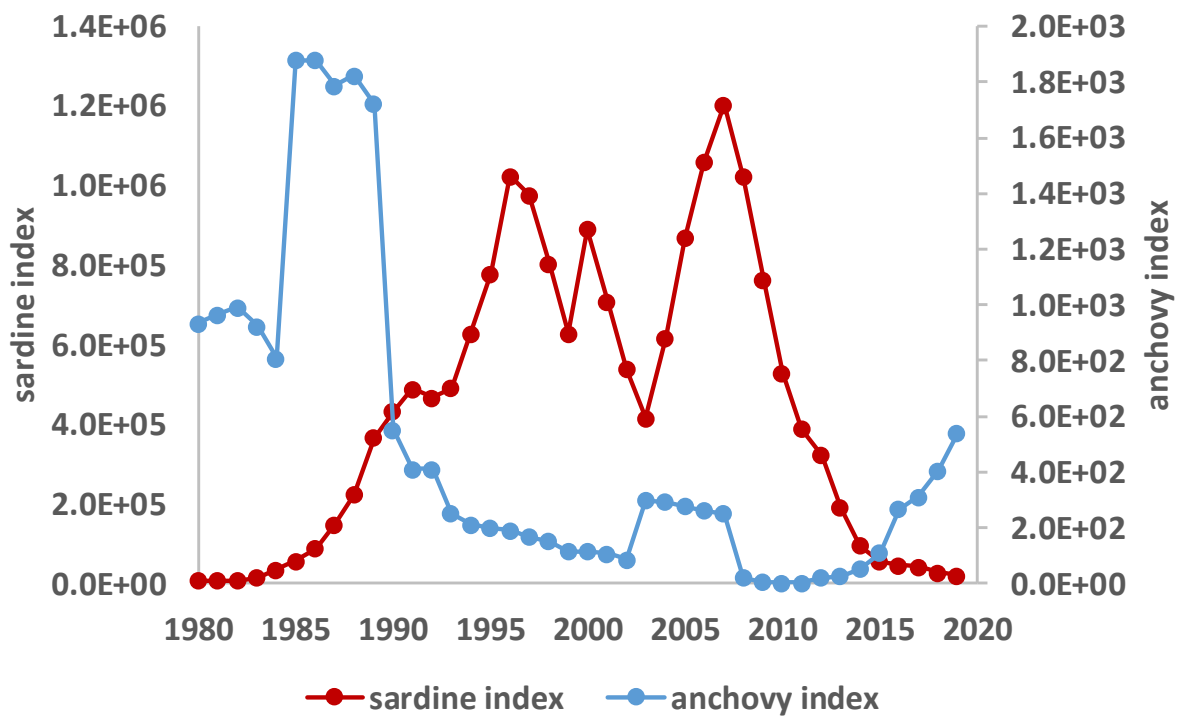**Supplementary Figure 1:** Sardine and anchovy biomass indices, calculated from sardine stock assessments, and anchovy larval abundances (see text for more details).
