Calibration by proximity: The key to index-based inference about abundance and distribution for marine bioregions

Authors (Order TBD): James Thorson, Robyn Forrest

Potential co-authors (must find task to quality as co-authors): Anne Hollowed, Lingbo Li

Target journal: ICES JMS

Abstract

Surveys of fish abundance are typically conducted over a pre-defined region that corresponds to biological or management boundaries. However, many fish species occur across multiple adjacent regions, and are therefore sampled by multiple, non-overlapping surveys. Interpreting results from multiple surveys is difficult due to differences in catchability, such that differences in sampling rates may be due to differences in density or catchability. Estimating differences in catchability from multiple surveys (termed "intercalibration") can be accomplished either by calibration experiments (where multiple survey gears are used at a single location and time) or by fitting an assessment model (where historical fishing and biological assumptions are used as a depletion experiment). However, neither approach is feasible for many species and regions. We therefore introduce a third approach ("calibration by proximity", CBP), which assumes that breakpoints between surveys are statistically independent of differences in species density. Given this assumption, we fit a single spatio-temporal model to multiple disjoint surveys while estimating a catchability ratio for each survey. We use a simulation experiment to show that CBP results in unbiased estimates of (1) catchability ratios, (2) the proportion of abundance in different surveys, and (3) changes in distribution over time. We then illustrate the approach for arrowtooth flounder in the Northeast Pacific, which shows increasing abundance region-wide but particularly for the California Current. We conclude by arguing that CBP represents "best available science" for estimating distribution shifts across survey regions for many low-value and ecosystem-component species. We also recommend ongoing efforts to compare calibration estimates between calibration-experiments, assessment models, and CBP models.

Introduction

Fisheries science has its roots in scientific attempts to inform fishing regulations (Smith 2007). In part due to this practical origin, fisheries scientists and managers often define a fish "stock" as the subset of the spatial distribution for a given fish species at which fishing activities can be regulated (Reiss et al. 2009). In many cases, the spatial scale of fisheries management (and resulting "stock" boundaries) will substantially differ from the spatial distribution for a given species, or a biologically distinct subset (termed a "population"). This mismatch may be entirely reasonable when the goal is regulating regional management, but many stakeholders, ecologists, and conservation groups are increasingly interested in understanding fish populations at scales larger than any single fisheries management authority.

For example, there has been considerable recent interest among non-profit groups, conservationists, and the media regarding "global ocean health" (Halpern et al. 2008, Worm et al. 2009). Of particular interest is whether the distribution of fish populations is shifting due to human-caused climate change and, if so, how to respond (Cheung et al. 2009, Pinsky et al. 2013). Global ocean health remains difficult to assess because many fish stocks lack the institutional resources or data to conduct formal stock assessments (Costello et al. 2012, Thorson et al. 2012). However, population-wide distribution shifts are also difficult to identify (even in well-studied parts of the world) due to a mis-match between the definition of stocks and the spatial distribution of ocean populations. For example, many commercially important species in the Pacific Ocean (e.g., Pacific cod) have a coastal distribution that ranges from California to the Bering Sea.

For marine fishes with a broad spatial distribution, distribution shifts must be identified by combining survey data from many different portions of their range. However, different surveys will use techniques (gear and operational protocols) that can increase or decrease the

sampling efficiency. Sampling efficiency is generally measured via "catchability", defined as the proportion of local biomass that is captured on average by a given sampling operation, and simultaneously analysing multiple surveys requires an estimate of differences in catchability. We see three main avenues to doing this when combining data from different regions with non-overlapping surveys:

1. *Ignoring catchability*: The simplest avenue is to ignore differences in catchability, and assume that each gear captures 100% of biomass that is within the spatial footprint of sampling (termed the "area swept" for a bottom trawl sampling gear). This will result in poor inference whenever different gear have large differences in catchability.
2. *Intercalibrate different surveys*: Fisheries scientists have long recognized the importance of estimating differences in catchability among sampling gear. Therefore, when sampling gear must be changed, scientists often conduct planned "calibration samples", wherein both gears are operated at the same place and time so that their expected catch rates can be compared (Miller 2013).
3. *Using stock assessment estimates*: Alternatively, a well-studied species may have existing stock assessments for each portion of its range. In this case, the stock assessment will generally provide an estimate of catchability for each survey. If these estimates are well-trusted, they could in theory be used to "weight" density samples from different regions in a combined model.

Each option could be used to make inference about distribution shifts and spatial dynamics using data from multiple surveys.

However, there are many cases where none of these three options are feasible, but where scientific information is still required sought regarding distribution shifts. For example, arrowtooth flounder [blurb about importance].

In this study, we therefore propose a fourth method for combining data from multiple surveys to infer distribution shifts and abundance trends at large spatial scales. This approach works best when different sampling programs are conducted near one-another (e.g., on both sides of a jurisdictional boundary between two nations), and requires the assumption that the location of breakpoints between surveys is independent of biogeographic differences in population density. If this assumption is met, then samples between adjacent but non-overlapping samples can be used to "intercalibrate" estimates of catchability differences. We therefore call this approach "calibration by proximity" (CBP). We test this CBP approach using a simulation experiment involving both biomass-dynamic and age-structured dynamics condition, which simulates the spatial distribution of sampling from thirteen separate data sources in the coastal Eastern Pacific Ocean from 1983-2014. We then apply the approach to data for arrowtooth flounder in this region, and show that the centroid of the population has shifted north by nearly 100 kilometers due largely to density increases in the Eastern Bering Sea.

Methods

We seek to develop a spatio-temporal statistical method to integrate data from multiple adjacent but non-overlapping surveys to estimate spatial variation in density, as well as annual changes in spatial distribution. As case-study, we integrate data from four marine ecosystems (Eastern Bering Sea, Gulf of Alaska, British Columbia, and US West Coast; see Fig. 1) and 10 different bottom trawl surveys occurring in the Northeast Pacific Ocean from 1983-2014 (Table 1, Fig. 2). The Eastern Bering Sea survey has been conducted in every year from 1983-2014, but every other survey is missing at least one year of data and the Queen Charlotte Sound Synoptic survey was conducted in only one year. We also split up two surveys into smaller time-series, reflecting changes in sampling gear or operation. Specifically, we split the Gulf of Alaska bottom trawl survey into two indices: 1980-1993,

and 1996-2013, where the seasonal timing and tow-length were changed between the 1993 and 1996 surveys. We also split the Triennial survey of the US West Coast into three indices: 1983-1992, 1995-2001, and 2004, where the seasonal timing was changed between the 1992 and 1995 surveys, and the operation of the survey was changed from Alaska to Northwest Fisheries Science Center between 2001 and 2004 surveys.

Data-integrated spatio-temporal index model

To integrate data from these multiple surveys, we adapt a spatio-temporal model that predicts biomass density $d(s,t)$ (in units kilograms per square-kilometer) for every location s within the sampling domain of any one survey, and every year $t \in \{1982,1983,\ldots,2015\}$.

