Using spatio-temporal models of tagging data to deal with incomplete mixing

Mark N Maunder, James T Thorson, Haikun Xu

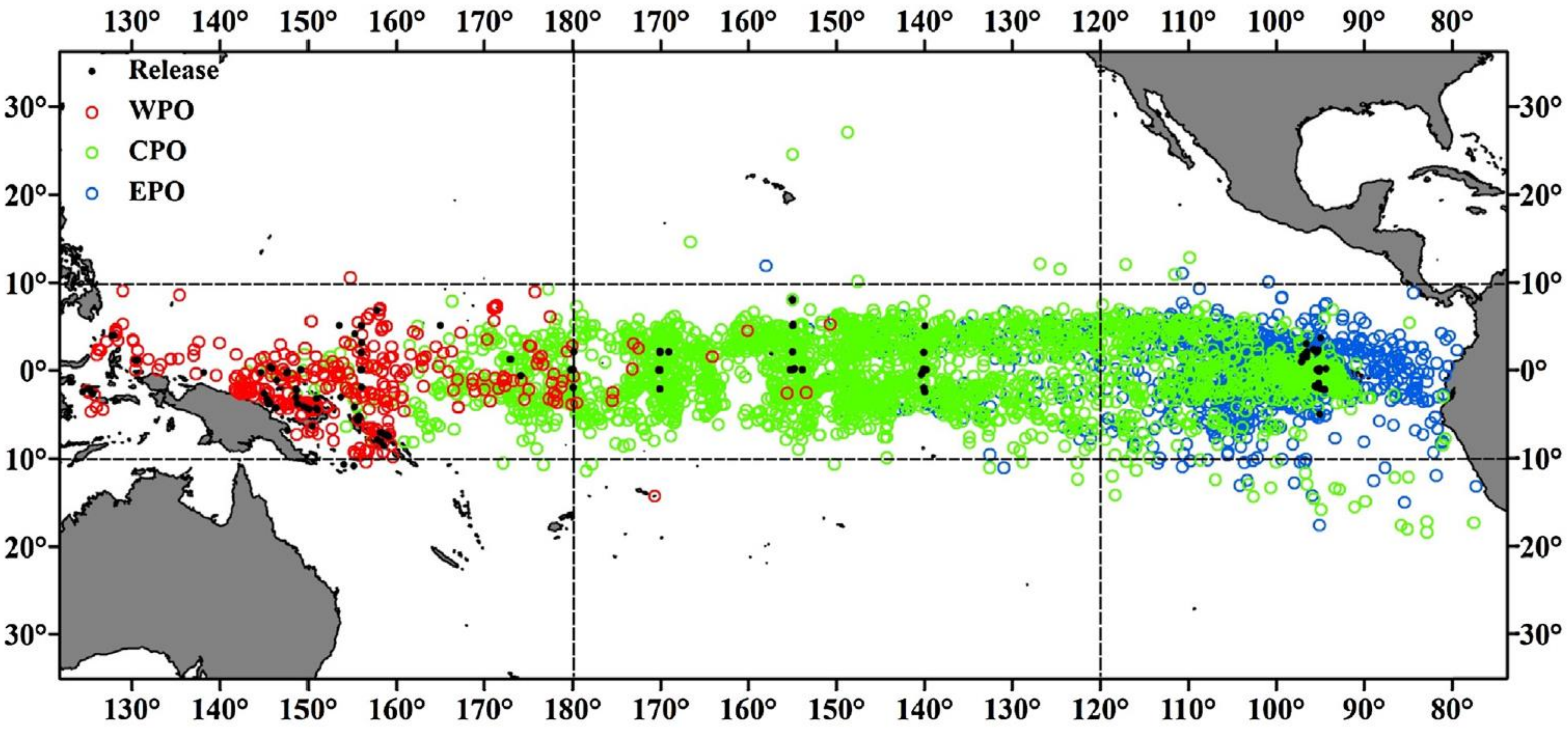

Fig. 13. Release and recovery positions of bigeye tuna, at liberty for >30 d, color coded for releases in the western (WPO), central (CPO), and eastern Pacific Ocean (EPO). The putative stock boundaries, discussed in the text, at longitudes 120° W and 180°, and latitudes 10° N and 10° S, are superimposed. (For interpretation of the references to color in figure legend, the reader is referred to the web version of the article.)

