Exxon Valdez Oil Spill
Long-Term Monitoring Program (Gulf Watch Alaska) Final Report

Nearshore Marine Bird Surveys: data synthesis, analysis and recommendations for sampling frequency and intensity to detect population trends

Exxon Valdez Oil Spill Trustee Council Project 12120114-F
Final Report

Heather A. Coletti
Tammy L. Wilson

National Park Service
240 W 5th Avenue
Anchorage, AK 99501

July 2018

Exxon Valdez Oil Spill
Long-Term Monitoring Program (Gulf Watch Alaska) Final Report

Nearshore Marine Bird Surveys: data synthesis, analysis and recommendations for sampling frequency and intensity to detect population trends

Exxon Valdez Oil Spill Trustee Council Project 12120114-F
Final Report

Heather A. Coletti
Tammy L. Wilson

National Park Service
240 W 5th Avenue
Anchorage, AK 99501

July 2018

Nearshore Marine Bird Surveys: data synthesis, analysis and recommendations for sampling frequency and intensity to detect population trends

Exxon Valdez Oil Spill Trustee Council Project 12120114-F
Final Report

Study History: Project 12120014-F originated from the need to have sufficient power to detect trends within a long-term monitoring program. Sustainability of long-term monitoring programs requires optimization of sampling intensity and efforts to minimize costs while concurrently having sufficient power to detect a trend. While there has been critical thought in the past regarding these issues, we attempted to use currently available analytical methods to allow for use of existing data in simulations to estimate number of samples and sample frequency required to detect a specified trend as well as to examine factors contributing to variation, such as imperfect detection.

The National Park Service Southwest Alaska Network Inventory and Monitoring Program and the U. S. Fish and Wildlife Service have been conducting skiff-based surveys of marine birds for over 5 and 20 years, respectively, along the Prince William Sound, Katmai and Kenai Fjords coastlines. These surveys do not currently account for imperfect detection nor do they focus on any single species in particular or nearshore habitat type. However, within the Southwest Alaska Network program, the goal is to estimate trends for a select group of marine bird species that rely on the nearshore food web and were impacted by the *Exxon Valdez* oil spill.

From preliminary analysis of Nation Park Service data, the current survey design does not provide power for detecting trends for the identified indicator species with suitable confidence (<0.50). We utilized coefficients of variance to determine within-year as well as across-year variation for each species. We determined that we may not be adequately surveying for some species because: (1) certain species are highly aggregated; (2) we are focusing on inappropriate habitat for species of interest; (3) our sample size is too small; or (4) the year-to-year variation in distribution and abundance is great enough that we should be conducting replicate surveys within a single season.

We proposed to continue to monitor existing transects to have continuity with legacy data. However, our objectives were to improve on existing protocols by minimizing variation and make recommendations to improve efficiency through sample intensity and frequency. This was done by examining the effects of sampling error and imperfect detection. Improving sampling methods will provide a better sense of population trends of specific species across the western Gulf of Alaska and increase efficiency as we move forward in our efforts to monitor species of interest within the *Exxon Valdez* spill area. This final report includes all analyses conducted for the Science Synthesis Report for the Gulf Watch Alaska Program (Neher et al. 2015).

Abstract: Sustainability of long-term monitoring programs requires optimization of sampling intensity and efforts to minimize costs while concurrently having sufficient power to detect a trend. While there has been critical thought in the past regarding these issues, we attempted to use currently available analytical methods to allow for use of existing data in simulations to estimate number of samples and sample frequency required to detect a specified trend as well as to examine factors contributing to variation, such as imperfect detection. We were unable to complete the full suite of analyses originally proposed, based on findings from preliminary analyses. Based on results from the Synthesis Report, we concluded that effort varied for each transect sampled and this would need to be modeled. This could include time-on-transect or actual length travelled during a single transect survey. There also was high model-selection uncertainty (all models had nearly the same AIC – Akaike information criterion). This indicated that there was still some un-modeled heterogeneity and this could be improved by calculating more appropriate habitat covariates, e.g., assigning habitat classes that take into account both shoreline type and bathymetry. Sample size also was an issue in the preliminary analysis. Although we had five replicate encounter histories for each transect, there were large uncertainties associated with estimates. Essentially, the limited number of transects did not capture the level of heterogeneity in the existing data. Despite this, the current sampling protocol represented the maximum effort that can be expended on surveys given logistical constraints. Because we cannot add more transects at this point, further discussion and analysis may have us: 1) reducing the scope of the monitoring program by focusing our efforts on specific habitats; 2) increasing the number of transects sampled; 3) changing the spatial grain of sampling (sample unit size); and 4) considering more complex model structures in a fully Bayesian framework. The optimal course of action will depend on refinement of monitoring objectives. Any discussion of objectives should address the following: spatial extent of analysis, spatial grain of analysis, target species, hypothesized population drivers, and feasible courses of action (e.g. management or conservation) if change is detected.

