

Ecological responses to observation error.

Andrew J. Tyre, Jonathan Rhodes, Tara Martin and Elizabeth VanWormer

Abstract

Introduction

Observations of reality are imprecise and fraught with error from many sources. Arguably, the science of statistics was invented to deal with this fact. In many disciplines, the error in observations can be well described by a Gaussian, or normal, distribution. Statistics is well equipped to deal with those cases! In a few disciplines, errors take on different structures that are not normal (pun intended). Sometimes simply changing the distribution we assume the observations are following is sufficient to regain a powerful suite of tools to draw conclusions. But in a few cases, even that is not sufficient.

Imagine trying to determine the prevalence of a disease in a population out in nature. The most basic observation, is this individual infected (or has been infected in the past), is a simple binomial variable. Even that observation has errors – there are few (maybe no?) methods for detecting disease in a tissue sample of an individual that have no false positive or negatives. But even if our method is perfect, going from a infected proportion of a sample of individuals back to the prevalence in the population is fraught with difficulty. We must assume that the individuals captured are a random sample of the population; that is, infected and uninfected individuals are both randomly distributed and equally likely to be captured. If we are taking samples from across a series of habitat patches in a fragmented landscape, we may also find that no individuals were captured at a particular patch at all. Does that mean the species is absent, and if not, what is the prevalence in that patch? The answers to that final question depend a great deal on the methods of sampling and how sampling effort is distributed in space.

A fundamental question in ecology broadly has to do with how species are distributed in space, and which biotic and abiotic characteristics of the environment explain that distribution. The sub-discipline of “species distribution modelling” is the art of answering this question. Historically, the details of how an observation of a species at a particular point in time and space arose were largely ignored (cite some early paper here). Species distribution models were built assuming that the presence and absence of a species at some series of points was known with certainty (or cite some examples here). Although this assumption was widely known to be false, there did not seem to be any methods both testing this assumption and mitigating its effects.

The primary issue arising from assuming observations are accurate when they are not is bias (Tyre et al. 2003). If a species is sometimes missed when present, a false negative, estimates of the effects of covariates on the probability of presence are biased towards zero. **is there a paper looking at effects of false positives in a similar broad way?**

Tyre, Andrew J, Brigitte Tenhumberg, Scott A Field, Darren Nijalke, Kirsten Parris, and Hugh P Possingham. 2003. “Improving Precision and Reducing Bias in Biological Surveys: Estimating False-Negative Error Rates.” *Ecological Applications* 13 (6). Ecological Society of America: 1790–1801.