Testing the Accuracy of Population Viability Analysis

M. A. McCARTHY,*†§ H. P. POSSINGHAM,*† J. R. DAY,+ AND A. J. TYRE†

*National Center for Ecological Analysis and Synthesis, University of California, 735 State Street, Suite 300, Santa Barbara, CA 93101, U.S.A.
†Department of Applied and Molecular Ecology, The University of Adelaide, Waite Campus, Adelaide, South Australia 5064, Australia
‡Centre for Research on Introduced Marine Pests, Commonwealth Scientific, Industrial, and Research Organization Marine Research, GPO Box 1538, Hobart TAS 7001, Australia

Abstract: Despite the growing use of population viability analysis (PVA), the predictions of these models rarely have been tested with field data that were not used in initially developing the model. We review and discuss a suite of methods that may be used to test the predictive ability of models used in PVA. In addition to testing mean predictions, appropriate methods must analyze the probability distribution of the model predictions. The methods we discuss provide tests of the mean predictions, the predicted frequency of events such as extinction and colonization, and the predicted probability distribution of state variables. We discuss visual approaches based on plots of observations versus the predictions and statistical approaches based on determining significant differences between observations and predictions. The advantages and disadvantages of each method are identified. The best methods test the statistical distribution of the predictions; those that ignore variability are meaningless. Although we recognize that the quality of a model is not solely a function of its predictive abilities, tests help reduce inherent model uncertainty. The role of model testing is not to prove the truth of a model, which is impossible because models are never a perfect description of reality. Rather, testing should help identify the weakest aspects of models so they can be improved. We provide a framework for using model testing to improve the predictive performance of PVA models, through an iterative process of model development, testing, subsequent modification and re-testing.

Probando la Precisión del Análisis de Viabilidad Poblacional

Resumen: A pesar del incremento en el uso del Análisis de Viabilidad Poblacional (AVP), las predicciones de estos modelos raramente han sido probadas con datos de campo que no fueron utilizados para el desarrollo inicial del modelo. Revisamos y discutimos una serie de métodos que pueden ser usados para probar la habilidad predictiva de los modelos utilizados en el AVP. Además de probar las predicciones promedio, los métodos apropiados deben analizar la distribución de probabilidades de las predicciones de los modelos. Los métodos que discutimos proporcionan pruebas de las predicciones promedio, de la predicción de la frecuencia de eventos como extinciones y colonizaciones y de la distribución de probabilidades de variables de estado prevista. Discutimos aproximaciones visuales basadas en parcelas de las observaciones versus las predicciones y aproximaciones estadísticas basadas en la determinación de diferencias significativas entre las observaciones y las predicciones. Se identifican las ventajas y desventajas de cada método. Los mejores métodos prueban la distribución estadística de las predicciones; las que ignoran la variabilidad no tienen sentido. Aunque el reconocimiento de que la calidad de un modelo no solo está en función de sus habilidades predictivas, las pruebas ayudan a reducir la incertidumbre inherente a los modelos. El papel de la prueba de los modelos no es el de probar la verdad de un modelo, lo cual es imposible porque los modelos nunca son una descripción perfecta de la realidad. Por lo contraria, probar los modelos debería ayudar a identificar sus aspectos más débiles de los modelos para que sean mejorados. Proporcionamos un marco de trabajo para probar modelos para mejorar el desempeño predictivo de los modelos de AVP, por medio de un proceso iterativo de desarrollo de modelos, probarlos, modificarlos y probarlos nuevamente.

‡Current address: School of Botany, University of Melbourne, Parkville Victoria 3010, Australia, email mamcca@unimelb.edu.au
Paper submitted July 12, 1999; revised manuscript accepted September 20, 2000.

