have to perform a complete factorial experiment if we are prepared to discard information about high-order interactions. The fitted curves which are then used to approximate the model are called response curves.

4   Testing models

Once we have studied our model and are satisfied with its performance, it is time to begin testing the model against observations from the physical system which it represents. This process is usually called validation. A similar word, verification, is generally reserved for the non-trivial problem of checking that the predictions from the model are faithful to the model description.

What data should we use for testing a model? There is a strong reason for not using the same data as we used in parameter estimation - it will make us think that the model gives better predictions than it is really capable of. In linear regression, we correct for the effect of estimating two parameters by dividing the residual sum of squares by $(n-2)$ instead of (n). In more complicated situations there is no genuine equivalent of the degrees of freedom argument so we have to proceed with greater caution.

In theory, we should test the model assumptions, structure, parameters and predictions. Usually it is necessary to assume a particular functional form; we should check that this isn't important for the predictions.

4.1   Testing the assumptions

We should check that we have made the correct assumptions when translating from a verbal model into a mathematical model. Assumptions can usually be tested with empirical data as they often are models themselves. If there are unwarranted assumptions made in the mathematical model, they should be relaxed and the resulting change in model fit checked. Assumptions that can be especially dangerous include:

- Linearity

- Deterministic dynamics (see Practical)

- 'Perfect spatial mixing' (see Practical)

If all is well there will be no qualitative (i.e. behavioural) change in the model predictions when relaxing these assumptions, in which case the specific assumption does not matter.

4.2   Model structure

Testing the model structure is usually the most difficult part of testing, particularly in a rigorous way. It often requires a change of methodology, for example, a rewrite of the code, or the need to back analytical work with numerical results. However, unless your model structure is backed by very strong empirical support, it is important to check that the observed features are not 'spurious' because you chose a bad model. Figure 11 illustrates a case when the (deterministic) model results are spurious because the model is a very special case. Making a small change can break the dynamics, and a stabilising mechanism is needed to truly explain predator-prey cycles. This can be achieved by making a small change to the functional form of the model; see Practical 1.

There is no standard way of dealing with model structure problems - it is necessary to simply try making changes to assumptions that are not well founded. Model parameters can however be dealt with in a standard way, but to do so we need some of the concepts introduced in the next section on Prediction.

4.3   Prediction of previously unused data

The most convincing way of testing a model is to use it to predict data which has no connection with the data used to estimate model parameters. In this way, we reduce to a minimum our chance of obtaining a spuriously good match between model predictions and data.
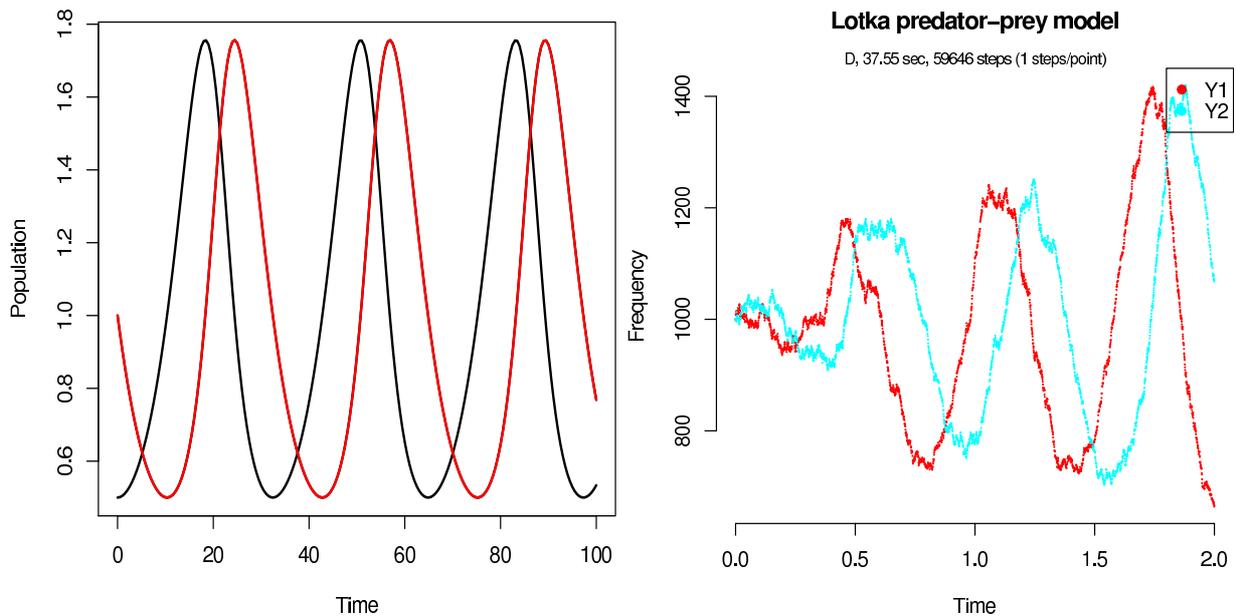
Figure 11: Left: The behaviour of the deterministic Lotka-Volterra predator-prey system. Right: The same model with stochastic birth and death events. The deterministic model predicts well defined cycles, but these are not stable to even tiny amounts of noise. The stochastic model predicts extinction of at least one type for large populations. If regular cycles are observed in reality, this means that some mechanism is missing from the model, even though the predictions may very well match reality.

