

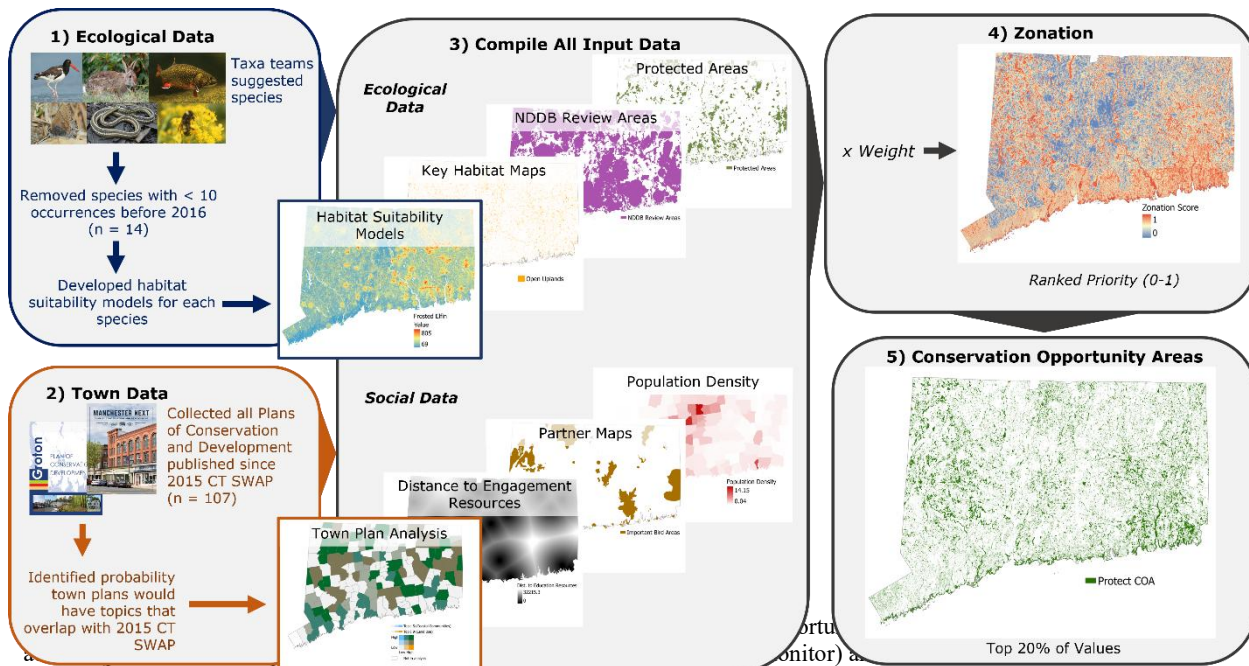
Appendix 4.6

Conservation Opportunity Area Mapping Methods

Prepared by: Katy Bischoff and Chadwick Rittenhouse, University of Connecticut

METHODS

Conservation opportunity areas (COAs) represent areas where partners can take voluntary actions to benefit wildlife populations and habitats and achieve conservation goals and objectives. The first step of mapping COAs was gathering the ecological, town, and social input data for the spatial prioritization process. The ecological data consisted of key habitat maps, Natural Diversity Database (NDDB) Review Areas, protected areas and habitat suitability maps for 14 species of greatest conservation need (SGCN) identified by taxa teams, that were representative of their taxa and/or key rare habitats and had at least 10 occurrences in the state since 2016. The town data identified towns that could be potential partners using topic models to gather common topics, a group of words representing an underlying theme, in town Plans of Conservation and Development that overlap with topics in the 2015 CT SWAP. The social data included maps of priority areas for other partners, such as the Important Bird Areas from the National Audubon Society, human population density, and proximity to engagement resources, such as libraries and nature centers. The spatial prioritization process involved including all ecological, town, and social data into the Zonation software, which iteratively ranks all grid cells based on their marginal loss in conservation value and gives each cell a score based on that ranking



(Moilanen et al. 2022). Zonation prioritizes areas with a high density of features, balance between features, and that minimize conservation loss. To reflect broad actions proposed by taxa teams, the PCRM-PI framework by Liberati et al. (2016) was followed and cross-walked to the Conservation Measures Partnership. COAs were created for each broad action: Protect, Connect, Restore, Manage, Partner and Inform (broken down further into Inform and Research and Monitor). Additionally, a combined COA was created with all 7 action COAs to identify where COAs overlap across the state (Figure 1).

Habitat Suitability Maps for Species of Greatest Conservation Need

The first step of mapping conservation opportunity areas (COAs) was developing the habitat suitability maps for SGCN in CT (Figure 1, Step 1). An initial email survey was sent to the six taxa teams (fish, plants, mammals, birds, herps, and invertebrates) explaining the purpose of the mapping process and asking for their recommendations of 2-3 species of greatest conservation need representative of key habitats and/or their taxonomic groups. Species were included within the model if their associated data had at least 10 occurrences, the current model minimum (Breiner et al. 2018), was collected since 2016 (to align with the temporal resolution of predictor data), and the species was considered a SGCN for the 2025 CT SWAP revision. The initial survey was followed up with a webinar that further explained the COA mapping process, why the suggestions were important, and examples of similar work done for other states. The habitat associations and data availability were used to narrow the initial list to 14 species eligible for modeling (Figure 2).

Biomod2 (Thuiller et al. 2025) and ecospat (Di Cola et al. 2017) packages in R 4.2.3 (R Core Team 2023) were used to predict potential suitable SGCN habitat across CT using the Storrs High Performing Computer Cluster. An ensemble of small models' approach was used with a two-step ensembling process also called double ensembling. This approach creates a series of bivariate models, based on all possible predictor combinations, and averages all the bivariate models to produce an ensemble model for each model tested. Models were ensembled into a final ensemble model using weights based on the AUC value (Lomba et al. 2010, Breiner et al. 2015, 2018, Di Cola et al. 2017). The advantage of using a series of small models is that the approach performs well with rare species data and can reduce overfitting by avoiding the violation of common standard of 10:1 number of observations to predictors (Breiner et al. 2015, 2018, Di Cola et al. 2017). Models were also tuned using the ecospat package (Di Cola et al. 2017). Tuning is an initial process of determining optimal parameters for each model by varying parameters for each model to determine which parameter settings maximize model performance and it also allows the parameters to align with the data used within the models (Breiner et al. 2018).

Nine models offered within the biomod2 package were tested for each SGCN. The models used within the package follow the two different approaches of viewing a statistical model; statistic-based where the data model is determined first and then fit into the model, and machine learning where the data determines the model (Breiman 2001a). The statistic-based models were Generalized Linear Model (GLM), Generalized Additive

Model (GAM), and Multivariate adaptive regression splines (MARS). The machine learning based models included Gradient Boosted Machine (GBM), Random Forest (RF), XGBoost, Artificial Neural Network (ANN), Classification Tree Analysis (CTA), and Flexible Discriminant Analysis (FDA). GLMs provide an advantage in species distribution models because the model can handle different distributions, such as Gaussian, Poisson, binomial, and negative binomial, thus different forms of data can be used within the species distribution model (McCullagh and Nedler 1989). GAMs are similar to GLMs but use non-parametric smoothing function for predictors which allows for a more complex relationship and do not require the shape of the distribution to be known (Hastie and Tibshirani 1986, Guisan et al. 2017). MARS are a piecewise linear regression where each data point is assessed for the ideal linear model, and if a different linear relationship reduces the error, then a knot or hinge point is formed (Friedman 1991). RF builds a series of independent decision-based trees using random subsamples drawn from the data that are then ensembled to create one model (Breiman 2001b). GBMs are also an ensemble machine learning approach where trees are created sequentially, instead of independently like RF, based on reducing the error in the previous tree while gradually increasing the focus on poorly modeled observations (Elith et al. 2008, Guisan et al. 2017). XGBoost also uses gradient boosting, similar to GBM, but includes a regularization parameter and parallel processing to increase computing power (Chen and Guestrin 2016). ANN uses a machine learning model to create a set of adaptive weights or interconnections between neurons or hidden layers to link responses to predictors (Ripley 1996, Guisan et al. 2017). CTA splits predictor variables into classes based on a decision tree (Breiman et al. 1984). FDA is also a classification method, based on a mixture of regression models (Hastie et al. 1994). Maximum Entropy (MAXENT) and Maximum Entropy fit with glmnet models were excluded from the analysis because they require a prevalence value, the calculation of which requires information on absences and is often not available for rare species. Additionally, the default prevalence value of 0.50 is only suitable for species with similar prevalence values otherwise the model error increases (Guillera-Arroita et al. 2014). Surface Rangle Envelop models were also excluded since it consistently performed the lowest in other species distribution models (Breiner et al. 2018, Guo et al. 2021). All 9 models were initially tested with each SGCN, however, if a model consistently failed, it was removed from the list of potential models. To save memory within the outputs, the probability of species occurrence was rescaled to vary from roughly 0-1000 (Thuiller et al. 2025).

To evaluate the performance of models, a combination of area under the receiver operating curve (AUC; Hanley and McNeil 1982), true skill statistic (TSS; Allouche et al. 2006), Kappa value (Cohen 1960), and the Boyce Index (Boyce et al. 2002) were used. Only the models that were better at predicting than random i.e., $AUC > 0.5$, were retained for the final ensemble model (Breiner et al. 2015). Independent model evaluation data was not able to be included within the ecospat package, so the data was split into training and testing data based on a 90:10 split (Breiner et al. 2015). Well performing models were identified with at least 3 performance metrics meeting the following standards: AUC values ≥ 0.80 (Hosmer et al. 2013), TSS values ≥ 0.60 (Komac et al. 2016), Kappa ≥ 0.6 (McHugh 2012), and Boyce ≥ 0.5 . The Boyce metric was designed for models with presence-only

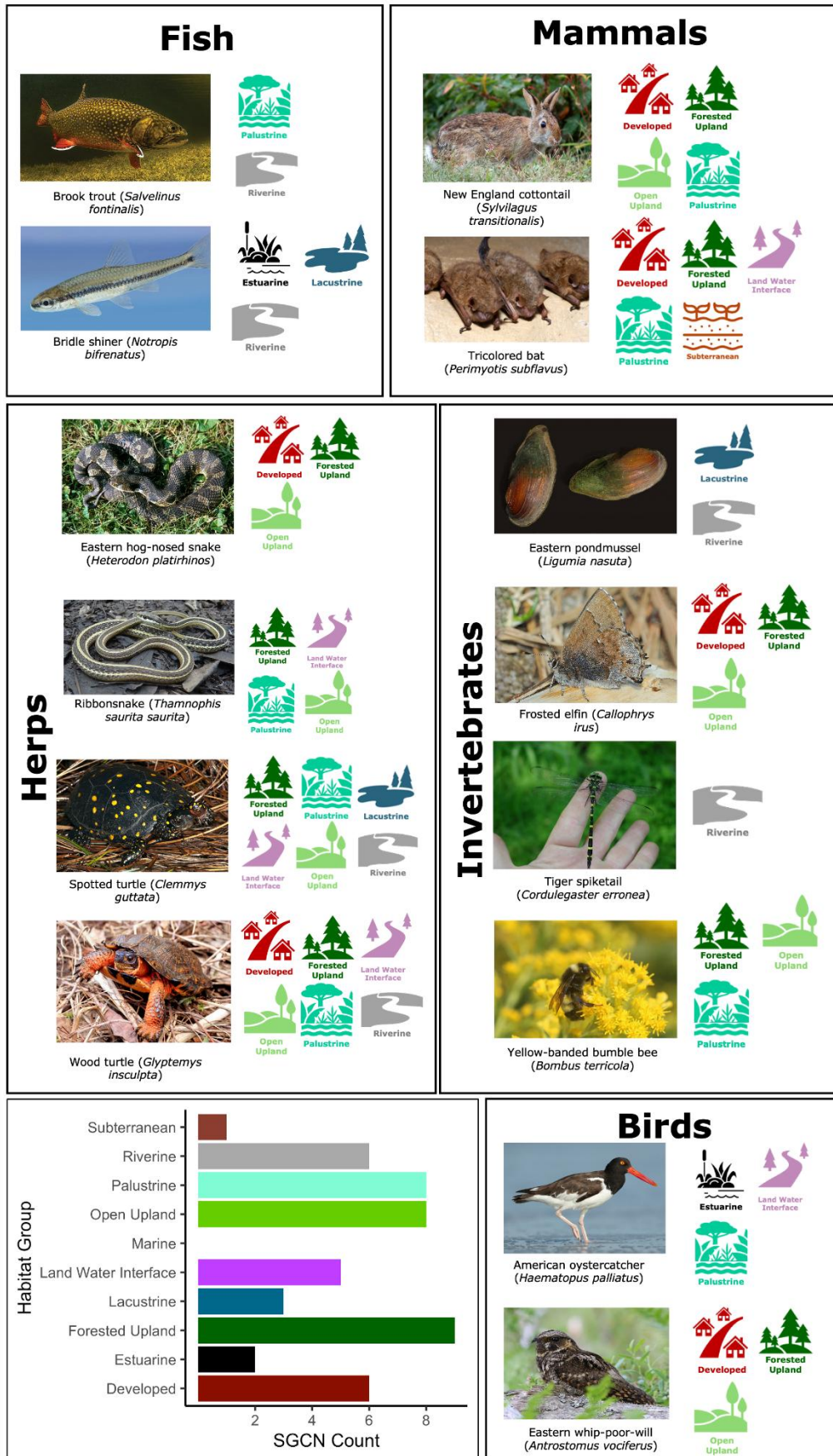
data (Boyce et al. 2002); thus, the Boyce index was not used for models with presence-absence data. The final model selected was the model with the highest performance across all metrics.

For species with both presence and absence location information, both presence and absence data were used within the models, since models with low amounts of presence-absence data have higher predictive performance than models with low amounts of presence-only data (Hirzel et al. 2006). However, species distribution models with low amounts of presence-only data and high numbers of pseudo-absences can still have high predictive performance (Barbet-Massin et al. 2012, Breiner et al. 2015, 2018). For species with only presence data, a large set (Barbet-Massin et al. 2012) of 5,000 random pseudo-absence locations was generated. Species data was randomly thinned by 1.0 km (average dispersal for most SGCN in the mapping process) to reduce spatial autocorrelation using the spThin package (Aiello-Lammens et al. 2015). Thinning some datasets caused the observations to be below the minimum standard for the model; thus, data that had < 15 species occurrences was not thinned. Locations within the same 10.0 m grid cell were removed from all models.

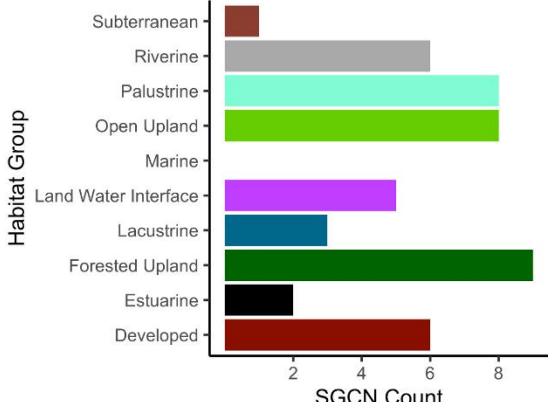
The CT state boundary was used as the extent for all SGCN models due to the availability of fine-scale spatial data that differentiated vegetation types and spatial patterns (Rittenhouse et al. 2022, Yang et al. 2023) and the availability of species occurrence data, recognizing that CT is a portion of each of the species' ranges. Previous studies using habitat suitability models have also limited the spatial extent to a portion of a species' range when modeling rare species or when specific areas are of interest (Warren et al. 2008, Maslo et al. 2016, Liroy et al. 2023, Santamarina et al. 2023).

Predictors were included in each model based on the habitat associations identified by the taxa teams and existing literature. Predictors ranged from large-scale processes at coarser resolutions, such as elevation and climate, to species specific fine-scale habitat characteristics, such as host plants, understory vegetation, young forest and shrubland vegetation, barrens, and soil characteristics. For most landcover and habitat characteristic predictors, Euclidean distance surfaces were created to measure proximity to these habitat features for each species. For species that use habitats specific to terrestrial, aquatic, or coastal areas, predictors were clipped to these extents to avoid identifying nonhabitat as habitat (Elith et al. 2011). To ensure minimal correlation between predictors, predictors from the model with a high pair-wise Pearson correlation coefficient $|r| \geq 0.60$ were excluded from the model, since not all models are robust to correlated predictors (Dormann et al. 2013). A stricter threshold than the standard of 0.70 (Dormann et al. 2013) was used due to the reduced spatial extent of the models. Variable contributions were measured for each predictor and model for each species. Variable contributions were the ratio between the sum of weights of the model with all predictors compared to the sum of weights of models without the selected predictors, thus, a ratio ≥ 1.0 indicates the selected covariate has a higher contribution than or equal contribution to average (Broennimann et al. 2025).

2025 Connecticut Wildlife Action Plan



Habitat Group



Habitat Group	SGCN Count
Subterranean	1
Riverine	6
Palustrine	8
Open Upland	8
Marine	0
Land Water Interface	5
Lacustrine	3
Forested Upland	9
Estuarine	2
Developed	6

Figure 2. The 14 SGCN included in the habitat suitability and conservation opportunity area mapping processes. These SGCN are representative of either key habitats and/or their taxa. The bar plot at the bottom demonstrates the count of SGCN in each broad habitat group, demonstrating the 14 SGCN represent all habitat groups excluding marine.

Birds

The SGCN modeled within the bird taxa group included 2 species: American oystercatcher (*Haematopus palliatus*) and eastern whip-poor-will (*Antrostomus vociferus*). American oystercatcher was the only modeled species found solely in coastal habitats, thus the model extent was the land and coastal waters identified as a coastal area by C.G.S. 22a-94 (CT General Statutes 2005). Restricting the extent allowed the models to distinguish fine scale differences along the coast more accurately than if the entire state was used as the extent. The eastern whip-poor-will is a terrestrial only species, thus all waterbodies identified in the US Geological Survey (USGS) Hydrography Dataset (USGS 2005) were removed from the predictors and the extent of the model.

American oystercatcher (Haematopus palliatus)

Species Data

The American oystercatcher presence locations were retrieved from the NDDB records from 2016-2017. This included 13 presence locations; thus, the location data was not spatially thinned.

Predictors

The final list of predictors included distance to impervious surfaces, estuaries, shellfish beds, and the presence of mixed forest. The 1.0 m resolution Coastal Change Analysis Program land cover map was used to extract mixed forest, impervious surfaces (impervious surfaces and impervious development were combined), and estuaries (emergent and scrub/shrub estuarine wetlands were combined; National Oceanic and Atmospheric Administration (NOAA) 2020). Because of the high prevalence of forest within the state, the mixed forest predictor was the presence of mixed forest rather than the distance to mixed forest, and the mixed forest predictor was resampled to a 10.0 m resolution. The shellfish bed location data from CT Department of Agriculture was used to determine the distance to shellfish beds (CT DEEP and CT Department of Agriculture 2019). American oystercatchers use a diversity of habitats and are an ideal umbrella species for coastal habitats (Maslo et al. 2016). Shellfish beds and estuaries were included in the model to reflect the availability of foraging opportunities (Gliesch et al. 2023) and impervious surfaces and mixed forest were included in the model to reflect proximity to threats and nonhabitat (US Fish and Wildlife Service (USFWS) 2024a). Other predictors evaluated in the model included distance to barrens, buildings, and grasslands, however these predictors reduced model performance and were highly correlated, so they were not included in the final model.

Model Evaluation and Selection

The highest performing model was the MARS model, with an AUC value of 1.00, Boyce value of 1.00, TSS value of 0.99, and Kappa value of 0.5 (Table 1), meeting all 4 criteria for evaluation metrics. The highest contributing predictors (values ≥ 1.00) were distance to shellfish beds (1.09) and estuaries (1.16), while distance to impervious development and the presence of mixed forest had low contributions (values < 1.00 ; Table 2). American oystercatcher suitable habitat decreased with increasing distance to estuaries and shellfish beds and the presence of mixed forest (Figure 3).

Tables and Figures

Table 1. Evaluation metrics output for each model tested for American oystercatcher, Artificial Neural Network (ANN), Flexible Discriminant Analysis (FDA), Gradient Boosted Machine (GBM), Multivariate adaptive regression splines (MARS), XGBoost, and the Ensembled approach (EF). The grey highlighted row is the final model chosen for the mapping process.

Models	AUC	Boyce	TSS	Kappa
ANN	0.99	0.68	0.99	0.22
FDA	1.00	1.00	0.99	0.33
GBM	0.99	0.22	0.99	0.40
MARS	1.00	1.00	0.99	0.50
XGBOOST	1.00	0.82	0.99	0.36
EF	1.00	0.98	0.99	0.44

Table 2. Variable contribution values for the predictors within the American oystercatcher models: Artificial Neural Network (ANN), Flexible Discriminant Analysis (FDA), Gradient Boosted Machine (GBM), Multivariate adaptive regression splines (MARS), XGBoost, and the Ensembled approach (EF). Contribution values ≥ 1.00 indicate the contribution value for that predictor is higher than or equal to the average contribution value. The grey highlighted row is the final model chosen for the mapping process.

Predictors	ANN	FDA	GBM	MARS	XGBOOST	EF
Impervious surface distance	0.83	0.94	0.91	0.88	0.94	0.90
Estuary distance	1.03	1.09	0.94	1.16	1.08	1.06
Shellfish bed distance	1.10	1.07	1.09	1.09	1.04	1.08
Mixed forest	1.06	0.92	1.08	0.89	0.95	0.98

2025 Connecticut Wildlife Action Plan

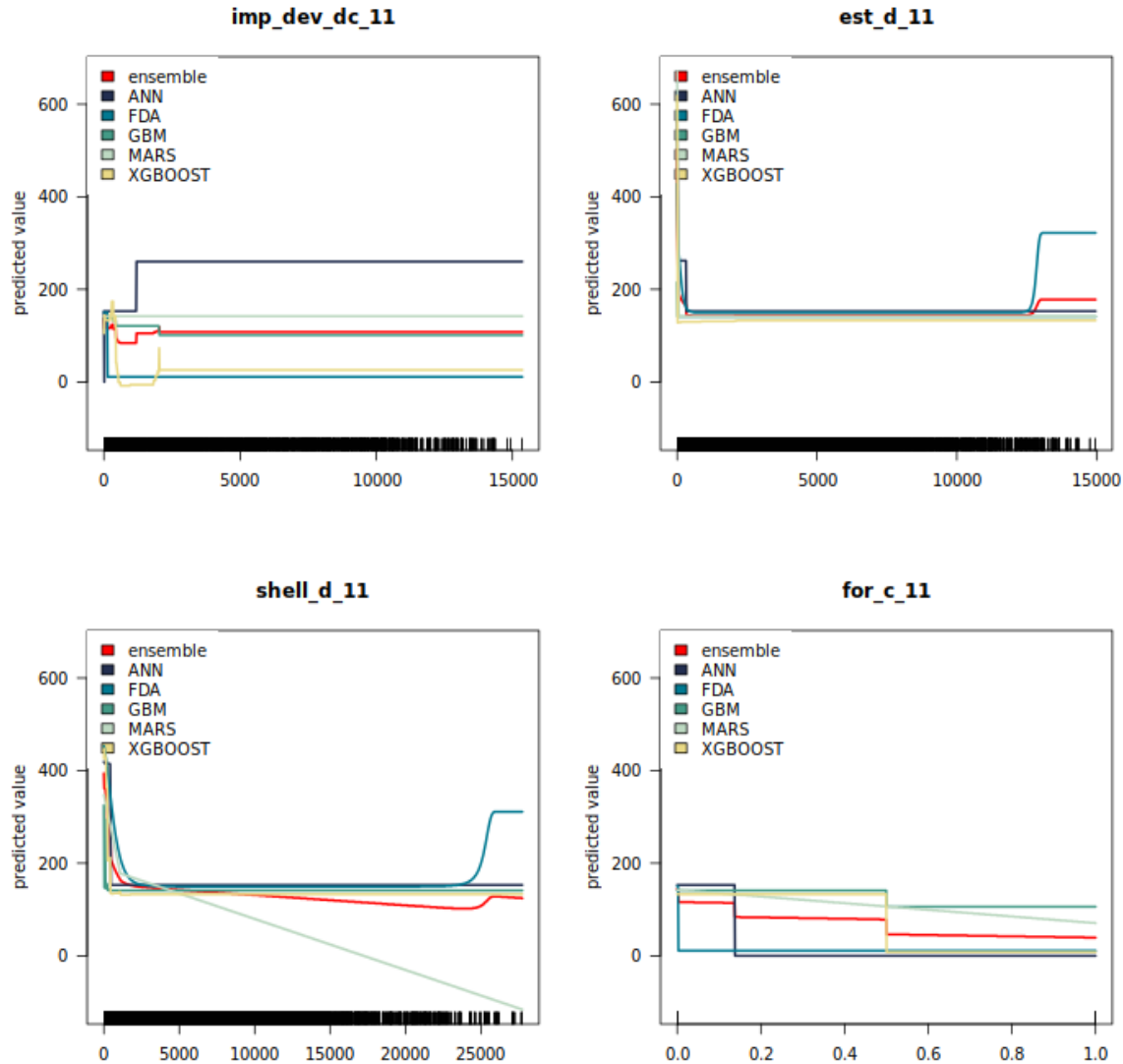


Figure 3. Response curves for the predictors in the American oystercatcher models: distance to impervious surfaces (imp_dev_dc_11), distance to estuaries (est_d_11), distance to shellfish beds (shell_d_11), and presence of mixed forest (for_c_11). All distance predictors are measured in meters. Each model evaluated represents a different colored line: Artificial Neural Network (ANN), Flexible Discriminant Analysis (FDA), Gradient Boosted Machine (GBM), Multivariate adaptive regression splines (MARS), XGBoost, and the Ensembled approach (EF).