1030
Introduction

Population viability analysis (PVA) has been used in conservation biology for approximately 20 years (Shaffer 1981; Gilpin & Soulé 1986). Development of stochastic population models for PVA has proliferated recently, but uncertainty about model parameters and processes make the predictions of PVA subject to considerable uncertainty (Possingham et al. 1993; Taylor 1995; McCarthy et al. 1996; Beissinger & Westphal 1998; Ludwig 1999). Despite this uncertainty, the quantitative predictions of PVA models have rarely been tested with field data (Lacy 1993; Brook et al. 1997; Lindenmayer et al. 1999; Brooker et al. 2000; McCarthy & Broome 2000; McCarthy et al. 2000). Testing procedures are available for deterministic models, but there are few established methods for testing stochastic population models (Reynolds et al. 1981). Conducting appropriate tests is difficult because of apparent requirements for large data sets to assess the statistical properties of the predictions (Burgman et al. 1993). For any given species or system, little replication is possible and extinctions may be rare, so it is difficult to test the risk of extinction, which is the primary prediction of PVA models (Beissinger & Westphal 1998). Instead, testing of PVA models is usually confined to secondary predictions of population dynamics prior to extinction. For PVA models to be useful tools for management, it is desirable to test their predictions. For many population models, annual changes in abundance are likely to be the most easily tested prediction (e.g., Lacy 1993). In metapopulation models, the most readily tested predictions are likely to be rates of local extinction and recolonization and patterns of patch occupancy (e.g., Hanski 1997).

In contrast to deterministic models that focus on central tendencies (e.g., mean population trajectories), stochastic population models usually predict probability distributions of population size. Testing PVA models by comparing mean predictions with observations is insufficient because variability, which is a critical factor influencing the risk of population decline, is ignored (Burgman et al. 1993; Lacy 1993). In addition, several authors have argued that the relative (and qualitative) results of PVA models are more important than the absolute predictions because predictions of models are subject to uncertainty (Possingham et al. 1993; Lindenmayer & Possingham 1994, 1996; Beissinger & Westphal 1998). Therefore, methods for testing these aspects of PVA models would also be useful.

At the outset, it is important to recognize that no model can be a perfect description of reality. Attempts to prove that the predictions of models are “true” or “false” are misguided. No amount of model testing can prove a model true because its predictions will always be inaccurate in at least some respects (Rytkönen 1996). The process of testing PVA models should aim to identify their weakest aspects so that they can be modified and their predictions improved. Testing may be used to identify a model’s limits of accuracy and to determine circumstances in which its predictions are unreliable as a conservation tool. Tests of competing models can indicate whether a particular model should be rejected in favor of an alternative.

Our aim is not to review methods of assessing model parameters and assumptions or selecting among competing models with different structures. These aspects are fundamental parts of the scientific process and are covered in detail throughout the literature (e.g., Sokal & Rohlf 1981; Lebreton et al. 1992; Hilborn & Mangel 1997; Burnham & Anderson 1998). Although such investigations should be a fundamental aspect of developing a PVA model, they can only assess known parameters and assumptions. Testing aggregate predictions, the aim of this paper, can help to identify errors that were previously unrecognized.

We used field data collected independently of the data used to construct the model to identify methods for testing secondary predictions of PVA models. The processes of parameter estimation (model fitting) and model testing should not be confused (Lacy 1993). Data sets used for parameter estimation tend to fit the model better than an independent data set used in testing, given equivalent data quality. Some of the methods we reviewed are suitable for testing either single-population models or metapopulation models, whereas others are suitable only for metapopulation models. Data we used to test the predictions included population size, patterns of patch occupancy, and colonization and extinction of patches. We drew on literature from several fields and suggest a set of techniques for examining the predictive performance of PVA models. We discuss the good and bad features of the different methods and make specific recommendations about which tests are likely to be most useful and those that are meaningless. We discuss how these methods may be used to help improve the predictions of the models.

Methods for Testing PVA Models

We reviewed a range of methods for testing PVA models (Table 1). The methods were reviewed in four different categories, reflecting the underlying philosophy of the tests and the data used.

Testing Mean Population Sizes

Brook et al. (1997) made one of the first attempts to test PVA models by comparing the predicted trajectory of the mean population size to the single observed population trajectory. Similarly, others compared predictions to observations using correlation or regression analyses
This approach can be applied to assessing the dynamics of individual populations or aggregate dynamics across a metapopulation. It is similar to testing the predictions of a deterministic model in that stochastic variation is ignored. For several reasons, we do not think this approach is meaningful. These methods may indicate whether there is a positive relationship between the average model predictions and the observations, but there are important deficiencies. First, if the model is stochastic and predicts variation in population size, a strong correlation between the mean and the observed trajectory might occur because the observed population dynamics are substantially less variable than the model is predicting.

To illustrate the limitations of testing only mean predictions, we constructed a simple stochastic population model (Fig. 1a) and calculated the mean predicted population (Fig. 1b). When the mean trajectory is compared to data generated by the same model from individual runs, the correlation between the two may be substantially positive or substantially negative, depending on the individual run (Fig. 1c). The data from each individual run can be thought of as potential "data," where the model is a perfect description of reality. Even in this case,
substantially negative correlations may occur because the actual trajectory can be very different from the mean trajectory.