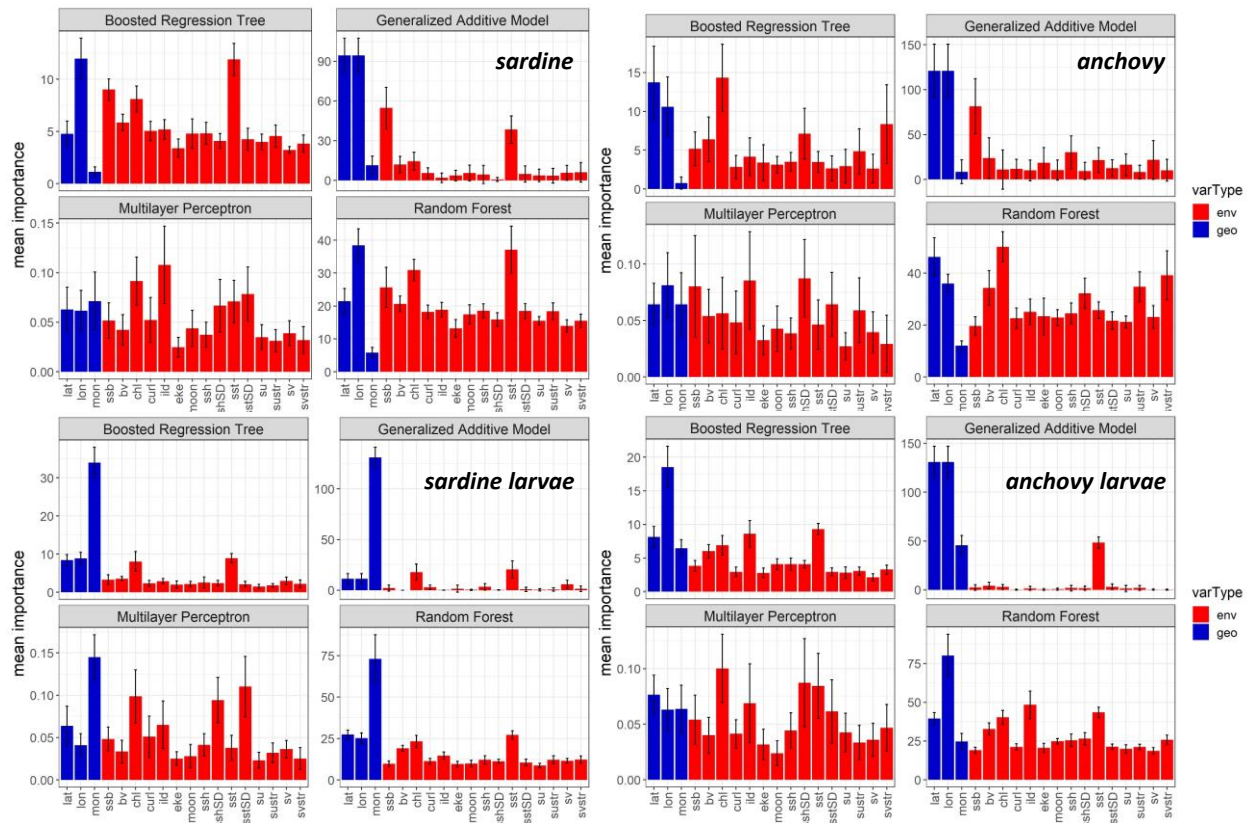
Supplementary Figure 2: Area Under the Receiver Operating Curve (AUC) skill metrics for Experiment 1 SDMs, broken out by season and focusing on months with the highest sampling coverage. Means and standard deviations across all SDM ensembles (see **Fig. 1**) are shown for each life stage of each species. Colors of bars denote the SDM configuration ("all", "env", or "geo"). The horizontal black dashed lines show AUC values of 0.5 (no better than a random model), and 0.75 (a rough approximation of a "useful" model), for reference.
