Viability analysis for multiple populations

Seth J. Wenger[a,*], Douglas R. Leasure[a], Daniel C. Dauwalter[b], Mary M. Peacock[c], Jason B. Dunham[d], Nathan D. Chelgren[d], Helen M. Neville[b]

[a] University of Georgia, 203 D.W. Brooks Drive, Athens, GA 30602, United States
[b] Trout Unlimited, 910 West Main Street #342, Boise, ID 83702, United States
[c] University of Nevada, Reno. Department of Biology, Reno, NV 89557, United States
[d] US Geological Survey, 3200 SW Jefferson Way, Corvallis, OR 97331, United States

ARTICLE INFO

ABSTRACT

Many species of conservation interest exist solely or largely in isolated populations. Ideally, prioritization of management actions among such populations would be guided by quantitative estimates of extinction risk, but conventional methods of demographic population viability analysis (PVA) model each population separately and require temporally extensive datasets that are rarely available in practice. We introduce a general class of statistical PVA that can be applied to many populations at once, which we term multiple population viability analysis or MPVA. The approach combines models of abundance at multiple spatial locations with temporal models of population dynamics, effectively borrowing information from more data-rich populations to inform inferences for data-poor populations. Covariates are used to explain population variability in space and time. Using Bayesian analysis, we illustrate the method with a dataset of Lahontan cutthroat trout (*Oncorhynchus clarkii henshawi*) observations that previously had been analyzed with conventional PVA. We find that MPVA predictions are similar in bias and higher in precision than predictions from simple PVA models that treat each population individually; moreover, the use of covariates in MPVA allows for predictions in minimally-sampled and unsampled populations. The basic MPVA model can be extended in multiple ways, such as by linking to a sampling and observation model to provide a full accounting of uncertainty. We conclude that the approach has great potential to expand the use of PVA for species that exist in multiple, isolated populations.

1. Introduction

Perhaps the single most interesting and important question in conservation biology is: "why do some small populations decline to extinction while others persist?" If this question could be reliably addressed for species of interest, managers would have the information necessary to strategically focus actions on populations most at risk. Conservation theory and empirical observations have produced rules of thumb on the minimum number of individuals (Frankham et al., 2014; Franklin, 1980; Traill et al., 2007) or minimum habitat extent (e.g., Hilderbrand and Kershner, 2000; Robbins et al., 1989) required for a viable population, but there are numerous examples of populations that defy these rules (Peterson et al., 2014; Shoemaker et al., 2013). Smallness alone is an insufficient predictor of risk; it is also critical to understand the factors that correlate with population declines and stochasticity, and thus extinction (Caughley, 1994).

A holistic understanding of the causes of population declines, and ultimately the processes contributing to extinction of small populations, is particularly important for species in highly fragmented habitats (Fagan and Holmes, 2006; Gilpin and Soule, 1986) — a category that covers many imperiled species of conservation interest. Examples include the island fox (*Urocyon littoralis*) in the Channel Islands (Kohlmann et al., 2005), "mountain island" species such as pika (*Ochotona* spp.; Beever et al., 2003), and water-associated species in arid landscapes (Kodric-Brown and Brown, 1993). For such species, the relative viability of individual populations is of fundamental concern to managers who must make hard decisions regarding the allocation of limited resources to prevent population and species extinctions. Ideally, such decisions would be guided by data-driven estimates of extinction probabilities under alternative scenarios of management action (or inaction) and varying environmental conditions, such as future climates.

Population viability analysis (PVA) is a class of analytical approaches that yields probabilistic estimates of population viability (or extinction) over specified time horizons (Beissinger and McCullough,

2002; Morris and Doak, 2002). However, traditional PVAs require temporally extensive demographic data (e.g. survival, reproduction, and maturation rates), and have generally been limited to populations that have been the focus of long-term monitoring programs. Such datasets are unusual for imperiled species, and it is particularly rare to have long-term data for every population of interest, which is required to evaluate relative risk and to target management actions effectively. Some traditional PVA approaches also require data from marked and tracked individuals; such data are generally expensive to collect.

One alternative is to use a species distribution modeling approach (Elith and Leathwick, 2009) or abundance modeling approach (e.g., Royle, 2004) to identify spatial or spatiotemporal environmental covariates to explain presence or abundance. These covariates allow projection of occurrence or abundance across broad geographies and under alternative management and climate conditions. However, they generally do not consider temporal population dynamics or density dependence, and are often made at the scale of the sampling unit rather than the full population. Thus, as typically employed they cannot provide estimates of population viability.

2. A multiple population viability model

We propose that statistical methods of modeling presences or abundances in space can be married with statistical time-series models of populations to provide a data-driven approach to population viability analysis that can be applied to many populations simultaneously. The method has four defining characteristics:

1. Some population parameters are shared among populations. One or more of these parameters are influenced by covariates that vary in space and in time (or both) so they can be used to describe spatial differences among populations as well as temporal fluctuations within populations.
2. Populations are modeled as autoregressive, meaning that abundance at one point in time depends in part on the abundance in a previous time step, as in a traditional PVA. A mechanism to account for density dependence should be included in most cases.
3. Modeling is at the scale of the full population, not the sampling unit. Counts must either be scaled to the full population before modeling, or (better yet) scaled to the full population during modeling via an observation model and a sampling model directly linked to a process model.
4. As a statistical method it is driven by empirical data, which sets it apart from simulation-based methods that usually rely on a mix of literature values and expert opinion.