Eastern whip-poor-will (Antrostomus vociferus)

Species Data

Using NDDB records, 12 eastern whip-poor-will presence locations were retrieved that were collected from 2017-2023. The data was not thinned due to the small number of presence locations.

Predictors

The final list of predictors included distance to buildings, young forest and shrubland, palustrine wetlands, and understory, elevation and presence of mixed forest. The building footprints layer was used as a proxy for development (Microsoft 2019). The Young Forest and Shrubland Vegetation Map was used to identify different types of young forest and shrubland vegetation within Connecticut (Rittenhouse et al. 2022). The map used ecological processes (succession, disturbance, regeneration, and hydrology), vegetation height, percent vegetation cover by height category, previous land cover type, and time since disturbance to classify different types of young forest and shrubland vegetation (Rittenhouse et al. 2022). Elevation was calculated from 30.0 m elevation tiles (National Aeronautics and Space Administration (NASA) 2000). The elevation predictor was resampled to 10.0 m to align with the spatial resolution of the models. The 1.0 m resolution Coastal Change Analysis Program land cover map to extract mixed forest and palustrine wetlands (combined emergent, scrub/shrub, and forested palustrine wetlands; NOAA 2020). Because of the high prevalence of forest within the state, the mixed forest predictor was the presence of mixed forest rather than the distance to mixed forest, and the mixed forest predictor was resampled to a 10.0 m resolution. The understory vegetation map was used to identify locations of understory species, greenbrier (*Smilax rotundifolia*), mountain laurel (*Kalmia latifolia*), Japanese barberry (*Berberis thunbergii*) and other mixed invasive understory species within deciduous forests of Connecticut (Yang et al. 2023). Young forest and shrubland, wetlands, and mixed forest predictors were included due to the positive influence of these predictors on eastern whip-poor-will occupancy and elevation, buildings, and the presence of understory was included due to the negative influence on occupancy in previous studies (Slover and Katzner 2016, Spiller and King 2021, Larkin et al. 2024). Distance to waterbodies and impervious surfaces were also evaluated as predictors, but these predictors reduced model performance, thus they were excluded from the final model.

Model Evaluation and Selection

The model with the highest performance was the ensemble model (ANN, GAM, GBM, MARS, RF), with an AUC of 1.00, Boyce of 0.51, TSS of 0.99, and Kappa of 0.67 (Table 3), meeting all 4 evaluation metric thresholds. Distance to young forest and shrublands and buildings and elevation had high contributions to the model with values of 2.32, 1.13, 1.00. The distance to palustrine wetland and understory vegetation, and presence of mixed forest all had low contribution, with values < 1.0 (Table 4). Eastern whip-poor-will habitat suitability was positively influenced by increasing distance to buildings and negatively influenced by increasing distance to young forest and shrubland (Figure 4).

Tables and Figures

Table 3. Evaluation metrics output for each model tested for eastern whip-poor-will, Artificial Neural Network (ANN), Generalized Additive Model (GAM), Generalized Boosting Model (GBM), Multivariate Adaptive Regression Splines (MARS), Random Forest (RF), and the Ensembled approach (EF). The grey highlighted row is the final model chosen for the mapping process.

Models	AUC	Boyce	MaxTSS	MaxKappa
ANN	1.00	1.00	0.99	0.44
GAM	0.98	0.79	0.95	0.40
GBM	0.99	0.72	0.97	0.40
MARS	0.99	0.60	0.98	0.50
RF	0.97	0.40	0.50	0.50
EF	1.00	0.51	0.99	0.67

Table 4. Variable contribution values for the predictors within eastern whip-poor-will models, Artificial Neural Network (ANN), Generalized Additive Model (GAM), Generalized Boosting Model (GBM), Multivariate Adaptive Regression Splines (MARS), Random Forest (RF), and the Ensembled approach (EF). Contribution values ≥ 1.00 indicate the contribution value for that predictor is higher than or equal to the average contribution value. The grey highlighted row is the final model chosen for the mapping process.

Predictors	ANN	GAM	GBM	MARS	RF	EF
Building distance	0.84	1.19	0.60	1.08	1.96	1.13
Young forest and shrubland distance	1.19	1.34	1.64	1.36	6.14	2.32
Mixed forest	1.30	1.17	1.47	1.02	0.00	0.99
Palustrine wetland distance	1.59	0.68	0.70	0.91	0.00	0.78
Elevation	0.79	0.76	0.59	0.86	2.00	1.00
Understory vegetation distance	0.53	0.98	1.37	0.83	0.67	0.88

2025 Connecticut Wildlife Action Plan

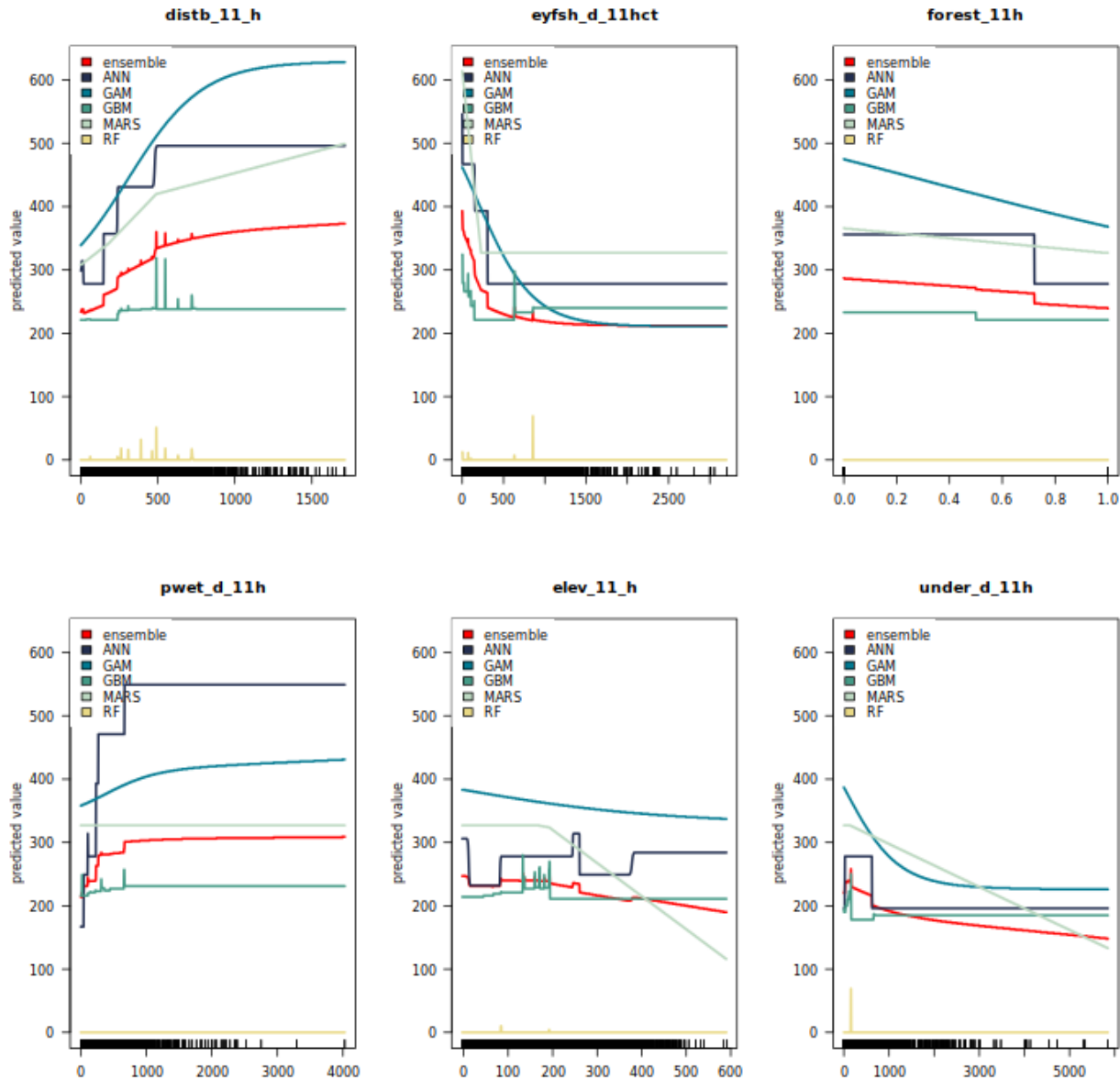


Figure 4. Response curves for the predictors in the eastern whip-poor-will models: distance to buildings (distb_11_h), distance to young forest and shrubland (eyfsh_d_11hct), presence of mixed forest (forest_11h), distance to palustrine wetlands (pwet_d_11h), elevation (elev_11_h), and distance to understory vegetation (under_d_11h). All distance and topographic predictors are measured in meters. Each model evaluated represents a different colored line: Artificial Neural Network (ANN), Generalized Additive Model (GAM), Generalized Boosting Model (GBM), Multivariate Adaptive Regression Splines (MARS), Random Forest (RF), and the Ensembled approach (EF).

Fish

The SGCN modeled within the fish taxa group included 2 species: bridle shiner (*Notropis bifrenatus*) and brook trout (*Salvelinus fontinalis*). The following SGCN models were aquatic only models, thus all terrestrial features were removed from the predictors and subsequent models using the National Hydrography Dataset Plus (Buto and Anderson 2020). Distance predictors for all aquatic predictors used Euclidean distance rather than a true surface distance following water due to lack of continuity between all above ground waterbodies, acknowledging this is a relative measure of proximity and not the physical distance an individual would travel to reach the habitat feature. The extent of brook trout was limited to rivers only, since brook trout survive primarily in rivers in CT (Jacobs 2009), using the USGS hydrography dataset to remove waterbodies within the National Hydrography Dataset Plus that intersect lakes, ponds, and reservoirs.

Bridle shiner (Notropis bifrenatus)

Species Data

Presence-absence data for bridle shiner was collected from 2018-2023 count data through long-term monitoring surveys conducted by CT DEEP. Some observation coordinates were located outside of the National Hydrography Dataset Plus (the dataset used to establish the extent of the model), thus any coordinates that were not within 10.0 m of the National Hydrography Dataset Plus grid cells were removed. Any locations that fell outside of the National Hydrography Dataset Plus rivers grid cell but were within 10.0 m received the center coordinate of the nearest National Hydrography Dataset Plus rivers grid cell. Initially the dataset contained 39 presences and 1,207 absences, but after removing coordinates outside of the extent, and thinning locations by 1 km, the dataset included 17 presence locations and 518 absence locations.

Predictors

Distance to impervious surfaces, palustrine aquatic beds, and dams, slope, annual total precipitation, and mixed forest proportion in local basins were included as predictors for the final models. Slope was calculated from 30.0 m elevation tiles (NOAA 2000). The slope predictor was resampled to 10.0 m to align with the spatial resolution of the models. The 1.0 m resolution Coastal Change Analysis Program land cover map was used to extract impervious surfaces and palustrine aquatic beds (NOAA 2020). Dam location data was retrieved from CT DEEP (DeLand 2025) to create a distance to nearest dam predictor. The 1.0 km resolution annual total precipitation (mm) data (Thornton et al. 2020) was used to average annual total precipitation across 2016-2019 and then resampled to 10.0 m. The 1.0 m resolution Coastal Change Analysis Program land cover map was also used to extract mixed forests (NOAA 2020) but the proportion of mixed forest within each hydrologic unit (HUC) 12 drainage basin (USGS 2021) was calculated instead of presence of mixed forest. A catchment-based predictor was used for mixed forest because the amount of forest upstream can influence water and habitat quality downstream and sub catchment predictors can improve model performance for small-scale freshwater species distribution models (Kuemmerlen et al. 2014). Distance to impervious surfaces was

included since this negatively influences the distribution of bridle shiner in Connecticut (Pregler et al. 2019). Palustrine aquatic bed, slope, annual precipitation, and dams were also included as predictors because macrophyte cover, velocity, and isolation were influential covariates for bridle shiner occupancy, depending on the season (Jensen 2012). Maximum temperature was also evaluated as a predictor within the model, but this reduced model performance, so it was excluded from the final models.

Model Evaluation and Selection

The highest performing bridle shiner model was the GLM model, with AUC value of 0.96, TSS value of 0.89, and Kappa value of 0.52 (Table 5), meeting all 3 of the evaluation criteria for presence-absence models. The predictors with the highest contribution to the model were slope and distance to impervious development, with values of 1.13 and 1.10, while distance to dams and palustrine aquatic beds, proportion of forest in local basin, and annual precipitation all had contribution values below 1.0 but above 0.90 (Table 6). Bridle shiner suitable habitat was positively associated with increasing distance to impervious surface and annual precipitation and negatively associated with increasing distance to palustrine aquatic bed, and dams, and slope (Figure 5)

Tables and Figures

Table 5. Evaluation metrics output for each bridle shiner model, Generalized Additive Model (GAM), Generalized Linear Model (GLM), Random Forest (RF), XGBoost, and the Ensembled approach (EF). The grey highlighted row is the final model chosen for the mapping process.

Model	AUC	TSS	Kappa
GAM	0.93	0.85	0.43
GLM	0.96	0.89	0.52
RF	0.95	0.85	0.52
XGBOOST	0.90	0.83	0.37
EF	0.95	0.90	0.41

Table 6. Variable contribution values for the predictors within the bridle shiner models: Generalized Additive Model (GAM), Generalized Linear Model (GLM), Random Forest (RF), XGBoost, and the Ensembled approach (EF). Contribution values ≥ 1.00 indicate the contribution value for that predictor is higher than or equal to the average contribution value. The grey highlighted row is the final model chosen for the mapping process.

Predictor	GAM	GLM	RF	XGBOOST	EF
Impervious surface distance	1.07	1.10	1.19	1.38	1.18
Dam distance	0.96	0.93	0.61	0.54	0.76
Slope	1.21	1.13	1.25	1.46	1.26
Forest proportion in local basin	0.91	0.97	1.10	0.89	0.97
Annual precipitation total	0.87	0.97	1.11	1.16	1.03

2025 Connecticut Wildlife Action Plan

Palustrine aquatic bed distance	1.01	0.91	0.84	0.79	0.89
------------------------------------	------	------	------	------	------

2025 Connecticut Wildlife Action Plan

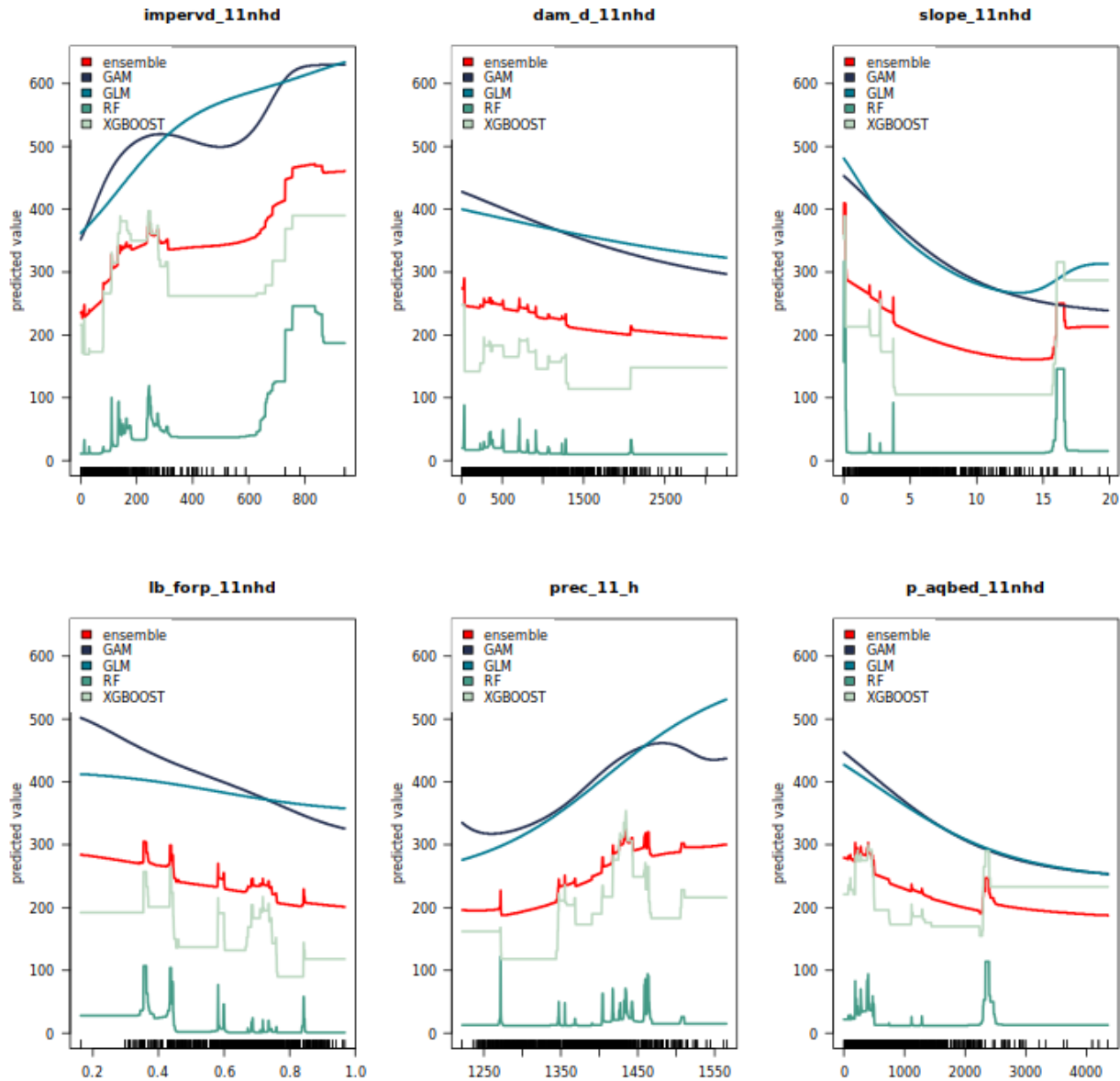


Figure 5. Response curves for the predictors in the bridge shiner models: distance to impervious surfaces (impervd_11nhd), distance to nearest dam (dam_d_11nhd), slope (slope_11nhd), forest proportion in local basin (lb_forp_11nhd), annual precipitation total (prec_11_h), and distance to palustrine aquatic bed (p_aqbed_11nhd). All distance and topographic predictors are measured in meters and annual precipitation is measured in millimeters. Each model evaluated represents a different colored line: Generalized Additive Model (GAM), Generalized Linear Model (GLM), Random Forest (RF), XGBoost, and the Ensembled approach (EF).

Brook trout (Salvelinus fontinalis)

Species Data

The presence-absence data from long term monitoring survey by CT DEEP in 2018-2019 and 2022-2023 was retrieved for this analysis. Only coordinates located within 10.0 m of the National Hydrography Dataset Plus rivers were kept to ensure the predictor extent and observation locations aligned. Any locations that fell outside of the National Hydrography Dataset Plus rivers grid cell but were within 10.0 m received the center coordinate of the nearest National Hydrography Dataset Plus rivers grid cell. Initially, the dataset included 196 presences and 115 absences. After removing coordinates outside National Hydrography Dataset Plus rivers, there were 77 presences and 57 absences. After thinning the data randomly by 1.0 km, the data in the final model consisted of 64 presences and 49 absences.

Predictors

Distance to impervious surface and cold-water sites, mixed forest proportion in local basins, average monthly winter precipitation, and slope were the predictors in the final model. Slope was calculated from 30.0 m elevation tiles (NASA 2000) and the predictor was resampled to 10.0 m to align with the spatial resolution of the models. The 1.0 m resolution Coastal Change Analysis Program land cover map was used to extract impervious surfaces (NOAA 2020). The 1.0 km resolution monthly total precipitation (mm) data (Thornton et al. 2022) was used to extract average winter (January, February, and December) precipitation across 2016-2022 and then resampled to 10.0 m. The 1.0 m resolution Coastal Change Analysis Program land cover map was also used to extract mixed forests (NOAA 2020) but the proportion of mixed forest within each hydrologic unit (HUC) 12 drainage basin (USGS 2021) was calculated instead of presence of mixed forest to capture the effects of habitat quality upstream and previous studies have found sub catchment predictors can improve model performance for small-scale freshwater species distribution models (Kuemmerlen et al. 2014). The location of cold-water sites was collected by CT DEEP from their Monitoring and Assessment and Inland Fisheries Programs (CT DEEP 2020a). The proportion of forest in the local basin, slope, and impervious surfaces were included because percent forest cover positively influenced brook trout occupancy and slope both positively and negatively influenced occupancy at the catchment-level (Kelly et al. 2021) and percentage impervious surface decreased brook trout occupancy at the landscape scale (Wagner et al. 2013). Precipitation was extracted only for winter months (January, February, and December), since high and low winter precipitation can influence brook trout recruitment the following year (Maitland and Latzka 2022). Distance to cold-water sites was included due to increased abundance of trout in cold water refuges (Sullivan and Vokoun 2023). Distance to dams, whether the location was in a large river (Housatonic, Connecticut, Naugatuck or Thames River), average monthly maximum temperature and maximum monthly temperature over a year, were also evaluated as predictors but these predictors reduced model performance.

Model Evaluation and Selection

The highest performing model was the GBM model, with an AUC value of 0.95, TSS value of 0.91, and Kappa value of 0.90 (Table 7), meeting the 3 evaluation criteria for the

presence-absence models. The predictors with the highest contributions (values ≥ 1.00) were proportion of forest within the local basin (1.44) and distance to the nearest cold-water site (1.86), slope, winter precipitation and distance to impervious surface had low contributions (values < 1.00 ; Table 8). Brook trout habitat suitability increased with higher forest proportion and decreased with increasing distance to cold-water sites and increasing average winter precipitation (Figure 6).

Tables and Figures

Table 7. Evaluation metrics output for each brook trout model, Generalized Additive Model (GAM), Gradient Boosted Machine (GBM), Generalized Linear Model (GLM), Random Forest (RF), XGBoost, and the Ensembled approach (EF). The grey highlighted row is the final model chosen for the mapping process.

Model	AUC	TSS	Kappa
GAM	0.88	0.80	0.81
GBM	0.95	0.91	0.90
GLM	0.92	0.72	0.71
RF	0.95	0.82	0.81
XGBOOST	0.85	0.72	0.71
EF	0.95	0.81	0.81

Table 8. Variable contribution values for the predictors within the brook trout models: Generalized Additive Model (GAM), Gradient Boosted Machine (GBM), Generalized Linear Model (GLM), Random Forest (RF), XGBoost, and the Ensembled approach (EF). Contribution values ≥ 1.00 indicate the contribution value for that predictor is higher than or equal to the average contribution value. The grey highlighted row is the final model chosen for the mapping process.

Predictors	GAM	GBM	GLM	RF	XGBOOST	EF
Impervious surface distance	0.62	0.92	0.93	1.01	1.25	0.95
Forest proportion in local basin	0.99	1.44	1.00	1.32	1.17	1.19
Cold-water site distance	2.86	1.86	1.39	1.92	1.54	1.91
Slope	0.49	0.41	0.91	0.43	0.55	0.56
Winter precipitation	1.04	0.84	0.84	0.78	0.74	0.85

2025 Connecticut Wildlife Action Plan

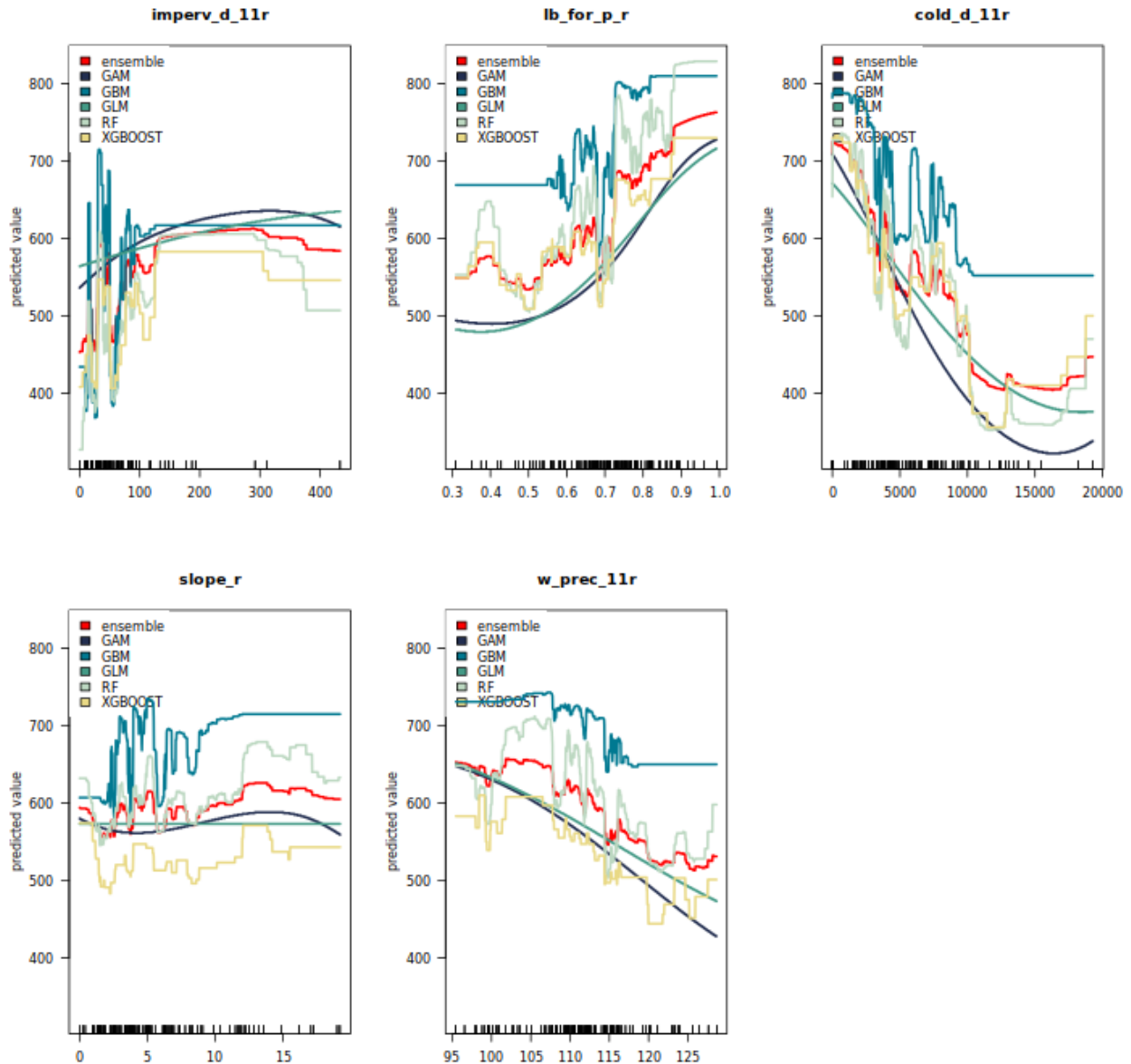


Figure 6. Response curves for the predictors in the brook trout models: distance to impervious surfaces (impervd_11r), forest proportion in local basin (lb_for_p_r), distance to cold-water site (cold_d_11r), slope (slope_r), and winter precipitation (w_prec_11r). All distance and topographic predictors are measured in meters and winter precipitation is measured in millimeters. Each model evaluated represents a different colored line: Generalized Additive Model (GAM), Gradient Boosted Machine (GBM), Generalized Linear Model (GLM), Random Forest (RF), XGBoost, and the Ensembled approach (EF).

