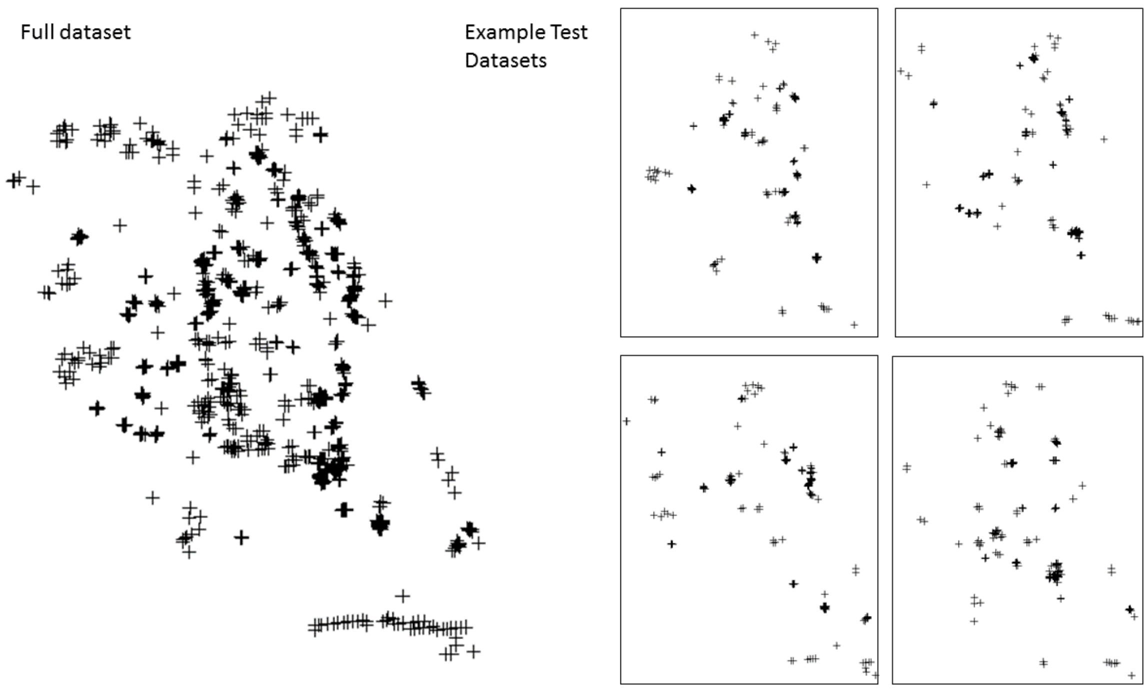
Appendix 1. Supplementary information on Boosted Regression Trees.

Spatially stratified sampling

We generated 20 training and test datasets for each modeled species in order to accurately assess model performance and to generate ensemble estimates of abundance robust to the data partitioning process (Barker et al. 2014). Training and test datasets were generated using a masked geographically structured (Radosavljevic and Anderson 2014) sub-setting routine that randomly assigns points to training and test datasets at a ratio of 5:1 based on a 5-km grid overlaid on the study area. This results in 20 unique sets of training and test data (Figure A1.1).

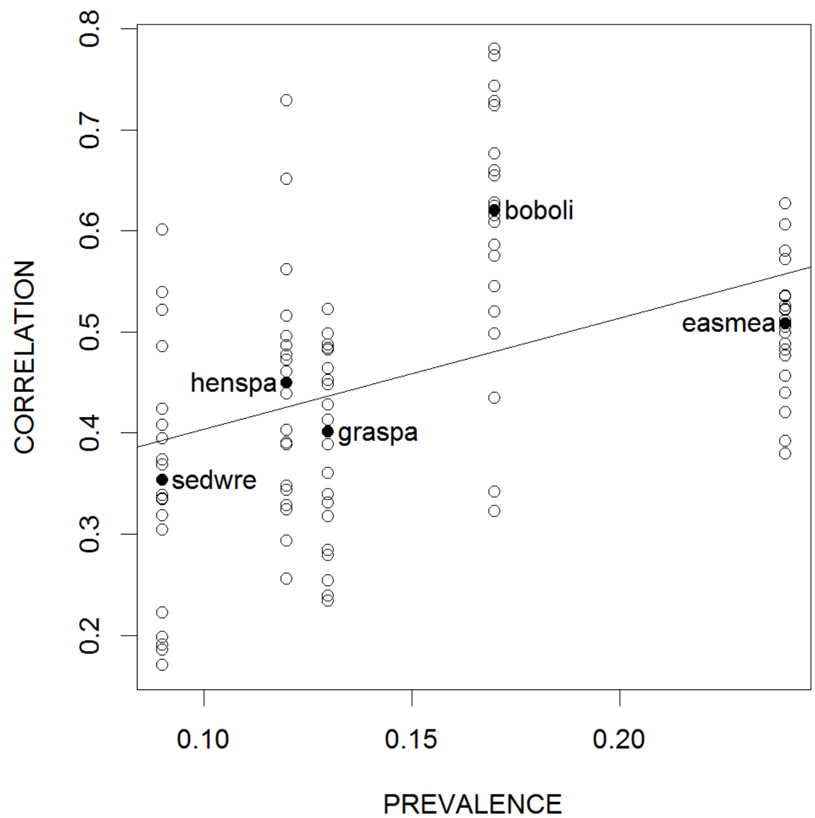
Figure A1.1. Sample training datasets generated by a spatial stratification algorithm used to generate datasets for model training and evaluation. Spatial stratification produces robust assessments of model performance. The algorithm assures that prevalence in the training and test datasets is similar to the full dataset.



