THE ROYAL SOCIETY

Super-sensitivity to structure in biological models

Simon N. Wood[1*] **and Matthew B. Thomas**[2]

[1]*Mathematical Institute, University of St Andrews, St Andrews, Fife KY16 9SS, UK*
[2]*NERC Centre for Population Biology and International Institute for Biological Control, Silwood Park, Ascot, Berkshire SL5 7PY, UK*

Applied scientific disciplines use mathematical models to make predictions. In the majority of cases these models are constructed using plausible mathematical characterizations of various component processes of the modelled system, rather than being based entirely on exact mathematical descriptions of proven mechanisms. We use general arguments and a specific example from applied ecology to demonstrate that model predictions can show alarming sensitivity to apparently tiny changes in model specification, in a manner that is counter-intuitive and entirely invisible to conventional model sensitivity analysis. This result has serious implications for practical prediction using biological models.

Keywords: super-sensitivity; mathematical modelling; ecology; prediction; risk assessment; impact assessment

1. INTRODUCTION

Mathematical models are widely used to make predictions about modelled systems. In the majority of cases for which predictions are required, it is not possible to construct models every element of which can be comprehensively justified on the basis of proven mechanism. Rather, the process of model construction is an attempt to arrive at a parsimonious and robust characterization of the system under study. Elements of this characterization will necessarily be 'phenomenological': they are bits of mathematical machinery that behave in accordance with what is known about a system without constituting any sort of explanation of that behaviour.

When evaluating models, it is standard practice to perform sensitivity analyses, in which the sensitivity of model predictions to parameter changes is assessed. It is quite usual for a model to display very high sensitivity to parameter variation in some directions in the model parameter space, while displaying robustness to variation in other directions. Similarly, it is well known that predictions from nonlinear models can display extreme sensitivity to small perturbations of state variables: the phenomenon of chaos. This paper is about a third way in which predictions can be fragile, which may occur in models that are non-chaotic and display no great sensitivity to parameter variation: predictions may display extreme sensitivity to apparently minor changes in model specification.

It has, of course, long been recognized (e.g. Morris 1990) that, given very noisy data, radically different models may appear equally plausible and that the substantial model misidentification that this fact permits can lead to serious errors in predictions. In a general statistical context such model misspecification error has recently received attention (e.g. Buckland *et al.* 1997), but it has again been assumed that the level of misspecification has to be fairly large to matter. This turns out not to be so.

In the interests of clarity, we will first give a specific applied example of super-sensitivity to specification and then present a general explanation.

2. SUPER-SENSITIVITY TO STRUCTURE: AN EXAMPLE

The example concerns an attempt to predict the success or failure of a biocontrol programme for grasshoppers and locusts, which have had a high profile as pest species for a long time (Exodus, chapter 10) and continue to be pests of global importance (Steedman 1990). Such predictive problems are central to applied ecology: losses to insect pests are estimated to account for between 20 and 30% of worldwide crop production (i.e. to around US$300 billion annually) (Hill 1997). Building on previous work (Thomas *et al.* 1995; Wood & Thomas 1996), we constructed a model to predict the outcome of a biological control programme against grasshoppers or locusts in a seasonal environment. The model was parameterized using field experiments in which the rice grasshopper, *Hieroglyphus daganensis*, was infected with the fungal entomopathogen *Metarhizium flavoviride*. The basic ecology of the system can be characterized as follows. Grasshoppers hatch from dormant eggs at the start of the rainy season. During the rainy season they grow, eat and are susceptible to fungal infection. At the end of the rainy season surviving adults lay eggs and die. Fungal infection (initially from spray application) causes grasshoppers to die some days after infection: cadavers then act as sources of further infection. A typical cadaver goes through an initial phase of increasing infectivity as the fungus develops, followed by a slower decline as the resources in the cadaver are exhausted and it physically decays. A

*Author for correspondence.

