

## Abstract

Models of population dynamics play a central role in theoretical and applied ecology where they are used for purposes such as testing hypotheses about density dependence and predicting species' responses to future environmental change or management actions. State-space models, are widely used for these purposes, and they allow for estimating parameters of classical growth models, such as exponential, logistic, and Gompertz, while accounting for demographic stochasticity and observation error. Conventional state-space models, however, have three important limitations: (1) the parameters are not identifiable in many common situations, (2) they do not admit spatial variation in population dynamics, and (3) there is no clear interpretation of the observation error. We demonstrate how each of these problems can be resolved using a class of hierarchical models for spatially-replicated time-series data recently proposed by Dail and Madsen (2011). We expand this class of models to accommodate classical growth models, zero-inflation, and random effects such as observer-specific detection probabilities. Furthermore, we describe classical and Bayesian methods of parameter estimation. These new developments will allow researchers to apply these methods to address important questions regarding the factors affecting spatial and temporal variation in abundance. We demonstrate these developments by analyzing data from the North American Breeding Bird Survey. Code to fit these models using the R package unmarked and JAGS is also provided.

**Key words:** abundance, Dail and Madsen model, density-dependence, Gompertz model, immigration, open population point count models, random observer effects, range, Ricker model, zero-inflated

Theoretical ecology requires robust models for testing hypotheses regarding density dependence and spatial variation in population dynamics. Such models are also required in applied ecological research for predicting species' responses to environmental change or management actions. For example, many researchers are currently investigating the influence of weather on population dynamics so that predictions can be made about the effects of future climate change. Several complicating factors must be confronted when addressing these questions. First, deterministic models of population dynamics are virtually always inadequate due to "process variation", the inherent stochasticity in demographic parameters and environmental conditions. Second, abundance—the natural state variable in studies of population dynamics—can rarely be observed perfectly in field studies because of "observation error", such as imperfect detection. Failure to account for and heterogeneity in observation error can bias the estimators of abundance and related parameters.

State-space models have become the most widespread approach for studying population dynamics while accounting for process variation and observation error (de Valpine and Hastings 2002, Buckland et al. 2004, Dennis et al. 2006). Conventional state-space models are simply time-series models in which the true state of the system is observed imperfectly. Originally developed for xxx (cite Kalman filter papers), the ability to model both the ecological state process and the observation process has made them relevant in numerous ecological applications [need to clean this up]. State-space models have been defined differently by different authors, but we follow Buckland et al. (2004) using the definition : "xxx" . Typically, the ecological process is described by a Markovian model that includes a deterministic description of population change as well as two random sources of variation for "process variation". The model for the observation process is typically phenomenological in the sense that observation error is not defined explicitly. As an example, let  $x_t$  be the abundance of a species during year  $t$ , and  $y_t$  be the observed data, which differs from  $x_t$  due to observation error  $\epsilon_t$  ( ). A simple state space model that includes exponential growth and process variation ( ) could be written as (1) Here is the intrinsic rate of increase. Typically, both random effects are assumed to follow normal distributions:  $\epsilon_t \sim N(0, \sigma^2)$  and  $\eta_t \sim N(0, \sigma^2)$  (de Valpine and Hastings 2002, Dennis et al. 2006). Even though this is the most widespread approach for modeling population dynamics using time-series data, several problems are immediately evident. First, if abundance is a non-negative integer, as it always is, there is clearly a problem with this formulation because the normally distributed random effects violate this constraint. To avoid this issue, ecologists often replace  $x_t$  with  $\log(x_t + 1)$  where  $A$  is the area surveyed—i.e., abundance is replaced by the natural log of population density, which is assumed to be normally distributed. This is desirable because this allows for the fitting of models assuming normally distributed residuals, which are computationally much less intensive than most alternatives. However, this transformation is problematic since abundance may be zero and thus  $\log(0)$  is undefined. Generally, zeros are replaced with some small, but positive number, although the effect of this is rarely discussed.

The problems mentioned thus far are minor in comparison to the more serious issue that the parameters of the model are often not identifiable. Specifically, the Markovian nature of the model implies the following likelihood

### LIKELIHOOD

The first term is the probability density for the first observation in the time series,  $y_1$ , yet it is intuitively not possible to estimate the parameter(s) of this distribution using only a single observation. This problem has resulted in several papers with interesting names like "multi-modal likelihoods and ridges etc..." The workaround here is to assume that the population is at equilibrium so that  $x_1$  can be replaced with the equilibrium distribution. However, assuming equilibrium defeats the objective of many studies of population dynamics, namely determining if and why a population is at equilibrium.

A third problem with these models is that they do not admit spatial variation. This is not a fair criticism because this lies beyond the scope of traditional state-space models, but it does restrict the utility of their since it is typically impossible assess the impacts of factors such as habitat fragmentation or climate change on a population without considering spatial variation. Indeed many populations are regulated by spatial processes such as source-sink dynamics.

Dail and Madsen (2011) developed a model (henceforth the DM model) that resolves each of these problems with traditional state-space models. The DM model allows for inference about spatial and temporal dynamics in abundance while accounting for observation error using only spatially and temporally replicated count data. Their model is an extension of the closed-population N-mixture model (Royle 2004)(2004), which was designed specifically to address the problem of modeling spatial variation in abundance when detection probability is less than unity. The DM extension relaxes the assumption of population closure and includes explicit parameters describing population change as a first-order Markovian process. [more details]

[applications / potential etc... ] The DM model is relevant to a huge number of problems confronting basic and applied ecological research. Answering questions about climate change etc... require spatially and temporally extensive datasets such as result from the North American Breeding Bird Survey (BBS) (Robbins et al. 1986). However, most of the datasets of the appropriate scale are collected by volunteers and using protocols that are not amenable to traditional approaches of modeling abundance and detection probability based on standard capture-recapture methods. Instead they have, at best,

simple count data for which, when it can be interpreted as counts of unique individuals, the DM model can be applied.. Yadi ya

Although the DM model was designed explicitly for these purposes, the original formulation of the model makes strict distributional assumptions and does not acknowledge many of the sources of variation inherent to existing ecological time-series data. We propose variations on these models that are likely to be more realistic and useful in many cases. We demonstrate the usefulness of these new models with simulated and real BBS data.

## References

Dail, D. and L. Madsen, 2011. Models for estimating abundance from repeated counts of an open metapopulation. *Biometrics* **67**:577–87.