

Original article

Comparing predictions of extinction risk using models and subjective judgement

Michael A. McCarthy^{a,b,*}, David Keith^c, Justine Tietjen^d, Mark A. Burgman^a, Mark Maunder^e, Larry Master^f, Barry W. Brook^g, Georgina Mace^h, Hugh P. Possinghamⁱ, Rodrigo Medellin^j, Sandy Andelman^k, Helen Regan^k, Tracey Regan^a, Mary Ruckelshaus^l

^a School of Botany, University of Melbourne, Parkville, Vic. 3010, Australia

^b Australian Research Centre for Urban Ecology, Royal Botanic Gardens Melbourne, South Yarra, Australia

^c New South Wales National Parks and Wildlife Service, Hurstville, Australia

^d Department of Ecology and Evolution, State University of New York, Stony Brook, NY, USA

^e Inter-American Tropical Tuna Commission, Scripps Institution of Oceanography, 8604 La Jolla Shores Drive, La Jolla, CA 92037-1508, USA

^f NatureServe, 11 Avenue de Lafayette, 5th Floor, Boston, MA 02111, USA

^g Key Centre for Tropical Wildlife Management, Northern Territory University, Darwin, NT 0909, Australia

^h Institute of Zoology, Zoological Society of London, Regents Park, London, NW1 4RY, UK

ⁱ The Ecology Centre, The University of Queensland, St. Lucia, Brisbane, Qld 4072, Australia

^j Instituto de Ecología, Universidad Nacional Autónoma de México, Mexico City, Mexico

^k National Centre for Ecological Analysis and Synthesis, University of California, Santa Barbara, CA 93101, USA

^l Northwest Fisheries Science Centre, National Marine Fisheries Service, 2725 Montlake Boulevard East, Seattle, WA 98112, USA

Received 14 May 2003; accepted 9 January 2004

Available online 15 July 2004

Abstract

Models of population dynamics are commonly used to predict risks in ecology, particularly risks of population decline. There is often considerable uncertainty associated with these predictions. However, alternatives to predictions based on population models have not been assessed. We used simulation models of hypothetical species to generate the kinds of data that might typically be available to ecologists and then invited other researchers to predict risks of population declines using these data. The accuracy of the predictions was assessed by comparison with the forecasts of the original model. The researchers used either population models or subjective judgement to make their predictions. Predictions made using models were only slightly more accurate than subjective judgements of risk. However, predictions using models tended to be unbiased, while subjective judgements were biased towards over-estimation. Psychology literature suggests that the bias of subjective judgements is likely to vary somewhat unpredictably among people, depending on their stake in the outcome. This will make subjective predictions more uncertain and less transparent than those based on models.

© 2004 Elsevier SAS. All rights reserved.

Keywords: Extinction; Population dynamics; Population viability analysis; Risk assessment; Stochasticity; Subjective judgement

1. Introduction

Models of population dynamics are often used to predict risks faced by threatened species, pests and harvested resources, and to assess the effectiveness of strategies to manage uncertainty (Boyce, 1992; Burgman et al., 1993; Possingham et al., 1993; Shea, 1998). There is often considerable (Taylor, 1995; McCarthy et al., 1996; Ludwig, 1999;

Fieberg and Ellner, 2000) and perhaps debilitating (Beissinger and Westphal, 1998) uncertainty associated with these predictions. Resource managers require information on ecological risks for effective decision making. The need for prediction is not removed by the presence of uncertainty associated with using models. Therefore, it is useful to know the relative performance of alternative approaches in estimating risk. One alternative of using a mathematical model, is for experts to make subjective judgements of risk (Burgman, 2000), a technique used to list endangered species in the US and elsewhere (Committee on Scientific Issues in the Endan-

* Corresponding author.

E-mail address: mamcca@unimelb.edu.au (M.A. McCarthy).

gered Species Act, 1995). Psychology literature suggests that subjective judgements of risk are likely to be at least somewhat unreliable (Tversky and Kahneman, 1974; Ayton and Wright, 1994; Gigerenzer and Hoffrage, 1995; Anderson, 1998), so it is important that they be assessed.

Ideally, predicted risks of population decline would be evaluated with field data by studying the persistence of replicated populations. However, attempts to test predicted risks of species decline with field data have limited replication (Hanski, 1997; Brook et al., 2000; McCarthy and Broome, 2000; McCarthy et al., 2000, 2001; Ellner et al., 2002; Lindenmayer et al., 2003). Consequently only certain aspects of the models can be tested, and regardless of the amount of data that can be realistically collected, the reliability of the predictions remains at least somewhat uncertain (McCarthy et al., 2001). In these circumstances, it is possible to use simulation models of hypothetical species to generate the kinds of data that might typically be available to ecologists (Fieberg and Ellner, 2001; McCarthy et al., 2003) and then invite other researchers to predict risks of population decline using these data (Cooke, 1995; Millner-Gulland et al., 2001). The accuracy of the predictions can be assessed by comparison with the forecasts of the original model. Given acceptable assumptions, this approach has the compelling advantage that the truth is known precisely, something that is invariably lacking in the real world. We used this approach to assess different methods for predicting risks of population decline. In particular, we compared predictions made using models with those based on subjective judgement.

