FATE Proposal Cover Sheet

1. Title – Improving salmon population forecasts by combining environmental drivers, variable age composition and spatial structure into hierarchical models

2. Principal Investigator – Mark Scheuerell

3. Principal Investigator Institution – NOAA/NMFS/NWFSC

4. Associated NMFS Fisheries Science Center – NWFSC

5. Research Priority – 6

6. Project Duration – 2 years

7. Total Funding Request – $149,050

8. Year 1 Request (all institutions) – $108,200

9. Year 2 Request (all institutions) – $40,850

10. Lead NMFS Investigator – Mark Scheuerell

11. Lead NMFS Investigator organization – NWFSC/Fish Ecology/Applied Ecology

12. List all other Co-PIs and Institutions – James Thorson (NOAA/NMFS/NWFSC)

13. PI email – mark.scheuerell@noaa.gov

14. PI phone – (206) 302-2437

15. PI address – NOAA NWFSC, 2725 Montlake Blvd E, Seattle, WA 98112

16. NMFS PI email – mark.scheuerell@noaa.gov

17. NMFS PI phone – (206) 302-2437

18. NMFS PI address – NOAA NWFSC, 2725 Montlake Blvd E, Seattle, WA 98112

19. List institutional breakdown of Year 1 budget (NOAA/NMFS – $108,200)

20. List institutional breakdown of Year 2 budget (NOAA/NMFS – $40,850)

Improving salmon population forecasts by combining environmental drivers, variable age composition and spatial structure into hierarchical models

Mark D. Scheuerell[1] and James T. Thorson[2]

[1]*Fish Ecology Division and* [2]*Fishery Resource Analysis and Monitoring Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Seattle, WA 98112*

Background

Scientists are increasingly responding to calls to improve forecasts of future ecological responses to perturbations such as climatic change, biological invasions, and resource exploitation (Clark et al. 2001, Carpenter 2002). To be useful to decision makers weighing the risks and benefits of different actions, ecological forecasts must combine two critical elements: (1) knowledge of underlying mechanisms necessary to refine the accuracy of predictions, and (2) a complete characterization of the associated uncertainty (precision) in the forecast (Clark et al. 2001). Often, however, forecasters, managers, and the public focus largely on accuracy and pay too little attention to the implications of uncertainty (Pielke and Conant 2003). Uncertainty arises from sampling or measurement error, estimation of parameters from a finite sample of data, (mis)specification of a model to represent unknown dynamics, and environmental process noise inherent in the system itself (Harwood and Stokes 2003). Hierarchical or state-space models have proven to be powerful tools for estimating the contributions of these various sources of uncertainty (Newman et al. 2006, Cressie et al. 2009). Likewise, Bayesian methods offer a flexible framework for fitting these complex models to data and communicating probabilistic results in an intuitive form (Punt and Hilborn 1997, Clark 2005, Stewart et al. 2013).

Fisheries science has often been at the forefront of applied ecology in developing forecasting methods (Harwood and Stokes 2003); indeed, a central goal of the FATE Program is to integrate process-based information into models to help predict population or ecosystem dynamics. For example, integrated population models (Maunder and Punt 2013) are used to forecast population changes while propagating uncertainty from multiple data sources and including environmental drivers (Methot and Wetzel 2013). In the context of Pacific salmon (*Oncorhynchus* spp.), the goal of forecasting is often to predict adult spawning run size at the population, watershed, or larger scales, typically one generation or less in advance (e.g., 1-3 years). Such short-term forecasts inform harvest policies (PFMC 2012) and provide an early warning of declines that may trigger mitigation actions for ESA-listed salmon ESUs (NMFS 2013). Moreover, salmon are keystone species (Helfield and Naiman 2006) and trends in their abundance are important indicators of the status of coastal ecosystems (Levin and Schwing 2011).

