Exploring the Risk of Dreissenid Invasion in USACE Reservoirs

Todd Swannack, Iris Foxfoot, Kiara Cushway

2025-02-07

Table of contents

Purpose

This document summarizes the exploratory analysis we have conducted so far to assess and predict Dreissenid (*i.e.*, *Dreissena polymorpha* and *Dreissena rostriformis bugensis*) invasion in USACE reservoirs.

Data exploration

Data description

A description of the datasets used can be found below:

DATA OVERVIEW:

The data contained in these files were generated from multiple sources: - USACE reservoir GIS data was sourced from a shapefile downloaded from https://geospatial-usace.opendata.arcgis.com/datasets/03e322d7e89b48a9b48e9c3f4bcaf29e/explore - recreational visitor information was provided by The U.S. Army Corps of Engineers Institute for Water Resources. - Dreissenid presenece/absence data was derived from USGS NAS data downloaded from https://nas.er.usgs.gov/queries/CollectionInfo.aspx?SpeciesID=3118 - Land use data was sourced from the NLCD land cover for 2021 downloaded from https://www.mrlc.gov/data?f%5B0%5D=category%3ALand%20Cover and USACE reservoir shapefiles. - Surface geology data was generated using USGS surficial geology data downloaded from https://pubs.usgs.gov/ds/425/ and USACE reservoir shapefiles. - Water hardness and pH data were interpolated for the conterminous US by the EPA and was obtained at: https://www.arcgis.com/home/item.html?id=ec9fe618d7e74c5ca68667b1c8d6e9bc and https://www.arcgis.com/home/item.html?id=bdc5b9ca9bfc454d8af22912ba150035. - Climate data was sourced from PRISM 30 year (1990-2020) normal climate data, which was downloaded using the prism R package. For information on PRISM data see https://prism.oregonstate.edu/. - Normalized Difference Chlorophyll Index (NDCI) data was derived from sentinel 2 imagery. Processed and downloaded from google earth engine.

Total records used in analysis: 352

NOTES ABOUT CHANGES TO USACE RESERVOIRS SHAPEFILE:

1. There are two Sardis Lakes (Vicksburg and Tulsa districts), which have been renamed Sardis Lake Vicksburg and Sardis Lake Tulsa to distinguish them from one another.
2. Aberdeen Lake, Columbus Lake, R.E. Bob Woodruff Lake, William Dannelly Reservoir, Bay Springs Lake, Tennessee-Tombigbee Waterway at Fulton, Tennessee-Tombigbee Waterway at G.V. Montogmery, and Tennessee-Tombigbee Waterway at Glover Wilkins in the Mobile District had repeated records that were a subset of the whole lake, and these were deleted for analysis.
3. Water Conservation Area 1, 2A, 3A, 2B, and 3B were merged into a single record: "Water Conservation Area Combined"
4. Arkport Reservoir was deleted because it is a dry dam that is only impounded after rain
5. Jadwin Reservoir was deleted because it is a dry dam that is only impounded after rain
6. Candy Lake Reservoir was deleted because it was never built according to USACE records
7. Hulah Lake, Lake Barkley, and Lake Shelbyville had extra disconnected fragments that were deleted
8. Southern disconnected portion of Lake Merrisach was deleted
9. Lake Merrisach, Arkansas River Pool 2, and Arkansas Post Canal Reservoir were merged into a single record because they appeared to be entirely connected
10. The section of Robert S. Kerr Reservoir between Robert S. Kerr Lock and Dam and W.D. Mayo Lock and Dam was deleted
11. Arkansas River Pool 13 was combined with the portion of Robert S. Kerr Reservoir that stretches upstream to W.D. Mayo Lock and Dam
12. 50 reservoirs were removed because they were dry according to USACE record (Dry=Yes in attribute table)

HARDNESS AND pH COLUMNS:

1. Water hardness and pH data were interpolated for the conterminous US by the EPA and was obtained at: https://www.arcgis.com/home/item.html?id=ec9fe618d7e74c5ca68667b1c8d6e9bc and

https://www.arcgis.com/home/item.html?id=bdc5b9ca9bfc454d8af22912ba150035. Data was generated by clipping the shapefile to the extent of the USACE reservoirs shapefile and converting to a raster, then using tabulate area to determine hardness class of each reservoir

2. hardness_range includes the range of hardness values occurring in the reservoir based on the hardness shapefile. If multiple classes occurred in the reservoir, the range was expanded to include all of them

COLUMNS PERC_OPEN_WATER_25MI - PERC_HERBACEOUSWETLAND_50MI:

This data was generated using NLCD land cover for 2021 downloaded from https://www.mrlc.gov/data?f%5B0%5D=category%3ALand%20Cover and USACE reservoir shapefiles.

In ArcGIS Pro, 25 mile and 50 mile buffers were created around each USACE reservoir, and the number of pixels belonging to each land use type was counted for each buffer using the "Tabulate Area" tool. The percent of each buffer representing each land use type was then calculated by dividing the number of pixels of each land use type by the total number of pixels in the buffer. Note that not all percentages add up to 100% because some buffers extended into Canada or the ocean, outside of the coverage of NLCD data (hence, pixels we considered "unclassed" and were excluded from analysis).

SURFICIAL GEOLOGY:

This data was generated using USGS surficial geology data downloaded from https://pubs.usgs.gov/ds/425/ and USACE reservoir shapefiles. In ArcGIS Pro, 25 mile buffers were created around each USACE reservoir, and the number of pixels belonging to each surficial geology type was counted for each buffer using the "Tabulate Area" tool. The percent of each buffer representing each land use type was then calculated by dividing the number of pixels of each land use type by the total number of pixels in the buffer.

CLIMATE DATA:

Climate data was extracted from PRISM 30 year (1990-2020) normal climate data, which was downloaded using the prism R package. For information on PRISM data see https://prism.oregonstate.edu/.

to produce this data, the center point of each reservior was found. Using this center point, values for "ppt" (precipitation), "tmean" (mean temperature), "tmin" (min temperature), "tmax" (max temperature), "tdmean" (mean dew point temp) "vpdmin" (minimum vapor deficit), and "vpdmax" (max vapor deficit) were extracted from the PRISM 30 year (1990-2020) year normal 800m resolution raster cells. These data were then condensed into seasonal totals (for precipitaiton), seasonal means (for temperature, dew point-called dewp), seasonal mins (for temperature and vapor pressure deficit –called vpd), and seasonal maxs (for temperature and vapor pressure deficit–called vpd)

temperature is in Celsius, precipitation is in millimeters, and vapor pressure deficit is in hectopascals.

seasons are as follows: - winter = dec, jan, feb - spring = mar, apr, may - summer = jun, jul, aug - fall = sep, oct, nov

NDCI DATA:

Sentinel-2 data acquisition was done using google earth engine. We used Sentinel 2: Level-2A orthorectified atmospherically corrected surface reflectance from Jan 1st 2023 to Jan 1st 2024. First, the USACE reservoir shapefile was uploaded to google earth engine as an asset. It was then imported and simplified to be accurate within 100m using the simplify function to reduce the complexity of the object. Within each simplified reservoir, we then selected pixels that were classified as water in sentinel-2 L2A's Scene Classification (SCL) image. Water pixels are designated a value of 6. This was done to remove potential clouds, ice, snow, and land pixels from the dataset. Then we calculated the mean of all bands of interest across all water pixels within the simplified reservoir. These values were then exported from google earth engine.

Then, using R we calculated Normalized Difference Chlorophyll Index (NDCI) from Mishra & Mishra 2012. The NDCI output ranges from -1 to 1, smaller values likely have less chlorophyll-a while higher values likely have an increased concentration of chlorophyll-a. Though the NDCI was developed for coastal and estuary waters, it has also been successfully used for North American reservoirs (see Kislik et al., 2022). NDCI was then aggregated into seasonal minimums, means, and maximums for each reservoir.