Predicted density is informed by biomass samples $b(i)$ (in units kilograms) for every sample i. The model calculates the probability of these data using a delta-model:

$$\Pr(b_i = B) = \begin{cases} 1 - p_i & \text{if } B = 0 \\ p_i \times \text{Gamma}\{B|\sigma_m^{-2}, r_i\sigma_m^2\} & \text{if } B > 0 \end{cases}$$

where $\text{Gamma}\{B|\sigma_m^{-2}, r_i\sigma_m^2\}$ is a gamma probability density function with shape σ_m^{-2} and scale $r_2(i)\sigma_m^2$, where σ_m^2 is the coefficient of variation for non-zero catches, p_i is the probability that $B > 0$, and r_i is the expected catch rate when $B > 0$. Both p_i and r_i are specified via a "Poisson-process" link function:

$$p_i = 1 - \exp(-a_i \times n_i)$$

$$r_i = \frac{a_i \times n_i}{p_i} \times w_i$$

where n_i is predicted number of groups, w_i is the predicted biomass per group for sample i, and a_i is the area-swept for sampling i. Numbers-density n_i and biomass per group w_i are in turn are modelled using spatial and spatio-temporal random fields:

$$\log(n_i) = \beta_n(t_i) + \omega_n(s_i) + \varepsilon_n(s_i, t_i) + \sum_{k=1}^{N_k} \lambda_n(k)\, q(i,k)$$

$$\log(w_i) = \beta_w(t_i) + \omega_w(s_i) + \varepsilon_w(s_i, t_i) + \sum_{k=1}^{N_k} \lambda_w(k)\, q(i,k)$$

where $\beta_n(t_i)$ and $\beta_w(t_i)$ are annually varying intercepts for numbers-density and average weight, $\omega_n(s_i)$ and $\omega_w(s_i)$ represent spatial variation among locations s, $\varepsilon_n(s_i, t_i)$ and $\varepsilon_w(s_i, t_i)$ represent spatial variation that varies for each year t, and $\lambda_n(k)$ and $\lambda_w(k)$ represent unmeasured differences in survey efficiency (termed "catchability") among surveys k used for sample i as indicated by the design matrix $q(i,k)$.

This model requires specifying a distribution for spatial and spatio-temporal random fields:

$$\boldsymbol{\omega}_n \sim MVN(\mathbf{0}, \sigma_{\omega n}^2 \mathbf{R}_n)$$

$$\boldsymbol{\omega}_w \sim MVN(\mathbf{0}, \sigma_{\omega w}^2 \mathbf{R}_w)$$

$$\boldsymbol{\varepsilon}_n(t) \sim \begin{cases} MVN(\mathbf{0}, \sigma_{\varepsilon n}^2 \mathbf{R}_n) & \text{if } t = 1 \\ MVN(\rho_n \times \boldsymbol{\varepsilon}_n(t-1), \sigma_{\varepsilon n}^2 \mathbf{R}_n) & \text{if } t > 1 \end{cases}$$

$$\boldsymbol{\varepsilon}_w(t) \sim \begin{cases} MVN(\mathbf{0}, \sigma_{\varepsilon w}^2 \mathbf{R}_n) & \text{if } t = 1 \\ MVN(\rho_w \times \boldsymbol{\varepsilon}_w(t-1), \sigma_{\varepsilon w}^2 \mathbf{R}_n) & \text{if } t > 1 \end{cases}$$

where $\boldsymbol{\omega}_n$ is the vector formed by $\omega_n(s)$ for all locations s, ρ_n is the estimated first-order autocorrelation in spatio-temporal variation, and $\sigma_{\omega n}^2$ is the variance of $\omega_n(s)$, and $\mathbf{R}_n$ is the spatial correlation matrix representing the correlation $\mathbf{R}_n(s, s+h)$ between location s and location $s+h$ for spatial and spatio-temporal terms. This spatial correlation is approximated as following a Matern function:

$$\mathbf{R}_n(s, s+h) = \frac{1}{2^{\nu-1}\Gamma(n)} \times (\kappa_n |h\mathbf{H}|)^{\nu} \times K_{\nu}(\kappa_n |h\mathbf{H}|)$$

where smoothness ν is fixed at 1.0, κ_n controls the distance at which spatial correlations drop to 10%, and $\mathbf{H}$ is an estimated linear-transformation that represents differences in decorrelation rate for different directions (termed "geometric anisotropy").

We estimate fixed-effects via maximum marginal likelihood, while treating spatial ($\boldsymbol{\omega}_n$ and $\boldsymbol{\omega}_w$) and spatio-temporal ($\boldsymbol{\varepsilon}_n(t)$ and $\boldsymbol{\varepsilon}_w(t)$) variation as random effects. Parameters include the decorrelation distance (κ_n and κ_w), two parameters governing geometric anisotropy (**H**), the variance of spatial and spatio-temporal variation ($\sigma^2_{\omega n}$, $\sigma^2_{\omega w}$, $\sigma^2_{\varepsilon n}$, $\sigma^2_{\varepsilon w}$), first-order autocorrelation for spatio-temporal variation (ρ_n and ρ_w), annual intercepts ($\beta_n(t)$ and $\beta_w(t)$), and catchability parameters ($\lambda_n(k)$ and $\lambda_w(k)$). Parameter estimation is conducted using R package VAST (Thorson and Barnett In press, Thorson et al. In press), which is freely available and fully documented online (https://github.com/James-Thorson/VAST). VAST uses Template Model Builder (Kristensen et al. 2016) to apply the Laplace approximation (Skaug and Fournier 2006) to the marginal likelihood of fixed effects, and this approximation to the marginal likelihood is then optimized using a conventional Nelder-Mead optimizer in the R statistical environment (R Core Team 2016).

After parameters are estimated, we predict biomass at each location s and year t:

$$\log\big(d(s,t)\big) = \beta_n(t) + \omega_n(s) + \varepsilon_n(s,t) + \beta_w(t) + \omega_w(s) + \varepsilon_w(s,t)$$

where catchability variables are excluded from this prediction computation such that biomass is implicitly calculated relative to the catchability for the reference survey. We then predict coastwide biomass I_t for each year t, as well as biomass $I_t(l)$ for each of the four individual ecosystems:

$$I_t(l) = \sum_{s=1}^{N_s} \big(a(s,l) \times d(s,t)\big)$$

where $a(s,l)$ is the area associated with location s for region l (Shelton et al. 2014, Thorson et al. 2015). We also predict the centroid of distribution (COG) for each year:

$$Z_t(m) = \sum_{s=1}^{N_s} \frac{\big(z(x,m) \times a(x) \times d(s,t)\big)}{I(t)}$$

where $z(s,m)$ is a matrix representing the location in Eastings and Northings of each location s such that $Z_t(m)$ represents shifts in distribution in these coordinates. This estimate of distribution shift accounts for differences in the spatial distribution of and sampling efficiency of different surveys, which would otherwise inhibit analysis of distribution shifts at large spatial scales, and has previously been used for surveys in the US West Coast ecosystem (Thorson et al. 2016). For both biomass $I_t(l)$ and $Z_t(m)$, we use the epsilon bias-correction estimator to ensure that these indices are unbiased despite their calculation as a nonlinear transformation of random effects (Thorson and Kristensen 2016).