Herps

The SGCN modeled within the herp group included 4 species: eastern hog-nosed snake (*Heterodon platirhinos*), ribbonsnake (*Thamnophis saurita saurita*), spotted turtle (*Clemmys guttata*), and wood turtle (*Glyptemys insculpta*). The eastern hog-nosed snake was a terrestrial only model, meaning all waterbodies identified in the USGS Hydrography Dataset (USGS 2005) were removed from the predictors. Ribbonsnake, spotted turtle, and wood turtle were both aquatic and terrestrial models, thus predictors were only clipped to the extent of the state.

Eastern hog-nosed snake (Heterodon platirhinos)

Species Data

NDDDB records collected from 2017-2023 were used as the presence locations for eastern hog-nosed snake and were thinned by 1.0 km, which resulted in the initial 25 presence locations being reduced to 15 presence locations after spatial thinning.

Predictors

The final list of predictors in the model included distance to buildings, presence of barrens and mixed forest, elevation, and percent sand in the soil. The building footprints layer was used to measure distance to buildings (Microsoft 2019). The 1.0 m resolution Coastal Change Analysis Program land cover map was used to extract bare lands, grasslands, and scrub/shrub to create a proxy for the presence of barrens and to also extract mixed forest (NOAA 2020). Due to the high amount of forest within the state, the mixed forest predictor was the presence of mixed forest rather than the distance to mixed forest, and the mixed forest predictor was resampled to a 10.0 m resolution. Elevation was retrieved from 30.0 m elevation tiles (NASA 2000). The elevation predictor was resampled to 10.0 m to align with the spatial resolution of the models. Percent sand data was retrieved from the Soil Survey Geographic Database (Natural Resources Conservation Service (NRCS) 2023) using the soilDB package in R (Beaudette et al. 2025). The percent sand, elevation, forest, and building distance predictors were included based on suggestions by taxa team experts and eastern hog-nosed snake's association with open habitats in other studies (Buchanan et al. 2017). Other predictors that were assessed in the models but were excluded due to reducing model performance included distance to wetlands, impervious development, grasslands, young forest and shrubland habitat, and edge of young forest and shrubland habitat, and percent canopy cover. Predictors that included waterbodies within the extent were also tested, due to their prey species being primarily toads, however, these predictors reduced model performance.

Model Evaluation and Selection

The model with the highest performance was the ensemble of all models (GAM, MARS, XGBOOST). The evaluation metrics met 3 of the 4 performance thresholds with an AUC of 0.94, Boyce of 0.65, TSS of 0.89, and Kappa 0.25 (Table 9). Predictors with high variable contributions included distance to buildings and presence or absence of barrens (value ≥ 1.0), while mixed forest, elevation and percent sand had variable contribution values < 1.0 (Table 10). Eastern hog-nosed snake habitat suitability was positively

associated with further distances from buildings, presence of barrens and percent sand (above 90%) and negatively associated with elevation (Figure 7).

Tables and Figures

Table 9 Evaluation metrics for each model tested for eastern hog-nosed snake, Generalized Additive Model (GAM), Multivariate Adaptive Regression Splines (MARS), XGBoost, and the Ensembled approach (EF). The grey highlighted row is the final model chosen for the mapping process.

Model	AUC	Boyce	TSS	Kappa
GAM	0.93	0.62	0.86	0.18
MARS	0.94	0.62	0.89	0.12
XGBOOST	0.89	0.68	0.79	0.18
EF	0.94	0.65	0.89	0.25

Table 10. Variable contribution values for the predictors within the eastern hog-nosed snake models. Contribution values ≥ 1.00 indicate the contribution value for that predictor is higher than or equal to the average contribution value. The grey highlighted row is the final model chosen for the mapping process.

Predictor	GAM	MARS	XGBOOST	EF
Building distance	1.69	1.32	1.67	1.56
Barrens	1.13	0.99	1.16	1.09
Mixed forest	0.71	0.82	0.66	0.73
Elevation	1.16	0.72	0.69	0.86
Percent Sand	0.59	1.26	1.06	0.97

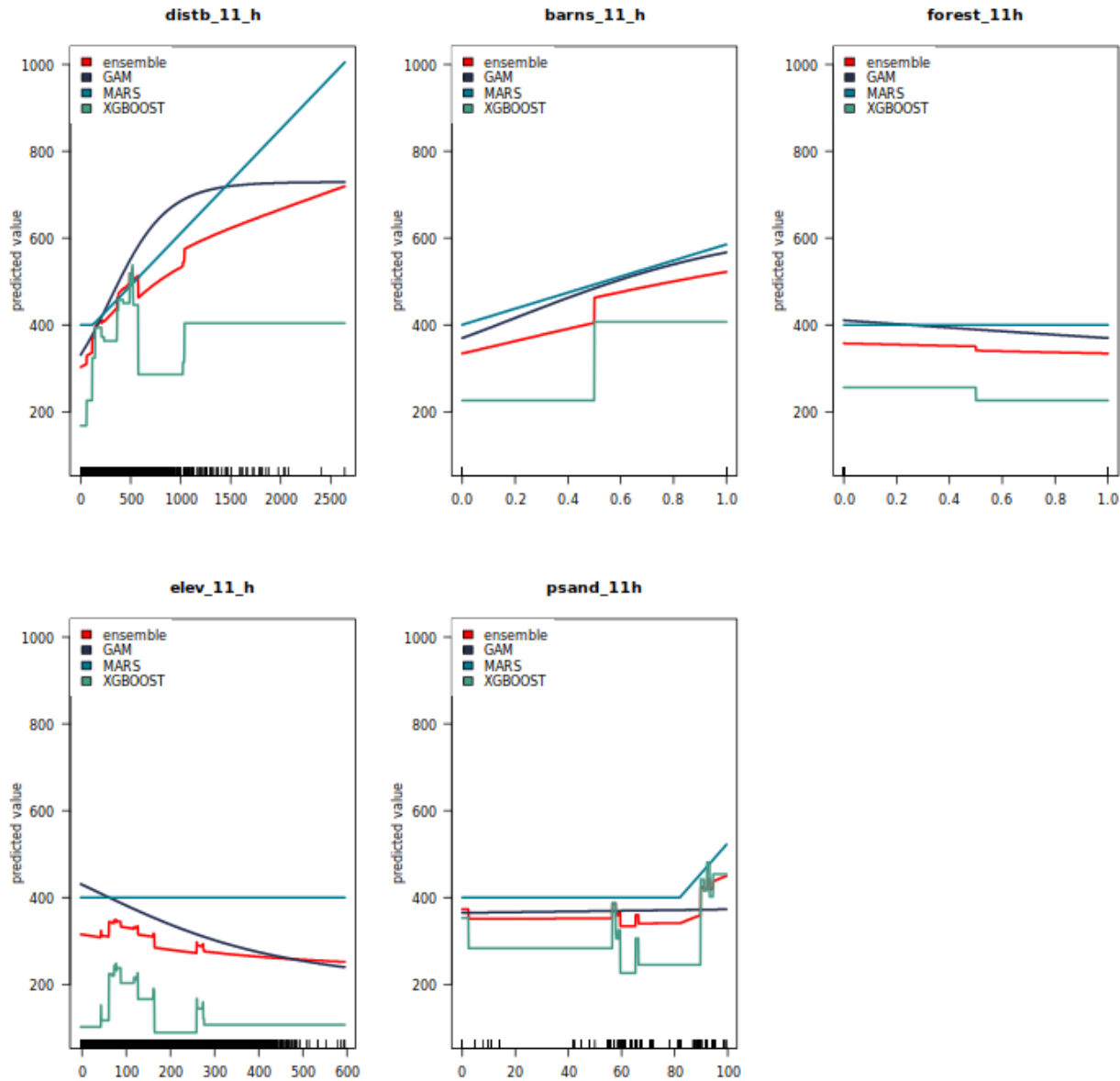


Figure 7. Response curves for the predictors in eastern hog-nosed snake models: distance to buildings (distb_11_h), presence of barrens (barns_11_h), presence of mixed forests (forest_11h), elevation (elev_11_h), and percent sand (psand_11_h). All distance and topographic predictors are measured in meters. Each model evaluated represents a different colored line: Generalized Additive Model (GAM), Multivariate Adaptive Regression Splines (MARS), XGBoost, and the Ensembled approach (EF).

Ribbonsnake (Thamnophis saurita saurita)

Species Data

Presence locations of ribbonsnake were retrieved from the NDDDB records from 2016-2023. Before thinning the data by 1 km, the dataset included 29 presence locations and after thinning, 21 presence locations were included in the final model.

Predictors

Distance to impervious surfaces, palustrine wetlands, and waterbodies, presence of mixed forest, annual total precipitation, and poorly drained soils were included as predictors in the model. The 1.0 m resolution Coastal Change Analysis Program land cover map was used to extract impervious surfaces, palustrine wetlands (emergent, scrub/shrub, and forested palustrine wetlands were combined), and mixed forests (NOAA 2020). Because of the high prevalence of forest within the state, the mixed forest predictor was the presence of mixed forest rather than the distance to mixed forest, and the mixed forest predictor was resampled to 10.0 m. The 1.0 km resolution annual total precipitation (mm) data (Thornton et al. 2020) was used to average annual total precipitation across 2016-2019 and was resampled to 10.0 m resolution. Poorly and very poorly drained soils classes were retrieved from the Soil Survey Geographic Database (NRCS 2023) using the soilDB package in R (Beaudette et al. 2025). The National Hydrography Dataset Plus (Buto and Anderson 2020) was used to determine distance to the nearest waterbody. Distance to wetlands, waterbodies, and annual precipitation and presence of poorly drained soils was included due to ribbonsnake's use of aquatic and semi-aquatic habitats (Bell et al. 2007). In Connecticut, ribbonsnakes are occasionally found in wet woodland habitat, thus the presence of mixed forest was included as a predictor (CT DEEP 2018a). Roadways are a source of mortality for ribbonsnakes (NH Fish and Game 2015), thus distance to impervious surfaces was included. Distance to young forest and shrubland and elevation were also evaluated as predictors in the model, but they were removed because they reduced model performance.

Model Evaluation and Selection

The model with the highest performance was GBM, with AUC value of 0.84, Boyce value of 0.58, TSS value of 0.70, and Kappa value of 0.14 (Table 11), meeting 3 of the 4 evaluation metric thresholds. Predictors with the highest contribution (value ≥ 1.00) were distance to wetlands (1.12), waterbodies (1.00), annual precipitation (1.00), and presence of forest (1.05), while distance to impervious surface and poorly drained soils variable contribution values were < 1.00 (Table 12). Ribbonsnake habitat suitability increased with increasing distance to impervious surfaces, precipitation, and the presence of poorly drained soils and suitability decreased with the presence of forest, and the increasing distance to palustrine wetlands and waterbodies (Figure 8).

Tables and Figures

Table 11 Evaluation metrics output for each ribbonsnake model, Generalized Additive Model (GAM), Gradient Boosted Machine (GBM), Generalized Linear Model (GLM), Multivariate adaptive regression splines (MARS), XGBoost, and the Ensembled approach (EF). The grey highlighted row is the final model chosen for the mapping process.

Model	AUC	Boyce	TSS	Kappa
GAM	0.82	0.72	0.60	0.05
GBM	0.84	0.58	0.70	0.14
GLM	0.66	0.71	0.37	0.02
MARS	0.80	0.66	0.62	0.04
XGBOOST	0.84	0.59	0.76	0.05
EF	0.80	0.69	0.60	0.06

Table 12. Variable contribution values for the predictors within the ribbonsnake models: Generalized Additive Model (GAM), Gradient Boosted Machine (GBM), Generalized Linear Model (GLM), Multivariate adaptive regression splines (MARS), XGBoost, and the Ensembled approach (EF). Contribution values ≥ 1.00 indicate the contribution value for that predictor is higher than or equal to the average contribution value. The grey highlighted row is the final model chosen for the mapping process.

Predictor	GAM	GBM	GLM	MARS	XGBOOST	EF
Impervious surface distance	0.88	0.87	1.58	0.78	0.86	0.98
Mixed forest	1.08	1.05	1.65	1.15	1.07	1.18
Palustrine wetland distance	1.17	1.12	0.00	1.30	1.18	0.98
Waterbody distance	1.05	1.00	1.58	0.85	0.96	1.07
Annual precipitation	0.93	1.00	1.35	1.14	0.98	1.07
Poorly drained soils	0.91	0.98	1.19	0.86	0.98	0.98

2025 Connecticut Wildlife Action Plan

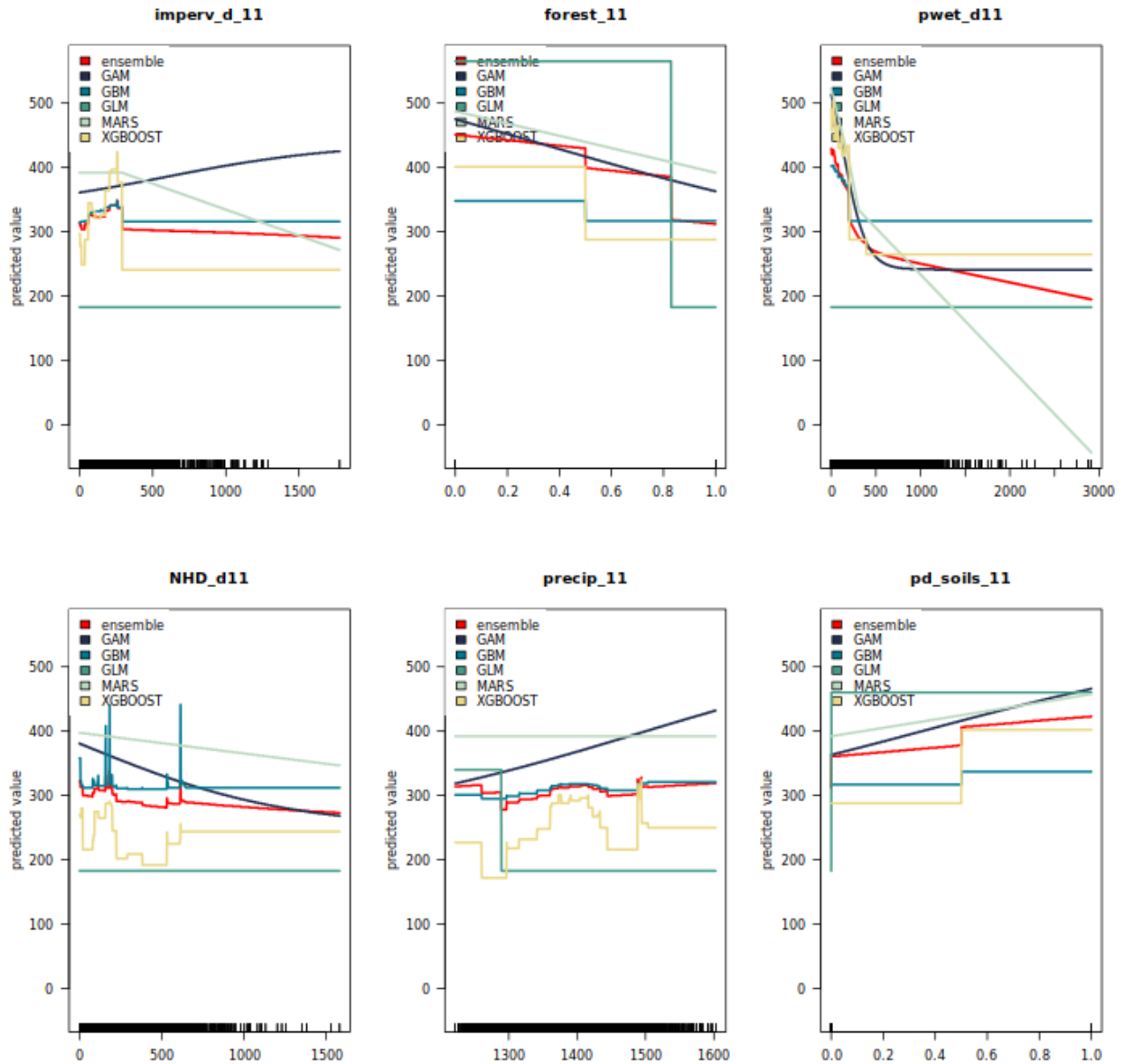


Figure 8. Response curves for the predictors in the ribbonsnake models: distance to impervious surfaces (imperv_d_11), presence of mixed forest (forest_11), distance to palustrine wetlands (pwet_d11), distance to waterbodies (NHD_d11), annual precipitation totals (precip_11), and presence of poorly drained soils (pd_soils_11). All distance and topographic predictors are measured in meters and annual precipitation is in millimeters. Each model evaluated represents a different colored line: Generalized Additive Model (GAM), Gradient Boosted Machine (GBM), Generalized Linear Model (GLM), Multivariate adaptive regression splines (MARS), XGBoost, and the Ensembled approach (EF).

Spotted turtle (*Clemmys guttata*)*Species Data*

Spotted turtle presence data was retrieved from NDDDB records collected from 2016-2024. In that time, there were 155 presence locations and after thinning data to 1.0 km, there was a total of 123 presence locations.

Predictors

The final list of predictors included distance to impervious surfaces and palustrine wetlands, elevation, percent sand, and presence of mixed forest. The 1.0 m resolution Coastal Change Analysis Program land cover map was used to extract impervious surfaces, palustrine wetlands (emergent, scrub/shrub, and forested palustrine wetlands were combined), and mixed forests (NOAA 2020). Due to the high prevalence of forest within the state, the mixed forest predictor was the presence mixed forest rather than the distance to mixed forest, and the mixed forest predictor was resampled to a 10.0 m resolution. Elevation was calculated from 30.0 m elevation tiles (NASA 2000). The elevation predictor was resampled to 10.0 m to align with the spatial resolution of the models. Percent sand data was retrieved from the Soil Survey Geographic Database (NRCS 2023) using the soilDB package in R (Beaudette et al. 2025). Mixed forests, palustrine wetlands, and percent sand were included since these habitat characteristics are associated with habitat use and nest site locations (Milam and Melvin 2001). Impervious surfaces were included since road mortality is a threat to spotted turtles (Beaudry et al. 2008). Distance to young forest and shrubland, open water, emergent wetlands, ponds, poorly drained soils, roads, slope, and canopy cover, elevation, and precipitation were also evaluated as predictors, but these predictors reduced model performance, so these predictors were excluded from the final models.

Model Evaluation and Selection

The highest performing model was the ensemble model, with an AUC of 0.83, Boyce of 0.86, TSS of 0.59, and Kappa of 0.17 (Table 13), meeting 3 of the 4 evaluation thresholds. Mixed forest and palustrine wetlands had a high variable contribution value (≥ 1.0), while impervious surface, elevation, and percent sand had low values at 0.77, 0.96, and 0.82 (Table 14). Spotted turtle habitat suitability was negatively associated with increasing distance to palustrine wetland and elevation (Figure 9).

Tables and Figures

Table 13. Evaluation metrics output for each spotted turtle model, Artificial Neural Network (ANN), Generalized Additive Model (GAM), Gradient Boosted Machine (GBM), Generalized Linear Model (GLM), Random Forest (RF), XGBoost, and the Ensembled approach (EF). The grey highlighted row is the final model chosen for the mapping process.

Model	AUC	Boyce	TSS	Kappa
ANN	0.82	0.75	0.58	0.18
GAM	0.83	0.83	0.60	0.12
GBM	0.74	0.78	0.40	0.19
GLM	0.82	0.72	0.62	0.11
RF	0.74	0.58	0.38	0.20

2025 Connecticut Wildlife Action Plan

XGBOOST	0.81	0.89	0.52	0.14
EF	0.83	0.86	0.59	0.17

Table 14. Variable contribution values for the predictors within the spotted turtle models: Artificial Neural Network (ANN), Generalized Additive Model (GAM), Gradient Boosted Machine (GBM), Generalized Linear Model (GLM), Random Forest (RF), XGBoost, and the Ensembled approach (EF). Contribution values ≥ 1.00 indicate the contribution value for that predictor is higher than or equal to the average contribution value. The grey highlighted row is the final model chosen for the mapping process.

Model	ANN	GAM	GBM	GLM	RF	XGBOOST	EF
Impervious surface distance	0.67	0.67	0.69	0.94	0.83	0.84	0.77
Mixed forest	1.18	1.15	1.24	0.99	0.81	1.05	1.07
Palustrine wetland distance	1.68	1.67	1.39	1.33	1.82	1.33	1.53
Elevation	0.94	1.05	1.11	0.85	0.92	0.90	0.96
Percent sand	0.76	0.70	0.73	0.94	0.87	0.93	0.82

2025 Connecticut Wildlife Action Plan

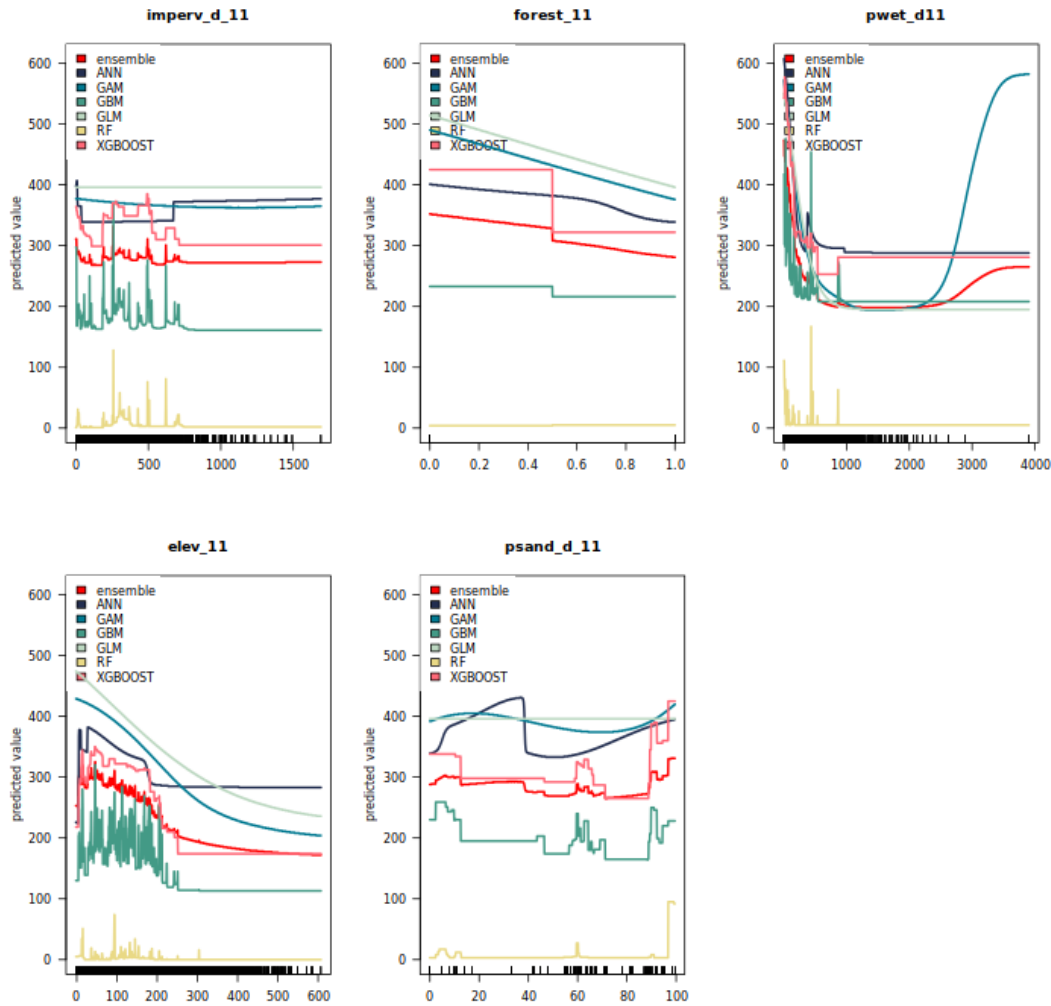


Figure 9. Response curves for the predictors in the spotted turtle models: distance to impervious surfaces (imperv_d_11), presence of mixed forest (forest_11), distance to palustrine wetlands (pwet_d11), elevation (elev_11), and percent sand (psand_d_11). All distance and topographic predictors are measured in meters. Each model evaluated represents a different colored line: Artificial Neural Network (ANN), Generalized Additive Model (GAM), Gradient Boosted Machine (GBM), Generalized Linear Model (GLM), Random Forest (RF), XGBoost, and the Ensembled approach (EF).