K. Schaefer et al. / Fisheries Research 161 (2015) 336–355

Estimating abundance from tagging data

- Petersen
  - $\frac{r}{n_2} = \frac{n_1}{N}$
  - n_1 number of individuals tagged
  - n_2 sample size for the recovery data
  - *r* the number of tagged individuals recovered
  - *N* is the population size
- $\widehat{N} = \frac{n_1 n_2}{r}$

Spatial estimates

- $\hat{N}_i = \frac{n_{1,i} n_{2,i}}{r_i}$
  - $n_{1,i}$ number of tagged individuals in area *i*
  - $n_{2,i}$ sample size for the recovery data in area *i*
  - r_i the number of tagged individuals recovered in area *i*
  - N_i is the population size in area *i*

Need to model movement of tagged individuals

Not all areas have tagged individuals so need to share information

Issues

- Need to model movement of tagged individuals
  - Advection diffusion model
  - e.g. Sibert et al. 1999. Can. J. Fish. Aquat. Sci. 56: 925-938.
- Need to share information on total abundance among space
  - Spatial (-temporal) model
  - e.g. Thorson et al. 2015. ICES J. Mar. Sci. 72: 1297–1310.
- The abundance of tagged fish changes over time
  - Multiple likelihood calculations for each area
  - e.g. Hilborn 1990. Can. J. Fish. Aquat. Sci. 47: 635–643.

Assumptions

- Tags are not removed from the population
- The total population does not move
- Catch is not removed from the total population

Advection diffusion model of tagged individuals

Thorson et al. 2016. J. Appl. Ecol. 54: 577-587

$$\boldsymbol{n}_{1,t} = \boldsymbol{m}\boldsymbol{n}_{\mathbf{1},t-1} + \boldsymbol{R}_t$$

Where R_t are the tag releases

Movement function m typically includes both random and directed components, termed diffusion and advection, respectively. This function can be calculated from an instantaneous movement rate:

$$\frac{\partial}{\partial t}\mathcal{B} = (\mathbf{u}^{\mathrm{T}}\nabla + \nabla \cdot \Sigma\nabla)\mathcal{B} \qquad \text{eqn 7}$$

where $\mathbf{u}^{\mathrm{T}}\nabla\mathcal{B}$ represents advective movement (where ∇ is the gradient operator, which yields a vector of length two when evaluated at location s because $\mathcal{B}$ is a function defined in two-dimensional space, and $\mathbf{u}$ is a direction vector of length two), and $\nabla \cdot \boldsymbol{\Sigma}\nabla$ represents diffusive movement (where $\boldsymbol{\Sigma}$ is a 2×2 rotation matrix governing the rate of diffusion in different directions, and if $\boldsymbol{\Sigma} = \mathbf{I}$ then $\nabla \cdot \boldsymbol{\Sigma}\nabla$ reduces to the Laplacian operator).

Appendix S2. Movement matrix computation on a triangulated mesh.

Spatial model

- $\widehat{N}_i = exp(d_0 + \gamma_i)$

$$\gamma \sim \text{MVN}(0, \sigma_\gamma^2 \cdot \text{R}_{\text{spatial}}) \quad (2)$$

where σ_γ is the marginal standard deviation (SD) of spatial variation γ and $\text{R}_{\text{spatial}}$ is spatial correlation for the random field:

$$\text{R}_{\text{spatial}}(s, s') = \text{Matérn}\left(\frac{|(s - s')|}{\kappa}\right) \quad (3)$$

where s and s' are the location of two spatial stations, κ defines the rate at which correlations drop with increasing distance, and Matérn $(|(s-s')|)$ is the Matérn correlation function, which calculates the correlation between γ at stations s and s' given their distance $|s-s'|$. We

