Spatio-temporal models for Arctic seal abundance: successes and computational challenges

Paul B. Conn[1]

[1] *Marine Mammal Laboratory, NOAA, Alaska Fisheries Science Center*

Email: paul.conn@noaa.gov

CAPAM spatio-temporal modeling workshop

NOAA Souteast Fisheries Science Center La Jolla, San Diego

March 1, 2018

Outline

- Ice-associated seals

Outline

- Ice-associated seals
- Seal aerial surveys

Outline

- Ice-associated seals
- Seal aerial surveys
- Abundance and distribution modeling

Outline

- Ice-associated seals
- Seal aerial surveys
- Abundance and distribution modeling
- Bayesian inference

Outline

- Ice-associated seals
- Seal aerial surveys
- Abundance and distribution modeling
- Bayesian inference
- “Bayesian" inference in TMB

Outline

- Ice-associated seals
- Seal aerial surveys
- Abundance and distribution modeling
- Bayesian inference
- “Bayesian" inference in TMB

Ice-associated seals

Bearded seal

Ribbon seal

Ringed seal

Spotted seal

Ice-associated seals

Bearded seal

Ribbon seal

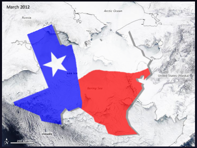
Ringed seal

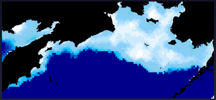
Spotted seal

Seal aerial surveys

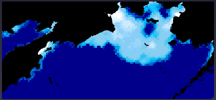
Joint U.S.-Russia Seal Surveys, 2012-2019

- Designed to provide spatially explicit, synoptic estimates of seal abundance over large areas (Okhotsk, Bering, Chukchi, and Beaufort Seas)
- Protocols that allow estimation of all forms of detection probability
- Analysis methods suitable for estimation under variable ice conditions





April 1, 2012

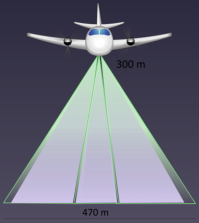


May 14, 2012

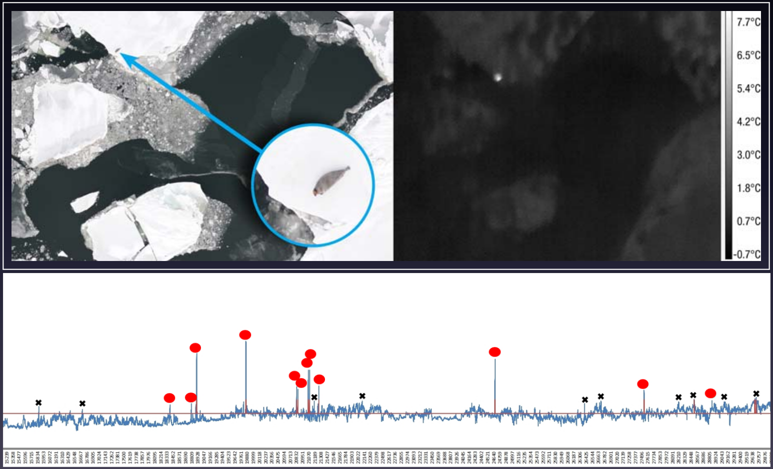
Joint U.S.-Russia Seal Surveys, 2012-2019

Instrument-based surveys

- Aerial surveys combining digital photography and thermal (infrared) video
- Model-based statistical inference: adjustment of flight tracks based on sea ice distribution