Wood turtle (Glyptemys insculpta)

Species Data

The wood turtle presence locations were retrieved from the NDDDB records from 2016-2023. The 135 presence locations were thinned to 1.0 km to reduce spatial autocorrelation, leading to 89 presence locations used within the final models.

Predictors

The final list of predictors included distance to roads, young forest and shrubland, palustrine wetland, waterbodies, and agriculture, and percent canopy cover. The Young Forest and Shrubland Vegetation Map was used to identify different types of young forest and shrubland vegetation within Connecticut (Rittenhouse et al. 2022). The map used ecological processes (succession, disturbance, regeneration, and hydrology), vegetation height, percent vegetation cover by height category, previous land cover type, and time since disturbance to classify different types of young forest and shrubland vegetation (Rittenhouse et al. 2022). The 1.0 m resolution Coastal Change Analysis Program land cover map was used to extract palustrine wetlands (combined emergent, scrub/shrub, and forested palustrine wetlands) and agriculture (hay/pasture and cultivated crops were combined; NOAA 2020). Road data was retrieved from CT Roads and Trails master layer, where only roads that could be traveled by cars (US and state highways and local roads) were selected (USGS 2005). The National Hydrography Dataset Plus (Buto and Anderson 2020) was used to determine distance to the nearest waterbody since this layer can identify small streams the wood turtle would use (Mothes et al. 2020). The 30.0 m USDA National Landcover Dataset of Tree Canopy Cover (USDA 2021) was used to extract percent canopy cover and then resampled to a 10.0 m resolution. Percent canopy cover was included as a predictor since it was a high contributing predictor in previous habitat suitability models (Mothes et al. 2020). Waterbodies and impervious surfaces were included since wood turtle relative abundance can be high in streams and lower in areas with more impervious surfaces (Willey et al. 2022). Agriculture, wetlands, and young forest and shrubland were also included since these are habitats wood turtles use in CT (CT DEEP 2011a). Mixed forest, impervious surfaces, precipitation, and poorly drained soils were also evaluated as predictors in the model, but these predictors reduced model performance.

Model Evaluation and Selection

The highest performing model was the GAM model with AUC of 0.82, Boyce of 0.83, TSS of 0.59, and Kappa of 0.30 (Table 15), meeting 3 of the 4 evaluation metrics thresholds. The predictors with high contributions (values ≥ 1.00) included distance to waterbodies (1.10), palustrine wetlands (1.04), and young forest and shrublands (1.00), with distance to roads and agriculture and percent canopy covers having low contribution values < 1.00 (Table 16). Wood turtle habitat suitability values decreased with increasing distances to roads, young forest and shrubland, waterbodies, and agricultural lands (Figure 10).

Tables and Figures

Table 15. Evaluation metrics output for each frosted elfin model, Generalized Additive Model (GAM), Gradient Boosted Machine (GBM), Generalized Linear Model (GLM), Multivariate adaptive regression splines (MARS), Random Forest (RF), XGBoost, and the Ensembled approach (EF). The grey highlighted row is the final model chosen for the mapping process.

Model	AUC	Boyce	TSS	Kappa
GAM	0.82	0.83	0.59	0.30
GBM	0.72	0.84	0.34	0.19
GLM	0.81	0.72	0.56	0.26
MARS	0.83	0.92	0.53	0.22
RF	0.67	0.82	0.32	0.12
XGBOOST	0.81	0.88	0.54	0.33
EF	0.82	0.90	0.56	0.31

Table 16. Variable contribution values for the predictors within the wood turtle models: Generalized Additive Model (GAM), Gradient Boosted Machine (GBM), Generalized Linear Model (GLM), Multivariate adaptive regression splines (MARS), Random Forest (RF), XGBoost, and the Ensembled approach (EF). Contribution values ≥ 1.00 indicate the contribution value for that predictor is higher than or equal to the average contribution value. The grey highlighted row is the final model chosen for the mapping process.

Predictor	GAM	GBM	GLM	MARS	RF	XGBOOST	EF
Road distance	0.93	0.95	0.93	0.91	1.02	0.92	0.94
Young forest and shrubland distance	1.00	0.89	1.02	0.94	0.71	0.97	0.93
Percent canopy cover	0.96	1.01	0.94	1.00	1.11	1.07	1.01
Palustrine wetland distance	1.04	1.10	1.02	1.04	1.20	1.01	1.06
Waterbody distance	1.10	1.10	1.12	1.16	1.28	1.11	1.14
Agriculture distance	0.98	0.97	0.99	0.95	0.77	0.93	0.93

2025 Connecticut Wildlife Action Plan

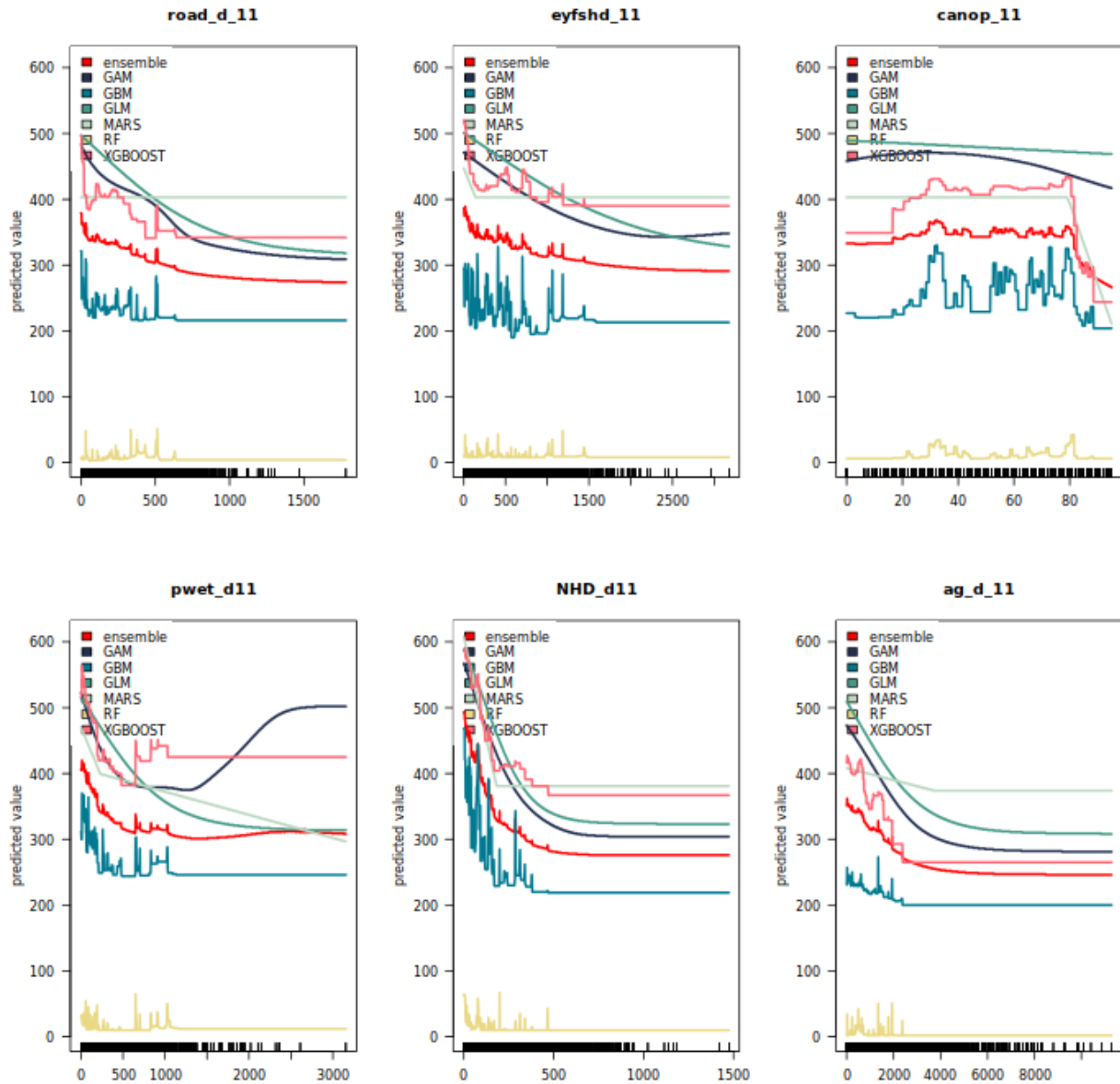


Figure 10. Response curves for the predictors in the wood turtle model: distance to roads (road_d_11), distance to young forest and shrubland (eyfshd_11), percent canopy closure (canop_11), distance to palustrine wetlands (pwet_d11), distance to waterbody (NHD_d11), and distance to agricultural area (ag_d_11). All distance and topographic predictors are measured in meters. Each model evaluated represents a different colored line: Generalized Additive Model (GAM), Gradient Boosted Machine (GBM), Generalized Linear Model (GLM), Multivariate adaptive regression splines (MARS), Random Forest (RF), XGBoost, and the Ensembled approach (EF).

Invertebrates

The SGCN modeled within the invertebrate taxa group included 4 species: eastern pondmussel (*Ligumia nasuta*), frosted elfin (*Callophrys irus*), tiger spiketail (*Cordulegaster erronea*), and yellow-banded bumble bee (*Bombus terricola*). The eastern pondmussel model was an aquatic-only model, thus all terrestrial features were removed from the predictors and subsequent model using the National Hydrography Dataset Plus (Buto and Anderson 2020). Distance predictors for all aquatic predictors used Euclidean distance rather than a true surface distance following water due to lack of continuity between all above ground waterbodies, acknowledging this is a relative measure of proximity and not the physical distance an individual would travel to reach the habitat feature. The frosted elfin and yellow-banded bumble bee were terrestrial only models, meaning all waterbodies identified in the USGS Hydrography Dataset (USGS 2005) were removed from the predictors. The tiger spiketail model was both a terrestrial and aquatic model, thus predictors were only clipped to the extent of the state.

Eastern pondmussel (Ligumia nasuta)

Species Data

Eastern pondmussel presence location data was retrieved from NDDDB records collected from 2015-2022. There are 10 presence locations within this timeframe, thus the species data was not spatially thinned.

Predictors

The final list of predictors included distance to impervious surfaces, dams, and palustrine aquatic beds, elevation, slope, and annual total precipitation. The 1.0 m resolution Coastal Change Analysis Program land cover map was used to extract palustrine aquatic beds and impervious surfaces (NOAA 2020). The 1.0 km resolution annual total precipitation data (Thornton et al. 2020) was used to average annual total precipitation across 2016-2019 and was resampled to a 10.0 m resolution. Dam location data was retrieved from CT DEEP (DeLand 2025) to create a distance to nearest dam predictor. Elevation and slope were calculated from 30.0 m elevation tiles (NASA 2000). The elevation and slope predictors were resampled to 10.0 m to align with the spatial resolution of the models. Slope, elevation, and precipitation were included as predictors since these factors affect flow, which can influence feeding and distribution of eastern pondmussel (Strayer 1993, Mistry and Ackerman 2017). Proximity of impervious surfaces and dams were included as a proxy for threats of poor water quality and limited connectivity (NY Department of Environmental Conservation 2014). The proximity to aquatic beds was included since eastern pondmussel is occasionally found near patches of aquatic vegetation (NC Wildlife Resources Commission n.d.). The initial model runs with the predictor set yielded high model performance, so additional predictors were not evaluated for this model.

Model Evaluation and Selection

The highest performing model was the MARS model, with an AUC of 0.98, Boyce of 0.63, TSS of 0.95, and Kappa of 0.67 (Table 17), meeting all 4 of the evaluation thresholds.

The top contributing predictors (variable contribution values ≥ 1.00) included distance to impervious surfaces, elevation, and distance to palustrine aquatic beds. The contribution values for slope and distance to dams were just below 1.0 at 0.91 and 0.86 and the value for annual precipitation was 0.60 (Table 18). Eastern pondmussel habitat suitability decreased with increasing distance to impervious surfaces, dams, and palustrine aquatic beds. Habitat suitability peaked at elevations around 50.0 m and 300.0 m and precipitation totals around 1400.0 mm (Figure 11).

Tables and Figures

Table 17. Evaluation metrics output for each eastern pondmussel model, Generalized Additive Model (GAM), Gradient Boosted Machine (GBM), Multivariate adaptive regression splines (MARS), XGBoost, and the Ensembled approach (EF). The grey highlighted row is the final model chosen for the mapping process.

	AUC	Boyce	TSS	Kappa
GAM	0.85	0.79	0.69	0.28
GBM	1.00	0.65	0.99	0.50
MARS	0.98	0.63	0.95	0.67
XGBOOST	0.98	0.66	0.96	0.40
EF	0.98	0.61	0.95	0.67

Table 18. Variable contribution values for the predictors within the eastern pondmussel models: Generalized Additive Model (GAM), Gradient Boosted Machine (GBM), Multivariate adaptive regression splines (MARS), XGBoost, and the Ensembled approach (EF).

Contribution values ≥ 1.00 indicate the contribution value for that predictor is higher than or equal to the average contribution value. The grey highlighted row is the final model chosen for the mapping process.

Predictor	GAM	GBM	MARS	XGBOOST	EF
Impervious surface distance	1.23	1.16	1.31	1.35	1.26
Elevation	0.98	2.21	1.49	1.51	1.56
Slope	0.96	2.04	0.91	1.14	1.27
Dam distance	1.00	0.80	0.86	0.54	0.79
Annual precipitation total	1.07	0.47	0.60	0.89	0.75
Palustrine aquatic bed distance	0.80	0.29	1.00	0.78	0.72

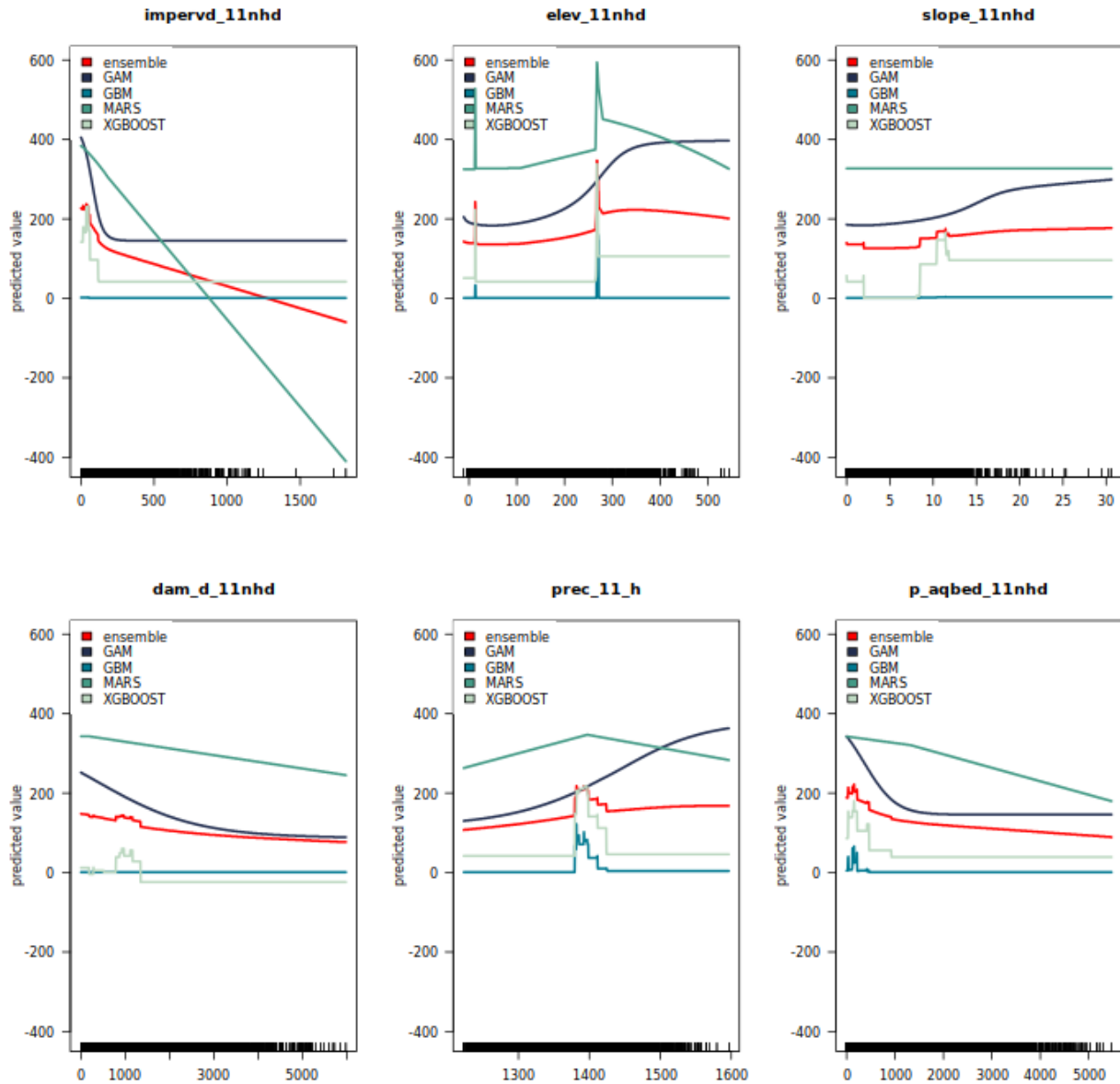


Figure 11. Response curves for the predictors in the eastern pondmussel models: distance to surfaces (impervd_11nhd), elevation (elev_11nhd), slope (slope_11nhd), distance to nearest dam (dam_d_11nhd), annual precipitation total (prec_11_h), and distance to palustrine aquatic bed (p_aqbed_11nhd). All distance and topographic predictors are measured in meters and precipitation is measured in millimeters. Each model evaluated represents a different colored line: Generalized Additive Model (GAM), Gradient Boosted Machine (GBM), Multivariate adaptive regression splines (MARS), XGBoost, and the Ensembled approach (EF).

Frosted elfin (Callophrys irus)

Species Data

Presence data of frosted elfin was retrieved from NDDDB records collected from 2022-2023 with a total of 37 locations. The frosted elfin locations were recorded as buffers, either around potential sightings or transect lines, with differing spatial and species certainty levels ranging from high to very low, thus the centroid of polygon was used as the presence location. Due to spatial and species uncertainty associated with each location and to keep as many records as possible, the data was not spatially thinned.

Predictors

The final list of predictors included elevation, distance to buildings, young forest and shrubland, and host plants, presence of barrens and powerlines, and percent sand in the soil as predictors in the model. Elevation was retrieved from 30.0 m elevation tiles (NASA 2000). The elevation predictor was resampled to a 10.0 m resolution to align with the spatial resolution of the models. The Microsoft building footprints was used as a proxy for development (Microsoft 2019). The 1.0 m resolution Coastal Change Analysis Program land cover map was used to extract bare lands, grasslands, and scrub/shrub to create a proxy for the presence of barrens (NOAA 2020). The Young Forest and Shrubland Vegetation Map was used to identify different types of young forest and shrubland vegetation within Connecticut (Rittenhouse et al. 2022). The map used ecological processes (succession, disturbance, regeneration, and hydrology), vegetation height, percent vegetation cover by height category, previous land cover type, and time since disturbance to classify different types of young forest and shrubland vegetation (Rittenhouse et al. 2022). Location of powerline right of ways were retrieved from Homeland Infrastructure Foundation Level Database (US Department of Homeland Security 2018). Percent sand data was retrieved from the Soil Survey Geographic Database (NRCS 2023) using the soilDB package in R (Beaudette et al. 2025). The locations of frosted elfin host plants, wild indigo (*Baptisia tinctoria*) and blue lupine (*Lupinus perennis*), were determined using iNaturalist. The spocc package in R (Owens et al. 2024) was used to retrieve iNaturalist records collected from 2016-2024. Presence of barrens, distance to buildings, and elevation were included based on conversations with taxa team members. Distance to young forest and shrublands and host plants and presence of powerlines were included based on previous frosted elfin adult and larvae habitat suitability studies finding positive associations with these features (Albanese et al. 2008, Shepard et al. 2021) and percent sand was included based on a study on the location of sandplains habitat within New York (Corbin and Flatland 2022). Predictors that were tested and reduced model performance included precipitation, northern and eastern aspect, slope, and soils known to support pine and sandplain barrens (76% sand at least 160 cm depth; Corbin and Flatland 2022), thus these were not included in the final model.

Model Evaluation and Selection

The highest performing model was the ensemble of all the models tested (ANN, GAM, GBM, MARS, and RF) with AUC of 0.98, Boyce of 0.78, TSS of 0.85, and Kappa of 0.83 (Table 19) meeting all 4 evaluation thresholds. Predictors with high contributions (≥ 1.00) included distance to frosted elfin host plant observation and buildings, while the rest were

< 1.00 (ranged from 0.90 to 0.99; Table 20). Based on the response curves for the ensemble model, frosted elfin suitable habitat increased with the presence of barrens and powerline corridors, increasing distance to buildings, and percent sand above 90% and decreased with increasing distance to young forest and shrubland and host plant observations (Figure 12).

Tables and Figures

Table 19 Evaluation metrics output for each model tested for frosted elfin, Artificial Neural Network (ANN), Generalized Additive Model (GAM), Generalized Boosting Model (GBM), Multivariate Adaptive Regression Splines (MARS), Random Forest (RF), and the Ensembled approach (EF). The grey highlighted row is the final model chosen for the mapping process.

Model	AUC	Boyce	TSS	Kappa
ANN	0.95	0.78	0.85	0.83
GAM	0.98	0.78	0.89	0.66
GBM	0.98	0.80	0.85	0.77
MARS	0.96	0.85	0.85	0.71
RF	0.99	0.50	0.89	0.63
EF	0.98	0.78	0.85	0.83

Table 20. Variable contribution values for the predictors within the frosted elfin models: Artificial Neural Network (ANN), Generalized Additive Model (GAM), Generalized Boosting Model (GBM), Multivariate Adaptive Regression Splines (MARS), Random Forest (RF), and the Ensembled approach (EF). Contribution values ≥ 1.00 indicate the contribution value for that predictor is higher than or equal to the average contribution value. The grey highlighted column is the final model.

Predictor	ANN	GBM	MARS	RF	EF
Elevation	0.91	0.91	0.94	0.98	0.93
Building distance	1.08	0.97	1.01	0.95	1.00
Barrens	1.04	0.94	0.95	0.96	0.97
Young forest and shrubland distance	0.93	0.99	0.93	1.03	0.97
Powerlines	1.05	0.98	0.96	0.98	0.99
Percent sand	0.84	0.98	0.95	0.84	0.90
Host plant distance	1.18	1.25	1.29	1.30	1.25

2025 Connecticut Wildlife Action Plan

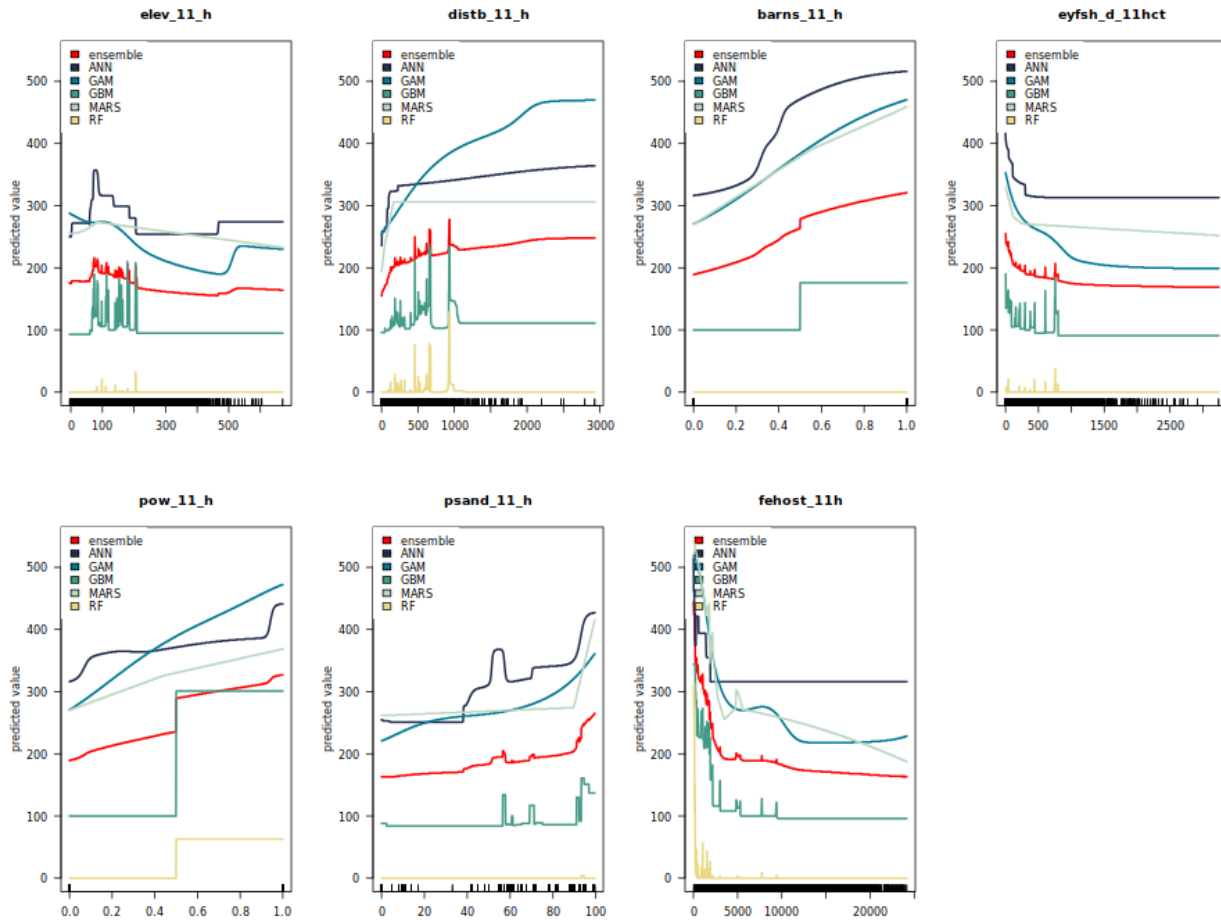


Figure 12. Response curves for the predictors in the frosted elfin models: elevation (elev_11_h), distance to buildings (distb_11_h), presence of barrens (barns_11_h), distance to young forest and shrubland (eyfsh_d_11hct), presence of powerlines (pow_11_h), percent sand (psand_11_h), and distance to host plant (fehost_11h). All distance and topographic predictors are measured in meters. Each model evaluated represents a different colored line: Artificial Neural Network (ANN), Generalized Additive Model (GAM), Generalized Boosting Model (GBM), Multivariate Adaptive Regression Splines (MARS), Random Forest (RF), and the Ensembled approach (EF).