2. Methods

We created a set of nine hypothetical species to represent a range of life history types (Table 1), and used an international web-based competition as an incentive for other researchers to make predictions of the risks of population decline for these species. Each species was represented by a complex model, which was used to simulate its population dynamics and generate data. The underlying model was not disclosed, and acted as the truth against which the predictions could be assessed. One component of the competition consisted of five species where spatially explicit data were provided (a butterfly, owl,

passerine bird, salmon, and shrub), and a second component consisted of four species where no spatial data were provided (a frog, snail, small mammal, and small plant).

Ten years of data were generated for each species and supplied to people making predictions. The data provided at least some information on population sizes in different years. To reflect the fact that biologists never have perfect information available to them, a range of sampling strategies were simulated, so that for some species the data were of high quality and in other cases it was of low quality (see Appendix). The people entering the competition used these data to make predictions about the risks of population decline under a range of scenarios. In addition, some extra information, such as body size and diet, was provided about each species. In an effort to simulate the role of an expert, the modellers were able to ask a “Sensible Biologist” (the person who created the true models) about each species. The Sensible Biologist judged what was likely to be known about each species and answered accordingly. The information available to each modeller, and the questions and answers to the Sensible Biologist were then made available at: <http://www.nceas.ucsb.edu/~mccarthy/CompIntro.htm>.

Invitations to enter the competition were sent to electronic mailing lists and to individuals who were likely to be interested in participating. Entries were open to anyone, and a range of people entered the competition using a range of methods. Entrants included professional population modellers, graduate students in population modelling courses, academic experts in population modelling, and an undergraduate student. There were six entries in the spatial part of the competition, with five using models and one using subjective judgement. There were 22 entries in the non-spatial part of the competition, 12 using models and 10 using subjective judgement.

To avoid confusion with the term “model”, the models of the hypothetical species that were used to generate the data will be referred to as “true models”, the data generated as “data” and the actual risks of population decline for the hypothetical species as the “truth” or the “true risks”. Models constructed by the assessors and used to make predictions will be referred to as “assessor’s models” and the predictions of all assessors (regardless of the method used to make the predictions) as “predicted risks”.

Table 1

The nine hypothetical species and associated features of the models that were used in the competition. These features were the presence of spatial structure, abiotic interactions, density dependence, the inclusion and type of trophic dynamics, the presence of inbreeding depression and the trend in the amount of habitat available to the species. Density dependence is classified according to its position on the scramble/contest competition continuum

Species	Spatial structure	Abiotic interactions	Density dependence	Trophic dynamics	Inbreeding depression	Trend in habitat
Small mammal	No	Weather	Contest	Prey	No	No
Small plant	No	Rainfall	Scramble	None	Yes	No
Snail	No	Rainfall	Scramble	Cannibalism	No	No
Frog	No	Rainfall	Scramble	None	Yes	No
Passerine	Yes	None	Contest	None	Yes	No
Salmon	Yes	Harvesting	Contest	None	No	Increasing
Butterfly	Yes	Weather	Scramble	Parasitoid	No	No
Shrub	Yes	Weather and fires	Scramble	None	No	Fire occurrence
Owl	Yes	Logging	Contest	None	No	Decreasing

The aim of the competition was to determine whether the accuracy of the predicted risks was influenced by the method used (models versus subjective judgement), characteristics of the models, differences between modellers, and the time frame of the predictions. We also investigated whether changes in risks due to changes in parameters (relative risks) could be predicted more reliably than the actual values (absolute risks, McCarthy et al., 2003). Entrants were required to predict the risk of extinction, risk of decline to 25% or less of the initial population size within the time frame of the simulations, and the risk of decline to 50% or less of the initial population size. In addition, the entrants were required to predict the threshold population size (as a proportion of the initial population size) such that the risk of decline to this level was 50%. This latter value is the median minimum population size (see McCarthy and Thompson, 2001). The initial population size was the population size in the final year for which data were provided. This figure was not provided to the entrants, so they had to estimate it. Predictions were made over 10 and 50 years. To determine how well the models could be used to assess the value of management strategies, the same predictions were also required for a 20% reduction in the fecundity rates, and also a 20% increase in the carrying capacity. We refer to the predicted risks of decline as the absolute predictions and the predicted changes in the risks of decline with changes in fecundity and carrying capacity as the relative risks (McCarthy et al., 2003). Thus, four measures of risk were required for each of the three management scenarios (no change in management, increase carrying capacity, reduce fecundity) and two time periods, leading to 24 predictions for each species.

The accuracy of the predicted risks was assessed by comparing them to the results from the true model. The true risks were obtained from 1000 iterations, over 10 and 50 years from the end of the 10-year survey period. The true risks and the predicted risks of decline can be compared by examining the quasi-extinction risk curve (Fig. 1; Burgman et al., 1993). However, this is difficult to present for all species because of the large number of comparisons (nine species and six curves for each species, representing two time frames and three management scenarios). Instead, overall accuracy was measured by taking the absolute value of the differences between each of the predictions and the truth, and averaging these differences over each species or each assessor (the average deviation). Bias was measured by subtracting the average of the true risks from the average of the corresponding predicted risks. In this case, the median minimum population size (which takes a value of 0 when the risk of extinction is 1) was multiplied by -1 to ensure that it was consistent with the probabilities of decline (i.e. it increased as risks increased). The same analyses were also done for the relative predictions to determine the accuracy with which changes in risks could be predicted. Therefore, positive values for bias indicate that risks (or changes in risks) were over-estimated.