Model performance

We characterized model performance by the median correlation across the 20 independent datasets. Correlation measures showed a significant positive relationship with the availability of non-zero counts, suggesting that more counts of rare birds would improve models of their relative abundance.

Figure A1.2. Model performance as measured by the correlation between predicted and observed abundance increased with increasing prevalence across five species of grassland birds: Sedge Wren (sedwre), Henslow's Sparrow (henspa), Grasshopper Sparrow (graspa), Bobolink (boboli), and Eastern Meadowlark (easmea). Open circles correspond to each of 20 models of relative abundance built for each species. Closed circles mark the median performance across those models. There is a significant positive relationship between performance and prevalence across these five species (F=20.11, p<0.0001).



Covariate relationships

Boosted regression tree models parameterize linear and non-linear responses between environmental covariates and observed relative abundance. Models were build using a backward stepwise procedure in which variables that did not contribute to improving model performance were removed. Figure A1.3 shows the frequency with which each environmental variable was included in models across species. Variable importance values within each model are estimated by random permutation of input variables for a given variable and estimating the impact on model accuracy. The top five environmental covariates for each species have the greatest impact on model accuracy and therefore are most closely related to relative abundance. Table A1.1 lists the top five predictor variables for each modeled species.

Variable response plots for boosted regression trees capture the shape of the average response to a given environmental predictor variable while other predictors are held at an intermediate value. Response plots (Figures A1.4a-e, see below) for each species are plot on a common scale. Values > 0 on the y axis suggest a positive response of abundance to a given predictor with negative values suggesting the opposite. Larger response values imply a greater effect size. Variables are listed in order of their permutation-based variable importance. Plots were generated using the gbm.plot function in package dismo of R (Hijmans et al. 2015).

Figure A1.3. Frequency at which environmental variables were included in models of relative avian abundance for five species of grassland birds.