Tiger spiketail (Cordulegaster erronea)

Species Data

Tiger spiketail presence locations were retrieved from NDDDB records collected in 2022, leading to a total of 13 presence locations in the final model. The data was not spatially thinned due to the low number of presence locations.

Predictors

The final list of predictors included distance to impervious surfaces and rivers, and presence of mixed forest, total precipitation in May, poorly drained soils, and slope. The 1.0 m resolution Coastal Change Analysis Program land cover map was used to extract mixed forest and impervious surfaces (NOAA 2020). Because of the high prevalence of forest within the state, the mixed forest predictor was the presence of mixed forest rather than the distance to mixed forest, and the mixed forest predictor was resampled to a 10.0 m resolution. Poorly and very poorly drained soils classes were retrieved from the Soil Survey Geographic Database (NRCS 2023) using the soilDB package in R (Beaudette et al. 2025). Slope was calculated from 30.0 m elevation tiles (NASA 2000). The slope predictor was resampled to 10.0 m to align with the spatial resolution of the models. River data was extracted using the USGS hydrography dataset (USGS 2005) to remove waterbodies within the National Hydrography Dataset Plus dataset (Buto and Anderson 2020) that intersect lakes, ponds, and reservoirs. The 1.0 km resolution annual May precipitation data (Thornton et al. 2022) was used to average May total precipitation across 2016-2022. Precipitation was extracted for May only, since May precipitation increased tiger spiketail habitat suitability (Howard and Schlesinger 2012). Species distribution models also determined slope was positively associated with tiger spiketail distribution (Howard and Schlesinger 2012). Mixed forest, impervious surfaces, poorly drained soils, and rivers were included in the models due to the presence of forests, streams, and buildings in home ranges and catchments occupied by tiger spiketail (Moskowitz and May 2017). Elevation and distance to waterbodies were also evaluated as predictors, however, these predictors reduced model performance, so they were removed from final models.

Model Evaluation and Selection

The highest performing model included the MARS model, with an AUC value of 0.99, Boyce value of 1.00, TSS value of 0.97, and Kappa of 0.40 (Table 21), meeting 3 of the 4 model evaluation thresholds. The highest contributing predictors (values ≥ 1.00) were distance to impervious development (1.14), presence of mixed forest (1.00), and distance to rivers (1.46), while the lowest contributing (values < 1.00) were May precipitation (0.76), poorly drained soils (0.99), and slope (0.76; Table 22). Tiger spiketail habitat suitability increased with increasing distance to impervious surfaces and presence of mixed forest and decreased with increasing distance to rivers (Figure 13).

Tables and Figures

Table 21. Evaluation metrics output for each tiger spiketail model, Flexible Discriminant Analysis (FDA), Generalized Linear Model (GLM), Multivariate adaptive regression splines (MARS), and the Ensembled approach (EF). The grey highlighted row is the final model chosen for the mapping process.

Model	AUC	Boyce	MaxTSS	MaxKappa
FDA	0.98	0.69	0.97	0.50
GLM	0.98	-0.07	0.95	0.50
MARS	0.99	1.00	0.97	0.40
XGBOOST	0.94	0.74	0.87	0.50
EF	0.98	0.43	0.97	0.67

Table 22. Variable contribution values for the predictors within the tiger spiketail models Flexible Discriminant Analysis (FDA), Generalized Linear Model (GLM), Multivariate adaptive regression splines (MARS), and the Ensembled approach (EF). Contribution values ≥ 1.00 indicate the contribution value for that predictor is higher than or equal to the average contribution value. The grey highlighted row is the final model chosen for the mapping process.

Predictor	FDA	GLM	MARS	XGBOOST	EF
Impervious surface distance	0.54	0.35	1.14	1.21	0.80
Mixed forest	1.18	1.64	1.00	1.02	1.21
May precipitation	0.85	0.76	0.76	0.33	0.68
River distance	2.45	2.09	1.46	1.58	1.90
Poorly drained soils	1.21	0.90	0.99	1.00	1.03
Slope	0.48	0.90	0.76	1.17	0.82

2025 Connecticut Wildlife Action Plan

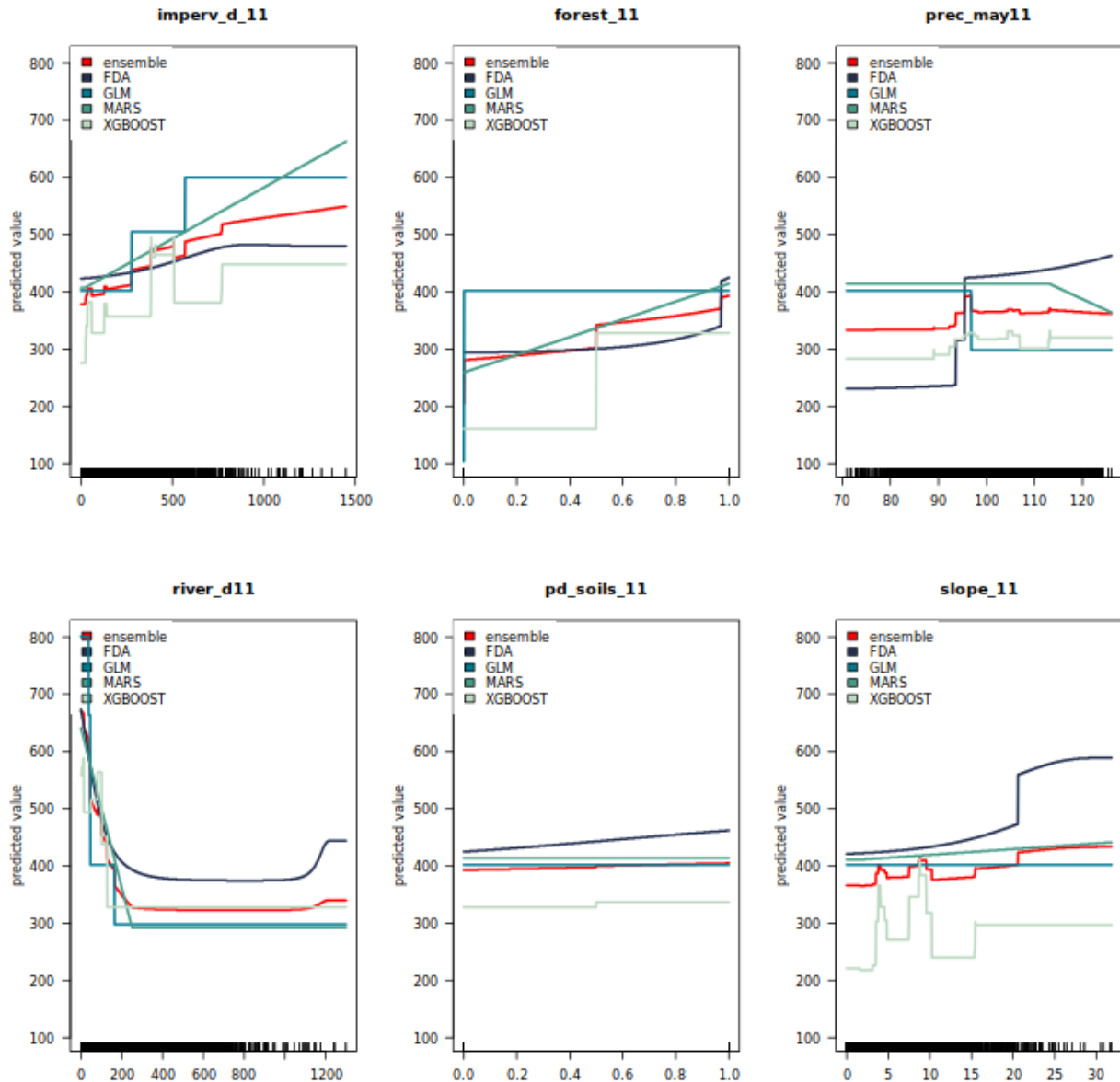


Figure 13. Response curves for the predictors in the tiger spiketail model: distance to impervious surfaces (imperv_d_11), presence of mixed forest (forest_11), total May precipitation (prec_may11), distance to rivers (river_d11), presence of poorly drained soils (pd_soils_11), and slope (slope_11). All distance and topographic predictors are measured in meters and precipitation is measured in millimeters. Each model evaluated represents a different colored line: Artificial Neural Network (ANN), Generalized Additive Model (GAM), Generalized Boosting Model (GBM), Multivariate Adaptive Regression Splines (MARS), Random Forest (RF), and the Ensembled approach (EF).

Yellow-banded bumble bee (*Bombus terricola*)

Species Data

The presence locations for the yellow-banded bumble bee were retrieved from NDDDB records and collected from 2016-2023. The total number of presence locations within this time was 14, thus species data was not thinned since this was below the threshold for thinning data.

Predictors

Distance to young forest and shrubland, agriculture, grasslands, and host plants, and the presence of mixed forest were used as predictors within the final model. The building footprints layer was used as a proxy for development (Microsoft 2019). The Young Forest and Shrubland Vegetation Map was used to identify different types of young forest and shrubland vegetation within Connecticut (Rittenhouse et al. 2022). The map used ecological processes (succession, disturbance, regeneration, and hydrology), vegetation height, percent vegetation cover by height category, previous land cover type, and time since disturbance to classify different types of young forest and shrubland vegetation (Rittenhouse et al. 2022). The 1.0 m resolution Coastal Change Analysis Program land cover map was used to extract grasslands, mixed forests, and agriculture (cultivated land and pasture/hay were combined; NOAA 2020). The spocc package in R (Owens et al. 2024) was used to extract iNaturalist records of yellow-banded bumble bee host plant locations collected from 2016-2024 of research grade quality. The list of host plants was collected from Canada Environment and Climate Change (2022) and included: *Anaphalis margaritacea*, *Aquilegia canadensis*, *Aralia*, *Arctostaphylos uva-ursi*, *Asclepias incarnata*, *Asclepias syriaca*, *Astragalus*, *Baptisia tinctoria*, *Berberis thunbergii*, *Caragana arborescens*, *Carduus nutans*, *Centaurea jacea*, *Chamaenerion angustifolium*, *Cirsium arvense*, *Crocus spp.* *Crocuses*, *Diervilla lonicera*, *Epigaea repens*, *Erigeron philadelphicus*, *Eupatorium fistulosum*, *Eupatorium maculatum*, *Eurybia macrophylla*, *Euthamia graminifolia*, *Echium vulgare*, *Gaylussacia*, *Heracleum lanatum*, *Hydrophyllum virginianum*, *Hypericum perforatum*, *Impatiens capensis*, *Kalmia angustifolia*, *Lactuca Canadensis*, *Rhododendron groenlandicum*, *Linaria vulgaris*, *Lonicera caerulea*, *Lonicera tatarica*, *Lupinus*, *Melilotus albus*, *Medicago sativa*, *Mertensia*, *Monarda fistulosa*, *Onopordum acanthium*, *Philadelphus*, *Pontederia cordata*, *Prunus cerasus*, *Prunus pensylvanica*, *Prunus tomentosa*, *Malus pumila*, *Rhexia virginica*, *Rhus typhina*, *Ribes grossularia*, *Ribes nigrum*, *Robinia hispida*, *Rosa*, *Rubus*, *Salix*, *Senecio*, *Solanum dulcamara*, *Solidago canadensis*, *Solidago flexicaulis*, *Solidago hispida*, *Solidago juncea*, *Sonchus oleraceus*, *Sorbus Americana*, *Spiraea latifolia*, *Symphyotrichum ericoides*, *Symphyotrichum lateriflorum*, *Symphyotrichum novae-anglia*, *Symphytum officinale*, *Syringa vulgaris*, *Taraxacum officinale*, *Thalictrum pubescens*, *Tilia americana*, *Tilia platyphyllos*, *Trifolium hybridum*, *Trifolium pretense*, *Trifolium repens*, *Vaccinium angustifolium*, *Vaccinium corymbosum*, and *Vicia cracca*. The mixed forest, host plant, and agricultural predictors were included based on yellow-banded bumble bee habitat being positively correlated with coniferous forest and with flowering plant cover and negatively correlated with agriculture (Liczner and Colla 2020). The young forest and shrubland and grassland predictors were included since these are areas of potential

habitat (USFWS 2019). Distance to wetlands and buildings were also evaluated as predictors within the model, but these predictors reduced model performance, so they were excluded from the final model.

Model Evaluation and Selection

The model with the highest performance was the ensemble model (including ANN, GAM, GBM, and MARS), with an AUC of 0.97, Boyce of 0.72, TSS of 0.93, and Kappa of 0.40 (Table 23), meeting 3 of 4 evaluation metric thresholds. The highest contributing predictors were presence of mixed forest, distance to agriculture and grasslands with variable contributions values ≥ 1.00 (1.50 and 1.06, 1.04), while distance to young forest and shrubland and host plant had contribution values < 1.0 (0.97 and 0.77; Table 24). Yellow-banded bumble bee suitable habitat increased with increasing distance to host plants and decreased with presence of mixed forest and increasing distance to young forest and shrubland habitat (Figure 14).

Tables and Figures

Table 23. Evaluation metrics output for each model tested for yellow-banded bumble bee, Artificial Neural Network (ANN), Generalized Additive Model (GAM), Generalized Boosting Model (GBM), Multivariate Adaptive Regression Splines (MARS), and the Ensembled approach (EF). The grey highlighted row is the final model chosen for the mapping process.

Model	AUC	Boyce	TSS	Kappa
ANN	0.96	0.54	0.92	0.22
GAM	0.94	0.35	0.89	0.11
GBM	0.94	0.31	0.92	0.05
MARS	0.94	0.52	0.90	0.04
EF	0.97	0.72	0.93	0.40

Table 24. Variable contribution values for the predictors within the yellow-banded bumble bee model for each model tested: Artificial Neural Network (ANN), Generalized Additive Model (GAM), Generalized Boosting Model (GBM), Multivariate Adaptive Regression Splines (MARS), and the Ensembled approach (EF). Contribution values ≥ 1.00 indicate the contribution value for that predictor is higher than or equal to the average contribution value. The grey highlighted column is the final model.

Predictor	ANN	GAM	GBM	MARS	EF
Young forest and shrubland distance	0.63	1.24	0.51	1.51	0.97
Mixed forest	1.57	1.16	1.61	1.65	1.50
Host plant distance	0.40	0.81	0.90	0.96	0.77
Agriculture distance	1.12	0.88	1.49	0.76	1.06
Grassland distance	1.86	0.96	0.83	0.49	1.04

2025 Connecticut Wildlife Action Plan

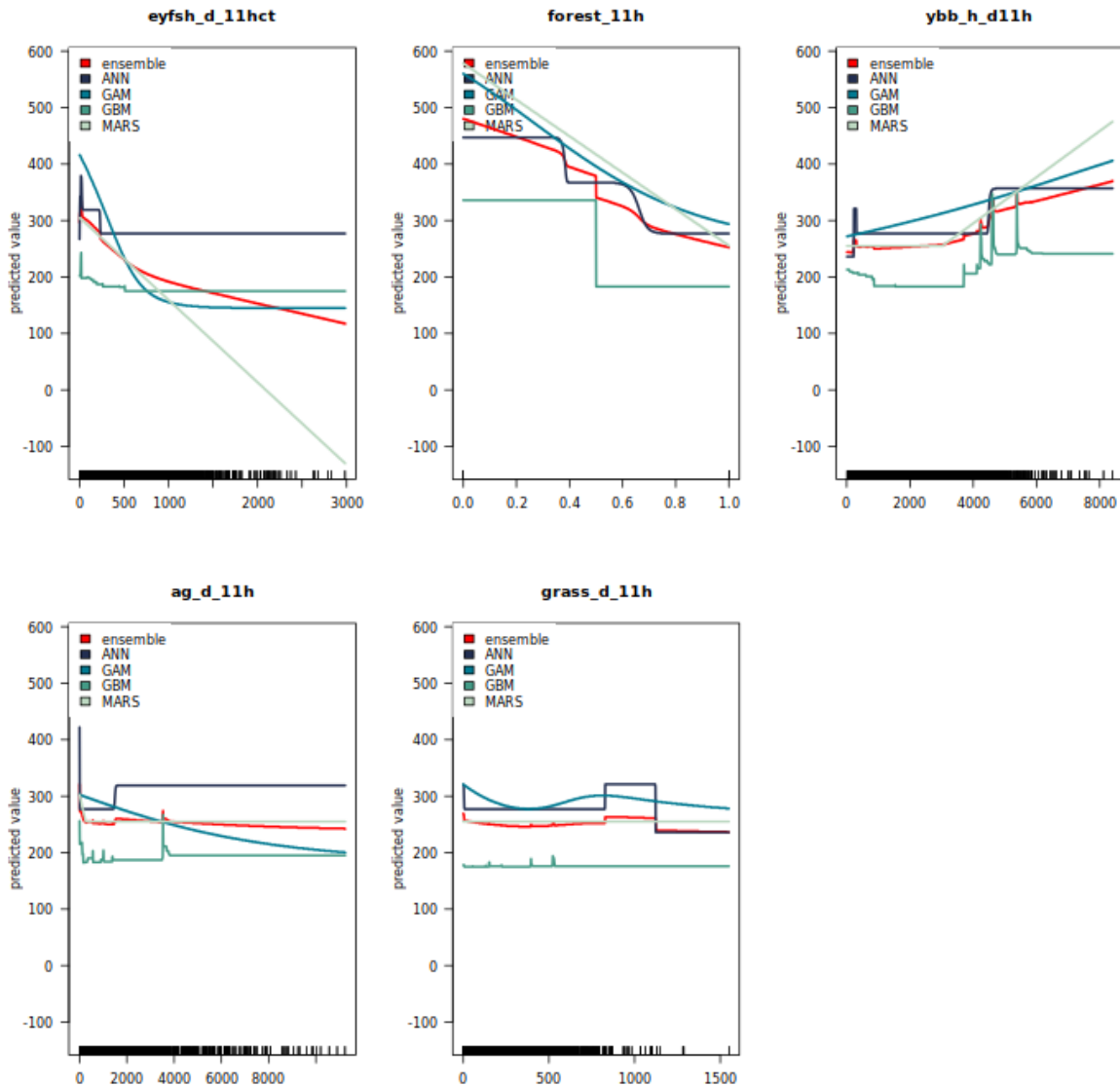


Figure 14. Response curves for the predictors in the yellow-banded bumble bee models: distance to young forest and shrubland (eyfsh_d_11hct), presence of mixed forest (forest_11h), distance to host plant (ybb_h_d11h), distance to agricultural lands (ag_d_11h), and distance to grasslands (grass_d_11h). All distance predictors are measured in meters. Each model evaluated represents a different colored line: Artificial Neural Network (ANN), Generalized Additive Model (GAM), Generalized Boosting Model (GBM), Multivariate Adaptive Regression Splines (MARS), and the Ensembled approach (EF).

Mammals

The SGCN modeled within the mammal taxa group included 2 species: New England cottontail (*Sylvilagus transitionalis*) and tricolored bat (*Perimyotis subflavus*). The New England cottontail model was a terrestrial only model, meaning all waterbodies identified in the USGS Hydrography Dataset (USGS 2005) were removed from the predictors. The tricolored bat model included both terrestrial and aquatic habitats, thus predictors were only clipped to the extent of the state.

New England cottontail (Sylvilagus transitionalis)

Species Data

The presence data used for the New England cottontail was collected by CT DEEP as part of the New England Cottontail Regional Monitoring Program from 2016-2022. This is a standardized monitoring protocol that collects cottontail pellets at designated sites and uses genetic analysis to determine species. The locations were thinned by 10.0 m, resulting in 1735 locations for New England cottontail (2086 presence locations before thinning). The monitoring protocol only records absences at the site-level, not locations, so absence information was not included within the model.

Predictors

The final list of predictors included a categorical predictor of young forest, shrubland, and understory spatial arrangement, average annual total precipitation, presence or absence of mixed forest, distance to buildings, shrublands, transitional to forest, regenerating forest and clearcut, mixed invasive, greenbrier, and Japanese barberry understory, and elevation, slope, northern aspect, and eastern aspect. The 1.0 m resolution Coastal Change Analysis Program land cover map was used to extract mixed forest (NOAA 2020). Because of the high prevalence of forest within the state, the mixed forest predictor was the presence mixed forest rather than the distance to mixed forest, and the mixed forest predictor was resampled to 10.0 m. The building footprints layer was used as a proxy for development (Microsoft 2019). The Young Forest and Shrubland Vegetation Map was used to identify different types of young forest and shrubland vegetation within Connecticut (Rittenhouse et al. 2022). The map used ecological processes (succession, disturbance, regeneration, and hydrology), vegetation height, percent vegetation cover by height category, previous land cover type, and time since disturbance to classify different types of young forest and shrubland vegetation (Rittenhouse et al. 2022). The classes of young forest and shrubland were included as separate predictors: shrubland (vegetation height primarily between 0.5-2.5m with no recent disturbance), transitional to forest (mixture of vegetation heights above and below 0.5 m with no recent disturbance), and regenerating forest and fields (young forest created from disturbance within the last 3-20 years; Rittenhouse et al. 2022). Predictors were generated for individual classes of deciduous understory vegetation, greenbrier (*Smilax rotundifolia*), Japanese barberry (*Berberis thunbergii*) and other mixed invasive understory species using the Understory Vegetation Map (Yang et al. 2023). The 30.0 m elevation dataset (NASA 2000) was used to measure elevation, slope, northern and eastern aspects. The elevation, slope, and northern and eastern aspect predictors were resampled to a 10.0

m resolution to align with the spatial resolution of the models. Using the National Wetlands Inventory (USFWS 2020), a specific list of forested/shrub wetland types found in sites known to be occupied by New England cottontail (Rittenhouse et al. 2022) was extracted. The 1.0 km resolution annual total precipitation (mm) data (Thornton et al. 2020) was used to average annual total precipitation across 2016-2019.

The young forest and shrubland vegetation map and the understory map was combined to run a morphological spatial pattern analysis (MSPA) using the Guidos Toolbox (Soille and Vogt 2009, 2022, Vogt and Riitters 2017, Vogt et al. 2022) and create a categorical predictor depicting spatial arrangement of habitat. The composite raster was inputted into Guidos Toolbox and set the presence of young forests, shrublands, and understory as the foreground and absence of those as the background. All habitat was classified into six categories: core indicated interior area, edge was the external object perimeter of the core, perforation was the internal object perimeter of the core, fragment was disjointed areas too small to contain core, margin was non-contiguous area that did not fit into the other categories, and core opening was the area within perforations (Vogt and Riitters 2017, Vogt et al. 2022). Core opening was the only background class created not from the presence of the input layer, but instead from the arrangement of the foreground around it (Vogt and Riitters 2017, Vogt et al. 2022; Figure 15). The output of the MSPA was included as a categorical predictor in the habitat suitability models. The New England cottontail model was created through a separate study (Bischoff et al. 2025), thus all distance surface predictors had an offset of 1.0 to replace all zero distance values, since zero-inflated data can bias another analysis used within the study, principal component analysis (Hellton et al. 2021).

Young forest and shrubland and understory predictors were included because the high stem density provides cover and protection from predators (Barbour and Litvaitis 1993, Litvaitis 1993, 2001, Cheeseman et al. 2018, 2019, 2021) and New England cottontail occupancy and relative abundance were positively associated with these habitat characteristics (Bischoff et al. 2023a, b). The spatial arrangement of young forest and shrubland habitat was included to understand the importance of the configuration of these habitat characteristics. Annual total precipitation was included due to the influence of winter precipitation on cottontail detection (Brubaker et al. 2014). Topographical predictors were included based on differences in topography in the two regions of the state where New England cottontail occurs. The presence of mixed forest was included since New England cottontail can be displaced into later successional forests when their introduced competitor, the eastern cottontail (*Sylvilagus floridanus*), is present (Cheeseman et al. 2018, 2021). The proximity to buildings was also included since development is a barrier to New England cottontail dispersal (Amaral et al. 2016). Another predictor tested was reverting fields, but this predictor reduced model performance and was excluded from the model.

Model Evaluation and Selection

This model was part of a separate analysis; thus, a different process was used for evaluation and selection compared to the other SGCN. The testing and training datasets were defined through a 10-fold cross validation, where the data is divided up into 10 folds or groups and each time the model is run (a total of 10 times) a different group is held out

for testing then the outputs from all runs are averaged. A cross-fold validation was used over selecting training and test data so all the data could be used for model validation (Phillips et al. 2017) and to incorporate randomness into the testing and training data that matches the randomness of the background data in the Maxent models (Elith et al. 2011). Maxent is a machine learning model that uses presence locations and predictors to estimate the unknown probability distribution of a species across geographic and environmental space, compared to the user-defined background distribution (Phillips et al. 2006). Maxent 3.4.4 (Phillips et al. 2020) was used to create the New England cottontail models. Maxent was chosen for New England cottontail because information on prevalence values across the landscape was known from previous occupancy studies (Bischoff et al. 2023a). Three separate Maxent models values were used for each of the mean, lower and upper credible intervals for the occupancy probabilities for New England cottontail from Bischoff et al. (2023a) as the prevalence values for each of the models. For the final habitat suitability map, the outputs from the mean, lower, and upper credible interval prevalence values were averaged but the results of the response curves and the percent contribution are from the mean prevalence value.