The phrase "has no connection with" is an important one. Suppose we were to estimate the parameters of a model by fitting to data collected at a farm over several years. To test the model, we collect more data from the same farm in the following years. What can this tell us about the model? At best, it will suggest that the model gives reliable predictions for that one farm. To justify statements about the use of models over a range of farms, the model must be tested over a selection of farms covering the entire range.

What summary statistics should we use to describe the discrepancy between data and model predictions? For predictions P_i of observations O_i, $i = 1...m$, we might use:

$$\text{Bias}(B) \quad = \quad \frac{1}{n}\Sigma_{i=1}^{n}(P_i - O_i)$$

$$\text{Standard deviation } (SD) \quad = \quad \sqrt{\frac{1}{n}\Sigma_{i=1}^{n}(P_i - O_i - B)^2}$$

$$\text{Prediction mean square error (MSE)} \quad = \quad \frac{1}{n}\Sigma_{i=1}^{n}(P_i - O_i)^2 = SD^2 + B^2$$

Each of these could be plotted against a chosen variable to test for homogeneity in performance. It is often a good idea to scale these summary statistics by the mean observation. This makes the summary statistics independent of units of measurement.

It seems intuitively reasonable that the graph of observations against predictions should show a straight line with unit slope, passing through the origin. Although it is certainly true that a good model would be expected to give rise to a straight line, it is not true that the expected regression has unit slope. The reasons for this apparent paradox are difficult to understand, but nevertheless it is a feature which is readily observable.

19

4.3.1 Reasons for prediction errors

What are the reasons for imperfect predictions? If we can understand why prediction errors occur, then we have a basis for deciding how to react to them. In general terms, there are three reasons.

Firstly, there is natural variability in the system and its environment. This is what we normally regard as being measurement error in our experiments. We may have a feel for the potential size of these errors from previous work. In this case, we will know when to look for other sources of error.

Secondly, there is the effect of factors which we have ignored. Whether such effects can be distinguished from natural variability depends on the amount of additional information recorded as a safeguard against this problem. It is advisable to spend time considering how graphical methods can be used to track down such defects, since actually making changes to the structure of a model can be time-consuming.

Thirdly, some proportion of the errors are attributable to mis-specification of the model. They may be due to either to errors in functional form, or to parameter estimates. In either case, we can proceed by making the minimum adjustments to the model equations which are necessary to fit well to the test data. We also have to decide whether such changes are compatible with previous data.

4.4 Estimating model parameters

Estimation of model parameters clearly comes before assessing model performance. However we will discuss it here because it is related to the measures of performance (B, SD and MSE) discussed above. If we have a set of data D and wish to determine model parameters from this data one obvious idea is to minimise one of our performance measures with respect to the parameter values. This will give a *best fit* set of parameters. This also helps to explain why we should use different data for validation and model fitting, since our parameter estimation ensures a good fit to the latter. A common choice for parameter estimation is to minimise the mean square error. One perceived advantage of this is that if one assumes that errors in the data are normally distributed and uncorrelated between observations O_i then the slope of the error surface around the minimum parameter values can be used to calculate the standard errors in the parameter estimates. The problem with this is that the assumption of normality may not be valid, and often the data points can not reasonably be considered uncorrelated (e.g. population size - or any other variable!) over time.