Joint U.S.-Russia Seal Surveys, 2012-2019

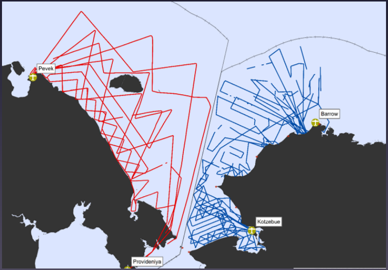


Joint U.S.-Russia Seal Surveys, 2012-2019

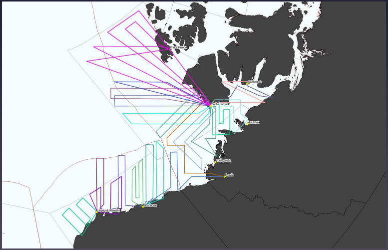
Bering-Okhotsk (BOSS) effort 2012-2013

CHESS survey crew

Chukchi (CHESS) effort 2016

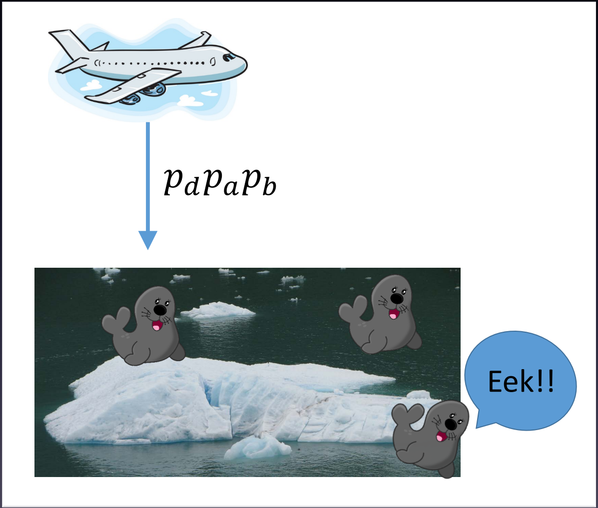


Planned Beaufort effort, 2019



Abundance and distribution modeling

Components of detection



Abundance and distribution modeling

Components of detection: Species misclassification

		True Species			
Obs	Conf.	Bearded	Ribbon	Spotted	Ringed
Bearded	Certain	$\pi^{[1\|1]}$	0	0	0
Bearded	Likely	$\pi^{[2\|1]}$	$\pi^{[2\|2]}$	$\pi^{[2\|3]}$	$\pi^{[2\|4]}$
Bearded	Guess	$\pi^{[3\|1]}$	$\pi^{[3\|2]}$	$\pi^{[3\|3]}$	$\pi^{[3\|4]}$
Ribbon	Certain	0	$\pi^{[4\|2]}$	0	0
Ribbon	Likely	$\pi^{[5\|1]}$	$\pi^{[5\|2]}$	$\pi^{[5\|3]}$	$\pi^{[5\|4]}$
Ribbon	Guess	$\pi^{[6\|1]}$	$\pi^{[6\|2]}$	$\pi^{[6\|3]}$	$\pi^{[6\|4]}$
Ringed	Certain	0	0	$\pi^{[7\|3]}$	0
Ringed	Likely	$\pi^{[8\|1]}$	$\pi^{[8\|2]}$	$\pi^{[8\|3]}$	$\pi^{[8\|4]}$
Ringed	Guess	$\pi^{[9\|1]}$	$\pi^{[9\|2]}$	$\pi^{[9\|3]}$	$\pi^{[9\|4]}$
Spotted	Certain	0	0	0	$\pi^{[10\|4]}$
Spotted	Likely	$\pi^{[11\|1]}$	$\pi^{[11\|2]}$	$\pi^{[11\|3]}$	$\pi^{[11\|4]}$
Spotted	Guess	$\pi^{[12\|1]}$	$\pi^{[12\|2]}$	$\pi^{[12\|3]}$	$\pi^{[12\|4]}$
Unknown	NA	$\pi^{[13\|1]}$	$\pi^{[13\|2]}$	$\pi^{[13\|3]}$	$\pi^{[13\|4]}$

Abundance and distribution modeling

Components of detection: Species misclassification

Confusion Matrix - all observers combined (U.S.)

	Observed Species				
True Species	Bearded	Ribbon	Spotted	Ringed	Unknown
Bearded	**0.83**	0.04	0.05	0.04	0.04
Ribbon	0.00	**0.98**	0.00	0.01	0.00
Spotted	0.03	0.23	**0.64**	0.01	0.09
Ringed	0.01	0.04	0.07	**0.86**	0.02

B. T. McClintock, E. E. Moreland, J. M. London, S. P. Dahle, G. M. Brady, E. L. Richmond, K. M. Yano, and P. L. Boveng. Quantitative assessment of species identification in aerial transect surveys for ice-associated seals. Marine Mammal Science, 31:1057–1076, 2015

Abundance and distribution modeling

Data for spatio-temporal modeling

Discrete in space and time

Count data:

Unit (s)	Time (t)	$R_{s,t}$	$O^1_{s,t}$	$O^2_{s,t}$	$\ldots$	$O^{13}_{s,t}$
20	1	0.02	0	5	$\ldots$	1
143	1	0.03	2	1	$\ldots$	0
150	2	0.02	0	3	$\ldots$	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

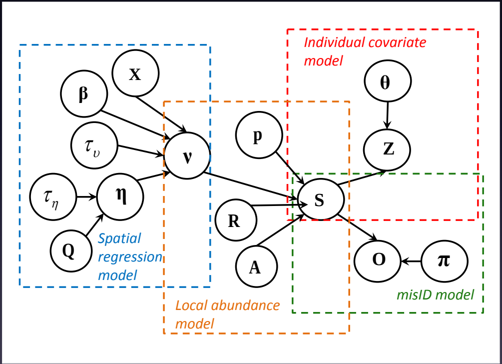
Environmental data:

Unit (s)	Time (t)	Size ($A_{s,t}$)	$X_{s,t}[1]$	$X_{s,t}[2]$	$\ldots$
1	1	1.0	0.5	3.4	$\ldots$
2	1	0.9	0.3	2.1	$\ldots$
3	1	1.0	0.8	0.1	$\ldots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		

Abundance and distribution modeling

Initial analysis: spatial only, eastern Bering 2012

First step: restrict flights to a “short" window; ignore time



Methods in Ecology and Evolution

Methods in Ecology and Evolution 2014, **5**, 1280–1293 doi: 10.1111/2041-210X.12127

SPECIAL ISSUE. MODELLING DEMOGRAPHIC PROCESSES IN MARKED POPULATIONS: PROCEEDINGS OF THE EURING 2013 ANALYTICAL MEETING

Estimating multispecies abundance using automated detection systems: ice-associated seals in the Bering Sea

Paul B. Conn*, Jay M. Ver Hoef, Brett T. McClintock, Erin E. Moreland, Josh M. London, Michael F. Cameron, Shawn P. Dahle and Peter L. Boveng

National Marine Mammal Laboratory, NOAA-NMFS, Alaska Fisheries Science Center, 7600 Sand Point Way NE, Seattle, WA 98115, USA

Summary

1. Automated detection systems employing advanced technology (e.g. infrared imagery, auditory recording systems, pattern recognition software) are compelling tools for gathering animal abundance and distribution data since investigators can often collect data more efficiently and reduce animal disturbance relative to surveys using human observers.

2. Even with these improvements, analysing animal abundance with advanced technology can be challenging because of potential for incomplete detection, false positives and species misidentification. We argue that double sampling with an independent sampling method can provide the critical information needed to account for such errors.

3. We present a hierarchical modelling framework for jointly analysing automated detection and double sampling data obtained during animal population surveys. Under our framework, observed counts in different sampling units are conceptualized as having arisen from a thinned log-Gaussian Cox process subject to spatial autocorrelation (where thinning accounts for incomplete detection). For multispecies surveys, our approach handles incomplete species observations owing to (i) structural uncertainties (e.g. in cases where the automatic detection data do not provide species observations) and (ii) species misclassification; the latter requires auxiliary information on the misclassification process.

4. As an example of combining an automated detection system and a double sampling procedure, we consider the problem of estimating animal abundance from aerial surveys that use infrared imagery to detect animals, and independent, high-resolution digital photography to provide information on species composition and thermal detection accuracy. We illustrate our approach by analysing simulated data and data from a survey of four ice-associated seal species in the eastern Bering Sea.

5. Our analysis indicated reasonable performance of our hierarchical modelling approach, but suggested a need to balance model complexity with the richness of the data set. For example, highly parameterized models can lead to spuriously high predictions of abundance in areas that are not sampled, especially when there are large gaps in spatial coverage.

6. We recommend that ecologists employ double sampling when enumerating animal populations with automated detection systems to estimate and correct for detection errors. Combining multiple data sets within a hierarchical modelling framework provides a powerful approach for analysing animal abundance over large spatial domains.