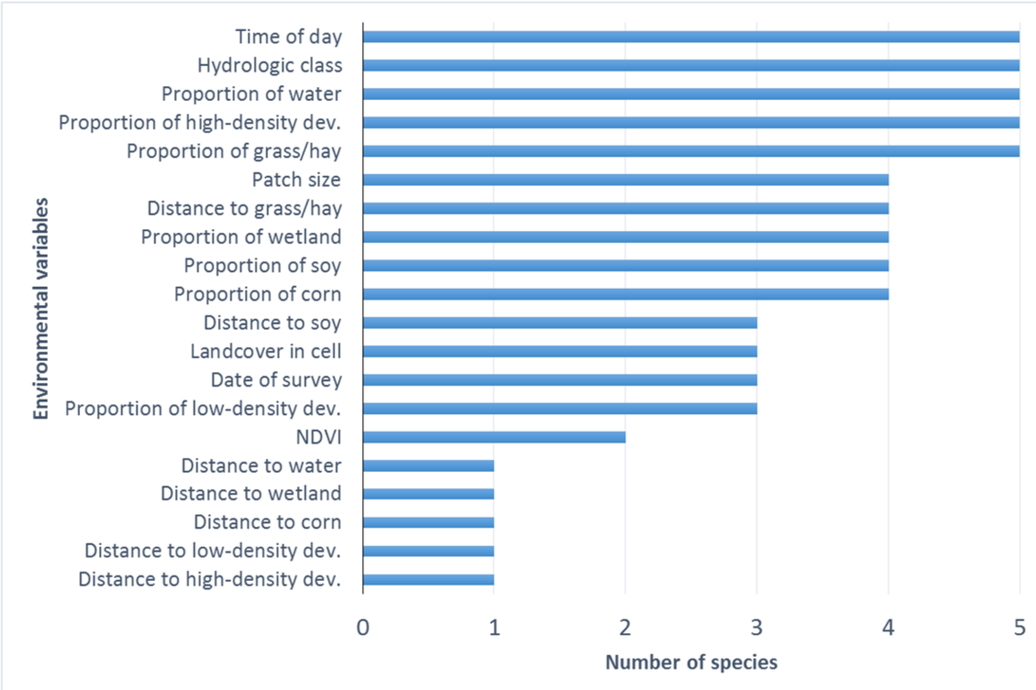


Table A1.1. Top five predictor variables for each species based on permutation-based variable importance.

Species	Predictor
Bobolink	Proportion of grass/hay
	Patch size
	Proportion of soy
	Proportion of water
	Proportion of high density development
Sedge Wren	Proportion of grass/hay
	Proportion of water
	Time of day
	Survey date
	Hydrological class
Henslow's Sparrow	Time of survey
	Proportion of grass/hay
	Proportion of soy
	Distance to grass/hay
	Hydrologic class
Eastern Meadowlark	Proportion of grass/hay
	Proportion of soy
	Distance to soy
	Proportion of low density development
	Landcover in cell
Grasshopper Sparrow	Proportion of grass/hay
	Proportion of soy
	Distance to soy
	Proportion of low density development
	Landcover in cell

References

Barker, N. K. S., S. G. Cumming, and M. Darveau. 2014. Models to predict the distribution and abundance of breeding ducks in Canada. Avian Conservation and Ecology 9. doi: 10.5751/ACE-00699-090207.

Robert J. Hijmans, Steven Phillips, John Leathwick and Jane Elith (2015). dismo: Species Distribution Modeling. R package version 1.1-1. https://CRAN.R-project.org/package=dismo

Radosavljevic, A., and R. P. Anderson. 2014. Making better Maxent models of species distributions: complexity, overfitting and evaluation. Journal of biogeography 41:629–643.

