The following supplement accompanies the article

Modeling habitat and bycatch risk for dugongs in Sabah, Malaysia

Dana K. Briscoe[1,*], Seth Hiatt[2], Rebecca Lewison[3], Ellen Hines[2,4]

[1]Stanford University, Hopkins Marine Station, 120 Ocean View Blvd, Pacific Grove, CA 93950, USA
[2]San Francisco State University, 1600 Holloway Avenue, San Francisco, CA 94132, USA
[3]San Diego State University, 5500 Campanile Drive, San Diego, CA 92182, USA
[4]Marine & Coastal Conservation and Spatial Planning Lab, Romberg Tiburon Center for Environmental Studies, Tiburon, CA 94920, USA

*Corresponding author: dbriscoe@stanford.edu

Endangered Species Research: 24: 237–247 (2014)

Supplement. Diagnostic testing is a necessary component in model assessment and evaluation. As discussed in greater detail in Phillips et al. (2006), such output diagnostics help to validate the model's overall predictive performance. In MaxEnt, presences are split into 'training' and 'test' data. Training data are used to create the predictive model; test data are used to assess model accuracy. In our study, 75% of the sightings (presences) were used as training data. The remaining 25% were used as test data. However, instead of partitioning the data into singular training and testing datasets, we used a process that randomly splits the presence data into a number of equal sized groups called "folds". Each fold is omitted in turn, and is used as the test data, thus testing the performance of the training data within that model's iteration (Friedlaender et al. 2011). Like Friedlaender et al. (2011), we used the replication function to randomly sample occurrences from each training run, and used the remaining occurrences to test the model. This type of cross-validation technique, known as 'K-folds' within MaxEnt, addresses the effects of spatial autocorrelation. For our models, we chose to run 10 iterations, similar to Phillips et al. (2006), and Friedlaender et al. (2011). The mean of the 10 replicates was then computed for the model output.

MaxEnt's output diagnostics generate a statistical 'threshold' of prediction, above which predicted values are considered to be suitable habitat; below the threshold predicted values are within unsuitable habitat. Figures S1 and S2 show the predictive performance of the mean MaxEnt ouput. Replicated rates of omission of training and testing samples (through the K-folds cross validation technique) were compared to MaxEnt's randomly generated prediction of suitable habitat.

The mean omission rate of test data (Fig. S1) showed a suitable match between the predicted omission rate randomly generated by MaxEnt itsel. This indicated that the training and test data are independent and thus showed no spatial autocorrelation with the presence data.

Performance measures can also be assessed in the area under the ROC curve, or AUC. The AUC is considered to be the probability that a randomly chosen site of occurrence will be ranked above a randomly chosen site of absence (Elith et al. 2006). The probability of species presence ranges from 0 (minimum) to high 1 (maximum) (Svenning et al. 2008). Since MaxEnt does not use absence data, the AUC is calculated by using randomly chosen 'background' data points from the study area. Given this aspect, the AUC thus involves rankings between the randomly chosen presence site and the randomly chosen background site (Elith et al. 2006, Phillips et al. 2006). Each model replicate was run with all background points available in the study area. The probability of species presence ranges from 0 (minimum) to high 1 (maximum) (Svenning et al. 2008). The mean AUC generated from the MaxEnt analysis indicated a high level of performance for correctly choosing a site of occurrence within the available background data points.

