Atlantic Marine Assessment Program for Protected Species (AMAPPS): Aerial Seabird Surveys 2010 – 2014.

U.S. Fish and Wildlife Service

Division of Migratory Birds, Northeast Region

Prepared by Malcolm "Tim" Jones and Jeffery B. Leirness

13 May 2015

Table of Contents

Executive Summary

Since August 2010, the U.S. Fish and Wildlife Service (USFWS) has conducted aerial seabird surveys as part of the Atlantic Marine Assessment Program for Protected Species (AMAPPS). All surveys were conducted using the fixed-wing Quest Kodiak aircraft flown at 110 knots and 70 m (200 ft) altitude, while an observer and pilot-observer counted all seabirds, sea turtles, and marine mammals within 400m-width strip transects. Transects paralleled the latitudinal lines and were spaced roughly every 5 nm at latitudes ending in xx˚x1' N and xx˚x6' N. Prior to August 2011, all surveys were flown with transects that extended east from the coastline to either the 30m depth contour or no more than 50nm from the nearest point of land. In August 2011, transects were extended to the 30 m depth contour and all subsequent surveys flew these longer transect lines. The August 2010 survey covered the Atlantic coast south of the North-South Carolina border (33˚51' N) to the southern tip of Florida (24˚30' N), including the Florida Keys. Surveys flown in December 2010 and January 2011 (Winter 1) targeted the mid-Atlantic region from Atlantic City, NJ (39˚11' N) to Wilmington, NC (34˚36' N), including Delaware Bay and Chesapeake Bay. Since August 2011, all surveys have covered the entire U.S. Atlantic coast from the U.S.-Canadian border (44˚46' N) to Cape Canaveral, FL (28˚26' N).

More than 780,000 observations have been recorded and entered into our database. These data represent a statistically rigorous survey design that monitors seabird occurrence and abundance from the coastline out to depth of 30m unless that depth exceeds more than 50nm from the nearest point of land. Surveys have been conducted across all seasons although we found the summer to have few seabirds present in our sampled area. After our initial summer surveys in 2010 and 2011 we shifted efforts to fall, winter and spring which have a greater abundance of seabirds.

This report summarizes the raw data from these surveys and presents initial, uncorrected densities per square kilometer surveys for Northern Gannets and a set of species groups. We are developing statistical methods and conducting additional surveys in order to better understand detectability issues and biases associated with our fleet aircraft.

Introduction

The assessment program described here was designed as a comprehensive effort to collect data required to estimate abundance and develop seasonally specific, localized density estimates for marine mammals, marine turtles, and seabirds. The program has coordinated the data collection and analysis efforts of the NMFS Northeast and Southeast Fisheries Science Centers and the U.S. Fish and Wildlife Service Division of Migratory Birds. The original proposal focused on objectives that addressed the immediate needs of the funding agencies, BOEM and The Navy, that included: 1) Collect broad-scale data over multiple years on the seasonal distribution and abundance of marine mammals (cetaceans and pinnipeds), marine turtles, and seabirds using direct aerial and shipboard surveys of coastal U.S. Atlantic Ocean waters; 2) Conduct tag telemetry studies within surveyed regions of marine turtles, pinnipeds and seabirds; 3) Explore alternative platforms and technologies to improve population assessment studies; 4) Assess the population size of surveyed species at regional scales; and 5) Develop models and associated tools to translate these survey data into seasonal, spatially-explicit density estimates incorporating habitat characteristics.

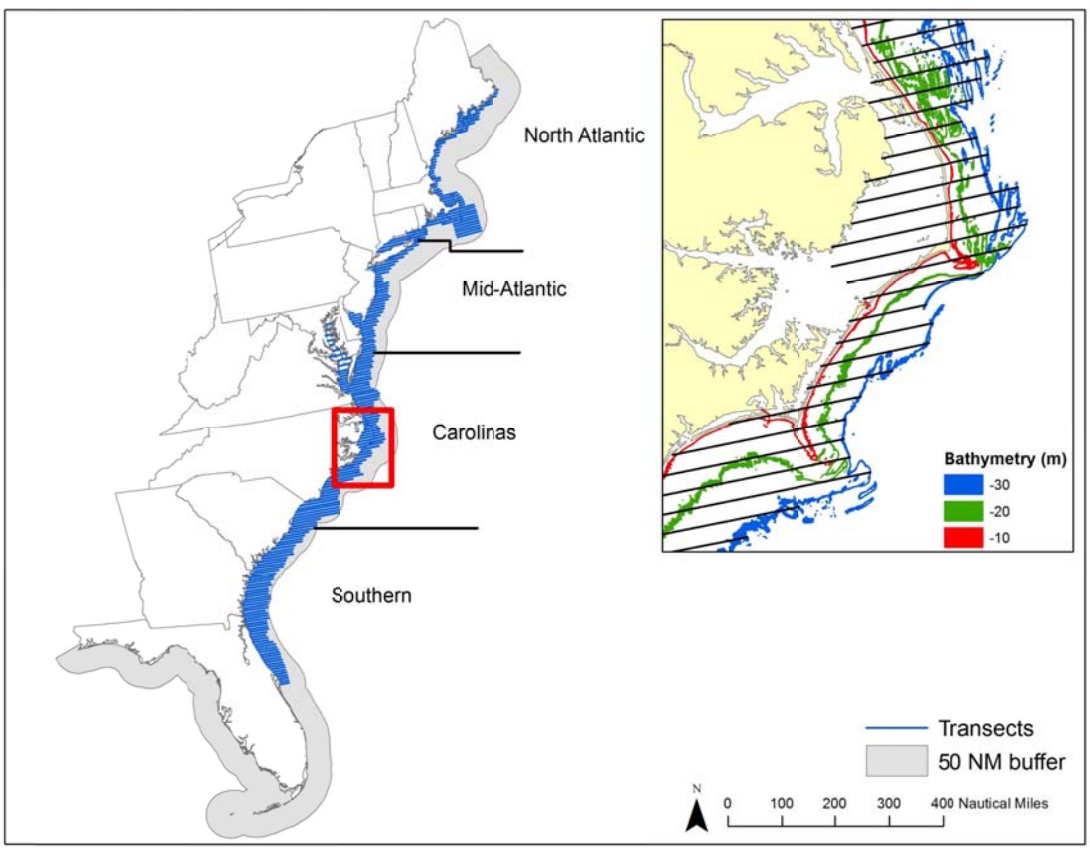
Nine surveys have been flown between August 2010 and October 2014. Crews flew 103,634km (55,958nm) of strip transects survey seabirds, sea turtles and marine mammals. Our database, not including two surveys, contains more than 780,000 records of seabird observations. Division of Migratory Bird staff are processing data from the final two surveys. All error-checked data have been submitted the Atlantic Seabird Catalog at the request of BOEM.

Methods

Survey design:

Transects are located at 5’ (~5 nm) intervals at every *1’ and *6’ minute of latitude and extend out to the 30m depth contour or nor more than 50 nm from the nearest land (Figure 1). Transects are numbered by their latitude ID (degrees/minutes, e.g., 3436 for 34o36’N); two additional digits that index multiple segments along a transect. Most transects have only a single segment, which is identified by 00. For transects with multiple segments (which may be separated by land or open water or may be contiguous), the digits 00 identify the most westerly transect, 01 indicates the next transect to the east, etc. Continuous transects that are divided into multiple sections cross survey strata (e.g, 365600, 365601, and 365602 at the mouth of the Chesapeake Bay); their start and stop points should be recorded, even if counting is not interrupted. Transect length depends on the survey being flown and the location along the coast. For the AMAPPs surveys, transects extend 30 nm offshore. Transects extend from Cape Canaveral, FL to the US-Canada border.

Figure 1. USFWS AMAPPS seabird survey design

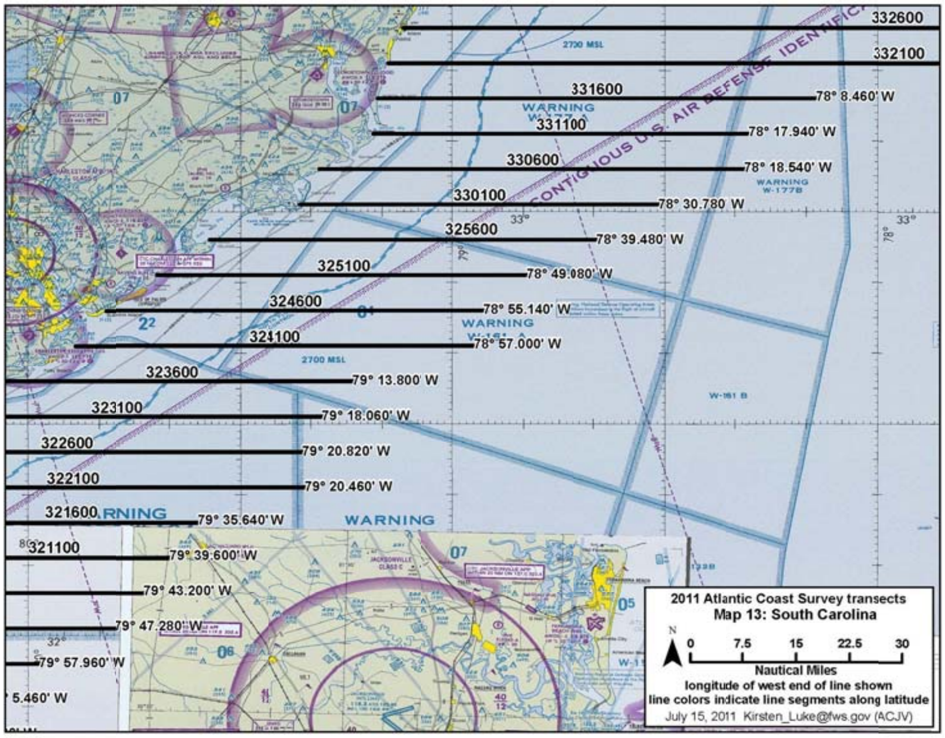


Survey Procedures (SOP):

Surveys were flown in USFWS fleet Kodiak amphibious aircraft, Quest Inc., with the exception of the summer 2010 surveys (Figure 2). The summer 2010 surveys were flown using older fleet aircraft; a twin engine Partenavia P68 Observer and a Cessna amphibious aircraft. All surveys are flown at a height of 200ft above ground level (AGL) and a speed of 110 kts. All seabirds, turtles and marine mammals observed within a 400m wide transect are recorded to the lowest taxonomic level possible along with number of individuals and for all observers, other than the pilot, the distance band. Our survey specific SOP is outlined below. Each survey crew is provided a set of maps depicting each transect overlaid on aviation maps as well as a set of GPX files that can be loaded on each aircraft's GPS unit for navigation (Figure 3).

Figure 2. USFWS fleet aircraft used in AMAPPS aerial seabird surveys. Counterclockwise from bottom left: Quest Kodiak, Partenavia P68 Observer, Cessna 206 Amphib.

Figure 3. Example crew map of transects overlaid on aviation charts for Charleston, SC area.



Transect width are marked on either the wing strut or window using a grease pencil or dry erase marker. Crew members are required to mark the 200m outer edge of the transect before starting the first transect in their crew area. This is done with the use of a standard clinometer and marking a 17° angle on some portion of the aircraft. Preferably this is done while on the ground to eliminate the effect of any turbulence. Observers on the right side of the aircraft also have to mark the 100m boundary (31° angle) since they record data into multiple distance bins.

All data are recorded to hard drives using software developed by the USFWS for aerial surveys. Observations are recorded using the program Record (Version 6.4, 2/11/2009) that stores each detection in a WAV file while GPS coordinates, GPS error and time since midnight are logged for every observation. Afterwards, the observer uses a program, Transcribe (Version 3.1, 3/13/2008), that allows the user to enter the data recorded on each WAV file and merges those data with the appropriate GPS location and timestamp.

<u>Survey SOPs</u>

- Surveys should be flown at 110 kts, at a height of 200 ft.
- Initiate surveys when wind speed ≤ 15 kts, discontinue if winds exceed 20 kts.
- Transect **width** is 200m on either side of the aircraft. (The approximate *actual* transect width will be determined by the observer's ability to see under the plane. If available, please use a

clinometer to estimate inner transect boundary.) At an altitude of 200 ft, the 200m boundary is at a 17° angle from horizontal and the 100m boundary is at 31°.

- Record "**BEGSEG"** and "**ENDSEG"** (required by recording program used in USFWS aircraft) at the start and finish of each east-west transect, including continuous lines broken into distinct transects, such as those over Pamlico Sound and the mouths of Chesapeake and Delaware Bay. Transcribe this code in the species/type field.
- Transcribe the six digit segment number (e.g., 382100) in the count field for every BEGSEG and ENDSEG record.
- If counting is suspended for any reason between the BEG and ENDSEG (technical difficulties, airsickness, flying over land, etc.), record "**ENDCNT**" to mark the break. Record "**BEGCNT**" when surveying resumes.
- Record all **sea ducks**, **diving ducks**, and other **seabirds** observed from the edge of the waterline eastward, including birds associated with ***exposed shoals***. Do not record birds sitting on pilings, jetties, beaches, boats, in trees, etc. Birds should be in the air over water or on the water.
- Birds flying above the plane should be recorded, if they are within the transect.
- Record all marine mammals and turtles observed on transect.
- All commercial fishing boats should be recorded with the code TRAWL along with the perpendicular distance (nm) from the transect line (including boats >200m from the plane).
- Balloons (deflated, floating on the surface) within the transect should be recorded with the code BALN.
- Report all sea ducks and seabirds to the lowest taxonomic level possible.
- See Appendix 5.1 for a list of species abbreviations.
- For **mixed flocks** seen within the transect boundaries:
    - Used species code MIXD
    - In the count field, code the entire mixed flock size
    - Add a comment to the end of that transcribed record with species proportions e.g., MIX,500, 25% SUSC; 50% BLSC; 25% WWSC. Exact counts of species are preferable, if known. It is also preferable to record the comment ***without commas***, as the *Transcribe* program uses commas to delimit variable fields.
- Record observation conditions on a 5-point Likert scale in the "Condition" header field. You should consider all factors influencing observation conditions when recording this (e.g., sea state, glare, observer alertness, etc). Use the following codes:
    - 1, 2, 3, 4, 5 with 1 = Worst observation conditions, 3 = Average condition, and 5 = Best observation conditions.
- Enter the condition code in the condition field at the start of each transect and for each observation; when conditions *change*, using the code **COCH** in the species/type field to indicate "condition change." Include the new condition value in both the condition field and the count field.
- *Observers* will record two additional pieces of information with each record (i.e., pilots are excluded):

(1) the 100 m band within which the bird was observed:
1 = [0,100m], 2 = (100,200m], 0 = within 200m, band unknown.
(Any birds recorded outside the 200m band should be coded as band = 3 and the offline code should be "y;" these records are not within the survey protocol.)

(2) if the bird was flying (F); Code non-flying birds with (S); and unknown as (0)

Data and File Management:

All files on the aircraft computer are backed up daily by the flight crew onto USB drives and then copied to hard drive on a laptop. At the end of the survey all files including the raw WAV, track files and transcribed data files are uploaded to the Department of the Interior (DOI) AMAPPS SharePoint site by crew area. Once the files are received by staff at Patuxent Wildlife Research Center, a series of R programs are run to check for data entry errors and to format the raw data for input into a Microsoft Access (MS Access Version 14.0.7143.5000) database (Appendix 5.2).

We also maintain a geodatabase representing all spatial information related to survey design transects, what we actually fly each survey (tracks) and observations. This geodatabase was created and maintained using ESRI ArcGIS (Version 10.2.2). We use this geodatabase to calculate how far from the coast, water depth and slope of the bottom for each observation.

- Make certain each computer's time and date are correct and that they are ***SYNCHRONIZED*** (clock synchronization is especially important for the data quality control for this survey, because we have no segment files to check against the entries you include for your location and because transect numbering can be prone to mis-entry).
- The Survey Name (and folder) for the Record and Transcribe programs will be provided to you at the beginning of the survey. If you switch survey planes or are relieved by a new observer, please move your files into a subfolder named with your initials.
- **Crew names**: crew names are designated by the four digit latitude of the northernmost transect. E.g., the northern crew will be Crew4446, which is the northernmost survey line.
- Please record any partial or missed transects in the *SurveyNotes.xls* file provided. Record any other deviations from the SOP and relevant survey details/comments in this file.
- Files names: each observer should have one data file for each survey day. The files should be named **Crew####**_MODAYEAR_birds.txt**, where Crew### is the crew name (see above), ** = lf for the pilot and rf (or rr, or lr) for the observer, MO = two digit month, DA = two digit day (e.g., 01 for the first day of the month), and YEAR = four digit year. For example, *Crew4446lf_02082012_birds.txt* includes the pilot's observations for Feb 8, 2012 Crew4446.
- There should be two track files submitted for every survey day. The corresponding track file names are **Crew####**_MODAYEAR_track.txt**.
- Backups of all files (track and sound files, as well as any transcribed files) from each computer should be made nightly onto the USB drive that belongs to that computer.
- Transcribed data files and pilot and observer track files should be uploaded regularly to the survey sharepoint site (the URL of this site will be included in the pre-survey materials), not less than every 5 survey days.
- At the end of the survey, a zipped file containing all ASCII data files, the pilot and observer track files, and the *SurveyNotes CREWAREA.xls* file should be sent to the survey coordinator.

Order of data fields in the transcribed file:

Header fields:

- Year (4 digits, 2011)
- Month (1 digit, 1 or 2) no leading zeros
- Day (1 or 2 digits) no leading zeros
- Seat (2 digits, lf, rf, rr, or lr)
- Observer (your 3 initials in lowercase letters … please use THREE initials!)
- Transect (6 digits, line #s will be the latitude degrees concatenated with the latitude minutes and then with the segment number [00, 01, etc.]. Typically there will be just one line segment "00," but when more than one segment occurs on the same latitude you might also have segment "01", etc., e.g., line on 36 deg 21 min, segment 00 = 362100.
- Observation condition (1 digit, 1-5)
- Offline (1 character, "n" = online/**within the 200m width while on transect**, "y" = offline)

Fields created by Hodges programs automatically:

- Species/type code
- Count (this is the count you enter into the count window in Hodge's program – if a flock crosses the 200m transect edge, include only those birds within the transect)

Additional fields:

- Distance Band (1 – 0-100m, 2- 100-200m, or 0 if unknown)
- Bird flight status (F = flying; S = sitting on water; 0 = unknown)
- Comment on composition of MIXD records

<u>Training</u>

In February 2012 we held a field training event for USFWS observers and pilots on the Outer Banks of North Carolina. The goals of this training were to a) increase the identification ability of our biologist-pilots and observers and b) expose new observers to aerial survey experience. USFWS biologist-pilots have been using aerial surveys to count breeding and non-breeding waterfowl for more than 50 years and seaducks for more than 4 years. Seabirds represented a new and unfamiliar group of birds rarely encountered by our biologist-pilots and observers. We contracted with Brian Patteson, an expert in seabird identification that runs a pelagic birding company out of Hatteras, NC.

The first day of the training consisted of reviewing our survey protocols and a presentation by Brian Patteson reviewing the likely seabirds that could be encountered along the U.S. Atlantic Coast. During this presentation, he reviewed the range of the species and identification tips. The second day was spent aboard the Stormy Petrel II for our biologist-pilots and observers to see a wide array of seabirds (Figures 4 and 5). The third day we introduced the observers who had never participated in an aerial survey the chance to experience survey conditions. All flights were conducted from Dare County Regional Airport in Manteo, NC. Due to water temperatures all participants were required to wear survival suits and practiced identifying and counting seabirds on two transects just off of the coast near the airport. Observers were introduced to aircraft safety procedures as well as using the computers for recording data.

Figure 4. Participants of seabird survey training in February 2012. Back row from left to right: Troy Wilson, Mark Koneff, Melanie Steinkamp, Emily Silverman, Jim Wortham, Dean Demarest, and Sarah Yates. Middle row from left to right: Walt Rhodes, Tim Jones, Steve Earsom, Mao Lin, Holiday Obrecht, and ?. Front row from left to right: Eric Kirshner, Jeff Shenot and Jeff Leirness.