Key words: Detection, encounter histories, Gulf of Alaska, Kenai fjords, marine bird, occupancy, population trend, Prince William Sound, sample size, transect length

Project Data: Data used for this analysis (Katmai 2013) have been posted on the AOOS workspace in accordance with Gulf Watch Alaska Data Standards. Data files for these at-sea surveys include species sightings, counts, behaviors, and coordinates as well as the coordinates of the vessel track line. The data release includes five data files (.csv format) per year per sampling region. 2012-2016 doi:10.5066/F7416V6H.

There are no limitations on the use of the data, however, it is requested that the authors be cited for any subsequent publications that reference this dataset. It is strongly recommended that careful attention be paid to the contents of the metadata file associated with these data to evaluate data set limitations or intended use.

Citation:

Coletti, H. A., and T. L. Wilson. 2018. Nearshore marine bird surveys: Data synthesis, analysis and recommendations for sampling frequency and intensity to detect population trends. *Exxon Valdez* Oil Spill Long-Term Monitoring Program (Gulf Watch Alaska) Final Report (*Exxon Valdez* Oil Spill Trustee Council Project 12120114-F). *Exxon Valdez* Oil Spill Trustee Council, Anchorage, Alaska.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

This final report includes all analyses conducted for the 2014 Science Synthesis Report for the Gulf Watch Alaska Program (Neher et al. 2015). We were unable to complete the full suite of analyses originally proposed, based on findings from preliminary analyses. Based on results from the Synthesis Report, we concluded that effort varied for each transect sampled and this would need to be modeled. This could include time-on-transect or actual length travelled during a single transect survey. There also was high model-selection uncertainty (all models had nearly the same AIC – Akaike information criterion). This indicated that there was still some un-modeled heterogeneity and this could be improved by calculating more appropriate habitat covariates, e.g., assigning habitat classes that take into account both shoreline type and bathymetry.

Sample size also was an issue in the preliminary analysis. Although we had five replicate encounter histories for each transect, there were large uncertainties associated with estimates. Essentially, the limited number of transects did not capture the level of heterogeneity in the existing data. Despite this, the current sampling protocol represented the maximum effort that can be expended on surveys given logistical constraints. Because we cannot add more transects at this point, further discussion and analysis may have us: 1) reducing the scope of the monitoring program by focusing our efforts on specific habitats; 2) increasing the number of transects sampled; 3) changing the spatial grain of sampling (sample unit size); and 4) considering more complex model structures in a fully Bayesian framework.

The optimal course of action will depend on refinement of monitoring objectives. Any discussion of objectives should address the following: spatial extent of analysis, spatial grain of analysis, target species, hypothesized population drivers, and feasible courses of action (e.g. management or conservation) if change is detected.

INTRODUCTION

The importance of marine birds to both pelagic and nearshore ecosystems has been discussed in the marine bird introduction and will not be repeated here (Kuletz and Esler 2015). However, the role that marine birds play in both these the ecosystems highlights how this project is linked to pelagic and benthic components of Gulf Watch Alaska (GWA). Nearshore marine bird monitoring currently has two ongoing efforts: 1) a synthesis and analysis of existing data to evaluate our ability to detect change (pelagic component) and 2) continuation of monitoring surveys (nearshore component). Results from the first effort will help direct any future modifications of nearshore marine bird surveys. Our objectives for this project were to improve on existing protocols by minimizing variation and make recommendations to improve efficiency through sample intensity and frequency.