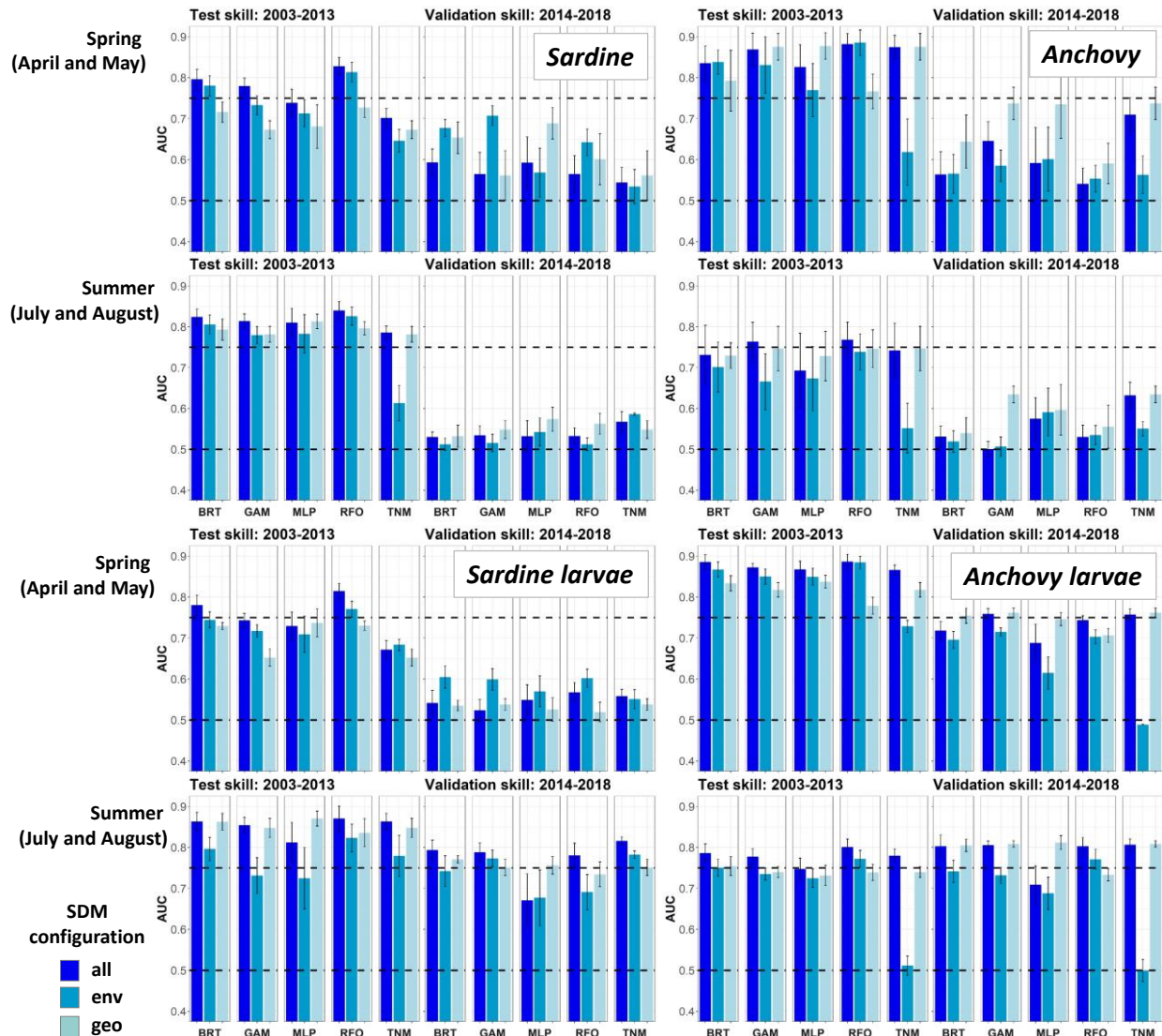
Supplementary Fig. 3: One-dimensional partial responses of each species and life stage to SST, integrated across all other predictors, for test years 2003 -2013, and validation years 2014 – 2018 (i.e. Experiment 1). Colored lines show SDM predictions, with each line representing a different SDM object (including the 10 repeated random splits of training and testing data: see Fig. 2). Black dots represent binned observations. Line colors distinguish different SDM types.