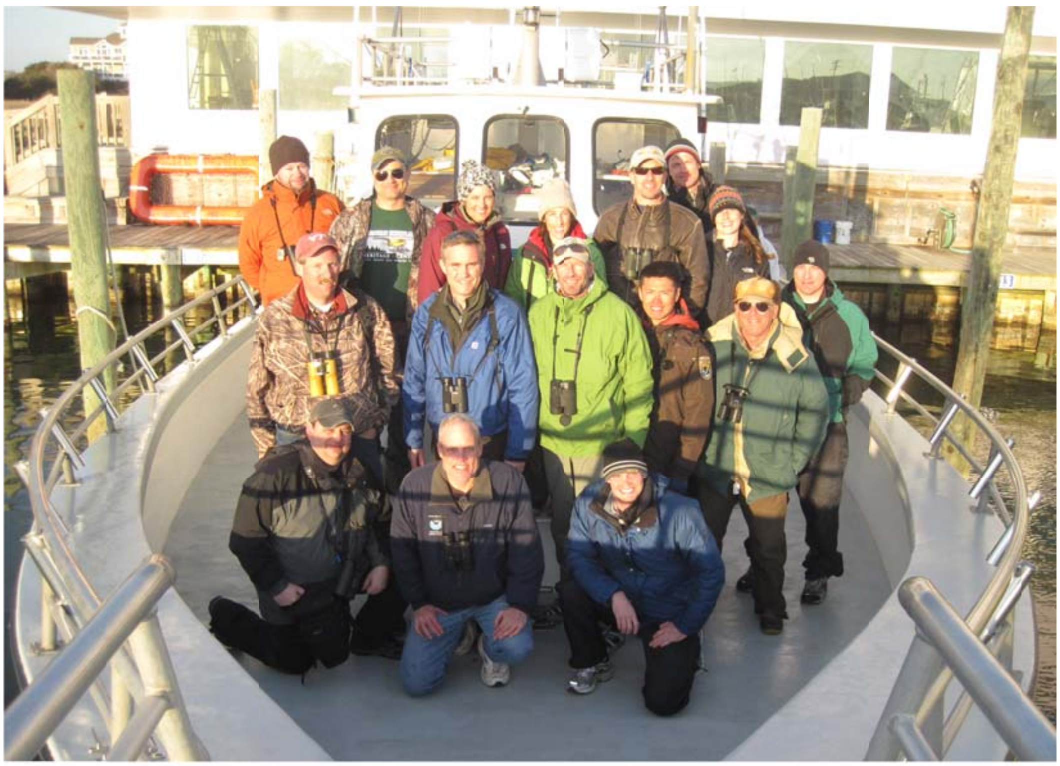


Figure 5. Seabirds observed during the seabird identification training session off Hatteras, NC. Clockwise from top left: Common Loon, adult Northern Gannet, juvenile Northern Gannet, Great Shearwater.

Analysis

Analyses have focused on how to describe sparse, yet highly aggregated counts of seabirds. Such patterns of abundance make estimating total numbers of individuals difficult and may provide biased estimates of uncertainty. Staff worked with personnel from the USGS Patuxent Wildlife Research Center to evaluate a set of statistical distributions that can describe highly right-skewed distribution of flock frequencies (Zipkin et al. 2012; Appendix 3). Current efforts are focused on understand detectability and biases associated with our fleet aircraft. We also are examining whether we can develop a statistical model that would allow us to impute species identification on certain guilds of species.

Results

USFWS conducted nine seabird surveys between August 2010 and October 2014 (Table 1). Crews flew 103,634km (55,958nm) of strip transects survey seabirds, sea turtles and marine mammals. The surveys

Table 1. Surveys flown by the U.S. Fish and Wildlife Service as part of the Atlantic Marine Assessment Program for Protected Species.

Survey	Start Date	End Date	Survey Distance (km)	# Transects	# Replicates
2010 August	August 3	August 24	5,421	115	62
2010 December	December 3	December 11	2,164	89	0
2011 January	January 16	January 17	619	22	0
2011 August	July 30	August 23	13,979	267	8
2012 March	March 15	March 31	13,784	282	0
2012 October	September 29	October 12	13,914	283	0
2013 September	September 16	September 28	17,112	266	0
2014 February	January 28	February 12	20,564	285	0
2014 October	October 6	October 22	16,077	189	0

in August 2010 did not conform to our survey design due to the Gulf Oil Spill. In response to that incident BOEM agreed to our shifting the survey transect south and into the eastern Gulf of Mexico (Figure 6A). In December 2010 we flew our first that went further than 8-10nm offshore (Figure 6B). The remainder of the surveys generally followed the survey design but varied due to weather or mechanical difficulties with the aircraft (Figures 6C-6H). Due to availability we are not able to maintain consistent crews over all the surveys but track observers to account for different detection probabilities among observers (Table 2).

Total counts and number transects for each species or species group observed are presented in Tables 3-7. Due to our data analyst leaving these summaries are not available for the 2013 and 2014 surveys at this time. Number of marine mammals and sea turtles observed during all surveys between 2010 and 2012 are presented in Table 8.

Raw density estimates per square kilometer were calculated for all aerial surveys from 2010 through 2012. Until we can correct these raw densities based on detectability the only species we are comfortable mapping individually is Northern Gannet. All other species were grouped into higher taxonomic groupings that included: alcids; gulls; loons; terns, sea and diving ducks; marine mammals and sea turtles. The results are shown in Figures 7 – 17.

Figure 6. Transect lines flown by the U.S. Fish and Wildlife Service as part of the Atlantic Marine Assessment Program for Protected Species surveys completed in (A) August 2010, (B) December 2010 and January 2011, (C) August 2011, (D) March 2012, (E) October 2012, (F) Septermber 2013, (G) February 2014 and (H) October 2014.

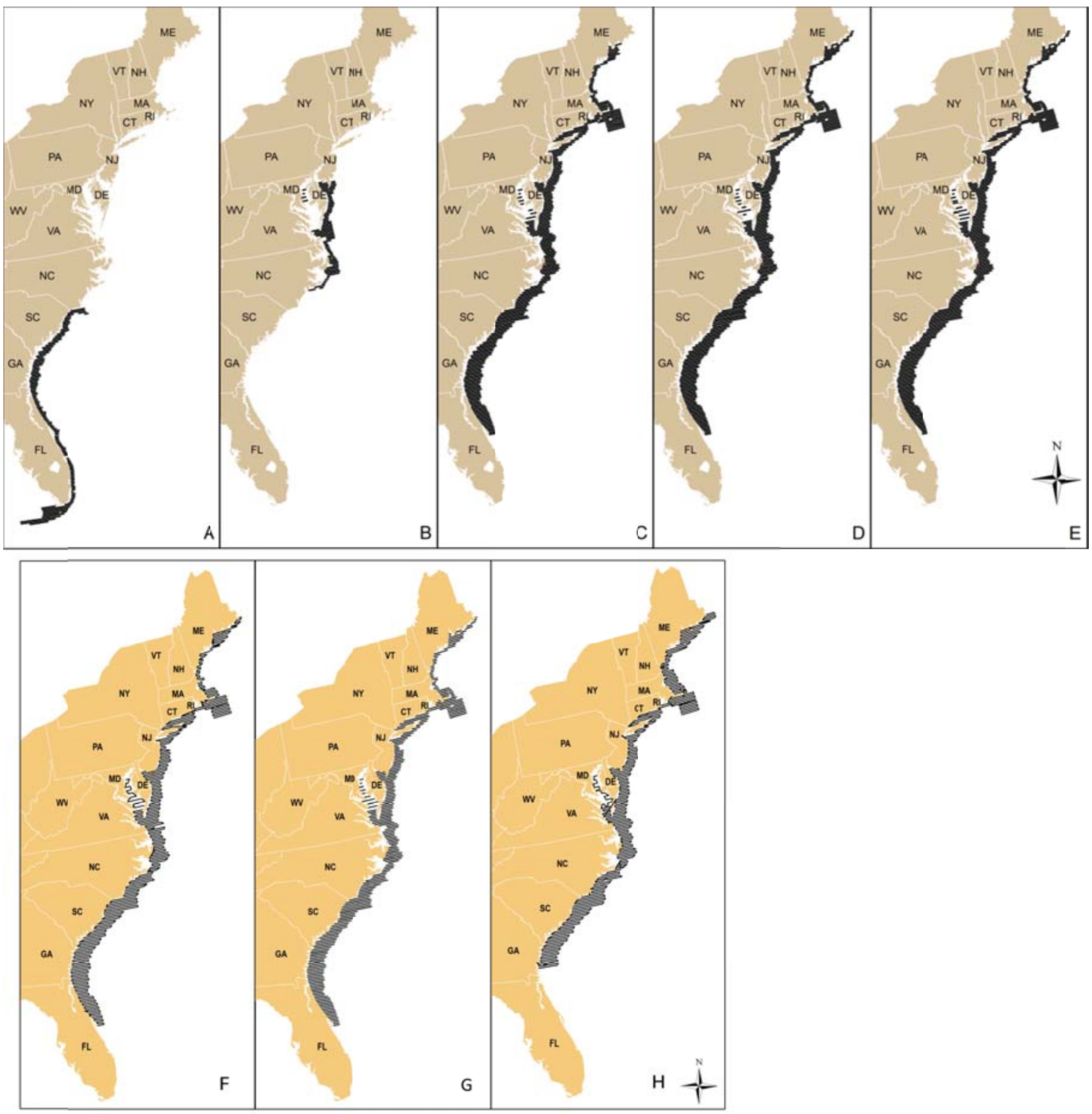


Table 2: Survey crews.

Survey	Crew*	Pilot(s)	Observer(s)
2010 August	2941	Mark D. Koneff	Doug J. Forsell
	3351	James S. Wortham	Emily R. Bjerre
2010 December	3916	James S. Wortham	Doug J. Forsell & M. Tim Jones
2011 January	3726	James S. Wortham	Timothy P. White
2011 August	3606	Walt E. Rhodes	M. Tim Jones
	4116	James S. Wortham	Dean W. Demarest
	4311	Fred H. Roetker	Holliday H. Obrecht
2012 March	3316	Walt E. Rhodes	M. Tim Jones
	3651	Stephen D. Earsom	Eric. L. Kershner
	4056	James S. Wortham & Mark D. Koneff	Caleb S. Speigel, Dean W. Demarest, & Melanie J. Steinkamp
	4446	Mark D. Koneff	Mao T. Lin & Sarah F. Yates
2012 October	3316	Walt E. Rhodes	M. Tim Jones
	3756	Stephen D. Earsom	Eric. L. Kershner
	4056	James S. Wortham	Mao T. Lin
	4446	Mark D. Koneff	Sarah F. Yates
2013 September	3316	Fred H. Roetker	M. Tim Jones
	3651	James S. Wortham	Pam Loring
	4056	Stephen D. Earsom	Mao T. Lin
	4446	Mark D. Koneff	Mao T. Lin
2014 February	3316	Fred H. Roetker	Caleb S. Speigel
	3651	James S. Wortham	Robert Simmons
	4056	Stephen D. Earsom	Mike Chouinard
	4446	Mark D. Koneff	Mao T. Lin
2014 October	3521	James S. Wortham	Fred H. Roetker & M. Tim Jones
	4126	Stephen D. Earsom	M. Tim Jones
	4446	Mark D. Koneff	Sarah F. Yates

*Numbers indicate the latitude (degrees-minutes) of the northernmost transect in the crew area.

Table 3: Total count (unique number of transects) for all seabirds identified during the August 2010 survey.

Species Group	Species	Crew2941	Crew3351	Total
Alcids	Unidentified large alcid	-	89 (20)	89 (20)
Larids	Black skimmer	15 (1)	-	15 (1)
	Herring gull	1 (1)	65 (24)	66 (25)
	Laughing gull	51 (15)	-	51 (15)
	Ring-billed gull	8 (5)	9 (1)	17 (6)
	Unidentified black-backed gull	-	183 (18)	183 (18)
	Unidentified large gull	10 (4)	8 (8)	18 (12)
	Unidentified small gull	-	42 (10)	42 (10)
	Unidentified gull	-	1,125 (32)	1,125 (32)
	Brown noddy	2 (2)	-	2 (2)
	Bridled tern	2 (2)	-	2 (2)
	Least tern	2 (1)	-	2 (1)
	Roseate tern	66 (21)	-	66 (21)
	Unidentified large tern	33 (12)	22 (10)	55 (22)
	Unidentified small tern	9 (8)	3 (3)	12 (11)
	Unidentified tern	10 (5)	33 (15)	43 (20)
Pelicaniforms	Northern gannet	5 (4)	284 (42)	289 (46)
	Brown booby	1 (1)	-	1 (1)
	Double-crested cormorant	109 (8)	-	109 (8)
	Unidentified cormorant	1 (1)	-	1 (1)
	Magnificent frigatebird	15 (8)	-	15 (8)
	Brown pelican	66 (15)	269 (20)	335 (35)
Tubenoses	Unidentified petrel	1 (1)	-	1 (1)
	Unidentified shearwater	4 (4)	1 (1)	5 (5)
	Wilson's storm-petrel	4 (3)	-	4 (3)
	Unidentified storm-petrel	-	6 (3)	6 (3)
	Unidentified seabird or diving duck	20 (24)	564 (37)	584 (61)
	Unidentified phalarope	-	1 (1)	1 (1)

Table 4: Total count (unique number of transects) for all seabirds identified during the December 2010 and January 2011 surveys.

Species Group	Species	Crew3726	Crew3916	Total
Sea ducks	Bufflehead	4 (1)	723 (16)	727 (17)
	Long-tailed duck	1 (1)	34 (8)	35 (9)
	King eider	-	6 (1)	6 (1)
	Common goldeneye	-	49 (6)	49 (6)
	Unidentified goldeneye	-	32 (4)	32 (4)
	Red-breasted merganser	-	9 (2)	9 (2)
	Unidentified merganser	-	3 (1)	3 (1)
	Unidentified goldeneye or merganser	-	9 (2)	9 (2)
	Black scoter	6 (1)	717 (24)	723 (25)
	Surf scoter	26 (5)	458 (12)	484 (17)
	White-winged scoter	-	3 (1)	3 (1)
	Dark-winged scoter	-	9 (1)	9 (1)
	Unidentified scoter	-	88 (4)	88 (4)
	Unidentified sea duck	7 (2)	10 (5)	17 (7)
Diving ducks	Unidentified scaup	-	197 (4)	197 (4)
Loons	Common loon	66 (15)	33 (11)	99 (26)
	Red-throated loon	16 (6)	194 (33)	210 (35)
	Unidentified loon	57 (14)	283 (45)	340 (51)
Alcids	Razorbill	3 (2)	-	3 (2)
	Unidentified large alcid	3 (3)	-	3 (3)
	Unidentified alcid	-	5 (1)	5 (1)
Larids	Bonaparte's gull	-	510 (50)	510 (50)
	Great black-backed gull	4 (2)	40 (17)	44 (19)
	Herring gull	55 (4)	71 (30)	126 (34)
	Laughing gull	-	36 (17)	36 (17)
	Ring-billed gull	-	508 (37)	508 (37)
	Unidentified black-backed gull	-	13 (9)	13 (9)
	Unidentified large gull	31 (1)	55 (16)	86 (17)
	Unidentified small gull	-	9 (7)	9 (7)
	Unidentified gull	35 (13)	536 (54)	571 (61)
	Black-legged kittiwake	17 (9)	-	17 (9)
	Forster's tern	-	6 (1)	6 (1)
	Royal tern	-	10 (3)	10 (3)
	Unidentified large tern	-	21 (8)	21 (8)
	Unidentified small tern	-	36 (8)	36 (8)
	Unidentified tern	-	37 (15)	37 (15)
Pelicaniforms	Northern gannet	2,237 (15)	1,566 (61)	3,803 (63)
	Double-crested cormorant	3 (2)	155 (9)	158 (11)
	Unidentified cormorant	1 (1)	167 (7)	168 (8)
	Brown pelican	-	23 (7)	23 (7)
Tubenoses	Unidentified shearwater	-	99 (3)	99 (3)
	Unidentified storm-petrel	-	1 (1)	1 (1)
	Unidentified seabird or diving duck	13 (5)	851 (59)	864 (59)

Table 5: Total count (unique number of transects) for all seabirds identified during the August 2011 survey.

Species Group	Species	Crew3606	Crew4116	Crew4311	Total
Sea ducks	Bufflehead	-	8 (1)	-	8 (1)
	Long-tailed duck	-	-	1 (1)	1 (1)
	Common eider	-	-	264 (7)	264 (7)
	Unidentified eider	-	-	153 (4)	153 (4)
	Unidentified merganser	-	-	6 (3)	6 (3)
	Black scoter	-	-	248 (17)	248 (17)
	Unidentified scoter	-	-	61 (2)	61 (2)
Loons	Common loon	-	-	2 (2)	2 (2)
	Red-throated loon	-	-	16 (7)	16 (7)
	Unidentified loon	-	-	3 (3)	3 (3)
Alcids	Dovekie	-	-	1 (1)	1 (1)
	Unidentified alcid	-	-	32 (7)	32 (7)
Larids	Great black-backed gull	10 (4)	113 (32)	204 (42)	327 (77)
	Herring gull	9 (5)	269 (50)	687 (65)	965 (116)
	Iceland gull	1 (1)	-	-	1 (1)
	Laughing gull	565 (77)	8 (5)	3 (1)	576 (83)
	Ring-billed gull	1 (1)	36 (16)	62 (22)	99 (39)
	Unidentified black-backed gull	-	1 (1)	135 (23)	136 (24)
	Unidentified large gull	5 (4)	60 (18)	478 (27)	543 (48)
	Unidentified small gull	-	1 (1)	283 (28)	284 (29)
	Unidentified gull	309 (13)	419 (26)	1,876 (48)	2,604 (83)
	Caspian tern	3 (3)	-	-	3 (3)
	Common tern	-	-	8 (3)	8 (3)
	Forster's tern	18 (8)	-	1 (1)	19 (9)
	Gull-billed tern	25 (6)	-	-	25 (6)
	Least tern	347 (36)	6 (2)	32 (16)	385 (54)
	Royal tern	38 (24)	4 (4)	-	42 (28)
	Unidentified large tern	690 (77)	218 (37)	3 (3)	911 (117)
	Unidentified medium tern	80 (31)	-	-	80 (31)
	Unidentified small tern	103 (25)	314 (56)	32 (7)	449 (88)
	Unidentified tern	352 (55)	7 (3)	958 (58)	1,317 (115)
Pelicaniforms	Northern gannet	-	2 (2)	252 (30)	254 (32)
	Double-crested cormorant	16 (6)	40 (8)	105 (14)	161 (27)
	Unidentified cormorant	3 (1)	51 (10)	402 (34)	456 (44)
	Brown pelican	470 (48)	121 (9)	-	591 (57)
	White-tailed tropicbird	1 (1)	-	-	1 (1)
Tubenoses	Northern fulmar	-	-	14 (6)	14 (6)
	Black-capped petrel	1 (1)	-	-	1 (1)
	Audubon's shearwater	2 (2)	-	-	2 (2)
	Cory's shearwater	169 (29)	-	23 (8)	192 (37)
	Great shearwater	6 (3)	-	221 (34)	227 (37)
	Sooty shearwater	-	-	14 (8)	14 (8)
	Unidentified shearwater	45 (9)	-	463 (40)	508 (49)
	Unidentified storm-petrel	-	90 (21)	274 (36)	364 (57)
	Unidentified seabird or diving duck	87 (30)	40 (6)	1 (1)	128 (37)
	Unidentified phalarope	186 (23)	130 (1)	61 (13)	377 (37)

Table 6: Total count (unique number of transects) for all seabirds identified during the March 2012 survey.

