SPATIOTEMPORAL TAGGING MODEL FOR SKIPJACK IN THE EPO

Tobias K. Mildenberger, Anders Nielsen
November 2022

CONTENTS

1. Summary

Skipjack tuna (SKJ, *Katsuwonus pelamis*) in Eastern Pacific Ocean (EPO) lacks a reliable index of relative abundance and age-composition data, challenging its assessment and thus sustainable management. A spatiotemporal population model utilizing available tagging data might allow estimating the population size, distribution and sustainable harvest levels for SKJ in the EPO. This model estimates the movement of SKJ as an advection-diffusion process and mortality rates based on tagging data and catch and effort data. The advection process can be based on a spatiotemporal habitat preference function dependent on environmental layers, such as temperature or bathymetry maps, as described by Thorson *et al.* (2021). Movement and distribution probabilities can then be estimated by means of the matrix exponential of instantaneous rates (Thorson et al., 2021) or based on the extended Kalman filter (Harvey, 1990). Results indicate that SKJ in the EPO prefers intermediate sea surface temperatures around 25-26°C and exhibits stronger undirected movement with higher temperatures. Further, the model estimates fishing mortality in space and time and a natural mortality rate of 3.85 $year^{-1}$.

2. Introduction

The lack of a reliable index of relative abundance and age-composition data are among the main reasons for the lack of a formal stock assessment for skipjack tuna (SKJ) in the Eastern Pacific Ocean (EPO), thus challenging its sustainable exploitation. On the other hand, available information from multiple tagging events might serve as the basis for a spatiotemporal population model that allows estimating the abundance and exploitation rates of SKJ. A core part of this model is the spatiotemporal tagging model that based on the recovered (and non-recovered) tags describes the most probable movement patterns of SKJ in the EPO. Here, we describe the approach used for the spatiotemporal tagging model to estimate the movement and mortality rates of SKJ in the EPO.

3. Data

The spatiotemporal tagging and population model requires information from recovered (and non-recovered) conventional tags, archival tags, environmental data, as well as catch and effort data.

3.1 Tagging data

Tagging data from four tagging events are available for SKJ in the EPO. In this analysis, we utilized the conventional tags from the last two events that took place from 2000-2006 and 2019-2022 but excluded the data from the other two tagging events that took place from 1955 to 1964 and from 1979 to 1981. While the six tuna tagging cruises in 2000 to 2006 targeted bigeye tuna, 3517 SKJ were tagged and released with plastic dart tags, of which 565 tags were recovered. By contrast, the IATTC multi-year Regional Tuna Tagging Program (RTTP-EPO 2019-2022, Project E.4.a) that was initiated in 2019 focused on SKJ. The RTTP included three tagging cruises in 2019, 2020, and 2022. A total of 6259 SKJ were tagged with plastic dart tags during the cruises of which 1619 were recovered at the time of writing this report. After applying a speed filter and data processing, a total of 9741 tags remained, of which the 2252 recovered tags were used for the analysis presented here (Figure 1). Although most tagged fish were released around the equator and 95°W, the recaptured fish span a wide region from 150°W to the coast of Peru and from 22°S to 20°N (Figure 1).

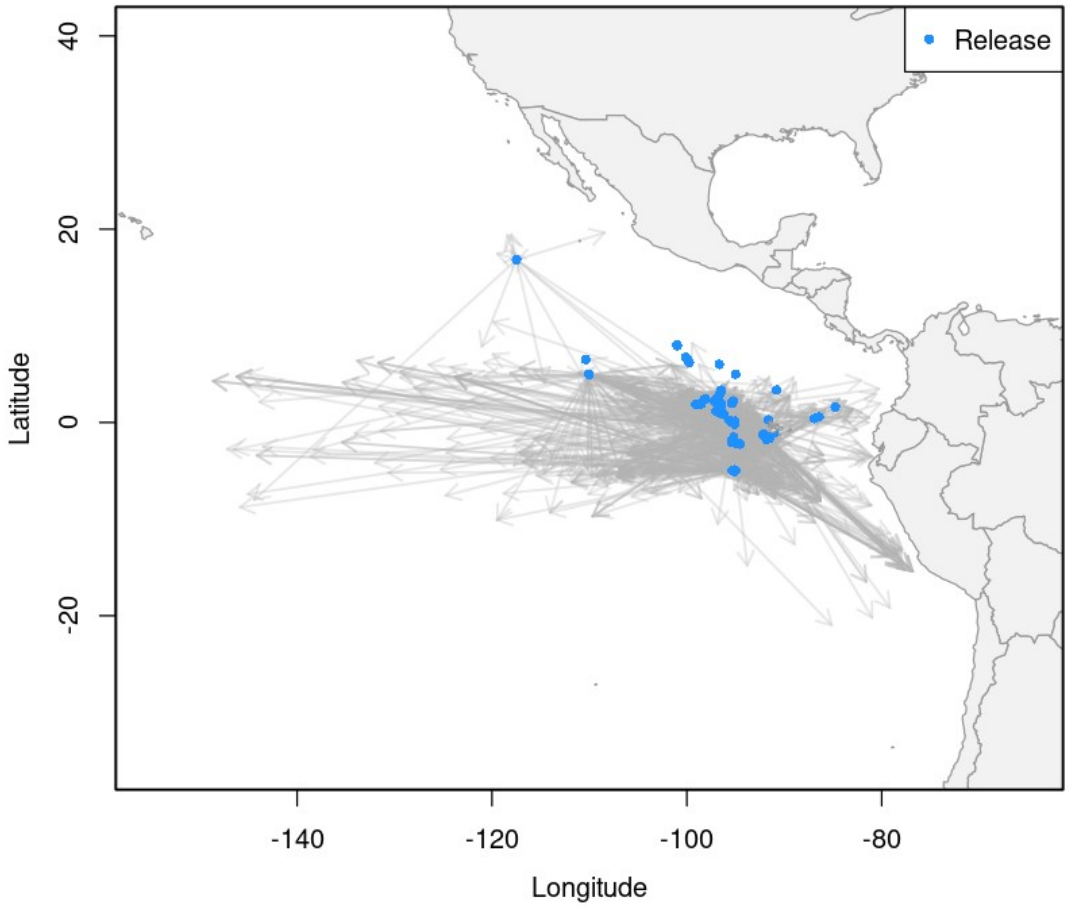

Figure 1: Release and recapture locations of the 2152 recovered conventional tags from the 2000 and 2019 tagging programs. Release locations are indicated by a blue circle and recapture locations by arrow heads.

In addition to the conventional tags, 35 archival tags from the 2019-2022 RTTP were processed and used for the analysis (Figure 2). The archival tags were recovered after 2 days to 8 months and travelled up to 15000 km. The most probable track was estimated with the unscented Kalman filter described in Lam et al. (2008).

