ARTICLE IN PRESS
Spatial Statistics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

A spatial–temporal double-hurdle model for extremely over-dispersed avian count data

Earvin Balderama[a,*], Beth Gardner[b], Brian J. Reich[c]

[a] Department of Mathematics and Statistics, Loyola University Chicago, Chicago, IL, United States
[b] School of Environmental and Forest Sciences, University of Washington, Seattle, WA, United States
[c] Department of Statistics, North Carolina State University, Raleigh, NC, United States

ARTICLE INFO

Article history:
Received 31 December 2015
Accepted 3 May 2016
Available online xxxx

Keywords:
Bayesian hierarchical modeling
Conditional autoregressive models
Extreme values
Generalized Pareto distribution
Seabirds
Zero-inflation

ABSTRACT

Several wind energy facilities are currently being planned for offshore Atlantic waters of the United States. However, relatively little is known about the distribution, abundance and spatio-temporal variability of marine birds in their offshore habitats and it is becoming increasingly necessary to accurately characterize these demographic parameters before assessing the influence of factors such as offshore energy development on populations. Thus, we incorporate a multi-scale approach to develop models for the space-time distribution and abundance of marine birds to identify potential high-use areas in need of further study. With data taken from past and ongoing survey efforts, we provide relative abundance and density estimates for marine birds over a wide geographical area during multiple years. Due to the excessive amount of zeros as well as extremely large counts exhibited in the data, a double-hurdle model is formulated that includes a negative binomial and a generalized Pareto distribution mixture. Spatial heterogeneity is modeled using a conditional auto-regressive (CAR) prior, and a Fourier basis was used for seasonal variation. We demonstrate our model by creating probability maps that show areas of high-abundance and aggregation for twenty-four species of marine bird.

* Corresponding author.
E-mail addresses: ebalderama@luc.edu (E. Balderama), bg43@uw.edu (B. Gardner), bjreich@ncsu.edu (B.J. Reich).

http://dx.doi.org/10.1016/j.spasta.2016.05.001
2211-6753/

Please cite this article in press as: Balderama, E., et al., A spatial–temporal double-hurdle model for extremely over-dispersed avian count data. Spatial Statistics (2016), http://dx.doi.org/10.1016/j.spasta.2016.05.001

1. Introduction

One of the fastest-growing segments of the energy market is wind power (American Council On Renewable Energy, 2014), particularly wind resources in offshore waters (American Wind Energy Association, 2013). The potential impacts of offshore wind facilities on marine bird species are not well understood, and may include exposing birds to increased mortality through turbine collisions, altering their behavior and flight pathways (Drewitt and Langston, 2006), and altering the habitat upon which marine birds depend for foraging. One of the first steps in evaluating potential interactions from wind energy facilities is to understand the distribution and abundance of marine birds (Huettman and Diamond, 2001). This information can be used in identifying sensitive and high-use areas of birds and is thus a key component of marine planning (Huettman and Diamond, 2001; Winiarski et al., 2014).

In marine bird research and ecology more generally, count data are often analyzed using Poisson, quasi-Poisson, or negative binomial (NB) distributions, or zero-inflated versions of these models (Ver Hoef and Boveng, 2007; Oppel et al., 2012). Given the level of over-dispersion that can occur in marine bird data, these distributions may not be robust enough for the data, but recent work has shown that other, less commonly used distributions, may be useful in capturing the high variance to mean ratio. Beauchamp (2011) found the power law distribution outperformed NB for modeling group sizes in seven marine bird species in the Western North Atlantic. Zipkin et al. (2014) compare and contrast a suite of distributions for sea duck data in the Western North Atlantic, finding that the discretized lognormal was the best fit over the geometric, logarithmic, zeta, Poisson, NB, and Yule–Simon distributions for modeling flock sizes. Other methods for trying to account for the large variation in marine bird data is using a Box–Cox hurdle model (Menza et al., 2012), which transforms count data to be more normally distributed. The main goal of these studies was to model the over-dispersion and zero-inflation of marine bird data without throwing data out, truncating, or mis-specifying the distribution, to better understand the main ecological drivers or spatial patterns of marine bird distributions.

Our objectives are to build on this foundation and further examine the extreme counts that arise in marine bird data (e.g., large aggregations, where counts can be 500–2000 birds at one location) and to explicitly account for spatial autocorrelation in marine bird data. By doing so, we aim to create more accurate predictions of sea bird distributions across a large spatial domain.

This paper is organized as follows: Section 2 gives details on how the data were collected, and what covariates were considered in the study. Section 3 describes the spatio-temporal double-hurdle model and our hierarchical Bayesian approach to parameter estimation. Model performance is evaluated for 24 different species in the western North Atlantic, and exposure maps are created in Section 4. We conclude with a discussion in Section 5.

2. Marine bird data

Between 1992 and 2010, a total of 43,701 boat and aerial transects made up an avian data collection effort that spanned the Atlantic coastline from Maine to Florida. We consider only strip sampling surveys, where observers record all observations within a fixed distance from a transect line (Williams et al., 2002). In the historical database, this amounted to over 2 million individual marine birds representing nearly 200 species in 133,890 separate sightings. For each sighting, the count, species, date and location were recorded (other information were also recorded, but not consistently across surveys).

We created approximate 4×4 km contiguous grid cells over the Atlantic coast region to match the resolution of biophysical covariate information obtained from the National Oceanographic and Atmospheric Administration (NOAA). These covariates, i.e., sea surface temperature (SST), chlorophyll-*a* concentration (CHL), and ocean depth (DEP), were used as a proxy for the variability in food resources that may influence where we expect birds to occur (Zipkin et al., 2010). Because the data are very sparse, observations which occurred in the same month and year were combined to give monthly counts for each grid cell. We analyze data from July 2002 to November 2010 (101 months), on 15,984 grid cells which satisfy the following constraints: north of 35.25° latitude, east of −76.5° longitude, and with an ocean depth of no deeper than 500 m. In defining this study region, we

Please cite this article in press as: Balderama, E., et al., A spatial–temporal double-hurdle model for extremely over-dispersed avian count data. Spatial Statistics (2016), http://dx.doi.org/10.1016/j.spasta.2016.05.001