Background

In the early 2000s, a holistic approach to nearshore marine ecosystem monitoring in the Gulf of Alaska (GOA) was developed (Dean and Bodkin 2006). Concurrently, in 2001, a network of five national park units in Southwest Alaska Network (SWAN) began the process of planning a long-term vital signs monitoring program (Bennett et al. 2006). In partnership, these two programs successfully implemented long-term monitoring of the nearshore at several spatial and temporal scales as well as across trophic levels (Dean et al. 2014). Nearshore marine birds were identified as a vital sign for monitoring by SWAN and surveys began in 2006 at Katmai National Park and Preserve (KATM) and in 2007 at Kenai Fjords National Park (KEFJ). Currently, eight years of nearshore geo-referenced survey data exist from KATM and KEFJ, and four years from Prince William Sound (PWS). Data from PWS are being collected by the U. S. Fish and Wildlife Service (USFWS). This work serves as a baseline for many aspects of the current GWA benthic monitoring program.

The original survey objective was to estimate long-term trends in seasonal abundance of seabirds and sea ducks (Dean et al. 2014). This can be a difficult task when data prove to be highly variable. Bennett et al. (2006) suggested summarizing data annually but acknowledged that trends should be estimated after 10 years of initial data collection. As indicated by Dean et al. 2014, the goal of the surveys was to be able to detect a significant decline (>50%) after 10 years of data collection. As we conducted annual summaries, several questions arose: 1) Is current survey intensity adequate to detect trends?; 2) How do we account for imperfect detection?; and 3) How do we correlate changes in abundance and distribution of marine birds to other metrics being measured by the nearshore component of GWA?

Early analyses of KATM and KEFJ marine bird survey results showed high between year variation in density estimates, making trend detection difficult (Coletti et al. 2009). These early analyses resulted in coefficients of variation (CVs) well over 0.50 (CV range of values from: 1.27 to 4.00) for all taxa, therefore confidence intervals for almost all species in all years encompassed zero, indicating little possibility to detect trends over time at our current sampling intensity. In an attempt to reduce CVs, post data collection, subpopulation (domain) analysis was conducted based on shoreline habitat type (Coletti 2009).

Classification of transects into specific habitat types or domains reduced the variability of the density estimates and improved the power to detect change (Coletti 2009). However, a consequence of conducting subpopulation (domain) analysis post survey was that the original sample size (number of transects) was reduced. In surveys similar to ours in Glacier Bay, Alaska, results showed that sample size was an important factor in determining CVs (Drew et al. 2008). Domain based designs generally have large samples sizes (Lehtonen and Pahkinen 2004) and by grouping each transect by habitat type prior to analysis, we essentially reduced sample size of the original survey, eliminating much of the efficiency gained by stratification. However, in our grouped analysis we did detect a decrease in variance, despite reduced sample sizes, which resulted in improved power to detect change. From the 2009 analysis, we recommended exploring the possibility of re-allocating sampling efforts to specific habitat types, reducing variation that may enhance our ability to detect trends for most species of interest.

In the survey's current form, we anticipate that we will be able to detect large (>50%) changes in abundance for relatively common species, but have considered whether we can detect smaller levels of change as well as answer other questions of interest. Hence, to increase power, sources of variation should be identified and removed by method standardization or data analysis.

We recognize that variability is influenced by several factors including, but not limited to:

1. Lack of independence of individuals in groups.

2. Imperfect detection.

3. Habitat preferences by species. Habitat was treated as homogeneous across transects.

4. Annual variation in distribution (i.e., availability) relative to our sampling area; by availability we mean birds present and subject to counts.

5. Within-season variation in distribution – birds may utilize home ranges that are larger than individual transects, and any individual that utilizes a given transect during the season may or may not be present and subject to being detected and counted at any given sampling occasion. Birds also may utilize home ranges that overlap multiple transects.

OBJECTIVES

The fundamental goal of this project was to improve on existing protocols by minimizing variation and make recommendations to improve efficiency through sample intensity and frequency. Specifically:

Objective 1. Develop occupancy model for given species incorporating detection probabilities (single species, multi-season approach).