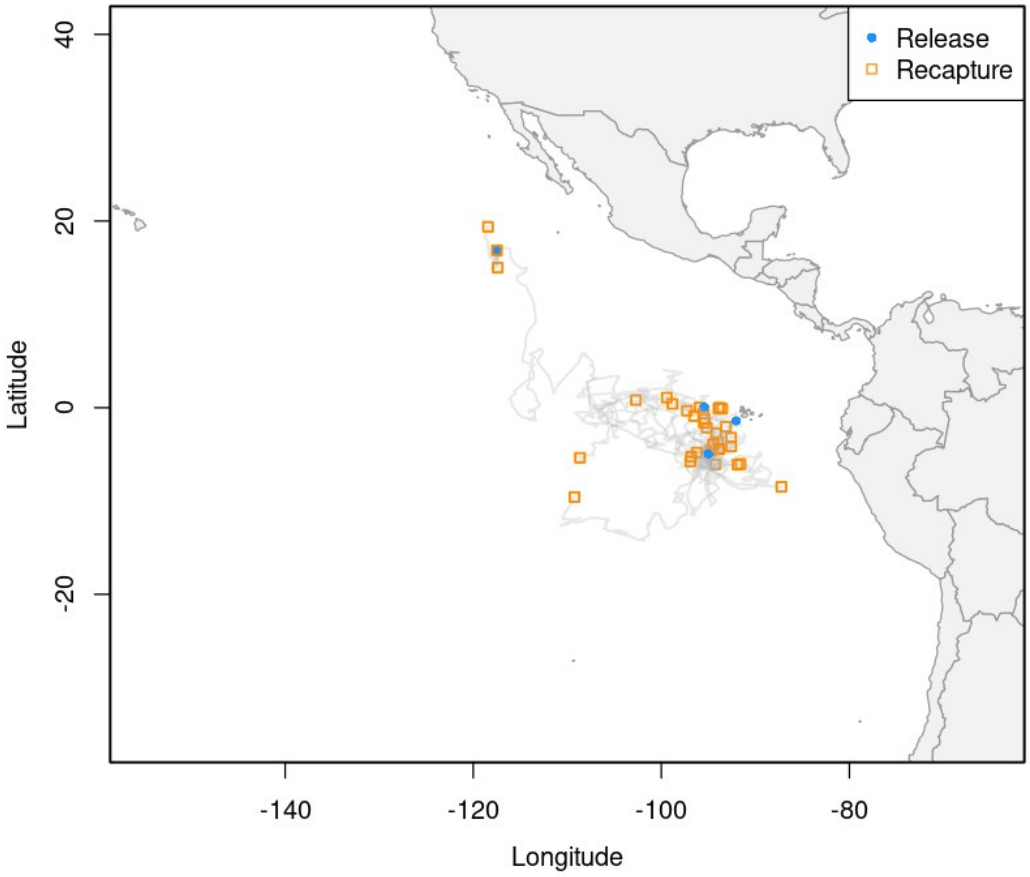

Figure 2: Tracks of the 35 archival tags used in this analysis, where blue points indicate the release location and orange squares the recapture locations.

The spatial domain of the model was defined by the Western management boundary at 150°W and the coastline of North and South America in the East as well as the 30°S and 35°N (darker blue area in Figure 3). An additional buffer zone of 5° around the model region was considered for the implementation of absorptive or reflective boundary conditions (lighter blue area in Figure 3).

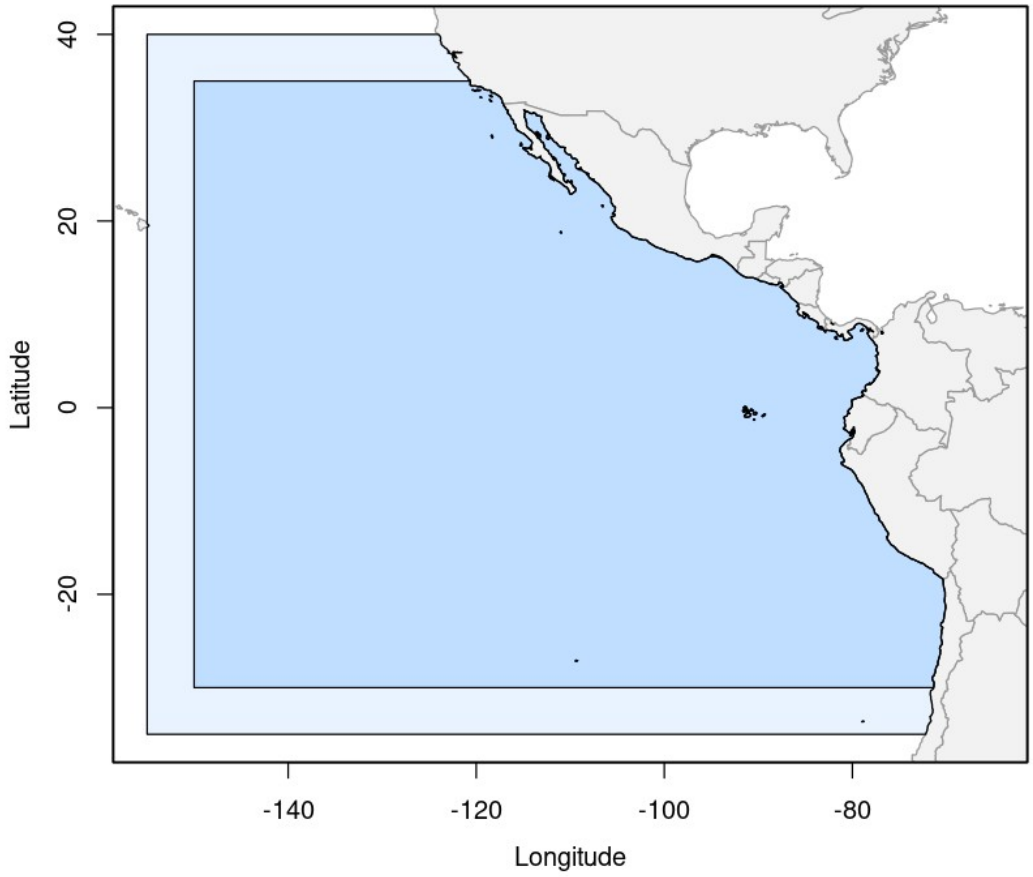

Figure 3: Spatial domain of the EPO considered for the tagging and abundance model (dark blue), with a buffer 5° buffer zone (in light blue).

Some tagged fish were recaptured close to the release location and within a short time after the release (even within the same day), these tags are not likely to contain a lot of information about the movement of the fish and cannot be assumed to be well mixed within the population. Thus, any tag that was recaptured within 14 days after the release, was removed from the analysis. Fourteen days were chosen as a conservative cut-off as the residence time of SKJ under FADs is likely longer (Dan Fuller, personal communication).