Species Group	Species	Crew3316	Crew3651	Crew4056	Crew4446	Total
Sea ducks	Bufflehead	-	593 (12)	1,036 (11)	75 (8)	1,704 (31)
	Harlequin duck	-	-	65 (2)	-	65 (2)
	Long-tailed duck	-	13 (2)	224 (9)	2,345 (43)	2,582 (54)
	Common eider	-	-	-	5,714 (53)	5,714 (53)
	Unidentified eider	-	-	-	25 (2)	25 (2)
	Common goldeneye	-	3 (2)	13 (3)	29 (6)	45 (11)
	Unidentified goldeneye	-	-	-	24 (4)	24 (4)
	Common merganser	-	12 (2)	-	14 (2)	26 (4)
	Red-breasted merganser	-	105 (6)	3 (1)	872 (33)	980 (40)
	Unidentified merganser	-	55 (10)	13 (8)	17 (3)	85 (21)
	Black scoter	40 (4)	2,285 (11)	370 (15)	283 (10)	2,978 (40)
	Surf scoter	-	279 (8)	1,855 (21)	213 (7)	2,347 (36)
	White-winged scoter	-	2 (2)	26 (5)	898 (24)	926 (31)
	Dark-winged scoter	-	-	110 (11)	1,375 (18)	1,485 (29)
	Unidentified scoter	-	1,004 (9)	1,385 (21)	468 (21)	2,857 (51)
	Unidentified sea duck	25 (1)	106 (6)	12 (4)	3 (1)	146 (12)
Diving ducks	Redhead	-	-	3 (1)	-	3 (1)
	Unidentified scaup	8 (1)	-	509 (4)	34 (3)	551 (8)
Grebes	Red-necked grebe	-	-	4 (2)	-	4 (2)
	Unidentified grebe	-	-	1 (1)	2 (1)	3 (2)
Loons	Common loon	28 (16)	381 (47)	155 (45)	320 (56)	884 (164)
	Red-throated loon	242 (34)	977 (56)	586 (46)	693 (37)	2,498 (173)
	Unidentified loon	69 (8)	10,084 (20)	278 (37)	25 (8)	10,456 (73)
Alcids	Dovekie	-	-	-	67 (12)	67 (12)
	Razorbill	-	1 (1)	-	527 (13)	528 (14)
	Black guillemot	-	-	-	60 (9)	60 (9)
	Unidentified murre	-	-	-	124 (11)	124 (11)
	Atlantic puffin	-	-	-	14 (4)	14 (4)
	Unidentified alcid	-	2 (2)	91 (9)	465 (27)	558 (38)
Larids	Bonaparte's gull	117 (25)	2,051 (30)	36 (4)	2 (2)	2,206 (61)
	Great black-backed gull	-	27 (17)	82 (19)	1 (1)	110 (37)
	Herring gull	28 (16)	3,174 (54)	477 (48)	3,636 (65)	7,315 (183)
	Laughing gull	95 (26)	4 (3)	3 (2)	-	102 (31)
	Lesser black-backed gull	1 (1)	-	1 (1)	-	2 (2)
	Ring-billed gull	221 (38)	-	15 (4)	333 (23)	569 (65)
	Unidentified black-backed gull	2 (2)	7 (4)	36 (10)	131 (37)	176 (53)
	Unidentified large gull	-	-	35 (14)	-	35 (14)
	Unidentified small gull	1 (1)	3 (2)	99 (20)	79 (8)	182 (31)
	Unidentified gull	66 (31)	3,125 (63)	109 (26)	1 (1)	3,301 (121)

Table 6 (cont): Total count (unique number of transects) for all seabirds identified during the March 2012 survey.

Species Group	Species	Crew3316	Crew3651	Crew4056	Crew4446	Total
Larids	Parasitic jaeger	-	1 (1)	-	-	1 (1)
	Pomarine jaeger	-	1 (1)	-	-	1 (1)
	Black-legged kittiwake	-	-	1 (1)	303 (15)	304 (16)
	Caspian tern	1 (1)	-	-	-	1 (1)
	Forster's tern	1 (1)	-	-	-	1 (1)
	Least tern	2 (2)	-	-	-	2 (2)
	Royal tern	223 (47)	114 (32)	-	-	337 (79)
	Unidentified large tern	92 (35)	5 (4)	14 (9)	-	111 (48)
	Unidentified small tern	21 (11)	38 (16)	-	-	59 (27)
	Unidentified tern	177 (40)	96 (33)	51 (11)	-	324 (84)
Pelicaniforms	Northern gannet	634 (51)	5,156 (67)	2,433 (62)	287 (38)	8,510 (218)
	Double-crested cormorant	98 (11)	25 (10)	50 (7)	-	173 (28)
	Unidentified cormorant	-	3 (1)	45 (5)	6 (2)	54 (8)
	Brown pelican	192 (33)	49 (15)	63 (2)	-	304 (50)
	Unidentified albatross	-	1 (1)	-	-	1 (1)
	Northern fulmar	-	-	7 (2)	-	7 (2)
	Audubon's shearwater	1 (1)	-	-	-	1 (1)
	Cory's shearwater	2 (2)	5 (3)	-	-	7 (5)
	Great shearwater	-	3 (3)	6 (3)	-	9 (6)
	Manx shearwater	-	1 (1)	-	-	1 (1)
Tubenoses	Unidentified shearwater	1 (1)	1 (1)	2 (2)	-	4 (4)
	Unidentified storm-petrel	-	-	1 (1)	-	1 (1)
	Unidentified seabird or diving duck	29 (11)	47 (19)	41 (20)	24 (14)	141 (64)
	Unidentified phalarope	1,592 (17)	4,484 (22)	21 (5)	-	6,097 (44)
	Unidentified albatross	-	1 (1)	-	-	1 (1)
	Northern fulmar	-	-	7 (2)	-	7 (2)
	Audubon's shearwater	1 (1)	-	-	-	1 (1)
	Cory's shearwater	2 (2)	5 (3)	-	-	7 (5)
	Great shearwater	-	3 (3)	6 (3)	-	9 (6)
	Manx shearwater	-	1 (1)	-	-	1 (1)
	Unidentified shearwater	1 (1)	1 (1)	2 (2)	-	4 (4)
	Unidentified storm-petrel	-	-	1 (1)	-	1 (1)
	Unidentified seabird or diving duck	29 (11)	47 (19)	41 (20)	24 (14)	141 (64)
	Unidentified phalarope	1,592 (17)	4,484 (22)	21 (5)	-	6,097 (44)

Table 7: Total count (unique number of transects) for all seabirds identified during the October 2012 survey.

Species Group	Species	Crew3316	Crew3756	Crew4056	Crew4446	Total
Sea ducks	Long-tailed duck	-	-	-	5 (1)	5 (1)
	Common eider	-	-	-	322 (19)	322 (19)
	White-winged scoter	-	-	-	8 (2)	8 (2)
	Unidentified scoter	-	-	-	98 (10)	98 (10)
	Unidentified sea duck	-	-	-	12 (1)	12 (1)
Grebes	Unidentified grebe	-	-	-	1 (1)	1 (1)
Loons	Common loon	-	-	-	41 (22)	41 (22)
	Red-throated loon	-	-	12 (5)	1 (1)	13 (6)
	Unidentified loon	-	-	3 (2)	3 (2)	6 (4)
Alcids	Dovekie	-	-	-	1 (1)	1 (1)
	Black guillemot	-	-	-	9 (5)	9 (5)
	Unidentified alcid	-	-	-	20 (6)	20 (6)
Larids	Bonaparte's gull	143 (6)	164 (11)	-	-	307 (17)
	Glaucous gull	-	7 (1)	-	-	7 (1)
	Great black-backed gull	1 (1)	248 (31)	75 (21)	-	324 (53)
	Herring gull	685 (14)	183 (31)	758 (49)	1,014 (73)	2,640 (167)
	Laughing gull	688 (45)	18 (8)	14 (4)	1 (1)	721 (58)
	Lesser black-backed gull	-	-	84 (13)	-	84 (13)
	Little gull	-	-	2 (2)	-	2 (2)
	Ring-billed gull	5 (3)	12 (2)	76 (26)	133 (22)	226 (53)
	Unidentified black-backed gull	-	14 (9)	1 (1)	169 (33)	184 (43)
	Unidentified large gull	4 (2)	20 (1)	-	-	24 (3)
	Unidentified small gull	2 (1)	2 (2)	75 (25)	-	79 (28)
	Unidentified gull	293 (14)	1,806 (51)	504 (14)	19 (12)	2,622 (91)
	Black-legged kittiwake	-	-	-	153 (20)	153 (20)
	Caspian tern	2 (1)	-	-	-	2 (1)
	Least tern	590 (14)	-	-	-	590 (14)
	Little tern	12 (1)	-	-	-	12 (1)
	Roseate tern	-	1 (1)	-	-	1 (1)
	Royal tern	71 (20)	193 (46)	-	-	264 (66)
	Unidentified large tern	1,835 (53)	30 (11)	1 (1)	59 (5)	1,925 (70)
	Unidentified medium tern	5 (2)	-	-	-	5 (2)
	Unidentified small tern	268 (19)	37 (11)	1 (1)	3 (2)	309 (33)
	Unidentified tern	978 (35)	418 (26)	38 (11)	21 (4)	1,455 (76)

Table 7 (cont): Total count (unique number of transects) for all seabirds identified during the October 2012 survey.

Species Group	Species	Crew3316	Crew3756	Crew4056	Crew4446	Total
Pelicaniforms	Northern gannet	-	1 (1)	247 (50)	240 (41)	488 (92)
	Double-crested cormorant	-	5,757 (14)	-	303 (3)	6,060 (17)
	Unidentified cormorant	-	-	90 (7)	64 (9)	154 (16)
	Magnificent frigatebird	1 (1)	-	-	-	1 (1)
	American white pelican	-	70 (1)	-	-	70 (1)
	Brown pelican	108 (20)	1,000 (22)	11 (4)	-	1,119 (46)
Tubenoses	Audubon's shearwater	8 (5)	-	-	-	8 (5)
	Cory's shearwater	97 (21)	81 (11)	-	-	178 (32)
	Great shearwater	2 (2)	-	-	6 (4)	8 (6)
	Unidentified shearwater	17 (6)	1 (1)	6 (6)	24 (12)	48 (25)
	Unidentified storm-petrel	-	-	1 (1)	1 (1)	2 (2)
	Unidentified seabird or diving duck	41 (3)	9 (1)	38 (18)	5 (4)	93 (26)
	Unidentified phalarope	149 (20)	17 (2)	-	5 (2)	171 (24)

Table 8: Total count for all marine mammals and sea turtles identified.

Species Group	Species	August 2010	Dec 2010 & Jan 2011	August 2011	March 2012	October 2012
Marine mammals	Bottlenose dolphin	24	16	-	-	48
	Risso's dolphin	-	-	-	-	6
	Unidentified spotted dolphin	2	-	-	-	-
	Unidentified dolphin	145	31	626	336	182
	West indian manatee	3	-	-	-	-
	Unidentified porpoise	-	-	5	1	2
	Unidentified seal	-	-	-	11	7
	Common minke whale	-	-	1	-	-
	Fin whale	-	-	-	1	-
	Humpback whale	-	-	1	1	-
	Unidentified whale	-	-	6	3	2
	Unidentified marine mammal	1	-	-	-	-
Sea turtles	Green sea turtle	7	-	-	5	15
	Kemp's ridley sea turtle	2	-	2	1	1
	Leatherback sea turtle	3	-	22	7	6
	Loggerhead sea turtle	152	1	184	92	184
	Unidentified sea turtle	182	-	248	262	72

Figure 7: Transect density (total count per sq. km) for (A) alcids, (B) gulls, (C) loons, (D) northern gannets, (E) terns, (F) sea ducks and diving ducks, (G) marine mammals, and (H) sea turtles from the August 2010 survey. Transects are colored according to density: gray (zero density), light blue (0.01 to 1.00 count per sq. km), yellow (1.01 to 10.00 count per sq. km), orange (10.01 to 100.00 count per sq. km), red (>100 count per sq. km).

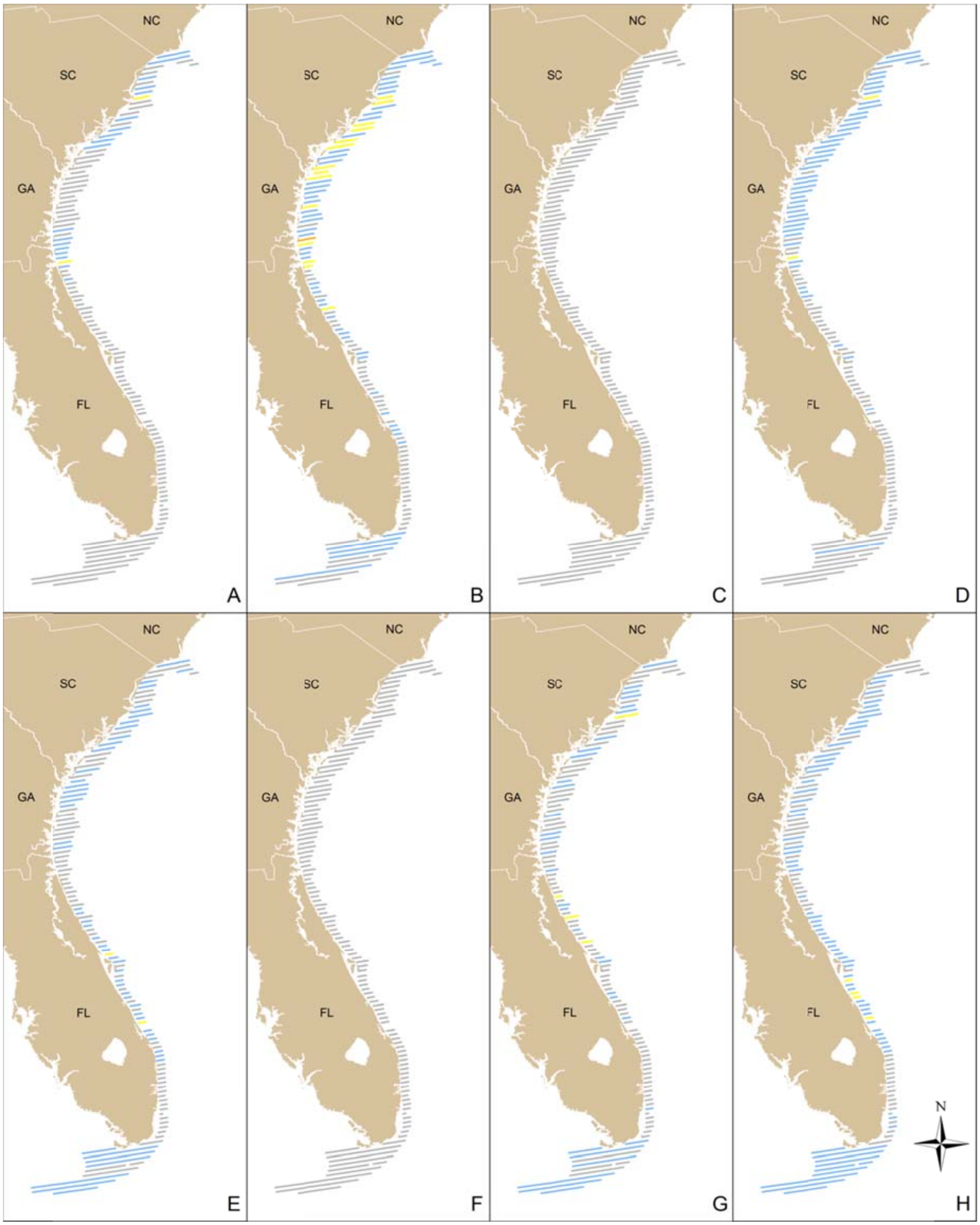


Figure 8: Transect density (total count per sq. km) for (A) alcids, (B) gulls, (C) loons, (D) northern gannets, (E) terns, (F) sea ducks and diving ducks, (G) marine mammals, and (H) sea turtles from the December 2010 and January 2011 surveys. Transects are colored according to density: gray (zero density), light blue (0.01 to 1.00 count per sq. km), yellow (1.01 to 10.00 count per sq. km), orange (10.01 to 100.00 count per sq. km), red (>100 count per sq. km).

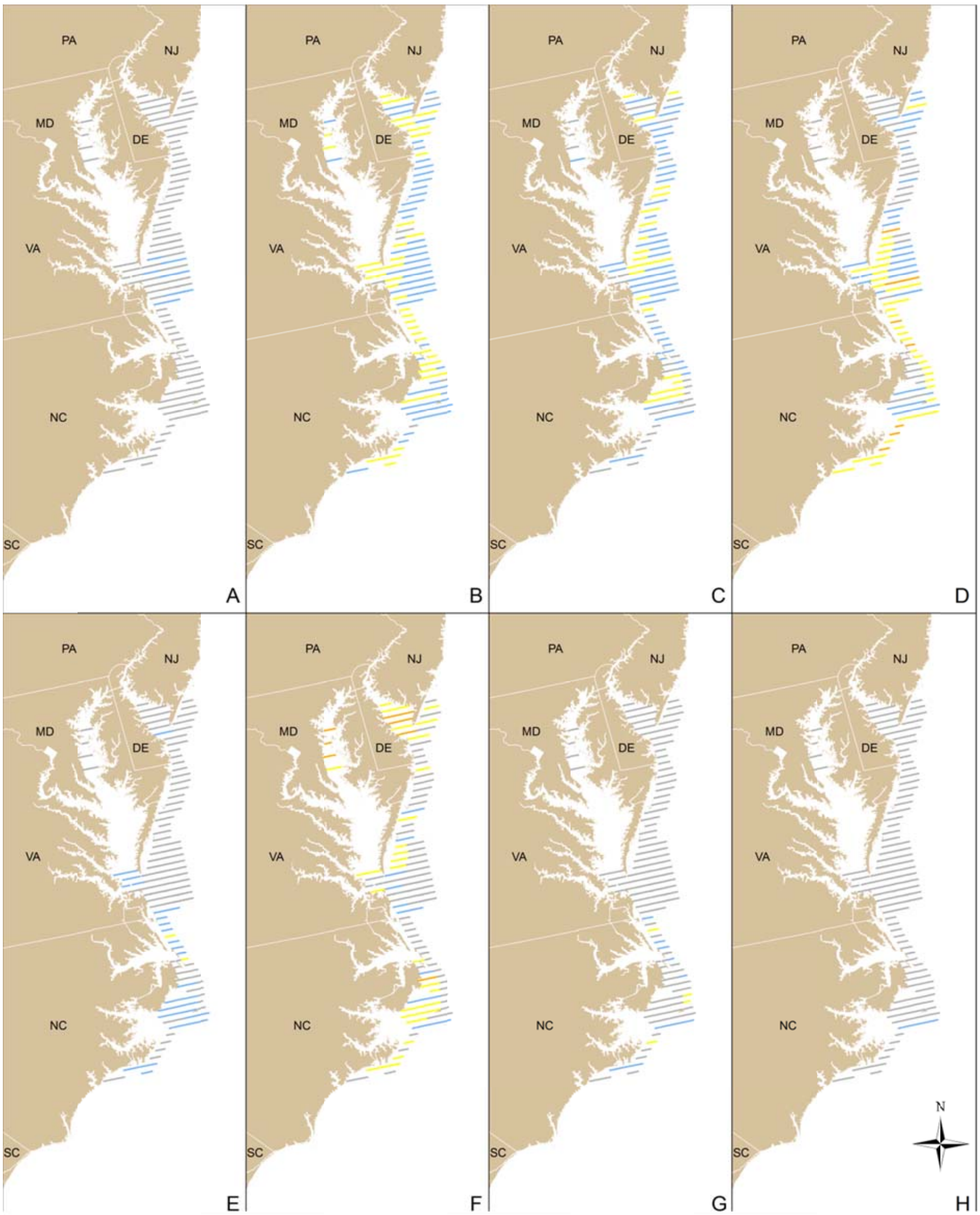


Figure 9: Transect density (total count per sq. km) for (A) alcids, (B) gulls, (C) loons, (D) northern gannets, (E) terns, (F) sea ducks and diving ducks, (G) marine mammals, and (H) sea turtles from the northern region of the August 2011 survey. Transects are colored according to density: gray (zero density), light blue (0.01 to 1.00 count per sq. km), yellow (1.01 to 10.00 count per sq. km), orange (10.01 to 100.00 count per sq. km), red (>100 count per sq. km).