Key-words: abundance estimation, aerial survey, automated detection, data augmentation, hierarchical models, pattern recognition, spatially restricted regression, species misidentification, thermal imagery

Introduction

Several promising approaches have been developed to monitor animal populations using advanced animal detection technology. Pattern recognition algorithms (e.g. Kogan & Margoliash 1998) applied to automated auditory collection systems (cf. Blumstein *et al.* 2011) are capable of discriminating different species, sexes and groups of animals. Ecologists have deployed acoustic arrays to study a range of taxa including terrestrial (Blumstein *et al.* 2011), marine (Moretti *et al.* 2010; Ward

*Correspondence author. E-mail: paul.conn@noaa.gov

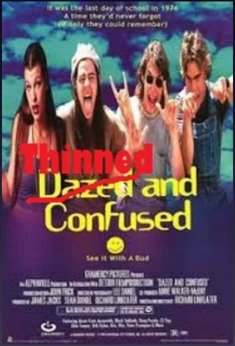
Published 2013. This article is a U.S. Government work and is in the public domain in the U.S.A.

Abundance and distribution modeling

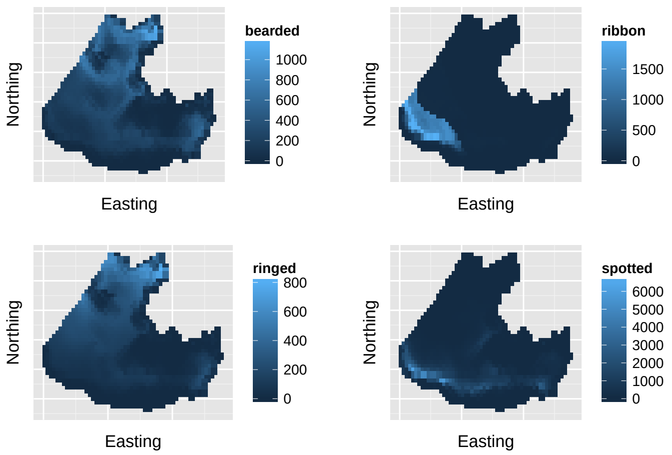
Initial analysis: spatial only, eastern Bering 2012

First step: restrict flights to a “short" window; ignore time

→ “Thinned and confused"
log-Gaussian Cox process



Abundance and distribution modeling



Spatio-temporal seal analysis

Space-time: the final frontier!

Want to allow spatio-temporal autocorrelation in abundance.

Many options [Cressie and Wikle, 2011]!

1. Descriptive vs. dynamical

Spatio-temporal seal analysis

Space-time: the final frontier!

Want to allow spatio-temporal autocorrelation in abundance.

Many options [Cressie and Wikle, 2011]!

1. Descriptive vs. dynamical
2. Strategy for dimension reduction (knot based, spectral decomposition)

Spatio-temporal seal analysis

Space-time: the final frontier!

Want to allow spatio-temporal autocorrelation in abundance. Many options [Cressie and Wikle, 2011]!

1. Descriptive vs. dynamical
2. Strategy for dimension reduction (knot based, spectral decomposition)
3. Boundary conditions (in the case of dynamical S-T models)

Spatio-temporal seal analysis

Space-time: the final frontier!

Want to allow spatio-temporal autocorrelation in abundance.
Many options [Cressie and Wikle, 2011]!

1. Descriptive vs. dynamical
2. Strategy for dimension reduction (knot based, spectral decomposition)
3. Boundary conditions (in the case of dynamical S-T models)

With regards to abundance estimation in particular:

1. Demographic closure

Spatio-temporal seal analysis

Space-time: the final frontier!

Want to allow spatio-temporal autocorrelation in abundance. Many options [Cressie and Wikle, 2011]!

1. Descriptive vs. dynamical
2. Strategy for dimension reduction (knot based, spectral decomposition)
3. Boundary conditions (in the case of dynamical S-T models)

With regards to abundance estimation in particular:

1. Demographic closure
2. Observation model and link function

P. B. Conn, D. S. Johnson, J. M. Ver Hoef, M. B. Hooten, J. M. London, and P. L. Boveng. Using spatio-temporal statistical models to estimate animal abundance and infer ecological dynamics from survey counts. Ecological Monographs, 85:235–252, 2015

Spatio-temporal models for abundance

Temporarily ignoring incomplete detection, how do several different forms of hierarchical spatio-temporal statistical models perform in estimating animal abundance from survey counts $(C_{s,t})$?