Received 26 October 1998 *Accepted* 16 December 1998

Table 1. *The three nonlinear transmission models compared*

(Parameter estimates are given with 95% confidence intervals following in parentheses.)

source	model equation	best fit r^2	parameter estimates
Hochberg (1991)	βA_1^{1+q}	0.73	$c = 0.077(0.069,0.082)$, $\beta = 0.40(0.36,0.44)$ $q = -0.32(-0.39,-0.27)$
Briggs & Godfray (1995)	$k\log(1+\alpha A_1/k)$	0.72	$c = 0.078(0.071,0.085)$, $\alpha = 1.10(0.97,1.50)$ $k = 0.16(0.10,0.19)$
Michaelis–Menten	$aA_1/(b+A_1)$	0.72	$c = 0.077(0.070,0.085)$, $a = 0.43(0.33,0.50)$ $b = 0.41(0.24,0.55)$

first-order gamma function provides a good description of the infectivity profile (Thomas *et al*. 1995), leading to the following simple set of delay differential equations governing the healthy host (grasshopper) density (m^{-2}) H within a rainy season:

$$\frac{dH}{dt} = -f(A_1)H,$$
$$\frac{dA_1}{dt} = c(A_0 - A_1),$$
$$\frac{dA_0}{dt} = c(f(A_1(t-\tau))H(t-\tau) - A_0).$$

A_1 and A_0 are auxilliary variables used to obtain the desired build-up and decline in individual infectivity (Thomas *et al*. 1995; Wood & Thomas 1996). A_1 is an index of pathogen density (m^{-2}). τ is the time lag (in days) from infection to death. $f(.)$ is a function used to model the saturation of pathogen infectivity at high densities. c controls the rate of build-up and decay of individual cadaver infectivity. Between rainy seasons the healthy host density is multiplied by a finite rate of increase, F, and supplemented by some small amount of immigration, m; the pathogen density, as represented by A_0 and A_1, is multiplied by a small survival rate γ. For the work reported here $\tau = 12\,\mathrm{d}$, $F = 4$, $m = 0.1\,\mathrm{m}^{-2}$ and $\gamma = 0.02$. A full discussion of this model for the case $f(A_1) = \beta A_1$ can be found in Thomas *et al*. (1995).

Experimental data suggest a saturating form for infectivity as a function of pathogen density. In the absence of a detailed mechanistic model of the infection process, we characterized the nonlinear transmission function f, using three different models suggested in the literature (Hochberg 1991; Briggs & Godfray 1995). These are detailed in table 1. To parameterize f, transmission experiments were performed in the field as detailed in Appendix A. Model parameters were estimated by directly fitting the population dynamic model to the resulting data on the proportions surviving three days of exposure to differing pathogen levels: details of the fitting method are also given in Appendix A.

It was statistically impossible to separate the three models in terms of goodness of fit, a result that is unsurprising when one examines the shape of the best fit functions in each case (figure 1). In order to investigate the distribution of dynamic behaviour implied by parameter uncertainty, 99 replicate parameter sets were produced by bootstrapping of the survival proportions data (Efron & Tibshirani 1993; Davison & Hinkley 1997). These replicate parameter sets represent parameter combinations consistent with the nonlinear transmission data. Model predictions that are consistent across all replicates hold at the 98% confidence level. Details of the bootstrapping method used are given in Appendix A.

