Using Bayesian state-space models with multiple sampling methods to infer the stock structure of Alaskan cod (*Gadus macrocephalus*) in the Bering Sea

Kelli F. Johnson1\*, André E. Punt1, and Kerim Y. Aydin2

*1School of Aquatic and Fishery Sciences, University of Washington, Box 355020, Seattle, WA 98195-5020, USA  
2Alaska Fisheries Science Center, National Marine Fisheries Service, NOAA, 7600 Sand Point Way N.E., Seattle, WA 98115-6349, USA*

*\*Corresponding author: tel: +1 206 543 4270; fax: +1 206 685 7471; email: kfjohns@uw.edu*

# Introduction

The management of modern fisheries relies on stock assessment models, which provide estimates of population abundance and potential catch, given management objectives. A typical assumption of stock assessment models is that the population dynamics of the harvested species is uniform throughout the area modeled. Arbitrary fisheries management boundaries, typically set for political ease, often decrease the likelihood that metapopulations freely mix within the boundaries at which they are managed (CITATION).

Stocks, or subpopulations, can be defined in almost an infinite number of ways: (1) political boundaries, (2) genetic differentiation, (3) synchronous growth rates or other population dynamics parameters, (4) geographic boundaries, or (5) life-history traits to name a few. The purpose for which one defines a stock often dictates the rules used to define it (CITATION). A priority of fisheries management is to increase yield, whether it be the maximum average yield, maximum sustainable yield, or the maximum economic yield. To ensure the previous it may be imperative that subpopulations of marine fishes are defined based on life history or biological parameters rather than just for the sake of political ease.

In 2010, Pacific cod *Gadus macrocephalus* was the 28th largest fishery by weight in the world, with 63% of catches coming from the United States (US), of which 78% were caught in the eastern Bering Sea (EBS) and the Aleutian Islands (AI) (www.fao.org). In addition to supporting a major commercial fishery, Pacific cod are ecologically important in the EBS and AIs as a predator of pollock, shrimp, and snow crab (CITATION) and as prey for fish, sea birds, and marine mammals, in particular the federally endangered Stellar sea lion *Eumetopias jubatus* (CITATION).

In the Bering Sea Aleutian Island (BSAI) management unit, Pacific cod are managed as a single stock, and have been since the introduction of the Magnusson-Stevens Fishery Conservation and management Act in 1977. Historical justification for their management as a single stock came from tagging studies documenting their migration between the regions and their lack of genetic variation between fish separated by large geographic distances (CITATION). Thanks to advances in genetic techniques and increased sample sizes, genetic differentiation between the two areas is now documented (Spies, 2012). Furthermore, probabilities of recapture in areas within close proximity to tagging locations have been shown to be much higher when Pacific cod were tagged and recaptured in the spawning season rather than during the summer (CITATION), suggesting site fidelity. MAYBE ALSO TALK ABOUT HOW EGGS ARE DEMERSAL and COD RARELY MIGRATE OVER GEOGRAPHIC BARRIERS.

Here we present an ecological assessment of stock structures in EBS Pacific cod from multivariate autoregressive state-space models. Their use in ecology began with time-series data for single populations (Staples, Taper, and Dennis, 2004; Dennis et al., 2006) and later extended into two-dimensional space for movement data (Jonsen, Myers, and Flemming, 2003) and multiple subpopulations (Ward et al., 2009) and species (NEED CITATIONS). Multivariate time-series analysis is one way to detect spatial patterns of correlation and synchrony across multiple locations (CITATION) and can provide inferences on stock structure based on ecological differences. Furthermore, state-space models can simultaneously estimate both process and observation error making them ideal for ecological data. Their ability to partition variation without needing prior estimates of variance or repeated measures facilitates analyses on data sets with missing years or sites within a year increases their applicability.

# Methods

**Data**

The EBS is a vast continental shelf with numerous deep marine canyons along its edge. Its wide shelf and northward currents make it a distinct ecological system from the Gulf of Alaska or the Aleutian Islands.