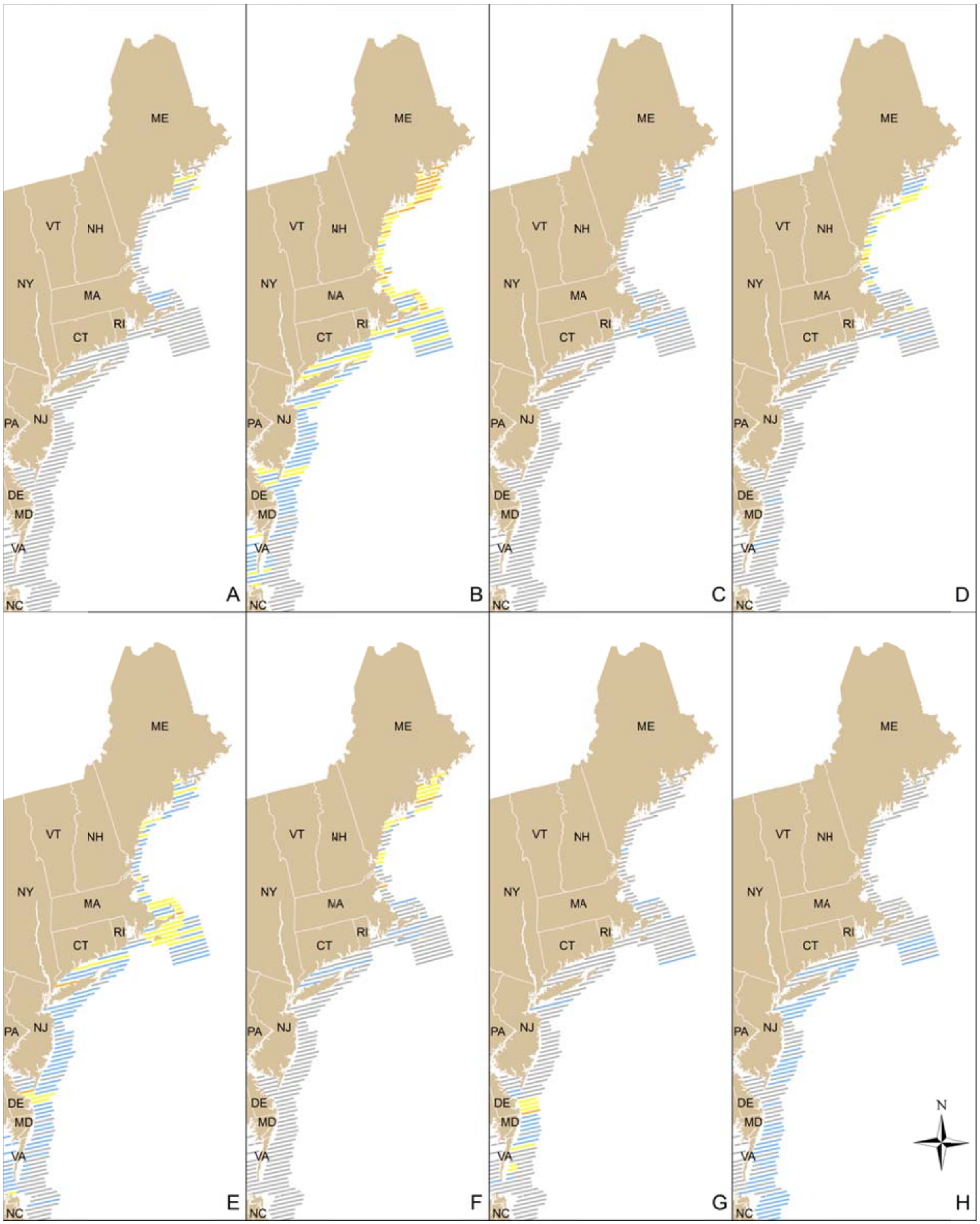


Figure 10: Transect density (total count per sq. km) for (A) alcids, (B) gulls, (C) loons, (D) northern gannets, (E) terns, (F) sea ducks and diving ducks, (G) marine mammals, and (H) sea turtles from the mid-Atlantic region of the August 2011 survey. Transects are colored according to density: gray (zero density), light blue (0.01 to 1.00 count per sq. km), yellow (1.01 to 10.00 count per sq. km), orange (10.01 to 100.00 count per sq. km), red (>100 count per sq. km).

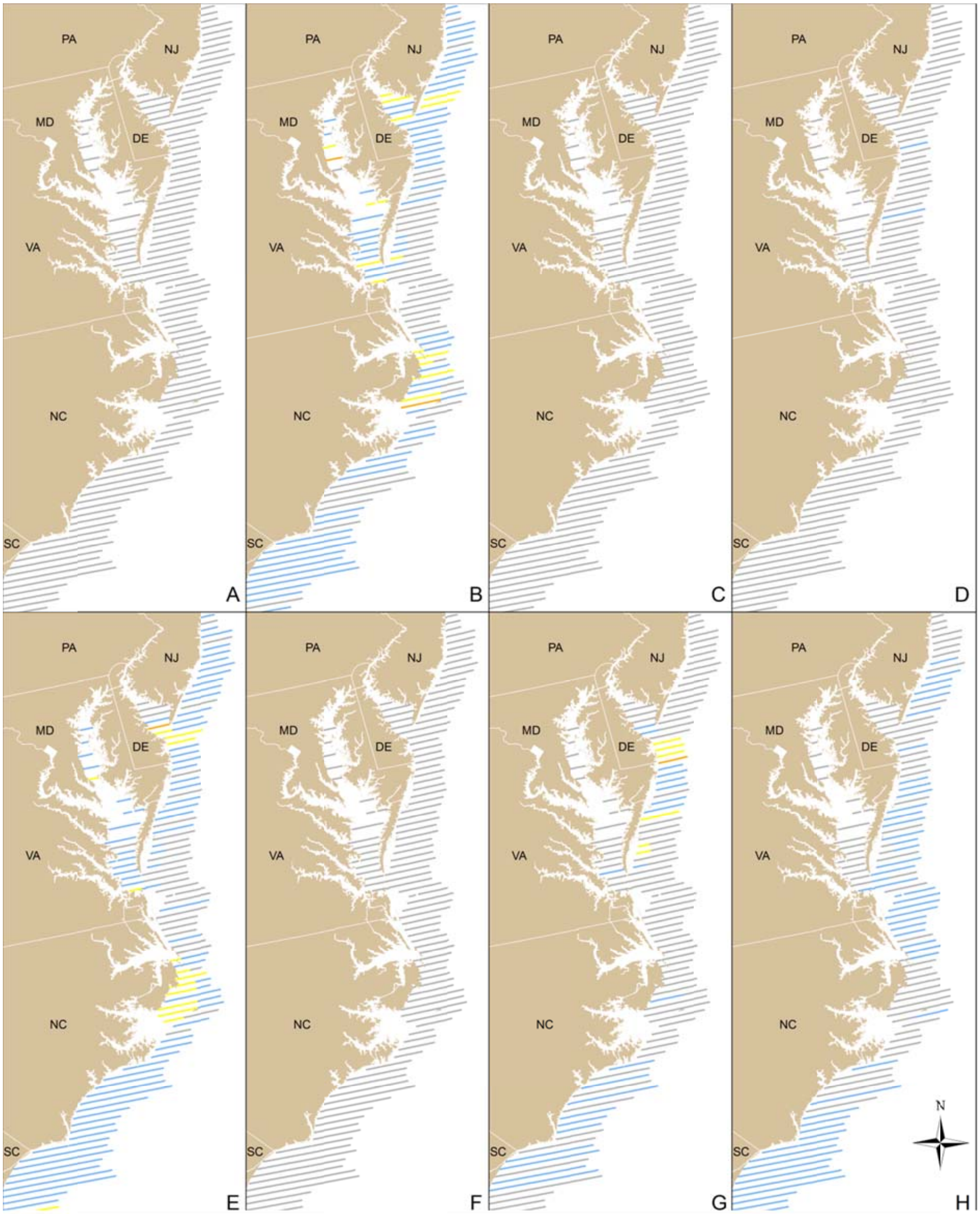


Figure 11: Transect density (total count per sq. km) for (A) alcids, (B) gulls, (C) loons, (D) northern gannets, (E) terns, (F) sea ducks and diving ducks, (G) marine mammals, and (H) sea turtles from the southern region of the August 2011 survey. Transects are colored according to density: gray (zero density), light blue (0.01 to 1.00 count per sq. km), yellow (1.01 to 10.00 count per sq. km), orange (10.01 to 100.00 count per sq. km), red (>100 count per sq. km).

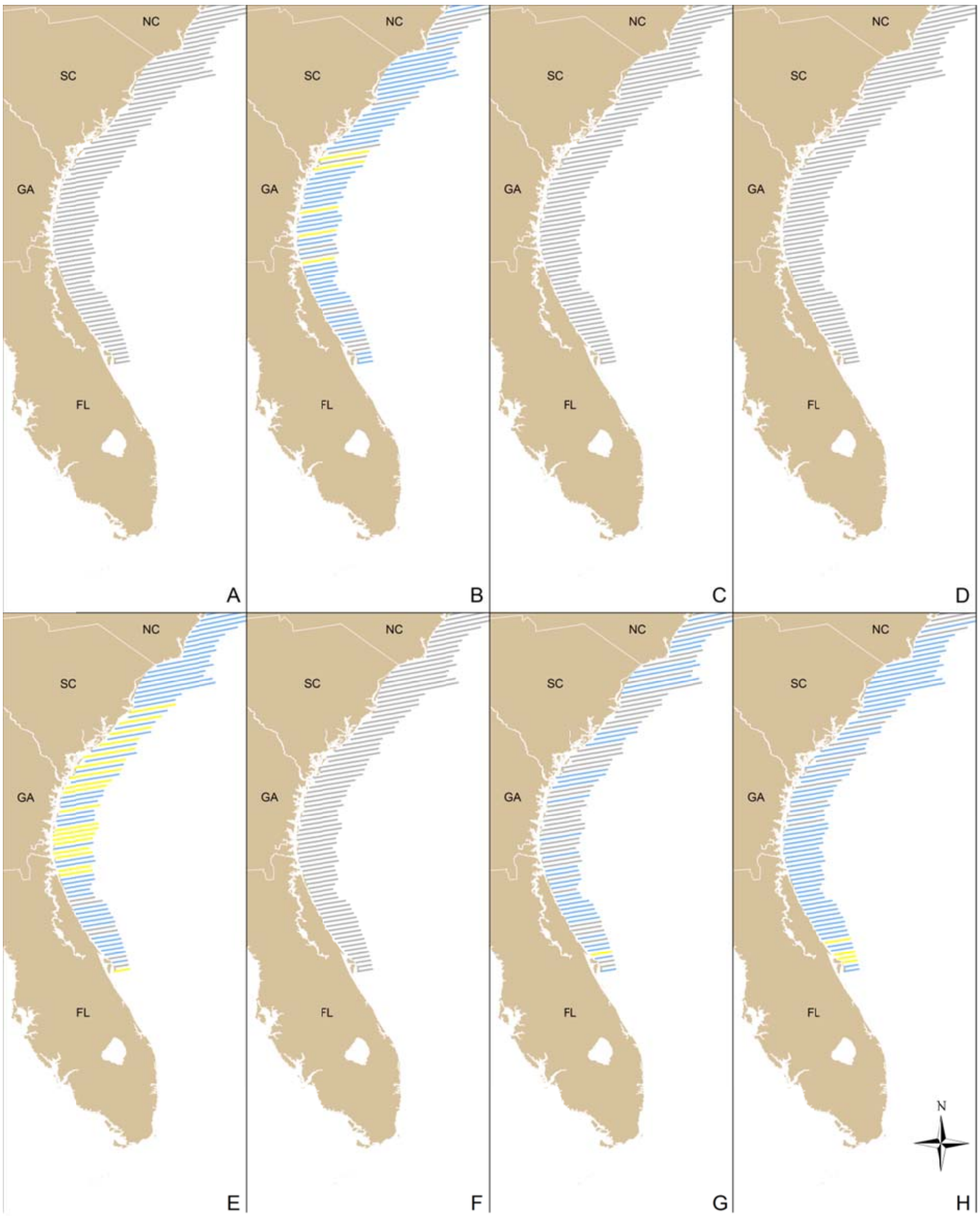


Figure 12: Transect density (total count per sq. km) for (A) alcids, (B) gulls, (C) loons, (D) northern gannets, (E) terns, (F) sea ducks and diving ducks, (G) marine mammals, and (H) sea turtles from the northern region of the March 2012 survey. Transects are colored according to density: gray (zero density), light blue (0.01 to 1.00 count per sq. km), yellow (1.01 to 10.00 count per sq. km), orange (10.01 to 100.00 count per sq. km), red (>100 count per sq. km).

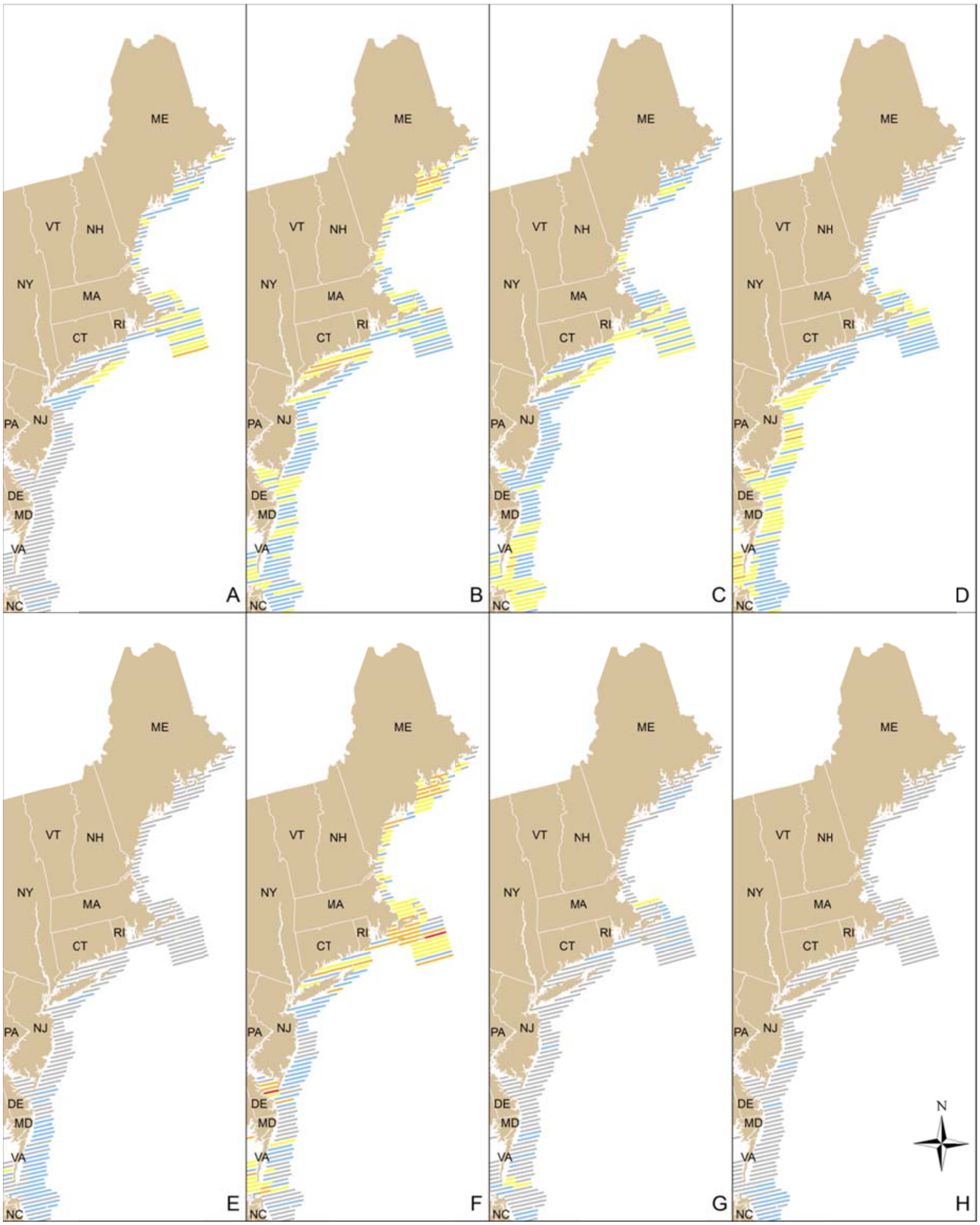


Figure 13: Transect density (total count per sq. km) for (A) alcids, (B) gulls, (C) loons, (D) northern gannets, (E) terns, (F) sea ducks and diving ducks, (G) marine mammals, and (H) sea turtles from the mid-Atlantic region of the March 2012 survey. Transects are colored according to density: gray (zero density), light blue (0.01 to 1.00 count per sq. km), yellow (1.01 to 10.00 count per sq. km), orange (10.01 to 100.00 count per sq. km), red (>100 count per sq. km).

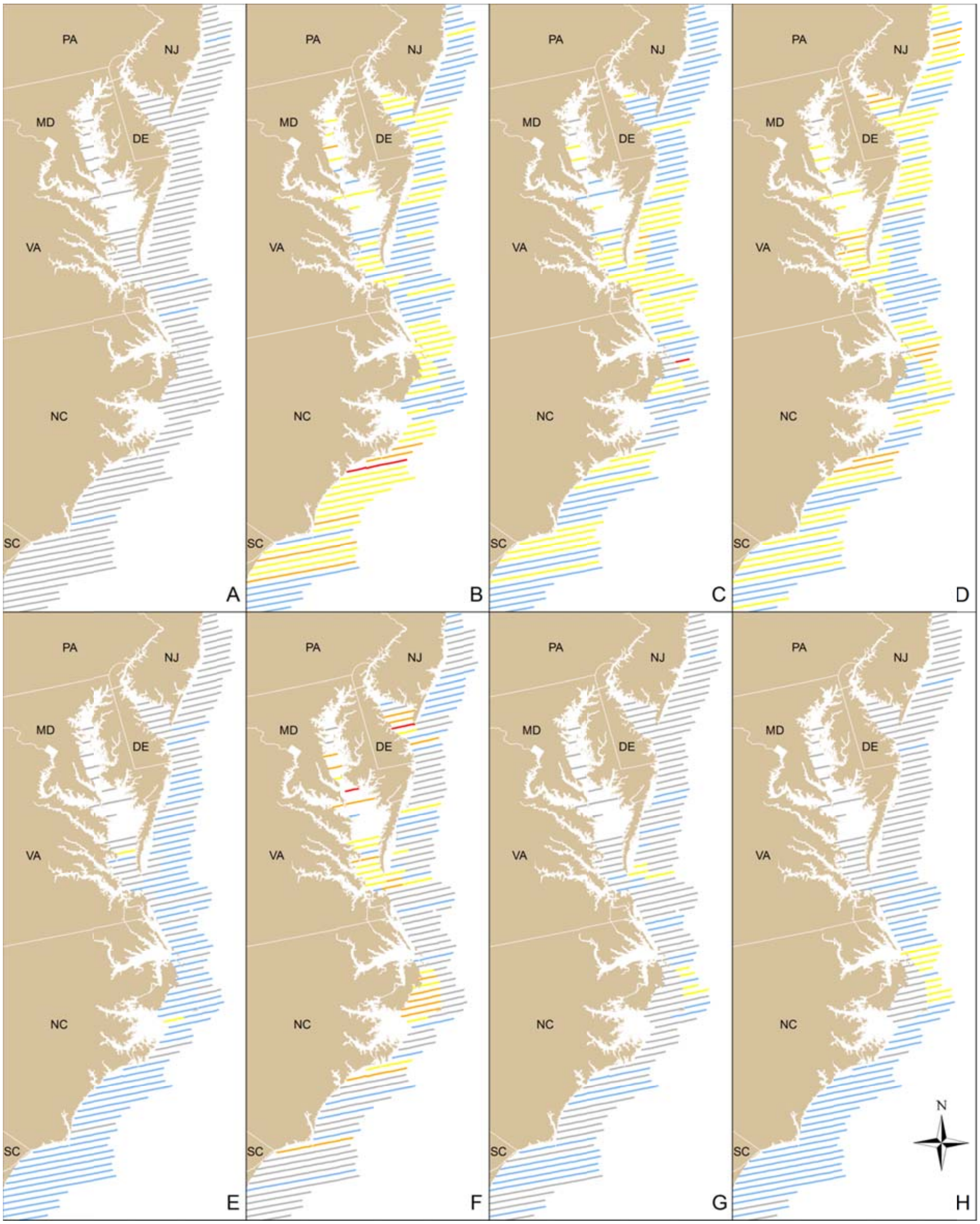


Figure 14: Transect density (total count per sq. km) for (A) alcids, (B) gulls, (C) loons, (D) northern gannets, (E) terns, (F) sea ducks and diving ducks, (G) marine mammals, and (H) sea turtles from the southern region of the March 2012 survey. Transects are colored according to density: gray (zero density), light blue (0.01 to 1.00 count per sq. km), yellow (1.01 to 10.00 count per sq. km), orange (10.01 to 100.00 count per sq. km), red (>100 count per sq. km).

