Sampling Designs and Statistical Models for Estimating the Occurrence, Spread, and Imperfect Detection of an Invasive Species

Robert M. Dorazio, USGS, Florida Integrated Science Center, Gainesville, Florida

J. Andrew Royle, USFWS, Division of Migratory Bird Management, Laurel, Maryland

Christopher K. Wikle, University of Missouri, Dept. of Statistics, Columbia, Missouri

Background and Need

The National Park Service (NPS) requires sampling methods and protocols for monitoring both the occurrence of newly established invasive species and the spread of individuals that have successfully colonized one or more areas of concern. It is natural then to consider spatial units of sampling and to base the sampling frame and the methods of sample selection (i.e., the design) on these units.

A variety of designs exist for selecting spatial sample units (e.g., see Thompson, 1992). Different designs employ different sampling strategies, such as stratification or clustering of sample units, and these strategies are often useful in partitioning expected sources of variation in the quantities being surveyed. However, estimators of population-level quantities (e.g., the population average) that are unbiased under a design (that is, unbiased with respect to all possible samples that may have been selected using the design) are unlikely to be useful in surveys of invasive species.

There are 2 reasons for this lack of utility. First, predictions of occurrence of invasive species are often required in regions that are smaller than the entire area being surveyed. The predicted spread of invasives along an advancing front, such as the westward spread of gypsy moths in North America (Sharov, Leonard, Liebhold, Roberts, and Dickerson, 2002), is an example. In these problems predicting the occurrence of invasives in individual sample units requires a model of the observations that includes one or more sources of spatial dependence

among observations. A second shortcoming of the design-unbiased estimators is that they apply only to quantities that are directly observable. In surveys of animal populations, one is rarely able to observe directly the abundance or density of animals in a typical sample unit because methods of capture or detection are usually imperfect. Instead, ecologists are often forced to combine clever sampling protocols, such as mark-recapture, removal, or distance sampling, with statistical models so that the observed counts of animals are properly related to scientifically relevant estimands, such as animal abundance or occurrence. Such models obviously have an important role to play in the prediction of invasive species.

Formal attempts to predict the dynamics of invading species have included sophisticated models of spread. Examples include reaction-diffusion models (Skellam, 1951; Okubo, 1980; Wikle, 2002), which specify the local dispersal of individuals as a continuous random walk but ignore discrete, long-distance dispersals, and integro-difference models (Van den Bosch, Metz, and Diekman, 1990; Van den Bosch, Hengeveld, and Metz, 1992), which incorporate the dispersal of individuals over larger spatial scales. While these dynamical models offer useful paradigms for predicting the spread of invasive species, the models are typically fitted to the spatially-indexed, observed counts of animals without considering the imperfect detectability or capture of animals. In other words, the observed counts have been used as surrogates for the actual abundances of animals. This practice may be appropriate in cases where the detection of animals remains constant over all sampling times and locations, but this condition is seldom satisfied in surveys of natural animal populations, as noted earlier.

Obviously, new types of sampling designs and models are needed to predict the occurrence and spread of invasive species while properly accounting for their imperfect detection.

Objectives

1. To develop sampling designs and statistical models for predicting the occurrence and spread of invasive species while properly accounting for the imperfect detection of animals encountered while sampling.

2. To illustrate this approach by analyzing the numbers of individuals observed in annual surveys of an invasive species. (Identification of the particular species depends on data availability (see Methods).)

Methods

Recently, a variety of statistical models have been developed for improving estimates of animal abundance by applying a particular survey protocol at several distinct locations that are selected to be representative of a larger region of interest. For example, models of spatially replicated removal samples, (Dorazio, Jelks, and Jordan, in review), distance

samples (Royle, Dawson, and Bates, in review), and even simple point counts (Royle and Nichols, 2003; Royle, 2004) have been developed to estimate either the local abundance of animals at a single sample location or the average abundance among sampling locations. These models provide a reasonable starting point for modeling spatial and temporal changes in abundance (and occurrence) of invading species. For example, we plan to extend the models to include both systematic and stochastic forms of spatial dependence that would provide predictions of abundance at both sampled and unsampled locations. In other words, a "map" of the spatial distribution of abundance will be estimated using an observed set of spatial covariates and the distances between locations on the map. Of course, the ability to compute such estimates (and their accuracy and precision) is expected to improve as the number of locations in the sample increases; therefore, careful attention to sampling design is crucial.