The final Maxent model did have a high AUC of 0.92 and TSS of 0.89 (Table 25). Predictors with the largest contribution were the morphological spatial pattern analysis (28.34%), distance to buildings (14.38%), shrublands (13.99%), mixed invasive understory (10.61%) and transitional to forest (9.64%; Table 26). New England cottontail habitat suitability decreased with increasing distance to shrublands, and mixed invasive understory vegetation, and increased with increasing distance to buildings and the presence of core, core openings, and perforations (Figure 16, 17).

More detailed information on this analysis and the results is in Bischoff et al. (2025).

Tables and Figures

Table 25. Average AUC and TSS values for all 10-fold cross validation New England cottontail Maxent models with standard errors.

	Metric	Standard Error
TSS	0.89	0.02
AUC	0.92	0.01

Table 26. Percent contribution for the New England cottontail Maxent model for all predictors included within the model. For the Morphological Spatial Pattern Analysis (MSPA) predictor, background indicates absence of young forest, shrubland, and understory vegetation classes. In the foreground, core is interior continuous vegetation classes, edge is the external perimeter of core, fragments are areas too small to contain core area, and margin is continuous vegetation classes that do not fit into any other category. Core openings are background classes found within core areas and perforations are the perimeter of those core openings.

Predictor	NEC % Contribution
MSPA	28.34
Building	14.38
Shrubland	13.99
Mixed invasive	10.61
Transitional to forest	9.64
Greenbrier	6.37
Barberry	5.55
Elevation	4.34
Regenerating forest	3.07
Mixed forest	1.61
Precipitation	1.36
Eastern aspect	0.24
Forested-shrub wetland	0.21
Northern aspect	0.19
Slope	0.10

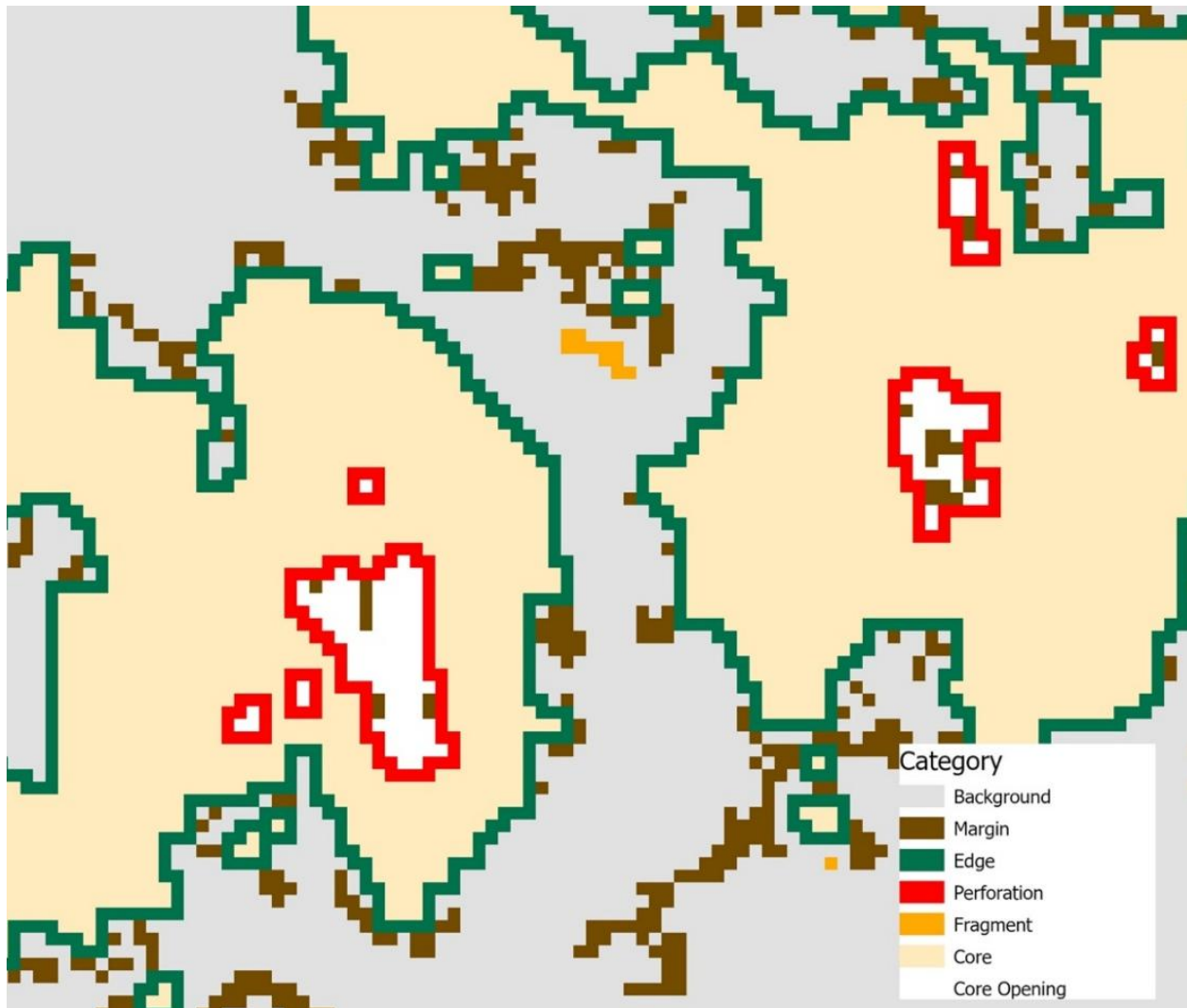


Figure 15. Examples of the morphological spatial pattern analysis categories. Background indicates absence of young forest, shrubland, and understory vegetation classes. In the foreground, core is interior continuous vegetation classes, edge is the external perimeter of core, fragments are areas too small to contain core area, and margin is continuous vegetation classes that do not fit into any other category. Core openings are background classes found within core areas and perforations are the perimeter of those core openings.

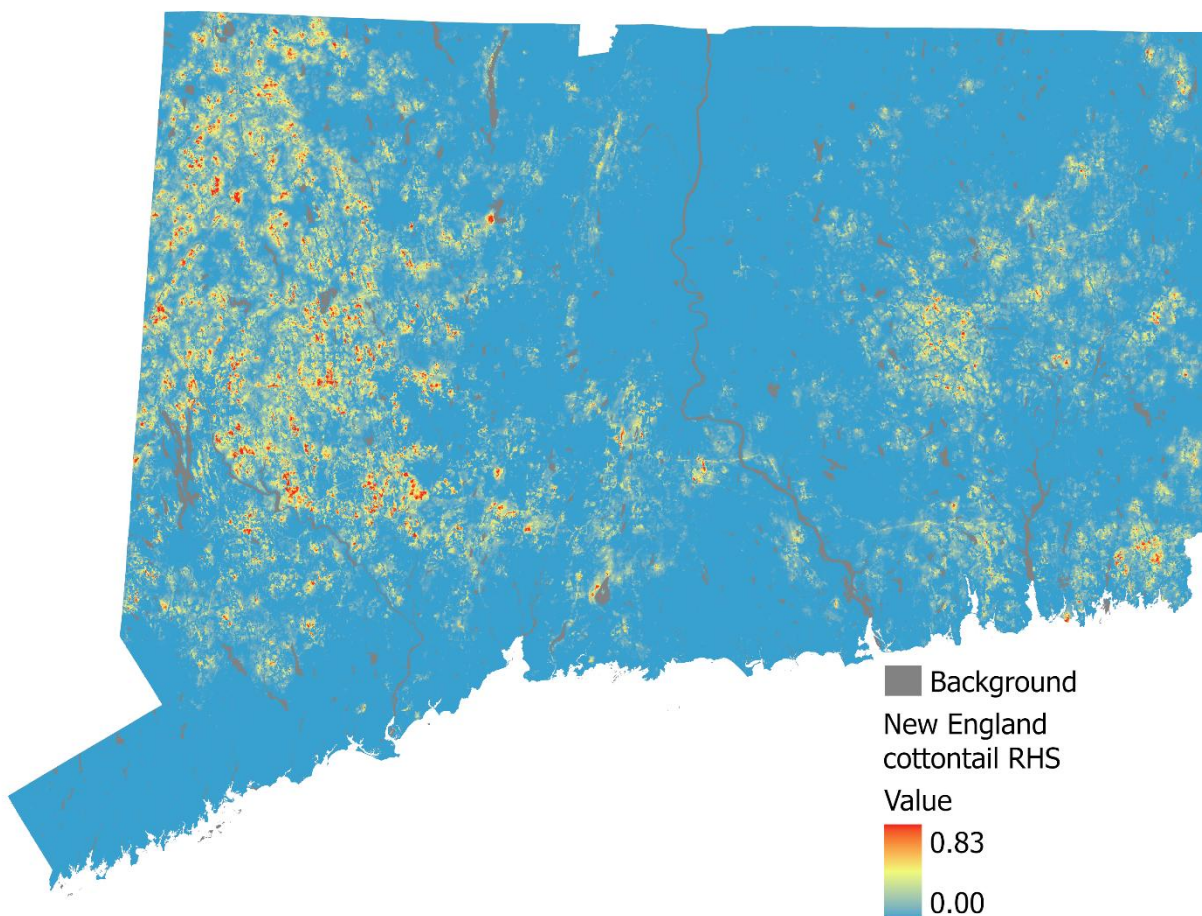


Figure 16. New England cottontail habitat suitability model where the warmer colors and higher values indicate higher relative habitat suitability (RHS) values.

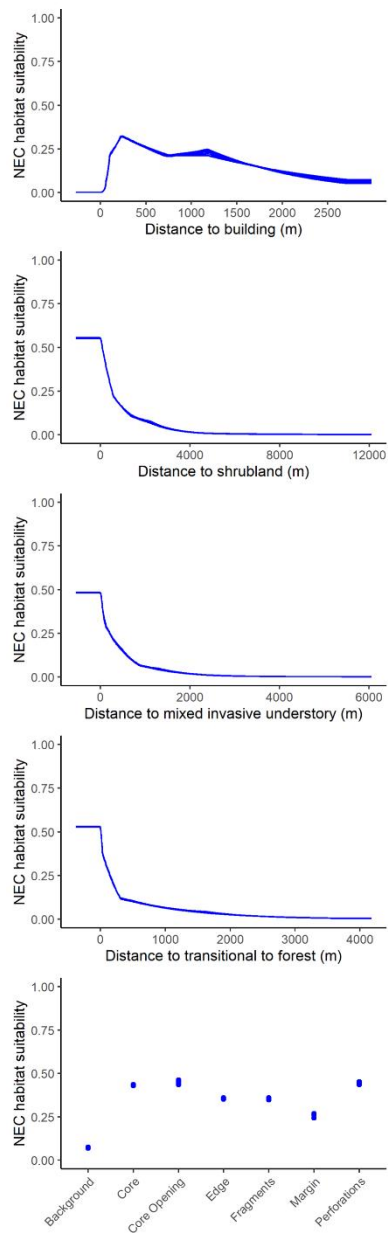


Figure 17. Response curves for the 5 top contributing predictors within the New England cottontail Maxent model, where the selected predictor was varied while all other predictors were held constant. Each dot or line represents all model runs (across all 10-fold cross validation runs). For the Morphological Spatial Pattern Analysis predictor, background indicates absence of young forest, shrubland, and understory vegetation classes. In the foreground, core is interior continuous vegetation classes, edge is the external perimeter of core, fragments are areas too small to contain core area, and margin is continuous vegetation classes that do not fit into any other category. Core openings are

background classes found within core areas and perforations are the perimeter of those core openings.

Tricolored bat (*Perimyotis subflavus*)

Species Data

Tricolored bat presence-absence data was retrieved from stationary acoustic monitoring conducted by CT DEEP, resulting in 38 presences and 93 absences collected from 2015 to 2024. After thinning the data by 1 km, the final dataset included 21 presences and 56 absences.

Predictors

Distance to buildings, young forest and shrubland, oak habitat, palustrine wetlands, and waterbodies, and roost suitability were included as predictors in the model. The building footprints layer was used as a proxy for development (Microsoft 2019). The Young Forest and Shrubland Vegetation Map was used to identify different types of young forest and shrubland vegetation within Connecticut (Rittenhouse et al. 2022). The map used ecological processes (succession, disturbance, regeneration, and hydrology), vegetation height, percent vegetation cover by height category, previous land cover type, and time since disturbance to classify different types of young forest and shrubland vegetation (Rittenhouse et al. 2022). The National Hydrography Dataset Plus (Buto and Anderson 2020) was used to determine distance to the nearest waterbody. The 1.0 m resolution Coastal Change Analysis Program land cover map was used to extract palustrine wetlands (combined emergent, scrub/shrub, and forested palustrine wetlands; NOAA 2020). The location of oak habitats was determined using the terrestrial habitat map at a 30.0 m resolution (The Nature Conservancy 2015). The summer roost suitability analysis was created by USGS for tricolored bat in North America at 250.0 m resolution (Inman et al. 2024) and this was resampled to a 10.0 m resolution to align with the spatial resolution of the model. Distance to oak habitat, waterbodies, and young forest and shrubland, were included as predictors in the model due to the association of suitable summer roost habitat and roost selection with oak trees, streams, and shrub cover (Cable and Willcox 2024, Williams et al. 2025). Distance to buildings was included since bat activity was higher in urban fragmented forests compared to rural fragmented forests (Johnson et al. 2008). Distance to palustrine wetlands was included since wetlands can be areas for foraging opportunities (USFWS 2024b). Presence of mixed forest, percent canopy cover, distance to barrens, deciduous forest, and bridges were also evaluated as predictors in the model, but these predictors reduced model performance.

Model Evaluation and Selection

The highest performing model was the RF model; however, the next highest performing model was chosen as the final model, the ensemble of all models tested (GAM, GBM, GLM, and RF) due to the habitat suitable map more accurately representing suitable summer roost habitat. The ensemble model still had a high AUC value of 0.93, TSS of 0.90, and Kappa of 0.81 (Table 27). The highest contributing predictors (variable contribution values ≥ 1.00) were distance to buildings (2.38), young forest and shrubland (1.08), palustrine wetlands (2.20), and waterbodies (1.75), while distance to oak habitat and roost suitability were the lowest contributing predictors (values < 1.00 ; Table 28). Tricolored bat suitable habitat increased with increasing distance from buildings and moderate distances

to palustrine wetlands (400-600 m), and decreased with increasing distance to waterbodies (Figure 18)

Tables and Figures

Table 27. Evaluation metrics output for each tricolored bat model, Generalized Additive Model (GAM), Gradient Boosted Machine (GBM), Generalized Linear Model (GLM), Random Forest (RF), and the Ensembled approach (EF). The grey highlighted row is the final model chosen for the mapping process.

Model	AUC	TSS	Kappa
GAM	0.83	0.60	0.57
GBM	0.90	0.80	0.65
GLM	0.60	0.33	0.43
RF	0.97	0.90	0.81
EF	0.93	0.90	0.81

Table 28. Variable contribution values for the predictors within the tricolored bat models Generalized Additive Model (GAM), Gradient Boosted Machine (GBM), Generalized Linear Model (GLM), Random Forest (RF), and the Ensembled approach (EF). Contribution values ≥ 1.00 indicate the contribution value for that predictor is higher than or equal to the average contribution value. The grey highlighted row is the final model chosen for the mapping process.

Predictor	GAM	GBM	GLM	RF	EF
Building distance	2.10	0.34	6.17	1.24	2.38
Young forest and shrubland distance	2.10	0.60	0.81	1.01	1.08
Oak habitat distance	0.00	1.15	0.41	0.28	0.49
Palustrine wetland distance	1.08	1.15	0.81	5.38	2.20
Waterbody distance	1.64	4.00	0.41	0.78	1.75
Roost suitability	0.42	0.60	1.20	0.35	0.64

2025 Connecticut Wildlife Action Plan

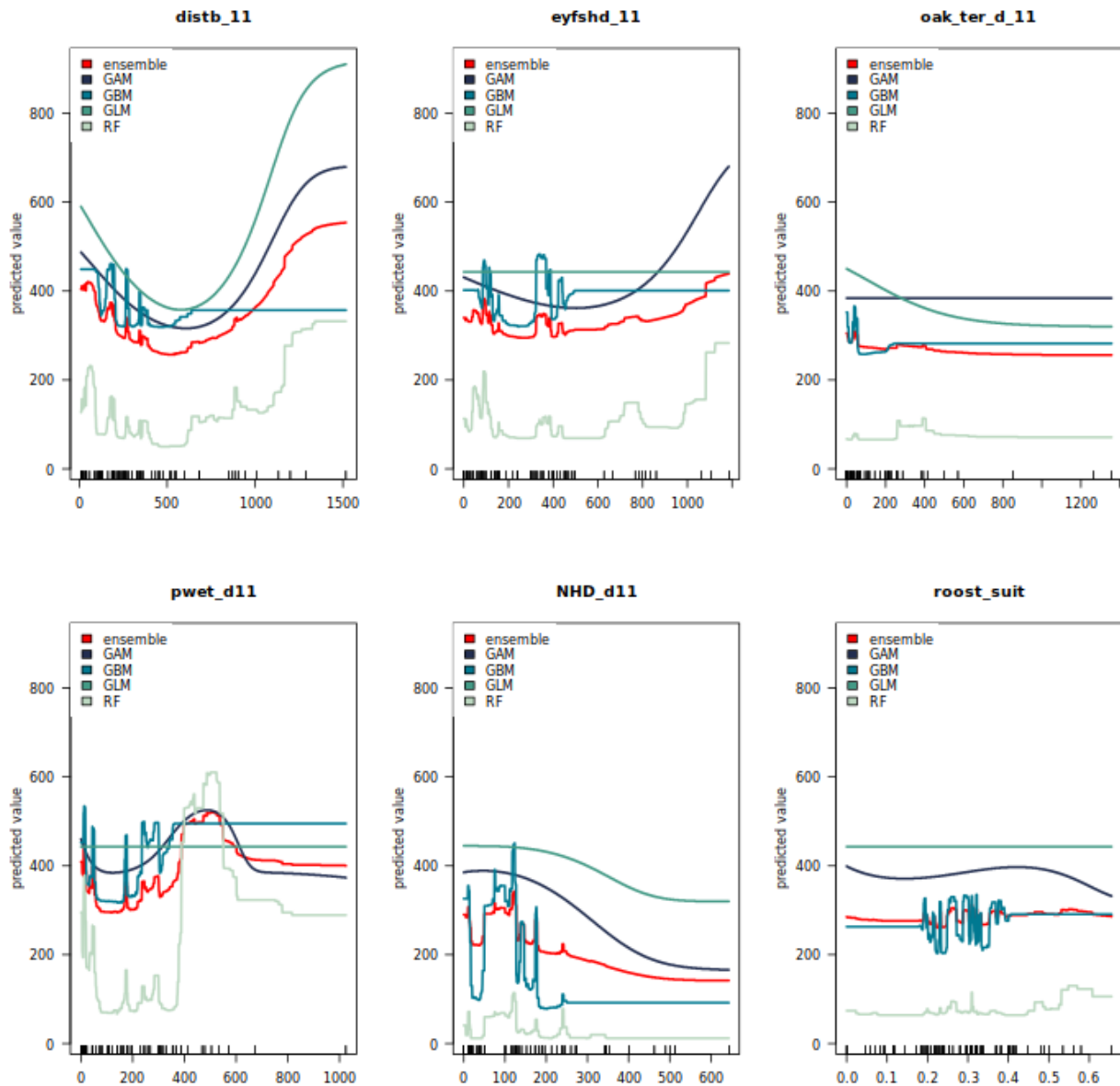


Figure 18. Response curves for the predictors in the tricolored bat model: distance to buildings (distb_11), distance to young forest and shrubland (eyfshd_11), distance to oak habitat (oak_ter_d_11), distance to palustrine wetlands (pwet_d11), distance to waterbodies (NHD_d11), and roost suitability score (roost_suit). All distance predictors are measured in meters. Each model evaluated represents a different colored line: Generalized Additive Model (GAM), Gradient Boosted Machine (GBM), Generalized Linear Model (GLM), Random Forest (RF), and the Ensembled approach (EF).

Key Habitat Maps

Key habitat maps were generated for the following habitat groups: open upland, forested upland, estuarine, palustrine, land water interface, lacustrine, and riverine, excluding development, subterranean, agriculture, marine, and unspecified. These key habitat maps were used to calculate the amount of habitat by town for Chapter 2. A combination of recent fine resolution landcover and vector data were used to generate the maps. For forested upland, the NOAA 1m Coastal Change Analysis Program Landcover map (NOAA 2020) was used to extract mixed forest and the Young Forest and Shrubland Vegetation Map (Rittenhouse et al. 2022) was used to extract recently disturbed forest, regenerating forest and clearcut, which are young forest classes that were forested 0-20 years ago. For open uplands, a combination of grasslands, scrub/shrub, and bare land from NOAA 1m Coastal Change Analysis Program Landcover map (NOAA 2020), and transitional to forest, reverting fields, and shrublands from the Young Forest and Shrubland Vegetation Map (Rittenhouse et al. 2022) was used. For palustrine, the NOAA 1.0 m Coastal Change Analysis Program Landcover map (NOAA 2020) was used to extract tidal and non-tidal palustrine emergent, scrub/shrub, and forested wetlands. For riverine, a combination of rivers, streams and brooks from the Connecticut Hydrography Dataset (USGS 2005) and the National Hydrography Dataset Plus (Buto and Anderson 2020) code 1 (rivers) water bodies were used. Both hydrography datasets were used to capture larger rivers as well as the small, ephemeral streams. For lacustrine, the CT hydrography dataset (USGS 2005) was also used to map lakes, ponds, reservoirs, and open water within the state. The CT hydrography dataset was chosen over the higher resolution National Hydrography Dataset Plus for lacustrine since it more accurately differentiated lakes and rivers. For land water interface, a combination of the NOAA 1.0 m Coastal Change Analysis Program Landcover map (NOAA 2020) unconsolidated shores, bare land within 100.0 m of the coastline (to capture beaches only), and delineated islands from the Northeast States Town Boundary Set (CT DEEP 2019) was used. For estuarine, the NOAA 1.0 m Coastal Change Analysis Program Landcover map was used to extract estuarine emergent, scrub/shrub, and aquatic vegetation (NOAA 2020).

Additional Ecological Data

Protected areas were compiled from CT protected open space data (CT DEEP 2011b), federal open space data (CT ECO 2019), National Protected Areas Database (USGS 2018), CT Aquifer Protection Areas (CT DEEP 2024b), CT DEEP property (CT DEEP 2024c), CT Fisheries Management areas (CT DEEP 2023a), and CT Surface Water classes with high water quality (class AA; CT DEEP 2023b). All the protected areas data were merged into one shapefile and then converted into a raster with a 10.0 m resolution. Distance to protected areas was created using the distance accumulation function in ArcPro 3.4.2 (ESRI 2025).

CT DEEP's Natural Diversity Database (NDDB) Review Areas (CT DEEP 2024d) are created for all state listed species and species of special concern and include potential suitable habitat and dispersal buffers. These maps are not specific to SGCN, where they are not maps of SGCN only, but are more specific areas of state listed endangered, threatened, and species of special concern. The NDDB Review Areas map was categorized into the low (1-2 species), medium (3-6 species), and high (> 7 species) number of overlapping layers for state threatened and endangered species, and species of special concern, based on the natural breaks in the data. The NDDB areas for SGCN already included in the SGCN mapping process were removed to avoid overrepresenting species.

The Nature Conservancy's Resilient and Connected Network was retrieved to include regional connectivity in the Connect COA (The Nature Conservancy 2016). All pixels in the Resilient and Connected Network in Connecticut were included and rasterized to a 10.0 m resolution raster.

Town Plan Analysis

Topic models were used to identify topics, groups of co-occurring words that represent an underlying theme, in town plans that overlap with the 2015 CT SWAP (Figure 3, Step 2). Town Plans of Conservation and Development (POCD) were retrieved from online resources. The 107 plans published since the 2015 CT SWAP revision were retrieved from online. Each town needs to complete a POCD to receive discretionary state funding, and the plans must reflect their vision of the future state of development of the town while balancing natural resource, residents' values, economic opportunities, transportation, and infrastructure (CT General Statutes 2018).

Plans were preprocessed, so the plans only included relevant information. The first preprocessing step was to remove entire pages that contained irrelevant or repetitive information, to avoid inclusion of irrelevant themes or artificially increasing the prevalence of certain themes. Specifically, pages containing solely title pages, tables of contents, executive summaries, literature cited, intentionally blank pages, welcome letters from town officials (unless containing a vision statement), background and overview of what a POCD is, description and acknowledgement of board members, statutory requirements, description of implementation tables, comparison to other plans, compliance with regional guidelines, reports from other agencies regarding the municipality, and public survey results were removed.

