Models, Problems and Algorithms: Perceptions about their Application to Conservation Biology?

H. P. Possingham

Departments of Mathematics and Zoology, University of Queensland, QLD 4072, Australia (hpossingham@zen.uq.edu.au)

Abstract: Conservation biologists are increasingly looking to quantitative tools to help them make decisions. I discuss the perceptions and misconceptions that surround two sorts of quantitative approach widely used in conservation biology: population viability analysis and reserve system design. The intent is that this discussion allows users to perceive these tools in a more balanced way – as neither useless nor the answer to all the world's problems – but rather as generally useful tools in supporting conservation management decisions. The difficulties that are experienced with respect to applying quantitative methods in conservation biology are no doubt common to their application to other disciplines where the majority of workers do not have an extensive mathematical background.

Keywords: Population modelling; Algorithms; Conservation biology; Reserve system design

1. INTRODUCTION

Mathematics is playing a growing role in conservation biology as evidenced by the recent book of chapters "Quantitative Methods for Conservation Biology" edited by Ferson and Burgman [2000]. This increasing role mirrors similar changes in many other biological disciplines and it is associated with a certain amount of pain. That pain revolves around concerns, perceptions and uncertainty regarding quantitative methods expressed by less quantitatively oriented ecologists and environmental managers. This paper discusses some of these issues based on experience in the application of two widely used quantitative tools: Population Viability Analysis (PVA) and reserve selection algorithms. Many of these concerns are issues of communication and semantics. The hope is that this paper will clarify some of these confusions. I suspect that our experience is more broadly applicable to all quantitative scientists working in fields that are traditionally less quantitative.

2. POPULATION VIABILITY ANALYSIS

2.1 Background

Population viability analysis is a modelling tool that is used to estimate extinction risks for species, or populations of species [Possingham et al., 2001b]. In most cases it relies on a suite of well-known population modelling methods including Monte Carlo simulation, Leslie matrix methods, Markov chains and stochastic differential equations [Beissinger and Westphal, 1998]. While these population models have existed for many decades, rooted in the paradigm that the size of a population is determined by birth, death, emigration and immigration, the idea of estimating extinction risk is relatively new [Shaffer, 1981]. The capacity to estimate extinction risk over different time frames is an appealing concept for conservation biologists, with the potential to underpin endangered species legislation and national programs for the management of endangered species. PVA models have been used for:

- Assessing the viability of a population.
- Determining minimum viable habitat areas.
- Ranking threatened species (often for funding purposes).
- Making endangered species management decisions.
- Bringing together experts to discuss a species within the context of a common goal.

These diverse roles are summarised in Burgman and Possingham [2000]. The use of PVA has increased rapidly. A search of the Web of Science using Population Viability Analysis as a "topic word" shows the following trend in the number of papers: 1990-3, 1995-22, 2000-39. Despite this growing popularity practitioners express considerable uncertainty and concern about the validity and use of PVA.

2.2 PVA Models are Wrong

There is growing criticism of PVA models with respect to their stated purpose of estimating the risk of extinction. Indeed Ludwig [1999] and Fieberg and Ellner [2000] have gone so far to say that for most wild populations the magnitude of uncertainty about the parameters needed in PVA models is so great that our estimates of mean time to extinction and probability of extinction are unlikely to be accurate for more than a few years in to the future. This contrasts the conclusions of Brook et al. [2000] who tested the capacity of several PVA models to predict the future using time series data on known populations. Coulson et al. [2000], McCarthy et al. [2001b] and others have made strong, and different, arguments that Brook et al.'s [2000] optimistic claims are unfounded leaving us with a modeling tool that, at first sight, appears to have limited predictive capacity. Given this recent literature and supporting comments at an international conference on PVA in 1998 it is tempting to dismiss PVA as generally useless, except in the few circumstances where the available data is exceptional in both quality and the length of time over which it has been gathered.

However this conclusion only holds if our interest in PVA is for purposes 1-3 listed above. Each of these purposes relies on an accurate assessment of extinction risk. Purpose 5, facilitating communication and data sharing, clearly does not rely on accurate predictions of risk. We have also found that the capacity to rank management options, purpose 4, is robust with respect to considerable uncertainty with respect to extinction risk [Lindenmayer and Possingham, 1996]. The utility of the approach is very context dependent. We now believe that the original applications of PVA were probably flawed, as the tool is unable to accurately answer the questions posed. PVA is poor at predicting absolute risk but does seem to be very good at determining the best way to manage a threatened species, a question which is probably equally, if not more, important than assessing viability. From the perspective of the ecologist and wildlife manager it is tempting to make the mistake of throwing the baby out with the bathwater. It can be challenging to convince ecologists and wildlife managers who read papers damning PVA as a tool, that by invoking a subtle change of question, a tool that was relatively useless, becomes essential.