Objective 2. Provide recommendations and discussion point to optimize survey design.

METHODS

We divided the shoreline of KATM into 5-km transects to obtain a population of sample units available for sampling. Transect lengths were adjusted to accommodate islands or groups of islands with less than 5 km of shoreline (minimum length = 2.5 km, as recommended by Drew et al 2008). Segments ≥ 2.5 km were large enough to contain entire bird aggregations (Bodkin 2011). Twenty-one transects were randomly selected for sampling (field method described below), such that a minimum of 20% of the park shoreline was surveyed. Five groups of transects that occurred on small islands summed to > 5 km. These small transects were later grouped into 5 km sample units, resulting in 19 5-km sample units. However, during standard skiff surveys, depending on tide height, conditions and the abilities of the skiff driver, those transects could be more or less than five km in length. Surveys were conducted once each year, and were not designed to account for imperfect detection and/or availability. To create spatial replicates for occupancy analysis, we divided sample units into 1 km segments.

Surveys were conducted from small vessels (5-8 m length) navigating along coastline transects at speeds of 8-12 knots. All transects were run 100 m offshore and parallel to the shoreline. Two observers searched for marine birds at distances up to 100 m on all sides of the vessel, including 100 m ahead of, behind, and over the vessel. One observer navigated the skiff, and surveyed the offshore portion of transects. The second observer surveyed the shore side of the survey transects. All marine birds within the sampled area were identified and counted. A third team member entered observations into a computer program (dLOG3) designed specifically for these surveys. Data collected included: species, count, sea state (Beaufort scale) and each observer's conditions (scale of 1-5, 1 being excellent and 5 being poor).

All harlequin duck (*Histrionicus histrionicus*) observations from 2013 in KATM were joined (ArcGIS Spatial Join Tool) to the appropriate 1-km spatial sampling unit, thereby creating a five-unit spatial encounter history of bird observations (counts > 0) at each of the 19 sites. Each unit was assigned detection attributes such as: Beaufort, tide height at the time of the survey, actual length surveyed, and occupancy attributes such as: exposure level, and latitude.

Analysis

Presence-absence data present a dilemma for analysis, because false absences can result in biased distribution and occurrence estimates. False absences can be broken into two components: detection (pertaining to the observer) and availability (pertaining to the animal). While it is possible to separate these factors, detection and availability are treated similarly by occupancy analysis. Typically, occupancy analysis is conducted on animal encounter histories generated through repeated surveys of sample units (sites) (MacKenzie et al. 2002, 2006). Observers visit sites a number of times, and record whether animals are detected or not during each visit. Sometimes spatial replicates are used in place of temporal replicates if sites are remote, and difficult to access (Hines et al. 2010, Reynolds and Renner 2014). The idea is that spatial variation is a surrogate for movement, and tells you something about the probability that animals will be in the unit during replicate surveys (Guillera-Arroita 2011).

Using harlequin duck data from 2013, we created encounter histories using the 1km segments from the 19 sample units (17 complete units, and 5 combined units), and fit single season, 2-state occupancy models using the Unmarked package in R. Using four detection and three state covariates (Table 1), we evaluated eleven single-variable models (Table 2). We produced graphs illustrating how detection and occupancy varied with important predictors. We used empirical Bayesian analysis to obtain estimates of the Proportion of Sites Occupied (PSO) using all models with model weight >0.05. We then used model averaging (Burnham and Anderson 2002) to produce the final PSO estimate.

Table 1. Description of covariates used in models of harlequin duck occupancy in KATM.

Variable	Detection	Occupancy	Data type	Description
Length	X	X	Interval	Measure of survey effort, and a surrogate measure of coastline complexity.
Beaufort	X		Ordinal	Wind and sea conditions (1 – 3)
Tide height	X		Interval	Tide height relative to MHHW
Segment type	X		Categorical	Most common habitat type in the 1-k segment
Unit type		X	Categorical	Most common habitat type in the 5-k unit
Latitude		X	Interval	Geographic Y coordinate

Table 2. Model selection table for models single season of harlequin duck occupancy in KATM where p refers to the detection probability while psi is the probability of occupancy.