Data were collected from 1982 to 2012 during the annual EBS shelf bottom trawl survey, a bottom trawl survey used to estimate biomass, abundance, and population structure of groundfish and crab for the North Pacific Management Council. The survey area encompasses approximately 463,000 km2 with sampling stations centered on a 37.04 km square grid pattern. However, more intensive sampling occurred in the areas near the Pribilof Islands and St. Matthew Island. Tows of 30 min (estimated on bottom duration) at speeds of 5.6 km · h-1 were conducted annually at nearly every station (*n* = 367). For each tow, species composition of the catch was determined and catch per unit effort (CPUE) was calculated as the weight of the catch (kg) divided by the area swept (ha) for the given trawl. The survey region was divided into six bathymetric strata bounded by the 50 m, 100 m, and 200 m isobaths and into two geographic strata that separate the northwest and southeast portions of the study area (Figure 1). For more details of the sampling scheme see Hoff (2013).

**Multivariate autoregressive state-space model**

Multivariate autoregressive state-spaces models were used to model each hypothesized stock structure and estimate their respective population parameters. Models were discrete, stochastic, and lacked any age structure. Growth rates were modeled using an exponential function described by

, (1)

where *t* = 1, 2, …, T; *Nt* is the population abundance at time *t*; and exp(*u*) is the deterministic per-unit-abundance growth rate in population density (Dennis *et al*., 1991). Exponential growth can be a valid assumption for depleted populations far from carrying capacity. Pacific cod are a heavily fished species, therefore the biases associated with the assumption of exponential growth are assumed to be minimal in this specific application. Here, process error (*et*) represents stochastic deviations from the long-term trend due to environmental effects. Process error was assumed to be normally distributed with mean of zero and a constant variance. After defining as the growth process for a single subpopulation, the growth process can be expressed as

, (2)

which represents a discrete-time random walk with drift. The model becomes multivariate by extending the equation to model *m* unknown subpopulations with potentially correlated growth rates and process error. Using capital letters for vectors and bold letters for matrices, the equation becomes

, (3)

where *U* is an *m x 1* vector of stochastic growth rates and *ωt* is an *m x 1* vector of serially uncorrelated process error with a mean of zero and *m x m* covariance matrix, ***Q***.

The observed values of log CPUE, *Yt*, are related to the true log abundances, *Xt*, through the following equation

, (4)

where *A* is a vector of length *n*, representing the mean bias between sites, ***Z*** is an *n x m* matrix of 0s and 1s, translating the *m* subpopulation sizes at time *t* into *n* observations at time *t*; and ε*t* is a vector of serially uncorrelated observation errors with a mean of zero and *n x n* covariance matrix **R.** Observation error can include several types of error: sampling error, measurement error, location error, or biases in availability. If ignored, measurement error can create bias in results, but it does not contribute to error in the predictions of future abundances only observations of those abundances.

**Candidate models**

Hypothesized stock structures were determined *a priori* based on the life history of Pacific cod and oceanography of the EBS. Flow through Unimak Pass results in a weak, but persistent, flow along the 50 m isobath, which contributes to approximately one third of the flow through the Bering Strait (0.35 m3/s) (Stabeno, Reed, and Napp, 2002). A lesser amount of the current coming through Unimak Pass results in a current that continues northwestward along the 100 m isobath, although the flow is not temporally consistent in its location (Reed, 1998). The following hypotheses were based on grouping the inner, middle, and outer areas of the continental shelf into two or three distinct groups (Figure 1).

**Inner slope** (*m* = 2): sites along the inner continental shelf (< 50 m) represent a single subpopulation, while sites in waters deeper than 50 m represent an additional subpopulation.

**Outer slope** (*m* = 2): sites along the outer continental shelf (101 - 200 m) represent a single subpopulation, while sites in waters shallower than 100 m represent an additional subpopulation.

**Independent** (*m* = 3): each of the three areas of the continental slope represents its own subpopulation.

Within the above hypotheses, we tested three levels of process variance complexity ((1) equal process variance, (2) unequal and uncorrelated process variance, and (3) unequal and correlated process variance) and one level of observation variance complexity (equal). We did not test for unequal and uncorrelated or unequal and correlated observation variances because observations were collected independently at each site.

**Priors**

The prior for observation error was specified as a beta distribution, providing a lower bound of zero. The shape and scale parameters for the prior were informed by analyzing sampling variation for all stations within the North Pacific Fishery Management Council that had multiple tows in a given year.