Previously introduced methodologies have incorporated some of these characteristics, but we know of no published approaches with all four. Clark and Bjørnstad (2004) introduced Bayesian methods to fit flexible state-space time series models that could account for hidden states, missing values, observation errors and other complexities, but applied them to only one population at a time. Zipkin et al. (2014) used a Bayesian modeling framework to extend the open N-mixture model (Dail and Madsen, 2011) to account for stage-structured time series population data. Kanno et al. (2015) further developed this to model abundance at multiple locations with density dependence as a function of climate covariates, but analysis was at the scale of the sampling unit rather than full populations, and the focus was not on viability.

We build on these antecedents to introduce a simple Bayesian multiple population viability analysis process model, or MPVA. While previous studies (Berliner, 1996; Clark and Bjørnstad, 2004; Staples et al., 2004; Zipkin et al., 2014) have emphasized the importance of linking population models to observation models to remove bias associated with incomplete detection, for the sake of simple exposition we focus here on the process model. We discuss methods for linking to observation and sampling models later.

To develop the model we assume a dataset collected from multiple isolated populations p, at least some of which have been sampled at sequential time steps t. Because our focus is on the process model, we assume that the population estimate for each population at each sampling occasion N_{pt} is known without error. We assume at least one covariate X_{pt} with a value corresponding to each sampling occasion. A traditional way to analyze such a dataset might be to use Poisson regression in a generalized linear modeling framework:

$$N_{pt} \sim Poisson(\overline{N}_{pt}) \tag{1}$$

$$log(\overline{N}_{pt}) = \beta_0 + \beta_1 X_{pt} \tag{2}$$

Such a model would probably not meet assumptions, as repeat samples of individual populations would not be independent. This could be corrected by including a random intercept for population identity, making it a multilevel model (Gelman and Hill, 2007; Raudenbush and Bryk, 2002). Alternatively, one could make the model explicitly autoregressive by making the population at the current time step dependent on the previous time step:

$$log(\overline{N}_{pt}) = log(N_{pt-1}) + \beta_0 + \beta_1 X_{pt} \tag{3}$$

The above applies to $N_{pt-1} > 0$, as the outcome $N_{pt} = 0$ is deterministic when $N_{pt-1} = 0$; this caveat also applies to all subsequent forms of this equation. Eq. (3) is simply an exponential growth model where the intrinsic population growth rate r is a linear function of covariate X. It can be rewritten as:

$$log(\overline{N}_{pt}) = log(N_{pt-1}) + r_{pt} \tag{4}$$

$$r_{pt} = \beta_0 + \beta_1 X_{pt} \tag{5}$$

While valid, the model lacks a term for density dependence, which means that populations are unbounded and could grow exponentially to infinity in projections.

One can solve this with a simple density dependence term:

$$log(\overline{N}_{pt}) = log(N_{pt-1}) + r_{pt}\left(1 - \frac{N_{pt-1}}{K_p}\right) \tag{6}$$

This is now a form of the Ricker model (Ricker, 1954) in which the realized population growth rate approaches the intrinsic growth rate r_{pt} when populations are small, but declines as populations approach carrying capacity K_p. We chose the Ricker model because it is simple, it has been widely used in both theoretical and applied ecology (May, 1974; Clark, 2007; Morris and Doak, 2002; Dail and Madsen, 2011; Kanno et al., 2015), and it has a linear form amenable to covariates on rates of recruitment and density-dependent mortality (Hobbs and Hooten, 2015). However, many other formulations are possible; the Gompertz model in particular has been shown to have desirable properties (Dennis and Taper, 1994). In our example K_p is indexed by population, indicating that each population has a unique carrying capacity that is constant through time. Carrying capacity could be allowed to vary temporally as a function of covariates, but we have found that when both r and K are allowed to vary spatio-temporally, the model can suffer identifiability issues (i.e. there are multiple optima).

A solution is to reformulate the Ricker equation using the term *phi* (φ) to represent r/K, the strength of density dependence (Hobbs and Hooten, 2015). One can include spatio-temporal covariates on both r and φ:

$$log(\overline{N}_{pt}) = log(N_{pt-1}) + r_{pt} - \varphi_{pt} N_{pt-1} \tag{7}$$

$$r_{pt} = \beta_0 + \beta_1 X_{pt} \tag{8}$$

$$\varphi_{pt} = \gamma_0 + \gamma_1 X_{pt} \tag{9}$$

In this formulation, φ represents the reduction in population growth rate associated with adding a single individual to last year's population, and will be a small positive number—generally much < 1 but > 0. The