Supplementary Material S3: Closure Creation and Evaluation

Here we detail how we created the spatially static and dynamic leatherback turtle closures in Phase 2 of the simulation, using the simulated observer data collected in Phase 1.

Static closures

Locations of the observed turtle bycatch events were used to create the 'Static-obs' (KDE) closures, using the 'kde2d' function in the 'MASS' R package (Venables and Ripley 2002), specifying n = 100 grid points in each direction. An iterative process was then used to find the value of the KDE function that enclosed the threshold percentage of observed bycatch (70% or 90%). We used this value to create a spatial contour (the closure boundary), and then used the 'extract' function from the 'raster' R package (Hijmans 2019) to identify which cells where inside the closure.

We used a binomial GAM to create the 'Static-pred' (GAM) static closure, with the form:

$$Pres = s(Lon, Lat)$$

Where *Pres* is probability of presence, and *s* is a thin plate regression spline. This was fitted using the 'mgcv' R package (Wood 2011). The maximum wiggliness of the bivariate smoother was specified k = 30, to avoid complex surfaces unsuitable for a spatial closure boundary. The fitted model was predicted for the entire domain, and all cells with *Pres* values > threshold value (0.1) were defined as inside the closure.

Dynamic closures

EcoCast

EcoCast is a decision support tool that calculates fishing suitability based on estimated habitat suitability of target and bycatch species. Each species is weighted according to the desired contribution to the suitability metric. The EcoCast metric can vary between -1 and +1 (or -1 and 0, if it only contains bycatch species), with more negative values indicating increased risk of bycatch and thus areas fishers should avoid.

The EcoCast equation in our simulation had the form (Hazen *et al.* 2018; Welch *et al.* 2020):

$$E = Pres_{sf} \times w_{sf} + Pres_{lb} \times w_{lb} + Pres_{bs} \times w_{bs}$$

E is the EcoCast risk value, $Pres_{sf}$ is the predicted probability of presence for swordfish (see next section), w_{sf} is the weighting for swordfish, and likewise for leatherback turtles (*lb*) and blue sharks (*bs*). Swordfish was a target species, so had a positive weighting (i.e. lowered the risk value), and the others were bycatch species so had a negative weighting (i.e. increased the risk value). The absolute value of the weighting values must sum to one. We evaluated two weighting schemes: a multi-species scheme and a single-species 'turtle' scheme. The multi-species scheme was used in the Dyn-multi$_s$ and Dyn-multi$_m$ scenarios (**Table 1**) and weighted the three species approximately evenly: w_{sf} = 0.25, w_{lb} = -0.50, w_{bs} = -0.25. The turtle scheme was used in Dyn-turt$_s$, Dyn-turt$_m$, and Dyn-turt$_s$P, and calculated risk based predominantly on turtle presence: w_{sf} = 0.10, w_{lb} = -0.80, w_{bs} = -0.10. The 'even' multi-species weights do not have the same absolute value because the mean probability of presence of turtles was considerably lower than the other two species. *E* was calculated for each day of Phase 3, with a fixed threshold (see below) defining which cells were unsuitable for fishing (i.e. closed) each day.

Fitting the BRTs for EcoCast

EcoCast requires predicted probability of presence for each species. We used the same approach as the real-world EcoCast, which uses binomial boosted regression trees (BRTs). These were fitted in Phase 2 to the simulated observer data from Phase 1 (with catches of each species as presence-absence data), using a BRT with the form:

$$Pres = SST + SST_{sd} + ILD + SSH + SSH_{sd} + EKE + Curl + BV + Lunar + Z + Z_{sd} + Lat$$