Given that measurements were not available for each subpopulation, only observations of those subpopulations, the initial simulated log cod CPUE (*X1*) values were drawn from random normal distributions with a mean of zero and a standard deviation of 0.01. NEED A BIOLOGICAL BASIS FOR THIS PRIOR.

Much of the process variability will arise from deviations in recruitment. Thus, the prior for ωwas specified as a gamma distribution. The gamma distribution restricts the variance from being negative and leads to low probability of high (>5, given the paramaterization) CITE SOMETHING HERE REGARDING PROCESS ERROR OF MARINE FISHES.

Priors were the same for all models except when ***Q*** was unequal and correlated. When the process variances were correlated we used the Wishart distribution. The Wishart distribution is guaranteed to produce positive definite draws, provided that ƞ ≥ *p,* where the smaller ƞ is, the more the variability (Kollo and Ruul, 2003). The Wishart was parameterized with an ƞ = *p* + 1 for all models.

Under the assumption that the priors for each parameter are independent, given that a Wishart distribution is used when variances are correlated, the priors were constructed individually and the joint prior is the product of the priors for each parameter:

, (5)

where *p(*θik) is the prior for parameter *i* (i = 1, 2, 3, …, S; where *S* is the number of parameters in model *k*) in the vector θk.

**Estimation**

The probability that a given hypothesis is true (θi) is conditioned on the data, (θi|data), and is given by

, (6)

where *p*(θi) is the prior probability density for the hypothesis *i* and *L(data|*θ*i*) is the likelihood function for the data evaluated at θ*i*. Given the problem at hand, the likelihood function was evaluated as:

(7)

also known as the normal density function minus parameters that remain constant.

**Sampling from the posterior**

It can be extremely difficult if not impossible to obtain an analytical solution of (6), thus numerical procedures are often necessary. To overcome this limitation most analyses use either Markov Chain methods (Gibbs sampler or Markov Chain Monte Carlo) or importance sampling.

With importance sampling or sampling importance resampling (SIR) one defines some density that closely approximates the posterior (often the prior). An importance function,

, (8)

is then used to weight the samples so that the approximation becomes as accurate as possible with the least amount of samples. Here, the importance function was set as samples from the joint prior distribution, thus the importance ratio becomes the likelihood of the data given the parameters.

With Markov Chain sampling the posterior is …

**Model fitting**

All models were fit with three chains with each chain having a final sample size of 1000 iterations. Convergence of the algorithms was checked using trace plots of the parameters, plots showing correlations between parameters, and Gweke statistics for convergence (CITATION).

**Model selection**

Bayes Factor, or the probability of alternative models, provides an estimate of the posterior probability for a given hypothesis (CITATION).

(9)

Deviance information criterion (DIC) is useful in Bayesian model selection when the posterior distribution is approximately multivariate normal and results were obtained using MCMC sampling. The formula is as shown,

, (10)

where smaller values indicate a better fit of the model to the data.

# Results

Growth rates between Pacific cod in the inner and outer slope regions appears to be similar, with both regions centered at log zero indicating a lack of any annual trend in their respective abundances (Figure 2). The more interesting result appears in the differentiation of process errors between the two regions. Pacific cod in the inner slope appear to be subject to higher variation in recruitment than Pacific cod located on the outer slope (Figure 2).

Observation error was quite high for all three models and degenerate, or unestimable, for the model hypothesizing Pacific cod on the outer shelf act independently from the two inner shelf regions, with each area displaying unique process variance (Figure not shown). PERHAPS TALK ABOUT OVERALL CONVERGENCE. Results were similar between MCMC and SIR samplers, though only results from the MCMC routine are shown.

The model with the most support from the data was one that grouped the middle and outer shelf areas as a single subpopulation with a unique process variance compared to the inner slope.

# Discussion

ANDRÉ – I obviously have not had time to properly think about the results, let alone a discussion section. My apologies. Thusly, I have bulleted a few ideas to think about in the future, which are written below.