In the case of stochastic models a more statistically well founded method of parameter estimation is available since in principle it is possible to construct the *likelihood* $L(D \mid p)$ which is simply the probability that the model with parameters p generated the observed data D. The likelihood is a function of the model parameters (and the observed data) and may therefore be considered to define the probability of the parameters given the data. By maximising with respect to the parameters we can obtain the *maximum likelihood estimates* of the parameters, and of course we also know their full probability distribution (including standard errors). Moreover, we have made no additional assumptions beyond those that went into the model. So where's the catch? Unfortunately it can be difficult to calculate the likelihood when the data contain missing events (e.g. births and deaths). In such cases, and in practice this is all cases, the missing data must be averaged over typically using computationally time consuming methods such as Markov Chain Monte Carlo (MCMC). This is a rapidly developing area, but currently most parameters are estimated via minimisation of the mean squared error.

Likelihood methods can be used for deterministic models, although in a theoretically unsatisfactory way. If we assume the the model is correct, then any deviation from that is observation error. In this way the model becomes stochastic by incorporating stochastic observations of a deterministic model. This has the advantage that a wide variety of theoretical tools become available; however, since not all the deviation does come from observation error, it may not always work as expected.
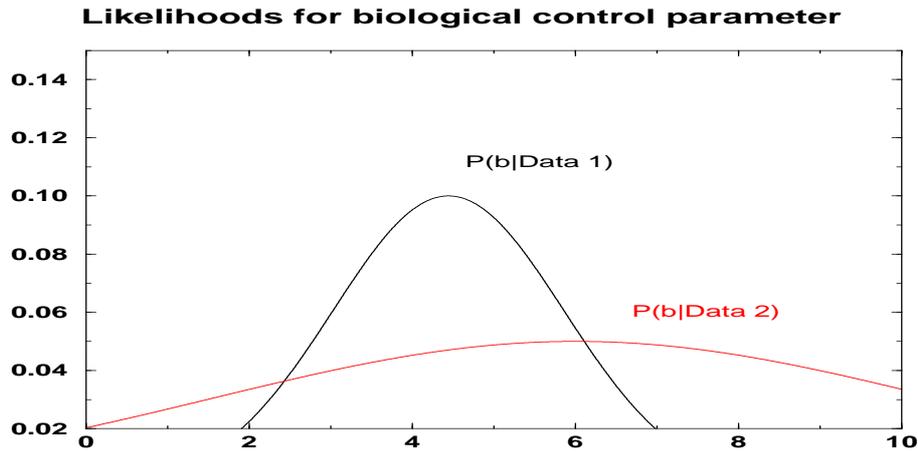
Likelihoods for biological control parameter

Figure 12: Comparison of two models via precision of parameter estimates.

4.5 Comparing two models for the same system

When two models of the same system are available, we may want to compare them with an eye to choosing one for future use. Such a comparison will always contain an element of subjectivity, since there are many different aspects on which the decision can be based. Examples of these are generality, predictive ability and computing requirements. To perform a systematic analysis, consider the building, studying and testing stages in turn.

A good place to begin any comparison is with the underlying assumptions. How do they differ? Presumably there must be some differences, or the models would be identical. If the differences are structural, it is likely that they are designed to be used in rather different circumstances. Which is closest to the proposed application? How important are missing interactions known to be?

Once we have tracked down the differences, we are in a position to ask how much effect they are likely to have on specific applications. Plot out predictions from the models under a range of carefully chosen scenarios. Do they differ sufficiently to cause concern? Two points are worth stressing here. Firstly, small differences are generally not worth worrying about since all models are only approximations. Secondly, beware of models where calculations are repeated iteratively because apparently small differences may accumulate to form substantial differences.

All things being equal, the final choice may be a pragmatic one. Try comparing model predictions with independent observations. Contrast the models by way of summary statistics of the errors. At the same time as comparing the models, this is also a direct test of whether either model is adequate for the application.

Figure 12 shows an example where model parameters are estimated from data form two treatments. In treatment 1 the biocontrol parameter b is estimated with reasonable precision. However, under the second treatment this parameter is poorly estimated (large standard error). This suggests that data set 2 does not support the biocontrol sub-model and therefore that an alternative model without this component should be preferred in this case.

There are standard ways of choosing between models, although they usually require some relationship between the models to be simply interpreted. If a model has Likelihood L, uses k parameters, and there are n independent data points available, then the following Information Criterion could be used:

- AIC (Akaike Information Criterion): Defined for nested models (that is, each model is a subsequent simplification), the AIC value is $-2\log(L) + 2k$.