Historically, most approaches to salmon run forecasting have fallen into two broad categories, which we refer to as "biological-oceanographic" and "sibling-based" methods. Biological-oceanographic forecasts are based on the well-documented role of ocean climate as a driver of salmon survival during marine residence (e.g., Beamish 1993, Mantua et al. 1997). These methods seek to predict adult abundance, marine survival, or recruits per spawner based on statistical relationships with oceanographic indicators such as coastal upwelling (CUI) or the Pacific Decadal Oscillation (PDO) (e.g., Logerwell et al. 2003, Scheuerell and Williams 2005, Peterman et al. 2009, Rupp et al. 2012). Forecasting models based on retrospective correlations, however, tend to fail without warning (Nickelson 1986). Forecasts can be improved by focusing on environmental drivers at appropriate spatial scales (Peterman et al. 2009, Schirripa et al. 2009) or mechanistically

linked to biological processes (Burke et al. 2013), or by adopting methods such as dynamic linear models that account for autocorrelation and allow the effects of drivers to vary through time (Scheuerell and Williams 2005). Nevertheless, this class of models generally ignores age structure, and instead the response variables are aggregated across calendar years or cohorts. *Post hoc* assumptions about adult age structure are then required to translate, say, a forecast of recruits per spawner into adult run size in a given year.

Sibling-based approaches, by contrast, focus on adult age structure to predict the future abundance of older adult age classes from observed returns of younger age classes, based on the conventional wisdom that cohort strength is largely determined during the first year of marine residence (Peterman 1982, Beamish and Mahnken 2001). This is similar to the projection of recruitment estimates during forecasts under assessment models (Methot and Wetzel 2013). However, forecasts based on these so-called sibling regressions, where the slope between the abundance of consecutive age classes within a cohort provides an estimate of the relative frequencies of those age classes, may also fail because age composition is not fixed over time, thus violating the assumption that the slope estimated from past data will remain constant in the future. For example, a large return of jacks (precocious mature males) might indicate high marine survival that portends large returns of older fish in future years, or it might simply reflect an unusually high prevalence of early male maturation in that cohort. Hybrid methods that supplement sibling-based predictions with historical averages offer improved forecasts in some cases (Haeseker et al. 2007, 2008), but they still do not explicitly account for temporal variation in age structure.

There is growing evidence that temporal fluctuations in the abundance and demographic rates of salmon populations are spatially correlated at scales ranging from watersheds to ocean basins, likely reflecting large-scale climatic forcing (Peterman et al. 1998, Hare et al. 1999, Mueter et al. 2002, Pyper et al. 2005). This synchrony reduces the stability of metapopulations or "portfolios" (Schindler et al. 2010, Moore et al. 2010, Thorson et al. 2013), but it could be used to improve forecasts by accounting for inter-population correlation structure rather than modeling each population in isolation (Su et al. 2004).

We propose to develop an integrated approach to salmon population forecasting that synthesizes the previously disparate methods summarized above. This parallels similar efforts in integrated assessment models for marine species, which increasingly incorporate environmental and age-structured information (Mäntyniemi et al. 2013). Like sibling-based models, our proposed model takes advantage of information contained in the abundance of previously observed adult age classes, but our state-space framework avoids the assumption of constant age structure by allowing the age distribution to vary through time. Our model will use data on juvenile (smolt) abundance, where available, to provide additional information on cohort size. We will include biological-oceanographic information through effects of environmental drivers on marine survival and adult age distribution. We will also incorporate hierarchical structures to represent spatial synchrony in survival and age parameters. Specifically, we propose to address the following questions:

(1) How much can we improve forecast performance by incorporating environmental drivers of marine survival and maturation rates?

(2) How does the inclusion of juvenile abundance and time-varying age structure improve forecast performance relative to sibling-based models that use only adult data and assume constant age structure?

(3) Does a multi-population model that includes spatial synchrony in demographic rates provide better forecasts than independent single-population models?

Approach

We view this project as consisting of three general steps to address our specific questions: (1) model development and software implementation, (2) data compilation, and (3) model fitting and simulation. Although we envision these activities proceeding iteratively and in parallel, we describe each separately.

Step 1: Develop an age-structured state-space forecasting model

The heart of our proposed forecasting approach is a simple age-structured model that projects a cohort of juveniles (smolts) into adults of various of ages returning to spawn some years later based on cohort-specific smolt-to-adult survival and age-at-return probabilities (Fig. 1). We will use a state-space framework, in which the demographic parameters (i.e., logit-transformed survival and additive log ratio-transformed age-at-return probabilities) jointly follow a multivariate logistic normal distribution at each time t (Aitchison 2003). The mean vector can include environmental drivers as regression terms (e.g., effects of ocean climate indicators on marine survival or the probability of returning at a given age). Multivariate normal process errors will allow us to model correlations between demographic rates that are not fully captured by environmental effects (e.g., a high marine survival in cohort t might be associated with a high probability of early maturation). Process errors will also include a first-order autoregressive component, allowing past patterns of autocorrelation to inform future predictions. The observation model for the vector of adult counts at each age in each cohort could be Poisson or multinomial, given the initial number of smolts and their survival and age-at-return probabilities.