Once irrelevant pages were removed, the documents were loaded into R 4.2.3 (R Core Team 2023) to create the corpus, remove irrelevant words, and run the topic models. A corpus is a group of documents, where each document contains a distribution of words (Blei 2003). Words were identified through the process of tokenization, where sentences are broken down into individual words, also called tokens (Banks et al. 2018). Topic models identify the topics or groups of similar tokens that represent a semantic theme within a corpus (Roberts et al. 2014). To identify the major topics within town plans, structural topic models (STM) were run. STM were chosen because they are similar to Latent Dirichlet Allocation (LDA) where they both view documents as random and fixed

mixtures of different topics or words but STM improves fit compared to LDA by incorporating covariates and removing the assumptions that topic prevalence is consistent amongst documents by incorporating a topic prevalence parameter (Blei 2003, Roberts et al. 2014). To prepare a corpus for topic modeling, common stop words such as “is” or “the,” punctuation, numbers (Schofield et al. 2017, Weston et al. 2023), and words with < 4 characters were removed. Words that describe the location of the plan (including town names) and words that refer to the structure of the plans such as chapter, page, key, and contents were also excluded.

Within the preprocessing phase there are a series of decision points that can alter the outcome of the topic models (Schofield and Mimno 2016), where best practices for selection are not always transparent (Banks et al. 2018) and can vary by researcher and the nature of the corpus (e.g., Roberts et al. 2014, 2019, Schofield and Mimno 2016, Müller-Hansen et al. 2020, Weston et al. 2023). For each of these decision points, a multiverse analysis was conducted (Steege et al. 2016) or a comparison of all possible data iterations based on the literature to determine how preprocessing choices influence model outcomes for each corpus.

The first decision point was determining whether to establish n-grams, or multi-word expressions, where n is the number of words within the expression (e.g., “open space”). N-grams can be established a priori or can be determined based on frequency of word co-occurrences. Establishing n-grams can aid analysis (Banks et al. 2018), but there is little guidance on the co-occurrence frequency thresholds to consider n-grams. Thus, a range of frequency thresholds to test within the multiverse analysis was chosen. A two-step process to determine n-gram thresholds was used, first running n-grams with a relatively low co-occurrence frequency for the corpus (less than 1% of the total number of tokens) and then used histograms to determine thresholds that had relatively large number of n-gram occurrences. The co-occurrence frequency decayed as the threshold decreased, thus the frequency was determined with the highest number n-grams and then iteratively halved the number until the n-grams being established were not meaningful. The final co-occurrence frequencies tested included 500, 250, and 125.

Another decision point in the preprocessing phase was the inclusion of rare tokens. Rare tokens were removed because if rare words are kept, they can dominate topics (Banks et al. 2018), and the goal of the study was to determine common topics shared amongst most plans. Rare tokens can be removed based on token frequency within the entire corpus (called minimum token frequency) or based on the number of documents the token occurs in (called minimum document frequency). Previous studies have used both methods, thus, values common within literature were chosen. The multiverse analysis was run with a minimum token frequency of 20 occurrences (Gan and Qi 2021, Weston et al. 2023), a sparsity coefficient, or a proportion approach for minimum token frequency, of 0.995 (Maurer et al. 2024), a minimum document frequency of the token occurring in at least 5% of the documents (5 documents for the town plans; Banks et al. 2018), and no minimum token or document frequency.

Once the corpus was preprocessed, the next important decision point within topic modeling was the number of topics. The researchers choose the number of topics. There is no right choice with topic number, too few topics can hide rare or unique topics, and too

many topics can bias towards rare words, thus topic number choice should balance the interests of the researchers and the performance of the models (Weston et al. 2023). First, models were run with a range of topic numbers, from 7 to 100 topics, for each n-gram and token frequency combination and then multiple diagnostic metrics and multiple evaluation tests were used to determine the ideal number of topics for each topic model. Four diagnostic metrics and three evaluation tests were used to compare the diagnostic metrics for each model in the multiverse analysis. Sematic coherence (S), how often common words in a topic co-occur, which often corresponds with human judgment (Mimno et al. 2011), held-out likelihood (H), which is the probability the model can generate held-out data accurately (Chang et al. 2009), residuals (R), which is the dispersion of the residuals (Taddy 2012), and lower bounds (L), which is the convergence of the model, were all used as diagnostic metrics (Weston et al. 2023). For the first evaluation test, the plots were assessed visually and identified where the addition of 1 topic had a strong influence on the diagnostic tests, as recommended by Weston et al. (2023). Second, the approach by Weston et al. (2023) was quantitatively applied. First, all outputs were scaled prior to using the equation, so the values vary from 0-1. Since S and H should be maximized, the scale was left as 0-1, however, an inverse scale was used for R and L since these values should be minimized (Roberts et al. 2019). The difference was summed in each of the scaled values of the model with number of topics (k) with the sum of the model with $k-1$ topics. Sums above 0 indicate the addition of a topic was influential compared to the previous number of topics. Thirdly, the difference was measured between the sum of k compared to the next lowest sum value and then identified the topic numbers with the highest differences in sum.

All ideal topic numbers determined by the evaluation tests were run. Besides varying n-grams and word or document frequency for the multiverse analysis, all model settings were kept consistent for each model. Spectral initialization or a technique using matrix decomposition based on a moments framework rather than a likelihood framework was used to initialize the model because it has been shown to perform best (Roberts et al. 2016, 2019). All default settings were used within the model since they are tuned for general use, except, the default prior was changed to a gamma prior to allow for regularization (Roberts et al. 2019), since there was a high number of covariates.

Covariates were included in topic modeling to incorporate the effects of the structure of the plan and motivations for including ecological themes in town plans. There are two types of covariates within STM, those that influence topic prevalence or the prevalence of a topic within a document and topic content or how much word use within a topic varies (Roberts et al. 2014, 2019). Only topical prevalence covariates were included within the models to reflect the goal of this study. Document level covariates were included that reflected attributes of the document itself, number of pages and year published. Covariates describing attributes of the town were also included to reflect why towns would focus on ecological themes within their plans. Covariates included annual income tax by town for 2020 (CT Open Data 2022), used as a proxy for town budget, the proportion of the town containing natural resources, forests (National Oceanic and Atmospheric Administration (NOAA) 2020) and water resources (US Geological Survey (USGS) 2005), the proportion of the town already protecting natural resources (CT DEEP 2011b, USGS 2018,

CT ECO 2019, CT DEEP 2024c), whether the town was considered a distressed community within the last 5 years (CT Department of Economic and Community Development 2023) and whether the town was on the coast. The regional council of government the town belongs to was also included since the towns are encouraged to be consistent with regional government's POCD (CT General Statutes 2018). Covariates were standardized to estimate and compare covariate effect sizes.

The similarity in topic composition or solution congruence was identified between all possible ideal model combinations (Weston et al. 2023) by computing Jaccard Similarity values for the 50 most common words within each topic across models. The similarity values were used rather than the beta values (per topic word probabilities) because beta values cannot be compared across models. The multiverse analysis determined how topic number, n-grams, and minimum word or document frequency influence the similarity amongst topics.

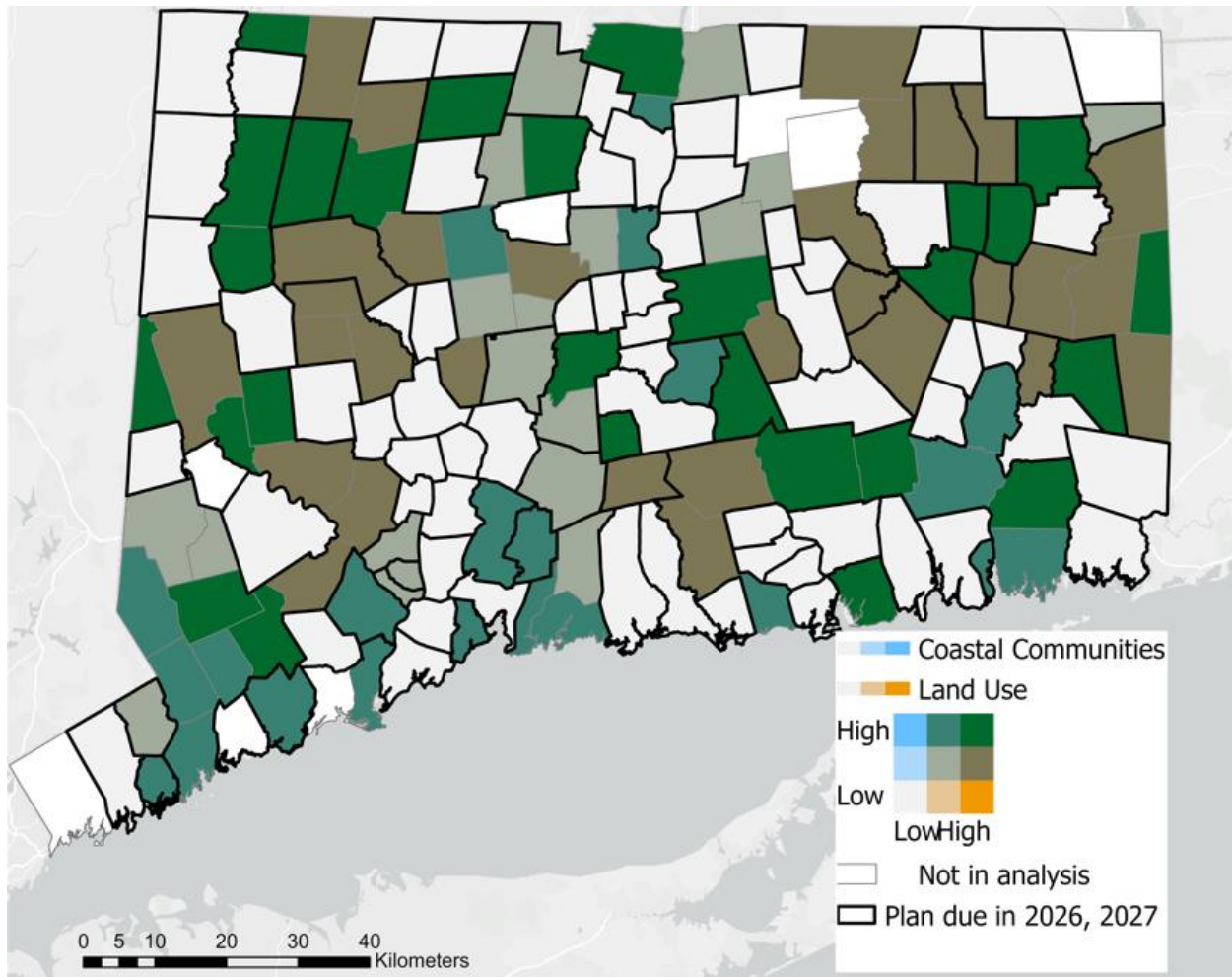


Figure 19. The results of the town plan analysis where colors correspond to the probability that coastal communities and land use are included within each town's plan of conservation and development (POCD). These topics were chosen because they had the most overlap the topic in the 2015 CT SWAP. Darker colors indicate a higher probability that one or both topics are found within the town plans. The towns in white were not included within the analysis and the towns with a black outline have a revision due within the next few years.

Once robust modeling criteria is determined, or models with different modeling parameters but similar topic compositions, the FREX and probability labels (top 10 words) for models with the highest similarity values were qualitatively analyzed. FREX is a composite score based on word frequency and exclusivity (Airoldi and Bischof 2016). A two-observer model was used, where both authors interpreted the overall themes of each topic based on the labels and then compared the interpretations. After independently labeling topics, the interpretations were discussed, and amendments were made to the initial interpretations based on the discussions. Example texts were used to aid in topic model selection (Müller-Hansen et al. 2020), especially for topics that were semantically difficult to interpret or had different interpretations amongst the observers. The final candidate models were chosen based on the amount of similarity between the observers' interpretation of topics and the ease of semantically inferring themes within topics.

To create data to input within the COAs, the topic probabilities for the town topics that had the highest similarity to the 2015 CT SWAP, land use and coastal communities, were

extracted for each town's plan. The values within the map represent the sum of the probabilities of the two topics (Figure 19).

Social Data

The social data included within the zonation process reflected proximity to engagement and education resources, partner priority maps, and population density. The locations of libraries were retrieved from the Library Technology Guides (Breeding 2024), which provides an address for each public library. This list was cross-referenced with each library's website to ensure the library was open and the address was correct. The addresses of non-state owned education and nature centers was retrieved from a tourism website (Visit New England 2025) and combined with the list of state-owned education and nature centers (CT DEEP 2024a). The addresses for libraries, education, and nature centers were converted to point locations using the geocode addresses function in ArcPro 3.4.2 (ESRI 2025) and then the distance to each location was calculated for each pixel within the 10.0 m grid of CT. The CT DEEP state property access point data (CT DEEP 2018b) was retrieved from online and a raster of distance to each location was created using a 10.0 m resolution. Town population density data was retrieved online (US Census Bureau 2023) and then the population was divided by the area of the town. Population density associated with each town was converted to a 10.0 m resolution raster.

All known priority area maps from federal, state, town, NGO, and private landowner partners were retrieved from online. The National Audubon Society Important Areas shapefile (National Audubon Society 2024) was clipped to CT and then converted into a raster with a 10.0 m resolution. The USFWS Highland Conservation Area Boundary (USFWS 2024c) was clipped to CT and then converted to a 10.0 m resolution raster. The core forest map was retrieved from Center for Land Use Education and Research (CLEAR) online (CLEAR 2016), which are the priority areas for the 2020 CT Forest Action Plan (CT DEEP 2020b), and then resampled the raster to a 10.0 m resolution. The map from Smith et al. (2024) that demonstrates zip codes with opportunities for private landowners engaging in forest management was also included. The opportunity score from this map was converted into a raster with a 10.0 m resolution. The list of Trout Unlimited priority waterbodies (Trout Unlimited 2023) was extracted from the named waterbodies layer (USGS 2005) and then converted into a raster with a 10.0 m resolution.

Identifying COAs

The next step of mapping conservation opportunity areas was to input the ecological, town, and social data into the Zonation software (Figure 1, Step 3 and 4). Zonation is a spatial prioritization software that ranks all grid cells in a study area based on their conservation value, which is calculated from multiple grid (raster) input layers given to the program (Moilanen et al. 2022). Zonation uses a priority rank algorithm which iteratively ranks all cells in the study area based on their marginal loss in conservation value, and continues to reorder cells based on the previous ranking until the model

converges (Moilanen et al. 2022). The rank score (varies from 0 to 1, with 0 representing the lowest and 1 representing the highest conservation priority) is given to each cell (Moilanen et al. 2005, 2022). Zonation prioritizes areas with more conservation coverage, high local density of features, balance between features, and aims to minimize conservation loss (Moilanen et al. 2022). Grid cells with high density of features, such as SGCN suitable habitat, key habitats, and other features (such as town and social data), are given the highest priority.

To associate conservation opportunities with broad actions, single-objective conservation opportunity areas were created based on the 7 actions in the PCRM-PI framework: Protect, Connect, Restore, Manage, Partner, Inform (divided into two separate action of Inform and Research and Monitor; Liberati et al. 2016), which was cross-walked to the Conservation Measures Partnership Level 2 actions. Each COA has different ecological, town, and/or social data included to reflect the unique aspects of each action. The weighting for each input data was based on the importance of the data. The weighting itself is arbitrary but does have meaning relative to the other layers included (Moilanen et al. 2022). For most layers, the weighting was kept equal and positive at 1, but specific layers were increased or made negative based on the interpretation of their values. To prioritize locations near protected areas, distance to protected areas was given a higher negative weight of -5.0 and to prioritize regional connectivity, the Resilient and Connected Network was given a higher positive weight of 5.0 in the Connect COA. The NNDB Review Areas layer was also given a higher weight of 21.0 to equal the combined weight of all the habitat suitability and key habitat maps in the Research and Monitor COA. To transform the continuous rank outputs from Zonation to binary COAs, rank values above 0.80 or the top 20% were extracted (Figure 1, Step 5; Moilanen et al. 2005).

Literature Cited

- Aiello-Lammens, M. E., R. A. Boria, A. Radosavljevic, B. Vilela, and R. P. Anderson. 2015. spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography* 38:541–545.
- Airoldi, E. M., and J. M. Bischof. 2016. Improving and Evaluating Topic Models and Other Models of Text. *Journal of the American Statistical Association* 111:1381–1403.
- Albanese, G., P. D. Vickery, and P. R. Sievert. 2008. Microhabitat use by larvae and females of a rare barrens butterfly, frosted elfin (*Callophrys irus*). *Journal of Insect Conservation* 12:603–615.
- Allouche, O., A. Tsoar, and R. Kadmon. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology* 43:1223–1232.
- Amaral, K. E., M. Palace, K. M. O'Brien, L. E. Fenderson, and A. I. Kovach. 2016. Anthropogenic habitats facilitate dispersal of an early successional obligate: implications for restoration of an endangered ecosystem. *PLOS ONE* 11:e0148842.
- Banks, G. C., H. M. Woznyj, R. S. Wesslen, and R. L. Ross. 2018. A Review of Best Practice Recommendations for Text Analysis in R (and a User-Friendly App). *Journal of Business and Psychology* 33:445–459.
- Barbet-Massin, M., F. Jiguet, C. H. Albert, and W. Thuiller. 2012. Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution* 3:327–338.
- Barbour, M. S., and J. A. Litvaitis. 1993. Niche dimensions of New England Cottontails in relation to habitat patch size. *Oecologia* 95:321–327.
- Beaudette, D., J. Skovlin, S. Roecker, and A. Brown. 2025. soilDB: Soil Database Interface. <<https://cran.r-project.org/web/packages/soilDB/index.html>>. Accessed 16 Mar 2025.
- Beaudry, F., P. G. deMaynadier, and M. L. Hunter. 2008. Identifying road mortality threat at multiple spatial scales for semi-aquatic turtles. *Biological Conservation* 141:2550–2563.
- Bell, S. L. M., T. B. Herman, and R. J. Wassersug. 2007. Ecology of *Thamnophis sauritus* (Eastern Ribbon Snake) at the Northern Limit of its Range. *Northeastern Naturalist* 14:279–292.
- Bischoff, K. E., D. Katz, C. D. Rittenhouse, and T. A. G. Rittenhouse. 2025. High Overlap in Niches and Suitable Habitat Between an Imperiled and Introduced Cottontail. *Ecology and Evolution* 15:e71083.
- Bischoff, K. E., T. A. G. Rittenhouse, and C. D. Rittenhouse. 2023a. Landscape context and spatial attributes matter for New England cottontail occupancy. *Journal of Wildlife Management* 87:e22417.
- Bischoff, K. E., T. A. G. Rittenhouse, and C. D. Rittenhouse. 2023b. Introduced competitor reduces abundance of an imperiled cottontail. *Biological Invasions* 25:3553–3566.
- Blei, D. M. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning* 993–1022.
- Boyce, M. S., P. R. Vernier, S. E. Nielsen, and F. K. A. Schmiegelow. 2002. Evaluating resource selection functions. *Ecological Modelling* 157:281–300.

- Breeding, M. 2024. Directory of Public Libraries in Connecticut. <<https://librarytechnology.org/libraries/public.pl?State=Connecticut>>. Accessed 15 Mar 2025.
- Breiman, L. 2001a. Statistical Modeling: The Two Cultures. *Statistical Science* 16:199–215.
- Breiman, L. 2001b. Random Forests. *Machine Learning* 45:5–32.
- Breiman, L., J. Friedman, R. A. Olshen, and C. J. Stone. 1984. *Classification and Regression Trees*. Chapman and Hall/CRC, New York.
- Breiner, F. T., A. Guisan, A. Bergamini, and M. P. Nobis. 2015. Overcoming limitations of modelling rare species by using ensembles of small models. *Methods in Ecology and Evolution* 6:1210–1218.
- Breiner, F. T., M. P. Nobis, A. Bergamini, and A. Guisan. 2018. Optimizing ensembles of small models for predicting the distribution of species with few occurrences. *Methods in Ecology and Evolution* 9:802–808.
- Broennimann, O., V. D. Cola, B. Petitpierre, F. Breiner, D. Scherrer, M. D`Amen, C. Randin, R. Engler, W. Hordijk, H. Mod, J. Pottier, M. D. Febbraro, L. Pellissier, D. Pio, R. G. Mateo, A. Dubuis, L. Maiorano, A. Psomas, C. Ndiribe, N. Salamin, N. Zimmermann, F. Collart, V. Verdon, T. Smith, and A. Guisan. 2025. *ecospat: Spatial Ecology Miscellaneous Methods*. <<https://cran.r-project.org/web/packages/ecospat/index.html>>. Accessed 16 Mar 2025.
- Brubaker, D. R., A. I. Kovach, M. J. Ducey, W. J. Jakubas, and K. M. O'brien. 2014. Factors influencing detection in occupancy surveys of a threatened lagomorph. *Wildlife Society Bulletin* 38:513–523.
- Buchanan, S. W., B. C. Timm, R. P. Cook, R. Couse, and L. C. Hazard. 2017. Spatial Ecology and Habitat Selection of Eastern Hognose Snakes. *The Journal of Wildlife Management* 81:509–520.
- Buto, S. G., and R. Anderson. 2020. NHDPlus High Resolution (NHDPlus HR) A hydrography framework for the Nation. Fact Sheet. U.S. Geological Survey.
- Cable, A. B., and E. V. Willcox. 2024. Summer habitat for the female Tricolored Bat (*Perimyotis subflavus*) in Tennessee, United States. *Journal of Mammalogy* 105:667–678.
- Canada Environment and Climate Change. 2022. Yellow-banded Bumble Bee (*Bombus terricola*): management plan proposed 2022. <<https://www.canada.ca/en/environment-climate-change/services/species-risk-public-registry/management-plans/yellow-banded-bumble-bee-proposed-2022.html>>. Accessed 23 Jul 2024.
- Chang, J., S. Gerrish, C. Wang, J. Boyd-graber, and D. Blei. 2009. Reading Tea Leaves: How Humans Interpret Topic Models. *Advances in Neural Information Processing Systems Proceedings*. Volume 22. Curran Associates, Inc.
- Chen, T., and C. Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. Pages 785–794 in. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16, Association for Computing Machinery, New York, NY, USA.
- Cheeseman, A. E., J. B. Cohen, S. J. Ryan, and C. M. Whipps. 2019. Determinants of home-range size of imperiled New England Cottontails (*Sylvilagus transitionalis*) and

- introduced Eastern Cottontails (*Sylvilagus floridanus*). *Canadian Journal of Zoology* 97:516–523.
- Cheeseman, A. E., J. B. Cohen, S. J. Ryan, and C. M. Whipps. 2021. Is conservation based on best available science creating an ecological trap for an imperiled lagomorph? *Ecology and Evolution* 11:912–930.
- Cheeseman, A. E., S. J. Ryan, C. M. Whipps, and J. B. Cohen. 2018. Competition alters seasonal resource selection and promotes use of invasive shrubs by an imperiled native cottontail. *Ecology and Evolution* 8:11122–11133.
- Center for Land Use Education and Research (CLEAR). 2016. CT Forest Fragmentation. <<https://clear.uconn.edu/projects/landscape/ct-forestfrag/>>. Accessed 15 Mar 2025.
- Cohen, J. 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* 20:37–46.
- Corbin, J. D., and E. L. Flatland. 2022. Mapping edaphic soils’ conditions to identify conservation targets for pine barren and sandplain ecosystems in New York State. *Ecology and Evolution* 12:e9282.
- CT Department of Economic and Community Development. 2023. Distressed Municipalities. CT.gov - Connecticut’s Official State Website. <https://portal.ct.gov/decd/content/about_decd/research-and-publications/02_review_publications/distressed-municipalities>. Accessed 1 Jun 2023.
- CT DEEP. 2011a. Wood Turtle. Endangered and Threatened Species Fact Sheet Series. <<https://portal.ct.gov/deep/wildlife/fact-sheets/wood-turtle>>. Accessed 17 Mar 2025.
- CT DEEP. 2011b. 2011 Protected Open Space Mapping Set. CT DEEP GIS Open Data Website. <<https://ct-deep-gis-open-data-website-ctdeep.hub.arcgis.com/search?q=water>>. Accessed 17 Mar 2025.
- CT DEEP. 2018a. Common Ribbonsnake. <<https://portal.ct.gov/deep/wildlife/fact-sheets/common-ribbonsnake#:~:text=Habitat%20and%20Diet%3A%20The%20ribbonsnake,sometimes%20near%20trap%20rock%20systems.>>>. Accessed 23 Jul 2024.
- CT DEEP. 2018b. DEEP Property Access Locations. CT DEEP GIS Open Data Website. <https://ct-deep-gis-open-data-website-ctdeep.hub.arcgis.com/datasets/3df3c20f4f2d415a8d728525c1d6d3f6_0/explore>. Accessed 15 Mar 2025.
- CT DEEP. 2019. Northeastern States Town Boundary Set. CT DEEP GIS Open Data Website. <<https://ct-deep-gis-open-data-website-ctdeep.hub.arcgis.com/search?q=water>>. Accessed 17 Mar 2025.
- CT DEEP. 2020a. Cold-water Sites Set. CT DEEP GIS Open Data Website. <<https://ct-deep-gis-open-data-website-ctdeep.hub.arcgis.com/search?q=water>>. Accessed 16 Mar 2025.
- CT DEEP. 2020b. CT Forest Action Plan. Connecticut’s Official State Website. <<https://portal.ct.gov/DEEP/Forestry/CT-Forest-Action-Plan>>. Accessed 15 Mar 2025.