Figure A1.4a. Response curves for Bobolink

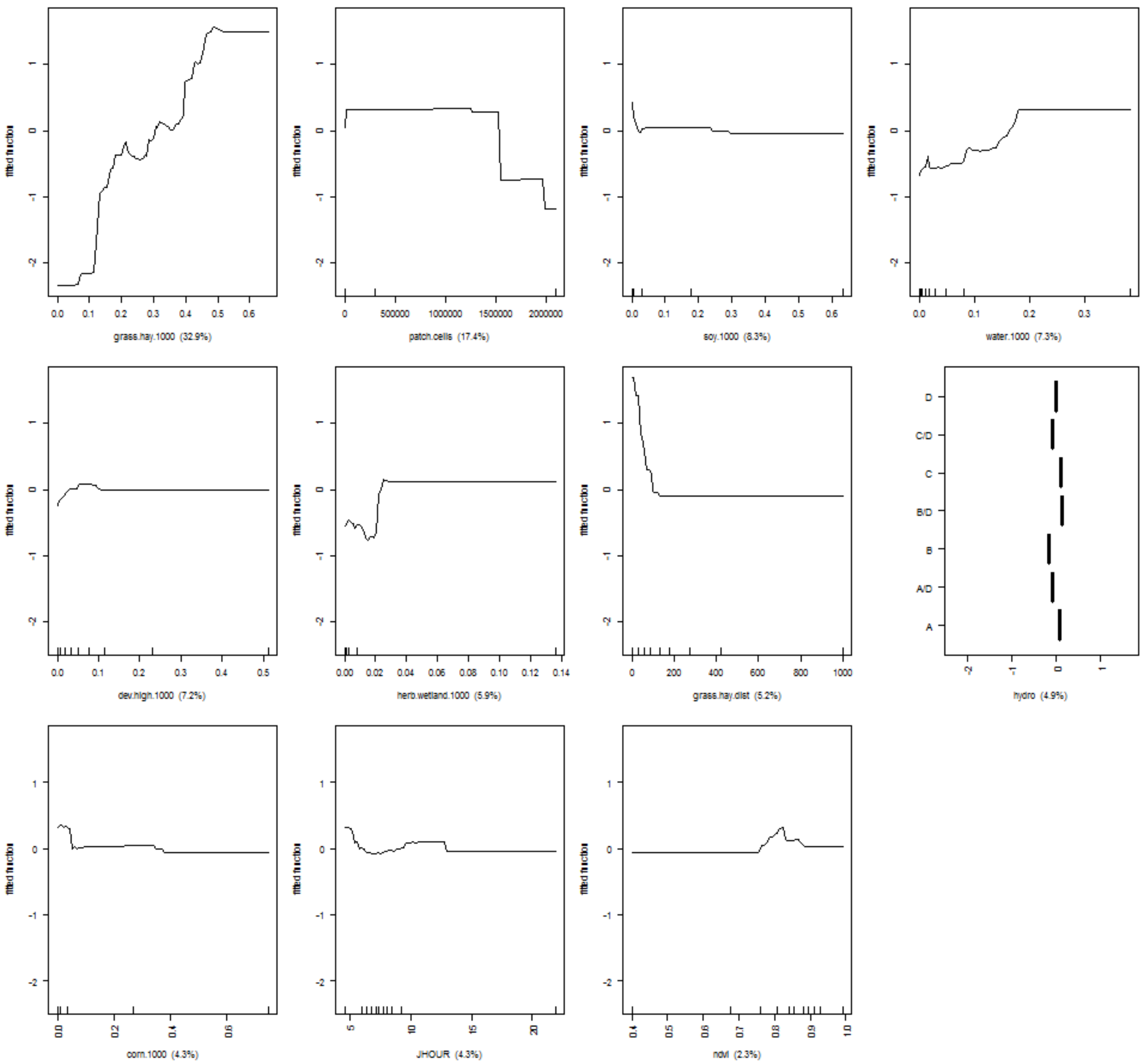


Figure A1.4b. Response curves for Eastern Meadowlark