Our aim was to evaluate the predictions of models made by experienced modellers, so three of the model entries for

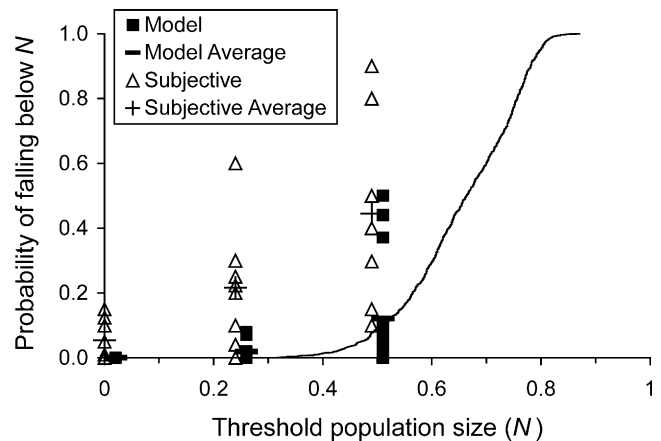


Fig. 1. The true quasi-extinction risk curve (line) and predicted risks of decline for the small plant, illustrating the smaller bias of models versus subjective judgement, with the average of the models close to the true quasi-extinction risk curve. Predictions using subjective judgement invariably over-estimated risk. The quasi-extinction risk curve shows the probability of declining to or below the threshold population size at least once within the next 10 years (Burgman et al., 1993). Population size (N) is expressed as a proportion of the initial population size, and in this case there were no changes in management strategies. The averages of each group are given by horizontal bars (models) and crosses (subjective judgements). The predictions have been offset slightly along the x-axis to improve clarity. The three model predictions that were analysed separately have been excluded from this figure.

the non-spatial part of the competition were analysed separately because they were expected to provide less reliable predictions. Two of these were models based on analytical diffusion approximations (Dennis et al., 1991; Foley, 1994), which we believed a priori would make inaccurate predictions for these species (Wilcox and Possingham, 2002). The third was an entry from an inexperienced modeller where the risks of decline were unreasonably high, given the data. These three entries performed consistently poorly when assessed in terms of absolute error and bias, although their relative predictions were comparable to the other entries. All of the subjective judgements were obtained from ecologists familiar with a variety of quantitative methods.

The modellers used a range of approaches for making predictions, ranging from Bayesian models constructed specifically for the task (Maunder, in press), simple models of population dynamics, and commercial packages for population modelling. The different assessors making subjective judgements also used a variety of approaches. These included simply reading the material and thinking of appropriate values for the answers, through to assigning points based on the attributes of the species (such as population size, trends, fluctuations, etc.), and then using subjective judgement to convert the sum of these points to a probability of decline. Providing details of the range of methods would be unwieldy; however, there was no obvious relationship between the characteristics of the method used and the accuracy of the predictions.

3. Results

In the non-spatial component of the competition, there were nine entries by people using models (not including the three

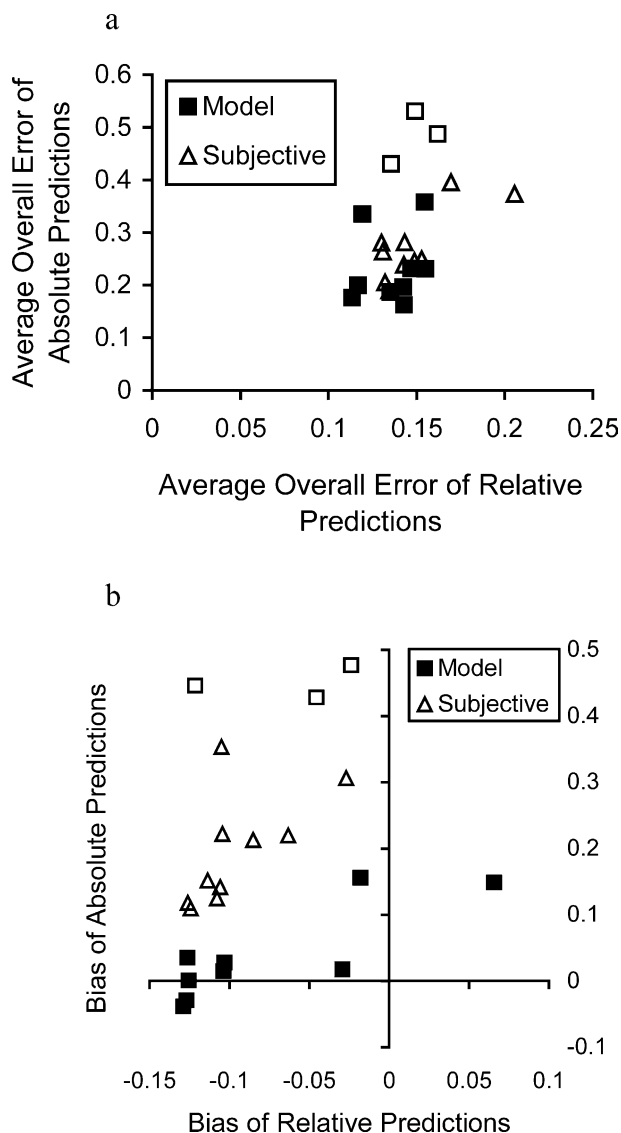


Fig. 2. Average overall error (a) and bias (b) of absolute predictions versus the average error of the relative predictions for models and subjective judgements in the non-spatial part of the competition. Overall error is measured by the absolute value of the difference between each of the predictions and the truth and averaging these values across species. Bias is measured by the difference between the average of the predictions and the truth. The three model entries considered separately are shown as open squares. Each point represents an individual assessor, with the average taken across all the predictions of the four species.