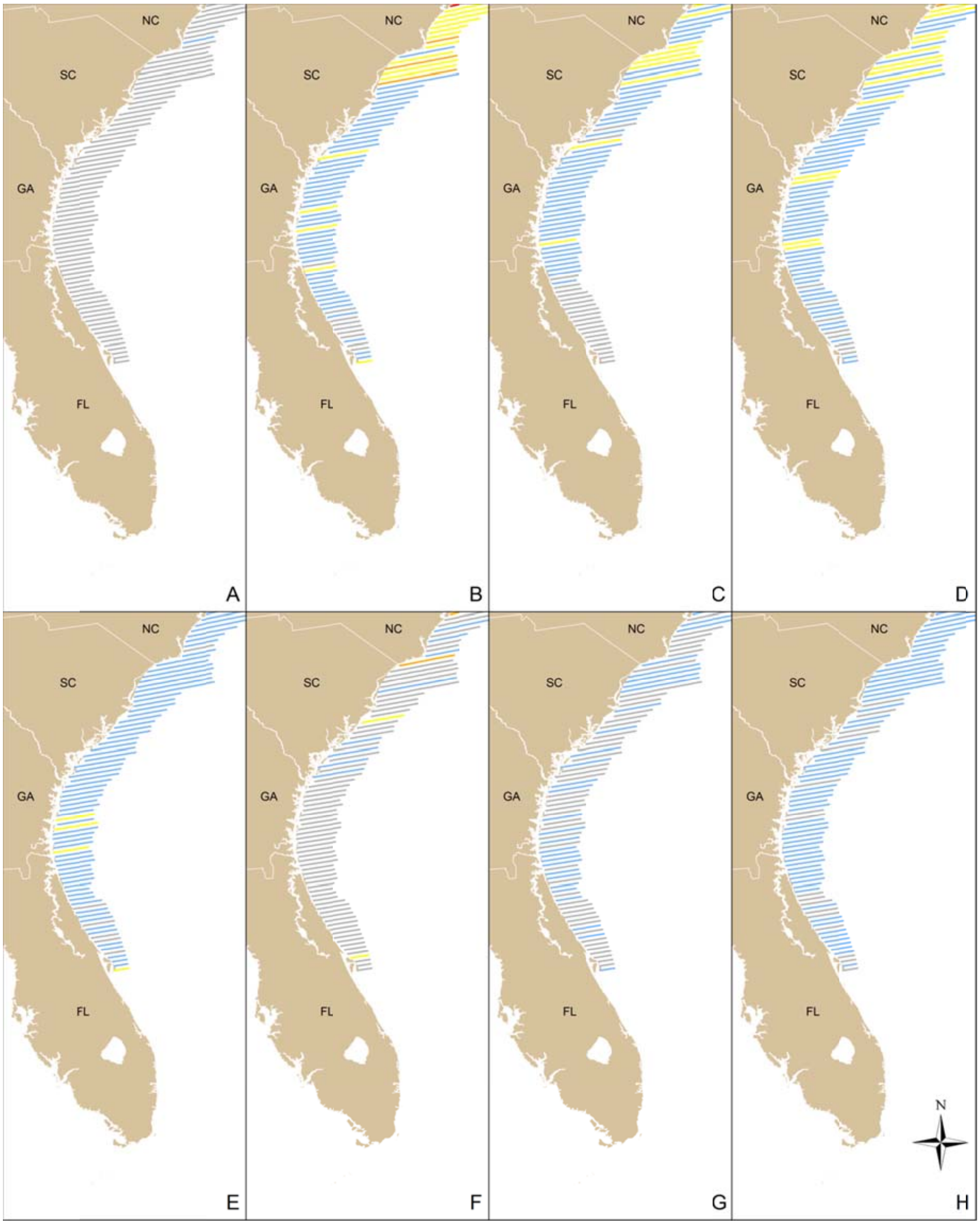


Figure 15: Transect density (total count per sq. km) for (A) alcids, (B) gulls, (C) loons, (D) northern gannets, (E) terns, (F) sea ducks and diving ducks, (G) marine mammals, and (H) sea turtles from the northern region of the October 2012 survey. Transects are colored according to density: gray (zero density), light blue (0.01 to 1.00 count per sq. km), yellow (1.01 to 10.00 count per sq. km), orange (10.01 to 100.00 count per sq. km), red (>100 count per sq. km).

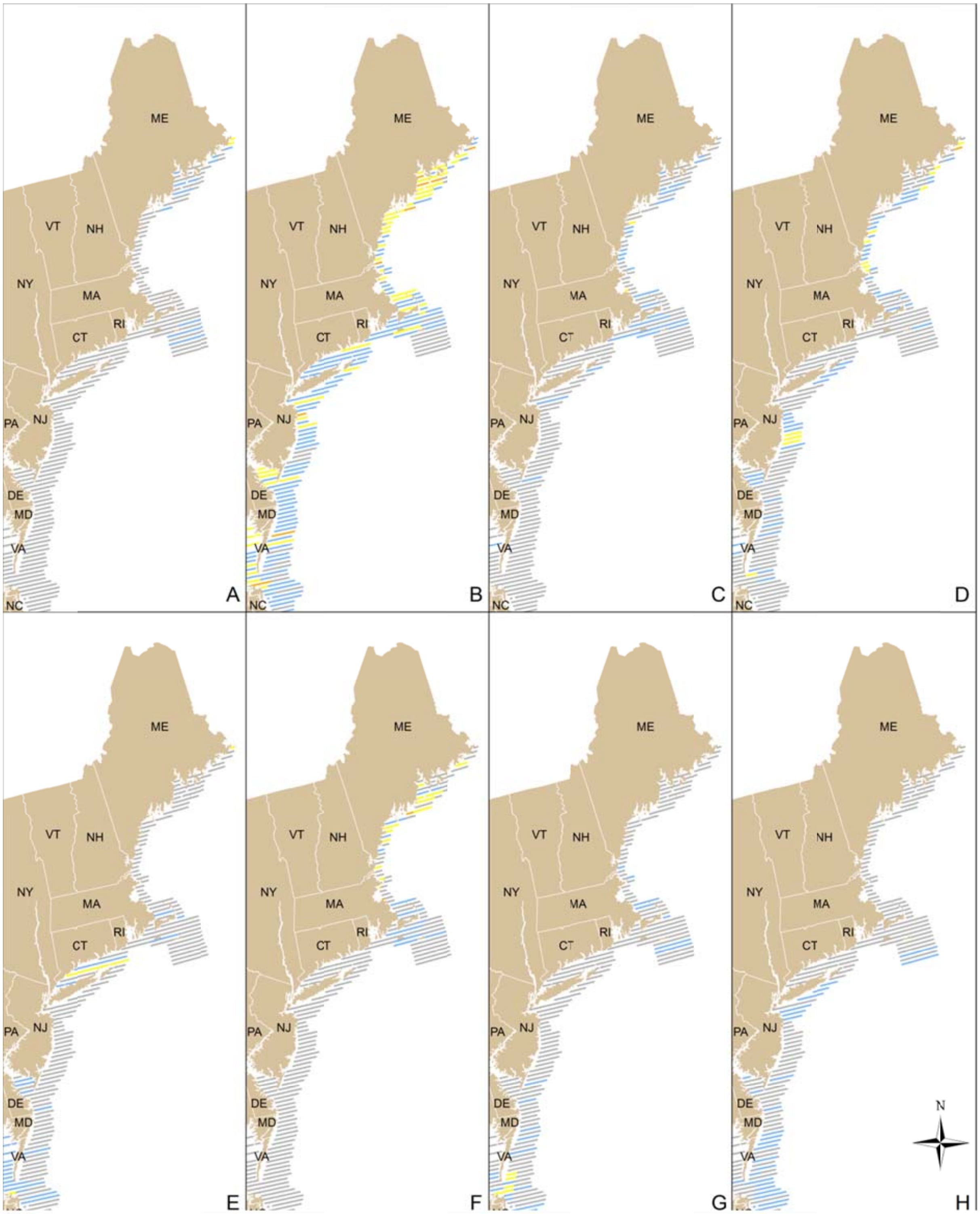


Figure 16: Transect density (total count per sq. km) for (A) alcids, (B) gulls, (C) loons, (D) northern gannets, (E) terns, (F) sea ducks and diving ducks, (G) marine mammals, and (H) sea turtles from the mid-Atlantic region of the October 2012 survey. Transects are colored according to density: gray (zero density), light blue (0.01 to 1.00 count per sq. km), yellow (1.01 to 10.00 count per sq. km), orange (10.01 to 100.00 count per sq. km), red (>100 count per sq. km).

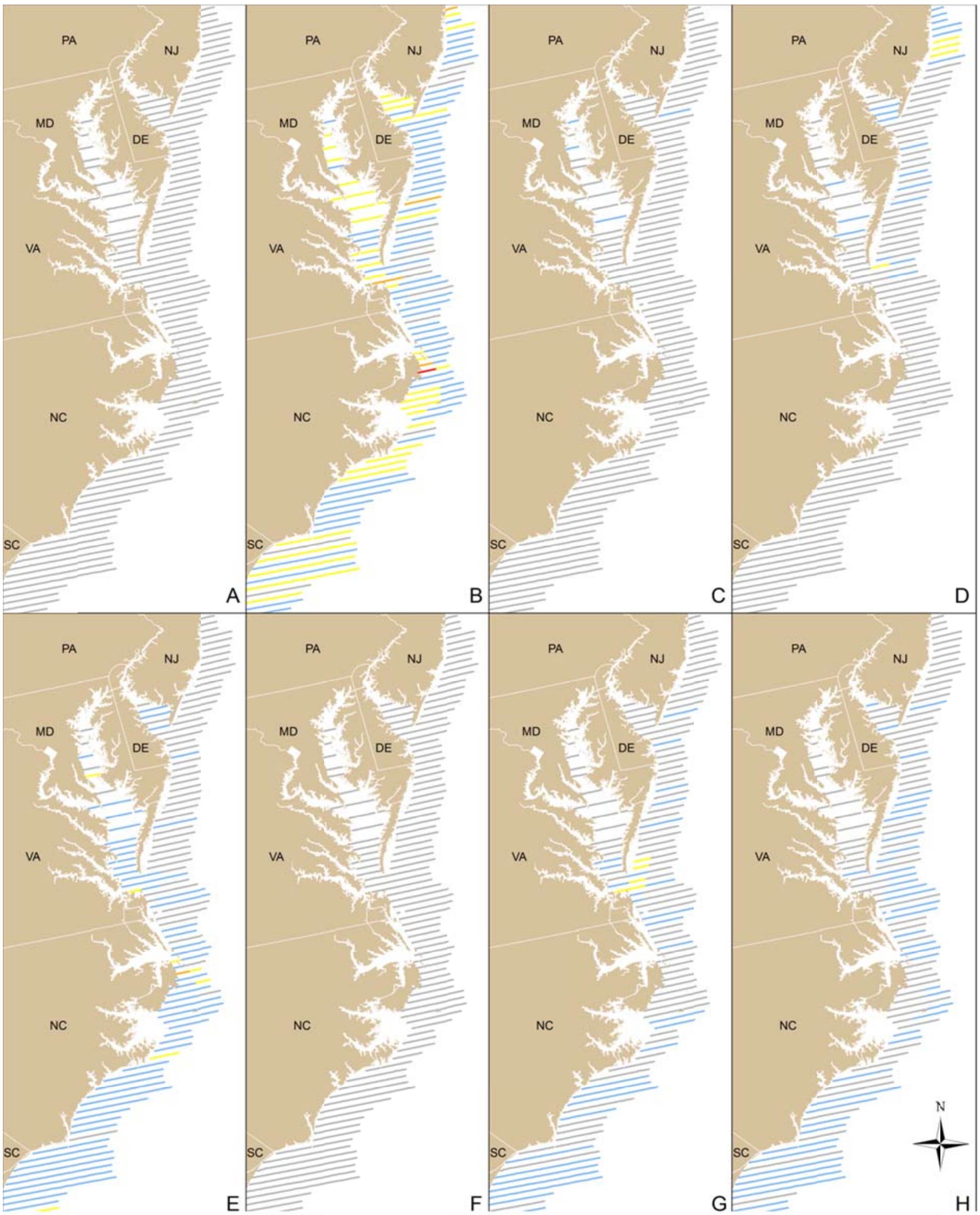
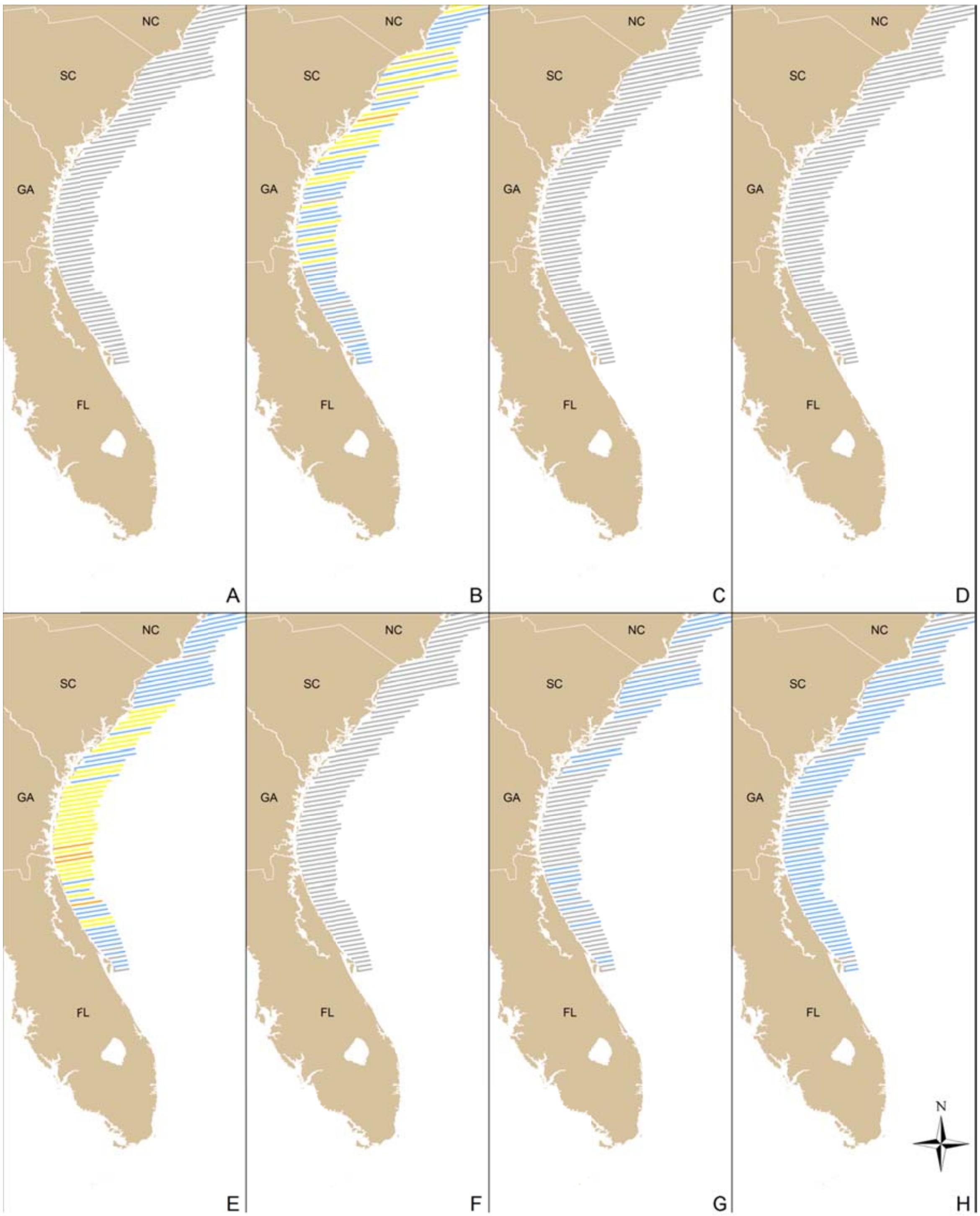


Figure 17: Transect density (total count per sq. km) for (A) alcids, (B) gulls, (C) loons, (D) northern gannets, (E) terns, (F) sea ducks and diving ducks, (G) marine mammals, and (H) sea turtles from the southern region of the October 2012 survey. Transects are colored according to density: gray (zero density), light blue (0.01 to 1.00 count per sq. km), yellow (1.01 to 10.00 count per sq. km), orange (10.01 to 100.00 count per sq. km), red (>100 count per sq. km).