- CT DEEP. 2023a. Fisheries Management Areas Set. CT DEEP GIS Open Data Website. <<https://deepmaps.ct.gov/maps/26da452f8ccd45319f3bd7b2e5e1b448/about>>. Accessed 17 Mar 2025.
- CT DEEP. 2023b. Surface Water Quality Classification Set. CT DEEP GIS Open Data Website. <<https://deepmaps.ct.gov/search?q=water>>. Accessed 17 Mar 2025.
- CT DEEP. 2024a. Directions to DEEP Facilities. Connecticut’s Official State Website. <<https://portal.ct.gov/deep/about/contact-us/directions-to-deep-facilities>>. Accessed 15 Mar 2025.
- CT DEEP. 2024b. Aquifer Protection Areas Set. CT DEEP GIS Open Data Website. <<https://ct-deep-gis-open-data-website-ctdeep.hub.arcgis.com/search?q=water>>. Accessed 17 Mar 2025.
- CT DEEP. 2024c. DEEP Property. CT DEEP GIS Open Data Website. <<https://deepmaps.ct.gov/datasets/deep-property-3/explore>>. Accessed 17 Mar 2025.
- CT DEEP. 2024d. Natural Diversity Database. CT DEEP GIS Open Data Website. <https://deepmaps.ct.gov/datasets/58c6d263e85647caaa82c466d5794227_0/explore>. Accessed 17 Mar 2025.
- CT DEEP, and CT Department of Agriculture. 2019. Shellfish Bed. CT DEEP GIS Open Data Website. <<https://ct-deep-gis-open-data-website-ctdeep.hub.arcgis.com/search?q=water>>. Accessed 16 Mar 2025.
- CT ECO. 2019. Federal Open Space. <<https://ct-deep-gis-open-data-website-ctdeep.hub.arcgis.com/datasets/CTDEEP::federal-open-space/about>>. Accessed 17 Mar 2025.
- CT General Statutes. 2005. Sec. 22a-94. Chapter 444. Coastal area; coastal boundary. <https://www.cga.ct.gov/2021/pub/chap_444.htm>. Accessed 16 Mar 2025.
- CT General Statutes. 2018. Sec. 8-23. Chapter 126: Municipal Planning Commissions. <https://www.cga.ct.gov/2023/pub/chap_126.htm>. Accessed 16 Mar 2025.
- CT Open Data. 2022. Personal Income Tax By Town, 2011-2020 | Connecticut Data. <https://data.ct.gov/Tax-and-Revenue/Personal-Income-Tax-By-Town-2011-2020/pvqv-e235/about_data>. Accessed 29 Aug 2024.
- DeLand, S. 2025. Connecticut Dams-Public Use. CT DEEP GIS Open Data Website. <<https://ct-deep-gis-open-data-website-ctdeep.hub.arcgis.com/datasets/CTDEEP::connecticut-dams-public-use/about>>. Accessed 16 Mar 2025.
- Di Cola, V., O. Broennimann, B. Petitpierre, F. T. Breiner, M. D’Amen, C. Randin, R. Engler, J. Pottier, D. Pio, A. Dubuis, L. Pellissier, R. G. Mateo, W. Hordijk, N. Salamin, and A. Guisan. 2017. ecospat: an R package to support spatial analyses and modeling of species niches and distributions. *Ecography* 40:774–787.
- Dormann, C. F., J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carré, J. R. G. Marquéz, B. Gruber, B. Lafourcade, P. J. Leitão, T. Münkemüller, C. McClean, P. E. Osborne, B. Reineking, B. Schröder, A. K. Skidmore, D. Zurell, and S. Lautenbach. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36:27–46.

- Elith, J., J. R. Leathwick, and T. Hastie. 2008. A working guide to boosted regression trees. *Journal of Animal Ecology* 77:802–813.
- Elith, J., S. J. Phillips, T. Hastie, M. Dudík, Y. E. Chee, and C. J. Yates. 2011. A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions* 17:43–57.
- ESRI. 2025. ArcGIS Pro 3.4 Patch 2 (3.4.2) Announcement. ESRI. <<https://support.esri.com/en-us/patches-updates/2025/arcgis-pro-3-4-patch-2-3-4-2-announcement>>. Accessed 17 Mar 2025.
- Friedman, J. H. 1991. Multivariate Adaptive Regression Splines. *The Annals of Statistics* 19:1–67.
- Gan, J., and Y. Qi. 2021. Selection of the Optimal Number of Topics for LDA Topic Model—Taking Patent Policy Analysis as an Example. *Entropy* 23:1301.
- Gliesch, L., B. de A. Linhares, C. P. Ozorio, P. H. Ott, J. Jacoby, L. Bugoni, and G. T. Nunes. 2023. Specialist shorebird respond to prey and habitat availability through trophic plasticity. *Food Webs* 36:e00300.
- Guillera-Aroita, G., J. J. Lahoz-Monfort, and J. Elith. 2014. Maxent is not a presence–absence method: a comment on Thibaud et al. *Methods in Ecology and Evolution* 5:1192–1197.
- Guisan, A., W. Thuiller, and N. E. Zimmermann. 2017. *Habitat suitability and distribution models: With applications in R*. Cambridge University Press.
- Guo, K., S. Yuan, H. Wang, J. Zhong, Y. Wu, W. Chen, C. Hu, and Q. Chang. 2021. Species distribution models for predicting the habitat suitability of Chinese fire-bellied newt *Cynops orientalis* under climate change. *Ecology and Evolution* 11:10147–10154.
- Hanley, J. A., and B. J. McNeil. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143:29–36.
- Hastie, T., and R. Tibshirani. 1986. Generalized Additive Models. *Statistical Science* 1:297–310.
- Hastie, T., R. Tibshirani, and A. Buja. 1994. Flexible Discriminant Analysis by Optimal Scoring. *Journal of the American Statistical Association* 89:1255–1270.
- Hellton, K. H., J. Cummings, A. O. Vik-Mo, J. E. Nordrehaug, D. Aarsland, G. Selbaek, and L. M. Giil. 2021. The Truth behind the Zeros: A New Approach to Principal Component Analysis of the Neuropsychiatric Inventory. *Multivariate behavioral research* 56:70–85.
- Hirzel, A. H., G. Le Lay, V. Helfer, C. Randin, and A. Guisan. 2006. Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling* 199:142–152.
- Hosmer, D. W. Jr., S. Lemeshow, and R. Sturdivant. 2013. Assessing the Fit of the Model. Pages 153–225 in. *Applied Logistic Regression*. John Wiley & Sons.
- Howard, T. G., and M. D. Schlesinger. 2012. *PATHWAYS Wildlife Habitat Connectivity in the Changing Climate of the Hudson Valley*. New York Natural Heritage Program. <<https://www.nynhp.org/projects/pathways/>>. Accessed 17 Mar 2025.
- Inman, R. D., A. N. Schuhmann, S. C. Sawyer, S. M. Gaulke, F. Tousley, H. T. Davis, B. J. Udell, B. R. Straw, J. Reichard, P. Cryan, and B. E. Reichert. 2024. Summer Roost Site Suitability Analyses of Four North American Bat Species in the Eastern United States. <<https://doi.org/10.5066/P1AEIUMU>>. Accessed 15 Mar 2025.

- Jacobs, R. P. 2009. A pictorial guide to freshwater fishes of Connecticut. Bulletin / Connecticut Department of Environmental Protection, Connecticut Department of Environmental Protection, Hartford, CT.
- Jensen, T. 2012. Movement, Habitat Use and Detection Probability of Bridle Shiner Estimated by Patch Occupancy Modeling in a Connecticut Watershed. Master's Theses. University of Connecticut, Storrs, USA.
- Johnson, J. B., J. E. Gates, and W. M. Ford. 2008. Distribution and activity of bats at local and landscape scales within a rural–urban gradient. *Urban Ecosystems* 11:227–242.
- Kelly, B., M. J. Siepker, and M. J. Weber. 2021. Factors Associated with Detection and Distribution of Native Brook Trout and Introduced Brown Trout in the Driftless Area of Iowa. *Transactions of the American Fisheries Society* 150:388–406.
- Komac, B., P. Esteban, L. Trapero, and R. Caritg. 2016. Modelization of the Current and Future Habitat Suitability of *Rhododendron ferrugineum* Using Potential Snow Accumulation. *PLOS ONE* 11:e0147324.
- Kuemmerlen, M., B. Schmalz, B. Guse, Q. Cai, N. Fohrer, and S. C. Jähnig. 2014. Integrating catchment properties in small scale species distribution models of stream macroinvertebrates. *Ecological Modelling* 277:77–86.
- Larkin, J. T., D. J. McNeil, L. Chronister, M. E. Akresh, E. B. Cohen, A. W. D'Amato, C. J. Fiss, J. Kitzes, J. L. Larkin, H. A. Parker, and D. I. King. 2024. A large-scale assessment of eastern whip-poor-will (*Antrostomus vociferus*) occupancy across a gradient of forest management intensity using autonomous recording units. *Journal of Environmental Management* 366:121786.
- Liberati, M. R., C. D. Rittenhouse, and J. C. Vokoun. 2016. Beyond protection: Expanding “conservation opportunity” to redefine conservation planning in the 21st century. *Journal of Environmental Management* 183:33–40.
- Liczner, A. R., and S. R. Colla. 2020. One-size does not fit all: at-risk bumble bee habitat management requires species-specific local and landscape considerations. *Insect Conservation and Diversity* 13:558–570.
- Lioy, S., L. Carisio, A. Manino, and M. Porporato. 2023. Climatic Niche Differentiation between the Invasive Hornet *Vespa velutina nigrithorax* and Two Native Hornets in Europe, *Vespa crabro* and *Vespa orientalis*. *Diversity* 15:495.
- Litvaitis, J. A. 1993. Response of early successional vertebrates to historic changes in land use. *Conservation Biology* 7:866–873.
- Litvaitis, J. A. 2001. Importance of early successional habitats to mammals in eastern forests. *Wildlife Society Bulletin* 29:466–473.
- Lomba, A., L. Pellissier, C. Randin, J. Vicente, F. Moreira, J. Honrado, and A. Guisan. 2010. Overcoming the rare species modelling paradox: A novel hierarchical framework applied to an Iberian endemic plant. *Biological Conservation* 143:2647–2657.
- Maitland, B. M., and A. W. Latzka. 2022. Shifting climate conditions affect recruitment in Midwestern stream trout, but depend on seasonal and spatial context. *Ecosphere* 13:e4308.

- Maslo, B., K. Leu, C. Faillace, M. A. Weston, T. Pover, and T. A. Schlacher. 2016. Selecting umbrella species for conservation: A test of habitat models and niche overlap for beach-nesting birds. *Biological Conservation* 203:233–242.
- Maurer, G., M. Chandelier, B. Mulot, and O. Gimenez. 2024. Polarized media coverage of conflicting, yet emblematic species: The ambivalent portrayal of the Asian elephant. *Biological Conservation* 289:110391.
- McCullagh, P., and J. A. Nedler. 1989. *Generalized Linear Models* 2nd Edition. Chapman and Hall.
- McHugh, M. L. 2012. Interrater reliability: the kappa statistic. *Biochemia Medica* 22:276–282.
- Microsoft. 2019. Building Footprints. Microsoft.
<<https://planetarycomputer.microsoft.com/dataset/ms-buildings>>. Accessed 5 Jun 2021.
- Milam, J. C., and S. M. Melvin. 2001. Density, Habitat Use, Movements, and Conservation of Spotted Turtles (*Clemmys guttata*) in Massachusetts. *Journal of Herpetology* 35:418–427.
- Mimno, D., H. Wallach, E. Talley, M. Leenders, and A. McCallum. 2011. Optimizing Semantic Coherence in Topic Models. *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*. Edinburgh, Scotland.
- Mistry, R., and J. D. Ackerman. 2017. Algal flux affects the clearance rates of recently metamorphosed freshwater mussels. *Aquatic Sciences* 79:139–148.
- Moilanen, A., A. M. A. Franco, R. I. Early, R. Fox, B. Wintle, and C. D. Thomas. 2005. Prioritizing multiple-use landscapes for conservation: methods for large multi-species planning problems. *Proceedings. Biological Sciences* 272:1885–1891.
- Moilanen, A., P. Lehtinen, I. Kohonen, J. Jalkanen, E. A. Virtanen, and H. Kujala. 2022. Novel methods for spatial prioritization with applications in conservation, land use planning and ecological impact avoidance. *Methods in Ecology and Evolution* 13:1062–1072.
- Moskowitz, D., and M. May. 2017. Adult tiger spiketail (*Cordulegaster erronea* Hagen) habitat use and home range observed via radio-telemetry with conservation recommendations. *Journal of Insect Conservation* 21:885–895.
- Mothes, C. C., H. J. Howell, and C. A. Searcy. 2020. Habitat suitability models for the imperiled wood turtle (*Glyptemys insculpta*) raise concerns for the species' persistence under future climate change. *Global Ecology and Conservation* 24:e01247.
- Müller-Hansen, F., M. W. Callaghan, and J. C. Minx. 2020. Text as big data: Develop codes of practice for rigorous computational text analysis in energy social science. *Energy Research & Social Science* 70:101691.
- National Aeronautics and Space Administration (NASA). 2000. Shuttle Radar Topography Mission (SRTM GL1) Global 30m.
<<https://portal.opentopography.org/raster?opentopoID=OTSRTM.082015.4326.1>>. Accessed 9 May 2023.

- National Audubon Society. 2024. iba polygons public. <<https://data-library-audubon.hub.arcgis.com/maps/087b591fd12d4dc482a4d9a3e2b60246>>. Accessed 15 Mar 2025.
- National Oceanic and Atmospheric Administration (NOAA). 2020. C-CAP High-Resolution Land Cover. <<https://coast.noaa.gov/digitalcoast/data/ccaphighres.html>>. Accessed 17 Feb 2023.
- NC Wildlife Resources Commission. n.d. Eastern Pondmussel. <<https://www.ncwildlife.gov/species/eastern-pondmussel>>. Accessed 16 Mar 2025.
- NH Fish and Game. 2015. NH 2015 Wildlife Action Plan, Appendix A. NH Fish and Game. <<https://www.wildlife.nh.gov/wildlife-and-habitat/nh-wildlife-action-plan/swap-2015>>. Accessed 16 Mar 2025.
- Natural Resources Conservation Service (NRCS). 2023. Soil Survey Geographic Database (SSURGO). Natural Resources Conservation Service. <<https://www.nrcs.usda.gov/resources/data-and-reports/soil-survey-geographic-database-ssurgo>>. Accessed 16 Mar 2025.
- NY Department of Environmental Conservation. 2014. Species Status Assessment Eastern Pondmussel. NY Department of Environmental Conservation. <https://extapps.dec.ny.gov/docs/wildlife_pdf/sgcnepondmuss.pdf>. Accessed 16 Mar 2025.
- Owens, H., V. Barve, S. Chamberlain, K. Ram, T. Hart, and rOpenSci (ropensci.org/). 2024. spocc: Interface to Species Occurrence Data Sources. <<https://cran.r-project.org/web/packages/spocc/index.html>>. Accessed 16 Mar 2025.
- Phillips, S. J., R. P. Anderson, M. Dudík, R. E. Schapire, and M. E. Blair. 2017. Opening the black box: an open-source release of Maxent. *Ecography* 40:887–893.
- Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190:231–259.
- Phillips, S. J., M. Dudík, and R. E. Schapire. 2020. Maxent software for modeling species niches and distributions (Version 3.4.4). <http://biodiversityinformatics.amnh.org/open_source/maxent/>. Accessed 1 Mar 2023.
- Pregler, K. C., N. Hagstrom, E. T. Schultz, and J. C. Vokoun. 2019. Landscape factors predict local extirpation in an imperilled minnow species, the bridle shiner. *Aquatic Conservation: Marine and Freshwater Ecosystems* 29:1227–1237.
- R Core Team. 2023. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Ripley, B. D. 1996. *Pattern Recognition and Neural Networks*. Cambridge University Press, Cambridge, UK.
- Rittenhouse, C. D., E. H. Berlin, N. Mickle, S. Qiu, D. Riordan, and Z. Zhu. 2022. An object-based approach to map young forest and shrubland vegetation based on multi-source remote sensing data. *Remote Sensing* 14:1091.
- Roberts, M. E., B. M. Stewart, and D. Tingley. 2016. Navigating the Local Modes of Big Data: The Case of Topic Models. Pages 51–97 in R. M. Alvarez, editor. *Computational*

- Social Science: Discovery and Prediction. Analytical Methods for Social Research, Cambridge University Press, Cambridge, UK.
- Roberts, M. E., B. M. Stewart, and D. Tingley. 2019. stm: An R Package for Structural Topic Models. *Journal of Statistical Software* 91:1–40.
- Roberts, M. E., B. M. Stewart, D. Tingley, C. Lucas, J. Leder-Luis, S. K. Gadarian, B. Albertson, and D. G. Rand. 2014. Structural Topic Models for Open-Ended Survey Responses. *American Journal of Political Science* 58:1064–1082.
- Santamarina, S., R. G. Mateo, E. Alfaro-Saiz, and C. Acedo. 2023. On the importance of invasive species niche dynamics in plant conservation management at large and local scale. *Frontiers in Ecology and Evolution* 10:1049142.
- Schofield, A., M. Magnusson, and D. Mimno. 2017. Pulling Out the Stops: Rethinking Stopword Removal for Topic Models. Pages 432–436 in M. Lapata, P. Blunsom, and A. Koller, editors. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*. Association for Computational Linguistics, Valencia, Spain.
- Schofield, A., and D. Mimno. 2016. Comparing Apples to Apple: The Effects of Stemmers on Topic Models. *Transactions of the Association for Computational Linguistics* 4:287–300.
- Shepard, C. A., L. C. Crenshaw, E. M. Baldwin, K. Sammon, K. M. Holman, D. A. Gazaway, N. E. Phelan, W. H. Baltosser, M. Lombardi, M. D. Moran, and M. R. McClung. 2021. Distribution and Habitat Preferences of a Frosted Elfin Subspecies (*Callophrys irus hadros*, Lycaenidae) in Arkansas. *The Journal of the Lepidopterists' Society* 75:104–112.
- Slover, C. L., and T. E. Katzner. 2016. Eastern Whip-Poor-Wills (*Antrostomus vociferus*) Are Positively Associated With Low Elevation Forest In the Central Appalachians. *The Wilson Journal of Ornithology* 128:846–856.
- Smith, A. A., C. D. Rittenhouse, E. Shoreman-Ouimet, and T. E. Worthley. 2024. An examination of opportunities and risks to private forest planning efforts among Connecticut private forest owners. *Trees, Forests and People* 17:100637.
- Soille, P., and P. Vogt. 2009. Morphological segmentation of binary patterns. *Pattern Recognition Letters* 30:456–459.
- Soille, P., and P. Vogt. 2022. Morphological Spatial Pattern Analysis: Open Source Release. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLVIII-4/W1-2022*:427–433.
- Spiller, K. J., and D. I. King. 2021. Breeding Habitat Associations of Eastern Whip-Poor-Wills in Managed Forests. *The Journal of Wildlife Management* 85:1009–1016.
- Steege, S., F. Tuerlinckx, A. Gelman, and W. Vanpaemel. 2016. Increasing Transparency Through a Multiverse Analysis. *Perspectives on Psychological Science* 11:702–712.
- Strayer, D. L. 1993. Macrohabitats of Freshwater Mussels (*Bivalvia:Unionacea*) in Streams of the Northern Atlantic Slope. *Journal of the North American Benthological Society* 12:236–246.
- Sullivan, C. J., and J. C. Vokoun. 2023. Disentangling effects of habitat on salmonid abundance in thermal refuges while accounting for imperfect detection. *Ecosphere* 14:e4727.

- Taddy, M. 2012. On Estimation and Selection for Topic Models. Pages 1184–1193 in. Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics. PMLR 22:1184-1193.
- The Nature Conservancy. 2015. Terrestrial Habitat Map for the Northeast US and Atlantic Canada. <<https://www.conservationgateway.org/ConservationByGeography/NorthAmerica/UnitedStates/edc/reportsdata/terrestrial/habitatmap/Pages/default.aspx>>. Accessed 16 Mar 2025.
- The Nature Conservancy. 2016. Resilient and Connected Landscapes. <<https://www.conservationgateway.org/ConservationByGeography/NorthAmerica/UnitedStates/edc/reportsdata/terrestrial/resilience/Pages/default.aspx>>. Accessed 13 Apr 2025.
- Thornton, M. M., R. Shrestha, Y. Wei, P. E. Thornton, and S.-C. Kao. 2020. Daymet: Annual Climate Summaries on a 1-km Grid for North America, Version 4. Oakridge National Laboratory Distributed Active Archive Center. <<https://doi.org/10.3334/ORNLDAAAC/1852>>. Accessed 23 Mar 2023.
- Thornton, M. M., R. Shrestha, Y. Wei, P. E. Thornton, and S.-C. Kao. 2022. Daymet: Monthly Climate Summaries on a 1-km Grid for North America, Version 4 R1. ORNL DAAC. <<https://doi.org/10.3334/ORNLDAAAC/2131>>. Accessed 16 Mar 2025.
- Trout Unlimited. 2023. Trout Unlimited Priority Waters Connecticut. Trout Unlimited. <<https://prioritywaters.tu.org/connecticut/>>. Accessed 13 Apr 2025.
- Thuiller, W., D. Georges, M. Gueguen, R. Engler, F. Breiner, B. Lafourcade, R. Patin, and H. Blancheteau. 2025. biomod2: Ensemble Platform for Species Distribution Modeling. <<https://cran.r-project.org/web/packages/biomod2/index.html>>. Accessed 16 Mar 2025.
- US Census Bureau. 2023. Census Bureau Data. <<https://data.census.gov/>>. Accessed 15 Mar 2025.
- US Department of Homeland Security. 2018. Homeland Infrastructure Foundation Level Database. <<https://hifld-geoplatfrom.hub.arcgis.com/>>. Accessed 16 Mar 2025.
- US Department of Agriculture. 2021. NLCD CONUS 2021 Tree Canopy Cover. <https://www.mrlc.gov/downloads/sciweb1/shared/mrlc/metadata/nlcd_tcc_conus_2021_v2021-4.tif.html>. Accessed 16 Mar 2025.
- US Fish and Wildlife Service (USFWS). 2020. National Wetlands Inventory. US Fish and Wildlife Service. <<https://www.fws.gov/program/national-wetlands-inventory>>. Accessed 2 Jul 2021.
- US Fish and Wildlife Service (USFWS). 2019. Yellow Banded Bumble Bee (*Bombus terricola*). US Fish and Wildlife Service. <<https://www.fws.gov/species/yellow-banded-bumble-bee-bombus-terricola>>. Accessed 16 Mar 2025.
- US Fish and Wildlife Service (USFWS). 2024a. American Oystercatcher (*Haematopus palliatus*). US Fish and Wildlife Service. <<https://www.fws.gov/species/american-oystercatcher-haematopus-palliatus>>. Accessed 16 Mar 2025.
- US Fish and Wildlife Service (USFWS). 2024b. Tricolored Bat (*Perimyotis subflavus*). US Fish and Wildlife Service. <<https://www.fws.gov/species/tricolored-bat-perimyotis-subflavus>>. Accessed 16 Mar 2025.

- US Fish and Wildlife Service (USFWS). 2024c. Official Current Highlands Region. US Fish and Wildlife Service. <https://gis-fws.opendata.arcgis.com/datasets/33939c4d654b4d21b4683b9b1adf2320_1/explore?location=41.019974,-74.668447,6.52>. Accessed 17 Mar 2025.
- US Geological Survey (USGS). 2005. Topographic Maps. US Geological Survey. <<https://www.usgs.gov/programs/national-geospatial-program/topographic-maps>>. Accessed 16 Mar 2025.
- US Geological Survey (USGS). 2018. Protected Areas Database of the United States (PAD-US) 2.0. US Geological Survey. <<https://www.usgs.gov/data/protected-areas-database-united-states-pad-us-20>>. Accessed 17 Mar 2025.
- US Geological Survey (USGS). 2021. Watershed Boundary Dataset. US Geological Survey. <<https://www.usgs.gov/national-hydrography/watershed-boundary-dataset>>. Accessed 16 Mar 2025.
- Visit New England. 2025. Connecticut Travel Vacation Guide, Attractions, & Events. Visit Connecticut. <<https://www.visitconnecticut.com/state/science-and-nature/>>. Accessed 16 Mar 2025.
- Vogt, P., and K. Riitters. 2017. GuidosToolbox: universal digital image object analysis. *European Journal of Remote Sensing* 50:352–361.
- Vogt, P., K. Riitters, P. Rambaud, R. d’Annunzio, E. Lindquist, and A. Pekkarinen. 2022. GuidosToolbox Workbench: spatial analysis of raster maps for ecological applications. *Ecography* 2022:e05864.
- Wagner, T., J. T. Deweber, J. Detar, and J. A. Sweka. 2013. Landscape-Scale Evaluation of Asymmetric Interactions between Brown Trout and Brook Trout Using Two-Species Occupancy Models. *Transactions of the American Fisheries Society* 142:353–361.
- Warren, D. L., R. E. Glor, and M. Turelli. 2008. Environmental Niche Equivalency Versus Conservatism: Quantitative Approaches to Niche Evolution. *Evolution* 62:2868–2883.
- Weston, S. J., I. Shryock, R. Light, and P. A. Fisher. 2023. Selecting the Number and Labels of Topics in Topic Modeling: A Tutorial. *Advances in Methods and Practices in Psychological Science* 6:25152459231160105.
- Willey, L. L., M. T. Jones, P. R. Sievert, T. S. B. Akre, M. Marchand, P. deMaynadier, D. Yorks, J. Mays, J. Dragon, L. Erb, B. Zarate, J. D. Kleopfer, K. D. Gipe, S. Parren, J. Andrews, H. P. Roberts, J. W. Tamplin, C. Raithel, L. Johnson, A. M. Ross, A. Breisch, G. Johnson, S. Angus, J. Drasher, and B. Wicklow. 2022. Distribution models combined with standardized surveys reveal widespread habitat loss in a threatened turtle species. *Biological Conservation* 266:109437.
- Williams, S. C., S. K. Krueger, G. A. Zirkle, and C. G. Haase. 2025. Summer roost site selection of a declining bat species. *Mammalian Biology* 105:57–68.
- Yang, X., S. Qiu, Z. Zhu, C. Rittenhouse, D. Riordan, and M. Cullerton. 2023. Mapping understory plant communities in deciduous forests from Sentinel-2 time series. *Remote Sensing of Environment* 293:113601.