Marine and Freshwater Research 2019, **70**, 33–42
http://dx.doi.org/10.1071/MF17387_AC

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 Champion[A,B,D], *Alistair J. Hobday*[B,C], *Xuebin Zhang*[B], *Gretta T. Pecl*[A,C] *and Sean R. Tracey*[A]

[A]Institute for Marine and Antarctic Studies, Hobart, Tas. 7001, Australia.

[B]CSIRO Oceans and Atmosphere, Hobart, Tas. 7000, Australia.

[C]Centre for Marine Socioecology, Hobart, Tas. 7001, Australia.

[D]Corresponding 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:

$$Response = s(SST) + s(SLA) + s(EKE) + (1|Year)$$

where *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 *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

Smoothing factors are indicated by ‘*s*’. Delta-AIC values denote differences in AICs between models relative to the most parsimonious model (bold). EKE, eddy kinetic energy; SLA, sea level anomaly; SST, sea surface temperature

Iteration	Model	Variable added	ΔAIC
0	$s(SST) + (1\|Year)$	–	194.432
1	$s(SST) + s(EKE) + (1\|Year)$	EKE	1.903
2	$s(SST) + s(SLA) + s(EKE) + (1\|Year)$	SLA	1.807
3	$\boldsymbol{s(SST) + SLA + s(EKE) + (1\|Year)}$	*Smoother removed from SLA	0

http://dx.doi.org/10.1071/MF17387_AC

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.

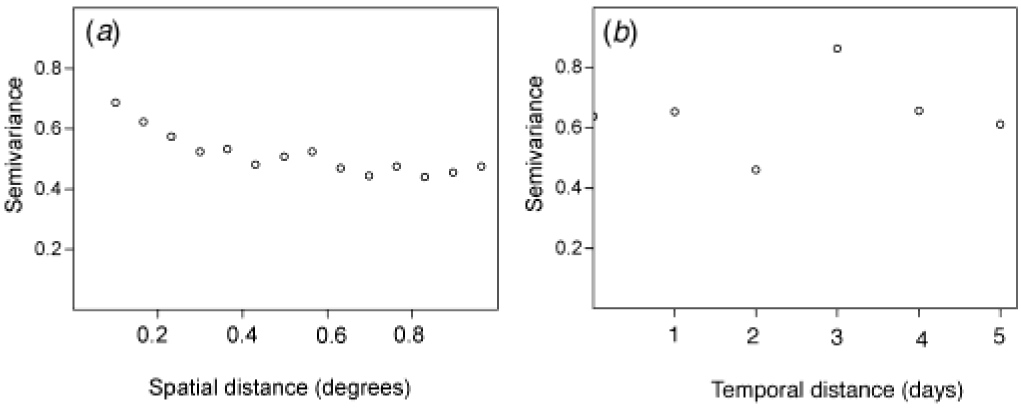

Fig. S1. (*a*) Spatial and (*b*) temporal variograms, showing the correlation of kingfish presence points with respect to the spatial and temporal distances that separate them. A straight line shows spatial independence, and autocorrelation is typically shown by a positive slope, usually at small *x*-values.

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 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.

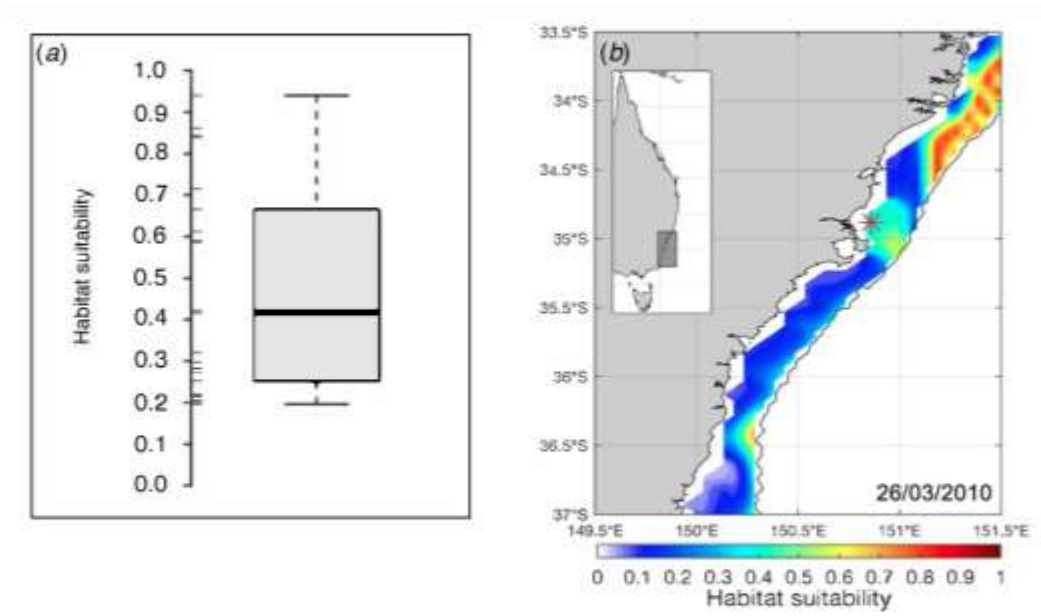

Fig. S2. (*a*) Distribution of habitat suitability values from Atlas of Living Australia kingfish-occurrence records (*n* = 22; min = 0.196) matched with day-specific habitat predictions used to differentiate between 'suitable' and 'unsuitable' oceanographic habitats (i.e. suitable ≥ 0.196 < unsuitable) 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*.