2.3 Testing PVAs

We are often taught to believe that all models should be tested, and eventually rejected and replaced. Those ecologists trained in the rigor of null hypothesis testing are keen to take PVA models and subject them to repeated tests until they fail: this approach is misguided.

First the primary output of interest is risk of extinction of a threatened species and since a species only goes extinct once, testing an expectation or probability distribution with one data point, the extinction of that species, is silly. Brook et al. [2000] tackled this problem by using the first section of a time series to predict the second part of a time series. Lindenmayer et al [2000] and McCarthy et al. [2001b] show how data on species distribution in fragmented habitats provides a way of testing a spatially explicit PVA. In these papers we recognize that all models are wrong, so rejecting a modeling approach is not useful. Instead we took the approach that where our initial tests exposed some significant differences between predictions and data then our role was to refine (calibrate) the model. Unfortunately we can change a large number of parameters and processes to fit a complex model to small sets of data. We have relied on our intuition to come up with the best plausible alternative model modifications. We see the process of model testing and refinement as iterative and never-ending. This is important for applied population biology but does not sit well with a more Popperian view of science [Hilborn and Mangel, 1997]. The fact that intuition and judgment are involved leads some ecologists to feel even less comfortable with PVA.

2.4 What if we Reject PVAs?

If we rush to reject PVA as a useful pursuit, either through its incapacity to accurately predict extinction risk, or through our inability to test and validate the models, it is prudent to ask ourselves how to answer the questions asked of PVA by any other means? For example in the case of choosing management options, without a PVA how can the best choice be made? Ultimately to make a decision we need to know how different management actions will affect the risk of extinction. In the absence of more formal PVA we can only rely on our intuition and experience. Intuition and experience can be formalized, and indeed used to form the basis of qualitative PVAs. Indeed it is not really a question of whether or not to do a PVA - you have to - it is more a question of how formal you wish to make the model that translates actions in to outcomes.

A recent focus of PVA as part of a decision-making process helps to put its role in perspective [Possingham et al., 2001b]. If we acknowledge that decisions need to be made, that actions must be tied to outcomes, and that the process must be transparent and justifiable, then we need a PVA in some form or other. This focus on decision-making leads us to our next example where the decision involves how to make an effective and efficient nature reserve system.

3. RESERVE SYSTEM DESIGN

Before discussing reserve system design let us step back and think through the process of making a decision.

3.1 Decision-Making in Conservation

In the broadest sense we can think of nature conservation decisions as involving three components:

- Defining the *problem.*
- Describing the way the system works using *models.*
- Using an *algorithm* to find good solutions to the problem.

These components are described in detail by Possingham et al. [2001a] in the context of conservation biology. I have found that a major challenge in delivering decision-support to conservation biologists and managers is exposing the role of each component. I use the case of reserve system design, or more generally ecoregional planning, to expose the differences between the three quantitative components of decision-making.

3.2 The Reserve System Design Problem

In a seminal paper Cocks and Baird [1989] defined the minimum set reserve design problem as one of representing all biodiversity features in a region for overall minimum cost. This notion has become a baseline problem for reserve selection methods. The simplest formulation of the problem is as follows.

Let the total number of sites be m and the number of different features (which may be species or vegetation types) to be represented in the final reserve system be n. The information about whether or not a feature is found in a site is contained in a site-by-feature (m x n) matrix **A** whose elements a_{ij}, are

$$a_{ij} = \begin{cases} 1 \text{ if feature j occurs in site i} \\ 0 \text{ otherwise} \end{cases}$$

for $i = 1,\ldots,m$ and $j = 1,\ldots,n$.

Next, define a control variable, that determines whether or not a site is included in the reserve, as the vector **X** with dimension m and elements x_i, given by

$$x_i == \begin{cases} 1 \text{ if site } i \text{ is included in the reserve} \\ 0 \text{ otherwise} \end{cases}$$

for $i = 1,\ldots,m$.

With these definitions, the minimum representation problem is an explicit objective:

minimise $\sum_{i=1}^{m} x_i$ {minimise the number of sites in the reserve system}

subject to a number of constraints:

$\sum_{i=1}^{m} a_{ij}x_i \geq 1$, for $j = 1,\ldots,n$ {subject to each species being represented at least once}

where $a_{ij}, x_i \in \{0,1\}$.