Pres is the probability of presence, and *SST*, SST_{sd}, *ILD*, *SSH*, *EKE*, *Curl*, Z_{sd}, and *Lat* are as defined in **Table S2.1**. *Z* is bottom depth, SSH_{sd} is the standard deviation of *SSH* (m), *BV* is bulk Brunt-Vaisala frequency (s^{-1}), and *Lunar* is lunar illumination (%). These covariates represent the set used for the real-world EcoCast, with the exception of *Lat*, which we added to provide opportunity to fit the spatially-structured LB1 model, and the exclusion of four current velocity and wind stress covariates. The number of covariates needed to be reduced to prevent occasional fitting issues, caused by the low number of simulated turtle bycatch events, and model selection showed the four velocity/stress covariates were least important. To ensure the closure was generalisable, we used the same set of covariates for each species and for each operating model and iteration.

The BRTs were fitted using a learning rate of 0.01, a tree complexity of 2, step size of 50, and a bag fraction of 0.6 (Elith *et al.* 2008), using the function 'gbm.step' in the 'dismo' R package (Hijmans *et al.* 2017). Due to the low number of turtle bycatch events, we needed to reduce the learning rate to 0.003, increase the bag fraction to 0.75, and reduce the step size to 20 for the turtle BRTs.

Closure thresholds

A threshold value for EcoCast was used to define a level of unsuitable risk, and thus which areas were closed. We defined a fixed threshold for 'good quality' leatherback turtle habitat

(*Pres* = 0.1), based on the frequency of this habitat in the LB1 and LB2 operating model predictions, and sensible amount of area closed. This value was used to calculate a threshold of the EcoCast metric to define the closure, which varied with each operating model and iteration and among EcoCast weightings. The threshold was calculated by randomly sampling 200 locations from each of ten randomly sampled days, and recording the turtle probability of presence (defined by the turtle operating model) and the EcoCast metric for that operating model and iteration. We then found the EcoCast value above which 50% or 90% of the 0.1 probability locations existed (**Table 1**). These two values were the thresholds used to define the closures, given moderate (50%) or strict (90%) protection of good quality turtle habitat. An example of EcoCast and the threshold closures are shown in **Fig. S3.1**.

EcoCast validation: simulated vs real-world

An important part of our simulation was ensuring the simulated EcoCast was similarly accurate at indicating bycatch risk as the real-world EcoCast. If the simulated EcoCast was more accurate, our MSE could overestimate the performance of dynamic spatial closures. We evaluated accuracy by comparing simulated and real-world model performance for the leatherback turtle species distribution models used to calculate EcoCast. We used k-folds cross validation to measure performance of our simulated models, as described in **Supp. Material S2**, and compared this with k-fold results for the real-world EcoCast reported in Welch et al. (2020). Cross validation for the simulated turtle models was done in Phase 2, at each level of the operating model, iteration, and observer program size. It was not computationally feasible to run k-folds validation for the other two species, but goodness-of-fit (AUC of fitted and observed for the full model) was recorded for a handful of swordfish and blue shark models. These AUC values were: 0.80 (SF1), 0.76 (SF2), and 0.82 (BS1); noting that these can be higher than values from k-folds cross validation. Performance was generally similar between the simulated and real-world EcoCasts. The cross-validated AUC of models used in the simulated EcoCast ranged from 0.66-0.83 for LB1 and LB2 (**Table S3.1**), and was 0.77 for the turtle model used in the real-world EcoCast (Welch *et al.* 2020). The AUC of the real-world swordfish and blue shark models were 0.72 and 0.75 respectively (Welch *et al.* 2020), which were similar to our approximations. During cross-validation we also calculated the true skill statistic (TSS), which ranged from 0.30-0.63 for the LB1 and LB2 models (**Table S3.1**), and was 0.41 for the real-world EcoCast (Welch *et al.* 2020).

Table S3.1. Summary of the k-folds (k=10) cross validation on the leatherback turtle binomial BRTs fitted in our simulation, for each combination of operating model (OM), and for the 20% and 50% observer coverage simulations. Performance was measured using the area under the receiver operating Curve (AUC) and true skill statistic (TSS). AUC ranges from 0-1, and values > 0.5 indicate better than random predictability. TSS is a measure of the accuracy predicting both presences and absences, and ranges from -1 to +1, where +1 indicates perfect accuracy and > 0 better than random predictability.