- BIC (Bayesian Information Criterion): mostly used for time series data and linear regression.
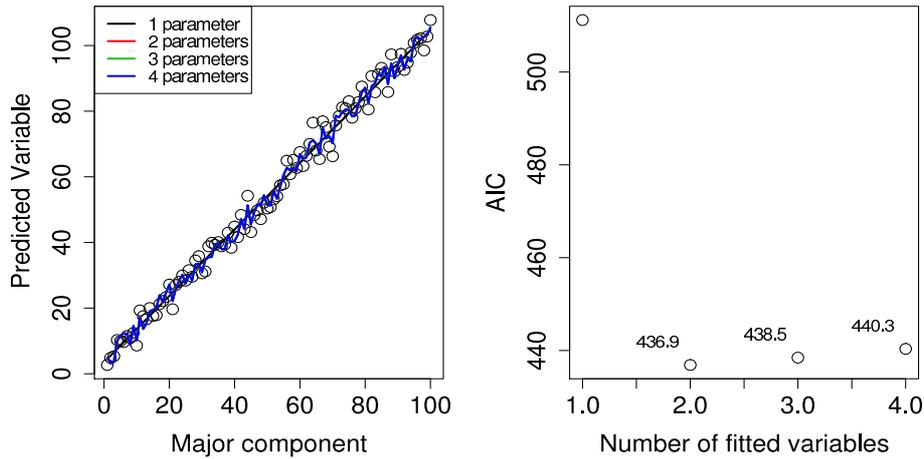
Figure 13: AIC use in a simple linear regression model. Left: The predictions of the model for 1,2,3 and 4 parameters, along with the real data (open circles) generated from a 4 parameter model with noise. Right: the AIC values for each number of parameters. The most parsimonious model is the 2 parameter model, as it has the lowest AIC.

The BIC value is $-2\log(L) + k\log(n)$.

- DIC (Deviance Information Criterion) is most applicable to parameter sets obtained via MCMC, and accounts for not just the best parameter fit but also the range of parameters that fit. The DIC value is the effective number of parameters minus (2 times) the average log likelihood.

In general, one should choose the model with the lowest information criterion. Often this will be in the same order for all information criteria. When the information criterion values are similar between models (within around 5 as a rough guide) then care must be taken. If the models have similar predictions then the simplest may be best for understanding, whilst the more complex retained for predictions. If the models have different predictions then 'model averaging' is appropriate. Model averaging involves taking predictions from several models, and trying to combine them in a sensible way that accounts for the degree of belief in each, as well as the variability between them.

An example of using AIC is given in Figure 13, for a linear model with 4 components: each a factor a less important than the last. Each component has normal distributed noise of variance 1 applied. The model is specified mathematically as $y_i = \sum_{j=1:4}(a^j x_{ij} + \sigma_{ij})$, where $a = 0.075$ and $\sigma_{ij} = 1$. Since the 3rd and 4th components change by less than 1 over the whole range, they are swamped by the noise and the predictive ability gained by adding the extra parameters is negligible. The AIC criterion accounts for this. However, the difference in the AIC value is small enough so that we must consider the meaning of the extra parameters before we discount them.

5 Using models

The method of presentation of a model to its eventual user depends to an extent on how much the user knows about the model. Since, in general, the use will know rather little about the details of the model, it is a good idea to present all relevant information in model output. This allows the user, not the programmer, to make the interpretation. It is almost invariably a good idea to check whether a prediction involves hidden extrapolation. Such extrapolation may be taking place either relative to the data used to build the model, or relative to the data used to test the model.

5.1 Predictions with estimates of precision

If the only output from a model is the prediction of some quantity, how can the user assess the accuracy of the prediction? Of course, this cannot be done, and the user is left in a take-it or leave-it situation. It would be better if the prediction were accompanied by an estimate of precision, such as a standard error or a confidence interval. These can be obtained from model studying or model testing.

If we have investigated the effect of errors in parameter estimates when studying the model, we can estimate the precision of a prediction by summarising the distribution of potential outcomes. This provides only a minimum estimate of error, since it takes no account of potentially erroneous forms of relationships used.

Alternatively, the estimates of error might be carried through from prediction errors analysed whilst testing the model. This is the best error to use, as it includes contributions from all possible sources.

A direct application of estimates of prediction errors is the calculation of safety margins in feed relations. We know from experience that even when animals are fed an identical diet, they will grow at slightly different rates. Any model which describes the growth of the group will thus have to contain a stochastic element. If the target is for animals to gain weight at 1kg per day, choosing a ration from which the average gain is 1kg per day will mean that, on average, half the animals will not meet the target. In order for 95% of animals to meet the target, we need to set the average gain to be 1kg per day plus 1.6 standard deviations (assuming a normal distribution).