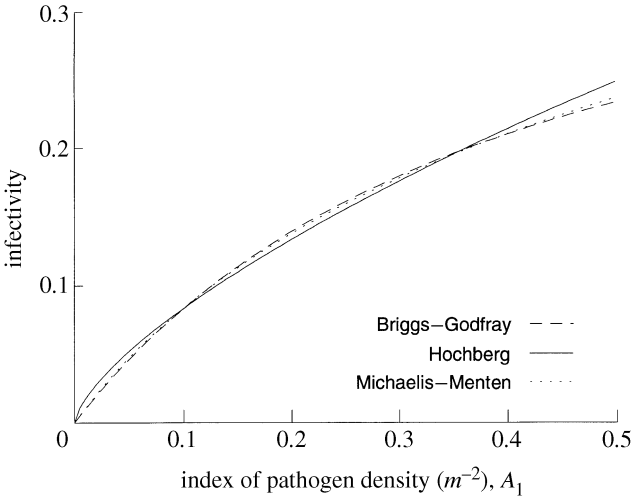

Figure 1. Comparison of the shapes of the best fitting nonlinear transmission models, over the range of pathogen density index present in the experiment (the Hochberg (1991) model pathogen density index range was 0–0.3; other models went as high as 0.8 for short periods—although one can repeat the entire analysis presented in this paper with lower spray action thresholds to bring this figure below 0.5: conclusions are unaltered). The index of pathogen density is the quantity A_1 in the model equations and is proportional to the density of fungal spores (m^{-2}). See table 1 for model descriptions and parameter values. Infectivity is an individual's instantaneous daily infection risk.

Despite substantial variability in the parameter sets the qualitative predictions made by each model were unaffected by parameter variation. Given the unusual amount of care devoted to estimating the uncertainty in parameters and its consequences, one might expect to be able to have reasonable confidence in the predictions of any of the models. By implication, given the statistical equivalence of the models one would expect their predictions to be in close agreement. They are not. Hochberg's (1991) model predicted sustained control at low levels after a single pathogen application for all replicates: figure 2*a* shows a typical replicate. The other two models predicted that repeated pathogen application would be necessary for all replicates: figure 2*b*,*c* shows typical replicates for the Briggs & Godfray (1995) model and the Michaelis–Menten model (with only repeated pathogen application holding the host populations in check).

As can be seen from figure 3, the variability introduced by parameter variation was substantial and on most measures