This is the integer linear programming formulation of the set-covering problem [Possingham et al., 2000]. It is NP-Complete so that the difficulty of guaranteeing an optimum solution increases exponentially with the number of constraints n. For the kinds of practical problems we have tackled where the number of sites is more than 10,000 and the number of features is in the hundreds, the number of possible solutions is so vast that any hope of finding an optimal solution can be discarded [Ball, 2000].

There are numerous ways in which this basic problem can be modified to allow for the things that conservation biologists see as important. For example if the data matrix includes information about the area of a vegetation type in a site, or the size of a population of a species in a site then the target is not just to represent the feature but to choose enough sites so the vegetation type of species is adequately conserved. In early work algorithms paid no attention to the spatial pattern of sites chosen for a reserve system. In our recent work our objective function is to minimize a linear combination of the total area (or costs) of sites selected and the boundary length of the final reserve system. This makes the problem non-linear and means that the final reserve system is more spatially cohesive – an important part of a good reserve system. My intent here is not to discuss all the possible ways of formulating the reserve design problem, but more to give an example of the kind and scale of problem concerned. In particular the Great Barrier Reef marine Park Authority is solving such a problem where the number of sites is about 13,000, and The Nature Conservancy are looking at problems with 20,000 sites and several hundred constraints.

3.3 The Algorithms to Solve the Problem

There are many ways to find good solutions to different versions of the reserve design problem. Early methods relied on heuristic methods such as richness, rarity and greedy algorithms [Pressey et al., 1997]. More recently Ball [2000] has used the simulated annealing algorithm to create a flexible tool, MARXAN, which finds good solutions to many variations of the basic reserve system design problem [Kirkpatrick et al., 1983].

In our application of MARXAN we are repeatedly queried about the validity of the "model". Our first task is to explain that MARXAN (and its variants) are not models, but algorithms. For small problems they almost certainly get very close to finding one or more optimal solutions. Our tests [Pressey et al., 1997] on systems where the true optimal value has been found using branch and bound integer linear programming methods shows these methods to be remarkably efficient and fast.

It would appear that the "black-box" and computer-based nature of algorithms like simulated annealing lead to suspicion. This suspicion regarding the algorithm detracts attention from where users real suspicion should lie – in the formulation of the problem and the nature of the "models" that invariable underlie the relationship between control variables and outcomes.

3.4 Models that Underlie Decision-Making Problems

Behind any problem there must be ways of connecting the decision variables, like what sites to reserve, with outputs, in this case our capacity to conserve viable populations or landscapes. These "models" are often vague. In practice there are two areas where "models" are hidden within the formulation of the problem. The first hidden model is the analysis of data that determines what features are in a site, whether those features be species or habitat types. These models are often statistical models that include considerable uncertainty; however, because they can be represented as maps they are rarely challenged! The second hidden suite of models concerns setting the targets for conservation. In practical examples the constraints are usually to represent x% of each habitat type, or x% of the range of a species. The justification of x is invariably quite rubbery. Given the enormous uncertainty about the models that underpin the data and problem formulation it has been remarkable to us how much attention is paid to the algorithm used to solve the problem.

3.5 Problems, Models and Algorithms

It is only recently that I have been able to dissect the concern that conservation biologists and nature conservation managers hold for the role of mathematics in decision support. I believe that most of their concerns stem from their inability to untangle the relationship between the three mathematical components of decision theory: problem, models and algorithm.

The definition of the problem should be of most concern. It is how we set objectives and constraints, or indeed what we call an objective or a constraint that will have the most profound affect on any conservation planning or management outcome. For example, in the reserve design problem we could have set the area to be conserved as a constraint and the objective could have been to maximize the number of features conserved. Reformulating the problem like this has profound social and economic implications and will change our final solution dramatically.

The statistical and verbal models that underpin the data used in the problem should also be subject to careful scrutiny. In particular any theory or model used to set the target area or size of a feature will have a huge impact on the result.

Despite the importance of the problem definition and the models hidden in the problem, there is always remarkable interest in the algorithm. I have often been asked to test the algorithms on real systems, yet in some cases there are mathematical proofs available to show the algorithm guarantees and optimal solution. Even where we cannot guarantee finding an optimal solution, the error incurred by the algorithm is invariably orders of magnitudes smaller than minor changes in problem formulation. I have often found that the only way to disentangle the three components and hence rectify misconceptions about their relative importance is to use an extended analogy that relates to a problem people encounter in everyday life.