1. Descriptive additive space-time (AST) (*)

Spatio-temporal models for abundance

Temporarily ignoring incomplete detection, how do several different forms of hierarchical spatio-temporal statistical models perform in estimating animal abundance from survey counts ($C_{s,t}$)?

1. Descriptive additive space-time (AST) (*)
2. Descriptive spatio-temporal process convolution (STPC) (*)

Spatio-temporal models for abundance

Temporarily ignoring incomplete detection, how do several different forms of hierarchical spatio-temporal statistical models perform in estimating animal abundance from survey counts ($C_{s,t}$)?

1. Descriptive additive space-time (AST) (*)
2. Descriptive spatio-temporal process convolution (STPC) (*)
3. Dynamic open population resource selection (OPRS) (*)

Spatio-temporal models for abundance

Temporarily ignoring incomplete detection, how do several different forms of hierarchical spatio-temporal statistical models perform in estimating animal abundance from survey counts ($C_{s,t}$)?

1. Descriptive additive space-time (AST) (*)
2. Descriptive spatio-temporal process convolution (STPC) (*)
3. Dynamic open population resource selection (OPRS) (*)
4. Descriptive multinomial closed population ideal free model (CPIF)

Spatio-temporal models for abundance

Temporarily ignoring incomplete detection, how do several different forms of hierarchical spatio-temporal statistical models perform in estimating animal abundance from survey counts ($C_{s,t}$)?

1. Descriptive additive space-time (AST) (*)
2. Descriptive spatio-temporal process convolution (STPC) (*)
3. Dynamic open population resource selection (OPRS) (*)
4. Descriptive multinomial closed population ideal free model (CPIF)

* based on Log-Gaussian Cox process

$$
\begin{aligned}
[\mathbf{C}] &= \text{Poisson}(\boldsymbol{\lambda}), \\
\boldsymbol{\lambda} &= \exp\left(\mathbf{o} + \mathbf{H}\boldsymbol{\mu}\right) \\
N_{s,t} &\sim \text{Poisson}(\exp(\mu)) \text{ [posterior prediction]}
\end{aligned}
$$

AST model

Additive space-time (AST) model [similar to Ver Hoef and Jansen, 2007]:

$$\begin{aligned}
\mu_{s,t} &= \mathbf{X}_{s,t}\boldsymbol{\beta} + \eta_s + \gamma_t + \epsilon_{s,t} \\
\boldsymbol{\eta} &= \mathbf{K}\alpha \\
\boldsymbol{\alpha} &\sim \mathcal{N}(0, \tau_\alpha^{-1}) \\
\gamma_t &\sim GMRF - RW2(\tau_\gamma) \\
\epsilon_{s,t} &\sim \mathcal{N}(0, \tau_\epsilon^{-1})
\end{aligned}$$

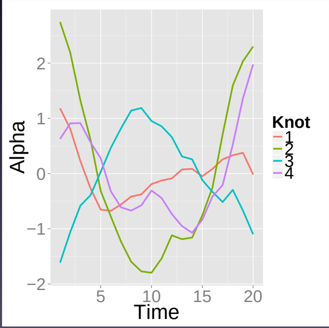
Here, $\mathbf{K}$ is an $S \times m$ matrix with entries $K_{ij} = \mathcal{N}(d_{ij}; 0, \sigma^2)$ - where d_{ij} is the distance from the centroid of the ith sampling unit to the jth knot and σ is set a priori to the average distance between knots.

STPC model

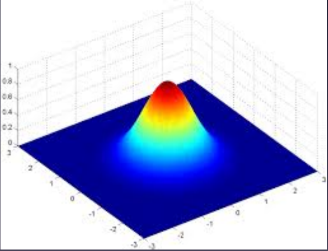
Spatio-temporal process convolution (STPC) model:

$$\begin{aligned} \boldsymbol{\mu} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\kappa} + \boldsymbol{\epsilon} \\ \epsilon_{s,t} &\sim \mathcal{N}(0, \tau_\epsilon^{-1}) \end{aligned}$$

Here, $\boldsymbol{\kappa}$ is spatio-temporal effect induced by dynamic process convolution [Calder et al., 2002]