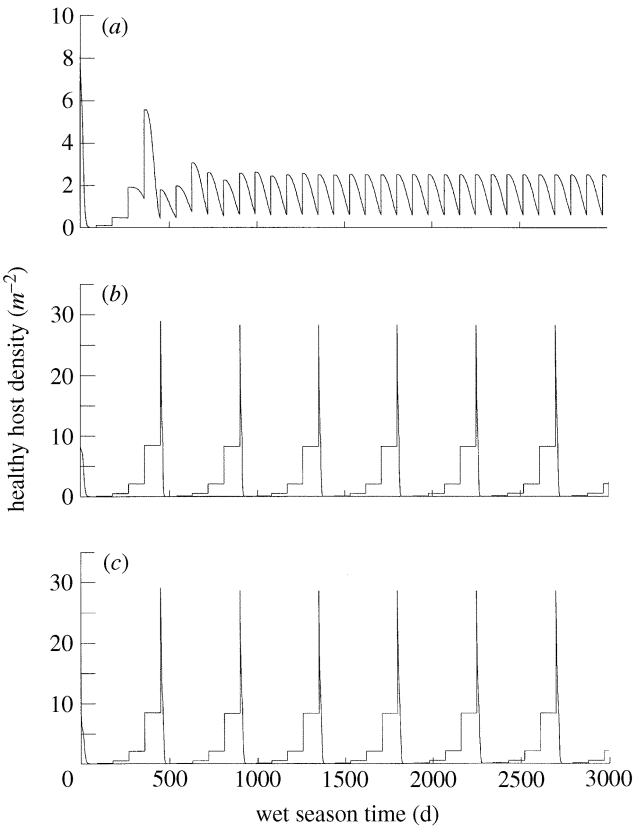

Figure 2. Dynamics implied by the best fit parameter sets for each alternative model. These dynamics are qualitatively similar to those observed in every bootstrap replicate run for their respective models. The vertical axes are healthy host densities (m^{-2}). The time axes are days within wet seasons: discontinuities occur for between-season dynamics. It is assumed that host densities above 10 m^{-2} trigger spraying, with an initial contact rate of 15% and a decay in residual infectivity as measured in Thomas *et al.* (1997). (*a*) Hochberg's (1991) model predicts sustained control with a single pathogen application in the first season. (*b*) The Briggs & Godfray (1995) model predicts that pathogen must be applied every six years. (*c*) The Michaelis–Menten model makes almost identical predictions to the Briggs & Godfray (1995) model. Note the change in vertical scale between (*a*) and the other plots. We also ran stochastic simulations in which the transmission model parameters were reset at the beginning of each rainy season by sampling from the population of parameter sets obtained by bootstrapping: the qualitative results were unchanged. Hochberg's (1991) model gives sustained control after a single application of pathogen, while the other two models require regular and frequent spraying to prevent population explosions.

greatly exceeded the variability between alternative models with their best fit parameters: however, predictions that were completely robust to substantial parameter variation given a particular model specification were not robust to small changes in that specification. This result demonstrates a key point: quantitatively similar models need not give even qualitatively similar predictions. Ecologists have tended to assume that, provided a model is constructed to capture the essence of a process or interaction qualitatively, then its qualitative conclusions will be correct, while in other disciplines it has generally been assumed that examination of the sensitivity of model predictions to parameter variation is sufficient to ensure probity of predictions: both beliefs appear suspect.

The models employed in the example all demonstrated non-chaotic dynamics that were robust to parameter variation on the scale implied by the amount of parameter uncertainty consistent with the experimental data. At the same time, the difference in specification between the models appeared negligible, particularly when compared to the range of variation that was consistent with experiment. Yet the model predictions were completely different. While most modellers would expect that changes in model specification would change predictions somewhat, such extreme sensitivity to specification relative to the commonly considered sensitivities is counter-intuitive.

3. HOW DOES THE PHENOMENON COME ABOUT?

Having observed the phenomenon of extreme sensitivity to model structure, the obvious questions are what causes it and how general is it likely to be? Two levels of explanation are possible and we give both below. The first is more intuitive than the second, but also less general.

First, suppose that providence has seen fit to reveal 'the correct model': in common with most models, it displays great sensitivity to variation in some directions in its parameter space and robustness in other directions. Any practical model constructed without this perfect knowledge will be an approximation to the full model. In terms of the full model, the practical model acts to restrict the combination of values which the full model's parameters can assume. In principle the practical model could be rewritten as the correct model and some (possibly very complicated) constraints on the correct model's parameters. The source of specification super-sensitivity is now clear: it is quite possible for two practical models to both constrain the correct model's parameters in such a way that both models give robust predictions. However, it is also possible that the differences between the two practical models amount to parameter changes of the correct model to which predictions are extremely sensitive.

For a more formal explanation (illustrated schematically in figure 4), consider a model intended to make quantitative predictions p. Suppose that the model combines some established mechanisms, with some phenomenological characterizations described by functions $f_1, f_2 \ldots$ and that the model may additionally depend on some parameters and other quantities, such as initial conditions, that can be written as a vector $\boldsymbol{b}$. The model embodiment of the known mechanisms can be viewed as a nonlinear functional, M, from the (product of) spaces containing all possible f_is and all possible $\boldsymbol{b}$s to the space of all possible predictions.

$$p = M(f_1, f_2, \ldots, \boldsymbol{b}).$$

The sensitivity of p to the elements of $\boldsymbol{b}$ has received much attention and nonlinear dynamicists have explored the consequences of extreme sensitivity to initial conditions in depth. Indeed, the analysis of the sensitivity of predictions to model parameters is a routine and important part of serious predictive modelling. Sensitivity to the f_is has been ignored except in the very limited sense