Model	nPars	AIC	delta	AICwt	cumltvWt
p(Length)psi(.)	3	108.70	0	0.32	0.32
p(Length)psi(Length)	4	109.04	0.34	0.27	0.60
p(Length)psi(Type)	6	109.81	1.11	0.19	0.79
p(Length)psi(Latitude)	4	110.70	2.00	0.12	0.91
p(.),psi(Length)	3	111.63	2.93	0.08	0.98
p(.)psi(.)	2	116.51	7.81	0.01	0.99
p(.),psi(Type)	5	117.08	8.38	0.01	1
p(tide)psi(.)	3	117.75	9.04	0.00	1
p(.),psi(Latitude)	3	118.51	9.80	0.00	1
p(type)psi(.)	5	119.83	11.13	0.00	1
p(Beaufort)psi(.)	4	120.32	11.62	0.00	1

RESULTS

Transect length was the most important predictor of both detection and occupancy, occurring in all of the models with AIC < the no-covariate model (Table 2). Both detection and occupancy increased with increasing transect length (Table 2). There was weak evidence of heterogeneity in occupancy with sites with different habitat types (Table 2). Although there was much variation, protected and semi-protected sites had a slightly lower probability of being occupied than exposed sites (Figure 1). A slight latitudinal gradient was observed, where the probability of occupancy increased with increasing latitude (Table 2). The model-averaged proportion of sites occupied was 0.87 (90% CI = 0.77 - 0.97).

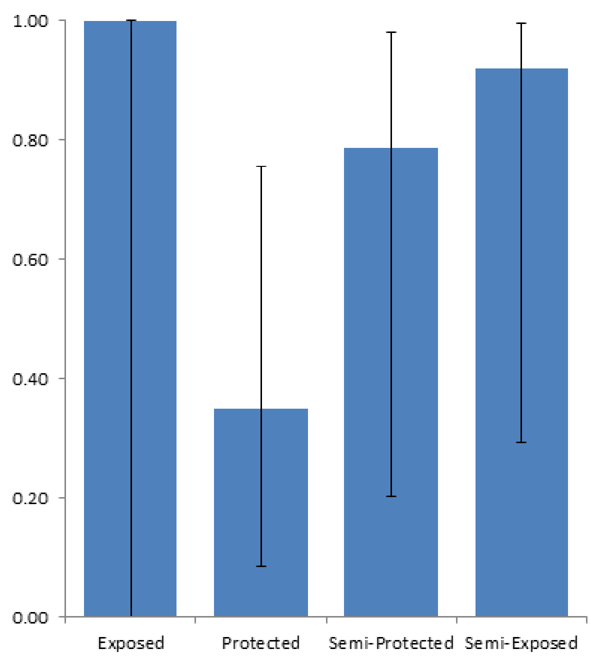
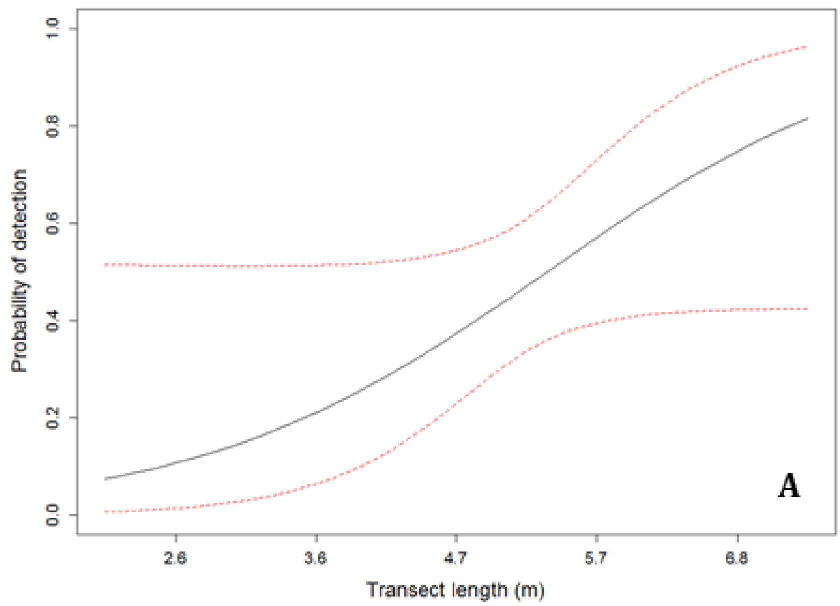


Figure 1. Probability of site occupancy for four habitat types in KATM.