When using sentinel 2 data, the formula for NDCI is

$(B5 - B4)/(B5 + B4)$

Where B5 is red edge 1 and B4 is red.

seasons are as follows: - winter = dec, jan, feb - spring = mar, apr, may - summer = jun, jul, aug - fall = sep, oct, nov

COLUMNS:

- name: Name of reservoir (from USACE shapefiles)
- district: District in charge of reservoir (from USACE shapefiles)
- dist_sym: District symbol (from USACE shapefiles)
- division: Division in charge of reservoir (from USACE shapefiles)
- div_sym: Division symbol (from USACE shapefiles)
- dry: Whether the reservoir is dry (from USACE shapefiles)
- connectivity: Type of barrier between reservoir and other reservoir polygons immediately adjacent (NOTE: this only includes reservoirs where polygons are immediately adjacent, not reservoirs that may be relatively connected according to imagery)
- dam_name: Name of dam (from USACE shapefiles)
- dist_to_infest_km: Distance to nearest ZQM presence record from NAS data in km, calculated using "Generate Near Table" in ArcGIS Pro using the geodesic measurement method. If the near distance was 0, a reservoir was infested, and the next closest record was selected. To account for measurement/GPS error, all records within 100m (0.1km) of a reservoir were checked visually. If a record was not clearly in a different waterbody, the record was considered within the range of GPS error and the next furthest record was selected until the record was either >100m away from the reservoir or clearly occupying another waterbody (e.g., above a dam)
- surface_area_km: Surface area of reservoir (km); calculated using "Calculate Geometry" tool in ArcGIS Pro
- connectivity: Type of connection between adjacent reservoirs (None, Lock and Dam, Dam)
- num_connections: Number of direct connections (e.g., lock and dam) a reservoir has to adjacent reservoirs
- perc_open_water_25mi: percent land cover within 25 miles of reservoir consisting of open water
- perc_perennial_snowice_25mi: percent land cover within 25 miles of reservoir consisting of perennial snow and ice
- perce_dev_openspace_25mi: percent land cover within 25 miles of reservoir consisting of developed open space
- perc_dev_lowintensity_25mi: percent land cover within 25 miles of reservoir consisting of low intensity development
- perc_dev_medintensity_25mi: percent land cover within 25 miles of reservoir consisting of medium intensity development
- perc_dev_highintensity_25mi: percent land cover within 25 miles of reservoir consisting of high intensity development

- perc_barren_25mi: percent land cover within 25 miles of reservoir consisting of barren land
- perc_forest_25mi: percent land cover within 25 miles of reservoir consisting of deciduous, evergreen, or mixed forest
- perc_shrubscrub_25mi: percent land cover within 25 miles of reservoir consisting of shrub and scrub
- perc_herbaceous_25mi: percent land cover within 25 miles of reservoir consisting of herbaceous land
- perc_haypasture_25mi: percent land cover within 25 miles of reservoir consisting of hay and pastureland
- perc_crops_25mi: percent land cover within 25 miles of reservoir consisting of crops
- perc_wetland_25mi: percent land cover within 25 miles of reservoir consisting of woody or herbaceous wetlands
- perc_open_water_50mi: percent land cover within 50 miles of reservoir consisting of open water
- perc_perennial_snowice_50mi: percent land cover within 50 miles of reservoir consisting of perennial snow and ice
- perce_dev_openspace_50mi: percent land cover within 50 miles of reservoir consisting of developed open space
- perc_dev_lowintensity_50mi: percent land cover within 50 miles of reservoir consisting of low intensity development
- perc_dev_medintensity_50mi: percent land cover within 50 miles of reservoir consisting of medium intensity development
- perc_dev_highintensity_50mi: percent land cover within 50 miles of reservoir consisting of high intensity development
- perc_barren_50mi: percent land cover within 50 miles of reservoir consisting of barren land
- perc_forest_50mi: percent land cover within 50 miles of reservoir consisting of deciduous, evergreen, or mixed forest
- perc_shrubscrub_50mi: percent land cover within 50 miles of reservoir consisting of shrub and scrub
- perc_herbaceous_50mi: percent land cover within 50 miles of reservoir consisting of herbaceous vegetation
- perc_haypasture_50mi: percent land cover within 50 miles of reservoir consisting of hay and pastureland
- perc_crops_50mi: percent land cover within 50 miles of reservoir consisting of cropland
- perc_wetland_50mi: percent land cover within 50 miles of reservoir consisting of woody or herbaceous wetland
- mean_slope: average slope of reservoir, calculated using PRISM slope raster and zonal statistics as table with reservoirs as zones. A few reservoirs were calculated by hand because zonal statistics did not work for them
- mean_elev_m: average elevation of each reservoir (in m), calculated using PRISM 800m elevation raster and zonal statistics as table with reservoirs as zones. A few reservoirs were calculated by hand because zonal statistics did not work for them
- perc_CaO_25mi: % lithological calcium oxide (CaO) content in surface or near surface geology in a 25 mile buffer around reservoir; obtained from https://www.sciencebase.gov/catalog/item/53543d10e4b0bab7f98ce7e0

- perc_alluvial_25mi: % surficial geology in a 25 mile buffer with a major classification of “alluvial” (UNIT CODE begins with 0)
- perc_eolian_25mi:% surficial geology in a 25 mile buffer with a major classification of “eolian” (UNIT CODE begins with 3)
- perc_glacial_glaciofluvial_25mi: % surficial geology in a 25 mile buffer with a major classification of “glacial till and glaciofluvial” (UNIT CODE begins with 4)
- perc_lacustrine_25mi colluvial_25mi: % surficial geology in a 25 mile buffer with a major classification of “colluvial” (UNIT CODE begins with 6)
- perc_organic_rich_25mi: % surficial geology in a 25 mile buffer with a major classification of “organic-rich” (UNIT CODE begins with 7)
- perc_proglacial_25mi:% surficial geology in a 25 mile buffer with a major classification of “proglacial” (UNIT CODE begins with 8)
- perc_resid_volcanic_art_wat_25mi: % surficial geology in a 25 mile buffer with a major classification of “residual, volcanic, artificial, and water” (UNIT CODE begins with 9)
- hardness_range: range of hardness values in milligrams per liter of CaCO3 observed based on EPA shapefile
- min_hardness: minimum hardness value in milligrams per liter of CaCO3 observed in reservoir based on EPA shapefile
- max_hardness: maximum hardness value in milligrams per liter of CaCO3 observed in reservoir based on EPA shapefile
- pH_range: range of pH values observed in reservoir based on interpolated shapefile from EPA
- pH_min: minimum pH observed near reservoir based on interpolated shapefile from EPA
- pH_max: maximum pH observed near reservoir based on interpolated shapefile from EPA
- winter_total_precip: Winter total precip (mm), calculated from PRISM 30 yr normal dataset.
- spring_total_precip: Spring total precip (mm), calculated from PRISM 30 yr normal dataset.
- summer_total_precip: Summer total precip (mm), calculated from PRISM 30 yr normal dataset.
- fall_total_precip: Fall total precip (mm), calculated from PRISM 30 yr normal dataset.
- winter_mean_temp: Winter mean air temperature (C), calculated from PRISM 30 yr normal dataset.
- spring_mean_temp: Spring mean air temperature (C), calculated from PRISM 30 yr normal dataset.
- summer_mean_temp: Summer mean air temperature (C), calculated from PRISM 30 yr normal dataset.
- fall_mean_temp: Fall mean air temperature (C), calculated from PRISM 30 yr normal dataset.

- winter_min_temp: Winter mean daily minimum air temperature (C), calculated from PRISM 30 yr normal dataset.

- spring_min_temp: Spring mean daily minimum air temperature (C), calculated from PRISM 30 yr normal dataset.

- summer_min_temp: Summer mean daily minimum air temperature (C), calculated from PRISM 30 yr normal dataset.

- fall_min_temp: Fall mean daily minimum air temperature (C), calculated from PRISM 30 yr normal dataset.

- winter_max_temp: Winter mean daily maximum air temperature (C), calculated from PRISM 30 yr normal dataset.

- spring_max_temp: Spring mean daily maximum air temperature (C), calculated from PRISM 30 yr normal dataset.

- summer_max_temp: Summer mean daily maximum air temperature (C), calculated from PRISM 30 yr normal dataset.

- fall_max_temp: Fall mean daily maximum air temperature (C), calculated from PRISM 30 yr normal dataset.

- winter_mean_dewp: Winter mean daily dew point (C), calculated from PRISM 30 yr normal dataset.

- spring_mean_dewp: Spring mean daily dew point (C), calculated from PRISM 30 yr normal dataset.

- summer_mean_dewp: Summer mean daily dew point (C), calculated from PRISM 30 yr normal dataset.

- fall_mean_dewp: Fall mean daily dew point (C), calculated from PRISM 30 yr normal dataset.

- winter_min_vpd: Winter mean daily minimum vapor pressure deficit (hectopascals), calculated from PRISM 30 yr normal dataset.