* The models were able to capture the mean log CPUE for each time series but failed to track the trends in temporal variation of Pacific cod in the EBS.
* Higher variation in recruitment could perhaps be due to the tendency of Pacific cod to lay their eggs in waters less than 100 m and the timing of the surveys.
* Cope and Punt (2009) was unable to differentiate between the middle – inner and middle – outer grouping.
* A stronger current arising from Unimak pass flows along the 50 m isobaths the current along the 100 m isobaths inconsistently located between years.
* Survey strata are designed to minimize the variances of population and biomass estimates for groundfishes across the different oceanographic domains encompassed in the EBS (Bakkala, 1993).

# Acknowledgements

We would like to thank Eric J. Ward, Elizabeth E. Holmes, and Mark D. Scheuerell for teaching their applied time-series class at the University of Washington, which this paper originated from.

# References

Bakkala, R.G. 1993. Structure and historical changes in the groundfish complex of the eastern Bering Sea. U.S. Dep. Commer., NOAA Tech. Rep. NMFS 114, 91 p.

Dennis, B., Munholland, P.L., and Scott, J.M. 1991. Estimation of growth and extinction parameters for endangered species. Ecological Monographs 61:115-143.

Dennis, B., Ponciano, J.M., Lele, S.R., Taper, M.L, and Staples, D.F. 2006. Estimating density dependence, process noise, and observation error. Ecological Monographs 76:323-341.

Hoff, G.R. 2013. Results of the 2012 eastern Bering Sea upper continental slope survey of groundfish and invertebrate resources. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC-258, 268 p.

Jonsen, I.D., Myers, R.A., and Flemming, J.M. 2003. Meta-analysis of animal movement using state space models. Ecology 84:3055-3063.

Kollo, T., and Ruul, K. 2003. Approximations to the distribution of the sample correlation matrix. Journal of Multivariate Analysis 85(2):318-334.

Reed, R.K. 1998. Confirmation of a convoluted flow over the southeastern bering sea shelf. Continental Shelf Research 18(1):99-103.

Stabeno, P.J., Reed, R.K., and Napp, J.M. 2002. Transport through Unimak Pass, Alaska. Deep Sea Research Part II: Topical Studies in Oceanography 49(26):5919-5930.

Staples, D.F., Taper, M.L., and Dennis, B. 2004. Estimating population trend and process variation for PVA in the presence of sampling error. Ecology 85:923-929.

Spies, I. 2012. Landscape genetics reveals population subdivision in Bering Sea and Aleutian Islands Pacific Cod. Transactions of the American Fisheries Society 141:1557-1573.

Ward, E. J., Chirakkal, H., González-Suárez, M., Aurioles-Gamboa, D., Holmes, E. E., and Gerber, L. 2010. Inferring spatial structure from time-series data: Using multivariate state-space models to detect metapopulation structure of California sea lions in the Gulf of California, Mexico. Journal of Applied Ecology 47:47-56.

# Tables

Table 1. Prior distributions for model parameters.

|  |  |  |
| --- | --- | --- |
| Parameter | Process error structure | Prior distribution |
| Um |  | Normal (0,0.2) |
| An |  | Normal (0,1) |
| X1 |  | Normal (0,0.01) |
| R |  | Gamma (0.01, 0.01) |
| Q | constant | Gamma (0.01, 0.01) |
| **Q(m,m)** | diagonal & uncorrelated | Gamma (0.01, 0.01) |
| **Q** | diagonal & correlated | Wishart (m + 1, diag(0.01)) |

Table 2. Deviance information criteria (DIC) for the candidate model set. The model in bold signifies the model with the lowest DIC.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | DIC | | Bayes Factor | |
|  | MCMC | SIR | MCMC | SIR |
| Outer slope – equal Q | 39154.12 |  |  | 0.30 |
| Outer slope – unequal Q | 39151.62 |  |  | 0.31 |
| **Inner slope – equal Q** | **38857.88** |  |  | **0.4** |
| Inner slope – unequal Q | 38854.88 |  |  | 0.3 |