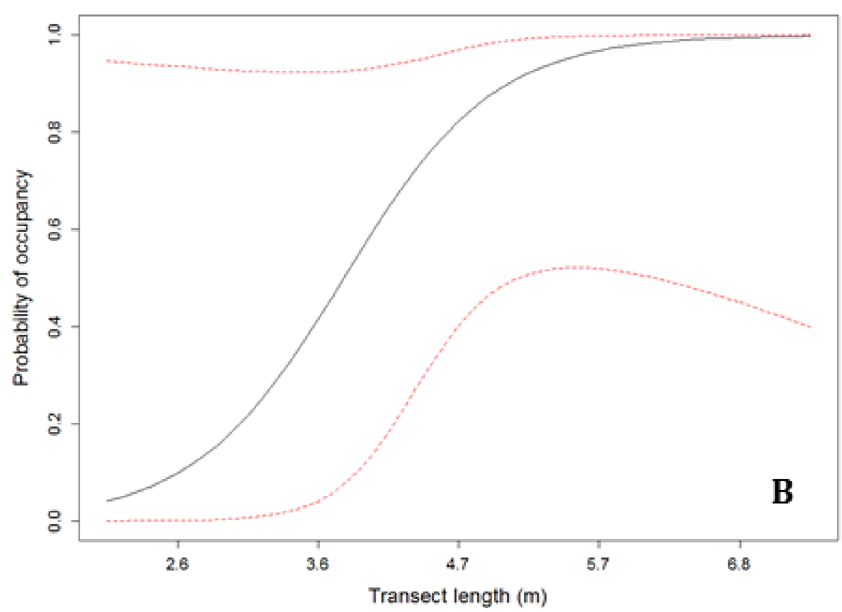


Figure 2. Relationship of harlequin duck (A) detection and (B) occupancy with transect length in KATM. The black line represents the likelihood estimate, and the red-dashed lines represent the 95% confidence intervals.

Because a unit of occupancy is spatially defined, we also assume we will be able to quantify metrics such as prey availability, habitat type, exposure, shoreline complexity, water quality parameters, etc. to that same spatial unit(s). Changes or shifts in site occupancy could theoretically be correlated to other physical or biological drivers of the system. This becomes particularly important in the face of climate change as potential stressors to a system increase. Understanding how a species or community is responding to those stressors through changes in distribution will be informative for resource managers to implement appropriate management actions.

DISCUSSION

From this preliminary analysis, how survey effort is allocated is critical. In the initial design, transects were five km long. However, during standard skiff surveys, depending on tide height, conditions and the abilities of the skiff driver, those transects could be significantly more or less than five km in length. This equates to variable effort per transect. While standardizing length would be ideal, we also recognize that it is not feasible. We suggest effort should be modeled rigorously. This could include time on transect or actual length travelled during a single transect survey. There was also high model-selection uncertainty (all models have nearly the same AIC). This indicates that there is still some un-modeled heterogeneity and this may be improved by calculating more appropriate habitat covariates. For example, assigning habitat classes that take into account both shoreline type and bathymetry.

Sample size was also an issue in the preliminary analysis. Even though we had five replicate encounter histories, there were large uncertainties associated with estimates (Figure 2). Essentially, the limited number of transects does not capture the level of heterogeneity in the existing data. Despite this, the current sampling protocol represented the maximum effort that can be expended on surveys given logistical constraints. While we are not in a position to recommend adding more transects at this point, further discussion and analysis may have us: 1) reducing the scope of the monitoring program by focusing our efforts in specific habitats; 2) increasing the number of transects sampled; 3) changing the spatial grain of sampling (sample unit size); and 4) considering more complex model structures in a fully Bayesian framework. The optimal course of action will depend on refinement of monitoring objectives. For example, the estimated proportion of sites occupied was close to one, and near the upper boundary of that considered to be "meaningful" for occupancy analysis (MacKenzie et al. 2006). Reducing the sample unit size could remedy this problem for harlequin ducks, but may reduce the effectiveness of the sampling design for a species that isn't as common. Any discussion of objectives should address the following: spatial extent of analysis, spatial grain of analysis, target species, hypothesized population drivers, and feasible courses of action (e.g. management or conservation) if change is detected.