5.2 Decision support

We now consider the task of embedding models in a economic framework to assist decision-making. Costs are attached to various inputs to the model, such as animal feed or plant fertiliser. Input levels are chosen in a way that satisfies any constraints on the system. Conditional on the level of these inputs, the model is used to estimate biological outputs. The biological outputs themselves will have a financial value (e.g. sale price) attached to them. The difference between output and input values is then an estimate of gross profit.

The aim of economic analysis may be to find the strategy which will be the most profitable. For simple models with a few control variables, this is usually not too difficult. When there are just two control variables, a graphical treatment is adequate. Numerical methods such as linear or quadratic programming may be applicable, but inaccuracies in the model predictions or economic conditions may render the extra number of significant figures calculated spurious.

For complex systems with many control variables, optimisation may be a difficult problem. In special cases, it may be possible to adapt the problem so that linear or quadratic programming are applicable. Alternatively, a response surface could be fitted, treating profit as the variable to be predicted. Advanced methods such as dynamic programming may be needed to search through the web of interrelated decisions.

Where the model outcome is stochastic, it is rarely possible to make statements like "strategy A is always more profitable than strategy B". Instead, we have to make statements about probabilities,

like "the average profit for strategy A is greater than the average profit for strategy B". If this is true, would we always be wise to choose strategy A? The answer of course is no. It might be that strategy A has a 40% chance of leading to bankruptcy, but a 60% chance of creating huge profits. Not everybody would be prepared to take the risks associated with strategy A.

One aid to making decisions between strategies with random outcomes is the concept of stochastic dominance. Strategy A is said to dominate strategy B if, for all possible outcomes x, the probability that the outcome exceeds x is greater for strategy A than for strategy B. To express this condition mathematically, we need to define the distribution function for strategy A as $F_A(x) = \text{Probability(outcome} < x)$ and, similarly, $F_B(x)$. Then strategy A dominates strategy B if $F_A(x) < F_B(x)$. This implies that the graph of $F_A(x)$ against x lies to the right of the graph $F_B(x)$ (Figure 5.2a).

The theory of stochastic dominance has been extended to cover cases where, in the above sense, neither A dominates B nor B dominates A. For example, second order stochastic dominance is exhibited by A if