Estimating differences in catchability via "calibration by proximity"

In the following, we specify a design-matrix for catchability $\mathbf{Q}$ where the row for sample i $\mathbf{q}(i)$ is a vector indicating which of 13 survey indices was used to obtain that sample. We fix $q(i,k) = 0$ for a single reference index, to ensure that $\lambda_n(k)$ and $\beta_n(t_i)$ are both identifiable. Given our use of a log-link for n_i and w_i, $\lambda_n(k)$ and $\lambda_w(k)$ can be interpreted as the log-ratio of expected numbers and biomass-per-group for survey k relative to the reference survey.

In general, these thirteen surveys do not operate in the same area and the same year. However, exceptions include the AFSC slope survey and the "early" Triennial survey in 1992/1995, and the "final" Triennial survey and the West Coast groundfish bottom trawl survey in 2004, see Fig. A1. These exceptions are able to estimate the ratio of catchability via average differences in encounter-probability and positive catch-rate for samples occurring at overlapping locations and years.

For non-overlapping surveys, however, the model cannot estimate catchability via average differences in catch-rate at the same location and time. It therefore informs estimated differences in catchability via the correlation in density $d(s,t)$ between nearby locations that are sampled in the same year. We therefore call this type of information "calibration by proximity" (CBP), in recognition that replaces "calibration sampling" with

model-based inference about calibration. Importantly, CBP requires assuming that the location of sampling (including the breakpoint between different survey operations) is "statistically independent" or the response variable. If this assumption is met, then statistical theory suggests that the spatio-temporal model will generally result in unbiased estimates of biomass trends and distribution shifts.

Simulation experiment

To explore the estimation properties of this model, we simulate density (in biomass) over the same four ecosystems from 1983-2014 (i.e., the same area and years) at a fine spatial scale (103,166 grid cells each with area of approximately 4 square-kilometers). Specifically, we simulate numbers-density:

$$n(x,c,t) = \begin{cases} \exp\big(\beta_n(t) + \omega_n(s) + \varepsilon_n(s,t)\big) & \text{if } c = 1 \\ \exp(-M(\mathrm{c}-1)) \times \exp\big(\beta(t) + \omega_n(s) + \varepsilon_n(s,t)\big) & \text{if } t = 1 \text{ and } c > 1 \\ \exp(-M) \times n(x,c-1,t-1) & \text{if } t > 1 \text{ and } c > 1 \end{cases}$$

and biomass-per-group:

$$w(x,c,t) = a(L_\infty(1-\exp(-kc))^b \times \exp\big(\beta_w(t) + \omega_w(s) + \varepsilon_w(s,t)\big)$$

where $a(L_\infty(1-\exp(-kc))^b$ is the expected biomass per individual of age c, $a = 1$ is the biomass-per-volume coefficient, $b = 3$ represents isometric weight-at-length, $L_\infty = 1$ represents asymptotic individual length, $k = 0.2$ is the Brody growth coefficient, and M is the individual mortality rate.

We then simulate samples of biomass $b(i)$ for each age grid cell x, year t, and age $c \in \{1, \dots, c_{max}\}$ using the Poisson-process link function using linear predictors:

$$n_i = n(x,c,t) \times \exp\big(\lambda_n(k)\big) \times S(k,c)$$

$$w_i = w(x,c,t)$$

where $\lambda_n(k)$ is the catchability coefficient for each survey, and $S(k,c)$ is the age-based selectivity for survey k:

$$S(k,c) = \frac{1}{1 + \exp(-(c - S_{50\%}(k)) \times v)}$$

where $S_{50\%}(k)$ is the age of 50% selection for survey k and v governs the rate-of-increase for selectivity at age $S_{50\%}(k)$.

We explore two scenarios using this simulation model:

Scenario #1: Biomass dynamic

Our first scenario uses this simulation model to closely mimic a conventional "biomass-dynamic" operating model. To do so, we fix $M \underset{lim}{\longrightarrow}$ Inf, such that all individuals have identical biomass, and $S_{50\%}(k) = 0$ and $v \underset{lim}{\longrightarrow}$ Inf, such that all individuals are fully selected by each survey (Mangel et al. 2013). We then model substantial differences in catchability among surveys:

$$\lambda_n(k) \sim Normal(0, 0.5^2)$$

and record the sampled biomass for age 1 individuals.

Scenario #2: Age-structured

Our second scenario uses this simulation to generate age-structured data. We specify that $c_{max} = 10$ and $M = 0.3$, and also model differences in catchability and selectivity among surveys:

$$\lambda_n(k) \sim Normal(0, 0.2^2)$$

$$S_{50\%}(k) \sim Normal(3, 2)$$

where $v = 1$. We then sum sampled biomass across all ages in a given sample, where differences in selectivity among surveys appears as spatial variation in catchability when analysed using a model that ignores age-structure.

For each simulation replicate, we simulate an identical sampling design as exists for all thirteen surveys. Specifically, we simulation biomass for each age c at the exact location (Latitude-Longitude, x), time (year, t) and survey (k) for each sample i. Our simulation

sampling data is therefore unbalanced across space and time in the same way as available bottom-trawl data for the Northeast Pacific. We generate 100 simulation replicates for each scenario, and for each replicate we record the true catchability $\lambda_n(k)$ for each survey, the true centroid-of-distribution $Z_t(m)$, and the true proportion of biomass in each of the four ecosystems, $P_t(l) = I_t(l)/I_t$. We then fit the spatio-temporal model to simulated data for each simulation replicate, and evaluate model performance via error and bias in estimation of these variables.

Case-study example: Arrowtooth flounder

[Fill in history of arrowtooth flounder, and importance among regions, including any stock assessments in the different regions, importance as predator for pollock, and previous research regarding environmental drivers]

Given (1) the growing interest regarding arrowtooth flounder as a predator for commercially important species, and (2) its distribution throughout the Northeast Pacific, we therefore model biomass of this species in all four marine ecosystems simultaneously.

Results

Simulation experiment

Our simulation experiment shows that the spatio-temporal model is able to accurately estimate the catchability coefficient $\lambda_n(k)$ relative to the average catchability for all thirteen surveys given the biomass-dynamic operating model (Fig 3 top-left panel). However, the biomass-dynamic operating model does not include variation in $\lambda_w(k)$, whereas the estimation model often estimates substantial variation in $\lambda_w(k)$ (Fig. 3 top-right panel). By contrast, the age-based operating model includes variation in $\lambda_n(k)$ and selectivity-at-age among surveys. These both contribute to "apparent" catchability, and this causes the estimation model estimates of $\lambda_n(k)$ to be highly variable relative to its true values (Fig. 3 bottom row).