3.2 Environmental data

The SKJ tagging model requires environmental data to inform the habitat preference of SKJ. A range of potential environmental covariates could be relevant for informing the habitat preference and thus movement of SKJ in the EPO. We considered sea surface temperature (SST), mixed layer depth

(MLD), chlorophyll-a, kinetic energy, and an indicator of primary production as potential covariates informing the habitat preference and diffusion of SKJ. The environmental data is aggregated to 1°x1°, 2.5°x2.5°, and 5°x5° grid cells and weekly and monthly time steps for the period from 2000 to 2022. Supplementary figures S1 and S2 show SST and MLD, respectively, for a 2.5°x2.5° grid and with a monthly resolution for 2021.

3.3 Catch and effort data

Catch and effort information was available from 2000 to 2021 as number of sets and catch in metric tons per 1° grid cell, day, and fleet. The three different purse seine fleets are differentiated based on the type of sets between dolphin associated sets (dph), unassociated sets (sch), and fishing aggregation device (FAD) associated sets (log).

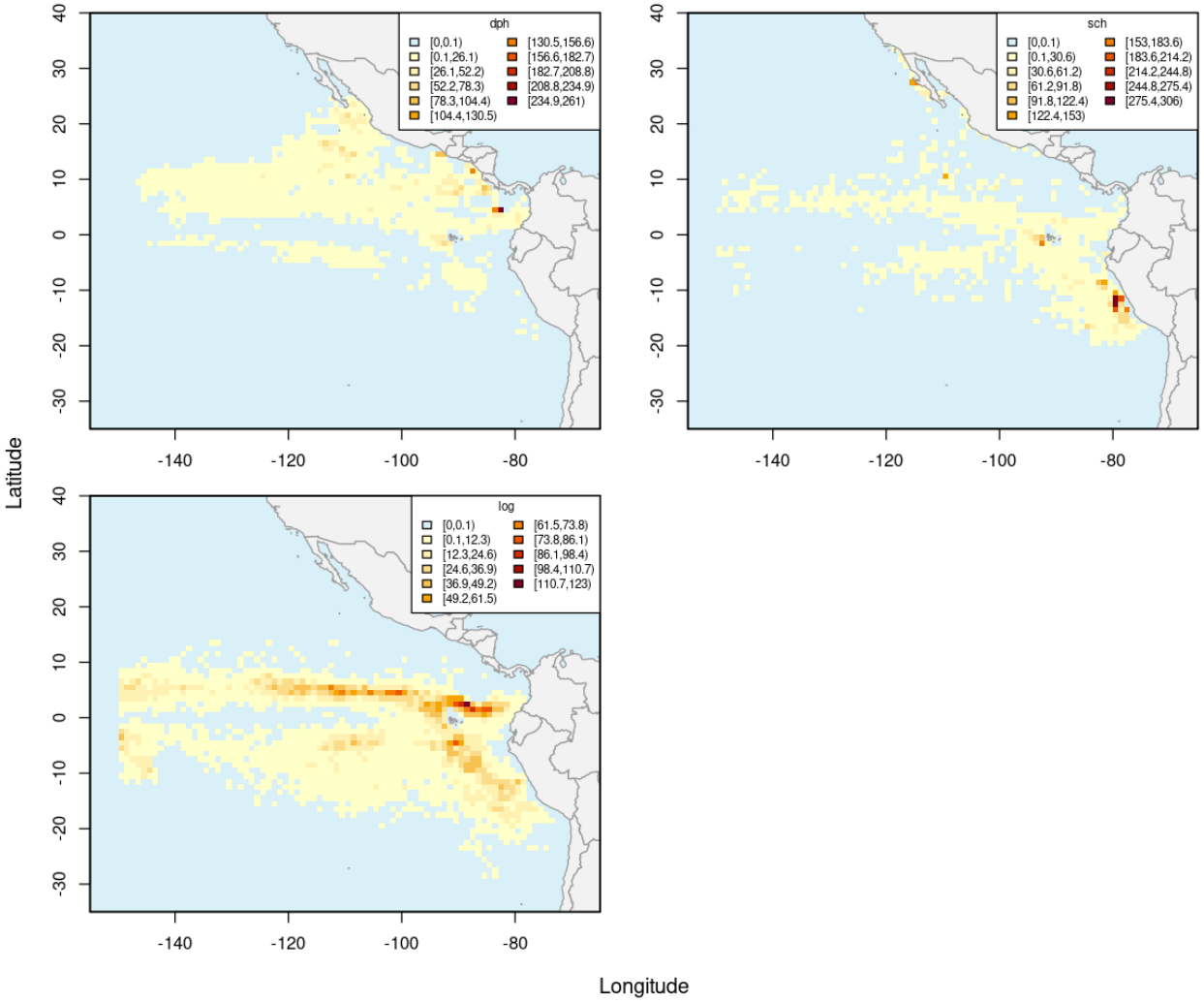

Figure 4: Annual effort in number of sets per 1° grid cell for 2021 for the three purse seine fleets: dph = dolphin associated sets, sch = unassociated sets, and log = FAD-associated sets.