If the mean predicted trajectory changes only a little over time relative to the variation in population size between runs, a weak correlation may be expected. In contrast, if the model predicts large variation within runs relative to the variation between runs (e.g., variation due to deterministic population cycles is greater than stochastic variation in population size), a strong relationship between the observed trajectory and the predicted mean trajectory may be expected. Although this may test the deterministic component of the model, the method remains deficient because the predicted variation in the model is ignored. When only mean predictions are compared to observations, there is no indication of how accurately the variation in population size is modeled. Thus, interpreting the importance of the relationship between the observed and predicted mean trajectory depends on the level of variability. Meaningful analyses should accommodate this aspect, but this is not achieved by comparing mean predicted trajectories, either visually, with correlations, or with regression analysis. Nevertheless, comparing the observed trajectory to the predicted distribution of trajectories (rather than just the mean or median) may give a qualitative indication of whether the predictions are reasonable.

Brook et al. (2000) tested the median prediction of 21 different PVA models by calculating the signed predictive bias, a measure of whether the models (on aggregate) over- or underpredicted the final population abundance. Such an analysis relies on the assumption that the separate predictions are independent of one another, which was a reasonable assumption for the 21 models of Brook et al. (2000). Brook et al. (2000) used this test to demonstrate that there was no consistent bias in the estimation of the median population size when it was averaged over all their studies. This is a useful test when combined with assessments of the predicted variation in population size.

Testing Extinction and Colonization Events and Patch Occupancy

Metapopulation models predict the occurrence of probabilistic events such as local extinction, colonization, and occupancy. For example, using an incidence-func-

\[
\text{variance of 0.1. (b) Mean population trajectory from this stochastic model (continuous line) compared with two different individual trajectories (broken lines). Actual trajectories may be very different from predicted mean trajectory, with the correlation coefficient between the two being highly variable. For the trajectories shown in (b), the correlation coefficient is } -0.50 \text{ for one (c, circles) and 0.82 for the other (c, crosses).}
\]
tion metapopulation model, Hanski (1997) predicted the annual extinction and colonization rates of three species of *Sorex* shrews over a 5-year period on 17 small islands. When compared to data that were not used to make the predictions, the predicted number of colonization and extinction events closely matched the observations for each of the three species, indicating that the mean extinction and colonization rates were predicted accurately for that time period.

Such tests are particularly useful because they assess predicted extinctions and colonizations, which are the fundamental mechanisms of metapopulation models. It would also be useful to test some of the relative predictions. For example, are patches with a low predicted probability of extinction less likely to become extinct than patches with a higher predicted probability of extinction? Such a test requires that individual predicted probabilities be compared with the incidence of events. Logistic regression may be used to perform this sort of analysis because it is designed for cases in which the prediction is a probability (e.g., probability of patch extinction) and the observation is a binary variable (e.g., extinction occurs or does not). To test whether there is a positive relationship between the predicted probability of extinction and actual extinction, a logistic regression equation of the following form could be used:

\[
\text{logit}(y) = \ln\left(\frac{y}{1-y}\right) = a + bx,
\]

where \(y\) is the fitted probability of extinction observed in the field, \(x\) is the probability of extinction predicted by the population model, and \(a\) and \(b\) are the regression coefficients. The regression analysis indicates whether the predictions (\(x\)) and the observations (\(y\)) are positively related. This is simply a test of whether the coefficient \(b\) is greater than 0, based on standard significance testing for logistic regression (Hosmer & Lemeshow 1989). In cases where extinctions are rare, tests of the predicted risk of quasiextinction (decline to a defined small population size; Ginzburg et al. 1982) are more appropriate. Lindenmayer et al. (1999) give an example of this method, in which they tested predicted probabilities of patch occupancy using data on the presence or absence of four species of arboreal marsupials in a system of eucalypt patches.

The above test indicates only whether the observations and predictions are positively related, not whether the predictions differ from the observations. Cox (1958) provides a modification of this approach by using a logistic regression of the form

\[
\text{logit}(y) = A + B\text{logit}(x).
\]

An imperfect relationship between predictions and observations is indicated if the regression coefficient \(A\) is different from 0 or if coefficient \(B\) is different from 1 (Cox 1958; Miller et al. 1991). The results indicate whether the probabilities are being under- or overpredicted across the range of the data (Cox 1958; Miller et al. 1991).

