Supplementary material

Changing windows of opportunity: past and future climate-driven shifts in temporal persistence of kingfish (Seriola lalandi) oceanographic habitat within south-eastern Australian bioregions

Curtis ChampionA,B,D, Alistair J. HobdayB,C, Xuebin ZhangB, Greta T. PeclA,C and Sean R. TraceyA

AInstitute for Marine and Antarctic Studies, Hobart, Tas. 7001, Australia.
BCSIRO Oceans and Atmosphere, Hobart, Tas. 7000, Australia.
CCentre for Marine Socioecology, Hobart, Tas. 7001, Australia.
DCorresponding author. Email: curtis.champion@utas.edu.au

Detailed description of kingfish oceanographic habitat modelling

The oceanographic habitat preference of kingfish from eastern Australia was described by applying generalised additive mixed modelling (GAMM) using the logistic link function to relate the binomially distributed response variable (i.e. presence or pseudo-absence) to environmental predictors. Information on fishing effort was not available in the tagging database, so calendar year was included as a random effect to account for inter-annual variability in kingfish catch per unit effort. To optimise smoothing functions and avoid over-fitting to the data, penalised regression spline type smoothers of moderate rank were applied using generalised cross validation. However, these were removed from individual predictors if their estimated degrees of freedom were approximately equal to 1, which indicates linearity with the log-of-odds transformed response variable (Zuur et al. 2009). The resulting GAMM available for selection has the form:

\[ \text{Response} = s(\text{SST}) + s(\text{SLA}) + s(\text{EKE}) + (1|\text{Year}) \]

where \( \text{Response} \) is the probability of kingfish occurrence modelled as a function of sea surface temperature (SST), sea level anomaly (SLA) and eddy kinetic energy (EKE), with \( \text{Year} \) included as a random factor. Smoothers are denoted by \( s \).

Forward model selection was applied using an information theoretic approach to identify single term additions from the available environmental predictors that most improved model quality (Warren and Seifert 2011). The resulting set of exploratory models contained nested covariate combinations of increasing complexity (Table S1), and the model in this set with the lowest Akaike information criterion (AIC) value was identified as the most parsimonious model.

Table S1. Summary of the full model and nested alternatives assessed using an Akaike information criterion (AIC)-informed model selection procedure on covariate combinations of increasing complexity

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Model</th>
<th>Variable added</th>
<th>ΔAIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>( s(\text{SST}) + (1</td>
<td>\text{Year}) )</td>
<td>–</td>
</tr>
<tr>
<td>1</td>
<td>( s(\text{SST}) + s(\text{EKE}) + (1</td>
<td>\text{Year}) )</td>
<td>EKE</td>
</tr>
<tr>
<td>2</td>
<td>( s(\text{SST}) + s(\text{SLA}) + s(\text{EKE}) + (1</td>
<td>\text{Year}) )</td>
<td>SLA</td>
</tr>
<tr>
<td>3</td>
<td>( s(\text{SST}) + \text{SLA} + s(\text{EKE}) + (1</td>
<td>\text{Year}) )</td>
<td>*Smoother removed from SLA</td>
</tr>
</tbody>
</table>
Spatial and temporal autocorrelation was a concern in the present study because presence points were recorded by recreational anglers whose fishing effort may be spatiotemporally biased (e.g. favour fishing locations or fish more on weekends or holidays). Autocorrelation was evaluated using spatial and temporal variograms to relate the semi-variance of points to the spatial (°) and temporal (days) distance separating them (Zuur et al. 2013). Cut-off distances were chosen to reflect the spatial and temporal limits that autocorrelation would arise from angler bias. Dates of fish captures were converted to Julian days in order to create a temporal semi-variogram with a cut-off distance of 5 days. Coordinates of fish captures were used to create a spatial semi-variogram with a cut-off distance of 1° (~111 km). In exploratory analyses, both spatial and temporal correlation was judged to be consistent across distances (Fig. S1), except at fine spatial scales (0.1–0.3°) where there was lower correlation (higher semi-variance) than at other distances. This is likely to reflect the spatial influence of pseudo-absence points existing close to presence points (i.e. between 0.1 and 0.2°), resulting in increased residual variation at fine spatial scales where a binary response characterises relatively similar environmental habitats. Regardless, there was no evidence to suggest positive spatial or temporal autocorrelation in the present study.