Appendices

Appendix 1. Species Codes used in AMAPPS database

Species codes:

BLSC = Black scoter
SUSC = Surf scoter
WWSC = White-winged scoter
DWSC = Dark-winged scoter (i.e., unidentified BL/SUSC)
SCOT = unidentified scoter
LTDU = Long-tailed duck
COEI = Common eider
KIEI = King eider
EIDE = unidentified eider
COME = Common merganser
RBME = Red-breasted merganser
HOME = Hooded merganser
MERG = unidentified merganser
BAGO = Barrow's goldenye
COGO = Common goldeneye
GOLD = unidentified goldeneye
GOME = unidentified goldeneye/merganser
BUFF = Bufflehead
HARD = Harlequin duck
CANV = Canvasback
REDH = Redhead
RNDU = Ring-necked duck
SCAU = Scaup spp.
GRSC = Greater scaup
LESC = Lesser scaup
DUCK = unidentified sea duck
HOGR = Horned grebe
RNGR = Red-necked grebe
UNGR = unidentified grebe
COLO = Common loon
RTLO = Red-throated loon
LOON = unidentified loon
ATPU = Atlantic puffin
BLGU = Black guillemot
COMU = Common murre
DOVE = Dovekie
RAZO = Razorbill
TBMU = Thick-billed murre
UNMU = unidentified murre
UNLA = unidentified large alcid
ALCD = unidentified alcid
BBGU = Black-backed gull
BLKI = Black-legged kittiwake
BOGU = Bonaparte's gull
GBBG = Greater black-backed gull
GLGU = Glaucous gull
HERG = Herring gull
ICGU = Iceland gull
LAGU = Laughing gull
LBBG = Lesser black-backed gull
LIGU = Little gull
RBGU = Ring-billed gull
UNLG = Large gull
UNSG= Small gull
GULL = unidentified gull
UNLT = unidentified large tern (e.g., Caspian, Royal, Roseate)
UNMT = unidentified medium tern (e.g., Forster's, Gull-billed, etc.)
UNST = unidentified small tern (e.g., Least, Arctic, Common)
UNTE = unidentified tern
ARTE = Arctic Tern
BRTE = Bridled Tern
COTE = Common Tern
FOTE = Forster's Tern
GBTE = Gull-billed Tern
LETE = Least Tern
ROST = Roseate Tern
ROYT = Royal Tern
SOTE = Sooty Tern
BLTE = Black Tern
CATE = Caspian Tern
BRNO = Brown Noddy
BLSK = Black skimmer
NOFU = Northern fulmar
AUSH = Audubon's shearwater
BCPE = Black-capped petrel
COSH = Cory's shearwater
GRSH = Greater shearwater
SOSH = Sooty shearwater
MASH = Manx shearwater
UNSH = unidentified shearwater
UNSP = unidentified storm-petrel
LHSP = Leach's Storm-petrel

WISP = Wilson's Storm-petrel
BSTP = Band-rumped Storm-petrel
NOGA = Northern gannet
DCCO = Double-crested cormorant
GRCO = Great cormorant
UNCO = unidentified cormorant
BRPE = Brown pelican
AWPE = American white pelican
MAFR = Magnificent frigatebird
RBTR = Red-billed Tropicbird
WTTR = White-tailed Tropicbird
BIRD = unidentified seabird or diving duck

Other species recorded:

Sharks and Rays:
GWSH = Great white shark
SHAR = unidentified shark
MARA = Manta ray
UNRA = unidentified ray

Sea Turtles:
GRST = Green sea turtle
LEST = Leatherback sea turtle
LOST = Loggerhead sea turtle
KRST = Kemp's ridley sea turtle
UIST = unidentified sea turtle

Marine Mammals:
BODO = Bottlenose dolphin
UNSD = unidentified spotted dolphin
DOLP = unidentified dolphin
PORP = unidentified porpoise
HUWH = Humpback whale
PIWH = Pilot whale
RIWH = Right whale
WHAL = unidentified whale
GRSE = Gray seal
SEAL = unidentified seal
WIMA = West Indian manatee
UNMM = unidentified marine mammal

Appendix 2. Database Field Glossary

Microsoft Access Database - Atlantic_Coast_Surveys

ACWSD	indicator for whether or not transect was surveyed as part of the Atlantic Coast Winter Sea Duck Survey
ACWSDreport	indicator for whether or not transect was included in 2009 - 2011 Atlantic Coast Winter Sea Duck Survey report analysis
AvgCondition	distance-weighted average observation condition
Band	survey band in which bird was located (perpendicular to flight path): 0 = unknown or not recorded 1 = less than 100 meters from plane 2 = 100 to 200 meters from plane
CommonName	species common name
Condition	observation condition (measured on a 5-point Likert scale: 1 = poor and 5 = excellent)
Crew	crew name (typically designated by the four digit latitude of their northern-most transect)
Day	day the transect was surveyed
Depth	water depth for each observation (units = meters); negative values are meters below sea level (e.g., -1 means water depth for this observation was 1 meter below sea level)
Dist2Coast_m	distance each observation is from the coast (units = meters)
Dist2Coast_nm	distance each observation is from the coast (units = nautical miles)
DistFlown	distance surveyed on a transect by an observer (units = nautical miles)
EndDt	date the transect survey ended
FlockSize	number of individuals observed at a given location
GpsError	error associated with geographic coordinates recorded during surveys (value of -1 indicates that latitude, longitude, or seconds value was interpolated based on surrounding data points)

ImputedDistFlown	indicator for whether or not distance flown was imputed (due to unknown transect BEG/END points) by using crew member's distance flown value
Lat	latitude in decimal degrees (GCS = WGS84)
LatinName	species Latin (scientific) name
Long	longitude in decimal degrees (GCS = WGS84)
MissingTrackFile	indicator for whether or not track file from observer was missing
Month	month the transect was surveyed
Obs	observer initials
ObsInitials	initials of non-pilot observer(s)
ObsName	name of non-pilot observer(s)
PilInitials	initials of pilot(s)
PilName	name of pilot(s)
Replicate	transect replicate number for a particular survey (1 = first time transect was flown, 2 = second time transect was flown, etc.)
Seat	observer seat in plane: lf = left front (i.e., pilot) rf = right front lr = left rear rr = right rear
Sec	time in seconds from midnight as recorded by the computers' internal clock (specific to each observer) NOTE: observers were asked to set computer clocks to local time, but this was not always done; therefore, this value should not be used as a proxy for time of day
Slope	steepness of the ocean bottom based on changes in water depths (units = degrees)
Species	four letter code used to identify observations during survey (AOU band code was used when possible; see Species_Information table for details)

StartDt	date the transect survey started
SurveyDescription	brief description of survey
SurveyNbr	unique survey ID: 1 = 2008 Preliminary ACWSD 2 = 2009 ACWSD 3 = 2010 ACWSD 4 = 2010 Preliminary AMAPPS 5 = December 2010 wind area additional flying 6 = January 2011 wind area additional flying 7 = 2011 ACWSD 8 = 2011 Summer AMAPPS 9 = 2012 Southern BLSC Survey 10 = 2012 Mid-Atlantic Detection Survey 11 = 2012 Spring AMAPPS 12 = 2012 Fall AMAPPS
SurveyEndDt	date the survey ended
SurveyStartDt	date the survey started
Transect	unique ID for each survey line; the first four digits represent latitude in degrees decimal minutes and the last two digits indicate segment number
Type	type of GPS track point: BEGTRAN = beginning of transect ENDTRAN = end of transect BEGCNT = start counting again ENTCNT = stop counting while on transect COCH = location where observation condition changed along transect WAYPNT = GPS point along transect
WindArea	indicator for whether or not transect covers proposed BOEM offshore wind development area off Chesapeake Bay

Year	year the transect was surveyed

ESRI ArcMap Geodatabase - Atlantic_Coast_Surveys

Observations	Point shapefile containing the location of seabird and sea duck flocks along the Atlantic Coast and the habitat covariates associated with each flock. Fields are the same as the Observations table located in the Atlantic_Coast_Surveys Access database.
Tracks	Point shapefile containing the location of each track point along a given transect. Fields are the same as the Tracks table located in the Atlantic_Coast_Surveys Access database.
Transect_Information	Polyline shapefile containing all transects surveyed during the 2008 - 2012 Atlantic Coast surveys. Fields are the same as the Transect_Information table located in the Atlantic_Coast_Surveys Access database.

Appendix 3. Fitting statistical distributions to sea duck count data: Implications for survey design and abundance estimation

Zipkin, E.F. J.B. Leirness, B.P. Kinlan, A.F. O'Connell and E.D. Silverman. 2012. Fitting statistical distributions to sea duck count data: Implications for survey design and abundance estimation. Statistical Methodology

ARTICLE IN PRESS

Statistical Methodology ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at SciVerse ScienceDirect

Statistical Methodology

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

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

A R T I C L E I N F O

Article history:
Received 8 June 2012
Received in revised form
24 September 2012
Accepted 8 October 2012

Keywords:
Common eiders
Discretized lognormal
Group size modeling
Long-tailed ducks
Marked point process
Model selection
Negative binomial
Scoters

A B S T R A C T

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).

1572-3127/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.stamet.2012.10.002

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

Please cite this article in press as: E.F. Zipkin, et al., Fitting statistical distributions to sea duck count data: Implications for survey design and abundance estimation, Statistical Methodology (2012), doi:10.1016/j.stamet.2012.10.002

or include habitat covariates) and/or data collection efforts are necessary. Identifying a parsimonious model is of primary importance because monitoring programs require repeated, timely estimates that are easy to explain and robust to unexpected data reduction or other survey changes. Thus, analytically complex and data-hungry approaches are ill-advised for management-oriented monitoring programs.

The most challenging problem we face is characterizing a count distribution with an extreme variance to mean ratio, as is often observed in sea duck data [52]. Identifying appropriate statistical distributions for analyzing count data of animal populations is an ongoing area of investigation in ecology. For reasons based on first principles and for convenience, the Poisson distribution has frequently been used [8] and is popular in modeling avian species (e.g., [14,28]). Yet the assumption that the variance equals the mean often does not hold for many seabird species, which are known to form large flocks. The negative binomial distribution, which allows the variance to exceed the mean, is used as an alternative to the Poisson to characterize the count distributions for species where spatial aggregation is known to occur (e.g., [2,11,49]). The negative binomial distribution is the result of a Poisson–Gamma mixture and converges to the Poisson distribution as the shape parameter, k, approaches infinity (Appendix A). Okubo [34] recommended the geometric distribution – a discrete analog to the exponential distribution and also a special case of the negative binomial where the shape parameter equals one – to handle extremely large group sizes and demonstrated its applicability for a number of taxa including birds. Empirical evidence suggests, however, that the negative binomial and geometric models do not adequately capture observed distributions of counts for some populations, especially those that are found in very large group sizes, such as some fish and bird species. Ma et al. [29] derived a logarithmic distribution from first principles based on rules for when individuals should join and leave groups; this model has outperformed the Poisson and negative binomial distributions in studies of house sparrows [17] and seabirds [21]. Ma et al. [29] additionally pointed out that the logarithmic can be derived as a limiting case of the negative binomial distribution as the shape parameter (k, Appendix A) approaches zero (see also [39]), placing it in the context of other distributions used to model ecological count data.

More recently, the power law distribution has been proposed for modeling group sizes when the variance to mean ratio is much larger than can be accommodated by the aforementioned models [3,4]. Several studies have demonstrated that the power law distribution fits well to a number of empirical examples including populations of fish, seabirds, and mammals [10,2,22,23,45]. However, the power law distribution (using ecologically relevant parameter ranges) is capable of producing extremely large counts (e.g., in the millions; [10]), which are not realistic for most sea duck species. The power law can be truncated or combined with an exponentially decaying function [33] to address this problem. In fact, Ma et al. [29] pointed out that the logarithmic distribution itself is a discrete form of a power law distribution with an exponential cutoff, where the power law exponent is -1 and the upper tail decays exponentially above a cutoff that is directly related to the average group size experienced by an individual. Bonabeau et al. [4] also presents mechanistic models of group size that lead to power law distributions with exponential decay.

Other heavy-tailed distributions exist and should be considered in a model selection context before concluding that "power law-like" behavior observed in empirical data necessarily indicates a power law distribution [10]. These include the Yule–Simon and the discretized lognormal distributions, which themselves can be viewed, respectively, as limiting distributions of stochastic preferential attachment or multiplicative growth processes [10,31]. Given the diversity of possibilities, a model selection framework would be useful to guide choices of appropriate distributions to model highly skewed ecological count data [2].

In this paper, we test the fit of a series of over-dispersed statistical distributions, from the negative binomial to the power law, to counts of sea duck flock sizes; we also assess the fit of a series of over-dispersed models to the distribution of flock frequencies. Our assessment is a critical first step in the applied statistical work needed for the development of rigorous survey designs, power analysis, risk and impact assessments, and optimal management strategies for sea ducks. Appropriate modeling of the basic underlying distributional characteristics of avian count data is critical for making strong inferences about the distribution of target populations, particularly in the marine environment where logistics are inherently more difficult than in terrestrial systems and reliance upon statistical models is correspondingly greater.

Please cite this article in press as: E.F. Zipkin, et al., Fitting statistical distributions to sea duck count data: Implications for survey design and abundance estimation, Statistical Methodology (2012), doi:10.1016/j.stamet.2012.10.002

2. Methods

2.1. Data collection

The USFWS aerial survey was conducted along the Atlantic coast from the US-Canadian border (44° 46′N) to Jacksonville, FL (30°21′N) between January and March, 2009–2011. Four fixed-wing aircraft were flown along east–west transects spaced systematically at intervals of five minutes of latitude (approximately 5 nm apart). These transects extended east from the coastline to the longer of two distances: 8 nm or the distance to 16 m depth. Transects ranged in length from 1 to 80 nm (with 95% of transects between 4.8 and 46.4 nm). The mean transect length was 17.9 nm (standard deviation: 12.8 nm) with transects less than 8 nm in areas that span bays and longer transects paralleling the shoreline in complicated coastal areas (e.g., Long Island Sound).

The survey crews, which consisted of an observer and pilot-observer, flew at 110 knots and 70 m altitude, while counting sea ducks and other aquatic birds within 400 m-width strip transects (the observer counts a 200 m strip on one side of the plane while the pilot does the same on the opposite side). After completing their entire set of transect lines, each crew flew north to their first east–west transect line and replicated every other transect from north to south. The replicate surveys were conducted approximately one week after the first surveys and do not duplicate the original track exactly, making the possibility of recounting the same individuals remote. The three scoter species are difficult to distinguish reliably in the field, leading to a large number of scoters identified only to genus (*Melanitta* spp.). As such, we focused our analyses on generic scoter species (records for all three species combined with unidentified scoters), along with the Common eider and Long-tailed duck. We refer to these two species and one genus as the "species groups" of interest.

Surveys were conducted from 1 to 18 February in 2009, 23 January to 2 March in 2010, and 31 January to 17 February in 2011. Due to the vagaries of field operations, transects and replicates varied somewhat between years. We use data from the 236 transects, and 76 replicates that were successfully surveyed in all three years. Common eider and Long-tailed ducks do not winter in the southern portions of the survey area, and so models fit for them are based on fewer transects (88 for Common eiders, of which 21 were replicated; 173 for Long-tailed ducks, of which 54 were replicated).

The data consist of observations along survey transects recording the (1) location, (2) species, and (3) number of birds seen at the location. We refer to the group of birds recorded at one location (including single birds) as a "flock", and the number of birds seen as the "flock size". Note that birds are counted only within the transect boundaries, while the actual flock might have extended well beyond.

2.2. Analysis

To estimate the abundance of sea ducks by species, we represent the data as a modified marked point process [12,20] where the flocks are the points and the size of the flocks, discrete and independent of the points, are the marks. The point process is summarized by transect: we first model the flock counts (i.e., number of flocks) on each transect, and then model the flock sizes, conditional on the number of flocks observed. Preliminary analyses indicated large variations and only small correlations in the number of species-specific flocks (points) among neighboring transects (0.23 for Common eiders, 0.41 for Long-tailed ducks, and 0.24 for scoters), due in part to zero-zero neighbors in areas of low density. This suggests that the number of flocks on one transect is not predictive of the flock count on neighboring transects. We additionally found no significant relationships between the number/density of flocks per transect and the sizes of those flocks, which fits our assumption of independence in marks and points.

To determine the appropriate model to describe the observed number of flocks per transect (the point process), we tested the fit of four distributions to the transect-level flock counts: zero-inflated Poisson, zero-inflated geometric, and zero-inflated negative binomial, as well as the standard negative binomial (Appendix A). The data were fit separately for Common eiders, Long-tailed ducks, and scoter species and we included an offset for transect area (to account for variable transect lengths), which was standardized by dividing the area of each transect by the mean of all transect areas. We fit each model using maximum likelihood estimation (MLE) in the program R (version 2.13.2; R development Core [40]) with the VGAM package [50].

For the flock size data (the marks), we fit seven discrete distributions with positive integer support (because there are no flocks of size zero): positive Poisson, positive negative binomial, positive geometric, logarithmic, discretized lognormal (a discretized version of the continuous lognormal, truncated to a minimum of one), zeta (discrete power law), and Yule–Simon (which we refer to as the Yule) distributions (Appendix B). We modeled the data for species groups separately using each statistical distribution [40]. We again estimated the parameters for distributions using MLE in the program R (version 2.13.2; [40]). We used the VGAM package [50] to estimate parameters for the positive Poisson, positive negative binomial, positive geometric, and logarithmic distributions. We used the methods and code provided in Clauset et al. [10] to estimate the parameters for the discretized lognormal, the zeta, and the Yule distributions. In applying the zeta distribution, both a shape parameter as well as a threshold (sometimes referred to as $x_{\min}$) can be estimated, below which data are excluded from the analysis. This is sometimes done because it is hypothesized that power law distributions may occur only above some minimum value for a given data set [10]. Because we were interested in fitting each of these distributions to the complete dataset, we set the threshold equal to one for the zeta distribution (and other distributions, where applicable).

For both the points and marks, we calculated the log-likelihood of each model. We used the likelihoods to calculate Akaike's Information Criterion corrected for finite sample sizes (AICc), which we then used to rank the models [6]. We further assessed model fit using the Vuong closeness test [47] for pair-wise comparisons of the best fitting models to the flock size data (marks). The Vuong is a likelihood-ratio test that measures whether one model is closer than the other to the unknown true model using the Kullback–Leibler information criterion [47] and can be derived for both nested and non-nested models. The benefit of using the Vuong test is that it allowed us to evaluate the hypothesis that models ranked higher based on AICc were significantly closer to the true data-generating model than lower-ranked models through estimation of a p-value. We implemented the Vuong test by generalizing the "vuong" function for non-nested models (because all top models turned out to be non-nested) in the pscl package in program R [51]. We then compared parameter estimates for the top models for each species group.

3. Results

There were 1742, 2709, and 4047 flocks observed from 2009 to 2011 for Common eiders, Long-tailed ducks, and scoters, respectively, with the total number of individuals being 28,968 Common eiders, 30,677 Long-tailed ducks, and 55,859 scoters. The number of flocks per transect ranged from 0 to 95 for Common eiders, 0–130 for Long-tailed ducks, and 0–104 for scoters. Even after accounting for species ranges, there were a large number of transects in which no flocks were observed: 166 out of 327 for Common eiders, 413 out of 681 for Long-tailed ducks, 525 out of 936 for scoters.

Flock size ranged from 1 to 2000 for Common eiders, 1–750 for Long-tailed ducks, and 1–5000 for scoters with the median flock size equal to three for Common eiders and Long-tailed ducks and four for scoters. However, the standard deviation of flock size was quite high: 94 for Common eiders, 39 for Long-tailed ducks, and 112 for scoters. These statistics and plots of log-frequency versus log-abundance (Fig. 1) demonstrate the right skew of the flock size distributions.

3.1. Distribution of number of flocks per transect

The negative binomial distributions (zero-inflated and standard) were the best fitting distributions for the data on the number of flocks per transect for all species groups (Table 1; this was also true for the three scoter species identified to species—results not shown). For the Common eider, the zero-inflated negative binomial distribution had a slightly higher log-likelihood (and hence lower AICc value) than the standard negative binomial. In the case of the Long-tailed ducks and scoters, the zero inflation parameter was estimated to be zero, collapsing to the standard negative binomial distribution. The zero-inflated geometric and Poisson distributions had considerably lower log-likelihoods and comparably poorer fits to the data (Table 1).

Please cite this article in press as: E.F. Zipkin, et al., Fitting statistical distributions to sea duck count data: Implications for survey design and abundance estimation, Statistical Methodology (2012), doi:10.1016/j.stamet.2012.10.002

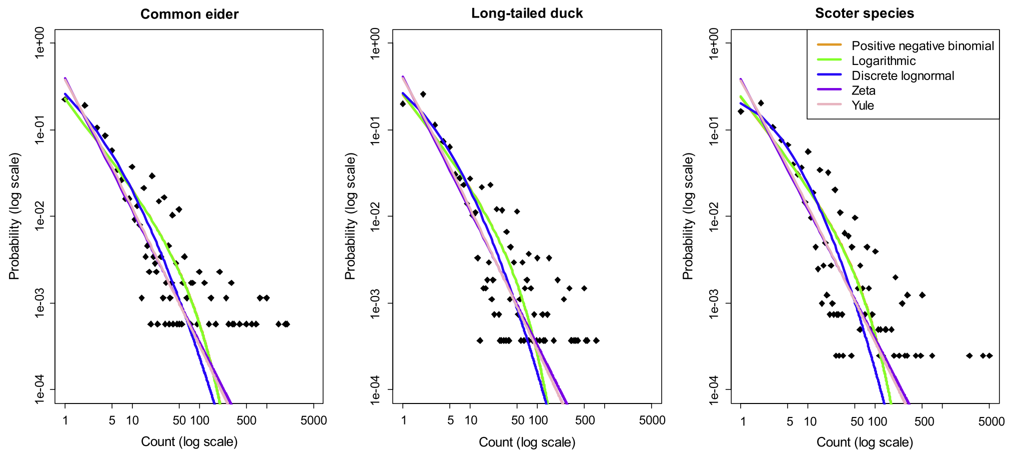


Fig. 1. Model fits (lines) and observed probabilities (black dots) for count data (marks) for the three species groups: Common eiders, Long-tailed ducks, and scoters. Fits are shown for the top 5 models: logarithmic, discretized lognormal, zeta, Yule, and positive negative binomial. The positive negative binomial fit is not visible because it is obscured by the logarithmic fit.