# Figures

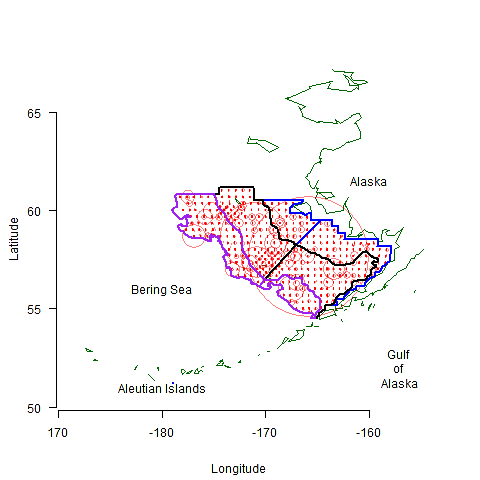
****

Figure 1. Eastern Bering Sea bottom trawl geographic strata, divided by the 50, 100, and 200 m isobaths (blue, black, and purple, respectively). Circles represent mean yearly CPUE per station, where the size of the circle is proportional to the magnitude of the CPUE. The east west delineation of the three isopleths indicate the Pacific Management Council strata and have no relevance in the study at hand.

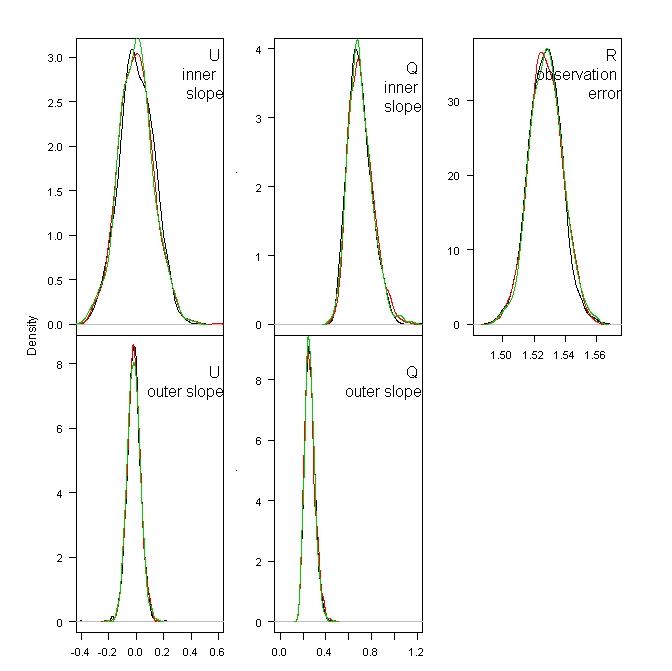


Figure 2. Posterior distributions for the five parameters in the best model, determined using DIC. Colours represent individual chains from the MCMC output, with each chain containing 1000 iterations.

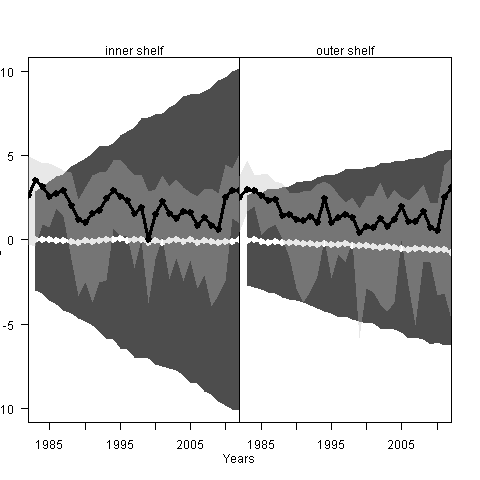


Figure 3. Posterior predictive plots from the best fitting model as determined by DIC. Dark gray polygons indicate the 5% and 95% credibility intervals, while the light gray polygons display the 5% and 95% quantiles for the observed log CPUE in the inner and outer shelves, respectively. White and black dots display the median trajectories for the posterior predictive and observed time series in the two areas.

# Extra text

**Potential hypotheses to test**

Panmictic (*m* = 1): all sites were part of one single large metapopulation, where each sites is an independent observations of a single population trajectory.

Inner slope (*m* = 2): sites along the inner continental shelf (< 50 m) represent a single metapopluation, while sites in waters deeper than 50 m represent an additional metapopulation.

Middle slope (*m* = 3): sites were grouped according to the inner, middle, and outer continental shelf zones, corresponding to depths of < 50 m, 50 – 100 m, and 101 – 200 m respectively.