LB OM	SF OM	AUC-20%	AUC-50%	TSS-20%	TSS-50%
LB1	SF1	0.815 (0.010)	0.820 (0.003)	0.577 (0.024)	0.549 (0.005)
LB1	SF2	0.834 (0.012	0.832 (0.015)	0.630 (0.033)	0.588 (0.029)
LB2	SF1	0.659 (0.033)	0.672 (0.015)	0.312 (0.050)	0.302 (0.026)
LB2	SF2	0.669 (0.020)	0.700 (0.012)	0.340 (0.031)	0.349 (0.023)
LB1seas	SF1	0.847 (0.014)	0.855 (0.006)	0.643 (0.025)	0.624 (0.014)
LB1seas	SF2	0.862 (0.013)	0.865 (0.010)	0.682 (0.025)	0.654 (0.017)
LB2scal	SF1	0.601 (0.011)	0.617 (0.016)	0.294 (0.018)	0.232 (0.011)
LB2scal	SF2	0.627 (0.023)	0.630 (0.017)	0.294 (0.036)	0.250 (0.021)

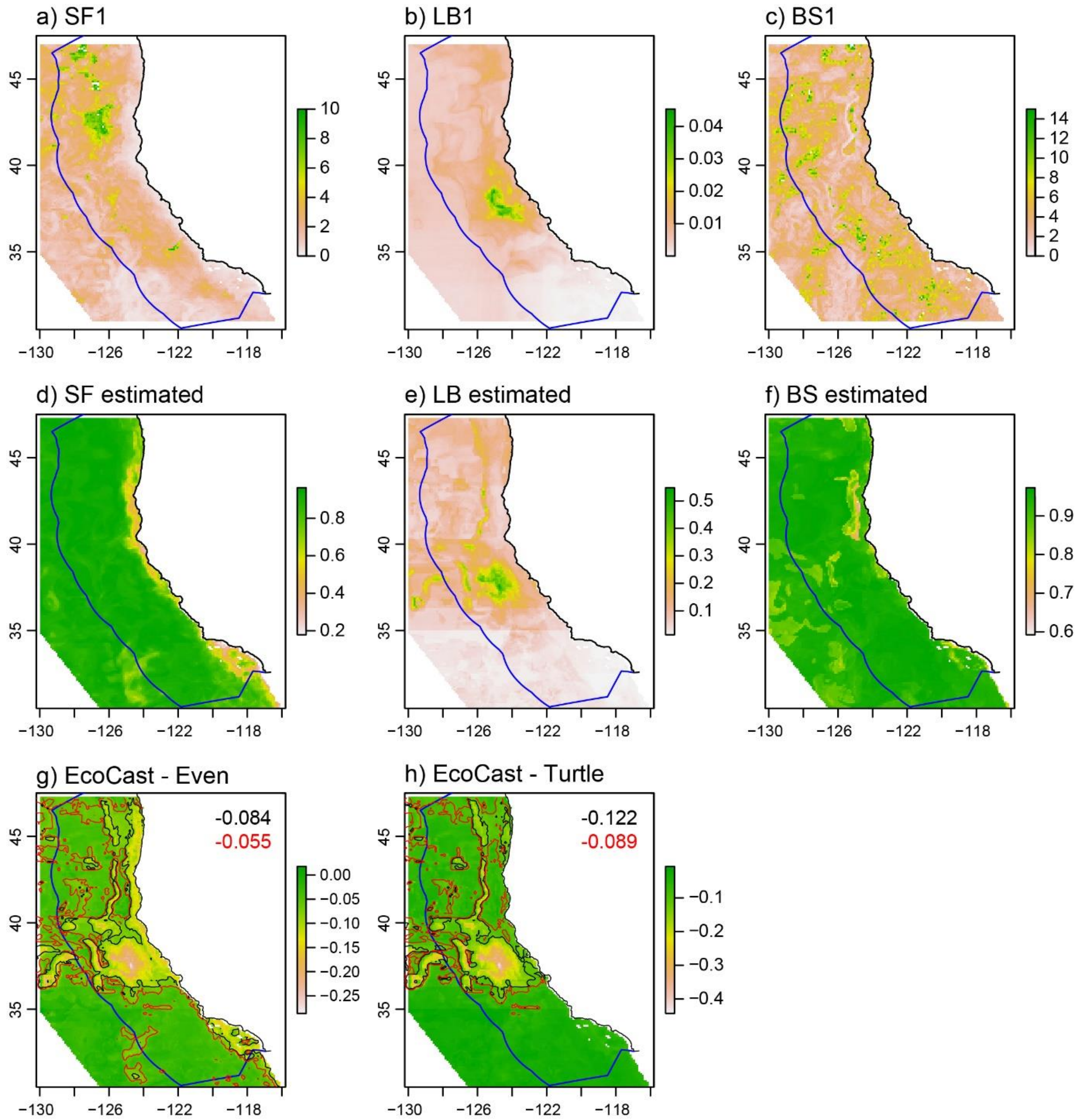

Fig. S3.1. Operating model catch distributions, estimated probability of presence, and calculated EcoCast, for an example date (1996-10-13). Panels a-c) show the distributions of catch for the three species (i.e. the LB1-SF1-BS1 operating model). Panels d-f) show the probability of presence estimated by the binomial BRT in Phase 2, using the simulated observer data from Phase 1. Panels g-h) show the EcoCast surfaces calculated using the estimated presence in d-f, with the two closure thresholds (moderate = black contour and text, strict = red contour and text). Areas inside those contour lines (i.e. with EcoCast values more negative than the threshold) were closed to fishing on this date. In this example, swordfish and blue sharks were estimated to be equally present in most of the domain (d, f), except lower presence inshore for swordfish. Thus, the ‘evenly’ weighted EcoCast surface looks very similar to the ‘turtle’ weighted EcoCast, except close inshore which was poor swordfish habitat and thus deemed less suitable for fishing when swordfish was equally weighted (g).