Table 1
Log-likelihood and parameter estimates for distributions fit to data on the number of flocks per transect for Common eiders, Long-tailed ducks, and all scoters combined. Likelihoods are presented because likelihood rankings were identical to AICc rankings (sample sizes were relatively large and the number of parameters for all fitted models ranged from 2 to 3). Specifications for each distribution are given in Appendix A. The parameter φ is the zero inflation parameter (ranging from 0 to 1) and is the probability of a structural zero. The second to last column shows the observed (sample) mean number of flocks per transect for each species (bold) and estimates of the mean under each distributional assumption. Note that the MLE of the negative binomial distribution is the sample mean by definition. The last column shows the observed proportion of transects without flocks (bold) and the proportion estimated under each distributional assumption. The zero inflated negative binomial is excluded from this table for the Long-tailed ducks and scoter species because the zero-inflated parameter was estimated to be zero, collapsing the distribution to a standard negative binomial.

	Log-likelihood	φ	Parameter estimates		Mean flocks per transect	Transects with no flocks
Common eiders					**5.33**	**0.51**
Zero inflated negative binomial	−727.72	0.19	$\mu = 7.20$	$k = 0.43$	5.81	0.43
Negative binomial	−743.24		$\mu = 5.33$	$k = 0.24$	5.33	0.48
Zero inflated geometric	−885.62	0.07	$p = 0.55$		1.12	0.57
Zero inflated Poisson	−1444.37	0.56	$\lambda = 9.57$		4.18	0.49
Long-tailed ducks					**3.98**	**0.61**
Negative binomial	−1162.43		$\mu = 3.98$	$k = 0.21$	3.98	0.54
Zero inflated geometric	−1644.99	0.05	$p = 0.66$		1.86	0.68
Zero inflated Poisson	−2270.05	0.45	$\lambda = 6.82$		3.73	0.45
Scoters					**4.32**	**0.56**
Negative binomial	−1782.63		$\mu = 4.32$	$k = 0.20$	4.32	0.53
Zero inflated geometric	−2286.72	0.07	$p = 0.59$		1.33	0.61
Zero inflated Poisson	−4280.94	0.49	$\lambda = 7.80$		4.00	0.49

3.2. Distribution of flock sizes

The discretized lognormal distribution produced the best fit to the data for flock sizes of all three species groups (Table 2; Fig. 1). This was a consistent result applying to all species together (Fig. 2), each species separately (including the three scoter species when identified to species; results not shown) and each species separately by year (2009–2011; results not shown). In all cases, the discretized lognormal had the lowest AICc value when compared to the other six candidate distributions and had a significantly better fit compared to the other top models as inferred from Vuong pair-wise closeness tests (Table 2). The next best models varied by species group with the logarithmic, Yule, zeta, and positive negative binomial distributions all producing reasonable (although inferior) fits to

Table 2
Model selection results for each model fit to non-zero flock size data for Common eiders, Long-tailed ducks, all scoter species combined. Log-likelihood values are shown in the diagonals. Likelihoods are presented because likelihood rankings were identical to AICc rankings (sample sizes were relatively large and the number of parameters ranged from 1 to 2 for all fitted models). The off-diagonals report the *p*-values from pair-wise Vuong closeness tests. In all pair-wise comparisons, the distribution with the lower log-likelihood value was also identified as the best (closest to unknown true model) by the Vuong test statistic. However, the values in grey show when the difference was not significant. The positive Poisson and geometric models are excluded from our comparison because their likelihoods indicated very poor fits to our data (Common eiders: −6585.6 geom, −61,046.0 pois; Long-tailed ducks: −9160.3 geom, −48,029.6 pois; scoters: −14,519.5 geom, −111,268.9 pois).

Common eiders					
	Discretized lognormal	Yule	Zeta	Logarithmic	Positive negative binomial
Discretized lognormal	−5227.0				
Yule	<0.001	−5347.9			
Zeta	<0.001	<0.001	−5404.8		
Logarithmic	<0.001	0.049	0.333	−5425.5	
Positive negative binomial	<0.001	0.041	0.304	<0.001	−5429.3
Long-tailed ducks					
	Discretized lognormal	Yule	Logarithmic	Positive negative binomial	Zeta
Discretized lognormal	−7718.0				
Yule	<0.001	−7922.1			
Logarithmic	<0.001	0.394	−7931.6		
Positive negative binomial	<0.001	0.352	<0.001	−7935.9	
Zeta	<0.001	<0.001	0.007	0.007	−8022.5
Scoters					
	Discretized lognormal	Logarithmic	Positive negative binomial	Yule	Zeta
Discretized lognormal	−12 312.9				
Logarithmic	<0.001	−12 764.7			
Positive negative binomial	<0.001	<0.001	−12 774.4		
Yule	<0.001	0.126	0.149	−12 901.7	
Zeta	<0.001	0.005	0.008	<0.001	−13 069.6

the data (Table 2; Fig. 1). For all three species, the positive negative binomial had a very similar, although slightly inferior fit as compared to the logarithmic distribution using AICc and Voung tests (e.g., the positive negative binomial model is obscured by the logarithmic in Fig. 1). This is consistent with the fact that the logarithmic distribution is a limiting case of the negative binomial [39,29] and that the shape parameter in the negative binomial for all species was close to zero (Table 3). This was also true for the Yule and zeta distributions, whose fits were qualitatively very similar, although the Yule outperformed the zeta for all species by AICc and Vuong tests (Table 2). The geometric and positive Poisson models were the worst fitting models in all cases with likelihoods much lower than the other models (see caption for Table 2) and were thus excluded from further consideration.

In all comparisons, the direction of the Vuong test statistic supported the ranking of model fits by their AICc values (and by their log-likelihoods). The discretized lognormal had a significantly better fit as compared to the other six distributions for all three species groups (Vuong tests, $p < 0.001$; Table 2). In all other pair-wise comparisons, the distribution with the highest likelihood value was judged closer to the true model than the inferior model, although in some situations the difference between models was not significant.

Fig. 2 shows log-probability versus log-abundance plots for each distribution for simulated data using parameter values as estimated by maximum likelihood fitting to combined flock size data from all species (Fig. 2 column 1) as compared to the actual data of all species groups combined (Fig. 2 column 2). The figure demonstrates that the positive Poisson, positive geometric, logarithmic, and positive negative binomial distributions are unable to account for the large flocks sizes that are observed in the data while the zeta and Yule are capable of producing flock sizes that are much larger than observed in the data. Fig. 2 highlights the superior fit of the discretized lognormal distribution –

Please cite this article in press as: E.F. Zipkin, et al., Fitting statistical distributions to sea duck count data: Implications for survey design and abundance estimation, Statistical Methodology (2012), doi:10.1016/j.stamet.2012.10.002

Table 3
Parameter estimates for the top five models to the flock size data for: all species combined, Common eiders, Long-tailed ducks, and scoters (listed in order by AICc). The values shown are the parameters for each distribution as described in Appendix B. The six right-most columns of the table give summary statistics of the observed flock size data for each species (bold) as well as summary statistics of simulations of flock size under each fitted distribution. The summaries for each distribution are the mean values based on 10,000 simulations using each species' parameter estimates and size of the sample data. The last column is the standard deviation of the maximum count over the 10,000 simulations. Note that the MLE parameters for the negative binomial and logarithmic distributions are such that the estimated mean of the distribution is the sample mean by definition.

	Parameter estimates		1st quartile	Median	Mean	3rd quartile	Max	SD (max)
All species			**2**	**3**	**13.59**	**9**	**5000**	
Discretized lognormal	$\mu = 1.093$	$\sigma = 1.478$	2.00	4.00	10.03	9.11	993.72	634.77
Logarithmic	$p = 0.982$		1.89	4.45	13.59	14.50	343.81	58.40
Positive negative binomial	$\mu = 0.438$	$k = 0.008$	1.96	4.63	13.54	14.57	338.26	59.22
Yule	$a = 0.610$		1.00	2.00	4.6E+06	7.88	3.9E+10	3.3E+12
Zeta	$a = 0.518$		2.00	4.00	2.0E+07	14.56	1.7E+11	9.6E+12
Common eiders			**2**	**3**	**16.63**	**9**	**2000**	
Discretized lognormal	$\mu = 0.866$	$\sigma = 1.680$	1.14	3.40	11.83	9.45	959.22	843.11
Yule	$a = 0.609$		**1.00**	2.03	3.6E+04	7.84	6.2E+07	1.7E+09
Zeta	$a = 0.521$		**2.00**	3.96	1.4E+08	14.37	2.4E+11	2.2E+13
Logarithmic	$p = 0.986$		1.97	5.05	16.63	17.34	347.07	75.37
Positive negative binomial	$\mu = 0.419$	$k = 0.006$	1.99	5.13	16.89	17.69	350.89	76.29
Long-tailed ducks			**2**	**3**	**11.32**	**7**	**750**	
Discretized lognormal	$\mu = 0.886$	$\sigma = 1.526$	1.03	3.01	9.13	8.16	649.26	459.21
Yule	$a = 0.652$		1.00	2.00	1.5E+04	6.84	4.1E+07	1.5E+09
Logarithmic	$p = 0.977$		1.16	4.00	11.33	12.33	231.47	47.87
Positive negative binomial	$\mu = 0.314$	$k = 0.008$	1.23	4.00	11.29	12.35	227.77	46.70
Zeta	$a = 0.548$		2.00	3.64	8.2E+05	12.56	2.2E+09	1.0E+11
Scoters			**2**	**4**	**13.80**	**10**	**5000**	
Discretized lognormal	$\mu = 1.286$	$\sigma = 1.369$	2.00	4.00	9.97	9.94	589.93	359.30
Logarithmic	$p = 0.982$		1.85	4.57	13.80	14.71	315.93	60.20
Positive negative binomial	$\mu = 0.919$	$k = 0.017$	1.98	4.90	14.04	15.15	313.52	61.90
Yule	$a = 0.586$		2.00	2.06	1.2E+05	8.48	4.9E+08	1.3E+10
Zeta	$a = 0.498$		2.00	4.00	9.0E+07	16.20	3.6E+11	2.4E+13

which best captures the range of variation observed in the right tail – to the sea duck data as compared to the other six distributions.

The parameter estimates for the top models were comparable among species groups with estimates generally being more similar between Common eiders and Long-tailed ducks as compared to scoters (Table 3). In the parameterization of the zeta and Yule distributions that we present (Appendix B), the mean is not finite for values of $a < 1$ [10,50], yet for all three species groups the maximum likelihood estimates for these parameters were less than one. Thus, in order to compare the output from the fit of each statistical distribution, we simulated count data for each species group that was the size of the sample data ($n_{\text{all}} = 8498$; $n_{\text{common eider}} = 1742$; $n_{\text{long-tailed ducks}} = 2709$; $n_{\text{scoters}} = 4047$) 10,000 times and report the mean values for the summary statistics (Table 3). These results demonstrate the relationship between sample moments and moments of MLE fitted distributions. Note that the mean of the fitted logarithmic and negative binomial distributions match the observed sample mean (as expected given that the sample mean is the maximum likelihood estimator of the negative binomial and logarithmic means), but result in too many moderately large groups (3rd quartile), too few very large groups (maximum), and an underestimation of the variance observed in the data. Thus, although the fitted negative binomial and logarithmic distributions describe the mean of the data well, they mischaracterize other aspects of the data distribution and underestimate uncertainty about the mean. On the other end of the spectrum, the Yule and zeta distributions have unrealistically heavy tails and overestimate the variance in the counts. For example, the average standard deviation of flock size for all species combined (as estimated from simulations) was 1.15E+09 for the zeta distribution as compared to 25.8 for the discretized lognormal and 23.6 for the negative binomial (and 91.1 in the observed data). Although the standard deviation of flock size is only slightly higher with the discretized lognormal as compared to the logarithmic and negative binomial

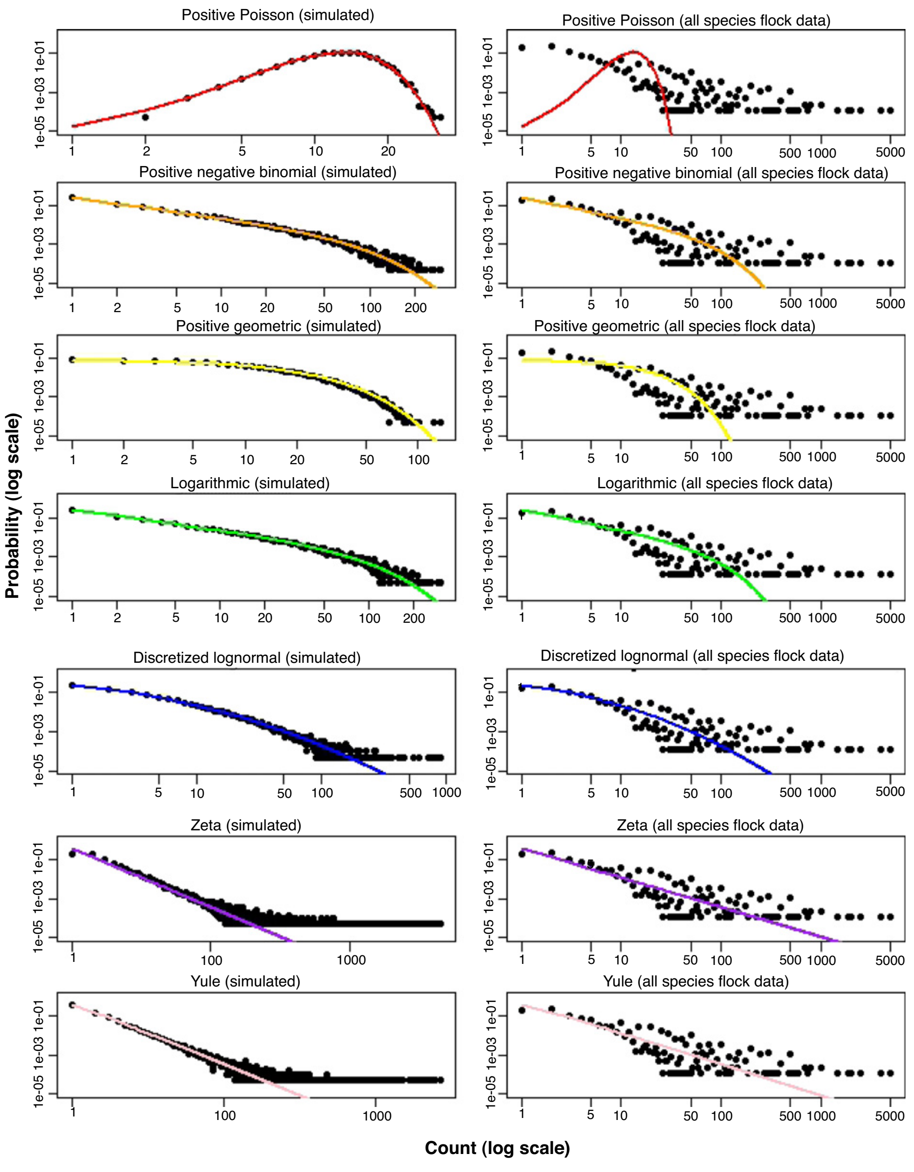


Fig. 2. Simulated (left column) and observed (right column) data for all species fitted using the seven distributions that we compared. Note the variable *x*-axes for the simulated data.

distributions, the latter two distributions are more likely to underestimate maximum flock size (last column, Table 3). The discretized lognormal distribution best matches the range of the observed data (Fig. 2, Table 3) but it also consistently underestimates the mean flock size, in part because it produces too few very large counts. Thus, while the discretized lognormal captures the variance and the upper tail probability of the data somewhat better than the other distributions (negative binomial and logarithmic underestimate upper tail probability and variance; zeta and Yule overestimate upper tail probability and variance), this comes at a cost to efficient estimation of the mean (negative bias of 20%–30% in our simulations). Given this result, Poisson mixture distributions may currently be preferable for abundance estimation, assuming reasonable variance corrections can be incorporated.

Please cite this article in press as: E.F. Zipkin, et al., Fitting statistical distributions to sea duck count data: Implications for survey design and abundance estimation, Statistical Methodology (2012), doi:10.1016/j.stamet.2012.10.002

4. Discussion

We described a marked point process framework for modeling flock numbers and flock sizes to characterize sea duck distribution and abundance in the Atlantic. We employed model selection techniques to choose appropriate models for skewed and zero-inflated distributions of flock numbers and highly right-skewed distributions of flock sizes. Our process-oriented approach should be useful in modeling other highly aggregated, patchily distributed species. The distributions that best fit the "points", i.e., the number of flocks per transect, (negative binomial and zero inflated negative binomial) and "marks", i.e., the flock sizes, (discretized lognormal) were surprisingly consistent across sea duck species and did not vary among years.

Our results have important implications for estimating annual abundances of wintering sea ducks and for designing future surveys that will be able to generate information on population statuses and trends. Inappropriate choice of the distribution family in a modeling framework can lead not only to bias in parameter estimates, but to inaccurate assessments of uncertainty and statistical power. Appropriate characterization of uncertainty and estimation of statistical power are of particular importance in a management context because uncertainty will be propagated through decision-making processes and will affect our understanding of population dynamics, as well as the design and implementation of future monitoring programs. For example, national harvest regulations for many species of ducks are set annually by the US Fish and Wildlife Service using population estimates derived from aerial surveys of breeding areas (e.g., [46,48]); these regulatory decisions are informed by predictions from models of population dynamics that are also derived from survey estimates. Because the sea ducks considered here breed in remote areas that are not covered by current surveys, estimates from winter areas may provide our best means of monitoring responses to exploitation and environmental change, but only if estimates from winter surveys can correctly and precisely estimate abundance. Our results are also particularly relevant to applications that require proper modeling of the extreme values of abundance observed for many species and where surveying presents logistical challenges, thereby limiting the number of samples collected. This includes risk and impact assessments, as well as detection of high-use areas. As marine environments along the eastern United States are currently being considered for development of wind energy production [5], sufficient survey methods and accurate maps are critically needed to assess the potential impacts of the proposed development on sea ducks and seabirds.

The best-fitting distributions for flock size in our study (discretized lognormal, logarithmic, negative binomial, Yule, and zeta) differ from each other primarily in the shape of the upper tail. The probability mass of the zeta distribution declines log-linearly in the tail (that is, linearly on doubly logarithmic axes), and the Yule distribution nearly so, making them the heaviest tailed distributions in our candidate set. This is evident in the relatively common occurrence of very large counts in these distributions (column one in Fig. 2, Table 3). The probability mass of the upper tail of the discretized lognormal distribution declines in a log-quadratic fashion, whereas the logarithmic and negative binomial display an exponential decay in the upper tail. Thus, the heaviness of tails in these distributions is ranked as follows: zeta $\approx$ Yule $>$ discretized lognormal $>$ logarithmic $\approx$ negative binomial. That the discretized lognormal distribution was consistently selected for our three sea duck species groups suggests that the upper tails of flock size distributions for these species are not exponentially bounded (logarithmic and negative binomial), but not as extreme as would be predicted under power law-type distributions (e.g., zeta, Yule). This is fortunate for abundance estimation, because power law behavior implies that the variance (for $a < 2$) and mean (for $a < 1$) are not finite; that is, that sample moments would increase with the area and time spent sampling rather than providing estimates of meaningful characteristic properties of the abundance distribution.