We will fit models to data using Bayesian Markov Chain Monte Carlo (MCMC) techniques, with noninformative priors specified for all parameters (Gelman 2004). A particular challenge is estimating survival and age-at-return for those cohorts whose older age classes have not yet returned. To address this, we will use particle filter MCMC, an advanced algorithm well suited to state-space models (Andrieu et al. 2010, Knape and de Valpine 2012), implemented using R (R Development Core Team 2013) and STAN (Stan Development Team 2013).

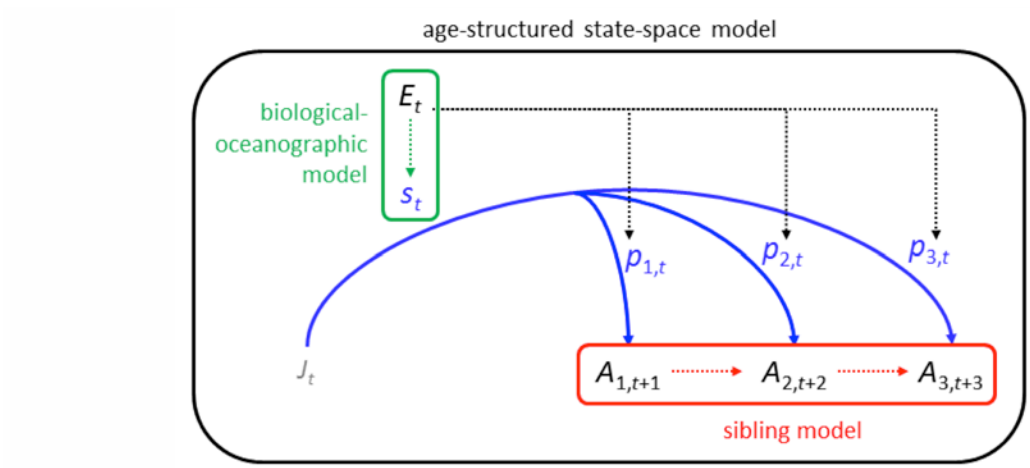

Figure 1. Schematic of the proposed age-structured, state-space forecasting model showing the integration of both biological-oceanographic (green box) and sibling models (red box). Transition probabilities (blue) correspond to unknown, time-varying states, which are a combination of cohort-specific marine survival (s_t) and probability of returning at age a ($p_{a,t}$), both of which may depend on environmental drivers (E_t). Model parameters will be estimated by fitting to observed time series of age-specific adult ($A_{a,t}$) and juvenile abundance (J_t), where available.

Step 2: Compile data on salmon populations and environmental drivers

We will begin our modeling with data for ESA-listed populations already compiled in a publicly accessible database maintained by NMFS[1]. This includes information on adult run size, age composition, and terminal harvest for a range of populations of Pacific salmon and steelhead. Specifically, we will focus on Snake River Spring/Summer Chinook salmon and Oregon coast coho salmon because they have been the focus of previous efforts, and we have additional data on juvenile outmigrants (smolts) for those populations as well. In addition, we will work with federal, state, and tribal agencies to obtain data for other, exploited populations (e.g., Bristol Bay sockeye salmon) with the goal of providing operational forecasts for harvest management.

We already have some time series of environmental drivers used in the existing literature (see Background). Our preliminary indicators of ocean climate include monthly PDO, monthly CUI, monthly North Pacific Gyre Oscillation (NPGO). We do not intend this to be an exhaustive list of potential environmental drivers, so we will work with regional scientists to assemble additional data sets related to marine food webs (Burke et al. 2013), as well as freshwater conditions that are thought to influence smolt-to-adult survival (Petrosky and Schaller 2010).

Step 3: Fit alternative models and compare forecast performance

To address each of our specific research questions, we will compare the retrospective forecast performance of alternative submodels or extensions of the general model described above by iteratively fitting them to subsets of the data and then forecasting subsequent observations.