Likelihood: Poisson

$$-lnL = \sum_{t,i} -r_{t,i}\, ln[\lambda_{t,i}] + \lambda_{t,i}$$

$$\lambda_{t,i} = \frac{n_{2,t,i}}{N_{t,i}} n_{1,t,i}$$

$n_{1,i}$ number of tagged individuals in area *i*
$n_{2,i}$ sample size for the recovery data in area *i*
r_i the number of tagged individuals recovered in area *i*
N_i is the population size in area *i*

Spatio-temporal model

- Accounts for movement and catch
- Does not use the information on movement from the tagging data
- Does not explicitly use the information on catch
- $\widehat{N}_i = exp\left(d_{0,t} + \gamma_i + \gamma_{t,i}\right)$

Improvements

- Removing tag recoveries from the tagged population;
- Covariates for N
- Using the advection-diffusion process to move the total population;
- Removing catch from the total population;
- Alternative likelihood functions;
  - zero inflation
- Including size information.

Journal of Applied Ecology 2016 doi: 10.1111/1365-2664.12664

Using spatio-temporal models of population growth and movement to monitor overlap between human impacts and fish populations

James T. Thorson*, Jason Jannot and Kayleigh Somers

Fisheries Resource Analysis and Monitoring Division, Northwest Fisheries Science Center, National Marine Fisheries Service, NOAA, 2725 Montlake Blvd. E, Seattle, WA 98112, USA

Spatiotemporal variation in size-structured populations using fishery data: an application to shortfin mako (*Isurus oxyrinchus*) in the Pacific Ocean[1]

Mikihiko Kai, James T. Thorson, Kevin R. Piner, and Mark N. Maunder

Spatial delay-difference models for estimating spatiotemporal variation in juvenile production and population abundance

James T. Thorson, James N. Ianelli, Stephan B. Munch, Kotaro Ono, and Paul D. Spencer

Movement function m typically includes both random and directed components, termed diffusion and advection, respectively. This function can be calculated from an instantaneous movement rate:

$$\frac{\partial}{\partial t}\mathcal{B} = (\mathbf{u}^{\mathbf{T}}\nabla + \nabla \cdot \Sigma\nabla)\mathcal{B} \qquad \text{eqn 7}$$

where $\mathbf{u}^{\mathbf{T}}\nabla\mathcal{B}$ represents advective movement (where ∇ is the gradient operator, which yields a vector of length two when evaluated at location s because $\mathcal{B}$ is a function defined in two-dimensional space, and $\mathbf{u}$ is a direction vector of length two), and $\nabla{\cdot}\boldsymbol{\Sigma}\nabla$ represents diffusive movement (where $\boldsymbol{\Sigma}$ is a 2×2 rotation matrix governing the rate of diffusion in different directions, and if $\boldsymbol{\Sigma} = \mathbf{I}$ then $\nabla{\cdot}\boldsymbol{\Sigma}\nabla$ reduces to the Laplacian operator).

In practical applications, the solution to movement and production defined for all possible locations within a population domain can only be calculated analytically given particular functions for density dependence (Okubo, Hastings & Powell 2001). More generically, however, the model can be approximated using techniques derived from finite element analysis. We first divide the entire spatial domain $\mathcal{D}$ into a set of n_r triangles such that every location $s \in \mathcal{D}$ is within exactly one triangle (see Appendix S1 in Supporting Information). The number of triangles represents a balance between numerical precision and computational speed, and we recommend that future studies confirm that results are unchanged when increasing the number of triangles (as we have done for the results presented in this study). Fish within each triangle are assumed to be homogenous and evenly mixed, such that every location s within triangle r has the same density, harvest rate, surplus production, process error, etc. Therefore, each function in the spatial surplus production model (e.g. density $\mathcal{B}_t$) is approximated as a piecewise constant function. Triangle r has area a_r (in units km^2) and this area contains abundance $b_{r,t}$ (in units kg.) such that population density at location s in that triangle is $b_{r,t}/a_r$, and we use vector $\mathbf{b}_t$ to refer to the abundance in every triangle. Population abundance changes among years as follows:

$$\mathbf{b}_{t+1} = g(\mathbf{M}(\mathbf{b}_t * \exp(-u_t\mathbf{f}_t))) * \exp(\varepsilon_t) \qquad \text{eqn 8}$$

where $\mathbf{M}$ is a matrix representing annual movement rates among triangles, $\mathbf{f}_t$ is proportional to the amount of fishing activity per unit area $f_{r,t}$ in triangle r (e.g. the total area swept by bottom trawlers divided by the total triangle area), and u_t is an estimated scaling coefficient such that $u_t f_{r,t}$ is the instantaneous fishing mortality rate. Process error $\boldsymbol{\varepsilon}_t$ again represents spatially correlated, unexplained variation in dynamics:

$$\boldsymbol{\varepsilon}_t \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon) \qquad \text{eqn 9}$$

where $\text{MVN}(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon)$ is a multivariate normal distribution with mean zero and covariance $\boldsymbol{\Sigma}_\varepsilon$, where process error covariance $\boldsymbol{\Sigma}_\varepsilon$ between triangles r_1 and r_2 follows a Matérn function of distance:

$$\text{Cov}(\varepsilon_{r1,t}, \varepsilon_{r2,t}) = \frac{\tau_\varepsilon^{-2}}{2^{\nu-1}\Gamma(\nu)}(\kappa_\varepsilon|s_1 - s_2|)^\nu K_\nu(\kappa_\varepsilon|s_1 - s_2|) \qquad \text{eqn 10}$$

where τ_ε governs the pointwise variance of $\boldsymbol{\varepsilon}_t$, κ_ε governs the geostatistical range of correlations, ν is the smoothness of the covariance matrix (we assume that $\nu = 1$ in the following), and K_ν is the Bessel function.

To approximate movement rates $\mathbf{M}$ among triangles within the population domain, we define a matrix $\mathbf{N}$ representing instantaneous movement rates among all triangles. $\mathbf{N}$ is negative on the diagonal and positive or zero everywhere else, and the off-diagonal $n_{r1,r2}$ is zero if triangles r_1 and r_2 do not share an edge. Instantaneous movement rate $\mathbf{N}$ is further decomposed into movement in each of four cardinal directions, where $\mathbf{m}$ is a vector of parameters representing movement in each cardinal direction. Further details regarding the computation of $\mathbf{N}$ given a set of n_r triangles in a population domain are given in Appendix S2, and code for computing these matrices is provided as an R package *MovementTools* on the first author's GitHub page (https://github.com/james-thorson/movement_tools). Given $\mathbf{N}$, the matrix $\mathbf{M}$ approximating annual movement rates can be calculated using the matrix exponential operator. During parameter estimation, however, we apply the Euler approximation movement to calculate annual movement rates $\mathbf{M}$ given instantaneous rates $\mathbf{N}$ (see Appendix S3).

Spatial (-te

- $\widehat{N}_i =$

$\gamma \sim$ MVI

$R_{\text{spatial}}(s$

$$d(s, t, q) = \exp\left(d_0(t, q) + \gamma(s) + \theta(s, t) + \omega(s, q) + \sum_{j=1}^{n_j} \beta_j x_j(s, t)\right) \quad (1)$$

$$\gamma \sim \text{MVN}(0, \sigma_\gamma^2 \cdot R_{\text{spatial}}) \quad (2)$$

where σ_γ is the marginal standard deviation (*SD*) of spatial variation γ and R_{spatial} is spatial correlation for the random field:

$$R_{\text{spatial}}(s, s') = \text{Matérn}\left(\frac{|(s - s')|}{\kappa}\right) \quad (3)$$

where s and s' are the location of two spatial stations, κ defines the rate at which correlations drop with increasing distance, and Matérn $(|(s-s')|)$ is the Matérn correlation function, which calculates the correlation between γ at stations s and s' given their distance $|s-s'|$. We