**Appendix 3 (edited)**

**Research Question**

What is the best MaxEnt setup for our sparse presence-only observation points, considering environmental predictor resolution, pseudo-absence selection methods and environmental predictor pre-processing?

**Method**

We control the following three variables:

1. Environmental predictor resolution: pixel (1 km) or grid (5 km) level. The grid-level coarser environmental predictor data matches the filter size (5 km) when we filter out observation datapoints in triatomine data preparation stage.
2. Pseudo-absence selection method: nobuffer or buffer. In the nobuffer case, we randomly selected 10,000 pseudo-absence points in the study region for each species. In the buffer case, we randomly selected 10,000 pseudo-absence points in the study region outside the 0.5 degree buffers created around observation points for each species.
3. Environmental predictor pre-processing: on or off. In the on case, we eliminated highly correlated environmental predictors and used the PCA results with 8 principle components to carry out the model training and prediction. In the off case, we used all the raw environmental predictors.

There are two situations for each of the 3 variables, and we have in total 8 cases. Inspired by Konowalik K and Nosol A1, for each case, we first trained MaxEnt on 14 species separately and calculated 4 model evaluation metrics values: AUC, TSS, MAE and Bias. Second, we performed Spearman Correlation Analysis to find two most independent evaluation metric scores. Third, we ploted a 2D graph using the two most independent metric scores as x and y values for each case, and found the best performing case from the graph.

We chose to plot the absolute value of Spearman Correlation Coefficient for two reasons: 1) we are interested in whether there is a monotonic relationship between metrics but not linear relationship, so we chose Spearman over Pearson; 2) we only want to know if the metrics have correlation or not, and whether it is positive or negative correlation is not a determining factor, so we plotted the absolute value of Spearman Correlation Coefficient.

The MaxEnt model for each case and each species is run with 10-fold cross validation, and this generates mean and standard deviation for each evaluation metrics. We use the mean value of evaluation metrics to conduct the correlation analysis.

**Results**

From Figure 1, MAE and Bias have the lowest absolute Spearman correlation value, and thus we pick them as our pair.

Since the best metric pair consists of Bias and MAE, we would be looking for which case has the smallest MAE and the smallest absolute Bias values. From Figure2, pixel\_buffer\_off and grid\_buffer\_off have the best Bias and MAE combinations. Pixel\_buffer\_off means this case uses pixel-level (1 km) data, picks pseudo-absence points outside 0.5 degree buffers, and does not perform any variable preprocessing. Grid\_buffer\_off means this case uses grid-level (5 km) data, picks pseudo-absence points outside 0.5 degree buffers, and does not perform any variable preprocessing. We picked Pixel\_buffer\_off as the best method to proceed because it has higher resolution, which is beneficial for our study.

**Figure 1**

A chart with numbers and colors

AI-generated content may be incorrect.

**Figure 2**

**A graph with colored lines and numbers

AI-generated content may be incorrect.**

**References**

1. Konowalik K, Nosol A. Evaluation metrics and validation of presence-only species dis- tribution models based on distributional maps with varying coverage. Scientific Reports. 2021;11(1):1482.