Appendix E. Model evaluation methods

AUC and Gain are threshold independent measures of model performance calculated in MAXENT on 25 percent of the data withheld for testing in each of the 20 replicate runs. Gain is a measure calculated in MAXENT that is closely related to deviance, a measure of goodness of fit used in generalized additive and generalized linear models (Elith et al., 2011). Gain is defined as the average log probability of the presence samples, minus a constant that makes gain of a uniform distribution equal to zero. Gain of zero indicates no model discrimination between background points and occurrence data, whereas higher values of gain indicate better discrimination ability. Gain essentially indicates how closely the model is concentrated around the presence samples and can be used to approximate an odds ratio of how much more likely the presence samples are than that of random background pixels by calculating exp(gain) (Phillips, 2005).

Overall accuracy, sensitivity, specificity and Kappa (K) are all threshold-dependent measures of model fit based on a confusion matrix (Fielding and Bell, 1997). This required that we decide on a threshold for continuous model output to make it binary. It also required known absences, which we represented by generating pseudo-absences (random points within the area available for selection that were >5km from any occurrence locations). The number of pseudo-absences produced for each niche model was set equal to the number of presence points for that niche model. We selected a model threshold which maximized Kappa (Freeman and Moisen, 2008), balancing omission and commission error. Values from final ENMs for each activity-type were extracted to presence and pseudo-absence points, confusion matrices were generated based on correct and incorrect classification of the ENM outputs at these points, and threshold-dependent accuracy measures were calculated using formulas from Fielding and Bell (1997).

For calculating the continuous Boyce index we used a moving window increment of 0.25 for calculating $B_{cont(0.1)}$ for each replicate, and a moving window increment of 0.01 for plotting P/E curves. We chose a larger increment for calculating the continuous Boyce index because we found that using a moving window increment of less than 0.25 resulted in a large proportion of zero values due to low sample sizes. Both Boyce index calculations and P/E curves were truncated at a MAXENT logistic value of 0.85 due to the lack of sufficient test samples at higher logistic values.

References

Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E. & Yates, C.J. (2011) A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, **17**, 43-57.

Fielding, A.H. & Bell, J.F. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, **24**, 38-49.

Freeman, E.A. & Moisen, G.G. (2008) A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. *Ecological Modelling*, **217**, 28-58.

Phillips, S.J. (2005) A brief tutorial on Maxent. AT&T Research, Florham Park, New Jersey. Available from *http://www.cs.princeton.edu/~schapire/maxent/tutorial/tutorial.doc* [Accessed April 23, 2013].