

Supporting Information. Daniel Fink, Tom Auer, Alison Johnston, Viviana Ruiz-Gutierrez, Wesley M. Hochachka, and Steve Kelling. 2020. Modeling avian full annual cycle distribution and population trends with citizen science data. *Ecological Applications*.

Appendix S3: Base Model Boosted Regression Tree Parameters

Within each spatiotemporal block, we fit a two-step boosted regression tree model designed to deal with zero-inflation to predict the observed counts (abundance) of each species. The boosted regression trees for both steps of the zero-inflation model were fit with the gbm package.

The strategy used to select the base-model parameters was based on statistical considerations under the constraint of a fixed computational budget. By relying on the variance-reducing properties from averaging across the ensemble, we did not need to worry about overfitting individual base models and could avoid costly base model cross validation to select gbm parameters. This facilitated a strategy geared towards learning as much of the signal as possible with a limited number of gbm trees ($n_{trees} = 1000$) for each base model. Based on experimentation fitting base models across a set of regions, seasons, and species we set $bag\ fraction = 0.80$ and learning rate or shrinkage $= 0.05$. The $tree.depth$ parameter was set to 5 for the occurrence model and 10 for the abundance model, giving both models the ability to adapt to nonlinear and interacting predictor effects.