Similarly, the spatio-temporal model is able to accurately estimate the proportion of biomass $P_t(l)$ residing in each of four ecosystems l in each year t in both simulation scenarios (Fig. 4). This is true, despite substantial variance in estimates of $\lambda_w(k)$, where an estimate of $\lambda_w(k) > 0$ will cause the model to estimate a smaller proportion of biomass for the ecosystem sampled by survey k and vice-versa. Unbiased estimates of this proportion $P_t(l)$ is, in turn, necessary to generate unbiased estimates of changes in the centroid of the population, $Z_t(m)$. The model does provide unbiased estimates of distribution shift for both scenarios (Fig. 5), although the age-based operating model has much lower precision (a correlation of 0.58-0.68 between true and estimated distribution shifts) than the biomass-dynamic model (which has a correlation of 0.8-0.86).

Arrowtooth flounder

Application of this spatio-temporal model to data for arrowtooth flounder shows that abundance substantially increased in the Northeast Pacific in two stages: from 1983-1988, and again from 1999-2005 (Fig. 6). However, abundance increases have been driven by increases in the Eastern Bering Sea: the US West Coast and British Columbia have shown fluctuations but little evidence of a secular trend, and the Gulf of Alaska has shown relatively stable biomass across the past three decades.

These increases in abundance in the Eastern Bering Sea have therefore driven a shift northward and westward for the center-of-distribution for arrowtooth flounder in the Northeast Pacific (Fig. 7). This pattern has resulted in the population shifting nearly 100 kilometers north between the average from 1983-1997 and the recent average 2010-2014.

Discussion

[Specification as "scale-free" and thus robust to different calculations of area-swept]

Acknowledgements

295

296

Works cited

Cheung, W.W.L., Lam, V.W.Y., Sarmiento, J.L., Kearney, K., Watson, R., and Pauly, D. 2009. Projecting global marine biodiversity impacts under climate change scenarios. Fish Fish. **10**(3): 235–251. doi:10.1111/j.1467-2979.2008.00315.x.

Costello, C., Ovando, D., Hilborn, R., Gaines, S.G., Deschenes, O., and Lester, S.E. 2012. Status and solutions for the world's unassessed fisheries. Science **338**(6106): 517–520. doi:10.1126/science.1223389.

Halpern, B.S., Walbridge, S., Selkoe, K.A., Kappel, C.V., Micheli, F., D'Agrosa, C., Bruno, J.F., Casey, K.S., Ebert, C., Fox, H.E., and others. 2008. A global map of human impact on marine ecosystems. Science **319**(5865): 948.

Keller, A.A., Wallace, J.R., and Methot, R.D. 2017. The Northwest Fisheries Science Center's West Coast Groundfish Bottom Trawl Survey: History, Design, and Description. NOAA Technical Memorandum, Northwest Fisheries Science Center, Seattle, WA.

Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H., and Bell, B.M. 2016. TMB: Automatic Differentiation and Laplace Approximation. J. Stat. Softw. **70**(5): 1–21. doi:doi: 10.18637/jss.v070.i05.

Mangel, M., MacCall, A.D., Brodziak, J.K., Dick, E.J., Forrest, R.E., Pourzand, R., and Ralston, S. 2013. A perspective on steepness, reference points, and stock assessment. Can. J. Fish. Aquat. Sci. **70**(6): 930–940.

Miller, T.J. 2013. A comparison of hierarchical models for relative catch efficiency based on paired-gear data for US Northwest Atlantic fish stocks. Can. J. Fish. Aquat. Sci. **70**(9): 1306–1316. doi:10.1139/cjfas-2013-0136.

Pinsky, M.L., Worm, B., Fogarty, M.J., Sarmiento, J.L., and Levin, S.A. 2013. Marine taxa track local climate velocities. Science **341**(6151): 1239–1242.

R Core Team. 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available from https://www.R-project.org/.

Reiss, H., Hoarau, G., Dickey-Collas, M., and Wolff, W.J. 2009. Genetic population structure of marine fish: mismatch between biological and fisheries management units. Fish Fish. **10**(4): 361–395. doi:10.1111/j.1467-2979.2008.00324.x.

Shelton, A.O., Thorson, J.T., Ward, E.J., and Feist, B.E. 2014. Spatial semiparametric models improve estimates of species abundance and distribution. Can. J. Fish. Aquat. Sci. **71**(11): 1655–1666. doi:10.1139/cjfas-2013-0508.

Skaug, H., and Fournier, D. 2006. Automatic approximation of the marginal likelihood in non-Gaussian hierarchical models. Comput. Stat. Data Anal. **51**(2): 699–709.

Smith, T.D. 2007. Scaling Fisheries: The Science of Measuring the Effects of Fishing, 1855-1955. *In* 1st edition. Cambridge University Press, Cambridge, UK.

Thorson, J.T., and Barnett, L.A.K. In press. Comparing estimates of abundance trends and distribution shifts using single- and multispecies models of fishes and biogenic habitat. ICES J. Mar. Sci. doi:10.1093/icesjms/fsw193.

Thorson, J.T., Branch, T.A., and Jensen, O. 2012. Using model-based inference to evaluate global fisheries status from landings, location and life history data. Can. J. Fish. Aquat. Sci. **69**(4): 645–655.

Thorson, J.T., Ianelli, J.N., and Kotwicki, S. In press. The relative influence of temperature and size structure on fish distribution shifts: a case study on walleye pollock in the Bering Sea. Fish Fish.

Thorson, J.T., and Kristensen, K. 2016. Implementing a generic method for bias correction in statistical models using random effects, with spatial and population dynamics examples. Fish. Res. **175**: 66–74. doi:10.1016/j.fishres.2015.11.016.

Thorson, J.T., Pinsky, M.L., and Ward, E.J. 2016. Model-based inference for estimating shifts in species distribution, area occupied and centre of gravity. Methods Ecol. Evol. **7**(8): 990–1002. doi:10.1111/2041-210X.12567.

Thorson, J.T., Shelton, A.O., Ward, E.J., and Skaug, H.J. 2015. Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. ICES J. Mar. Sci. J. Cons. **72**(5): 1297–1310. doi:10.1093/icesjms/fsu243.

Von Szalay, P.G., Wilkins, M.E., and Martin, M.M. 2010. Data report: 2009 Gulf of Alaska bottom trawl survey. US Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Alaska Fisheries Science Center. Available from http://www.afsc.noaa.gov/Publications/AFSC-TM/NOAA-TM-AFSC-189/Main%20Body%20of%20Report.pdf [accessed 9 May 2014].

Workman, G.D., Rutherford, K.I., and Olsen, N. 2008. Hecate Strait Groundfish Bottom Trawl Survey, May 25th to June 29th, 2005. Fisheries and Oceans Canada, Nanaimo, British Columbia.