analysed separately) and 10 entries where people used subjective judgement. The results were initially analysed by examining the average error for each assessor. Models and subjective judgements performed similarly in terms of absolute errors and relative errors when averaged over the four species (Fig. 2a). Four of the five best predictions (as measured by absolute error averaged over the species) were from models, although this apparently superior performance of the models could have occurred by chance alone ($P = 0.119$, based on combinatorial probabilities using the hyper-geometric distribution). If a real difference in accuracy exists, it is unlikely to be substantial (Fig. 2a). In contrast, the predictions of assessors using models were noticeably less biased than those using

subjective judgement. The five least-biased predictions were all from models, an unlikely chance occurrence ($P = 0.011 = (9 \times 8 \times 7 \times 6 \times 5) / (19 \times 18 \times 17 \times 16 \times 15)$), based on combinatorial probabilities). The subjective judgements tended to over-estimate risks (Fig. 2b), a result that may have been predicted by research on cognitive psychology in risk perception (Anderson, 1998). In contrast, the predictions of the models tended to be unbiased, which is consistent with a previous test of population models (Brook et al., 2000).

As expected (McCarthy et al., 2003), the changes in risks of decline were predicted more accurately than the absolute risks (Fig. 2a). The absolute error for assessors using models was 0.23 while the relative error was 0.14, averaged over the four species and all nine assessors. For subjective judgements, the corresponding values were 0.27 and 0.15. However, the quality of the absolute predictions by an entrant did not necessarily reflect the quality of the relative predictions; there was relatively weak correlation between the two (Fig. 2; for the model predictions, Pearson's $r = 0.20$ for overall error). Similarly as expected, predictions over shorter time frames tended to be more accurate than over longer time frames. The average error for model-based predictions was 0.18 over 10 years, and 0.28 over 50 years. For subjective judgements, the corresponding values were 0.25 and 0.29.

The results obtained in the spatial component of the competition were consistent with those obtained in the non-spatial component. The five entries from assessors using models were less biased than the single subjective judgement, averaged over the five species in this part of the competition. Relative risks were more accurate and less biased than absolute risks when averaged over the five species (Fig. 3). Risks that were predicted by the models over 10 years were more accurate than that over 50 years (average overall error of 0.13 versus 0.28).

The results are qualitatively the same when each species is analysed separately, averaging over the different assessors (Table 2). The absolute errors are not consistently different between subjective judgements and model-based assessments, and there is little difference between species, with the errors being approximately between 0.1 and 0.3 (Table 2). However, the bias towards over-estimation of risk by subjective judgement is apparent, with the average bias being positive for all but one of the nine species. For three of the four species for which there were multiple assessors, which allows 95% confidence intervals ($\text{mean} \pm t \times \text{S.E.}$) to be constructed, there is clear evidence of bias because the confidence intervals are far from zero (Table 2). In the case of predictions using models, six of the nine species had positive bias, which is not very different from the 4.5 that would be expected. Additionally, only one of the 95% confidence intervals did not encompass zero (the owl), providing only a little evidence of bias, although the bias could be large for some of the other species (e.g. the small plant and the snail).

There was a bias towards under-estimating relative risks (changes in risks) in both components of the competition, with the bias (averaged over the species) being negative for