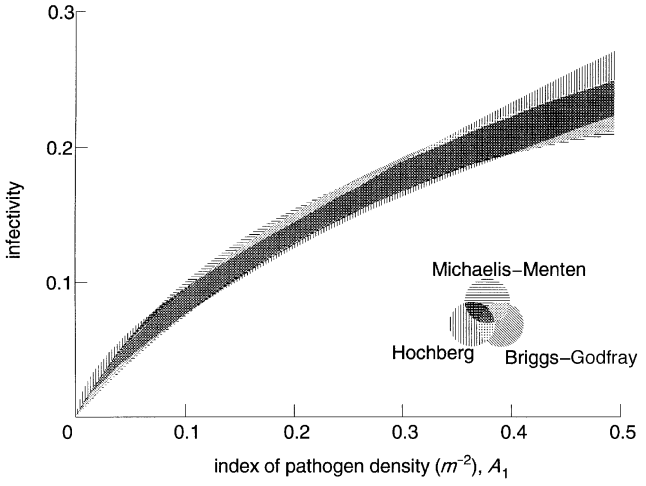

Figure 3. The 98% confidence bands obtained by parametric bootstrapping for the three models. The level and type of shading indicates the three bands and the regions of overlap of the different bands. (It may be helpful when viewing this figure to imagine that the bands are made of light-polarizing material whose polarization direction is given by the directions of the shading lines.)

that it can be explored by varying the parameters of some particular functional form for particular f_is.

The importance of sensitivity to the structure of the f_is can be seen by considering the straightforward case in which

$$p = M(f).$$

Let P denote the space of possible predictions and let F denote the space of plausible fs and assume that appropriate metrics can be associated with both spaces. Now consider perturbing some f_0 in F which corresponds to p_0 in P, with two perturbations that are equal according to the metric on F: only in exceptional special cases will these two perturbations give rise to equal perturbations of p_0 as judged by P's metric. Indeed, since F will almost always be of higher dimension than P, it will usually be the case that an infinite number of perturbations exist which will lead to no change in p_0, while an infinite number also exist that will change p_0. In short, although p_0 may show great sensitivity to some changes in f_0 there will almost always exist an extensive manifold containing f_0 corresponding to a very small region around p_0. If a phenomenological function used for f restricts variation in F to such a manifold, then predictions may appear very robust, even if p is very sensitive to other changes of similar magnitude within F. In general this means that model predictions can be very sensitive to changes in the parametric form of functions, despite being insensitive to variation in the parameters of any given function, even if the variability induced by the latter appears to dwarf the variability induced by the former (according to F's metric).

It is fairly straightforward to demonstrate how this explanation applies to the example in § 2. To do so we must first construct a space of functions that is a better description of our state of knowledge about the correct functional form for transmission than any of the individual models used. We then examine the sensitivity of model predictions to variations in function shape that are confined to the manifold representing a single functional form, relative to the sensitivity to movements in the rest of the space.

Let f_H be Hochberg's (1991) function and f_B^* be the Briggs & Godfray (1995) function with the best fit parameters given in table 1. We now construct a space F of functions on [0,0.5] of the form

$$f_H(A_1)\delta + (1-\delta)f_B^*(A_1) \quad 0 \leqslant \delta \leqslant 1,$$

with metric

$$d(f,g) = \int |f(x) - g(x)| \mathrm{d}x.$$

d measures the distance between two functions (f and g, for example) in F. For illustrative purposes we have made this space quite restrictive, but it is clearly a better candidate for the space of all plausible model functions than that implied by any of the individual models used.

Having constructed the space F, we can investigate variation within the manifold implied by Hochberg's (1991) model, by variation of the parameters β and q of that model and investigate other sensitivities by variation of δ. We use the peak of the limit cycle of host density as the quantity to be predicted by the model, since this relates directly to whether control is achieved or not.

Starting from the best fit parameters of the Hochberg (1991) model with $\delta = 1$, we perturbed β, q and δ so as to produce perturbations of equal size according to d, the metric of F. We then measured the change in predictions that each of these perturbations produced. This procedure was repeated for a range of perturbations up to the point at which control was lost altogether for the δ perturbation (according to the criterion used in § 2). Figure 5 shows the results of this exercise. Clearly the model predictions are much more sensitive to changes in function shape off the manifold of the Hochberg (1991) model, than to changes on that manifold.