- spring_min_vpd: Spring mean daily minimum vapor pressure deficit (hectopascals), calculated from PRISM 30 yr normal dataset.

- summer_min_vpd: Summer mean daily minimum vapor pressure deficit (hectopascals), calculated from PRISM 30 yr normal dataset.

- fall_min_vpd: Fall mean daily minimum vapor pressure deficit (hectopascals), calculated from PRISM 30 yr normal dataset.

- winter_max_vpd: Winter mean daily maximum vapor pressure deficit (hectopascals), calculated from PRISM 30 yr normal dataset.

- spring_max_vpd: Spring mean daily maximum vapor pressure deficit (hectopascals), calculated from PRISM 30 yr normal dataset.

- summer_max_vpd: Summer mean daily maximum vapor pressure deficit (hectopascals), calculated from PRISM 30 yr normal dataset.

- fall_max_vpd: Fall mean daily maximum vapor pressure deficit (hectopascals), calculated from PRISM 30 yr normal dataset.

- mean_ndci_fall: Fall mean Normalized Difference Chlorophyll Index.
- mean_ndci_spring: Spring mean Normalized Difference Chlorophyll Index.

- mean_ndci_summer: Summer mean Normalized Difference Chlorophyll Index.

- mean_ndci_winter: Winter mean Normalized Difference Chlorophyll Index.

- min_ndci_fall: Fall minimum Normalized Difference Chlorophyll Index.

- min_ndci_spring: Spring minimum Normalized Difference Chlorophyll Index.

- min_ndci_summer: Summer minimum Normalized Difference Chlorophyll Index.

- min_ndci_winter: Winter minimum Normalized Difference Chlorophyll Index.

- max_ndci_fall: Fall maximum Normalized Difference Chlorophyll Index.

- max_ndci_spring: Spring maximum Normalized Difference Chlorophyll Index.

- max_ndci_summer: Summer maximum Normalized Difference Chlorophyll Index.

- max_ndci_winter: Winter maximum Normalized Difference Chlorophyll Index.
- mean_total_visits: The average annual number of recreational visitors from 2014-2024.
- min_Ca:
- max_Ca:
- avg_Ca:
- min_pH:
- max_pH:

- avg_pH:
- infest_status: Whether or not a reservoir is infested by ZQM or not according to NAS data (determined based on dist_to_infest_km=0 or <100m and not clearly in some other water body); 0=uninfested, 1=infested

Infestation patterns in USACE reservoirs

Approximately 24% of the 352 reservoirs we examined have recorded Dreissenid presences according to USGS Non-indigenous Aquatic Species (NAS) data (Benson et al. 2024). The remaining 76% are currently considered un-infested (Table 1). Note that NAS data provides information on reports of Dreissenid presence only, and presence does not necessarily imply successful establishment. Absences are not reported. Note also that the absence of positive records in a state (e.g., Michigan) does not imply that Dreissenids are not present there, just that no Dreissenids have been recorded in USACE reservoirs within the state.

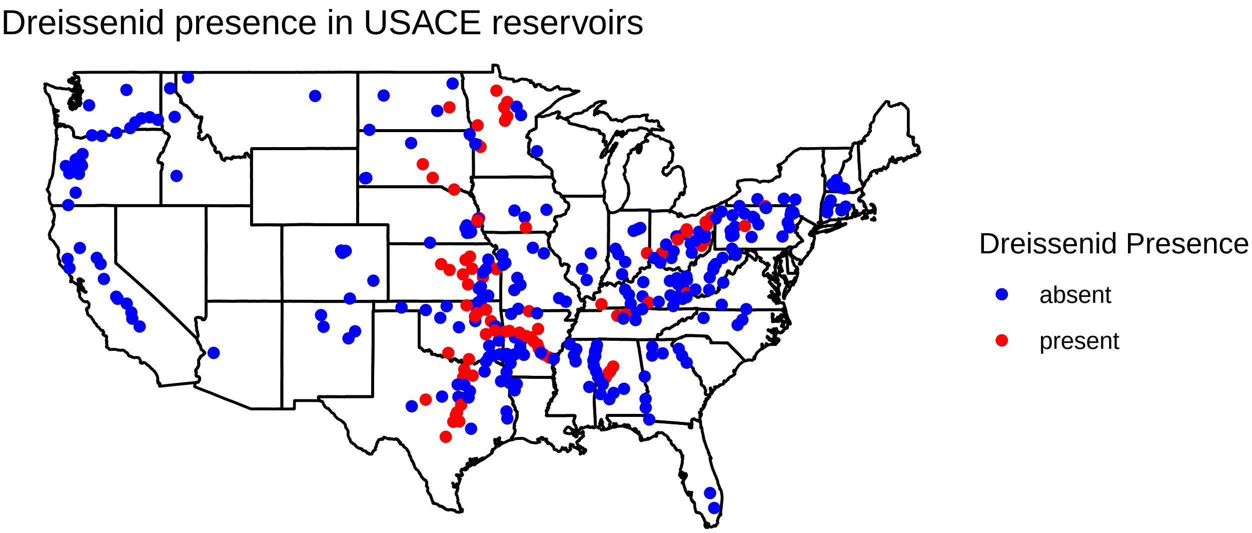

Figure 1. Location of Dreissenid infested and un-infested USACE reservoirs (n=352). Infestation status was determined using USGS Non-indigenous Aquatic Species data (Benson et al. 2024).

Table 1: Number of Dreissenid infested and un-infested USACE reservoirs (n=352). Infestation status was determined using USGS Non-indigenous Aquatic Species data (Benson et al. 2024).

Infestation status	Number of reservoirs
Un-infested	268
Infested	84

Environmental thresholds

We attempted to predict and assess Dreissenid infestation in USACE reservoirs using environmental data and assessed invasion risk based on discrete environmental thresholds that have been identified in the literature as facilitating or preventing Dreissenid invasion (Doll 1997; Sorba and Williamson 1997; Cohen and Weinstein 1998; Cohen 2005; Creamer et al., 2025). We used these environmental thresholds to define reservoir characteristics that may lead to low, moderate, or high risk of Dreissenid infestation (Table 2). Since not all literature reported the same thresholds, we chose the most conservative values such that we were more likely to overestimate than underestimate risk of infestation. For example, if one study defined low risk habitat as areas with a hardness < 45 mg/L and another as < 55 mg/L, we selected 45 mg/L as the maximum hardness describing low risk habitat, resulting in a wider moderate risk category.

Initially, we included maximum summer temperature in our risk assessment, but we found that the risk categories identified were not necessarily a good representation of invasion risk, as a higher proportion of moderate-risk reservoirs were already invaded by Dreissenids compared to high risk reservoirs. While maximum temperature is likely a factor that does limit Dreissenid range, the duration of high temperature events may be an important factor that was missing from our assessment. Since we have no information about how long maximum summer temperatures occurred, it is difficult to assess their effect on mussels, as Dreissenids may be able to survive short-term exposure to high temperatures but could have higher mortality during longer high temperature events (White et al., 2015). Hence, we removed maximum temperature from our risk assessment but included it in our statistical analyses investigating the relationship between environmental parameters and Dreissenid presence.

Table 2: Descriptions of environmental characteristics used in this analysis identified as contributing to low, moderate, or high risk of Dreissenid invasion. Note that in some cases, limited categorical data representing ranges was available for variables.

Metric	**Low risk**	**Moderate risk**	**High risk**	**Reference**
Calcium	< 9	9.0-15	> 15	Doll 1997 (in Cohen 2005)
pH	< 6.8 or > 9.5	6.8-7.4 or 8.7-9.5	7.4-8.7	Doll 1997 (in Cohen 2005)
Mean summer temperature (C)	< 15 or > 32	31-32	15-31	Doll 1997 (in Cohen 2005)
Hardness (mg/L)	< 45	45-90	>= 90	Sorba & Williamson 1997 (in Cohen 2005)
Distance to nearest infestation (km)	> 125	51-125	< 51	Creamer et al. 2025

Using the data we compiled, we assigned each USACE reservoir as low, moderate, or high risk for each of the variables in Table 2 (Table 3). We then assessed how many reservoirs falling into each risk category were currently infested (Table 4, Figure 3) and assigned an overall risk for each reservoir by selecting the lowest risk category for any variable (Table 4, Table 5, Figure 3). Any reservoirs categorized as low risk for any of the variables were assigned a low overall risk, while reservoirs with no low risk variables and at least one moderate risk variable were assigned moderate risk. Reservoirs assigned high risk status for all variables were defined as having high overall risk.