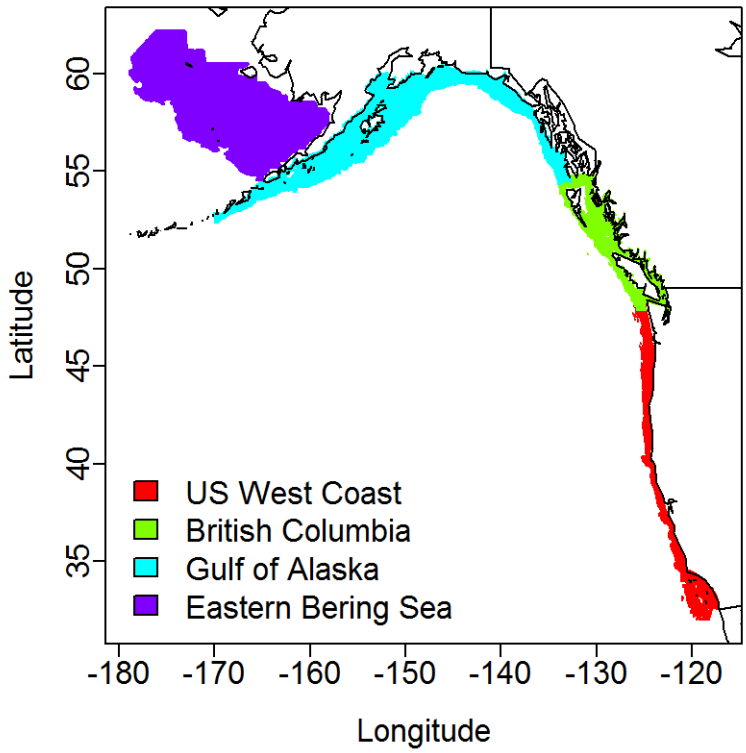
Worm, B., Hilborn, R., Baum, J.K., Branch, T.A., Collie, J.S., Costello, C., Fogarty, M.J., Fulton, E.A., Hutchings, J.A., Jennings, S., Jensen, O.P., Lotze, H.K., Mace, P., McClanahan, T.R., Minto, C., Palumbi, S.R., Parma, A.M., Ricard, D., Rosenberg, A.A., Watson, R., and Zeller, D. 2009. Rebuilding global fisheries. Science **325**(5940): 578.

369 Table 1 – Description of surveys used in this analysis

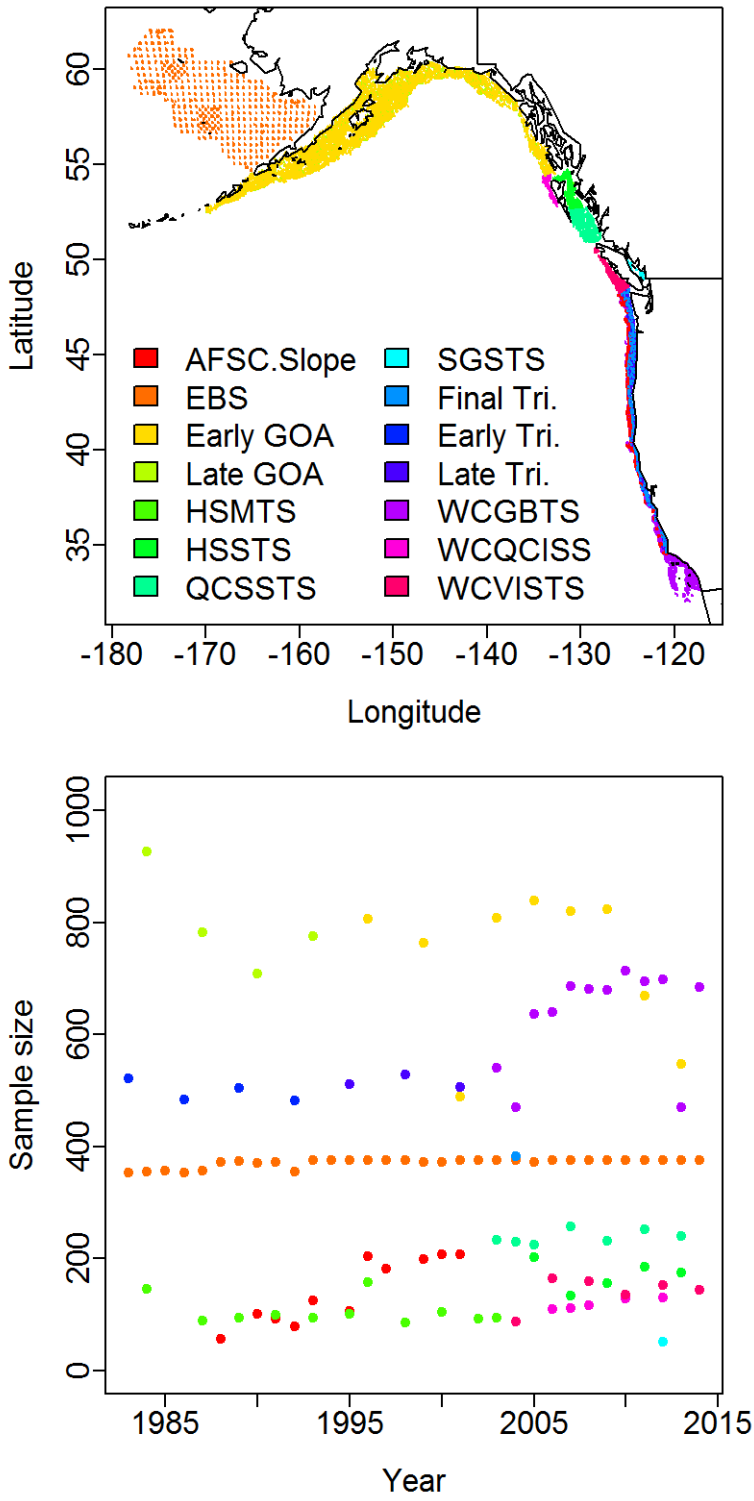
Survey code	Name	Region	Treatment	Survey design	Depth range (m)	References
WCGBTS	West Coast groundfish bottom trawl survey	West Coast of US	Single index	Random stratified design	55-1300	Keller et al. (2017)
Triennial	Triennial West Coast groundfish survey	West Coast of US	Three index: 1. "Early": 1983-1992 2. "Late": 1995-2001 3. "Final": 2004	Transect-based design	<500	Keller et al. (2017)
HSMTS	Hecate Strait Multispecies	British Columbia	Single index	Depth-stratified systematic design	10-500	
HSSTS	Hecate Strait Synoptic	British Columbia	Single index	Random stratified design	10-500	Workman et al. (2008)
QCSSTS	Queen Charlotte Sound Synoptic	British Columbia	Single index	Random stratified design	50-500	Olsen et al. 2007
SGSTS	Strait of Georgia Synoptic	British Columbia	Single index	Random stratified design	10-500	King et al. 2013
WCHGSS	West Coast Haida Gwaii Synoptic	British Columbia	Single index	Random stratified design	150-1300	
WCVISTS	West Coast Vancouver Island Synoptic	British Columbia	Single index	Random stratified design	50-500	
GOA	Gulf of Alaska bottom trawl survey	Gulf of Alaska	Two index: 1. "Early": 1984-1993 2. "Late": 1996-2013	Random stratified design	10-1000	von Szalay et al. (2010)
EBS	Eastern Bering Sea bottom trawl survey	Eastern Bering Sea		Random stratified design	50-300	Bakkala 1993

370

Fig. 1 – Domain used when estimating population density, aggregated across all four regions (see inset legend at bottom-left for color code).