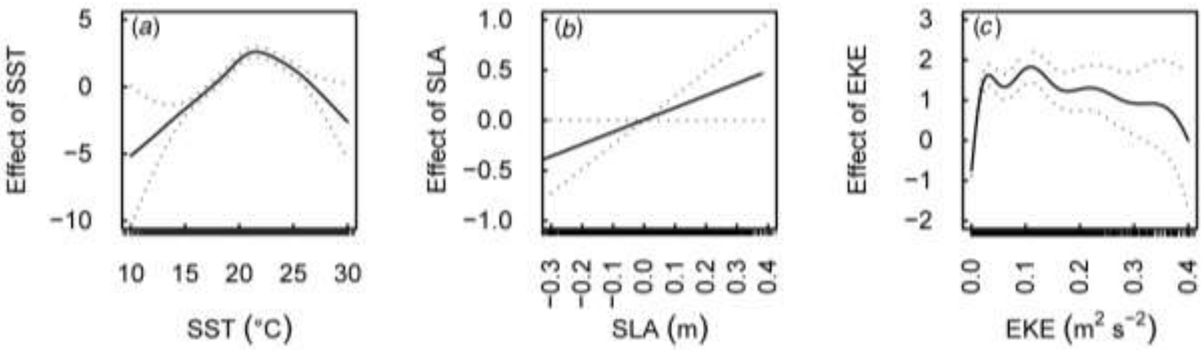

Fig. S3. Partial effects of (*a*) sea surface temperature (SST), (*b*) sea level anomaly (SLA) and (*c*) eddy kinetic energy (EKE) on the fitted values of the optimal kingfish habitat model (GAMM) detailed in C. Champion, S. R. Tracey, G. T. Pecl, and A. J. Hobday (in review). Dashed lines denote 95% 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*. Mea1 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

Variable	Effective degrees of freedom	Coefficient estimate	*P*-value	Mean AUC (±s.d.)	Mean TSS (±s.d.)
s(SST)	5.01	–0.25	<0.001	0.887 (0.002)	0.645 (0.013)
SLA	–	1.21	0.040		
s(EKE)	7.78	2.28	<0.001		
$\text{Year}_{\text{(intercept)}}$	–	–5.55	<0.001		

References

Allouche, O., Tsoar, A., and Kadmon, R. (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology* **43**(6), 1223–1232. doi:10.1111/j.1365-2664.2006.01214.x

Araújo, M. B., Pearson, R. G., Thuiller, W., and Erhard, M. (2005). Validation of species–climate impact models under climate change. *Global Change Biology* **11**(9), 1504–1513. doi:10.1111/j.1365-2486.2005.01000.x

Barbet-Massin, M., Jiguet, F., Albert, C. H., and Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution* **3**(2), 327–338. doi:10.1111/j.2041-210X.2011.00172.x

Brodie, S., Hobday, A. J., Smith, J. A., Everett, J. D., Taylor, M. D., Gray, C. A., and Suthers, I. M. (2015). Modelling the oceanic habitats of two pelagic species using recreational fisheries data. *Fisheries Oceanography* **24**(5), 463–477. doi:10.1111/fog.12122

Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R. J., Huettmann, F., Leathwick, J. R., Lehmann, A., Li, J., Lohmann, L. G., Loiselle, B. A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J. M. M., Townsend Peterson, A., Phillips, S. J., Richardson, K., Scachetti-Pereira, R., Schapire, R. E., Soberón, J., Williams, S., Wisz, M. S., and Zimmermann, N. E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**(2), 129–151. doi:10.1111/j.2006.0906-7590.04596.x

Pearce, J., and Ferrier, S. (2000). Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling* **133**(3), 225–245. doi:10.1016/S0304-3800(00)00322-7

Smith, J. A., Cornwell, W. K., Lowry, M. B., and Suthers, I. M. (2017). Modelling the distribution of fish around an artificial reef. *Marine and Freshwater Research* **68**(10), 1955–1964 doi:10.1071/MF16019.

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science* **240**(4857), 1285. doi:10.1126/science.3287615

Warren, D. L., and Seifert, S. N. (2011). Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications* **21**(2), 335–342. doi:10.1890/10-1171.1

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., and Smith, G. M. (2009) 'Mixed Effects Models and Extensions in Ecology with R.' (Springer, New York, NY, USA.)

Zuur, A. F., Hilbe, J., and Iepo, E. (2013) 'A Beginner's Guide to GLM and GLMM with R: a Frequentist and Bayesian Perspective for Ecologist.' (Highland Statistics: Newburgh, UK.)