References

Elith, J., Leathwick, J.R., and Hastie, T. (2008) A working guide to boosted regression trees. *Journal of Animal Ecology* **77**(4), 802-813. doi:10.1111/j.1365-2656.2008.01390.x

Hazen, E.L., Scales, K.L., Maxwell, S.M., Briscoe, D.K., Welch, H., Bograd, S.J., Bailey, H., Benson, S.R., Eguchi, T., Dewar, H., Kohin, S., Costa, D.P., Crowder, L.B., and Lewison, R.L. (2018) A dynamic ocean management tool to reduce bycatch and support sustainable fisheries. *Science Advances* **4**(5), eaar3001. doi:10.1126/sciadv.aar3001

Hijmans, R.J. (2019). raster: Geographic Data Analysis and Modeling. R package version 2.8-19. https://CRAN.R-project.org/package=raster.

Hijmans, R.J., Phillips, S., Leathwick, J., and Elith, J. (2017). dismo: Species Distribution Modeling. R package version 1.1-4. https://CRAN.R-project.org/package=dismo.

Venables, W.N., and Ripley, B.D. (2002) 'Modern applied statistics with S.' 4th edn. (Springer: New York)

Welch, H., Brodie, S., Jacox, M.G., Bograd, S.J., and Hazen, E.L. (2020) Decision-support tools for dynamic management. *Conservation Biology* **34**(3), 589-599. doi:10.1111/cobi.13417

Wood, S.N. (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73**(1), 3-36. doi:10.1111/j.1467-9868.2010.00749.x