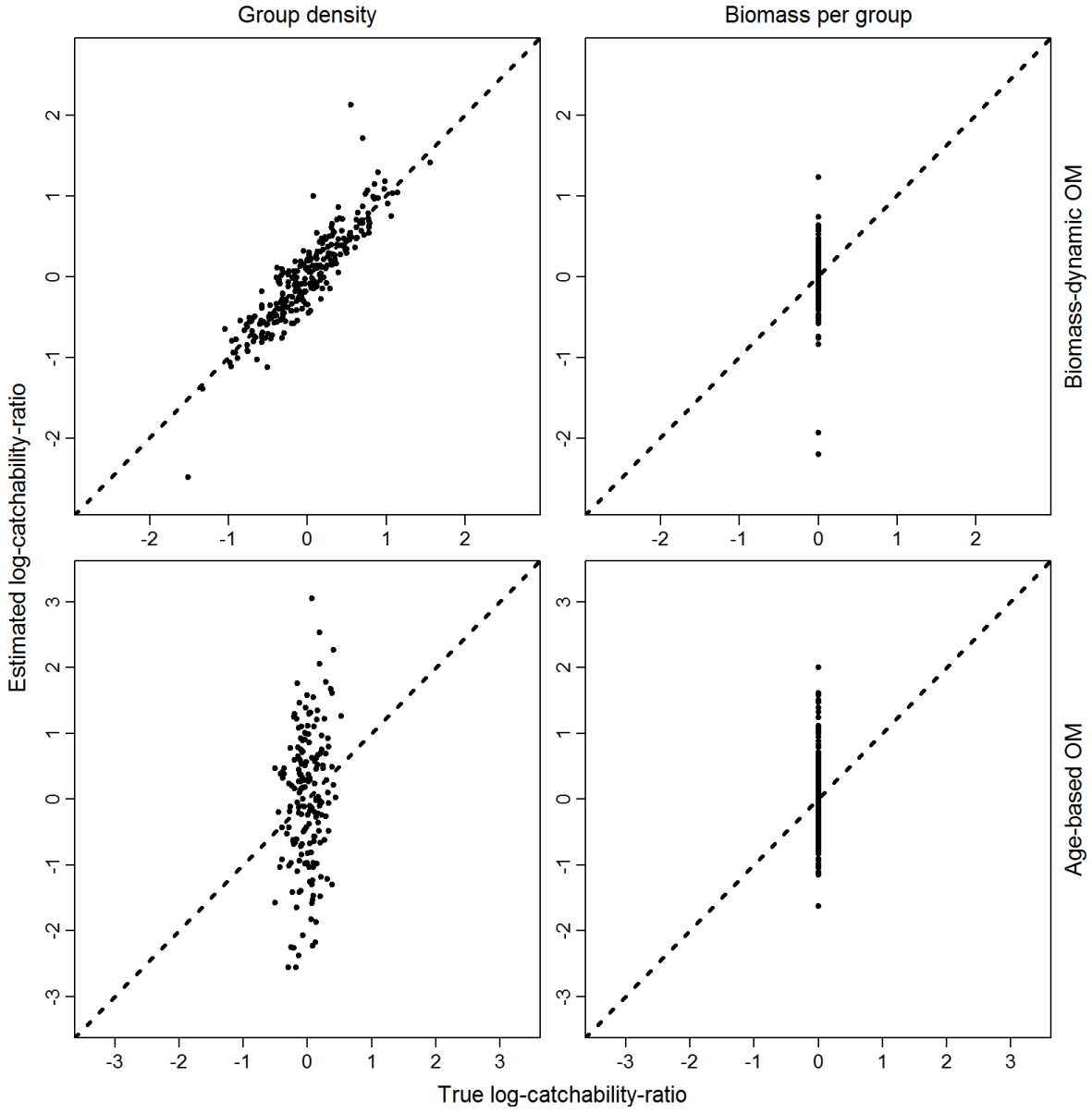


375 Fig. 2 – Map of sampling locations for each survey (top panel), and annual sample size for
376 each survey (bottom panel), where color code (legend in top panel) is identical between both
377 panels.



378

379 Fig. 3 – Comparison of true and estimated log-catchability in each survey relative to the
380 average log-catchability across all surveys in a simulation experiment (dashed line: one-to-
381 one relationship, representing perfect estimate).



382

383

384

Fig. 4 – Comparison of true and estimated proportion of biomass by year in each of four regions, $P_t(l)$ in a simulation experiment involving a biomass-dynamic (top panel) or age-structured (bottom panel) operating model (dashed line: one-to-one relationship, representing a perfect estimate).