Table 3: Risk classifications for USACE reservoirs based on environmental thresholds known to influence Dreissenid survival.

Reservoir	Mean summer temp risk	pH risk	Hardness risk	Calcium risk	Distance risk	Overall risk	Overall risk without distance	Infestation status
Aberdeen Lake	High	Moderate	High	High	Moderate	Moderate	Moderate	Not infested
Abiquiu Reservoir	High	High	High	High	Low	Low	High	Not infested
Alamo Lake	Low	High	High	High	Low	Low	Low	Not infested
Aliceville Lake	High	Moderate	High	High	Moderate	Moderate	Moderate	Not infested
Allegheny Reservoir	High	High	High	High	High	High	High	Not infested
Almond Lake	High	High	High	High	High	High	High	Not infested
Alum Creek Lake	High	High	High	High	High	High	High	Infested
Applegate Lake	High	High	High	High	Low	Low	High	Not infested
Aquilla Lake	High	High	High	High	High	High	High	Not infested
Arcadia Lake	High	High	High	High	High	High	High	Not infested
Arkabutla Lake	High	Moderate	High	High	High	Moderate	Moderate	Not infested
Arkansas River Pool 3	High	Moderate	High	High	High	Moderate	Moderate	Infested
Arkansas River Pool 4	High	Moderate	High	High	High	Moderate	Moderate	Not infested
Arkansas River Pool 5	High	Moderate	High	High	High	Moderate	Moderate	Infested
Arkansas River Pool 6	High	Moderate	High	High	High	Moderate	Moderate	Infested
Arkansas River Pool 7	High	Moderate	High	High	High	Moderate	Moderate	Infested
Arkansas River Pool 8	High	Moderate	High	High	High	Moderate	Moderate	Infested
Arkansas River Pool 9	High	Moderate	High	High	High	Moderate	Moderate	Infested
Arkansas River Pool 13	High	Moderate	High	High	High	Moderate	Moderate	Infested
Atwood Lake	High	High	High	High	High	High	High	Not infested
Aylesworth Creek Lake	High	Moderate	High	High	High	Moderate	Moderate	Not infested
B. Everett Jordan Lake	High	High	Moderate	High	Low	Low	Moderate	Not infested
Ball Mountain Lake	High	Moderate	High	Low	High	Low	Low	Not infested
Bankhead Lake	High	High	High	High	High	High	High	Infested
Bardwell Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Barren River Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Bay Springs Lake and Divide Cut	High	Moderate	High	Moderate	High	Moderate	Moderate	Not infested
Bayou Bodcau Reservoir	High	Moderate	High	Moderate	Low	Low	Moderate	Not infested
Beach City Lake	High	High	High	High	High	High	High	Not infested
Bear Creek Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Beaver Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Beech Fork Lake	High	Moderate	High	High	High	Moderate	Moderate	Not infested
Belton Lake	High	High	High	High	High	High	High	Infested
Beltzville Lake	High	Moderate	High	Moderate	High	Moderate	Moderate	Not infested
Benbrook Lake	High	High	High	High	High	High	High	Not infested
Berlin Lake	High	High	High	High	High	High	High	Infested

Big Hill Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Big Sandy Lake Reservoir	High	High	High	High	High	High	High	Not infested
Birch Lake	High	High	High	High	High	High	High	Not infested
Black Butte Lake	High	High	High	High	Low	Low	High	Not infested
Black Rock Lake	High	Moderate	High	Moderate	High	Moderate	Moderate	Not infested
Blue Marsh Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Blue Mountain Lake	High	Moderate	Moderate	Low	High	Low	Low	Not infested
Blue River Lake	High	High	Moderate	Low	Low	Low	Low	Not infested
Blue Springs Lake	High	High	High	High	High	High	High	Infested
Bluestem Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Bluestone Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Bowman-Haley Lake	High	High	High	High	Low	Low	High	Not infested
Branched Oak Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Broken Bow Lake	High	Moderate	High	Moderate	Moderate	Moderate	Moderate	Not infested
Brookville Lake	High	High	High	High	High	High	High	Infested
Buckhorn Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Bull Shoals Lake	High	High	High	High	High	High	High	Infested
Burnsville Lake	High	Moderate	High	Moderate	High	Moderate	Moderate	Not infested
Burr Oak Reservoir	High	High	High	High	High	High	High	Not infested
Caddo Lake	High	Moderate	High	High	Low	Low	Moderate	Not infested
Caesar Creek Lake	High	High	High	High	High	High	High	Infested
Cagles Mill Lake	High	High	High	High	High	High	High	Not infested
Canton Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Canyon Lake	High	High	High	High	High	High	High	Infested
Carlyle Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Carr Creek Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Carters Lake	High	Moderate	Moderate	High	Moderate	Moderate	Moderate	Not infested
Cave Run Lake	High	Moderate	High	High	Moderate	Moderate	Moderate	Not infested
Cecil M. Hardin Lake	High	High	High	High	High	High	High	Not infested
Center Hill Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Charles Mill Lake	High	High	High	High	High	High	High	Infested
Chatfield Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Cheatham Lake	High	High	High	High	High	High	High	Infested
Cherry Creek Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Claiborne Lake	High	Moderate	High	High	Moderate	Moderate	Moderate	Not infested
Clarence J. Brown Reservoir	High	High	High	High	High	High	High	Not infested
Clearwater Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Clendening Lake	High	High	High	High	High	High	High	Infested
Clinton Lake	High	High	High	High	High	High	High	Infested
Cochiti Reservoir	High	High	High	High	Low	Low	High	Not infested
Coffeeville Lake	High	Moderate	High	High	High	Moderate	Moderate	Not infested
Cold Brook Lake	High	High	High	High	Moderate	Moderate	High	Not infested

Colebrook River Lake	High	Moderate	High	Moderate	High	Moderate	Moderate	Not infested
Columbus Lake	High	Moderate	High	Moderate	Moderate	Moderate	Moderate	Not infested
Conchas Reservoir	High	High	High	High	Low	Low	High	Not infested
Conemaugh River Lake	High	Moderate	High	High	High	Moderate	Moderate	Not infested
Conestoga Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Copan Lake	High	High	High	High	High	High	High	Not infested
Coralville Lake	High	High	High	High	High	High	High	Not infested
Cordell Hull Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Cottage Grove Lake	High	High	Moderate	Low	Low	Low	Low	Not infested
Cottonwood Springs Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Cougar Lake	High	High	Moderate	Low	Low	Low	Low	Not infested
Council Grove Lake	High	High	High	High	High	High	High	Infested
Cowanesque Lake	High	High	High	High	High	High	High	Infested
Crooked Creek Lake	High	High	High	High	High	High	High	Not infested
Curwensville Lake	High	Moderate	High	High	Moderate	Moderate	Moderate	Infested
Dale Hollow Lake	High	High	High	High	High	High	High	Not infested
Dardanelle Lake	High	Moderate	High	High	High	Moderate	Moderate	Infested
Deer Creek Lake	High	High	High	High	High	High	High	Not infested
DeGray Lake	High	Low	High	Moderate	Moderate	Low	Low	Not infested
DeGray Reservoir	High	Low	Low	Low	Moderate	Low	Low	Not infested
Delaware Lake	High	High	High	High	High	High	High	Infested
Demopolis Lake	High	High	High	High	High	High	High	Not infested
DeQueen Reservoir	High	Moderate	High	Low	Low	Low	Low	Not infested
Detroit Lake	High	High	High	Low	Low	Low	Low	Not infested
Dewey Lake	High	High	High	High	High	High	High	Infested
Dexter Lake	High	High	Moderate	Low	Low	Low	Low	Not infested
Dierks Reservoir	High	Low	High	Low	Low	Low	Low	Not infested
Dillon Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Dorena Lake	High	High	Moderate	Low	Low	Low	Low	Not infested
Dworshak Reservoir	High	High	High	Moderate	Low	Low	Moderate	Not infested
East Branch Clarion River Lake	High	Moderate	High	Moderate	Moderate	Moderate	Moderate	Not infested
East Lynn Lake	High	High	High	High	High	High	High	Not infested
East Sidney Lake	High	High	High	Low	High	Low	Low	Not infested
Eau Galle Reservoir	High	High	High	High	High	High	High	Not infested
Edward MacDowell Lake	High	Low	Moderate	Low	Moderate	Low	Low	Not infested
El Dorado Lake	High	High	High	High	High	High	High	Infested
Elk City Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Englebright Lake	High	High	High	Moderate	Low	Low	Moderate	Not infested
Enid Lake	High	Low	High	High	Moderate	Low	Low	Not infested
Eufaula Lake	High	High	High	High	High	High	High	Infested
F. E. Walter Reservoir	High	Low	High	Low	High	Low	Low	Not infested
Fall Creek Lake	High	High	Moderate	Low	Low	Low	Low	Not infested