+



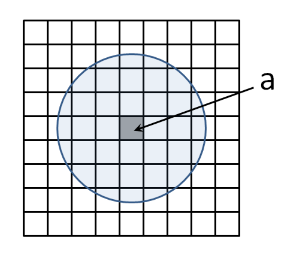
OPRS model

Open population resource selection (OPRS) model:

$$\mu_{s,t} = \begin{cases} \mathbf{X}_1\boldsymbol{\beta} + \boldsymbol{\eta} + \epsilon_1 & \text{if } t = 1 \\ \mathbf{M}_t\boldsymbol{\mu}_{t-1} + \gamma_t + \boldsymbol{\epsilon}_t & \text{if } t > 1. \end{cases}$$

The elements of the transition matrix $\mathbf{M}_t$ are determined via the “weighted distribution" typical in resource selection studies:

$$\begin{aligned} \psi_t^{ab} &= \frac{w_{b,t}\varphi_{a,b}}{\sum_s w_{s,t}\varphi_{a,s}} \\ \varphi_{a,b} &\propto \mathrm{Normal}(d(a,b), \tau_d^{-1}), \\ \log(\mathbf{w}_t) &= \mathbf{X}_t\boldsymbol{\beta}. \end{aligned}$$

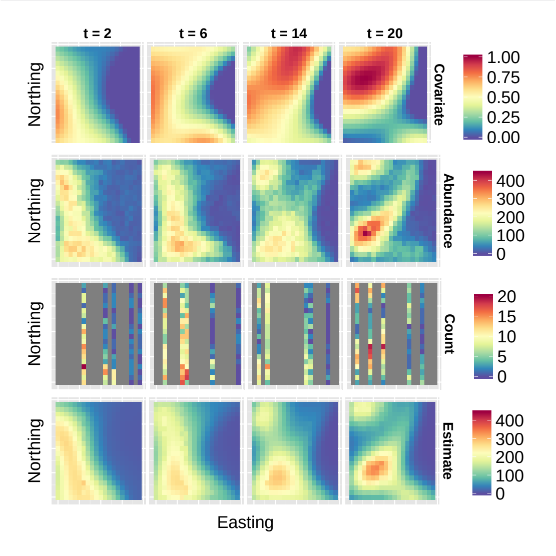


CPIF model

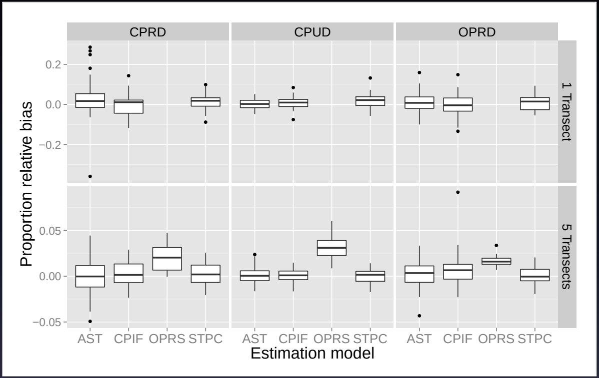
Closed population ideal free (CPIF) model

$$
\begin{aligned}
C_{s,t} &\sim \text{binomial}(N_{s,t}, p_{st}) \\
N_{s,t} &\sim \text{multinomial}(N, \boldsymbol{\pi}) \\
\pi_{i,t} &= \frac{\exp(\omega_{i,t})}{\sum_s \exp(\omega_{s,t})} \\
\boldsymbol{\omega_t} &= \mathbf{o} + \mathbf{X}_t \boldsymbol{\beta} + \boldsymbol{\kappa}_t + \boldsymbol{\epsilon_t} \\
N &\sim 1/N
\end{aligned}
$$

Generic simulated data



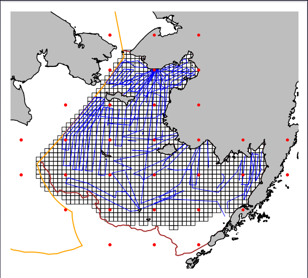
Generic simulated data



- Coverage often poor when different model used for estimation than for simulating data
- Computing hours/sim: ≈ 0.2; CPIF, STPC ≈ 1.5; OPRS 27-72
- Performance similar in a 2nd simulation study emulating seal surveys

Spatio-temporal seal analysis

We're currently working on applying the multinomial model to data from 2012 and 2013 Bering Sea aerial surveys, having added back in detection components (imperfect detection, availability < 1.0, species misidentification.