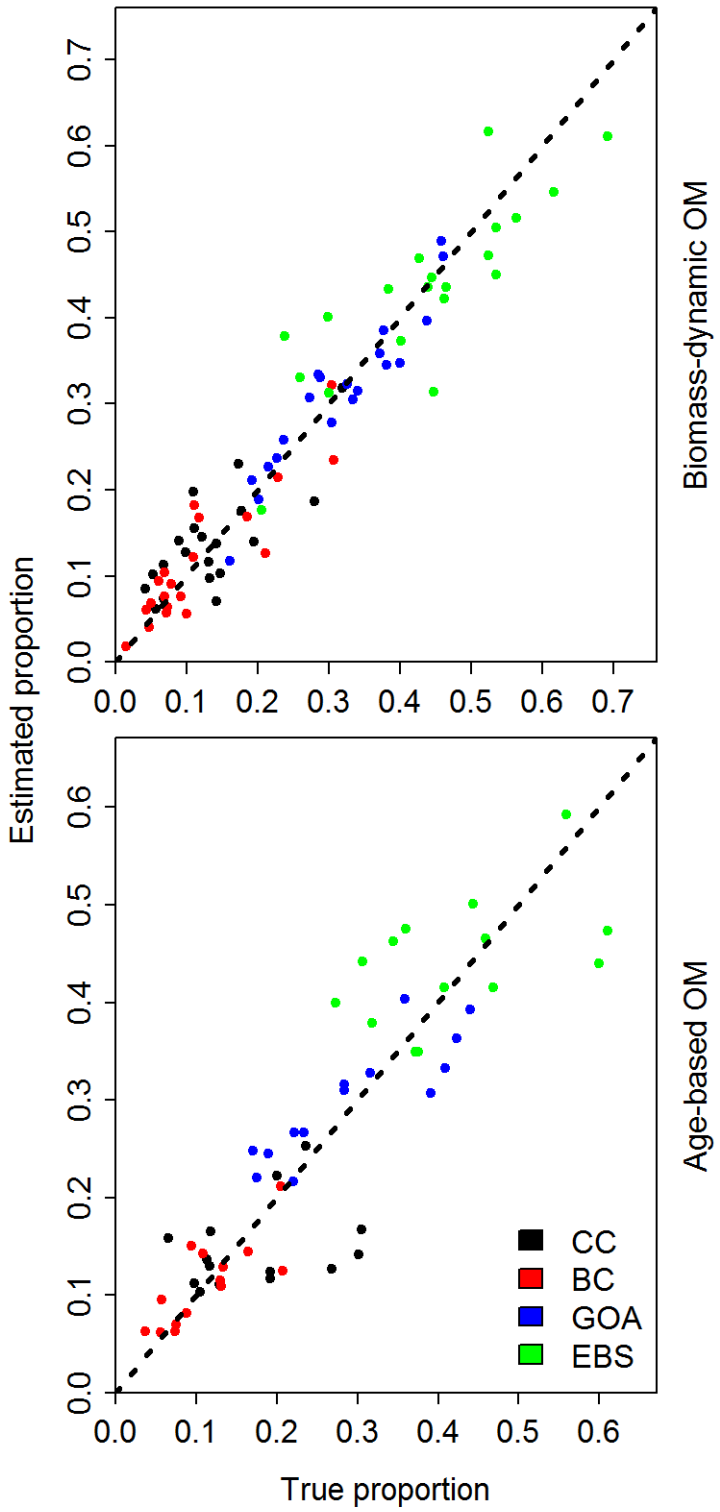


Fig. 5 – Estimated deviations of population center-of-gravity, $Z_t(m)$, in Eastings (left column) and Northings (right column) coordinates away from the average for all years in a given simulation replicate (dashed line: one-to-one relationship, representing perfect estimate).

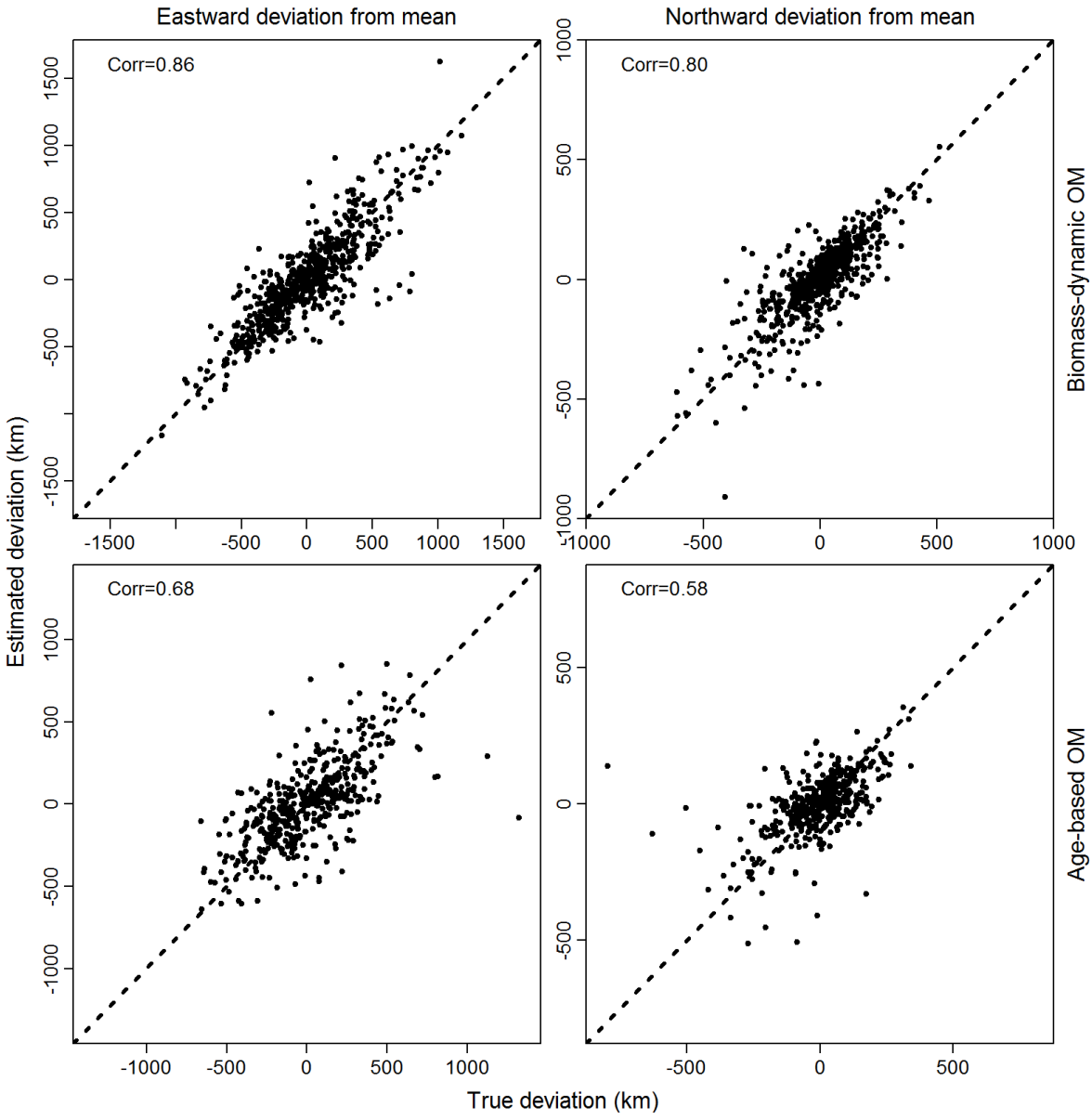
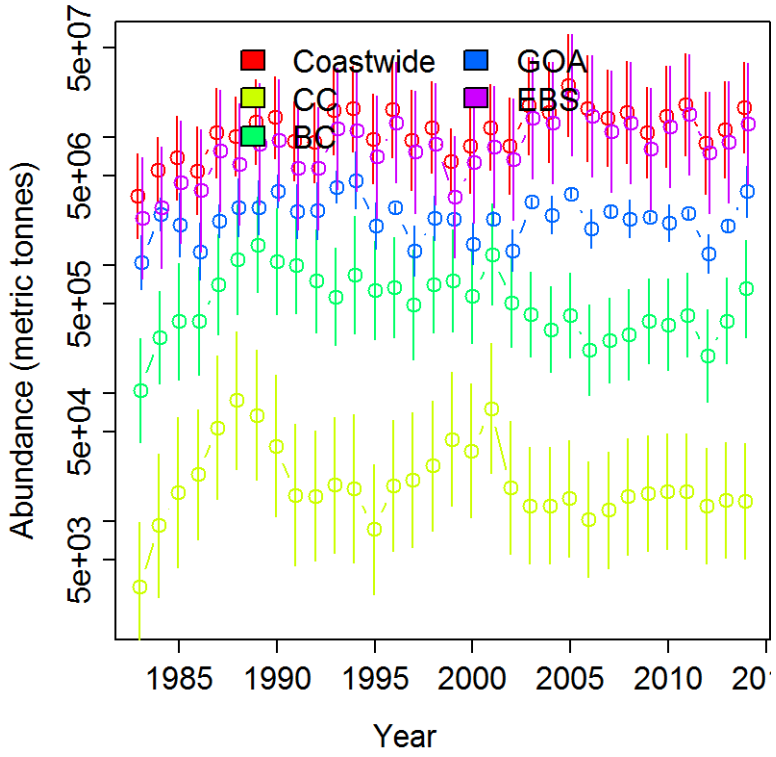