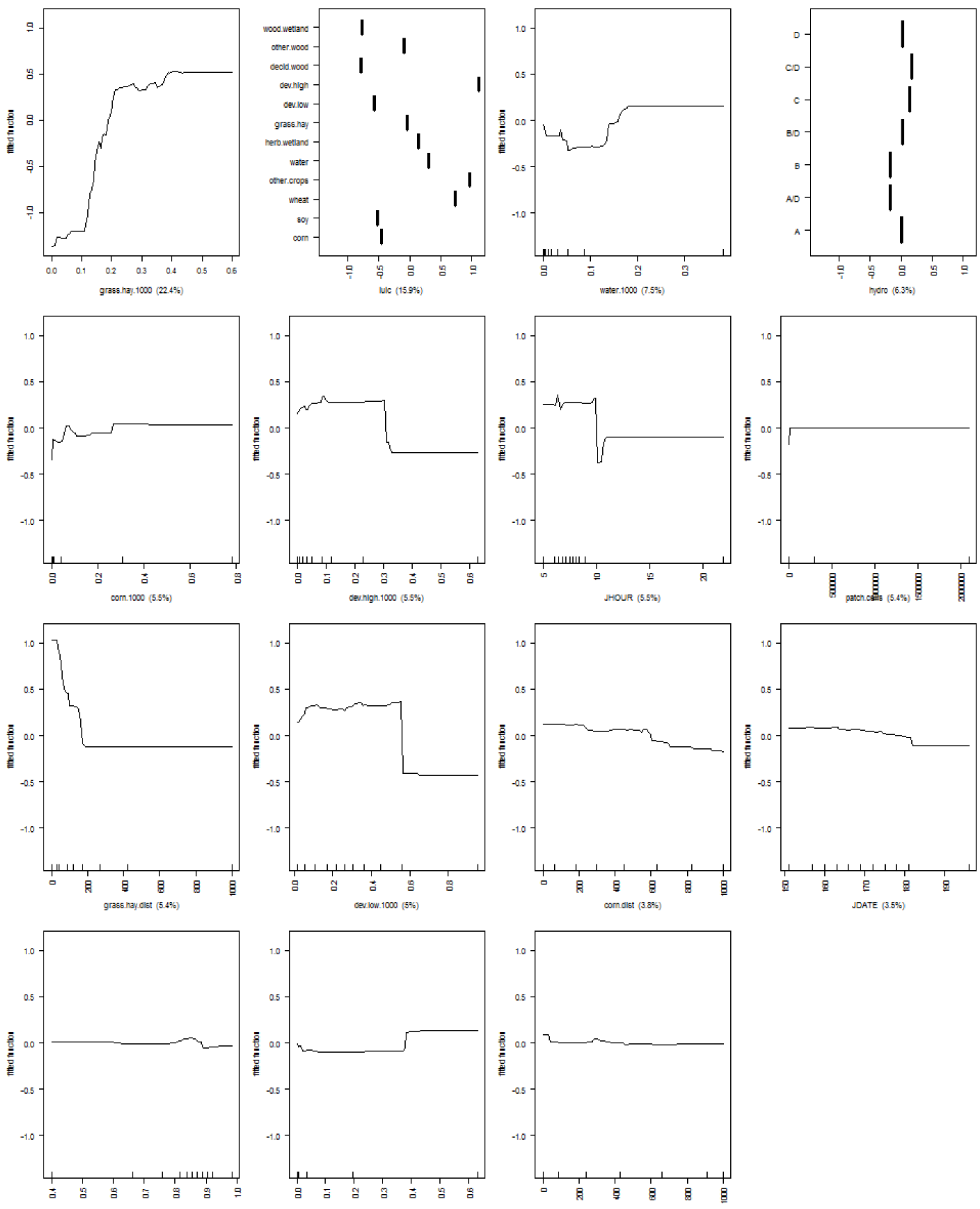


Figure A1.4c. Response curves for Grasshopper Sparrow

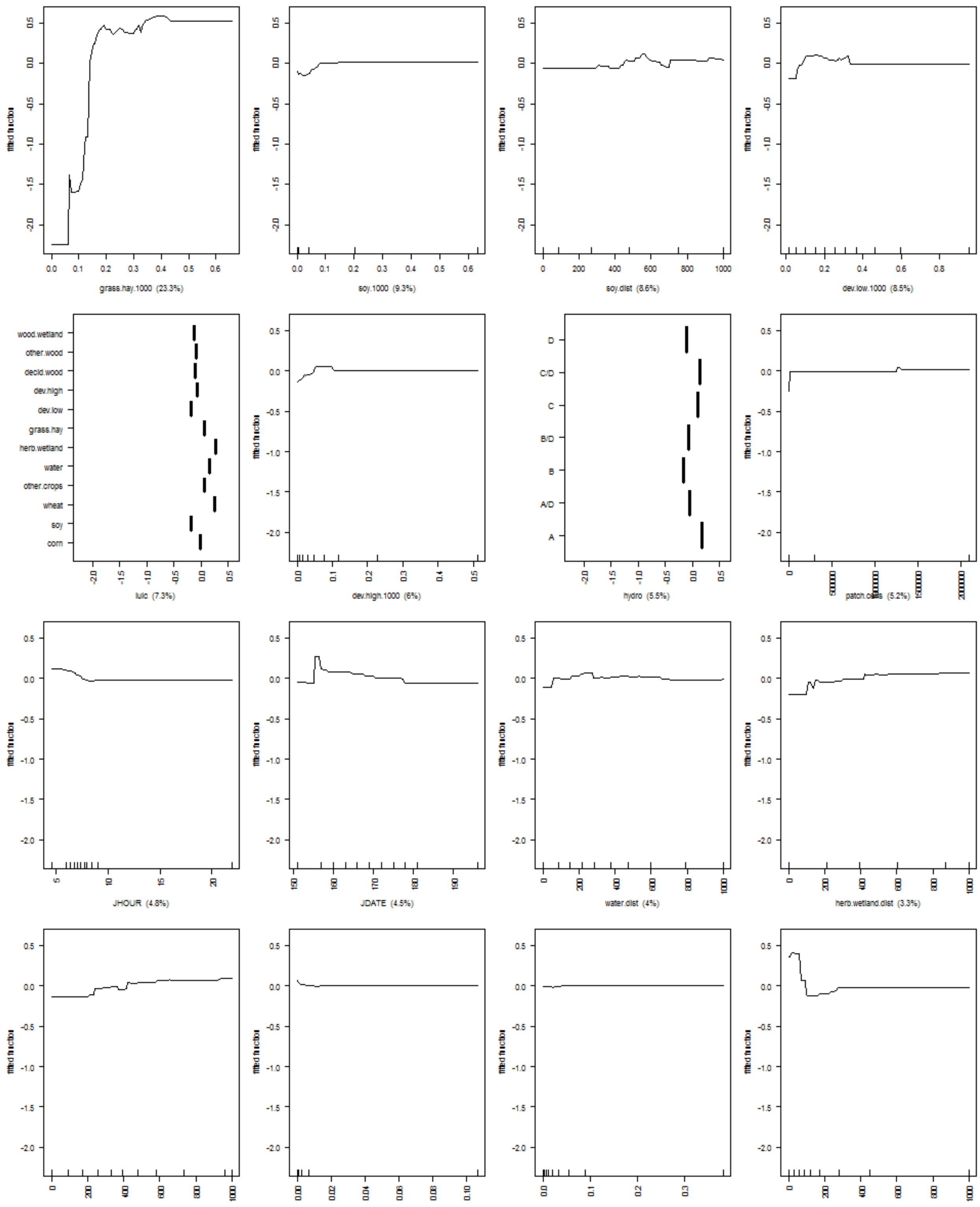