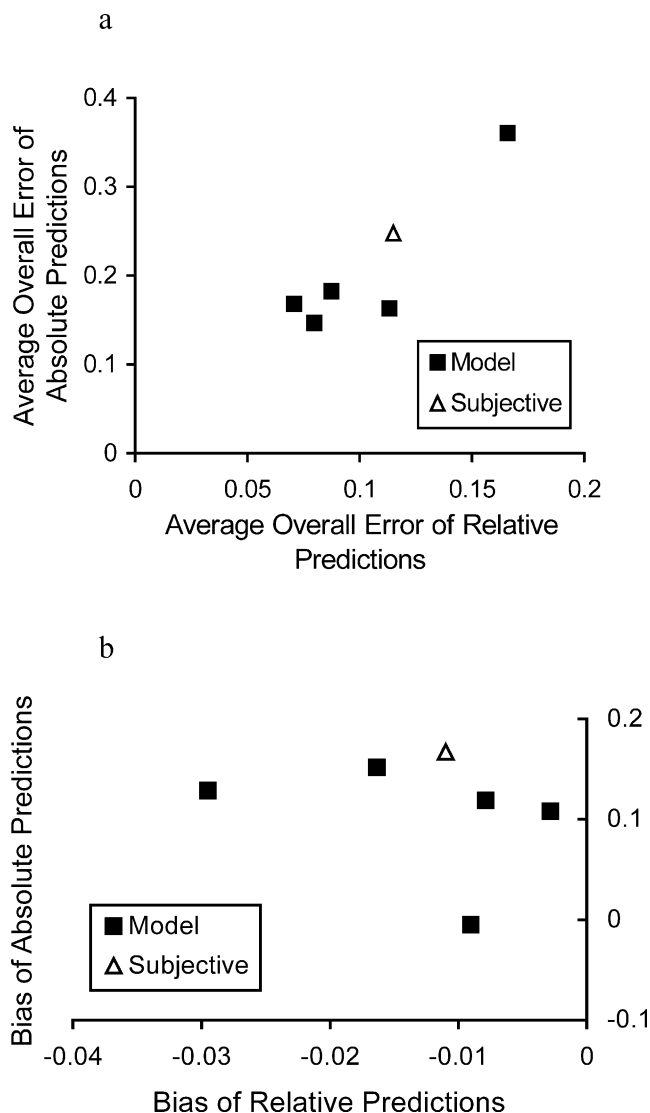


Fig. 3. Average overall error (a) and bias (b) of absolute predictions versus the average error of the relative predictions for models and subjective judgements in the spatial part of the competition. Each point represents an individual assessor, with the average taken across all the predictions of the five species.

all but one assessor (Figs. 2b and 3b). In the non-spatial component, this bias was predominantly due to the predicted effect of parameter changes on the small plant being under-

estimated. The bias in the relative predictions of the models for this species was -0.26 when averaged over assessors, while it was -0.03 for the small mammal, -0.03 for the frog and 0.02 for the snail. Similar biases were evident for the subjective judgements when averaged over the assessors (-0.34 for the small plant, -0.05 for the small mammal, -0.02 for the frog and -0.01 for the snail). Biases in the relative predictions for the species in the spatial component of the competition were of a similar magnitude for the five assessors who used models (0.02 for the butterfly, 0.07 for the owl, -0.14 for the passerine, 0.01 for the salmon and -0.02 for the shrub). The large biases in the relative risks for the passerine and small plant appear to have occurred because the inbreeding effects were large when fecundity was reduced, leading to larger relative risks than predicted by the assessors. Although the frog had an inbreeding component in the true model, it had little influence on the dynamics even when fecundity was reduced.

4. Discussion

Subjective judgements can be thought of as predictions based on a mental model of the dynamics of the species. The main question being addressed in this paper is whether the predictions of these mental models are superior to those that are explicitly written down and analysed mathematically. Our results suggest that the quality of the predictions of these methods is similar, although the models were slightly superior. Predictions of the mathematical models were marginally more accurate than subjective judgements and considerably less biased. The use of simulated species to evaluate the methods means that the results should be treated with some caution. For example, models may have been at an advantage, because the modellers knew that the data were themselves derived from models. However, people making subjective judgements had the same information, and the detail in the models meant that it was impossible for entrants to replicate the models by chance. The models used by the assessors were always considerably less detailed than the true model. Therefore, we do not believe that predictions based on models had an unfair advantage.

A second possible limitation is that a much greater range of information is usually used when making subjective judge-

Table 2

The average overall error and bias (S.E. in brackets) for the predictions for each of the nine hypothetical species, separated on the basis of predictions made with models. Averages are taken over the predictions of the assessors ($n = 9$ assessors for species without spatial structure and $n = 5$ for those with) and subjective judgement ($n = 10$ and $n = 1$)

Species	Average error for subjective judgements	Average error for models	Average bias for subjective judgements	Average bias for models
Small mammal	0.28 (0.04)	0.17 (0.03)	0.26 (0.05)	-0.02 (0.06)
Small plant	0.31 (0.02)	0.24 (0.03)	0.03 (0.03)	-0.15 (0.06)
Snail	0.19 (0.04)	0.24 (0.10)	0.18 (0.04)	0.23 (0.11)
Frog	0.31 (0.03)	0.27 (0.03)	0.31 (0.03)	0.09 (0.09)
Passerine	0.34 (-)	0.20 (0.04)	0.33 (-)	0.13 (0.06)
Salmon	0.26 (-)	0.12 (0.08)	0.26 (-)	0.06 (0.09)
Butterfly	0.17 (-)	0.26 (0.04)	0.02 (-)	0.12 (0.09)
Shrub	0.16 (-)	0.15 (0.10)	-0.09 (-)	-0.11 (0.10)
Owl	0.31 (-)	0.30 (0.01)	0.31 (-)	0.30 (0.01)

ments than that presented in the competition. However, we do not believe this is important because when available, this same information should also be used to construct models and interpret their predictions. Our study ensured that the same information was available to all entrants, regardless of the method they used to make their predictions. The question is whether the use of explicit models improves the predictions, because modellers would rarely ignore subjective judgements, in that they construct models that are thought to be reasonable. Compared to subjective judgement alone, our results suggest that the use of models can improve predictions of risk, especially by reducing bias, but the benefits may not be particularly large in terms of the accuracy of the predictions.

While the subjective judgements took between 1 and 2 h for most entrants (most of that time involved analysing the information provided), the predictions using models took on the order of 1–2 days. Given the comparatively good performance of the subjective judgements in terms of overall error, a question remains; is it worthwhile developing population models to predict risks? In terms of the usefulness of the predictions, it is possible that subjective judgements are cost-effective. This is especially true when one considers that unforeseen events have been largely ignored in this study (e.g. novel changes in the environment), and any difference in predictive performance may in reality be masked by these additional uncertainties. A noticeable case where assessors were surprised was when fecundity was reduced and inbreeding caused a larger than expected decline in the passerine and small plant. In this case, assessors under-estimated the change in the risk of decline. Unforeseen events, whether they have a positive or negative impact, are likely to lead to greater errors in both subjective judgements and model-based assessments than those reported here.