Fall River Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Falls Lake	High	Moderate	Moderate	Moderate	Low	Low	Moderate	Not infested
Fern Ridge Lake	High	High	Moderate	High	Low	Low	Moderate	Not infested
Fishtrap Lake	High	High	High	High	High	High	High	Infested
Fort Gibson Lake	High	High	High	High	High	High	High	Infested
Fort Peck Lake	High	Moderate	High	High	Low	Low	Moderate	Not infested
Fort Supply Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Foster Joseph Sayers Reservoir	High	High	High	High	High	High	High	Not infested
Foster Lake	High	High	Moderate	Low	Low	Low	Low	Not infested
Gainesville Lake	High	Moderate	High	High	High	Moderate	Moderate	Not infested
George B. Stevenson Reservoir	High	Moderate	High	Low	Moderate	Low	Low	Not infested
George W Andrews Lake	High	Moderate	High	Moderate	Low	Low	Moderate	Not infested
Georgetown Lake	High	High	High	High	High	High	High	Infested
Gillham Lake	High	Low	High	Low	Moderate	Low	Low	Not infested
Glenn Cunningham Lake	High	High	High	High	High	High	High	Infested
Granger Lake	High	High	High	High	High	High	High	Infested
Grapevine Lake	High	High	High	High	High	High	High	Infested
Grayson Lake	High	High	High	High	High	High	High	Not infested
Great Salt Plains Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Green Peter Lake	High	High	Moderate	Low	Low	Low	Low	Not infested
Green River Lake	High	High	High	High	High	High	High	Not infested
Greers Ferry Lake	High	Moderate	High	High	High	Moderate	Moderate	Infested
Grenada Lake	High	Low	High	High	Moderate	Low	Low	Not infested
Gull Lake	High	High	High	High	High	High	High	Infested
H. V. Eastman Lake	High	High	High	High	Low	Low	High	Not infested
Hammond Lake	High	High	High	High	High	High	High	Not infested
Harlan County Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Harry S. Truman Lake	High	High	High	High	High	High	High	Not infested
Hartwell Lake	High	Moderate	High	Low	Moderate	Low	Low	Not infested
Hensley Lake	High	High	High	High	Low	Low	High	Not infested
Heyburn Lake	High	High	High	High	High	High	High	Not infested
Highway 75 Reservoir	High	High	High	High	High	High	High	Not infested
Hills Creek Lake	High	High	Moderate	Low	Low	Low	Low	Not infested
Hillsdale Lake	High	High	High	High	High	High	High	Infested
Holmes Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Holt Lake	High	Moderate	High	High	High	Moderate	Moderate	Infested
Homme Lake	High	High	High	High	High	High	High	Not infested
Hords Creek Lake	High	High	High	High	High	High	High	Infested
Howard A. Hanson Reservoir	High	High	Moderate	Low	Low	Low	Low	Not infested
Hugo Lake	High	Moderate	High	High	Moderate	Moderate	Moderate	Not infested
Hulah Lake	High	High	High	High	High	High	High	Not infested
Isabella Lake	High	High	High	High	Low	Low	High	Not infested

J. Edward Roush Lake	High	High	High	High	High	High	High	Not infested
J. Percy Priest Lake	High	High	High	High	High	High	High	Not infested
J. Strom Thurmond Lake	High	Moderate	High	Moderate	Low	Low	Moderate	Not infested
Jennings Randolph Lake	High	Moderate	High	Moderate	Moderate	Moderate	Moderate	Not infested
Jim Chapman Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Joe Pool Lake	High	High	High	High	High	High	High	Not infested
John H. Kerr Reservoir	High	Moderate	High	High	Low	Low	Moderate	Not infested
John Martin Reservoir	High	High	High	High	Low	Low	High	Not infested
John Redmond Reservoir	High	High	High	High	High	High	High	Infested
John W. Flannagan Reservoir	High	High	High	High	High	High	High	Not infested
Kanopolis Lake	High	High	High	High	High	High	High	Infested
Kaw Lake	High	High	High	High	High	High	High	Infested
Kettle Creek Lake	High	Moderate	High	Low	Moderate	Low	Low	Not infested
Keystone Lake	High	High	High	High	High	High	High	Infested
Lac Qui Parle Lake	High	High	High	High	High	High	High	Infested
Lake Allatoona	High	Moderate	High	Moderate	Moderate	Moderate	Moderate	Not infested
Lake Ashtabula	High	High	High	High	Moderate	Moderate	High	Infested
Lake Barkley	High	High	High	High	High	High	High	Infested
Lake Bonneville	High	High	High	High	Low	Low	High	Not infested
Lake Bryan	High	High	High	High	Low	Low	High	Not infested
Lake Celilo	High	High	High	High	Low	Low	High	Not infested
Lake Clementine	High	Moderate	High	Low	Low	Low	Low	Not infested
Lake Cumberland	High	High	High	High	Moderate	Moderate	High	Infested
Lake Francis Case	High	High	High	High	High	High	High	Infested
Lake Greeson	High	Low	Moderate	Moderate	Moderate	Low	Low	Not infested
Lake Herbert G West	High	High	High	High	Low	Low	High	Not infested
Lake Kaweah	High	High	High	Moderate	Low	Low	Moderate	Not infested
Lake Koocanusa	High	High	High	High	Low	Low	High	Not infested
Lake Mendocino	High	High	High	High	Low	Low	High	Not infested
Lake Merrisach/Arkansas Post Canal/Pool 2	High	High	High	High	High	High	High	Infested
Lake Moomaw	High	High	High	Moderate	Moderate	Moderate	Moderate	Not infested
Lake O' The Pines	High	Moderate	High	High	Moderate	Moderate	Moderate	Not infested
Lake Oahe	High	High	High	High	Moderate	Moderate	High	Not infested
Lake Okeechobee	High	High	High	High	Low	Low	High	Not infested
Lake Ouachita	High	Moderate	High	High	Moderate	Moderate	Moderate	Not infested
Lake Pend Oreille	High	High	High	Moderate	Low	Low	Moderate	Not infested
Lake Ray Roberts	High	High	High	High	High	High	High	Infested
Lake Red Rock	High	High	High	High	Moderate	Moderate	High	Not infested
Lake Sacajawea	High	High	High	High	Low	Low	High	Not infested
Lake Sakakawea	High	High	High	High	Low	Low	High	Not infested
Lake Seminole	High	High	High	High	Low	Low	High	Not infested
Lake Sharpe	High	High	High	High	High	High	High	Infested