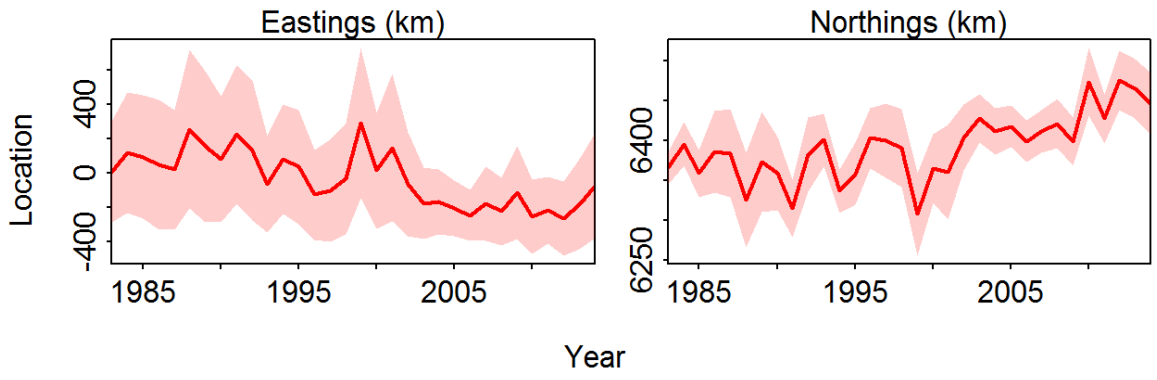
Fig. 6 – Arrowtooth flounder log-abundance in each of four ecosystems, $I_t(l)$, as well as aggregated across all regions, I_t, showing bias-corrected estimates of the mean biomass (circles) and +/- one estimated standard error (whiskers).



400 Fig. 7 – Centroid of distribution, $Z_t(m)$, for arrowtooth flounder in each modelled year,

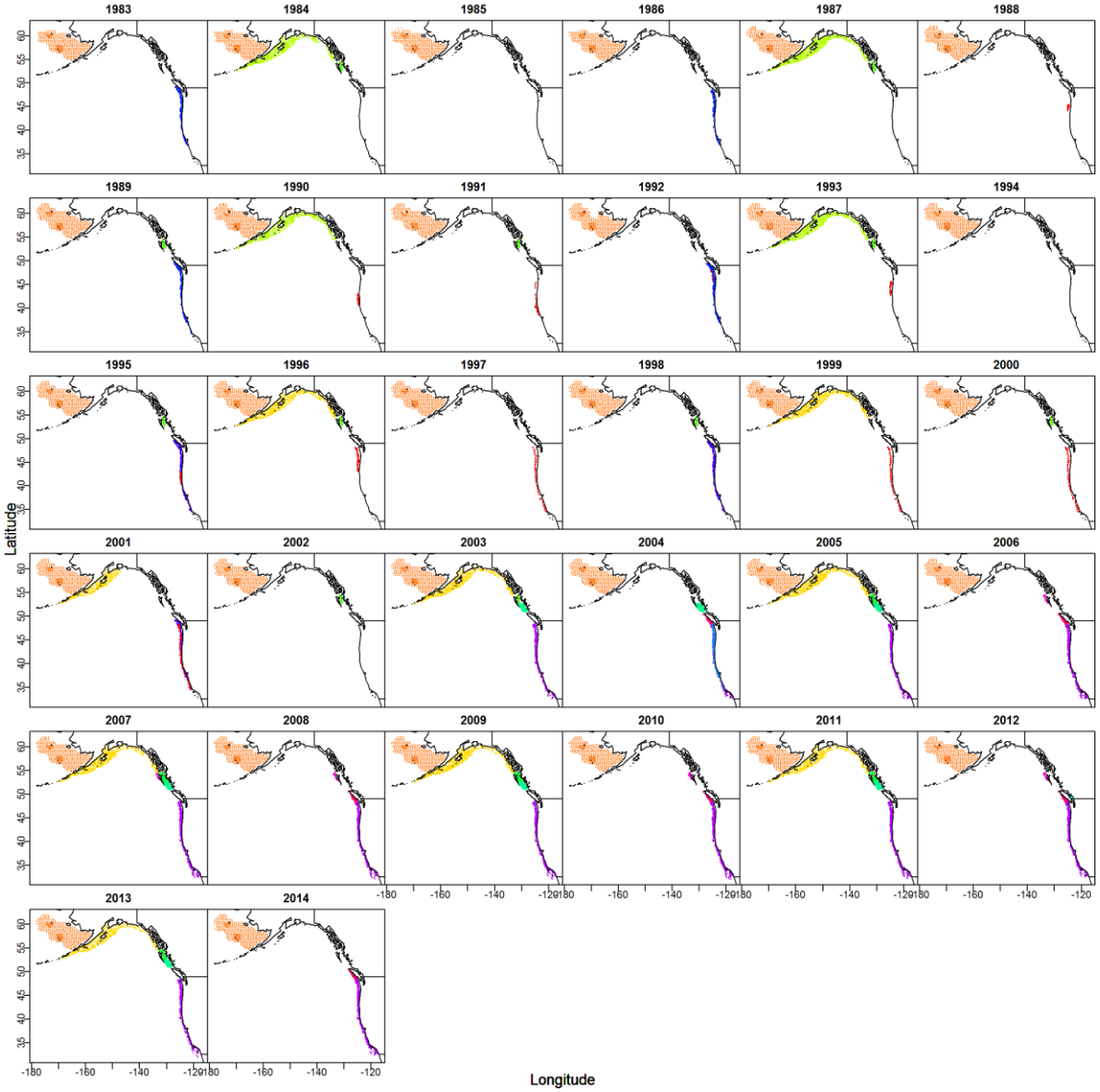
401 measured as kilometres east (relative to Universal Transverse Mercator 5) and north (relative

402 to equator), as well as +/- 1 standard error.



403

404 Appendix Fig. A1



405