USING SPATIOTEMPORAL STATISTICAL MODELS TO ESTIMATE ANIMAL ABUNDANCE AND INFER ECOLOGICAL DYNAMICS FROM SURVEY COUNTS

Paul B. Conn[1*], Devin S. Johnson[1], Jay M. Ver Hoef[1], Mevin B. Hooten[2,3,4], Josh M. London[1], and Peter L. Boveng[1]

[1] *National Marine Mammal Laboratory, Alaska Fisheries Science Center, NOAA National Marine Fisheries Service, Seattle, Washington 98115 U.S.A.*

[2] *U.S. Geological Survey, Colorado Cooperative Fish and Wildlife Research Unit, Colorado State University, Fort Collins, CO 80523 U.S.A.*

[3] *Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, CO 80523 U.S.A.*

[4] *Department of Statistics, Colorado State University, Fort Collins, CO 80523 U.S.A.*

Appendix C: Details of spotted seal abundance analyses employing spatiotemporal statistical models

In this supplement, we describe technical details of the analysis of spotted seal count data, including (1) information on data processing, (2) details on Markov chain Monte Carlo (MCMC) samplers and diagnostics, and (3) more detailed results than are presented in the main article text.

1 Spotted seal count data

We compiled spotted seal count data by georeferencing spotted seal detections made during aerial surveys from April 10, - May 8, 2012, and intersecting them with $25km \times 25km$ sample units used in the analysis (see Fig. 4 of main article). As described by Conn *et al.* (Online Early), detections were first made using an automatic detection algorithm applied to thermal (infrared) video; species information was then obtained by examining paired high resolution photographs. Technicians conducting species determinations recorded species (bearded seal, ribbon seal, ringed seal, spotted seal, unknown seal, walrus, or other), as well as an associated confidence level

("certain," "likely," or "guess"). Since the purpose of this article is to examine spatiotemporal modeling structures on spotted seal counts, we simply used all detections that were recorded as spotted seals, regardless of confidence. Integrating more complex models for the detection process incorporating incomplete detection and species misclassification (see e.g., Conn *et al.* Online Early) will be a focus of future research.

We calculated daily effort - the proportion of each sample unit that was surveyed in a given day- by overlaying global positioning system (GPS) records from each flight over the study area. This approach was used to determine the length of a flight in each cell. The average width of the thermal swath while transiting a particular sample unit was determined by applying trigonometric calculations to GPS altitude measurements and known field-of-view angles. As in a previous analysis of these data (Conn *et al.* Online Early), effort was restricted to times and locations where altitude was $228.6 - 335.3m$ and roll was $< 2.5°$ from center.

We compiled a number of covariates to help explain variation in spotted seal counts:
Sea ice concentration– We obtained daily sea ice concentration values at the resolution of our sampling units from the National Snow and Ice Data Center, Boulder, CO, USA. Sea ice concentration was recorded as the proportion of each sampling unit that consisted of sea ice.
Distance from southern ice edge– For each day in the study, we used sea ice concentration values to construct a daily line indicating the position of the southern ice edge. We then calculated the distance from the centroid of each study unit to the southern ice edge line. This covariate was standardized to have a mean of 1.0 prior to analysis.
Distance from 1000m depth contour– We calculated the distance from the centroid of each sample unit to the 1000m depth contour. The 1000m depth contour was selected to represent the Bering shelf break. This covariate was standardized to have a mean of 1.0 prior to analysis.

All models fit to spotted seal data included linear and quadratic effects of sea ice concentration, linear and square root effects of distance from southern ice edge, and a linear effect for the distance from the 1000m depth contour. We used the same set of prior distributions as indicated in Appendix S1 and applied in simulation analyses. We attempted to fit AST, CPIF, and STPC models to spotted seal data; MCMC execution details are provided in Table C1. We ran two chains for each model to help assess convergence; after discarding a burn-in, samples from

both chains were combined for posterior inference.

1.1 Results

The precision τ_ϵ for AST models all converged towards zero, resulting in extremely large variance for the log of abundance intensity, so much so that many posterior predictions of abundance resulted in numerical overflows (owing to exponentiating a very large number). This model resulted in nonsensical estimates of abundance. As such, results from this model will not be discussed further.

The STPC model also resulted in low posterior values for τ_ϵ and high predictions of abundance in unsampled areas (Fig. C1). The low values of τ_ϵ allowed predictions to match up well with observed data in sampled cells, but led to an extremely right skewed predictions of abundance (Fig. C1). For the STPC model, the posterior median for abundance on April 18, 2012 was 430,400 while the posterior mean was 1,936,000. Both values are considerably larger than expected from previous analyses of seal abundance in the Bering Sea (e.g., Conn *et al.* Online Early, Ver Hoef *et al.* 2014), especially when acknowledging that estimates from spatio-temporal models do not yet factor in incomplete detection (which is likely to make apparent abundance 30-50% of absolute abundance). There was also a large, unrealistic increasing trend in abundance as a function of survey day (Fig. C2).

By contrast to the AST and STPC models, the CPIF model resulted in lower estimates of posterior abundance that largely resembled our a prior expectations. The posterior median and mean were 108,751 and 109,425, respectively. Plots of this distribution, as well as as the posterior distribution for covariate effects, are available in the main article.