McCarthy et al. (2000) used the logistic regression method to test patch occupancy predicted by metapopulation models of two species of Australian treecreepers (Aves: *Climacteridae*: *Climacteris erythrops* and *Cormobates leucophaea*). The tests illustrated that both models underestimated occupancy by these species in a system of remnant eucalypt patches embedded within an exotic pine plantation. Modifying the models, using results of the tests and additional information on the biology of the species, improved their fit. These modified models were subsequently tested by use of observed extinctions and colonizations of patches over a 1-year period, although this second comparison was limited because of the presence of unknown sampling error.

Testing Risks of Population Decline

Perhaps the main prediction of PVA models is the risk of population decline (quasiextinction risk), including the risk of extinction. A test of this nature requires that numerous populations be assessed to obtain sufficient replication. Brook et al. (2000) used 21 different populations for assessing the predicted probability of population decline. Although only one of the populations became extinct, risks of population decline were assessed by comparing the proportion of populations that actually declined below the population sizes corresponding to predicted quasiextinction risks of 0.0, 0.1, 0.2, ..., 1.0. The predicted frequency of declines was not significantly different from the observed frequency of declines when tested by a \(G\) test and aggregated over the 21 cases (Brook et al. 2000). The differences between the predicted risks and the actual proportion of populations declining were less than approximately 0.2 when averaged over the 21 species.

Testing Probability Distributions of Predicted Population Size

Calculation of the predicted probability distribution of population size and visual comparison with actual data provides one method of determining whether observations are different from predictions. For each patch, or for each time step of a model, observations can be compared with the predicted distributions to gain a qualitative impression of the performance of the model. In such a comparison, it is important to be mindful that some extreme observations are expected by chance alone when many different predictions are being examined.

A quantitative comparison of predicted distributions and observations is hampered because we rarely have replicated observations for each distribution. Where multiple predictions are available, they are usually made at different times and in different places, so the predictions will rarely
be identical. Such circumstances mean that comparisons between predictions and observations need to be standardized to allow replication, which requires a transformation such that all the observations have the same statistical properties. A common method of rescaling variables is to use standard deviates, which measure each variable in terms of number of standard deviations away from the mean (Sokal & Rohlf 1981). Similarly, each observation can be rescaled as a standard deviate by subtracting the observed value from the predicted mean for that observation and dividing by the predicted standard deviation (Reynolds et al. 1981; Gilpin & Diamond 1982). The predicted means and standard deviations will typically be different for each observation. If the mean and standard deviation are predicted correctly, the resulting standard deviate will have a mean of zero and a variance of one (Sokal & Rohlf 1981). Each observation may be transformed in this way to obtain a set of standard deviates.

This set of data may be compared to the expected mean (0) and variance (1) by standard statistical methods (Sokal & Rohlf 1981) or asymptotic approximations (Reynolds et al. 1981). Nonparametric statistics may be required when the predicted distributions are strongly skewed, for example, if extinctions are predicted to be common. Overprediction by the model is indicated if the mean is >0; underprediction is indicated if the mean is <0. When the variance of the standard deviates is <1, this suggests that the predicted variation in population size is too large; the predicted stochasticity tends to be too small when the variance is >1.

McCarthy and Broome (2000) used standard deviates to test a model of the population dynamics of *Burramys parvus*, a terrestrial possum confined to alpine areas of Australia. The parameters of the model were estimated with data from one population and tested with data from three other populations. Models that assumed the habitat quality (measured by the equilibrium population size per unit area) was equal in all populations made predictions that were significantly different from observations for two of the three test populations. The analyses demonstrate that it was necessary to account for differences in habitat quality between populations to adequately simulate the population dynamics of the species. Use of standard deviates is appropriate for testing annual changes in abundance in populations or time periods that are different from those used in developing the model.

Brook et al. (2000) used the standard-deviates test of the final mean predicted population size of 21 different PVA models and found that the mean predictions were not significantly different from the observations when aggregated over the 21 models. Although they used the standard-deviates test of the mean, they did not test the predicted variance in population size and did not report the results of tests for individual models.

