SUPPORTING INFORMATION. Huggler, K., J. D. Holbrook, M. M. Hayes, P. W. Burke, M. Zornes, D. J. Thompson, J. G. Clapp, P. Liongerger, M. Valdez, and K. L. Monteith. 2022. Risky business: How an herbivore balances spatiotemporal aspects of risk from competitors and predators. Ecological Applications.

APPENDIX S1

Random forest encounter and predation risk models

Methods

We used a random forest (RF) approach to modelling probability of encounter for elk, coyotes, and mountain lions, and predation risk by mountain lions. For each risk source, we used 80 percent of the sampled GPS data as the training dataset for the RF model, and withheld the remaining 20 percent as our test dataset. We tuned RF models with our training dataset using 10-fold cross-validation in the caret package (version 6.0-84) in R version 3.5.2. Due to imbalance of used versus available locations (1:5 for encounter models, 1:70 for predation risk), we used synthetic minority over-sampling technique (SMOTE) to balance the distribution of used and available locations within the cross-validation step (Chawla et al., 2002). SMOTE is an approach the over-samples the minority class (in our case, used locations) and under-samples the majority class (available locations) (Chawla et al., 2002). We used the cross-validation results to choose the hyperparameter that optimized both specificity (true negatives) and sensitivity (true positives) to train the final RF model for each species. In RF only one hyperparameter is available to be tuned, 'mtry'. The 'mtry' parameter represents the number of variables to randomly sample at each split of a tree. Consequently, we developed a 'tunegrid' that included a range of values of 'mtry' to test from 1 to the max number of predictor variables (16). The 'mtry' value that optimized sensitivity/specificity was 7 for elk and coyotes, 8 for mountain lions, 6 for the predation risk model. After the optimal 'mtry' value was selected, we ran the RF model with the full dataset using the respective 'mtry' parameter from the cross-validation results.

We evaluated the predictive ability of the RF model by using the final trained RF model to predict values from the withheld test dataset using a confusion matrix (Kuhn 2008). A confusion matrix is simply a cross-tabulation of the number of correctly and incorrectly classified predictions for each category (1 = used, 0 = available). The RF models for elk,

coyotes, mountain lions, and predation risk were able to differentiate used from available locations as indicated by the specificity (true negative rate > 0.85) and sensitivity (true positive rate > 0.85). Lastly, we created a final RF model with the entire dataset using the best hyperparameters identified from the tuning process. The final predictive model was used to predict probability of use by elk, coyotes, mountain lions, and risk of predation separately.

REFERENCES

Chawla, N. v., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/JAIR.953