Lake Shelbyville	High	High	High	High	Moderate	Moderate	High	Not infested
Lake Sidney Lanier	High	Moderate	High	Moderate	Low	Low	Moderate	Not infested
Lake Sonoma	High	High	High	High	Low	Low	High	Not infested
Lake Success	High	High	High	High	Low	Low	High	Not infested
Lake Texoma	High	High	High	High	High	High	High	Infested
Lake Traverse	High	High	High	High	High	High	High	Not infested
Lake Umatilla	High	High	High	High	Low	Low	High	Not infested
Lake Wallula	High	High	High	High	Low	Low	High	Not infested
Lake Winnibigoshish	High	High	High	High	High	High	High	Infested
Laurel River Lake	High	Moderate	High	High	High	Moderate	Moderate	Not infested
Lavon Lake	High	High	High	High	High	High	High	Infested
Leech Lake	High	High	High	High	High	High	High	Infested
Leesville Lake	High	High	High	High	High	High	High	Not infested
Lewis and Clark Lake	High	High	High	High	High	High	High	Infested
Lewisville Lake	High	High	High	High	High	High	High	Infested
Littleville Lake	High	Moderate	Moderate	Low	High	Low	Low	Not infested
Long Branch Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Longview Lake	High	High	High	High	High	High	High	Not infested
Lookout Point Lake	High	High	Moderate	Low	Low	Low	Low	Not infested
Lost Creek Lake	High	High	High	Low	Low	Low	Low	Not infested
Lower Granite Lake	High	High	High	High	Low	Low	High	Not infested
Loyalhanna Lake	High	High	High	High	High	High	High	Not infested
Lucky Peak Lake	High	High	High	Moderate	Low	Low	Moderate	Not infested
Mahoning Creek Lake	High	High	High	High	High	High	High	Not infested
Mansfield Hollow Lake	High	Moderate	Moderate	Low	Moderate	Low	Low	Not infested
Marion Lake	High	High	High	High	High	High	High	Infested
Mark Twain Lake	High	High	High	High	High	High	High	Not infested
Marsh Lake	High	High	High	High	High	High	High	Not infested
Martins Fork Lake	High	Moderate	High	High	Moderate	Moderate	Moderate	Not infested
Melvern Lake	High	High	High	High	High	High	High	Not infested
Michael J. Kirwan Reservoir	High	High	High	High	High	High	High	Infested
Milford Lake	High	High	High	High	High	High	High	Infested
Millwood Lake	High	Moderate	High	High	Low	Low	Moderate	Not infested
Mississinewa Lake	High	High	High	High	High	High	High	Not infested
Monroe Lake	High	High	High	High	High	High	High	Not infested
Mosquito Creek Lake	High	High	High	High	High	High	High	Infested
Navarro Mills Lake	High	High	High	High	Moderate	Moderate	High	Not infested
New Hogan Lake	High	Moderate	High	High	Low	Low	Moderate	Not infested
Nimrod Lake	High	Low	High	Low	High	Low	Low	Not infested
Nolin River Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Norfork Lake	High	High	High	High	High	High	High	Not infested
North Branch of Kokosing River Lake	High	High	High	High	High	High	High	Not infested

North Fork of Pound Lake	High	High	High	High	High	High	High	Not infested
North Springfield Reservoir	High	High	High	High	Moderate	Moderate	High	Not infested
O. C. Fisher Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Okatibbee Lake	High	Low	High	High	Moderate	Low	Low	Not infested
Old Hickory Lake	High	High	High	High	High	High	High	Infested
Olive Creek Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Oliver Lake	High	Moderate	High	High	High	Moderate	Moderate	Infested
Oologah Lake	High	High	High	High	High	High	High	Infested
Optima Lake	High	High	High	High	Low	Low	High	Not infested
Orwell Lake	High	High	High	High	High	High	High	Infested
Otter Brook Lake	High	Low	Moderate	Low	Moderate	Low	Low	Not infested
Ozark Lake	High	Moderate	High	High	High	Moderate	Moderate	Infested
Paint Creek Lake	High	High	High	High	High	High	High	Not infested
Paintsville Lake	High	High	High	High	High	High	High	Not infested
Pat Mayse Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Patoka Lake	High	High	High	High	High	High	High	Not infested
Pawnee Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Perry Lake	High	High	High	High	High	High	High	Infested
Philpott Lake	High	Moderate	High	Low	Low	Low	Low	Not infested
Piedmont Lake	High	High	High	High	High	High	High	Infested
Pine Creek Lake	High	Low	High	Moderate	Moderate	Low	Low	Not infested
Pine Flat Lake	High	High	High	Low	Low	Low	Low	Not infested
Pine River Reservoir	High	High	High	High	High	High	High	Infested
Pipestem Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Pleasant Hill Lake	High	High	High	High	High	High	High	Infested
Pokegama Lake	High	High	High	High	High	High	High	Not infested
Pomme de Terre Lake	High	High	High	High	High	High	High	Not infested
Pomona Lake	High	High	High	High	High	High	High	Infested
Pool A (Amory)	High	Low	High	High	Moderate	Low	Low	Not infested
Pool B (Wilkins)	High	Low	High	High	Moderate	Low	Low	Not infested
Pool C (Fulton)	High	Low	High	High	Moderate	Low	Low	Not infested
Pool D (Rankin)	High	Low	High	High	Moderate	Low	Low	Not infested
Pool E (Montgomery)	High	Low	High	Low	High	Low	Low	Not infested
Proctor Lake	High	High	High	High	High	High	High	Not infested
Prompton Lake	High	Moderate	High	Low	High	Low	Low	Not infested
R. D. Bailey Lake	High	High	High	High	Moderate	Moderate	High	Not infested
R. E. Bob Woodruff Lake	High	Moderate	High	High	Moderate	Moderate	Moderate	Not infested
Rathbun Lake	High	High	High	High	Low	Low	High	Infested
Raystown Lake	High	High	High	High	High	High	High	Not infested
Rend Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Richard B. Russell Lake	High	Moderate	Moderate	Moderate	Low	Low	Moderate	Not infested
Robert S. Kerr Lake	High	High	High	High	High	High	High	Infested

Rough River Lake	High	High	High	High	High	High	High	Not infested
Rufus Woods Lake	High	High	High	High	Low	Low	High	Not infested
Salamonie Lake	High	High	High	High	High	High	High	Not infested
Sam Rayburn Reservoir	High	Moderate	High	High	Moderate	Moderate	Moderate	Not infested
Santa Rosa Reservoir	High	High	High	High	Low	Low	High	Not infested
Sardis Lake Tulsa	High	Moderate	High	Moderate	Moderate	Moderate	Moderate	Not infested
Sardis Lake Vicksburg	High	Low	High	High	Moderate	Low	Low	Not infested
Savage River Reservoir	High	Low	High	High	Moderate	Low	Low	Not infested
Saylorville Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Senecaville Lake	High	High	High	High	High	High	High	Infested
Shenango River Lake	High	High	High	High	High	High	High	Not infested
Skiatook Lake	High	High	High	High	High	High	High	Infested
Smithville Lake	High	High	High	High	High	High	High	Infested
Somerville Lake	High	High	High	High	High	High	High	Not infested
Stagecoach Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Standing Bear Lake	High	High	High	High	High	High	High	Not infested
Steinhagen Lake	High	Moderate	High	Low	Moderate	Low	Low	Not infested
Stillhouse Hollow Lake	High	High	High	High	High	High	High	Infested
Stillwater Lake	High	Moderate	High	Moderate	High	Moderate	Moderate	Not infested
Stockton Lake	High	High	High	High	High	High	High	Not infested
Stonewall Jackson Lake	High	High	High	High	High	High	High	Not infested
Summersville Lake	High	Moderate	High	High	High	Moderate	Moderate	Not infested
Surry Mountain Lake	High	Low	Moderate	Low	Moderate	Low	Low	Not infested
Sutton Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Table Rock Lake	High	High	High	High	High	High	High	Not infested
Tappan Lake	High	High	High	High	High	High	High	Not infested
Taylorsville Lake	High	High	High	High	High	High	High	Not infested
Tenkiller Ferry Lake	High	High	High	High	High	High	High	Not infested
Tioga Lake	High	High	High	High	High	High	High	Not infested
Tionesta Lake	High	Moderate	High	Moderate	High	Moderate	Moderate	Not infested
Toronto Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Townshend Reservoir	High	Moderate	High	Low	Moderate	Low	Low	Not infested
Trinidad Reservoir	High	High	High	High	Moderate	Moderate	High	Not infested
Tuttle Creek Lake	High	High	High	High	High	High	High	Infested
Twin Lakes	High	High	High	High	Moderate	Moderate	High	Not infested
Tygart Lake	High	Moderate	High	High	High	Moderate	Moderate	Not infested
Upper & Lower Red Lake	High	High	High	High	High	High	High	Infested
W. Kerr Scott Reservoir	High	Moderate	High	Low	Moderate	Low	Low	Not infested
Waco Lake	High	High	High	High	High	High	High	Infested
Wagon Train Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Wallace Lake	High	Moderate	High	High	Low	Low	Moderate	Not infested
Walter F. George Lake	High	Moderate	High	Moderate	Low	Low	Moderate	Not infested