The accuracy and predictive skill of the optimal model was evaluated using k-fold cross-validation. This was done by randomly partitioning the full dataset into five subsets (k = 5) containing an equal number of presence points and a random selection of 10 000 pseudo-absences (Barbet-Massin et al. 2012). To compute a set of confusion matrices for calculating measures of model accuracy (Swets 1988), the optimal model was trained on each of the four subsets and each model tested against the 5th subset. Five-fold cross-validation was selected because of concern that too few presence data would be used to create the evaluation models if data were partitioned into a greater number of folds (Smith et al. 2017). The area under the receiver operating characteristic curve (AUC) and true skill statistic (TSS) are appropriate measures of model accuracy for predictions of species presence and absence in geographic space (Allouche et al. 2006), and are commonly used in combination when evaluating overall model skill (Brodie et al. 2015). Rates of true positive (sensitivity) and true negative (specificity) predictions were used to calculate the mean AUC value. The AUC avoids the need to assume an arbitrary cut-off probability to differentiate between predictions of suitable and unsuitable oceanographic habitat, and is thus a valuable measure of the accuracy of species distribution models (Elith et al. 2006). AUC values range from 0 to 1, where an AUC of 0.5 indicates the prediction is no better than random and an AUC of 0.8 indicates a very good prediction.
greater than 0.8 indicates good model accuracy (Araújo et al. 2005; Swets 1988). Additionally, the mean TSS was calculated as an alternative, threshold dependent, measure of model accuracy obtained from average measures of model sensitivity and specificity (i.e. TSS = sensitivity + specificity – 1). TSS values ranges from −1 to 1, where 0 reflects models with no predictive skill. These procedures revealed that the optimal model had good predictive accuracy (mean AUC = 0.887 ± 0.002 s.d.) according to the AUC interpretation criteria of Swets (1988), and that predictive skill (mean TSS = 0.645 ± 0.013 s.d.) exceeded the acceptable standard for conservation planning applications (Pearce and Ferrier 2000). Mean values of the TSS and AUC statistics indicate that the optimal model contained an appropriate number and combination of environmental predictors to effectively describe suitable environmental habitat for kingfish from south-east Australia.

Fig. S2. (a) Distribution of habitat suitability values from Atlas of Living Australia kingfish-occurrence records \((n = 22; \text{min} = 0.196)\) matched with day-specific habitat predictions used to differentiate between ‘suitable’ and ‘unsuitable’ oceanographic habitats (i.e. suitable \(\geq 0.196 <\) unsuitable) for temporal-persistence analyses. (b) Example of a day-specific kingfish habitat prediction matched with an Atlas of Living Australia kingfish-occurrence record (red asterisk) used to create the distribution in a.

Fig. S3. Partial effects of (a) sea surface temperature (SST), (b) sea level anomaly (SLA) and (c) eddy kinetic energy (EKE) on the fitted values of the optimal kingfish habitat model (GAMM) detailed in C. Champion, S. R. Tracey, G. T. Pecl, and A. J. Hobday (in review). Dashed lines denote 95% confidence intervals and rugs on the x-axes indicate presence and pseudo-absence data for each predictor.
Table S2. Summary of results for the kingfish oceanographic habitat suitability model presented in C. Champion, S. R. Tracey, G. T. Pecl, and A. J. Hobday (in review) and utilised within the present study

Smoothing factors are denoted by \( s \). Mean Area Under the receiver-operating Curve (AUC; scale 0–1) and True Skill Statistic (TSS; scale –1 to 1) are indicative of the predictive accuracy of the model. EKE, eddy kinetic energy; SLA, sea level anomaly; SST, sea surface temperature.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Effective degrees of freedom</th>
<th>Coefficient estimate</th>
<th>( P )-value</th>
<th>Mean AUC (±s.d.)</th>
<th>Mean TSS (±s.d.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( s(\text{SST}) )</td>
<td>5.01</td>
<td>–0.25</td>
<td>&lt;0.001</td>
<td>0.887 (0.002)</td>
<td>0.645 (0.013)</td>
</tr>
<tr>
<td>SLA</td>
<td>–</td>
<td>1.21</td>
<td>0.040</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( s(\text{EKE}) )</td>
<td>7.78</td>
<td>2.28</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year(\text{intercept})</td>
<td>–</td>
<td>–5.55</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

References