The lognormal distribution has a long history in ecology (e.g., [38]) and a diversity of other fields [27] where it often arises as a plausible alternative to other heavy-tailed distributions like power laws (e.g., in birds; [1]). One classical generative process for a lognormal distribution is the multiplicative stochastic growth process first proposed by Gibrat [16], in which the size of an entity changes by successive multiplicative random effects; if the multiplicative random effects are independent and lognormally distributed, then the size distribution will be lognormal. The lognormal distribution arises even more generally as a direct consequence of the Central Limit Theorem for

Please cite this article in press as: E.F. Zipkin, et al., Fitting statistical distributions to sea duck count data: Implications for survey design and abundance estimation, Statistical Methodology (2012), doi:10.1016/j.stamet.2012.10.002

products of random variables; any process that involves the product of a sufficiently large number of independent and identically distributed random variables having any distribution with finite mean and variance has a limiting lognormal distribution. Thus, a discretized lognormal distribution of counts could arise from a variety of plausible ecological mechanisms. However, the lognormal distribution is known to produce biased estimates of the mean and variance when it is "contaminated" with even small amounts of data from other distributions [32]. In our dataset of flock sizes, the discretized lognormal underestimated the sample mean for all three species (Table 3), which suggests that our data may not conform perfectly to a lognormal distribution. One possible reason for small deviations from lognormality might be nonstationarity in the underlying process. It may be possible to control for this problem by stratifying areas of high/low abundance or adding covariates that account for changes in group sizes, such that the conditional distribution is closer to lognormal. The lack-of-fit of the lognormal may also reflect the manner in which observers count birds in aerial surveys: singles and pairs have a higher probability of being undetected [37], whereas flocks with more birds are typically undercounted [35]. Further exploration of the counting process and the relationship of the observed counts to actual sea duck flock sizes might help explain the disparity between the observed and lognormal tails. The ultimate choice of which distribution is the most appropriate depends on the modeling purpose. In our case, the discretized lognormal was identified as the best fitting distribution overall, and therefore might be the best choice for simulation modeling that requires a compact representation of the whole distribution. Yet, given the sensitivity of moment estimators to slight deviations from the lognormal distribution [32], one might be justified in choosing a statistical distribution with a lower total log-likelihood that can provide more robust mean abundance estimates, such as the logarithmic or negative binomial distributions. Simulation studies could help to choose the optimal distribution for particular applications.

Bonabeau et al. [4] suggested that an exponentially decaying power law may be a useful distribution for dealing with heavy-tailed data that is bounded. To determine the appropriateness of the exponentially decaying zeta distribution compared to our top performing models, we additionally fit this distribution to flock size data for the three species groups. While the exponentially decaying zeta distribution had greater log-likelihood values (−5324.3 for Common eiders, −7854.3 for Long-tailed ducks, and −12 713.0 for scoters) than either the zeta or Yule (suggesting a comparatively better fit; Table 2), it was still outperformed by the discretized lognormal ($p < 0.001$ in Vuong pair-wise comparison tests and lower AICc) for all three species groups, supporting the hypothesis that our data, while skewed, are less heavy-tailed than distributions in the power law family. Although the exponentially decaying power law may not produce a better fit to our data than the discretized lognormal, it may provide a useful alternative because of the above mentioned problems associated with estimating the moments of lognormal distributions when the lognormal is not a perfect fit. By no means did we present an exhaustive list of possible statistical distributions for modeling skewed count data. We suggest further exploration of the exponentially decaying zeta distribution, as well as other distributions as possible alternatives to the discretized lognormal, when abundance estimation is the objective.

It is important to note that selection among statistical distribution models that differ primarily in their tails is notoriously difficult with small sample sizes and noisy data [2,10]. We have used data from a very large survey, but many ecological datasets are substantially smaller and would not allow discrimination among the more similar of the models studied here [32]. This suggests a useful role for meta-analysis, synthetic analysis of large databases, and validation of mechanistic models of processes determining group size distributions, so that recommendations for appropriate choices of distributions can be made for selection of distributions on the basis of taxonomy, life history, environment, etc. The similarity in model fits among species, species groups, and years is encouraging, as it suggests that model power and estimator precision for individual species groups can be gained by borrowing information both over time and across species [48].

Many mechanistic models of group size formation and aggregation have been proposed to give rise to several of the distributions studied here. For example, Caraco [8], Niwa [33] and Ma et al. [29] have each demonstrated how differing rules related to the decision on when to join or leave groups can lead to negative binomial, decaying power law, and logarithmic distributions of group size, respectively. However, in our sea duck example, flock detection and flock size counts are likely the result not only of the biological processes associated with flocks coalescing, but also the specific fixed-width

Please cite this article in press as: E.F. Zipkin, et al., Fitting statistical distributions to sea duck count data: Implications for survey design and abundance estimation, Statistical Methodology (2012), doi:10.1016/j.stamet.2012.10.002

sampling protocol used during the surveys (i.e., the observation process). In this case, the negative binomial distribution combined with the discretized lognormal produced the best fit to our marked point process for observed number of flocks and flocks sizes, but it is possible that other sampling approaches could yield different combinations. Counting large flocks on the ocean within a 200 m strip while in a fast moving airplane is a difficult task, but one that can be improved through training and revised protocols. Beauchamp [2] noted that rough conditions at sea could bias counts and possibly alter which statistical distribution fits best to observed flock sizes. Further exploration of how to minimize and account for the effects of the observation process, such as including covariates, detection functions, and upper limits imposed by the size of the observation unit, may lead to more accurate and precise counts and better estimates of uncertainty, allowing for improved understanding of the biological mechanisms that produce variation in sea duck flock sizes.

Statistical models of ecological count data can be far more complex than those presented here. It is common to include spatial, temporal, and habitat strata, environmental and biological covariates influencing ecological processes leading to the presence or absence of a species, and sampling covariates, which can affect the detection process of individuals during surveying. We intentionally focused our study on simple distributional models for avian count data, neglecting additional complexity that may in some cases improve model explanatory power. It is fundamental to first determine what form of the underlying statistical distribution is appropriate before real world complexities can be incorporated into models. Our marked point process approach matches the observational process (e.g., seeing a flock, then determining its size) and readily allows for inclusion of covariates for both flock detection and flock size estimation.

A parsimonious approach is recommended for a second reason: large scale monitoring programs often do not have the capacity to collect, maintain, and utilize extensive ancillary data sets, and long-term changes in distribution, abundance, or phenology may make models calibrated to fixed strata (e.g., the study area; areas of high density) inappropriate or inefficient at large scales. Thus, simple descriptions that generalize across species and years are extremely valuable, when possible. Our results suggest that the sea duck counts based on our survey methodology have similar statistical properties, and comparable models can be used over time and across species. These models will form the basis for continued exploration aimed at identifying the covariates affecting wintering sea duck populations, and providing decision makers with the best possible description of sea duck distributional patterns and trends.

Acknowledgments

The authors would like to thank the many pilots and observers for collecting the data for this study: pilots J. Bredy, J. Bidwell, C. Ferguson, M. Koneff, T. Liddick, W. Rhodes, F. Roetker, J. Solberg, and J. Wortham; observers N. Carle, S. Earsom, D. Forsell, T. Jones, T. Lewis, H. Obrecht, P. Padding, M. Perry, and T. White. We would also like to thank K. Fleming, K. Luke, and D. Saalfeld for data and GIS support and A. Royle and D. Rypkema for comments on an earlier draft. Data collection was funded in part by the Sea Duck Joint Venture, the Atlantic Coast Joint Venture, and the Bureau of Ocean Energy Management. Data analysis and model development were funded in part by the Bureau of Ocean Energy Management, Office of Renewable Energy Programs through Interagency Agreement M12PG00068 with the US Department of Commerce, National Oceanic and Atmospheric Administration, National Ocean Service, National Centers for Coastal Ocean Science. B.P.K. was supported under NOAA Contract No. DG133C07NC0616 with Consolidated Safety Services, Inc. The findings and conclusions in this article are those of the author(s) and do not necessarily represent the views of the US Fish and Wildlife Service. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the US Government.

Appendix A

Parameters and probability mass functions for the four distributions that we compare using the data on the number of sea duck flocks per transect. In all cases, the support is $x \in \{0, 1, 2, 3, \ldots\}$.

Specifications of all distributions are as in the VGAM R package [50].

Distribution	Parameters	Probability mass function	
		$P[X=0]$	$P[X=x]$
Zero-inflated Poisson	$0 \le \varphi \le 1$ $\lambda > 0$	$\varphi + (1-\varphi)\, e^{-\lambda}$	$(1-\varphi)\ \dfrac{\lambda^x e^{-\lambda}}{x!}$
Zero-inflated geometric	$0 \le \varphi \le 1$ $0 < p \le 1$	$\varphi + (1-\varphi)\, p$	$(1-\varphi)\, p\, (1-p)^x$
Zero-inflated negative binomial	$0 \le \varphi \le 1$ $\mu > 0$ $k > 0$	$\varphi + (1-\varphi)\left(\dfrac{k}{\mu+k}\right)^k$	$(1-\varphi)\, \mathrm{dnbinom}\,(x, \mu, k)$
Negative binomial	$\mu > 0$ $k > 0$	$\mathrm{dnbinom} = \dbinom{x+k-1}{x}\left(\dfrac{\mu}{\mu+k}\right)^x\left(\dfrac{k}{\mu+k}\right)^k$	

Appendix B

Parameters and probability mass functions for the seven distributions that we compare using the sea duck flock size data. In all cases, the support is $x \in \{1, 2, 3, \ldots\}$. Specifications of all distributions are as in the VGAM R package [50] except for the discretized lognormal which is specified as in Clauset et al. [10].

Distribution	Parameters	Probability mass function
Positive Poisson	$\lambda > 0$	$\dfrac{\frac{\lambda^x}{x!} e^{-\lambda}}{1-e^{-\lambda}}$
Positive negative binomial	$\mu > 0$ $k > 0$	$\dfrac{\left(\dfrac{\Gamma(x+k)}{x!\,\Gamma(k)}\right)\left(\dfrac{\mu}{\mu+k}\right)^x\left(\dfrac{k}{\mu+k}\right)^k}{1-\left(\dfrac{k}{\mu+k}\right)^k}$
Geometric	$0 < p \le 1$	$p\,(1-p)^{x-1}$
Logarithmic	$0 < p < 1$	$\dfrac{-1}{ln\,(1-p)}\ \dfrac{p^x}{x}$
Discretized lognormal	$-\infty < \mu < \infty$ $\sigma > 0$	$\dfrac{\dfrac{\exp\left(-\dfrac{(\ln(x-0.5)-\mu)^2}{2\sigma^2}\right)}{(x-0.5)\sqrt{2\pi\sigma^2}} - \dfrac{\exp\left(-\dfrac{(\ln(x+0.5)-\mu)^2}{2\sigma^2}\right)}{(x+0.5)\sqrt{2\pi\sigma^2}}}{\sqrt{\dfrac{2}{\pi\sigma^2}}\exp\left(-\dfrac{(\ln(0.5)-\mu)^2}{2\sigma^2}\right)}$
Zeta	$a > 0$	$\dfrac{1}{x^{a+1}} \Big/ \sum_{n=1}^{\infty} \dfrac{1}{n^{a+1}}$
Yule	$a > 0$	$\dfrac{a\,\Gamma(x)\,\Gamma(a+1)}{\Gamma(x+a+1)}$

Please cite this article in press as: E.F. Zipkin, et al., Fitting statistical distributions to sea duck count data: Implications for survey design and abundance estimation, Statistical Methodology (2012), doi:10.1016/j.stamet.2012.10.002

References

[1] A.P. Allen, B.L. Li, E.L. Charnov, Population fluctuations, power laws and mixtures of lognormal distributions, Ecol. Lett. 4 (2001) 1–3.
[2] G. Beauchamp, Fit of aggregation models to the distribution of group sizes in Northwest Atlantic seabirds, Mar. Ecol. Prog. Ser. 425 (2011) 261–268.
[3] E. Bonabeau, L. Dagorn, Possible universality in the size distribution of fish schools, Phys. Rev. E 51 (1995) R5220–R5223.
[4] E. Bonabeau, L. Dagorn, P. Freon, Scaling in animal group-size distributions, Proc. Natl. Acad. Sci. 96 (1999) 4472–4477.
[5] C. Bowes, J. Allegro, The turning point for Atlantic offshore wind energy: time for action to create jobs, reduce pollution, protect wildlife, and secure America's energy future, Report produced by the National Wildlife Federation, 2012, p. 56.
[6] K.P. Burnham, D.R. Anderson, Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, second ed., Springer-Verlag, 2002.
[7] D.F. Caithamer, M. Otto, P.I. Padding, J.R. Sauer, G.H. Haas, Sea ducks in the Atlantic flyway: population status and a review of the special hunting seasons. US Fish and Wildlife Service, Laurel, Maryland, USA, 2000, pp. 1–40.
[8] T. Caraco, Stochastic dynamics of avian foraging flocks, Am. Nat. 115 (1980) 262–275.
[9] G. Certain, E. Bellier, B. Planque, V. Bretagnolle, Characterising the temporal variability of the spatial distribution of animals: an application to seabirds at sea, Ecography 30 (2007) 695–708.
[10] A. Clauset, C.R. Shalizi, M.E.J. Newman, Power-law distributions in empirical data, SIAM Rev. 51 (2009) 661–703.
[11] J.E. Cohen, Markov population processes as models of primate social and population dynamics, Theor. Popul. Biol. 3 (1972) 119–134.
[12] D.J. Daley, D. Vere-Jones, An Introduction to the Theory of Point Processes, vol. I, second ed., Springer, New York, 2003.
[13] A.L. Drewitt, R.H.W. Langston, Assessing the impacts of wind farms on birds, Ibis 148 (2006) 29–42.
[14] I. Fujisaki, E.V. Pearlstine, M. Miller, Detecting population decline of birds using long-term monitoring data, Popul. Ecol. 50 (2008) 275–284.
[15] S. Garthe, O. Huppop, Scaling possible adverse effects of marine wind farms on seabirds: developing and applying a vulnerability index, J. Appl. Ecol. 41 (2004) 724–734.
[16] R. Gibrat, Les Inégalités Economiques, Recueil Sirey, Paris, 1931.
[17] M. Griesser, Q. Ma, S. Webber, K. Bowgen, D.J.T. Sumpter, Understanding animal group-size distributions, PLoS One 6 (2011) http://dx.doi.org/10.1371/journal.pone.0023438.
[18] M. Guillemette, J.H. Himmelman, C. Barette, Habitat selection by common eiders in winter and its interaction with flock size, Can. J. Zool. 71 (1993) 1259–1266.
[19] D. Hall, Zero-inflated Poisson and binomial regression with random effects: a case study, Biometrics 56 (2000) 1030–1039.
[20] M. Jacobsen, Point Process Theory and Applications: Marked Point and Piecewise Deterministic Processes, Birkhäuser, Boston, Massachusetts, 2006.
[21] R. Jovani, R. Mavor, D. Oro, Hidden patterns of colony size variation in seabirds: a logarithmic point of view, Oikos 117 (2008) 1774–1781.
[22] R. Jovani, D. Serrano, E. Ursua, J.L. Tella, Truncated power laws reveal a link between low-level behavioural processes and grouping patterns in a colonial bird, PLoS One 3 (2008) http://dx.doi.org/10.1371/journal.pone.0001992.
[23] T.H. Keitt, H.E. Stanley, Dynamics of North American breeding bird populations, Nature 393 (1998) 257–260.
[24] M. Kirk, D. Esler, S.A. Iverson, Movements of wintering surf scoters: predator responses to different prey landscapes, Oecologia 155 (2008) 859–867.
[25] J.K. Larsen, M. Guillemette, Effects of wind turbines on flight behaviour of wintering common eiders: implications for habitat use and collision risk, J. Appl. Ecol. 44 (2007) 516–522.
[26] T.L. Lewis, D. Esler, W.S. Boyd, Foraging behavior of surf scoters and white-winged scoters in relation to clam density: inferring food availability and habitat quality, The Auk 125 (2008) 149–157.
[27] E. Limpert, W.A. Stahel, M. Abbt, Log-normal distributions across the sciences: keys and clues, BioScience 51 (2001) 341–352.
[28] W.A. Link, J.R. Sauer, A hierarchical analysis of population change with application to cerulean warblers, Ecology 83 (2007) 2832–2840.
[29] Q. Ma, A. Johansson, D.J.T. Sumpter, A first principles derivation of animal group size distributions, J. Theoret. Biol. 283 (2011) 35–43.
[30] T.G. Martin, B.A. Wintle, J.R. Rhodes, P.M. Kuhnert, S.A. Field, S.J. Low-Choy, A.J. Tyre, H.P. Possingham, Zero tolerance ecology: improving ecological inference by modelling the source of zero observations, Ecol. Lett. 8 (2005) 1235–1246.
[31] M. Mitzenmacher, A brief history of generative models for power law and lognormal distributions, Internet Math. (2003) 226–251.
[32] R.A. Myers, P. Pepin, The robustness of lognormal-based estimators of Abundance, Biometrics 46 (1990) 1185–1192.
[33] H.S. Niwa, Power-law versus exponential distributions of animal group sizes, J. Theoret. Biol. 224 (2003) 451–457.
[34] A. Okubo, Dynamical aspects of animal grouping: Swarms, schools, flocks and herds, Adv. Biophys. 22 (1986) 1–94.
[35] A.T. Pearse, P.D. Gerard, S.J. Dinsmore, R.M. Kaminski, K.J. Reinecke, Estimation and correction of visibility bias in aerial surveys of wintering ducks, J. Wildl. Manag. 72 (2008) 808–813.
[36] M.C. Perry, A.S. Deller, Waterfowl population trends in the Chesapeake Bay area, in: Hill, S. Nelson (Eds.), Proceedings of the 1994 Chesapeake Research Conference; Toward a Sustainable Watershed: The Chesapeake Experiment, in: CRC Publication, vol. 149, Chesapeake Research Consortium, Edgewater, MD, 1995, pp. 490–504.
[37] K.H. Pollock, W.L. Kendall, Visibility bias in aerial surveys: a review of estimation procedures, J. Wildl. Manag. 51 (1987) 502–510.
[38] F.W. Preston, The commonness and rarity of species, Ecology 29 (1948) 254–283.
[39] M.H. Quenouille, A relationship between the logarithmic, Poisson, and negative binomial series, Biometrics 5 (1949) 162–164.
[40] R Development Core R: a language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2011.

Please cite this article in press as: E.F. Zipkin, et al., Fitting statistical distributions to sea duck count data: Implications for survey design and abundance estimation, Statistical Methodology (2012), doi:10.1016/j.stamet.2012.10.002

[41] US Fish and Wildlife Service, Atlantic Coast Wintering Sea Duck Survey 2008–2011: Summary Report, US Department of the Interior, Washington, DC, USA, 2012.
[42] Sea Duck Joint Venture, Species Status Report. 2003. http://www.seaduckjv.org/meetseaduck/species_status_summary.pdf.
[43] E.D. Silverman, M. Kot, E. Thompson, Testing a simple stochastic model for the dynamics of waterfowl aggregation, Oecologia 128 (2001) 608–617.
[44] N.J. Silvy (Ed.), The Wildlife Techniques Manual, seventh ed., The Johns Hopkins University Press, 2012.
[45] M. Sjoberg, B. Albrectsen, J. Hjalten, Truncated power law: a tool for understanding aggregation patterns in animals? Ecol. Lett. 3 (2000) 90–94.
[46] US Fish and Wildlife Service, Adaptive harvest management: 2012 hunting season, US Department of Interior, Washington, DC, 2012, p. 58. Available online at: http://www.fws.gov/migratorybirds/mgmt/AHM/AHM-intro.htm.
[47] Q.H. Vuong, Likelihood ratio tests for model selection and non-nested hypotheses, Econometrica 57 (1989) 307–333.
[48] B.K. Williams, J.D. Nichols, M.J. Conroy, Case study: management of the sport harvest of North American waterfowl, in: Analysis and Management of Animal Populations, Academic Press, San Diego, 2002, p. 817.
[49] C.C. Wood, Aggregative response of common mergansers (Mergus merganser): predicting flock size and abundance on Vancouver Island salmon streams, Can. J. Fish. Aquat. Sci. 42 (1985) 1259–1271.
[50] T.W. Yee, The VGAM package for categorical data analysis, J. Stat. Soft. 32 (2010) 1–34.
[51] A. Zeileis, C. Kleiber, S. Jackman, Regression models for count data in *R*, J. Stat. Soft. 27 (2008) 1–25.
[52] E.F. Zipkin, B. Gardner, A. Gilbert, A.F. O'Connell, J.A. Royle, E.D. Silverman, Distribution patterns of wintering sea ducks in relation to the North Atlantic Oscillation and local environmental characteristics, Oecologia 163 (2010) 893–902.

Please cite this article in press as: E.F. Zipkin, et al., Fitting statistical distributions to sea duck count data: Implications for survey design and abundance estimation, Statistical Methodology (2012), doi:10.1016/j.stamet.2012.10.002