4. DISCUSSION

The preceding sections reveal a new type of model sensitivity to add to the well-known and well-studied phenomena that model predictions can be highly sensitive to their parameters and to initial conditions. This sensitivity may be of little significance in the physical sciences for models built on a well-tested mechanism, but in the biological sciences the implications are quite serious. Here it is rarely the case that models are so solidly mechanistic that there is no doubt as to the most appropriate specification of model terms and, in many cases, models contain elements that are purely phenomenological. As the general arguments of the last section indicate, the potential for extreme sensitivity to specification is very wide. Furthermore, the example illustrates that the issue is not merely theoretical. As both theory and example demonstrate, the crucial problem is that even the most careful conventional sensitivity analysis can completely fail to reveal the extreme structural fragility of a model's predictions.

Super-sensitivity to structural specification may provide partial explanation for the chequered record of attempts to

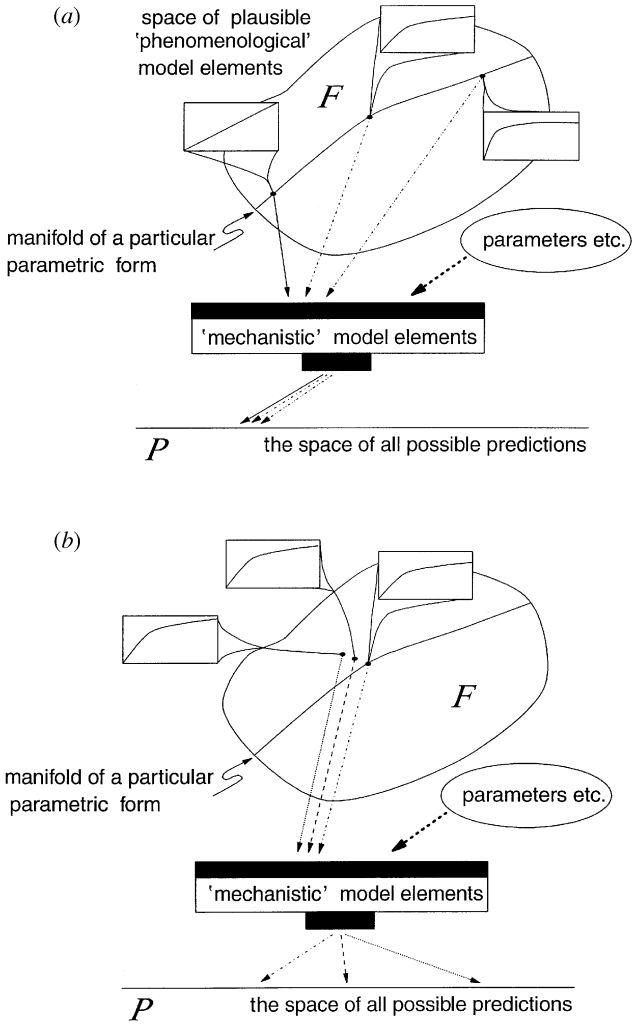

Figure 4. Approximate schematic representation of how super-sensitivity to structure is generated. The mechanistic elements of a model can be thought of as a way of combining phenomenological model elements and other model inputs to produce predictions. In the diagrams the space F of all plausible phenomenological elements is a space of one-dimensional functions—each 'point' in this space is one such function (a functional form and parameter values). The panels show functions that are five elements of the space. The arrows illustrate how incorporating a particular function from F into the model generates a particular prediction, with the type of dashing on the arrows indicating which function gives which prediction. In general, the change in predictions made by the model as we move within F will depend not only on how far we move, but also in which direction. (*a*) How substantial movement in F along the path generated by varying the parameters of a particular functional form might make little difference to predictions. (*b*) A contrasting scenario in which comparatively small movements within F, off the manifold of the particular functional form, produce large changes in predictions.