Wappapello Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Warrior Lake	High	Moderate	High	High	High	Moderate	Moderate	Infested
Water Conservation Area Combined	High	High	High	High	Low	Low	High	Not infested
Waurika Lake	High	High	High	High	Moderate	Moderate	High	Infested
Wehrspann Lake	High	High	High	High	High	High	High	Not infested
West Fork Lake	High	High	High	High	High	High	High	Not infested
West Point Lake	High	Moderate	High	Low	Low	Low	Low	Not infested
West Thompson Lake	High	Moderate	Moderate	Low	Moderate	Low	Low	Not infested
White River	High	High	High	High	High	High	High	Not infested
Whitney Lake	High	High	High	High	High	High	High	Not infested
Whitney Point Lake	High	High	High	High	High	High	High	Not infested
William Dannelly Reservoir	High	High	High	High	Moderate	Moderate	High	Not infested
William H Harsha Lake	High	High	High	High	High	High	High	Not infested
Wills Creek Lake	High	High	High	High	High	High	High	Not infested
Wilson Lake	High	High	High	High	Moderate	Moderate	High	Infested
Wister Lake	High	Moderate	High	Moderate	High	Moderate	Moderate	Not infested
Woodcock Creek Lake	High	High	High	High	High	High	High	Not infested
Wright Patman Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Yankee Hill Lake	High	High	High	High	Moderate	Moderate	High	Not infested
Yatesville Lake	High	High	High	High	High	High	High	Not infested
Youghiogheny River Lake	High	Moderate	High	Moderate	High	Moderate	Moderate	Not infested
Zorinksy Lake	High	High	High	High	High	High	High	Infested

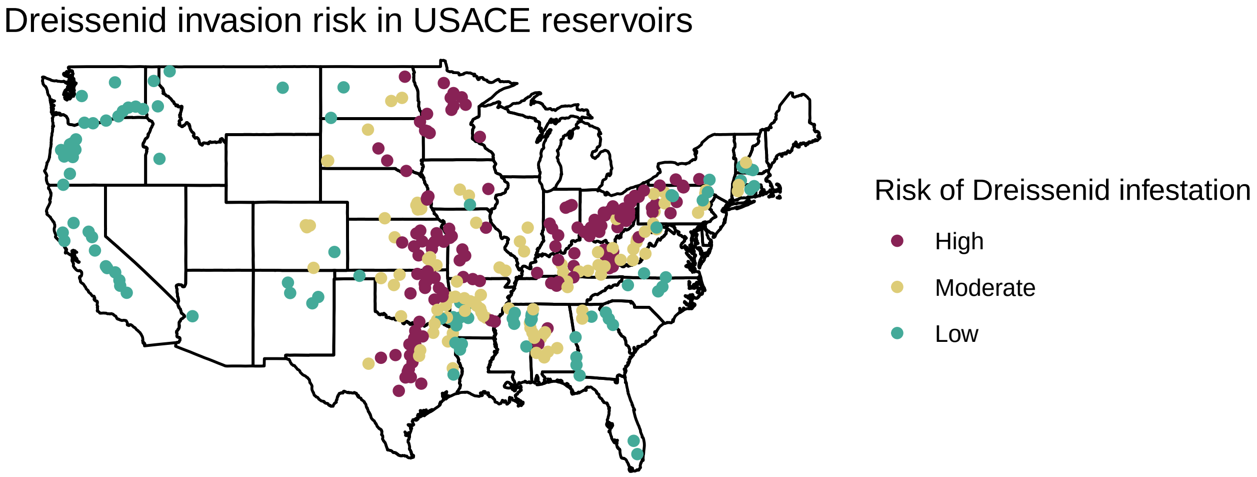
Dreissenid invasion risk in USACE reservoirs
Risk of Dreissenid infestation
High
Moderate
Low

Figure 2. Locations of low, moderate, and high risk reservoirs for dreissenid infestation. Risk was based on calcium levels, mean summer temperature, total water hardness, water pH, and distance to the nearest infestation

Since distance to infestation is likely to change over time as Dreissenids become established in new locations, we also assessed whether risk status changed if distance was not considered as a risk factor. Not considering distance resulted in 30 reservoirs with a low overall risk being elevated to high risk, and an additional 20 being elevated to moderate risk. Additionally, 58 reservoirs that were initially considered moderate risk based on distance to infestation were elevated to high risk.

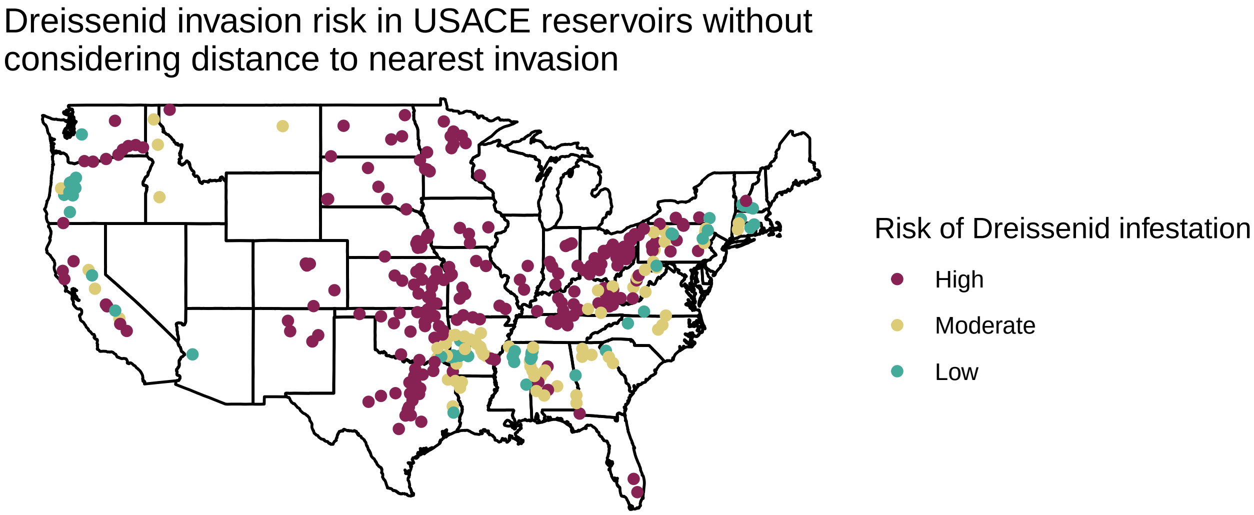
Dreissenid invasion risk in USACE reservoirs without considering distance to nearest invasion
Risk of Dreissenid infestation
● High
● Moderate
● Low

Figure 3. Location of low, moderate, and high risk reservoirs for Dreissenid infestation when not considering distance to nearest infestation as a risk factor. Risk was based on mean summer temperature, total water hardness, calcium concentration, and pH. Low risk reservoirs are classified as such by potentially unsuitable water chemistry. One reservoir in Arizona is considered unsuitable because of summer temperatures.

Table 4: Number of Dreissenid infested and un-infested reservoirs characterized as low, moderate, or high risk of invasion based on environmental characteristics.

Environmental variable	Infestation status	Low Risk	Moderate Risk	High Risk
Mean temperature	Infested	0	0	84
	Not infested	1	0	267
pH	Infested	0	14	70
	Not infested	21	65	182
Hardness	Infested	0	0	84
	Not infested	1	24	243
Calcium	Infested	0	0	84
	Not infested	40	31	197
Distance	Infested	1	5	78
	Not infested	69	97	102