$$\int^r F_B(x)dx > \int^r F_A(x)dx \quad \text{for all values of } r.$$

This implies that, averaging over all outcomes up to r, $F_A(x) < F)B(x)$. Graphically, it means that, to the left of any value $x = r$, the area above the curve $F_B(x)$ but below $F_A(x)$ (Figure 5.2b). Although we would expect everybody to choose strategy A if it dominates strategy B, not everybody would choose strategy A if it only displayed second order dominance. People who choose a strategy based on second order dominance are said to be "risk averse" because it indicates the safest bet.
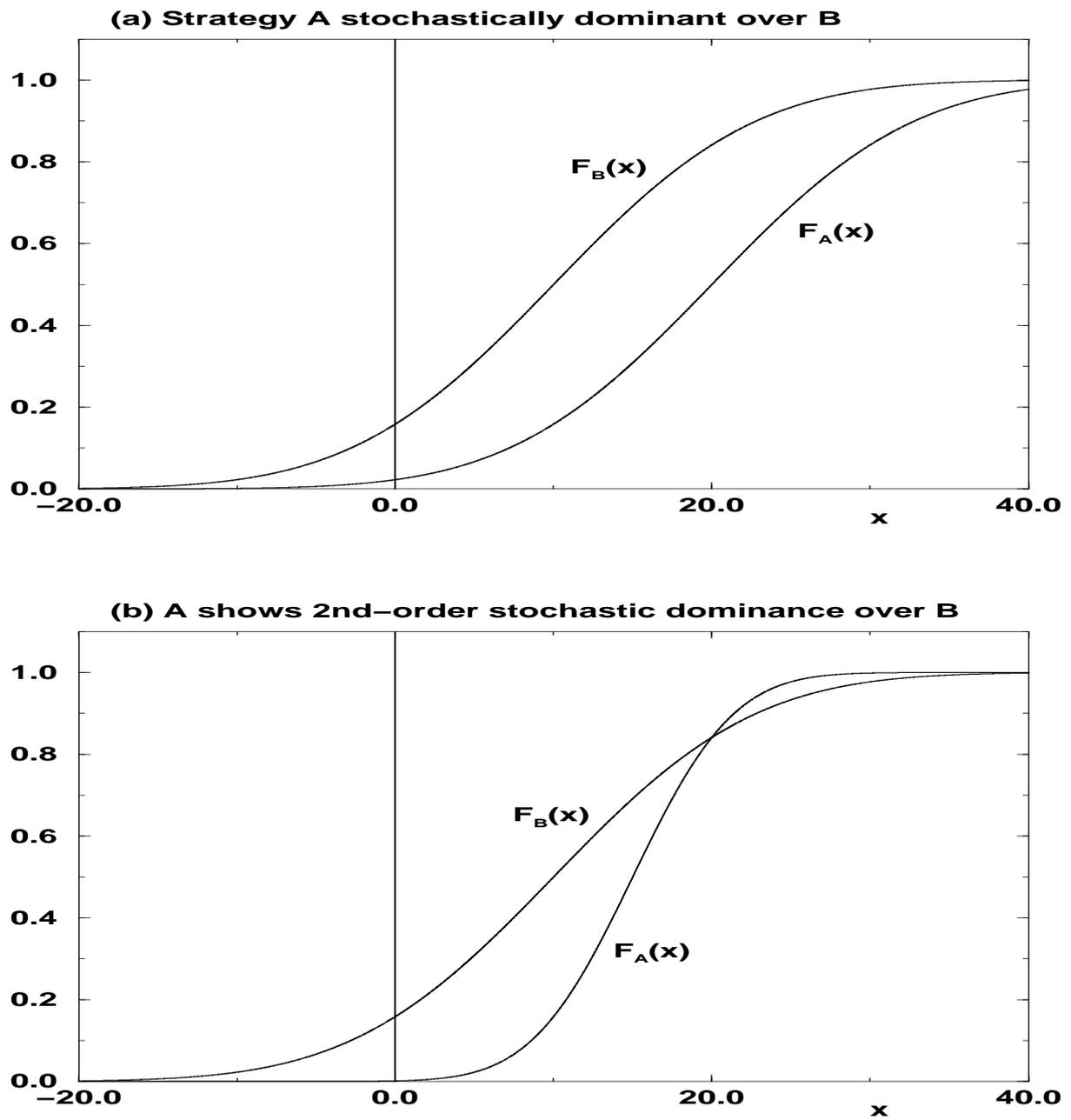
(a) Strategy A stochastically dominant over B

(b) A shows 2nd−order stochastic dominance over B

Figure 14: Distribution functions $F(x) = \text{Probability(outcome\textlangle x)}$ comparing two scenarios A and B.

6   Discussion

This course has displayed the range of activities involved in mathematical modelling. We shall reinforce the ideas introduced previously by discussing how to describe a model. Our final comments reach right back to the outset, when we give some comments about identification of systems which are amenable to modelling and the signals that most of the benefits of a modelling exercise have been achieved.

6.1   Description of a model

There are tow questions to be addressed. The first of these is what information needs to be present. The second is how to go about it.

A good start is a description of the system being modelled. This should put the system into context, reviewing the existing literature and describing the known features which have either been included in the model or discarded as unimportant. It is no embarrassment to exclude known effects from the model. The ultimate judgement is likely to be pragmatic: is the model useful.

This leads directly to a statement of the assumptions made. If they are contentious, then some justification will be necessary. Remember that assumptions are made in sequence. Firstly, there are the assumptions which relate to breaking an interacting system down into a form which can be modelled. Secondly, there are assumptions about the interactions between the components of the model. These may be either qualitative or quantitative.

Quantitative assumptions are usually centred on the choice of equations used, together with parameter estimates and standard errors. It is a good idea to present a table which contains details of the source of the equations and their range of application.

The mathematical model itself is the synthesis of these qualitative and quantitative assumptions. If the model calculations are performed algebraically, give details of the calculations in an appendix. If the numerical methods are used, they will need describing carefully including an appraisal of the accuracy of any approximations. Well-chosen graphs can be a tremendous help.

How was the model studied and tested? It is a good bet that no corrections will have been made to defects unless they have been considered in detail. Given your experiences, what future work will be of most value in improving the model. What levels of precision can you confidently ascribe to the predictions. If none is given, the reader should assume that either none have been calculated, in which case the modeller is flying blind, or there is something to hide.