(1) To quantify the utility of environmental indicators in forecasting, we will compare models with and without covariate effects on marine survival and age-at-return probabilities, and use model averaging to develop an ensemble forecast. We will evaluate retrospective forecasts by iteratively fitting the models to years 1:T of the observed time series (i.e., smolt and adult counts and environmental indicators), where T ranges from some minimal series length (e.g., 10 years) to the full length of the dataset. Forecasts will then be generated by simulating from the posterior distributions of $A_{k,T+1}$, ..., $A_{k,T+K}$, where K is the maximum ocean age of adults (note that the number of age classes that can be predicted declines from K in year $T + 1$ to one in year $T + K$). We will use a variety of metrics to compare forecast skill, such as root mean squared error (RMSE), posterior predictive deviance, and the predicted vs. observed risk of abundance falling below a specified threshold (Murphy 1993, Gelman 2004).

(2) To quantify the information gained by including juvenile abundance data, we will compare the general model with a variant fitted to adult abundance only. In the latter case marine survival cannot be estimated, but the process model will still allow the relative adult age proportions to vary through time. Likewise, we will also include models where these proportions are time-invariant (i.e., the process error is zero), with and without juvenile data. The latter case is essentially a Bayesian version of the standard sibling regression model. These comparisons will be restricted to populations where juvenile data are available. Forecast performance will be assessed as described for question 1.

(3) To assess the usefulness of spatially correlated population dynamics for forecasting, we will expand our model to include multiple populations. Each population will follow the dynamics shown in Fig. 1, but the underlying parameters governing the process model for survival and age structure, including environmental covariate effects, will be treated as random effects drawn from common distributions. This allows parameter estimates to "share information" across

[1] Data available at https://www.webapps.nwfsc.noaa.gov/apex/f?p=261:home:0

multiple stocks, potentially increasing the accuracy and precision of predictions (Cressie et al. 2009). We will compare the forecast performance of this hierarchical model with that of the standard model where population-level parameters are independent. In addition, the estimated variance components can be used to compute an index of among-population synchrony in demographic rates that provides another measure of the information gained when populations are analyzed together rather than separately (Thorson et al. 2013). For these comparisons, we will focus on populations of particular species within regional domains that are expected to share common environmental drivers.

Benefits

Our proposed modeling approach is novel in that it unites two longstanding traditions of salmon forecasting within an integrated population-dynamics framework well suited for propagating uncertainty about environmental and spatial relationships. Our model builds on the emerging literature linking oceanographic conditions to marine survival and maturation rates of salmon, while also taking advantage of the demographic information contained in abundance estimates for observed adult age classes to provide synthetic forecasts that fully quantify the uncertainty associated with parameter estimation, observation error, and underlying year-to-year variability in demographic rates. Such forecasts, presented as probability distributions of possible abundance given current environmental conditions, directly support fishery management decisions (e.g., setting preseason catch limits) and decisions based on risk to ESA-listed populations (NMFS 2013), and thus contribute directly to the goals of the FATE program. In addition, forecasts that explicitly incorporate covariation among groups of populations may be useful in designing management portfolios that address the role of spatial structure in fisheries management or recovery planning. Salmon are important indicators of ecosystem status in the California Current Large Marine Ecosystem Integrated Ecosystem Assessment (Levin and Schwing 2011), so improved population forecasts linked to broader ecological indicators will help inform the IEA process. Our proposed methodology also helps to bridge the gap between approaches traditionally used for salmon and the integrated assessment methods developed for marine fish stocks (Fournier and Archibald 1982, Maunder and Punt 2013). We are actively involved in assessments for salmon and West Coast groundfish, and we will work closely with other scientists in these communities.

Deliverables

We will produce adult run forecasts for the populations in our analysis, for the year following the most recent year of data. The dataset of salmon and environmental time series that we compile will be archived. Code for our models will be also provided to facilitate future development and specific applications to new data. We will report our findings at the FATE meeting and one other scientific conference, and in at least two peer-reviewed publications, one of which will be a review synthesizing traditional approaches to salmon forecasting and integrated population assessment.

Literature Cited

Aitchison, J. 2003. The Statistical Analysis of Compositional Data. Blackburn Press, Caldwell, N.J.

Andrieu, C., A. Doucet, and R. Holenstein. 2010. Particle Markov chain Monte Carlo methods. Journal of the Royal Statistical Society Series B-Statistical Methodology 72:269–342.

Beamish, R. 1993. Climate and exceptional fish production off the West Coast of North America. Canadian Journal of Fisheries and Aquatic Sciences 50:2270–2291.