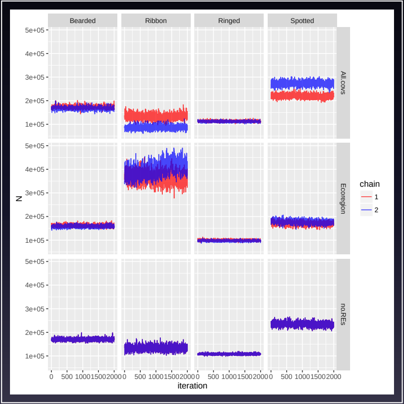
Predictive covariates

- Sea ice concentration (linear, quadratic)
- Distance from southern ice edge
- Distance from 1000m shelf break
- Distance from mainland
- Distance from a 90% sea ice contour

Spatio-temporal seal analysis

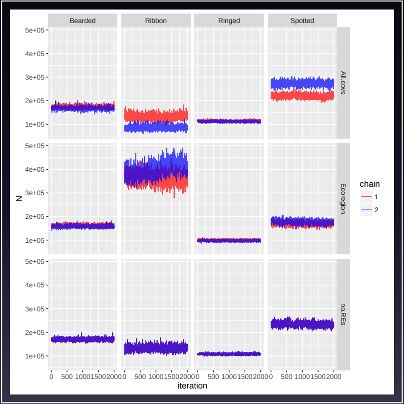
Videos!

Spatio-temporal seal analysis



- Large uncertainty about spotted, ribbon seal abundance in 2012
- Lack of convergence and evidence of multimodality in “complicated models"

Spatio-temporal seal analysis



Much better diagnostics in 2013.

- Large uncertainty about spotted, ribbon seal abundance in 2012
- Lack of convergence and evidence of multimodality in “complicated models"

Reasons to switch to TMB

- Bespoken MCMC sampler takes 7-10 days to run
- Lack of convergence
- Difficulty debugging and conducting adequate model testing
- It would be nice to implement new features quickly (e.g. pup recruitment processes, more flexibility in availability).

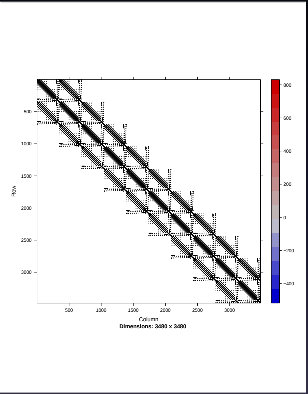
Implementation in TMB

Simple spatio-temporal Poisson model:

$$
\begin{aligned}
C_{stk} &\sim \text{Poisson}(p\tilde{\lambda}_{stk}) \\
\tilde{\boldsymbol{\lambda}}_{st} &= \boldsymbol{\psi}\boldsymbol{\lambda}_{st} \\
\log(\boldsymbol{\lambda_{tk}}) &= \mathbf{o} + \mathbf{X}_{tk}\boldsymbol{\beta}_k + \boldsymbol{\eta}_{tk} \\
\boldsymbol{\eta}_k &\sim MVN(\mathbf{0}, \boldsymbol{\Sigma}_{AR1,k} \otimes \boldsymbol{\Sigma}_{GMRF,k}) \\
p &\sim [p] \\
\boldsymbol{\psi} &\sim [\boldsymbol{\psi}]
\end{aligned}
$$

Implementation in TMB

Simple spatio-temporal Poisson model:



- 3 species, 1600 cells, 10 time steps: >48,000 REs. Okay!
- 3 species, 2500 cells, 10 time steps: >75,000 REs. Error in sparseHessianFun(env, skipFixedEffects = skipFixedEffects) : Memory allocation fail in function 'MakeADHessObject2'

Possible fix: knot-based interpolation such as piece-wise constant (VAST) or predictive process [Banerjee et al., 2008], basis function decomposition