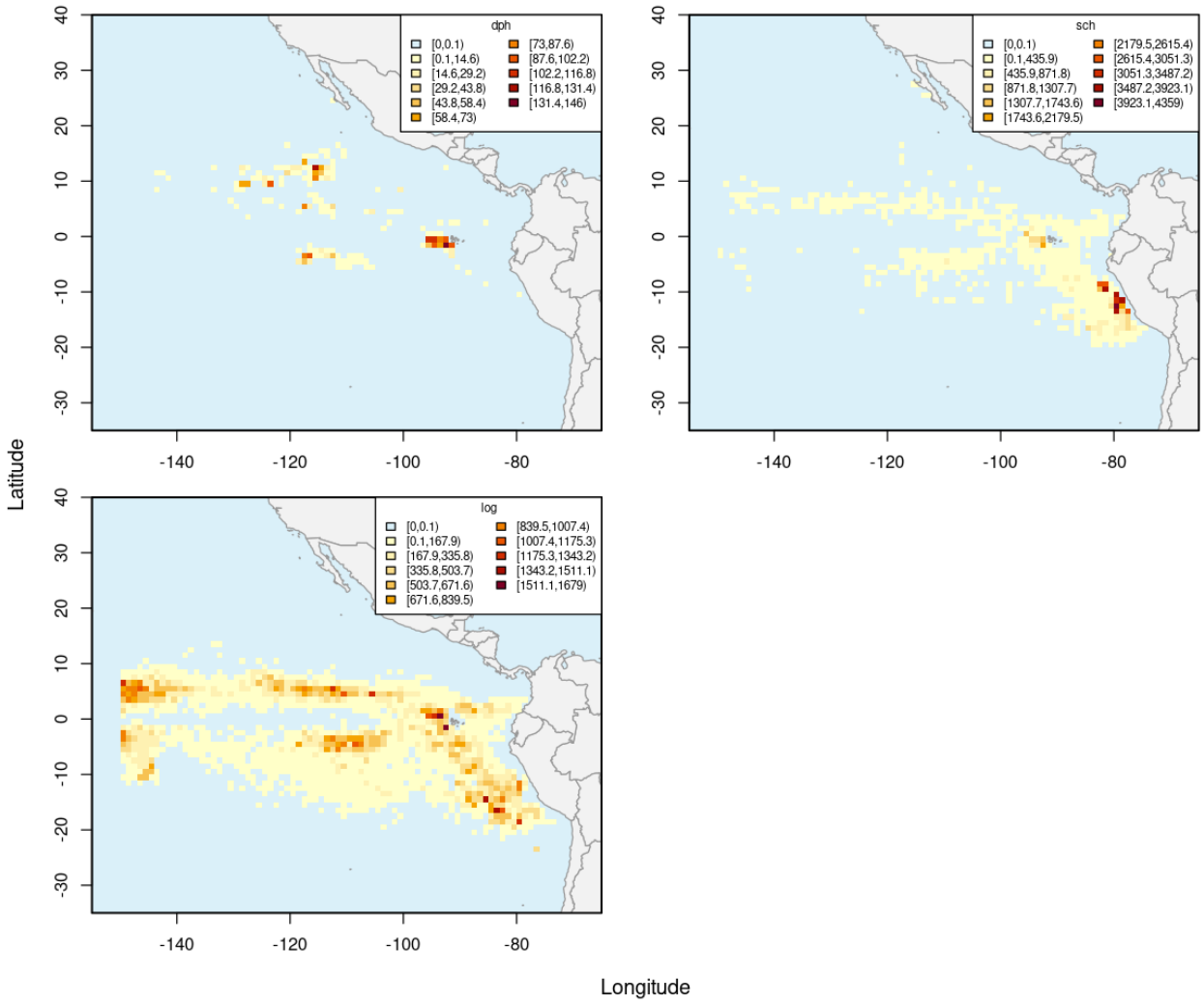

Figure 5: Annual catch in metric tons per 1° grid cell for 2021 for the three purse seine fleets: dph = dolphin associated sets, sch = unassociated sets, and log = FAD associated sets.

4. Movement model

The movement model describes an advection-diffusion process that utilizes environmental fields and smooth functions to inform advection and diffusion rates (Thorson *et al.* 2021). Two approaches are considered for the estimation of parameters: the matrix exponential and the extended Kalman filter.

The matrix exponential approach can be summarized by setting up the movement intensity matrix, which describes the intensity of moving from one cell i to another cell j. Note that this definition of M* includes the fishing and natural mortality rates, which addressed in more detail in the next section.

$$M^{\star}_{i\to j} = \begin{cases} D_i/\Delta^2 + 0.5\alpha_i^{(x)}/\Delta, & \text{if cell } j \text{ right of cell } i \\ D_i/\Delta^2 - 0.5\alpha_i^{(x)}/\Delta, & \text{if cell } j \text{ left of cell } i \\ D_i/\Delta^2 + 0.5\alpha_i^{(y)}/\Delta, & \text{if cell } j \text{ above cell } i \\ D_i/\Delta^2 - 0.5\alpha_i^{(y)}/\Delta, & \text{if cell } j \text{ below cell } i \\ F_i, & \text{if cell } j \text{ is 'caught'} \\ M_i, & \text{if cell } j \text{ is 'dead'} \\ -\sum_{j\neq i} M^{\star}_{i,j}, & \text{if } i = j \quad \text{(obviously calculated last in row)} \\ 0, & \text{otherwise} \end{cases}$$

The matrix is constructed from a diffusion field D and an advection vector-field α. The fields are defined by splines applied to the environmental fields. The vector-field for advection is defined as the gradient of a habitat field h, which is defined as:

$$h(i) = S_1^{(h)}(I_1(i)) + \cdots + S_m^{(h)}(I_m(i)) \text{ used as } \alpha(i) = \nabla h(i)$$

Where S denotes spline functions and I denote environmental input fields. As such the advection field is defined in each grid cell and the gradient is approximated by finite differences. The diffusion field is similarly but more directly defined as:

$$\log D(i) = S_1^{(D)}(I_1(i)) + \cdots + S_m^{(D)}(I_m(i))$$

As the environmental input fields are defined only at a discrete grid, then the advection and diffusion fields are also restricted to a discrete grid. This matches well with the conventional tags, which are observed to move from one specific cell (i) to another (j) and possibly with a hidden Markov model approach to the archival tags (Pedersen, 2010). The discrete definition of the advection and diffusion fields does not match well a more continuous approach to modelling the tags. For such approaches to work we need the advection and diffusion fields to be defined – and be differentiable – in all points in our study area. If we imagine that we could find a differentiable representation of each field (e.g. replace I_1 with $\widetilde{I}_1$), then the fields for advection and diffusion would also be continuous and differentiable defined everywhere.

Two approaches were used to represent the environmental field in a differentiable way. First a neural network with 3 inputs (lon, lat, and '1' (representing an intercept)), 15 hidden nodes, and one output (the corresponding environmental value) was set up. Such a neural network has 60 model parameters and after they have been estimated, then the environmental field can be approximated by evaluation the network at any (lon, lat)-point. The network representation is continuous and differentiable, so it can be used as basis for more continuously defined movement models.

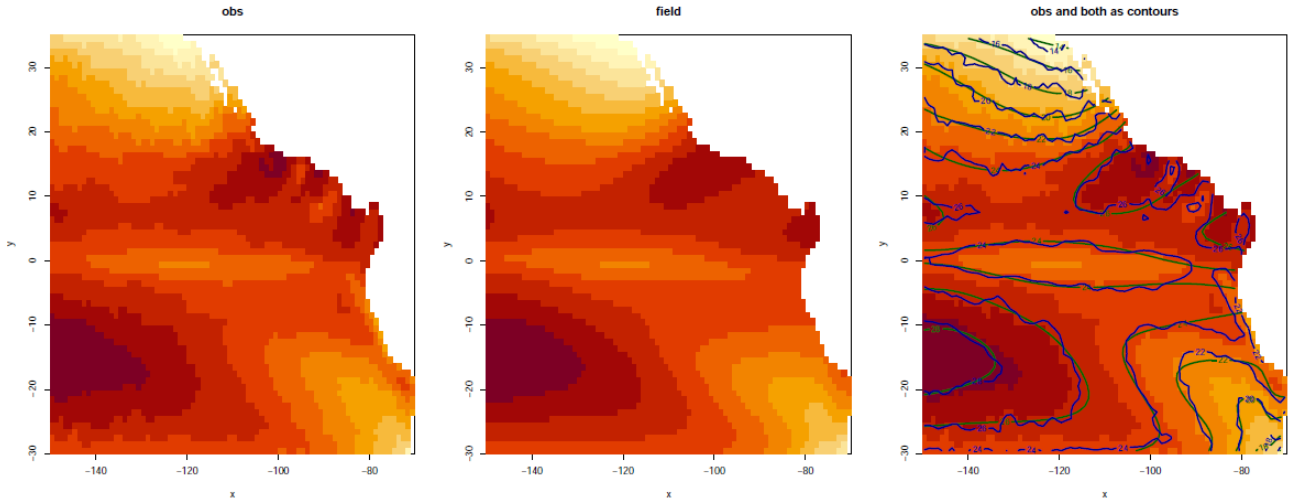

Figure 6: Raw temperature field (left), smooth neural network representation (middle), and compared contours (right).