4. CONCLUSIONS

The tension between conservation and quantitative decision theory is both intriguing and frustrating. It challenges us to be even more explicit about how we express mathematical ideas and use them to aid decision-making. One of the most important messages to sell at an early stage of any dialogue is that mathematics and computers deliver decision support not decision-making. No one will ever feel comfortable with computer programs making decisions where there is interplay of social values, uncertainties and complex interactions. In this sense decision support tools are probably best treated as tools for education not prescription.

5. ACKNOWLEDGEMENTS

The following people have contributed, through ideas and discussion, to the content of this paper: Ian Ball, David Lindenmayer, Jonathan Rhodes, Sandy Andelman, Bob Pressey, Drew Tyre, Brigitte Tenhumberg, Scott Field, Stephen Ball, Michael McCarthy, Mark Burgman, Kat Shea and Michael Westphal.

6. REFERENCES

Ball, I.R., Mathematical applications for conservation ecology: the dynamics of tree hollows and the design of nature reserves, PhD thesis, The University of Adelaide, 2000.

Beissinger, S.R. and M.I. Westphal, On the use of demographic models of population viability in endangered species management, *Journal of Wildlife Management*, 62, 821-841, 1998.

Brook B.W., J.J. O'Grady, A.P. Chapman, M.A.Burgman, H.R. Akçakaya and R. Frankham, Predictive accuracy of population viability analysis in conservation biology, *Nature* 404, 385-387, 2000.

Burgman, M.A. and H.P. Possingham, Population viability analysis for conservation: the good, the bad and the undescribed, In: *Genetics, Demography and Viability of Fragmented Populations*, A.G. Young and G.M. Clarke (eds.), Cambridge University Press, London, 97-112, 2000.

Cocks, K. D. and I.A. Baird, Using mathematical programming to address the multiple reserve selection problem: an example from the Eyre Peninsula, South Australia, *Biological Conservation,* 49, 113-130, 1989.

Coulson T., G.M. Mace, E. Hudson and H. Possingham, The use and abuse of population viability analysis, *Trends in Ecology and Evolution*, 16, 219-222, 2001.

Ferson, S. and M. Burgman (eds), *Quantitative Methods for Conservation Biology*, Springer-Verlag, New York, 2000.

Feiberg, J. and S.P. Ellner, When is it meaningful to estimate an extinction probability? *Ecology*, 81, 2040-2047, 2000.

Kirkpatrick, S., C.D. Gelatt and M.P. Vecchi, Optimization by simulated annealing, *Science*, 220, 671-680, 1983.

Hilborn, R. and M. Mangel, *The Ecological Detective*, Princeton University Press, Princeton, 1997.

Lindenmayer, D.B, M.A. McCarthy, H.P. Possingham and S. Legge, A simple landscape-scale test of a spatially explicit population model: patch occupancy in fragmented south-eastern Australian forests, *Oikos*, 92, 445-458, 2000.

Lindenmayer, D.W. and H.P. Possingham, Ranking conservation and timber management options for Leadbeater's Possum in southeastern Australia using population viability analysis, *Conservation Biology*, 10, 235-251, 1996.

Ludwig, D., Is it meaningful to estimate a probability of extinction? *Ecology*, 80, 298-311, 1999.

McCarthy M.A., D.B. Lindenmayer and H.P. Possingham, Assessing spatial PVA models of arboreal marsupials using significance tests and Bayesian statistics, *Biological Conservation*, 98, 191-200, 2001a.

McCarthy M.A., H.P. Possingham, J.R. Day, and A.J. Tyre, Testing the accuracy of population viability analysis, *Conservation Biology*, 15, 1030-1038, 2001b.

Possingham, H.P., I.R. Ball and S.J. Andelman, Mathematical methods for identifying representative reserve networks, In:

Quantitative Methods for Conservation Biology, S. Ferson, and M. Burgman (eds.), Springer-Verlag, New York, 291-305, 2000.

Possingham, H.P., S.J. Andelman, B.R. Noon, S. Trombulak, and H.R. Pulliam, Making smart conservation decisions, In: *Research Priorities for Conservation Biology,* G. Orians, and M. Soulé (eds.), Island Press, California, 2001a.

Possingham, H.P., D.B. Lindenmayer, and M.A. McCarthy, Population viability analysis, In: *The Encyclopedia of Biodiversity*, S. Levin (ed.), Academic Press, San Diego, 831-844, 2001b.

Pressey, R. L., H.P Possingham, and J.R. Day, Effectiveness of alternative heuristic algorithms for identifying indicative minimum requirements for conservation reserves, *Biological Conservation*, 80, 207-219, 1997.

Shaffer, M.L. Minimum population sizes for species conservation, *Bioscience*, 31,131-134, 1981.