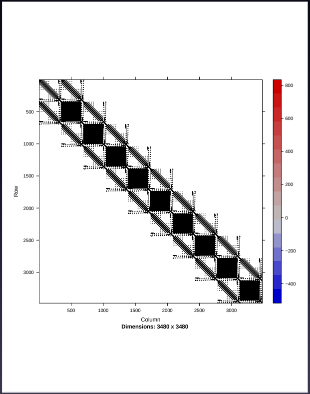
Implementation in TMB

CPIF-like species redistribution model using Poisson approx to multinomial and separable covariance:

$$
\begin{aligned}
C_{stk} &\sim \text{Poisson}(p\tilde{\lambda}_{stk}) \\
\tilde{\boldsymbol{\lambda}}_{st} &= \boldsymbol{\psi}\boldsymbol{\lambda}_{st} \\
\lambda_{stk} &= N_k \pi_{stk} \\
\pi_{stk} &= \frac{\exp(\omega_{stk})}{\sum_s \exp(\omega_{stk})} \\
\boldsymbol{\omega}_{tk} &= \mathbf{o} + \mathbf{X}_{tk}\boldsymbol{\beta}_k + \boldsymbol{\eta}_{tk} \\
\boldsymbol{\eta}_k &\sim MVN(\mathbf{0}, \boldsymbol{\Sigma}_{AR1,k} \otimes \boldsymbol{\Sigma}_{GMRF,k}) \\
p &\sim [p] \\
\boldsymbol{\psi} &\sim [\boldsymbol{\psi}]
\end{aligned}
$$

Implementation in TMB

CPIF-like species redistribution model



25×25 grid, 3 species, $> 18750 REs$: Runs but slow

30×30 grid, 10 time steps, 3 species, $> 27000 REs$: Memory allocation fail

$\rightarrow$ Need to use some form of dimension reduction.

Summary

There are many spatio-temporal modeling options

1. Choice of link function
2. Basis choice
3. Descriptive vs. dynamical

Summary

There are many spatio-temporal modeling options

1. Choice of link function
2. Basis choice
3. Descriptive vs. dynamical

And software options!

1. Simplest: Canned packages using GLMM-like syntax (mgcv, INLA)
2. More complex: Nonlinear and “non-standard" models in TMB
3. Most complex: Fully Bayesian: multiple levels of hierarchy, discrete latent variables, dynamical models

Summary

There are many spatio-temporal modeling options

1. Choice of link function
2. Basis choice
3. Descriptive vs. dynamical

And software options!

1. Simplest: Canned packages using GLMM-like syntax (mgcv, INLA)
2. More complex: Nonlinear and “non-standard" models in TMB
3. Most complex: Fully Bayesian: multiple levels of hierarchy, discrete latent variables, dynamical models

There are a lot of tradeoffs! (Ease of implementation, flexibility, realism)

Acknowledgments

- Funding: NOAA, BOEM
- Many coauthors and partners on parts of this work, including: E. Moreland, M. Cameron, J. Ver Hoef, J. London, M. Hooten, D. Johnson, E. Richmond, S. Dahle, B. McClintock, P. Boveng, I. Trukhovana
- All researchers and technicians who assisted with Arctic seal surveys

Literature Cited

S. Banerjee, A. E. Gelfand, A. O. Finley, and H. Sang. Stationary process approximation for the analysis of large spatial datasets. Journal of the Royal Statistical Society B, 70:825–848, 2008.

C. A. Calder, C. Holloman, and D. Higdon. Exploring space-time structure in ozone concentration using a dynamic process convolution model. In C. Gatssonis, R. E. Kass, A. Carriquiry, A. Gelman, D. Higdon, D. K. Pauler, and I. Verdinelli, editors, Case Studies in Bayesian Statistics 6, pages 165–176. Springer-Verlag, 2002.

P. B. Conn, D. S. Johnson, J. M. Ver Hoef, M. B. Hooten, J. M. London, and P. L. Boveng. Using spatio-temporal statistical models to estimate animal abundance and infer ecological dynamics from survey counts. Ecological Monographs, 85:235–252, 2015.

N. Cressie and C. K. Wikle. Statistics for spatio-temporal data. Wiley, Hoboken, New Jersey, 2011.

B. T. McClintock, E. E. Moreland, J. M. London, S. P. Dahle, G. M. Brady, E. L. Richmond, K. M. Yano, and P. L. Boveng. Quantitative assessment of species identification in aerial transect surveys for ice-associated seals. Marine Mammal Science, 31:1057–1076, 2015.

J. M. Ver Hoef and J. K. Jansen. Space-time zero-inflated count models of harbor seals. Environmetrics, 18: 697–712, 2007.