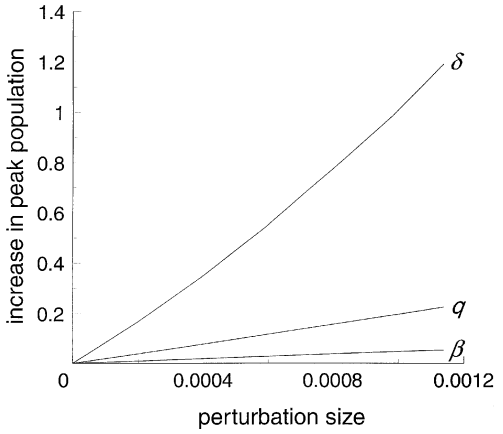

Figure 5. Relative sensitivity of predictions of the model of § 2 to perturbations of equal magnitude, produced by different parameters. The vertical axis is the predicted increase in peak host population. The horizontal axis is the magnitude of the perturbation as measured by the metric d given in § 3. These perturbations were produced by variation of δ, β and q, so each perturbation represents a different direction in the space of functions used in the model. Perturbations of β and q restrict variation to the manifold of the Hochberg (1991) model, while perturbation of δ causes variation off that manifold. Clearly predictions are much more sensitive to δ than to the other parameters.

use models for prediction in fields such as applied ecology. What the results indicate very clearly is that considerable caution is required when predicting with models that are not entirely mechanistic, unless they have been extensively validated using data with a time span that is lengthy in relation to the time-scale of prediction.

This paper has concentrated on the implications of extreme sensitivity to structure for predicting with biological models, but there is another implication. The results indicate that apparently small changes to the mechanisms governing a system could have major effects on the behaviour of that system, which could be very difficult to predict given the original unaltered mechanisms. Again, while physical mechanisms are usually fixed, biological mechanisms may not be.

We thank Cheryl Briggs, Bruce Kendall, Ed McCauley, Shea Gardner and Graeme Ruxton for many useful comments on the manuscript, discussion of model fitting issues and for suggesting that we examine stochastic versions of the models, Bruce Kendall for help in designing figure 4, Bill Murdoch and John Lawton for useful discussions and comments about the general implications of these results and an anonymous referee for suggestions on linking § 2 and § 3. The work was supported by a Nuffield foundation grant to S.N.W., an NSF grant supporting the National Center for Ecological Analysis and Synthesis (NCEAS) complex population dynamics working group, NERC Center for Population Biology funding to M.B.T. and the LUBILOSA programme.

APPENDIX A

(a) *The transmission experiments*

The transmission experiments were performed at a field site in Northern Benin during the 1994 rainy season. Cadavers of grasshoppers that had just died from infection were placed in field plots at densities of $2\,m^{-2}$, $4\,m^{-2}$, $8\,m^{-2}$ and $16\,m^{-2}$ (four replicates at each density). On six occasions over the next 51 days the infectivity of the cadavers was assessed by a bioassay in which field cages containing a cohort of 20 initially healthy grasshoppers were placed over the plots for three days. These exposed grasshoppers were then held in a laboratory for 21 days and monitored for infection (Thomas *et al.* 1996,

1997). During any such incubation some grasshoppers inevitably die of miscellaneous causes, so an estimator of infected proportion is required which corrects for this: let n_i be the number of live grasshoppers at start of incubation day i, and d_i be the number of deaths by pathogen on day i. An estimator of the probability of dying on day i given survival until then is $p_i = d_i/(d_i + n_i)$, the estimated probability of surviving that long is $P_{i-1} = \prod_{j=1}^{i-1}(1 - p_j)$. So the estimator for probability of death by pathogen is $\sum_i P_{i-1} p_i$.