Beamish, R. J., and C. Mahnken. 2001. A critical size and period hypothesis to explain natural regulation of salmon abundance and the linkage to climate and climate change. Progress in Oceanography 49:423–437.

Burke, B. J., W. T. Peterson, B. R. Beckman, C. Morgan, E. A. Daly, and M. Litz. 2013. Multivariate models of adult Pacific salmon returns. Plos One 8.

Carpenter, S. R. 2002. Ecological futures: Building an ecology of the long now. Ecology 83:2069–2083.

Clark, J. S. 2005. Why environmental scientists are becoming Bayesians. Ecology Letters 8:2–14.

Clark, J. S., S. R. Carpenter, M. Barber, S. Collins, A. Dobson, J. A. Foley, D. M. Lodge, M. Pascual, R. Pielke, W. Pizer, C. Pringle, W. V. Reid, K. A. Rose, O. Sala, W. H. Schlesinger, D. H. Wall, and D. Wear. 2001. Ecological forecasts: An emerging imperative. Science 293:657–660.

Cressie, N., C. A. Calder, J. S. Clark, J. M. V. Hoef, and C. K. Wikle. 2009. Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. Ecological Applications 19:553–570.

Fournier, D., and C. P. Archibald. 1982. A general theory for analyzing catch at age data. Canadian Journal of Fisheries and Aquatic Sciences 39:1195–1207.

Gelman, A. 2004. Bayesian data analysis. Chapman & Hall/CRC, Boca Raton, Florida.

Haeseker, S. L., B. Dorner, R. M. Peterman, and Z. Su. 2007. An improved sibling model for forecasting chum salmon and sockeye salmon abundance. North American Journal of Fisheries Management 27:634–642.

Haeseker, S. L., R. M. Peterman, and Z. Su. 2008. Retrospective evaluation of preseason forecasting models for sockeye and chum salmon. North American Journal of Fisheries Management 28:12–29.

Hare, S. R., N. J. Mantua, and R. C. Francis. 1999. Inverse production regimes: Alaska and West Coast Pacific salmon. Fisheries 24:6–14.

Harwood, J., and K. Stokes. 2003. Coping with uncertainty in ecological advice: lessons from fisheries. Trends in Ecology & Evolution 18:617–622.

Knape, J., and P. de Valpine. 2012. Fitting complex population models by combining particle filters with Markov chain Monte Carlo. Ecology 93:256–263.

Levin, P. S., F. B. Schwing. (Eds.) 2011. Technical background for an integrated ecosystem assessment of the California Current: Groundfish, salmon, green sturgeon, and ecosystem health. U.S. Dept. of Commerce, NOAA Tech. Memo., NMFS-NWFSC-109.

Logerwell, E. A., N. Mantua, P. W. Lawson, R. C. Francis, and V. N. Agostini. 2003. Tracking environmental processes in the coastal zone for understanding and predicting Oregon coho (*Oncorhynchus kisutch*) marine survival. Fisheries Oceanography 12:554–568.

Mantua, N. J., S. R. Hare, Y. Zhang, J. M. Wallace, and R. C. Francis. 1997. A Pacific interdecadal climate oscillation with impacts on salmon production. Bulletin of the American Meteorological Society 78:1069–1079.

Mäntyniemi, S., L. Uusitalo, H. Peltonen, P. Haapasaari, and S. Kuikka. 2013. Integrated, age-structured, length-based stock assessment model with uncertain process variances, structural

uncertainty, and environmental covariates: case of Central Baltic herring. Canadian Journal of Fisheries and Aquatic Sciences:1–10.

Maunder, M. N., and A. E. Punt. 2013. A review of integrated analysis in fisheries stock assessment. Fisheries Research 142:61–74.

Methot, R. D., and C. R. Wetzel. 2013. Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. Fisheries Research 142:86–99.

Moore, J. W., M. McClure, L. A. Rogers, and D. E. Schindler. 2010. Synchronization and portfolio performance of threatened salmon. Conservation Letters 3:340–348.

Mueter, F. J., R. M. Peterman, and B. J. Pyper. 2002. Opposite effects of ocean temperature on survival rates of 120 stocks of Pacific salmon (*Oncorhynchus* spp.) in northern and southern areas. Canadian Journal of Fisheries and Aquatic Sciences 59:456–463.

Murphy, A. 1993. What is a good forecast? An essay on the nature of goodness in weather forecasting. Weather and Forecasting 8:281–293.