LITERATURE CITED

Bennett, A. J., W. L. Thompson, and D. C. Mortenson. 2006. Vital signs monitoring plan, Southwest Alaska Network Inventory and Monitoring Program. Anchorage, AK.

Bodkin, J. L. 2011. SOP for Conducting Marine Bird and Mammal Surveys - Version 4.1. Southwest Alaska Inventory and Monitoring Network. Natural Resource Report NPS/SWAN/NRR-2011/392. Fort Collins, Colorado.

Burnham, K. P., and D. R. Anderson. 2002. Model Selection and Multi-model Inference: a practical information-thoretic approach. Second Edition. Springer, New York, New York, USA.

Coletti, H. A., J. L. Bodkin, T. A. Dean, and K. A. Kloecker. 2009. Nearshore Marine Vital Signs Monitoring in the Southwest Alaska Network of National Parks. Natural Resource Technical Report NPS/SWAN/NRTR - 2009/252. Fort Collins, Colorado.

Dean, T. A., and J. L. Bodkin. 2006. Sampling Protocol for the Nearshore Restoration and Ecosystem Monitoring (N-REM) Program (Nearshore Restoration and Ecosystem Monitoring Research Project G-050750), U. S. Geological Survey, Alaska Science Center. Anchorage, AK.

Dean, T. A., J. L. Bodkin, and H. A. Coletti. 2014. Protocol Narrative for Nearshore Marine Ecosystem Monitoring in the Gulf of Alaska: Version 1.1. Natural Resource Report NPS/SWAN/NRR - 2014/756. Fort Collins, Colorado.

Drew, G. S., S. G. Speckman, J. F. Piatt, J. M. Burgos, and J. L. Bodkin. 2008. Survey Design Considerations for Monitoring Marine Predator Populations in Glacier Bay, Alaska: Results and Post-hoc Analyses of Surveys Conducted in 1999-2003. USGS Final Report.

Guillera-Arroita, G. 2011. Impact of sampling with replacement in occupancy studies with spatial replication. Methods in Ecology and Evolution 2(4):401–406.

Hines, J. E., J. D. Nichols, J. A. Royle, D. I. MacKenzie, A. M. Gopalaswamy, N. S. Kumar, and K. U. Karanth. 2010. Tigers on trails: occupancy modeling for cluster sampling. Ecological Applications 20(5):1456–1466.

Kuletz, K., and D. Esler. 2015. Spatial and temporal variation in marine birds in the northern Gulf of Alaska: the value of marine bird monitoring as part of Gulf Watch Alaska. In: Quantifying temporal and spatial variability across the northern Gulf of Alaska to understand mechanisms of change. Gulf Watch Alaska Synthesis Report to the Exxon Valdez Oil Spill Trustee Council, Projects 14120114 and 14120120.

Lehtonen, R., and E. Pahkinen. 2004. Practical Methods for Design and Analysis of Complex Surveys. Page 349. Second Edition. John Wiley and Sons.

MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. A. Royle, and C. A. Langtimm. 2002. Estimating site occupancy rates when detection probabilities are less than one. Ecology 83(8):2248–2255.

MacKenzie, D. I., J. D. Nichols, J. A. Royle, and C. A. Langtimm. 2006. Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence. Page 324. Academic Press.

Neher TH and others. 2015. Quantifying temporal and spatial ecosystem variability across the northern Gulf of Alaska to understand mechanisms of change. Science Synthesis Report for the Gulf Watch Alaska Program. Programs 14120114 and 14120120. *Exxon Valdez* Oil Spill Trustee Council, Anchorage, Alaska.

Reynolds, J. H., and H. M. Renner. 2014. Using patch occupancy models to estimate area of crevice-nesting seabird colonies. The Condor 116(3):316–324.