Use of standard deviates provides a test of the mean and variance of the population size, but it does not test the shape of the distribution. One way to test this feature is to represent each observation by the predicted cumulative probability—the predicted probability of the population size being less than or equal to the observation. In this way, each observation is transformed to a cumulative probability on a scale between zero and one. If the predicted distributions are accurate, these cumulative probabilities will be distributed uniformly between zero and one (Fisher 1954). A range of statistical methods can be used to examine whether a set of variables is uniformly distributed, such as the Kolmogorov-Smirnov test (Sokal & Rohlf 1981; Daniels 1990) or Fisher’s (1954) test of uniform random numbers. If a total of *n* variables (\(p_i\)) are distributed uniformly between zero and one, the quantity

\[
L = -2 \sum_{i=1}^{n} \ln p_i
\]

has a chi-square distribution with \(2n\) degrees of freedom (Fisher 1954). Use of cumulative distributions to test predicted annual changes in abundance is appropriate in the same circumstances as those required for the standard-deviates test.

**Discussion**

Testing procedures are typically used to assess quantitative predictions, but the quality of models can also be assessed according to their usefulness for management, the ecological insights they provide, and/or their simplicity, generality, robustness (insensitivity to assumptions), and efficiency (Haefner 1996). If models are used for explanatory purposes, their predictions may be irrelevant. Simple models that summarize what is thought to be true may be useful, even if their predictions are known to be uncertain (Burgman et al. 1993). Such models can help clarify ideas and highlight the most important components of population dynamics. When used to help direct further data collection and analysis, models may improve understanding and be useful to management (Possingham et al. 1993). Such benefits may occur irrespective of the accuracy of quantitative predictions. But many PVA models are used to assess the habitat requirements of species or to rank management strategies. In these cases, at least the relative accuracy of predictions, if not their absolute accuracy, is always important.

Although we recognize that the quality of models should not be assessed only by their predictions, methods for testing the predictions of PVA models are required. Such testing can indicate the greatest deficiencies in a model, which can be corrected by further data collection and model development. In this way, testing can contribute to a better understanding of the species being modeled and assist the effective use of the model.
in the management of threatened species (Posingh et al. 1993). Model testing is most useful as part of an integrated process of model development and selection, experimental design, data collection, and parameter estimation. In this sense PVA becomes part of a program of active adaptive management (Walters & Holling 1990).

Our review identifies two different approaches to testing PVA models: (1) qualitative comparisons based on visual comparisons of predictions and observations, and (2) quantitative comparisons based on statistical analyses. All the methods have advantages and disadvantages and are suitable for testing different aspects of a model with different types of data (Table 1). Visual comparisons may indicate trends obscured by quantitative comparisons, but assessing the importance of differences remains subjective. The use of statistical tests provides a quantitative basis but may create an impression of a false dichotomy where a model is considered to be “true” or “false.” It is important that models are never regarded as being true, even after their predictions have been subjected to several tests. Additional data or a new test can always identify a model error. We suggest using a range of suitable methods to test PVA models, depending on the available data, although a test that ignores stochastic aspects of PVA models is meaningless. Stochastic population models are rarely if ever built to make mean predictions; testing only these predictions is of limited value when the primary prediction is the risk of population decline.

When PVA models are tested with significance tests, the probability of obtaining a significant result will depend, in part, on the sampling effort. The more data collected in the field, the greater the chance of identifying a significant difference between predictions and observations. In contrast to hypothesis testing, a significant result should not lead to the immediate rejection of the model. Instead, the model should be modified to determine why the predictions differ from reality. For example, if population density within patches is underpredicted, the dispersal rate or population growth rate could be increased to determine whether a better match is obtainable. If variation in population size is predicted incorrectly, the level of stochasticity or the relative contribution of different forms of stochasticity (e.g., demographic vs. environmental) could be modified. Such modifications rely on ecological knowledge and intuition, an understanding of the sensitivity of the model to changes in parameters, and consideration of the relative uncertainty associated with different parameters.

A range of parameter estimates should be examined to explore the parameter space and the resulting fit of the model. In some cases it may be impossible to obtain predictions that match observations. For example, population densities observed in small patches may be higher than in contiguous areas. Such a result cannot be replicated (even qualitatively) with a simple population model of a single species, indicating that basic structural changes are required. In this case, including a positive edge effect or interspecific interactions may improve predictions of the model. Such modifications should be based on ecological understanding, and the modified models should be retested against new data in an iterative process of model development.