Finally, it is important to give a clear statement of the range over which the model is believed to be valid. Remember that this range depends on both the data used to build the model and the data used to test the model. Scatter-plots are again useful, since these ranges may not be simple to describe.

6.2   Deciding when to model and when to stop

Back in Section 1 we gave some objectives which mathematical modelling may help achieve. But how can we assess whether a project will be a success, or at what point we should decide that sufficient gain has been made so that we should stop modelling and turn elsewhere? Although the best aid to making these decisions is experience, some pointers can still be given.

Modelling is most profitable when the interactions between components of a system have been studied, but the system has not yet been considered as a coherent unit. If a small number of interactions are known, it may be possible to estimate them from studying the system as a whole. In the end, it is generally necessary to sketch out the structure of a model and see how many gaps there are in our knowledge.

It is time to stop modelling either when the goals of a modelling project have been achieved, or when

there is a dearth of information of a sufficient quality. If information about some components simply does not exist, it may be worth performing experiments to plug this gap. Only when the gap has been filled is it worth reconsidering the decision not to model. There is always a danger that, once the model has satisfied one set of criteria, a more stringent set of criteria is introduced on the assumption that it is worth trying to do better. If a better model is really needed, try to identify the weakest link in the model and see how this can be improved. Working away at the remainder of the model will not help.

Further Reading

For a fuller discussion of material covered on this course, the following books are recommended. Modelling being such a broad subject, each book covers only some aspects.

Bender, E.A. 1978. *An introduction to mathematical modelling.* Wiley, New York.
> An Outline of basic mathematical techniques available to modellers. This is a mathematical text.

Cross, M. and Moscardini, A.O. 1985. *Learning the art of mathematical modelling.* Ellis Horwood Ltd. Chichester.
> A readable, non-technical book on how to start modelling, and how to teach others. It takes a distinctive approach, emphasising that modelling is more art than science.

France, J. and Thornley, J.H.M. 1984. *Mathematical models in agriculture.* Betterworths, London.
> A compendium of published agricultural models, with most emphasis on heavily mathematical, mechanistic ones.

Turchin, P. 2003. *Complex Population Dynamics*, Princeton University Press, 3 Market Place, Woodstock, Oxfordshire, OX20 1SY.
> An overview of models for ecology, with emphasis on mechasistic models. Covers general theory, types of models, possible behaviour, and fitting to empirical data.

References

Sources mentioned in the text

Abramowitz, M. and Stegun, I.A. 1968. *Handbook of Mathematical functions*, Dover, New York.

Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. 1987. *Numerical Recipes*, University Press, Cambridge, UK.

A   Modelling energy requirements for cattle growth

Modelling output from the model of energy requirements for cattle growth.

Here we use the energy model to predict animal weight (kg) after 100 days for the 27 combinations of: starting weight, $W0$ (150, 20, 250kg): feed intake I (30, 4050 MJ/d); feed quality, q (0.5, 0.6, 0.7). We then fit an empirical regression model to these predictions, using explanatory variables which are linear or quadratic in $W0$, I and q.

$W0$	I	q	predicted	F-*statistic*
150	30	0.5	179.97	179.25
150	30	0.6	187.50	187.24
150	30	0.7	195.15	195.24
150	40	0.5	209.79	109.95
150	40	0.6	220.28	220.20
150	40	0.7	230.65	230.45
150	50	0.5	232.65	233.00
150	50	0.6	245.99	245.51
150	50	0.7	258.95	258.02
200	30	0.5	210.97	210.68
200	30	0.6	217.51	217.93
200	30	0.7	224.30	225.19
200	40	0.5	242.73	243.06
200	40	0.6	251.96	252.57
200	40	0.7	261.26	262.08
200	50	0.5	267.31	267.79
200	50	0.6	279.16	279.56
200	50	0.7	290.86	291.32
250	30	0.5	240.35	242.10
250	30	0.6	249.73	248.62
250	30	0.7	255.90	255.13
250	40	0.5	277.21	276.16
250	40	0.6	285.41	284.93
250	40	0.7	293.82	293.70
250	50	0.5	303.05	302.57
250	50	0.6	313.69	313.60
250	50	0.7	324.34	324.62

Calculate $W0 = W0 - 200$, $I = I - 40$, $q = q - 0.6$ to aid numerical stability, then form quadratic terms and use regression to fit a quadratic surface.