(b) *Model solution and fitting*

The model was fitted to data as follows. The population model was used to predict the proportion of each experimental cohort that would become infected. Using a least-squares objective, the parameters of the model were adjusted to achieve the best fit between model predicted proportions and experimentally determined predictions. This was achieved using a constrained quasi-Newton method (Gill *et al*. 1981), but best fit parameters were also checked using a Gauss–Newton method backed by steepest descent. Care must be taken in integrating delay differential equation models, in order to meet the continuity assumptions of the optimization methods. Integration with an (adaptive) RK2(3) scheme, coupled to interpolation of delay variables by cubic hermite polynomials ensures numerical probity (Highman 1993). General methodology for fitting this sort of population model is given in Wood (1999).

(c) *Bootstrapping*

The bootstrapping was performed parametrically using an approximating normal error model (Davison & Hinkley 1997), with variance estimated individually for each cadaver density treatment (a non-parametric bootstrap leads to similar results). By refitting the population model to each of the 99 bootstrap replicate data sets (once for each of the three different transmission models), we obtained bootstrap replicate parameter sets which could be used to examine both the degree of overlap in the sampling distributions of the competing transmission models and the robustness of their predictions. To evaluate this robustness each model was run for 3000 within-season days with each of its 99 replicate parameter sets and qualitative dynamic features were checked (3000 within-season days is more than 30 model years, which is more than enough time for transient effects to have become undetectable). Initial healthy host population densities of $10\,m^{-2}$ were used, but separate tests over the density range $1–20\,m^{-2}$ demonstrated insensitivity of long-term dynamics to initial conditions, using the best fit parameter values for each model. For Hochberg's (1991) model the exercise was also repeated using a version slightly modified to have finite gradient at zero population—the conclusions were not altered.

REFERENCES

Briggs, C. J. & Godfray, H. C. J. 1995 The dynamics of insect–pathogen interactions in stage structured populations. *Am. Nat.* **145**, 855–887.

Buckland, S. T., Burnham, K. P. & Augustin, N. H. 1997 Model selection: an integral part of inference. *Biometrics* **53**, 603–618.

Davison, A. C. & Hinkley, D. V. 1997 *Bootstrap methods and their application*. Cambridge University Press.

Efron, B. & Tibshirani, R. J. 1993 *An introduction to the bootstrap*. New York: Chapman & Hall.

Gill, P. E., Murray, W. & Wright, M. H. 1981 *Practical optimization*. London: Academic Press.

Highman, D. 1993 Error control for initial value problems with discontinuities and delays. *Appl. Numer. Math.* **12**, 315–330.

Hill, D. S. 1997 *The economic importance of insects*. London: Chapman & Hall.

Hochberg, M. E. 1991 Non-linear transmission rates and the dynamics of infectious disease. *J. Theor. Biol.* **153**, 310–321.

Morris, W. F. 1990 Problems in detecting chaotic behaviour in natural populations by fitting simple discrete models. *Ecology* **71**, 1849–1862.

Steedman, A. 1990 *Locust handbook*, 3rd edn. Chatham, UK: Natural Resources Institute.

Thomas, M. B., Wood, S. N. & Lomer, C. J. 1995 Biological control of locusts and grasshoppers using a fungal pathogen: the importance of secondary cycling. *Proc. R. Soc. Lond.* B **259**, 265–270.

Thomas, M. B., Langewald, J. & Wood, S. N. 1996 Evaluating the effects of a biopesticide on populations of the variegated grasshopper *Zonocerus variegatus*. *J. Appl. Ecol.* **33**, 1509–1516.

Thomas, M. B., Wood, S. N., Langewald, J. & Lomer, C. J. 1997 Persistence of biopesticides and consequences for the biological control of grasshoppers and locusts. *Pestic. Sci.* **49**, 47–55.

Wood, S. N. 1999 Partially specified ecological models: how to formulate the model that the biology implies and fit it to data. *Ecology*. (Submitted.)

Wood, S. N. & Thomas, M. B. 1996 Space, time and persistence of virulent pathogens. *Proc. R. Soc. Lond.* B **263**, 673–680.