Literature Cited

Conn, P. B., Ver Hoef, J. M., McClintock, B. T., Moreland, E. E., London, J. M., Cameron, M. F., Dahle, S. P. & Boveng, P. L. (Online Early). Estimating multispecies abundance using automated detection systems: ice-associated seals in the eastern Bering Sea. *Methods in Ecology and Evolution*, pp. doi: 10.1111/2041–210X.12127.

Ver Hoef, J. M., Cameron, M. F., Boveng, P. L., London, J. M. & Moreland, E. E. (2014). A hierarchical model for abundance of three ice-associated seal species in the eastern Bering Sea. *Statistical Methodology*, **17**, 46–66.

Table C1. Execution details for MCMC analysis of spotted seal spatiotemporal abundance data as a function of estimation model. For each estimation model, we present the number of MCMC chains, the number of iterations used to summarize the posterior distribution for each chain ('MCMC length'), the number of additional iterations at the beginning of the MCMC chain that are discarded as a burn-in ('Burn-in'), the number of additional iterations conducted per simulation in order to tune Metropolis-Hastings proposal distributions prior to start of the main MCMC run ('Adapt length'), the number of consecutive iterations discarded between a saved iteration ("Thin"; used to conserve disk space), and the estimated number of computing hours per MCMC chain ('Hrs/chain'; running on a 2.93-GHz Dell Precision T1500 desktop with 8.0 GB of RAM.)

	Estimation model		
	AST	CPIF	STPC
Number of chains:	2	2	2
MCMC length:	30,000	310,000	410,000
Burn-in:	10,000	10,000	10,000
Adapt length:	5,000	5,000	5,000
Thin:	9	149	199
Hrs/chain:	0.3	45.0	55.3

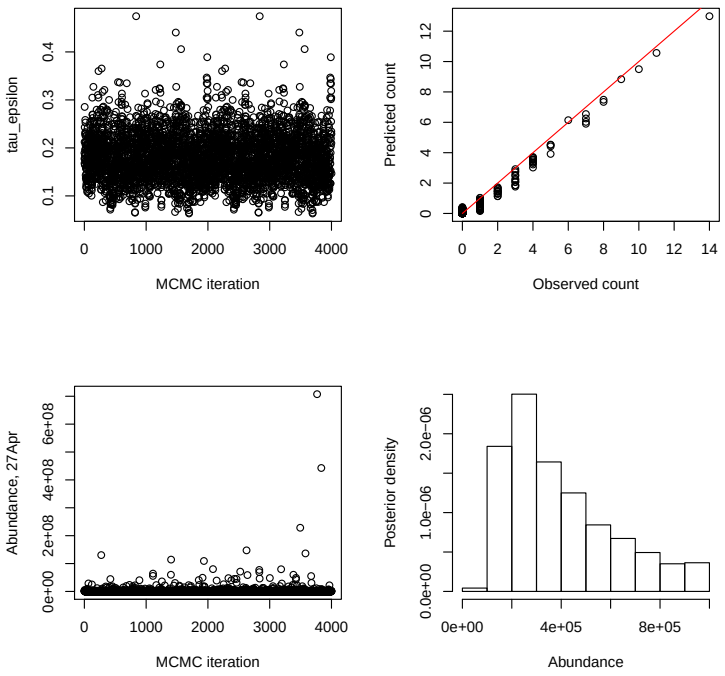

Figure C1. Parameter estimates and diagnostics from combined STPC MCMC chains. The top left panel shows MCMC samples for τ_ϵ, which converges to low values, implying a large variance on the log of abundance intensity. This solution allows the model to fit observed data well (top right panel; predicted counts are posterior mean predictions), but also permits large predictions of abundance in unsampled cells. Posterior predictions of abundance on April 27, 2012 (bottom panels) indicate an extremely right skewed distribution (note that the histogram on the bottom right is truncated at one million; 23% of MCMC samples were greater than this amount).

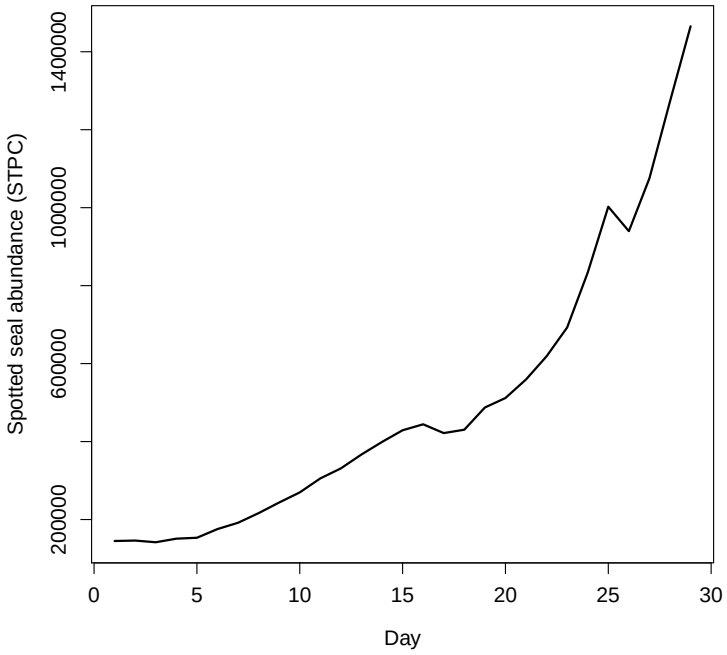

Figure C2. Posterior median predictions for spotted seal abundance as a function of the day of survey. For reference, April 27 corresponds to day 18 (this is the day that surveys were conducted in the epicenter of spotted seal range).