Protected species interaction estimation

Chris Long, Rob Ahrens

Monitoring protected species bycatch is required under the Endangered Species Act, Magnuson-Stevens Fishery Conservation and Management Act, and Marine Mammal Protection Act. This is typically achieved through the placement of fisheries observers on vessels and a linear extrapolation of observed bycatch to the full fleet. However, these methods to do not take advantage of the considerable oceanographic data available and the assumptions inherent to linear methods may not be appropriate for rare events like protected species interactions. Our goal was to use ensemble random forests, a machine learning algorithm, to estimate protected species bycatch and compare these estimates to existing methods.

We trained 25 independent ensemble random forest (ERF) models using all shallow-set longline (SSLL) sets (n=18,988) from 2005-2020, which as a result of 100% observer coverage have known amounts of interaction with three protected species: oceanic whitetip sharks (*Carcharhinus longimanus*)(n=639 sets with interaction), loggerhead sea turtles (*Caretta caretta*)(n=190), and leatherbacks (*Dermochelys coriacea*)(n=102). We provided the algorithm with a suite of 26 oceanographic variables (e.g., SST, chlorophyll a, current velocity) gathered for the polygon defined by set and retrieval GPS coordinates, along with bycatch data for the three species of interest, for all SSLL sets from 2005-2020. The algorithm used these data to assess and learn which variables were most associated with sets that had bycatch in the training data. We then used the resulting models to predict protected species bycatch for all 2021 SSLL sets and adjusted these predictions by accounting for known rates of Type 1 and Type 2 error in out-of-bag predictions from the 2005-2020 data. Finally, we compared these ERF-derived estimates to those produced using a Horvitz-Thompson estimator of bycatch at multiple coverage levels.

Overall, the ERF algorithm was more effective at predicting sets with bycatch for oceanic whitetips than loggerheads and leatherbacks. This is likely a result of very low bycatch rates, and therefore less information regarding environmental covariates of bycatch, for the two turtle species. However, after accounting for Type 1 and Type 2 error, the oceanic whitetip estimates were very good; mean predictions of number of sets with bycatch for the 25 models was 28.2 sets (SD: 0.49) with bycatch, closely matching the actual 2021 value of 28 sets with little variation among model runs. The loggerhead (mean prediction = 8.18 sets, actual = 14) and leatherback (mean prediction = 4.59, actual = 3) estimates were biased low and high respectively. Using an estimate derived from 2005-2020 data of 1.37 whitetip sharks per set given that there was a whitetip interaction, the ERF estimates of total whitetip interactions for 2021 had a mean of 38.75 (SD: 0.67), close to the real value of 37 but slightly biased high. Mean ERF estimates were substantially more precise among model runs than those produced by Horvitz-Thompson estimators, while the accuracy and level of bias using ERF bycatch estimation depends largely on the method of accounting for Type 1 and Type 2 error as well as estimator accuracy for species per set with interaction when this number is greater than 1.