While the models performed only marginally better, they have other advantages over subjective judgements. When using models, the rationale behind the predictions is explicit, and models are open to analysis, criticism and modification when new information becomes available. Models can help ensure that predictions are not arbitrary or capricious, although there may be arbitrary choices when deciding on the structure of models. Nevertheless, these choices are stated in the structure of the model but are difficult to disclose when making subjective judgements. Assessments based on subjective judgement can give the illusion that they are not scientifically rigorous (Burgman, 2000), regardless of whether they are or not. The assumptions underlying models can be tested. Models can be used to help design data collection strategies. They can help to resolve and avoid inconsistencies, and the rigorous analysis of data can help to clarify thoughts. These benefits are difficult, if not impossible to achieve with subjective judgement.

An additional advantage of using models instead of subjective judgement is that the biases of the latter are likely to vary somewhat unpredictably among people, depending on their stake in the outcome (Tversky and Kahneman, 1974; Ayton and Wright, 1994; Gigerenzer and Hoffrage, 1995;

Anderson, 1998). In contrast, our results and previous work (Brook et al., 2000; Lindenmayer et al., 2003) suggest that provided there are no surprises (such as unexpected declines), models of population dynamics can be used to make unbiased predictions of risks and changes in risks due to management. We conclude that when the aim is to assess risks of decline or extinction, development of a model appears to be the most transparent, accurate and precise method of those available. Subjective judgements may have substantial utility, especially when resources and appropriate expertise for modelling are unavailable. However, models are often most valuable for their heuristic rather than predictive capacities (Brook et al., 2002).

Acknowledgements

We are grateful to all the people who contributed predictions, including those not listed as authors (R. DiGiovanni, B. Allen, S. Ball, S. Bekessy, D. Brito, Y. Chee, D. Johnson, J. Dyson, J. Elith, P. Foley, J. Fox, M. Ippolito, B. Kendall, E. Milliman, K. Parris, and B. Wintle). Thanks to E.J. Millner-Gulland for discussing ideas about the competition, to M. McDonnell, K. Parris, R. van der Ree and two anonymous referees for helpful comments on the manuscript, and to NCEAS for sponsoring the research and competition web-site.

Appendix. Descriptions of species

A range of simulated hypothetical species was chosen to reflect a variety of life history types (Table 1). Very detailed population models with various characteristics were written for each of these species. The characteristics considered were inbreeding depression, deterministic trends in abundance, trophic dynamics, and dynamics in response to weather, although not all features were included in all the models (Table 1). The aim of the model development was to produce a range of hypothetical species that had sufficient detail such that they could be considered as being representative of real species. Not all the attributes were included in each hypothetical species. This was partly to ensure that there was some diversity in the combination of the attributes.

The population dynamics of the small mammal were influenced by weather conditions, the availability of prey, age structure, density dependence and mate-finding ability. Weather was modelled by drawing standard normal deviates for rainfall and temperature for 6-month periods ("winter" and "summer"). Survival rates and fecundity rates both declined with population density according to a negative exponential function. Temperatures in each 6-month period were positively correlated, and precipitation was negatively correlated with temperature in winter and summer. Prey (nominally thought of as insects) were modelled using a stochastic Ricker density dependence function. Stochasticity in the prey dynamics entered through the influence of the temperature over the winter and the rainfall of the current summer. The quality of the

environment for the small mammal in summer was influenced by the abundance of prey and by an extra source of stochasticity. This influenced their survival and reproduction over summer. Survival of adults and newborns over winter was influenced by a combination of winter rainfall, winter temperatures and the quality of the preceding summer. Intermediate temperatures and rainfall lead to the highest survival rates. To incorporate senescence, survival of older adults was reduced by 5% for each year beyond their fifth. The rate of pregnancy of females declined with the density of males (Dennis, 1989; McCarthy, 1997). Inbreeding was not incorporated because the populations have been isolated for long periods on an ecological time scale. The data provided to the competition entrants were the number of individuals in each sex and age class immediately prior to breeding.

The population dynamics of the small plant were influenced by rainfall, age structure, inbreeding depression, a soil seed bank, and density dependence in survival, seed-production rates and selfing rates (proportion of a plant's seeds that are self-pollinated). Rainfall was modelled as a correlated process in a manner similar to that used for the small mammal. Most plants died after 1 year, although some lived into a second year. Inbreeding depression was modelled for each individual by assigning alleles to 10 genes for each individual. Genes were passed from adult plants to seeds by randomly assigning one allele for each gene from each parent. The selfing rate increased as population sizes declined. For non-selfed seeds, the source of the pollen was determined by choosing a plant randomly from the population. The level of heterozygosity in individuals (proportion of genes with the same two alleles) influenced their survival as seeds in the soil, their survival as adults, and their production of seeds. Germination rates and survival rates of adults to their second year were enhanced by annual rainfall. Seed-production and survival rates declined as population size increased. Sampling of the plant population was simulated by making annual counts in quadrats, assuming that the number of plants per quadrat was approximately distributed as a negative binomial distribution.

Rainfall, age structure, density dependence in survival, and cannibalism of eggs influenced the population dynamics of the snail. Recruitment and survival rates increased with rainfall. Younger individuals had lower fecundity rates. Survival rates declined and cannibalism rates of eggs increased with population density. The population was sampled using theory for estimating abundance in closed populations, with the chance of detection varying randomly with an average of approximately 10%.

