Fitting statistical distributions to sea duck count data: Implications for survey design and abundance estimation

Elise F. Zipkin [a,*], Jeffery B. Leirness [b,c], Brian P. Kinlan [d,e], Allan F. O'Connell [f], Emily D. Silverman [b]

[a] USGS Patuxent Wildlife Research Center, 12100 Beech Forest Rd., Laurel MD, 20708, United States

[b] USFWS Division of Migratory Bird Management, 11510 American Holly Dr., Laurel MD, 20708, United States

[c] Department of Entomology and Wildlife Ecology, University of Delaware, Newark, DE 19716, United States

[d] NOAA National Ocean Service, National Centers for Coastal Ocean Science, Center for Coastal Monitoring and Assessment, Biogeography Branch, SSMC-4, N/SCI-1, 1305 East-West Hwy., Silver Spring, MD 20910-3281, United States

[e] Consolidated Safety Services, Inc. 10301 Democracy Lane, Suite 300, Fairfax, VA 22030, United States

[f] USGS Patuxent Wildlife Research Center, BARC East, Bldg. 308, 10300 Baltimore Ave., Beltsville, MD 20705, United States

ARTICLE INFO

ABSTRACT

Determining appropriate statistical distributions for modeling animal count data is important for accurate estimation of abundance, distribution, and trends. In the case of sea ducks along the U.S. Atlantic coast, managers want to estimate local and regional abundance to detect and track population declines, to define areas of high and low use, and to predict the impact of future habitat change on populations. In this paper, we used a modified marked point process to model survey data that recorded flock sizes of Common eiders, Long-tailed ducks, and Black, Surf, and White-winged scoters. The data come from an experimental aerial survey, conducted by the United States Fish & Wildlife Service (USFWS) Division of Migratory Bird Management, during which east-west transects were flown along the Atlantic Coast from Maine to Florida during the winters of 2009–2011. To model the number of flocks per transect (the points), we compared the fit of four statistical distributions (zero-inflated Poisson, zero-inflated geometric, zero-inflated negative binomial and negative binomial) to data on the number of species-specific sea duck flocks that were recorded for each transect flown. To model the flock sizes (the marks), we compared the fit of flock size data for each species to seven statistical distributions: positive Poisson, positive negative binomial, positive geometric, logarithmic, discretized lognormal, zeta and Yule–Simon. Akaike's Information Criterion and Vuong's

* Corresponding author. Tel.: +1 301 497 5810; fax: +1 301 497 5545.
E-mail address: ezipkin@usgs.gov (E.F. Zipkin).

closeness tests indicated that the negative binomial and discretized lognormal were the best distributions for all species for the points and marks, respectively. These findings have important implications for estimating sea duck abundances as the discretized lognormal is a more skewed distribution than the Poisson and negative binomial, which are frequently used to model avian counts; the lognormal is also less heavy-tailed than the power law distributions (e.g., zeta and Yule–Simon), which are becoming increasingly popular for group size modeling. Choosing appropriate statistical distributions for modeling flock size data is fundamental to accurately estimating population summaries, determining required survey effort, and assessing and propagating uncertainty through decision-making processes.

Published by Elsevier B.V.

1. Introduction

Effective management of wildlife populations requires high quality estimates of population abundance and distribution with associated measures of uncertainty. Managers use abundance estimates to determine population status, for comparison to environmental carrying capacities, and to monitor population trends [44]. Understanding patterns of abundance and aggregation is necessary at both regional and local scales to evaluate the impacts of conservation actions and human disturbance. Obtaining accurate population indices is difficult, however, because animals are often unevenly and unpredictably distributed [8,9,43]; for example, counts often include many zeros [19,30] and distributions of count data can be extremely right skewed [4,17]. The problem is compounded by a need for consistent repeated estimates over time; yet, sufficient data to characterize highly aggregated species distributions are expensive to collect and maintain. The choice of appropriate statistical models for wildlife count distributions is fundamental for consistency and efficiency of abundance and distribution estimation and to facilitate more reliable uncertainty assessments [48].

Waterfowl managers are especially interested in population estimates for five species of North American sea ducks (Tribe Mergini) that winter in large numbers off the Atlantic coast of the United States (Sea Duck Joint Venture 2003). Data from a variety of sources suggest that Common eiders (*Somateria mollissima*), Long-tailed ducks (*Clangula hyemalis*), and Black, Surf, and White-winged scoters (*Melanitta nigra*, *M. perspicillata*, and *M. fusca*) may be declining [36,42], and proposed offshore energy development has the potential to significantly alter their wintering habitat [13,15,25]. Waterfowl managers need accurate and precise coast-wide winter abundance indices to assess trends and set annual harvest regulations, while energy regulators need predictions of spatial variation in abundance to inform responsible site placement of offshore structures and to guide future development activities.

During the winter, sea ducks form large foraging flocks, but can also be found alone or in small groups [7]. Their distributions can shift within and between years, due to changes in habitat, weather, and prey availability [18,24,26,52], and they can be found up to 40 miles from land [41]. As a result, effective monitoring surveys are expensive, dangerous, and fraught with logistical challenges. If the resulting data are to be worth collecting, then appropriate statistical models to interpret the data need to be available and accessible.

The United States Fish and Wildlife Service (USFWS) Division of Migratory Bird Management initiated an experimental aerial survey, conducted from Maine to Florida in the winters of 2009–11, to assess the feasibility and effectiveness of a long-term winter sea duck monitoring program along the Atlantic coast. Determining whether precise estimates of regional annual abundance are possible for the five target species is necessary to evaluate the effectiveness of the survey. To meet these objectives, we explore the fit of a set of statistical models to data from the Atlantic coast wintering sea duck survey. Our goals are: (1) to identify a model, or models, that accurately describes the distribution of counts, characterized by an unusually heavy right tail and an excessive number of zeros; (2) to determine if the best model choice varies by species; and (3) to compare parameter estimates among species and assess whether more refined models (e.g., that stratify regions by high and low density