While the neural network approach does work and gives a parametric approximation of the field it uses a relatively high number of parameters and could possibly be poorly defined outside the data area. The second approach to represent the environmental field in a differentiable way uses local interpolation. The locally interpolated field is defined in any position (lon, lat) by a weighted average of the input field values from a radius 'R' around the position. The distance-weighting of the local points is defined by an iterated cosine function to ensure differentiability (to a high enough order) when observation points are smoothly added or excluded from the average, as the position changes. If the radius is defined to be exactly equal to the distance between neighboring positions, then the value of the differentiable representation evaluated at an actual observation position will be exactly equal to the observed value (because exactly one observation is included), and at any other position the value will be a smooth weighted average of neighboring points.

The advection field is defined as the gradient to the habitat field, and it is possible (easy) to evaluate the gradient of the local interpolation calculations. That gradient is, however, not what we want. If the radius is defined to be equal to the distance between neighboring points, then the gradient of the smooth representation will be zero if evaluated at the position of an observation (because exactly at such a point the calculation only involves taking the average of one point). To get useful gradient fields (one in the longitude direction and one in the latitude direction) the input fields for delta-longitude and delta-latitude were computed from each of the environmental fields and then the local interpolation method applied to get smooth versions. Hence each discretely given environmental field I is converted into 3 smooth fields ($\widetilde{I_{\square}}$, $\widetilde{I_{dx}}$, and $\widetilde{I_{dy}}$). Once these are defined, the gradient of the habitat field itself can be computed and that is what is needed to define advection and diffusion everywhere in a differentiable way.

With the advection and diffusion fields defined it is possible to formulate (and estimate) movement models, which are not defined on a grid. For the archival tag observations, the model becomes:

$$o_t \sim N(\psi_t, \Sigma_o)$$

$$\psi_{t+\Delta t} \sim N(\psi_t + \alpha(\psi_t, t)\Delta t, 2D(\psi_t, t)\Delta t I_{2\times 2})$$

Here ψrepresents the true unobserved position, o is the observed position with observation noise Σ_o. Starting from the known release position the model likelihood is computed via a classic Kalman filter (Harvey, 1990).

For the conventional tags only the release and recapture positions are known. In order to better capture the nonlinearities in the spatial fields and the temporally changing environmental fields a number of intermediate time points are inserted between release and recapture time. The likelihood contribution of each conventional tag is ten computed exactly at for the archival tags. Starting from the release location a Kalman filter is used to step from each timepoint and to the next (updating the distribution of the true unobserved position) and finally evaluating the likelihood of the recapture position. The only difference is that no observations are available at all the intermediate timesteps.

As multiple recovered tags are linked to the same release locations (but various recovery times), the computations are optimized for computation speed and memory allocation by estimating the movement matrices dependent on the unique release locations.

5. Estimation of mortality rates

The recovery of a tag at a given location and time does not only depend on the movement from the release to recapture location, but also the probability of the survival of the fish until being recaptured and the probability of capture at the recapture location by fleet f. Defining the instantaneous fishing mortality $F_{g,t,f}$ in grid cell g at time t of fleet f proportional to the effort of that fleet ($F_{g,t,f}=\lambda_f E_{g,t,f}$) and the instantaneous natural mortality rate M as a constant rate in space and time, allows to calculate the likelihood of recapture of tag h at time m as:

$$\boldsymbol{L}_h(M, \lambda|E) = \frac{F_{g_m,t_m,f}}{M + \sum_{f=1}^{F} F_{g_m,t_m,f}}(1 - e^{-(M+\sum_{f=1}^{F} F_{g_m,t_m,f})\Delta t_m}) \prod_{s=1}^{m-1} e^{-(M+\sum_{f=1}^{F} F_{g_s,t_s,f})\Delta t_s}$$

On the other hand, a conventional tag that was not recovered, is either lost with the death of the fish due to natural causes or still attached to the living fish. Note, that this assumes that the tag would have been reported if it was caught by any fleet (non-reporting=0). Thus, the likelihood of a non-recaptured tag h with assumed maximum age A:

$$\begin{aligned}\boldsymbol{L}_h(M,\lambda|E) = &\frac{M}{M+\sum_{f=1}^{F} F_{g_1,t_1,f}}(1-e^{-(M+\sum_{f=1}^{F} F_{g_1,t_1,f})\Delta t_1}) \\ &+\sum_{a=1}^{A}\frac{M}{M+\sum_{f=1}^{F} F_{g_a,t_a,f}}(1-e^{-(M+\sum_{f=1}^{F} F_{g_a,t_a,f})\Delta t_a})\prod_{s=1}^{a-1} e^{-(M+\sum_{f=1}^{F} F_{g_s,t_s,f})\Delta t_s} \\ &+\prod_{a=1}^{A} e^{-(M+\sum_{f=1}^{F} F_{g_a,t_a,f})\Delta t_a}\end{aligned}$$

Similar to the environmental fields, the effort must be defined in a continuous differentiable space for the Kalman filter approach. Therefore, we used the same local interpolation approach described above for the interpolation of the effort.

6. Population model

A rough estimate of the population biomass in space and time and based on information from the different fleets ($B^r_{g,t,f}$) can be calculated from observed catches and estimated fishing and natural mortality rates corresponding to fleet f by means of the Baranov catch equation:

$$B^r_{g,t,f} = \frac{C_{g,t,f}}{F_{g,t,f}/\left(M+\sum_{f=1}^{F} F_{g,t,f}\right)\left(1-e^{-(M+\sum_{f=1}^{F} F_{g,t,f})\Delta t}\right)}$$

However, this biomass is likely an underestimation of the population biomass as it is only defined for grid cells where the effort and catch are larger than 0. A spatio-temporal model can be used to intra- and extrapolate to those grid cells and areas and derive a smoother estimate of the population biomass in space and time ($B^s_{g,t}$). This model can be defined as a state-space model with the biomass in space and time as the random effect. The random effect biomass can be described by an autocorrelated process in time where the residual spatio-temporal variation in abundance corresponds to a Gaussian Markov random field:

$$\log\left(B^s_{g,t}\right) = \rho\log\left(B^s_{g,t-1}\right) + \epsilon_t$$

$$\text{where } \epsilon_t \sim N(0, Q^{-1}) \text{ and } Q = Q_0 + \delta I.$$

The spatiotemporal tagging and abundance model is implemented in the Template Model Builder (TMB; Kristensen *et al.* 2016) and parameter optimization is done in R 4.0.2 (R Core Team 2020).