Rainfall, age structure, density dependence in reproductive opportunities, and inbreeding depression influenced the population dynamics of the frog. Recruitment rates were maximised at intermediate levels of rainfall. Low levels limited tadpole development and high levels led to loss of eggs and tadpoles that were washed into a downstream dam. Younger individuals had lower fecundity rates. Survival rates were greatest for frogs of an intermediate age. Females had higher survival rates than males. Inbreeding was simulated in

the same way as for the small plant. Inbreeding depression acted only on survival of tadpoles. The population was sampled each year using simple surveys that mostly sampled a small proportion of the total population. The actual proportion seen, varied randomly and as a function of rainfall levels. No frogs were marked in the assessment.

The population dynamics of the passerine bird were influenced by age structure, density dependence in reproductive opportunities, inbreeding depression, and the spatial structure of the available habitat. The model was based on helmeted honeyeaters (McCarthy, 1996). Inbreeding depression (using the same method of assigning alleles as in the small plant) influenced survival of offspring, and territoriality was influenced by competition for breeding habitat that was arranged on a hexagonal grid. Dispersal rates between discrete habitat patches depended on the sex of the individual, with females dispersing further than males. The population was sampled by counting the number of breeding pairs and tracking the dispersal of banded offspring from a single population.

The population dynamics of the salmon were influenced by age structure, density dependence, fishing rates and movement of individuals among breeding populations. Fish returned to breed at either 3 years (half of the surviving fish), 4 years (80% of the surviving fish) or 5 years (all the surviving fish) of age. Most returned to spawning in their natal stream, but a small proportion (<5%) returned to other breeding sites. In 13 areas, spawning occurred in spring (headwater streams), while spawning occurred in autumn in two areas. Dispersal between these breeding areas was limited but did occur. The number of offspring produced, depended on the availability of suitable spawning sites, and their subsequent survival was also density dependent. All individuals died after spawning. Fishing removed spawning individuals from the population. In the future, a dam was to be removed and the quality of that stream as breeding habitat increased. The population was censused by counting the number of spawning individuals.

The population dynamics of the butterfly were influenced by weather, density dependence, parasitism and movement of individuals among breeding populations. The dynamics of a specialist parasitoid was controlled by its attack rate on the butterfly larvae and rate of spread among the 52 habitat patches. After metamorphosis, female butterflies searched for mates, laid eggs and possibly moved between habitat patches on a daily basis. Dispersal rates depended on the size of the current patch and the proximity and size of potential target patches. The daily survival rate (and hence the annual fecundity) depended on weather conditions. The survival rates of caterpillars depended on population density and the amount of rainfall. Impacts of generalist parasitoids were embedded within the density dependent function. The population was sampled by establishing line transects within each patch. The number seen was sampled from a negative binomial distribution, assuming that the average would be proportional to the density of butterflies within each patch. Additionally, results of a mark-recapture analysis from a single population were available.

The population dynamics of the shrub were influenced by weather, age structure, density dependence in reproductive opportunities, random fires, and the spatial structure of the available habitat. Habitat occurred continuously throughout a network of gullies, and was represented by an array of cells on a square grid. Germination only occurred following fires, when all burnt individuals died. Occurrence of fires within grid cells was spatially correlated (McCarthy and Lindenmayer, 1998). There was no soil-stored seed bank. The amount of seed on each plant was maximised at 40 years of age. The probability of fire increased with time since the last fire, and as the amount of rainfall in the previous year decreased. Flower production decreased with increases in the density of adult plants, and germination rates declined with increases in the density of germinants. The germination rate and survival rate of adults increased with rainfall. A small number of seeds dispersed to adjacent cells. The population was sampled by establishing quadrats in which the number of individuals were counted, assuming the counts were distributed as a negative binomial distribution.

The population dynamics of the owl were influenced by age structure, density dependence in reproductive opportunities, and logging of habitat. Suitable habitat was distributed among 16 patches. The owl was relatively long-lived with low reproductive output. Juveniles were more likely to disperse to nearby rather than more distant habitat patches, but the majority stayed in their natal patch. Once a breeding territory was established, adults did not move from that patch. The number of breeding territories within each patch was limited by the amount of suitable breeding habitat. Logging at various rates in different patches reduced the amount of habitat available. Sampling was conducted by recording the response of owls to amplified territorial calls at several points within each patch.