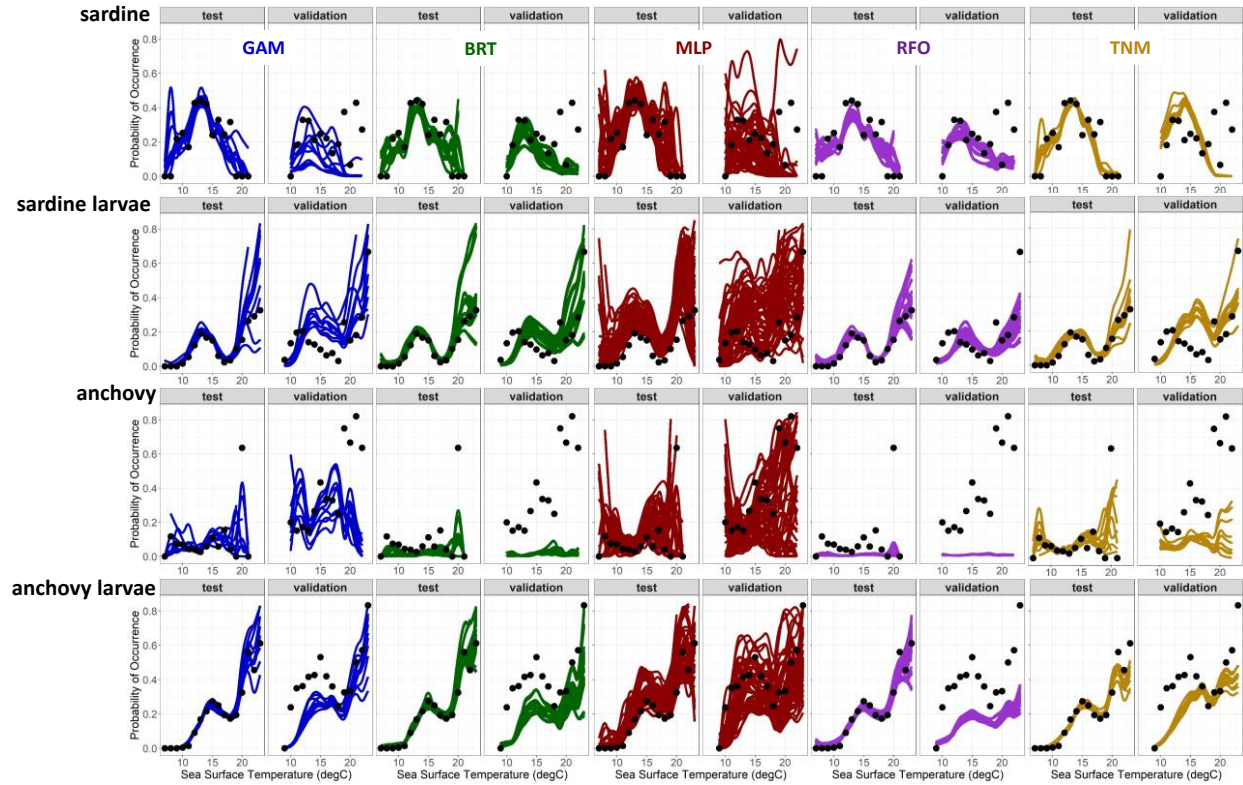
Supplementary Figure 4: Mean variable importance within each SDM, for each species and life stage. The three geospatial predictors (latitude, longitude, and month) are shown in blue, while environmental and biomass variables ("ssb") are shown in red. "bv" denotes bulk buoyancy frequency, "chl" is surface chlorophyll, "curl" is wind stress curl, "ild" is isothermal layer depth, "eke" is eddy kinetic energy, "moon" is moon phase, "ssh" is sea surface height, "sst" is sea surface temperature, "su" is surface eastward current velocity, "sustr" is surface eastward wind stress, "sv" and "svstr" are the equivalent variables for northward current and wind stress. "SD" denotes the spatial standard deviation. Boosted regression tree predictor importance was assessed based on the number of times each variable was selected for splitting, weighted by the squared improvement to the model as a result of each split, and averaged over all trees (Friedman & Meulman 2003; Elith et al. 2008). GAM variable importance was based on chi-square statistics. MLP variable importance was estimated using absolute value of Olden's connection weights algorithm (Olden & Jackson 2002). Random Forest variable importance estimated based on mean decrease in node impurity, from the Gini index

Supplementary Figure 5: Area Under the Receiver Operating Curve (AUC) skill metrics for Experiment 1 SDMs, broken out by season and focusing on months with the highest sampling coverage. Means and standard deviations across all SDM ensembles (see Fig. 1) are shown for each life stage of each species. Colors of bars denote the SDM configuration ("all", "env", or "geo"). The horizontal black dashed lines show AUC values of 0.5 (no better than a random model), and 0.75 (a rough approximation of a "useful" model), for reference. Only results from external validation years (2014 – 2018) are shown. Results where "dummy" negative stations were added before AUC calculation are shown on the left (see text for more details), with AUCs from original data (i.e. the same as Supplementary Fig. 1) are shown on the right.