7. Results

Among tested environmental fields, such as mixed layer depth, primary productivity, or kinetic energy, (sea surface) sea temperature showed the strongest signal in terms of the habitat preference of SKJ in the EPO. Results indicate that SKJ might prefer SST around 25-26°C and have a lower preference for lower and higher temperatures (left panel in Fig. 7). At the same time, the results suggest that the undirected movement (diffusion) increases with increasing temperature (right panel in Fig. 7).

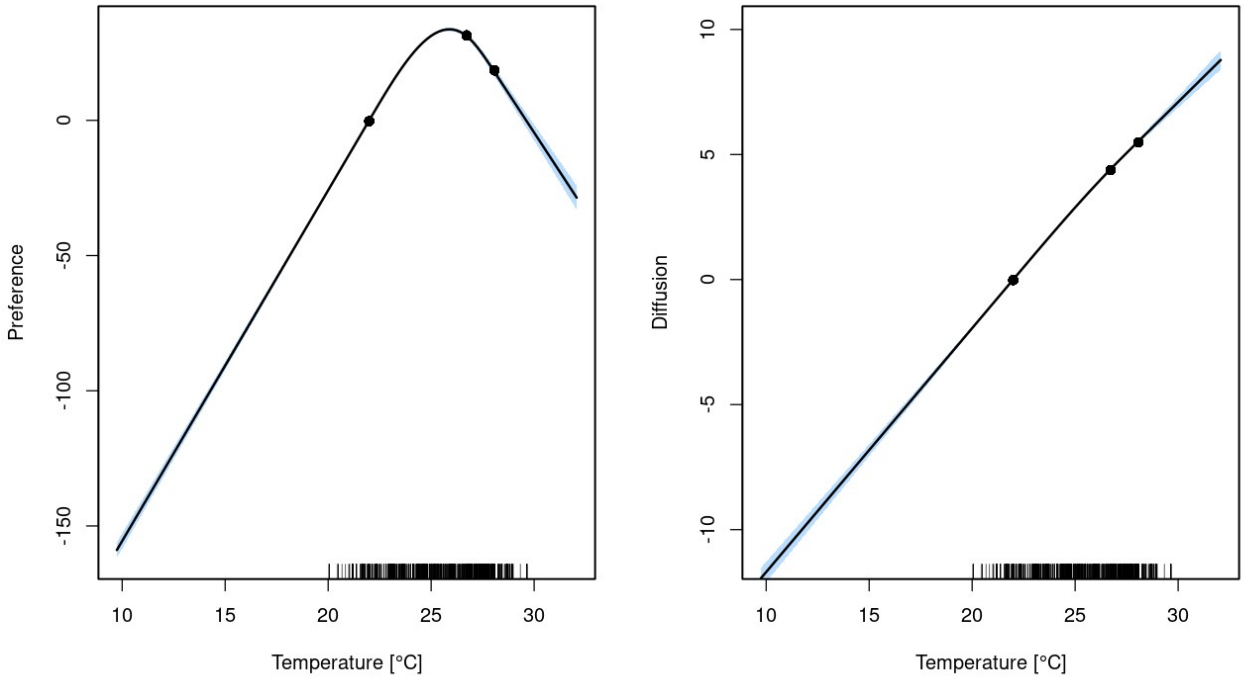

Figure 7: Estimated preference and diffusion as a function of sea surface temperature. The black dots indicate the locations of the 3 knots and the rugs on the x axis indicate "observed" temperature (I.e. temperature values in grid cells were SKJ was released or recaptured).

These temperature preferences and diffusion equate to preferred areas around and south of the equator and around 15° N over a wide range of longitudes (left panel in Fig. 8). Areas with high diffusion are along the coastline of Mexico and Central America, along 10°N and at 10° S at the western boundary of the spatial domain. Note, that Figure 8, shows an annual average and as temperature changes seasonally (Supplementary Fig. S1), so do the preferred habitats.

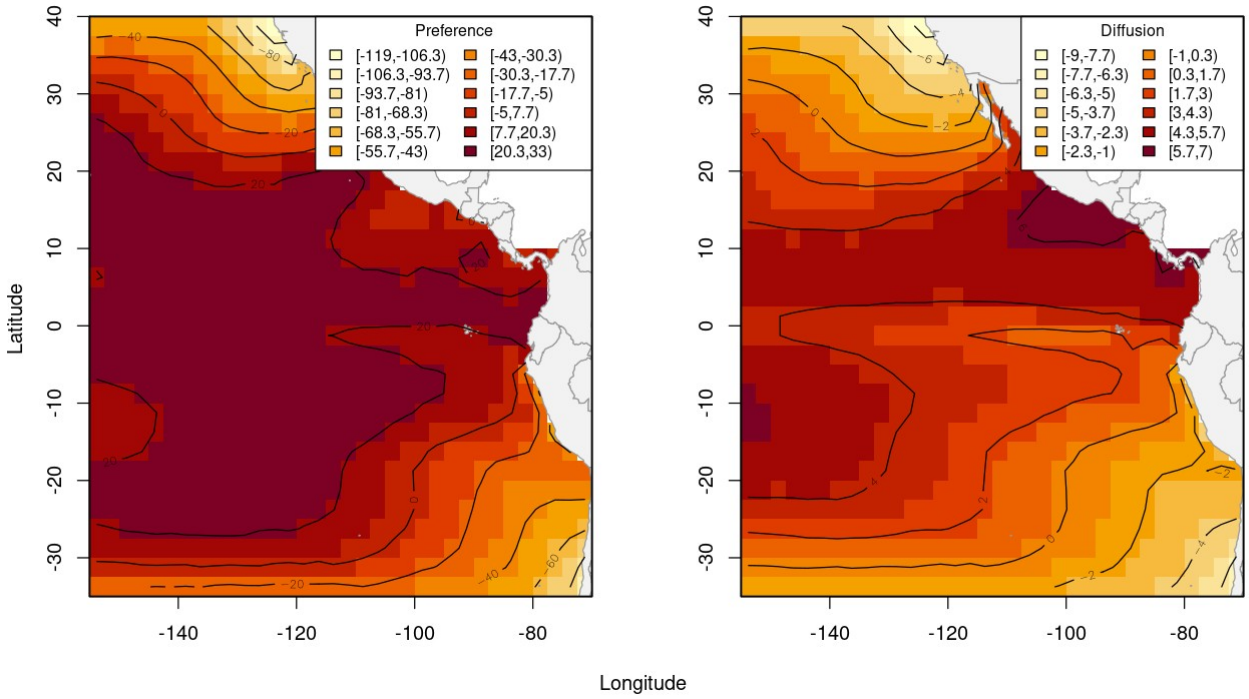

Figure 8: Estimated mean annual habitat preference and diffusion in 2021 based on the Kalman filter model with a 2.5°x 2.5° grid, 12 time steps per year, and natural splines with 3 knots based on sea surface temperature.