References

- Anderson, J.L., 1998. Embracing uncertainty: the interface of Bayesian statistics and cognitive psychology. online, URL.
- Ayton, P., Wright, G., 1994. Subjective probability: what should we believe? In: Wright, G., Ayton, P. (Eds.), *Subjective Probability*. Wiley, New York, pp. 163–184.
- Beissinger, S.R., Westphal, M.I., 1998. On the use of demographic models of population viability in endangered species management. *J. Wildlife Manage.* 62, 821–841.
- Boyce, M.S., 1992. Population viability analysis. *Ann. Rev. Ecol. Syst.* 23, 481–506.
- Brook, B.W., O'Grady, J.J., Chapman, A.P., Burgman, M.A., Akçakaya, H.R., Frankham, R., 2000. Predictive accuracy of population viability analysis in conservation biology. *Nature* 404, 385–387.
- Brook, B.W., Burgman, M.A., Akçakaya, H.R., O'Grady, J.J., Frankham, R., 2002. Critiques of PVA ask the wrong questions: throwing the heuristic baby out with the numerical bath water. *Conserv. Biol.* 16, 262–263.
- Burgman, M.A., 2000. Population viability analysis for bird conservation: prediction, heuristics, monitoring and psychology. *Emu* 100, 347–353.
- Burgman, M.A., Ferson, S., Akçakaya, H.R., 1993. *Risk Assessment in Conservation Biology*. Chapman Hall, London.
- Committee on Scientific Issues in the Endangered Species Act, 1995. *Science and the Endangered Species Act*. National Academy Press, Washington, DC.
- Cooke, J.G., 1995. The International Whaling Commission's Revised Management Procedure as an example of a new approach to fisheries management. In: Blix, A.S., Walløe, L., Ulltang, Ø. (Eds.), *Whales, Seals, Fish and Man*. Elsevier, Amsterdam, pp. 647–657.
- Dennis, B., 1989. Allee effects: population growth, critical density, and the chance of extinction. *Natural Resource Model.* 3, 481–538.
- Dennis, B., Munholland, P.L., Scott, J.M., 1991. Estimation of growth and extinction parameters for endangered species. *Ecol. Monogr.* 61, 115–143.
- Ellner, S.P., Fieberg, J., Ludwig, D., Wilcox, C., 2002. Precision of population viability analysis. *Conserv. Biol.* 16, 258–261.
- Fieberg, J., Ellner, S.P., 2000. When is it meaningful to estimate an extinction probability? *Ecology* 81, 2040–2047.
- Fieberg, J., Ellner, S.P., 2001. Stochastic matrix models for conservation and management: a comparative review of methods. *Ecol. Lett.* 4, 244–266.
- Foley, P., 1994. Predicting extinction times from environmental stochasticity and carrying capacity. *Conserv. Biol.* 8, 124–137.
- Gigerenzer, G., Hoffrage, U., 1995. How to improve Bayesian reasoning without instruction: frequency formats. *Psychol. Rev.* 102, 684–704.
- Hanski, I., 1997. Metapopulation dynamics: from concepts and observations to predictive models. In: Hanski, I., Gilpin, M.E. (Eds.), *Metapopulation Biology: Ecology, Genetics and Evolution*. Academic Press, San Diego, pp. 69–91.
- Lindenmayer, D.B., Possingham, H.P., Lacy, R.C., McCarthy, M.A., Pope, M.L., 2003. How accurate are population models? Lessons from landscape-scale tests in a fragmented system. *Ecol. Lett.* 6, 41–47.
- Ludwig, D., 1999. Is it meaningful to estimate a probability of extinction? *Ecology* 80, 298–310.
- Maunder, M.N., 2004. Population viability analysis based on combining Bayesian, integrated, and hierarchical analyses. *Acta Oecol.* 26 this issue.
- McCarthy, M.A., 1996. Extinction dynamics of the helmeted honeyeater: effects of demography, stochasticity, inbreeding and spatial structure. *Ecol. Model.* 85, 151–163.
- McCarthy, M.A., 1997. The Allee effect, finding mates and theoretical models. *Ecol. Model.* 103, 99–102.
- McCarthy, M.A., Lindenmayer, D.B., 1998. Multi-aged mountain ash forest, wildlife conservation and timber harvesting. *Forest Ecol. Manage.* 104, 43–56.
- McCarthy, M.A., Broome, L.S., 2000. A method for validating stochastic models of population viability: a case study of the mountain pygmy-possum (*Burramys parvus*). *J. Anim. Ecol.* 69, 599–607.
- McCarthy, M.A., Burgman, M.A., Ferson, S., 1996. Logistic sensitivity and bounds for extinction risks. *Ecol. Model.* 86, 297–303.
- McCarthy, M.A., Lindenmayer, D.B., Possingham, H.P., 2000. Testing spatial PVA models of Australian treecreepers (Aves: Climacteridae) in fragmented forest. *Ecol. Appl.* 10, 1722–1731.
- McCarthy, M.A., Possingham, H.P., Day, J.R., Tyre, A.J., 2001. Testing the accuracy of population viability analysis. *Conserv. Biol.* 15, 1030–1038.
- McCarthy, M.A., Thompson, C., 2001. Expected minimum population size as a measure of threat. *Anim. Conserv.* 4, 351–355.
- McCarthy, M.A., Andelman, S.A., Possingham, H.P., 2003. Reliability of relative predictions in population viability analysis. *Conserv. Biol.* 17, 982–989.
- Millner-Gulland, E.J., Shea, K., Possingham, H., Coulson, T., Wilcox, C., 2001. Competing harvesting strategies in a simulated population under uncertainty. *Anim. Conserv.* 4, 157–167.
- Possingham, H.P., Lindenmayer, D.B., Norton, T.W., 1993. A framework for improved threatened species management using population viability analysis. *Pacific Conserv. Biol.* 1, 39–45.
- Shea, K., 1998. Management of populations in conservation, harvesting and control. *Trends Ecol. Evol.* 13, 371–375.
- Taylor, B.L., 1995. The reliability of using population viability analysis for risk classification of species. *Conserv. Biol.* 9, 551–558.
- Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: heuristics and biases. *Science* 185, 1124–1131.
- Wilcox, C., Possingham, H., 2002. Do life history traits affect the accuracy of diffusion approximations for mean time to extinction? *Ecol. Appl.* 12, 1163–1179.