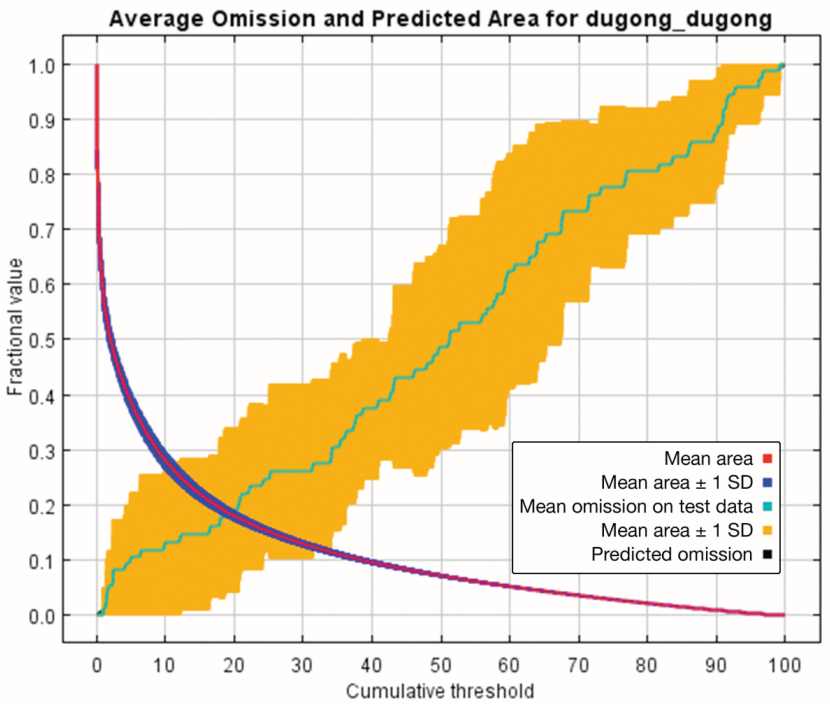

Fig. S1. Omission rates for mean output of MaxEnt model

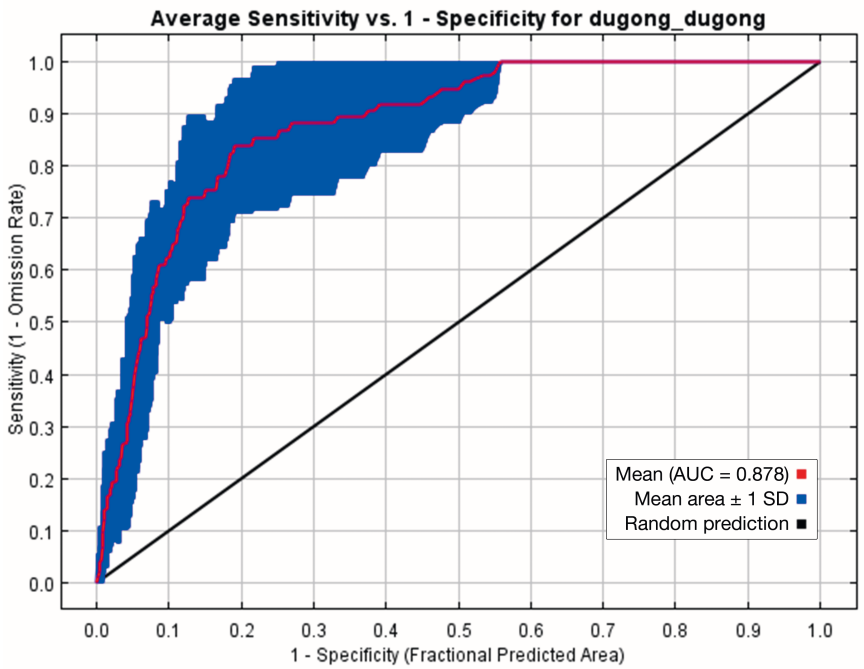

Fig. S2. Receiver Operator Characteristic (ROC) for MaxEnt model. AUC: area under curve

LITERATURE CITED

Elith J, Graham CH, Anderson RP, Dudík M and others (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography 29:129–151

Friedlaender AR, Johnston DW, Fraser WR, Burns J, Halpin PN, Costa DP (2011) Ecological niche modeling of sympatric krill predators around Marguerite Bay, Western Antarctic Peninsula. Deep-Sea Res II 58:1729–1740

Phillips S, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecol Model 190:231–259

Svenning JC, Normand S, Kageyama M (2008) Glacial refugia of temperate trees in Europe: insights from species distribution modeling. J Ecol 96:1117–1127