These estimate habitat preferences imply stronge r movement directed away from high latitudes towards lower latitudes and away from the coastline (left panel Fig. 9). While the directed movement between 20°N and 20°S is lower, the undirected movement is larger in these areas (Fig. 9).

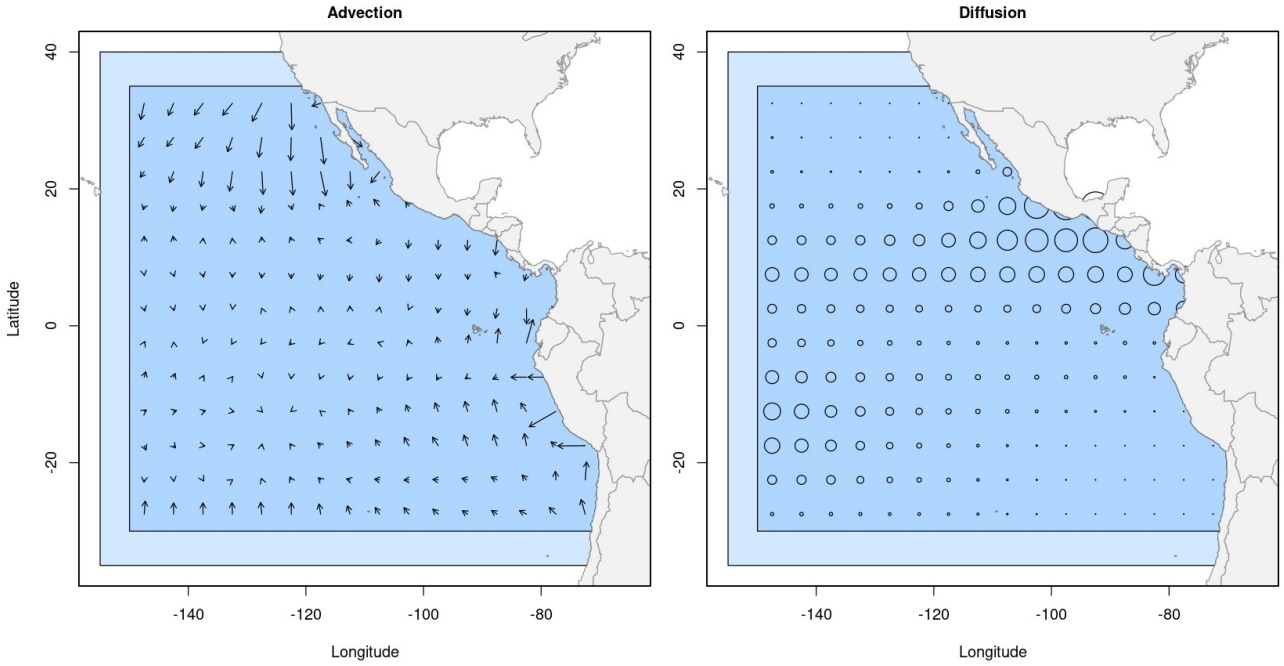

Figure 9: Estimated mean annual advection and diffusion in 2021 based on the Kalman filter model with a 2.5°x 2.5° grid, 12 time steps per year, and natural splines with 3 knots based on sea surface temperature. The length of the arrows indicates the strength of the directed movement (advection), while the size of the circles indicates the square root of the diffusion.

These results are consistent over a range of different assumptions regarding the splines, such as the location and number of knots, as well as the size of the temperature grid. The estimated natural mortality rate is 3.85 year^{-1} (Table 1).

Table 1: Estimated model parameters of the movement model with estimated standard error (SE) and lower and upper 95% confidence intervals (CI).

Parameter	Estimate	SE	Lower CI	Upper CI
α_1	31.2	0.447	30.3	32
α_2	17.9	0.814	16.3	19.5
β_1	4.42	0.0202	4.38	4.46
β_2	5.53	0.042	5.45	5.61
σ_{Obs}	0.396	0.018	0.383	0.411
λ_{dph}	0.000883	0.578	0.000284	0.00274
λ_{sch}	0.00217	0.289	0.00123	0.00383
λ_{log}	0.141	0.0318	0.133	0.15
M	3.85	0.0229	3.68	4.03

Results of the biomass estimation will be presented at the meeting but are too preliminary to be included in the report.

8. Next steps

The next steps include the development of model diagnostics and further sensitivity testing of the estimation of movement and mortality rates as well as the refinement of the population model. For example, double-tagging and seeding experiments suggest that shedding, tagging-related mortality, and non-reporting can amount to substantial rates (Dan Fuller, personal communication). Instead of the auto-correlated process and Gaussian Markov random field, the estimated movement should inform the population model about the spatial and temporal development of SKJ in the EPO. The biomass-aggregated or length-structured population model could then estimate biological reference points for sustainable harvest.

9. Acknowledgments

This work was funded by IATTC. We acknowledge discussions with and helpful comments from Mark Maunder, Dan Fuller, Kurt Schaefer, Kasper Kristensen, and Jon Lopez. We are grateful for the environmental data provided by COPERNICUS (https://www.copernicus.eu/en) and everyone involved in the tagging of SKJ in the EPO including recovery, reporting and analysis of these tags.

10. References

Harvey, A.C. (1990). Forecasting, structural time series models and the Kalman filter.

Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H., Bell, B.M. (2016). TMB: Automatic Differentiation and Laplace Approximation. Journal of Statistical Software, 70(5), 1-21. doi:10.18637/jss.v070.i05

Lam, C. H., Nielsen, A., & Sibert, J. R. (2008). Improving light and temperature based geolocation by unscented Kalman filtering. Fisheries Research, 91(1), 15-25.

Pedersen, M.W. (2010). Hidden Markov modelling of movement data from fish. DTU Informatics.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Thorson, J.T., Jannot, J., and Somers, K., 2017. Using spatio-temporal models of population growth and movement to monitor overlap between human impacts and fish populations. Journal of Applied Ecology, 54(2), 577-587.

Thorson, J.T., Barbeaux, S.J., Goethel, D.R., Kearney, K.A., Laman, E.A., Nielsen, J.K., Siskey, M.R., Siwicke, K. and Thompson, G.G., 2021. Estimating fine-scale movement rates and habitat preferences using multiple data sources. Fish and Fisheries, 22(6), pp.1359-1376.