National Marine Fisheries Service (NMFS). 2013. Endangered Species Act Section 7(a)(2) Biological Opinion: Supplemental consultation on remand for operation of the Federal Columbia River power system. U.S. Department of Commerce, NWR-2013-9562.

Newman, K. B., S. T. Buckland, S. T. Lindley, L. Thomas, and C. Fernandez. 2006. Hidden process models for animal population dynamics. Ecological Applications 16:74–86.

Nickelson, T. 1986. Influences of upwelling, ocean temperature, and smolt abundance on marine survival of coho salmon (*Oncorhynchus kisutch*) in the Oregon Production Area. Canadian Journal of Fisheries and Aquatic Sciences 43:527–535.

Pacific Fishery Management Council (PFMC). 2012. Preseason Report I: Stock abundance analysis and environmental assessment part 1 for 2012 ocean salmon fishery regulations. Pacific Fishery Management Council, Portland, Oregon

Peterman, R. 1982. Model of salmon age structure and its use in preseason forecasting and studies of marine survival. Canadian Journal of Fisheries and Aquatic Sciences 39:1444–1452.

Peterman, R. M., B. J. Pyper, F. J. Mueter, S. L. Haeseker, Z. Su, and B. Dorner. 2009. Statistical models of Pacific salmon that include environmental variables. Page 2008 American Fisheries Society Symposium.

Peterman, R. M., B. J. Pyper, M. F. Lapointe, M. D. Adkison, and C. J. Walters. 1998. Patterns of covariation in survival rates of British Columbian and Alaskan sockeye salmon (*Oncorhynchus nerka*) stocks. Canadian Journal of Fisheries and Aquatic Sciences 55:2503–2517.

Petrosky, C. E., and H. A. Schaller. 2010. Influence of river conditions during seaward migration and ocean conditions on survival rates of Snake River Chinook salmon and steelhead. Ecology of Freshwater Fish 19:520–536.

Pielke, R. A., and R. T. Conant. 2003. Best practices in prediction for decision-making: Lessons from the atmospheric and earth sciences. Ecology 84:1351–1358.

Punt, A. E., and R. Hilborn. 1997. Fisheries stock assessment and decision analysis: The Bayesian approach. Reviews in Fish Biology and Fisheries 7:35–63.

Pyper, B. J., F. J. Mueter, and R. M. Peterman. 2005. Across-species comparisons of spatial scales of environmental effects on survival rates of Northeast Pacific salmon. Transactions of the American Fisheries Society 134:86–104.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Rupp, D. E., T. C. Wainwright, P. W. Lawson, and W. T. Peterson. 2012. Marine environment-based forecasting of coho salmon (*Oncorhynchus kisutch*) adult recruitment. Fisheries Oceanography 21:1–19.

Scheuerell, M. D., and J. G. Williams. 2005. Forecasting climate-induced changes in the survival of Snake River spring/summer Chinook salmon (*Oncorhynchus tshawytscha*). Fisheries Oceanography 14:448–457.

Schindler, D. E., R. Hilborn, B. Chasco, C. P. Boatright, T. P. Quinn, L. A. Rogers, and M. S. Webster. 2010. Population diversity and the portfolio effect in an exploited species. Nature 465:609–612.

Schirripa, M. J., C. P. Goodyear, and R. M. Methot. 2009. Testing different methods of incorporating climate data into the assessment of US West Coast sablefish. ICES Journal of Marine Science 66:1605–1613.

Stan Development Team. 2013. Stan: A C++ Library for Probability and Sampling, Version 1.3. http://mc-stan.org/.

Stewart, I. J., A. C. Hicks, I. G. Taylor, J. T. Thorson, C. Wetzel, and S. Kupschus. 2013. A comparison of stock assessment uncertainty estimates using maximum likelihood and Bayesian methods implemented with the same model framework. Fisheries Research 142:37–46.

Su, Z. M., R. M. Peterman, and S. L. Haeseker. 2004. Spatial hierarchical Bayesian models for stock-recruitment analysis of pink salmon (*Oncorhynchus gorbuscha*). Canadian Journal of Fisheries and Aquatic Sciences 61:2471–2486.

Thorson, J. T., M. D. Scheuerell, E. R. Buhle, and T. Copeland. 2013. Spatial variation buffers temporal fluctuations in early juvenile survival for an endangered Pacific salmon. Journal of Animal Ecology. Available online. DOI: 10.1111/1365-2656.12117