Dynamical models for estimating and predicting colonizations of invasive species require spatially replicated samples at two or more times (e.g., annually). The specific functional form of the spatio-temporal process in these models depends, in part, on the nature of the dispersal. For example, do dispersals occur over short or long distances? Are there specific agents or vectors of dispersal (e.g., wind, flowing water, or highly mobile vertebrates) that can be monitored and used in prediction? Do features of the landscape or habitat serve as invasion corridors? All of these questions are relevant in the specification of dynamical models of invasive species spread, even if answers to these questions are incomplete. Therefore, we anticipate that our sampling design and model of invasive species spread will necessarily be tailored to the biology of the species under investigation.

Our approach is motivated by our experience in the design and analysis of surveys of animal populations. However species of plants, particularly invasives, also may be difficult to detect and therefore can benefit from our approach. We believe that the separation of detection and abundance of individuals in a population is fundamental to the problem's solution and, to our knowledge, has never been considered in models of invasive species spread.

Expected Products

- Sampling designs and models for predicting the occurrence and spread of invasive species.
- At least one publication in peer-reviewed journal.

Technology/Information Transfer

- Presentation of sampling protocol, sampling design, and other forms of monitoring required to satisfy the modeling assumptions.
- Presentation of results in scientific conferences.

Work Schedule

Date	Product
Aug 2004	Acquisition of suitable database
Dec 2004	Development of sampling design, model, and associated software
Dec 2004	Analysis of data (simulated and actual) using the model
Apr 2005	Presentation of results (oral and manuscript)

Budget

In-kind contribution	Dorazio	Royle
Salary	27,900	14,000
Administrative expenses	10,000	5,000
Total	37,900	19,000

Requested	Dorazio	Royle	Wikle
Salary	0	0	37,000
Travel	2,000	2,000	2,000
Supplies and materials	1,000	1,000	1,000
Administrative expenses	1,000	1,000	14,300
Total	4,000	4,000	54,300

Total Request = 62,300

References

Dorazio, R. M., Jelks, H. J., and Jordan, F. (in review), "Improving removal-based estimates of abundance by sampling a population of spatially distinct subpopulations," *Biometrics*.

Okubo, A. (1980), *Diffusion and ecological problems: mathematical models*, New York: Springer-Verlag.

Royle, J. A. (2004), "N-mixture models for estimating population size from spatially replicated counts," *Biometrics*, 60, in press.

Royle, J. A., Dawson, D. K., and Bates, S. (in review), "Estimating abundance effects in distance sampling models," *Ecology*.

Royle, J. A. and Nichols, J. D. (2003), "Estimating abundance from repeated presence-absence data or point counts," *Ecology*, 84, 777–790.

Sharov, A. A., Leonard, D., Liebhold, A. M., Roberts, E. A., and Dickerson, W. (2002), "'Slow the Spread', a national program to contain the gypsy moth," *Journal of Forestry*, 30–35.

Skellam, J. G. (1951), "Random dispersal in theoretical populations," *Biometrika*, 38, 196–218.

Thompson, S. K. (1992), *Sampling*, New York: John Wiley.

Van den Bosch, F., Hengeveld, R., and Metz, J. A. J. (1992), "Analyzing the velocity of animal range expansion," *Journal of Biogeography*, 19, 135–150.

Van den Bosch, F., Metz, J. A. J., and Diekman, O. (1990), "The velocity of spatial population expansion," *Mathematical Biology*, 28, 529–565.

Wikle, C. K. (2002), "Hierarchical Bayesian models for predicting the spread of ecological processes," *Ecology*, 84, 1382–1394.