11. Supplementary information

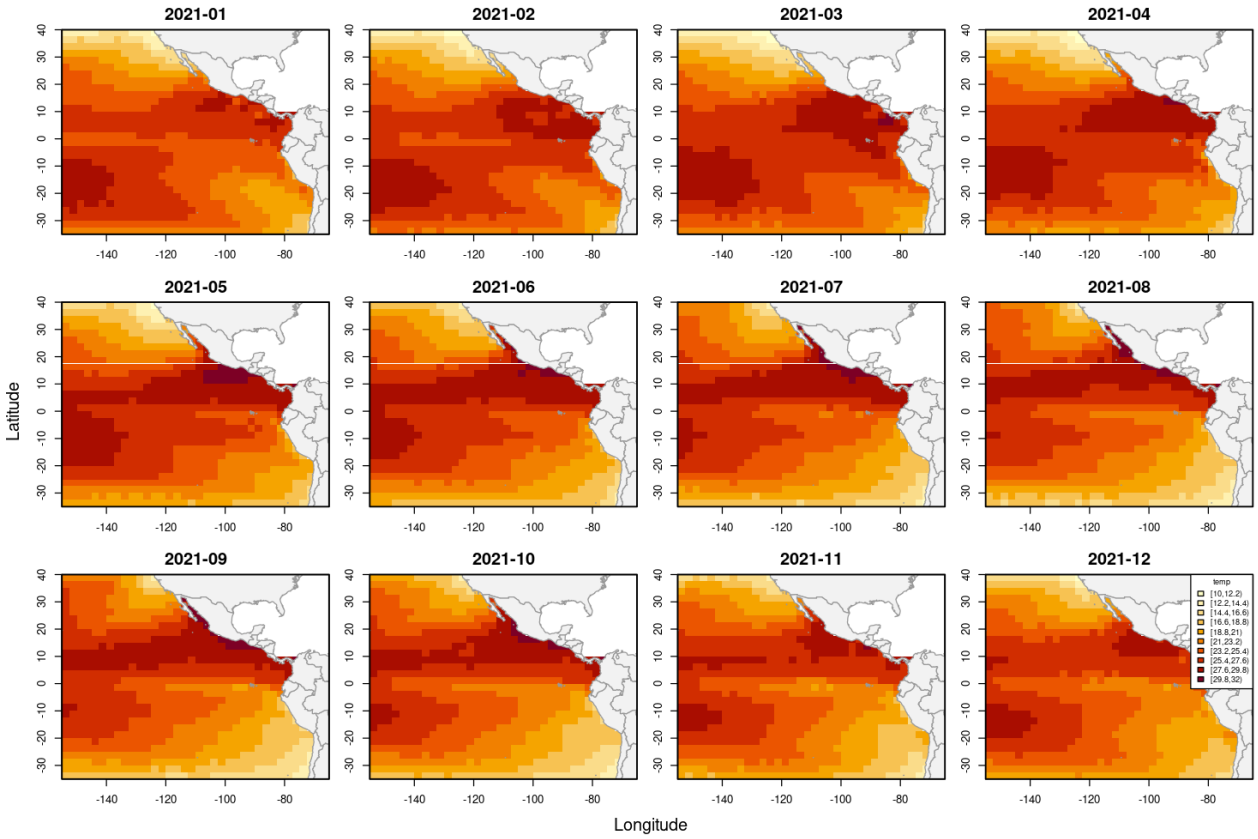

Figure S1: Monthly sea surface temperature (SST) for the EPO in 2021 on a 2.5° grid.

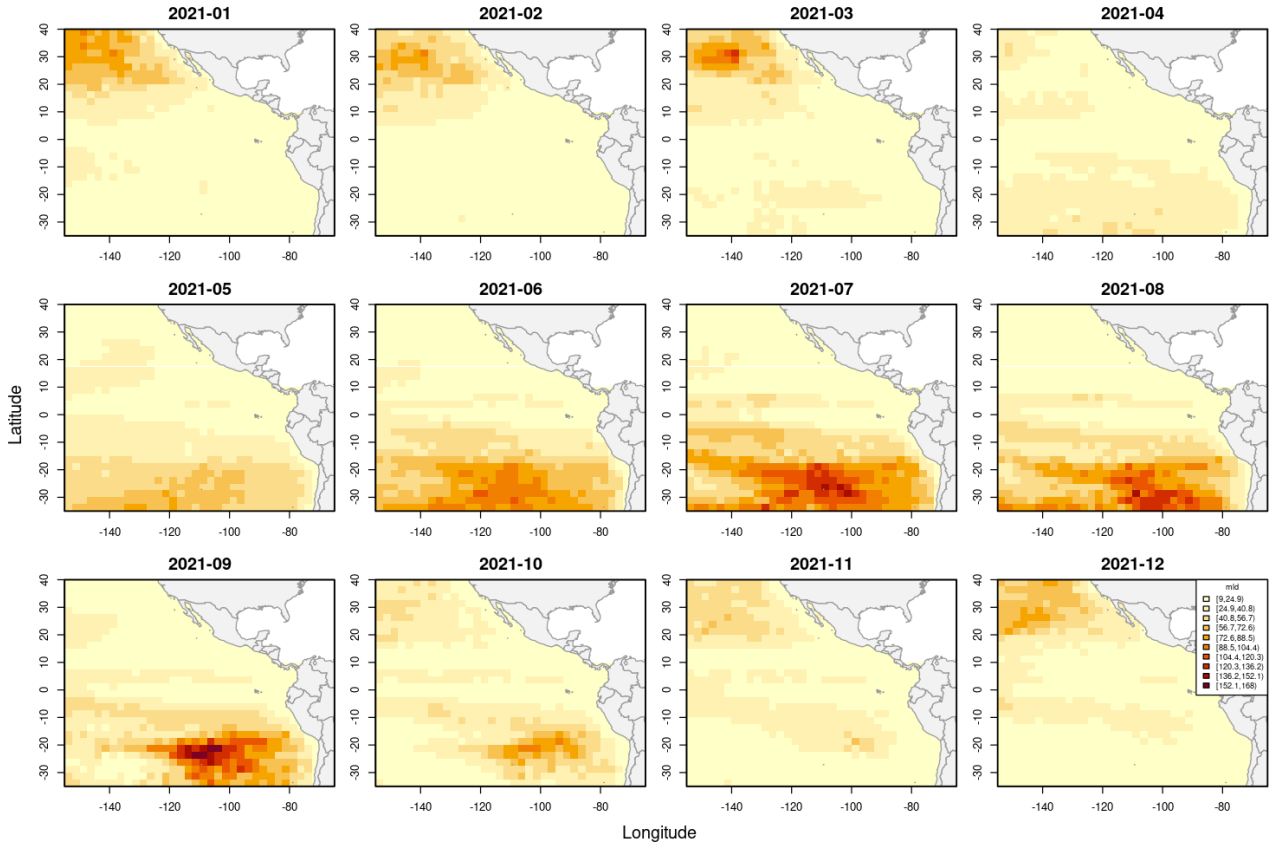

Figure S2: Monthly mixed layer depth (MLD) for the EPO in 2021 on a 2.5° grid.