Regression Analysis: Accumulated analysis of variance

Change	d.f.	s.s.	m.s.	v.r.
$+W$	1	18854.8789	18854.8789	38902.04
$+I$	1	17089.0742	17089.0742	35258.77
$+q$	1	1628.3027	1628.3027	3359.57
$+W0 \times W0$	1	2.8062	2.8062	5.79
$+I \times I$	1	87.7583	87.7583	181.07
$+q \times q$	1	0.1896	0.18960.39	
$+W0 \times I$	1	33.7680	33.7680	6967
$+W0 \times q$	1	6.5860	6.5860	13.59
$+I \times q$	1	61.1107	61.1107	126.09
Residual	17	8.2395	0.4847	
Total	26	37772.7148	1452.7968	

Repeat the above, dropping W0xW0 and qxq

Regression Analysis: Estimates of regression coefficients

	Estimate	s.e.	t(19)
Constant	252.568	0.256	985.33
$W0$	0.64730	0.00363	178.56
I	3.0812	0.0181	170.00
q	95.11	1.81	52.47
$I \times I$	-0.03824	0.00314	-12.18
$W0 \times I$	0.003355	0.000444	7.56
$W0 \times q$	-0.1482	0.0444	-3.34
$I \times q$	2.257	0.222	10.17

Figure

B Comparing models for cattle growth

Two models for prediction of liveweight gain in growing cattle

Notation:

MEI	=	metabolisable energy of daily ration (MJ/d)
q	=	ration of metabolisable to gross energy in the diet
E_m	=	energy of maintenance (MJ/d)
k_m	=	efficiency of utilisation of dietary ME for maintenance
L	=	$MEI * k_m/E_m$ (level of feeding)
E_g	=	energy retained in daily weight change (MJ/d)
k_g	=	efficiency of utilisation of dietary ME for weight change
EV_g	=	the energy value of tissue lost or gained (MK/kg)
W	=	liveweight (kg)
ΔW	=	liveweight change (kg/d)

General

The daily energy balance in growing cattle may be represented as follows:

$$MEI = \frac{E_m}{k_m} + \frac{E_g}{k_g} \tag{B.1}$$

writing $E_g = EV_g \times \Delta W$ we obtain

$$MEI = \frac{E_m}{k_m} + \frac{EV_g \times \Delta W}{k_g} \tag{B.2}$$

MODEL 1: (*AFRC 1980*)

$$
\begin{aligned}
E_m &= 0.53(W/1.08)^{0.67} + 0.0043W \\
k_m &= 0.35q + 0.503 \\
k_f &= 0.78q + 0.006 \\
k_g &= \frac{k_f}{L - 1} \\
EV_g &= \frac{(4.1 + 0.0332W - 0.000009W^2)}{1 - 0.1475\Delta W}
\end{aligned}
$$

substituting for EV_g into (B.2) and rearranging gives

$$\Delta W = \frac{E_g}{4.1 + 0.0332W - 0.000009W2 + 0.1475E_g}$$

where from (B.1,

$$E_g = k_g \times (MEI - \frac{E_m}{k_m}).$$

MODEL 2: (*TB33*)

$$
\begin{aligned}
E_m &= 5.67 + 0.061W \\
k_m &= 0.72 \\
k_g &= 0.9q \\
EV_g &= 6.28 + 0.3E_g + 0.0188W
\end{aligned}
$$

Rearranging (B.2) and substituting for EV_g gives

$$\Delta W = E_g/(6.28 + 0.0188W + 0.3E_g)$$

where $E_g = k_g \times (MEI - E_m/k_m)$.

Table A
Predictions of liveweight gains (g/d) according to the competing models

W (kg)	100				500			
MEI(MJ/d)	20		40		60		100	
q	.46	.68	.57	.68	.46	.68	.57	.68
Model 1	170	351	1112	1322	197	408	1044	1239
Model 2	157	226	947	1070	215	307	1010	1137

Table B
Bias in predicting liveweight gain (g/d) using independent data (standard deviations in parentheses)

		Data Set	Mean liveweight gain (g/d)	Model 1	Model 2
	1.	(Food, Reading)	1080	210(80)	130(80)
	2a.	(Hinks, Edinburgh)	660	180(100)	150(100)
	2b.		890	130(180)	100(170)
Sources	2c.		820	70(120)	50(120)
	2d.		970	160(130)	220(130)
	3.	(Drayton EHF)	730	−10(120)	0(120)
	4.	(MLC, Nottingham)	910	170(220)	140(210)