Bayesian methods are conducive to an iterative process of model development and testing, with the posterior probabilities becoming the prior probabilities for the next set of tests. This approach is well suited to examining the plausibility of different models or different parameter values given particular sets of data (Hilborn & Mangel 1997). Bayesian statistics provide a formal framework for comparing models and determining the relative level of confidence in their predictions. This comparison provides a formal basis for rejecting a particular model and adopting an alternative: adopt the alternative when it makes demonstrably better predictions than the original model (Hilborn & Mangel 1997; Burnham & Anderson 1998). Bayesian methods, however, compare a finite set of user-defined models, and the best of these may still provide only poor predictions. Thus, direct comparisons of predictions and observations and identification of significant biases and errors remain useful for evaluation of models and for determining their limits of accuracy. In addition, many of the approaches to model testing outlined in this review could be applied in a Bayesian framework. A complete review of Bayesian methods is beyond the scope of our paper and is covered elsewhere (e.g., Ludwig 1996; Hilborn & Mangel 1997; Wade 2000).

An important issue in testing PVA models is the independence of the data between years and across space. In general, the statistical procedures discussed here rely on the assumption that predictions of a model are independent of one another. For example, if time series data are used for comparisons, it is necessary that the standard deviates be independent to ensure a valid statistical analysis. This can be ensured, in the absence of delayed density dependence, if the predictions in each year are based only on the abundance in the previous year. If the standard deviates are not independent, the observed variance in the standard deviates will tend to be greater than expected. Similarly, predicted abundances in adjacent patches may be interdependent if dispersal between patches is important or if a factor such as weather has a regional influence on a species. When the observed data are dependent, the effective degrees of freedom of the statistical tests will be less than the number of observations. Ignoring these dependencies tends to make the predictions appear more different from the observations than they really are. The possibility of observations being correlated should be considered when models are tested, but it may be difficult to account for the problem if it exists. Correlated data could be pooled,
but this is not entirely satisfactory because some information for evaluating the model will be lost, and the appropriate level of pooling would require a subjective judgment.

The aim of testing PVA models is to determine how accurately the stochastic processes are modeled. Invariably when field data are collected, there will also be sampling error. When the occupancy of patches by species is determined, it is possible that false absences will be recorded even with intensive sampling effort. False absences are likely to be the most important error in occupancy data, although a kind of false presence may be recorded if an individual is detected while dispersing through an area rather than remaining resident in it. These errors should be quantified so that the predicted probabilities of occupancy can be modified appropriately (Lindenmayer et al. 1999). Errors in estimating abundance are also likely to occur and if present will make the population size appear to fluctuate more than it does in reality. Again, if observer error can be quantified, it can be accommodated by including an additional error in the model predictions.

Several of the methods test the absolute predictive ability of PVA models, but the value of many models lies in their relative predictions. The use of logistic regression to determine whether there is a significant positive relationship between the predicted and actual occupancy of patches provides an example of using the methods to determine the relative predictive ability of the model. The test is whether the model can distinguish between patches that are more or less likely to be occupied. The relative probability of a patch being occupied is of conservation interest because, among other reasons, this probability is often correlated with the importance of patches for metapopulation persistence. Such information can assist in the design of conservation reserves (Lindenmayer & Possingham 1994), and the test of the model is a test of the ability to make the correct decision. If different management strategies were employed in different areas, it would be possible to test the relative predictions under the different management strategies. The test would indicate the probability of making the correct decision. To obtain meaningful results, such large-scale experiments would need to be designed with appropriate stratification and replication (Walters & Holling 1990). If only a single management strategy is employed, the methods cannot be used directly to assess the relative predicted merits of different management strategies. This can be achieved only by relatively weak inference if the predictions of the model are not shown to be incorrect.

Different types of data allow different types of tests. The methods we have discussed require the types of data that we believe are typically collected for most species for which PVAs are conducted. An effort to test the predictions of PVA models encourages the systematic collection of new data on the distribution and status of modeled species. It has been several years since many PVA models were developed, and it is likely that some predictions of these models can now be tested with field data to determine their accuracy and identify ways in which they can be improved. Although large data sets may not be available, some testing is better than none.

**Acknowledgments**

This work was conducted in part while M.A.M. was a postdoctoral associate with the Extinction Risk Working Group at the National Center for Ecological Analysis Synthesis, a center funded by the National Science Foundation (grant DEB–94–21535), the University of California at Santa Barbara, the California Resources Agency, and the California Environmental Protection Agency. We are grateful to D. Lindenmayer and I. Broome, who collaborated with us on testing of population viability analysis; to S. Beissinger, R. Akcakaya and various anonymous referees for critical comments, and to B. Brook and M. Burgman for interesting discussions.

**Literature Cited**