Figure A1.4d. Response curves for Henslow’s Sparrow

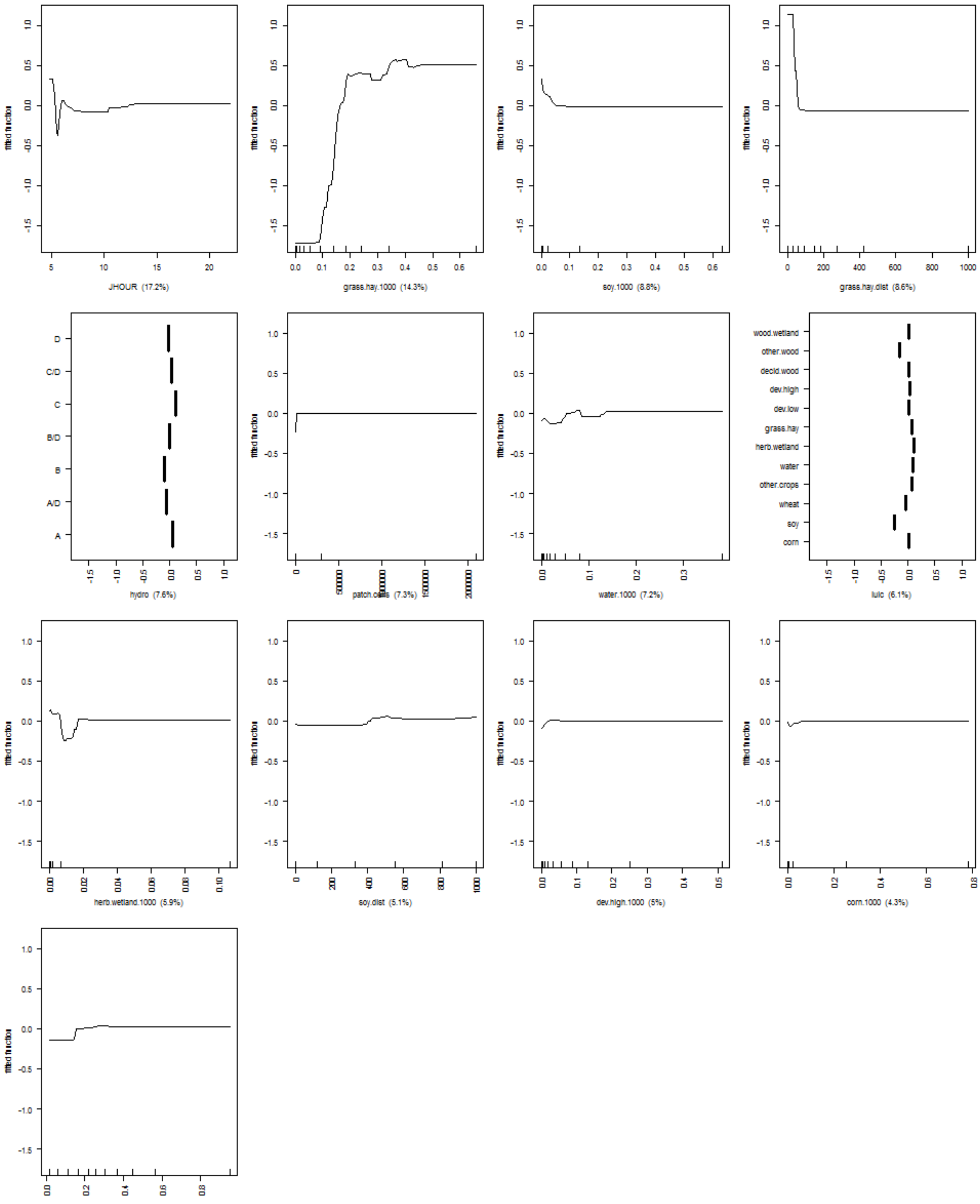


Figure A1.4e. Response curves for Sedge Wren