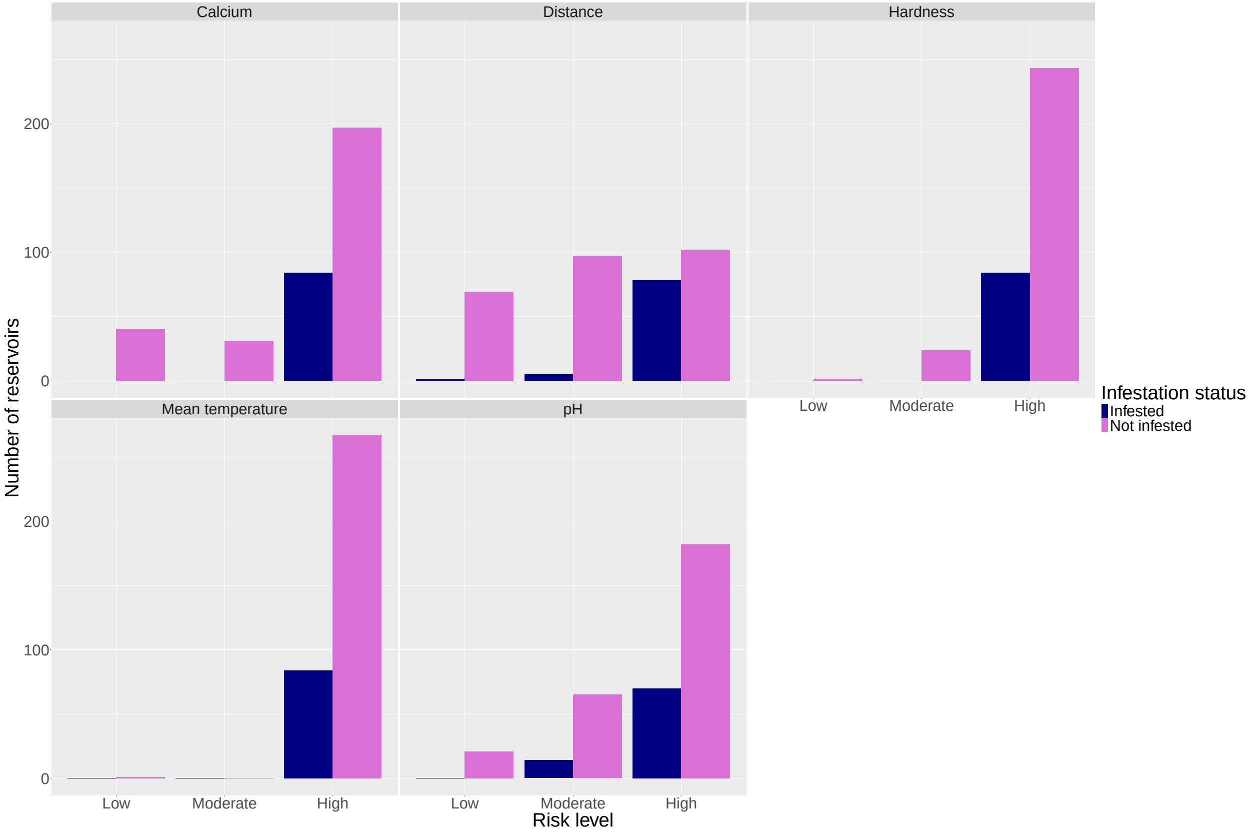

Figure 4. Number of Dreissenid infested and un-infested USACE reservoirs falling into each risk category (low, moderate, or high) for specified variables.

Table 5: Percent of USACE reservoirs falling into each risk category for Dreissenid infestation.

Low risk (%)	Moderate risk (%)	High risk (%)
29.26	30.97	39.77

Table 6: Number of Dreissenid infested or un-infested reservoirs categorized as low, moderate, or high risk for Dreissenid invasion based on factors including calcium concentration, mean summer temperature, pH, hardness, and distance to other infested water bodies.

Infestation status	Low risk	Moderate risk	High risk
Infested	1	18	65
Not infested	102	91	75

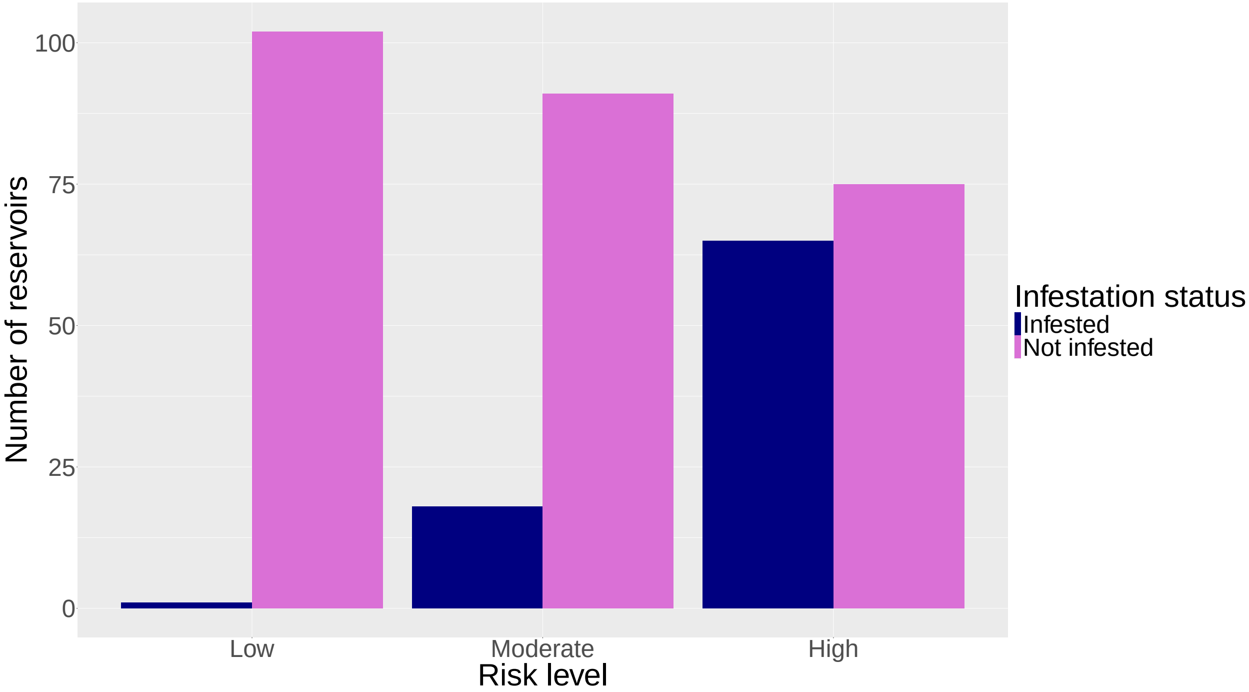

Figure 5. Number of USACE reservoirs falling into each overall risk category that are infested or un-infested with Dreissenids.

Based on our risk categories, approximately 46% of high risk reservoirs have already recorded Dreissenid presence. Approximately 16% of moderate risk reservoirs have also recorded Dreissenid presence. These reservoirs were ranked as moderate risk based on pH (n=13), pH and distance (n=1), or distance alone (n=4). One low risk reservoir, lake Ruthbun, is infested. Lake Rathbun is classified as low risk due to distance to nearest infestation. It is 128 km from the nearest Dreissenid observation.

Most of the reservoirs limited by pH (e.g. moderate risk but still infested) were located in the Arkansas River drainage or in the Black-Warrior Tombigbee system in Alabama. One pH-limited and distance reservoir was located in Pennsylvania (Curwensville Lake). The five distance limited reservoirs (e.g. low or moderate risk but still infested) were in Iowa (Rothbun lake - low risk), North Dakota (Lake Ashtabula - moderate risk), Oklahoma (Waurika Lake - moderate risk), Kentucky (Lake Cumberland - moderate risk), Kansas (Wilson Lake - moderate risk), and Pennsylvania (Curwensville Lake - moderate risk). With the exception of Curwensville Lake (Pennsylvania), Oliver Lake (Alabama), and Rathbun Lake (Iowa), all invaded moderate risk reservoirs have established Dreissenid populations, with some having been invaded as early as the 1990s.*

The fact that pH limited reservoirs are infested suggest that 1) more fine scale environmental data may be required to understand how intra-reservoir variation in environmental characteristics such as underlying geology could contribute to suitable conditions in reservoirs, or 2) that Dreissenids may be more tolerant of pH than expected. For the three remaining lakes where invasion status is unknown but positive Dreissenid records have occurred, it is possible that Dreissenids have been introduced to these reservoirs but have not been able to establish strong populations due to sub-optimal environmental conditions. For reservoirs with a low or moderate risk of invasion based on distance to the next nearest positive record, it is possible that Dreissenids were transported long distances by boaters given that few source populations were present nearby to fuel population expansion into these areas. Information on recreational boater movement is not systematically collected, and thus there is a need to further study how boaters move and how they affect the risk of infestation.

Table 7: Reservoirs with positive Dreissenid records despite being considered at moderate risk of Dreissenid invasion.

Reservoir	**Mean summer temp risk**	**pH risk**	**Hardness risk**	**Calcium risk**	**Distance risk**	**Overall risk**	**Overall risk without distance**	**Infestation status**
Arkansas River Pool 3	High	Moderate	High	High	High	Moderate	Moderate	Infested
Arkansas River Pool 5	High	Moderate	High	High	High	Moderate	Moderate	Infested
Arkansas River Pool 6	High	Moderate	High	High	High	Moderate	Moderate	Infested
Arkansas River Pool 7	High	Moderate	High	High	High	Moderate	Moderate	Infested
Arkansas River Pool 8	High	Moderate	High	High	High	Moderate	Moderate	Infested
Arkansas River Pool 9	High	Moderate	High	High	High	Moderate	Moderate	Infested
Arkansas River Pool 13	High	Moderate	High	High	High	Moderate	Moderate	Infested
Curwensville Lake	High	Moderate	High	High	Moderate	Moderate	Moderate	Infested
Dardanelle Lake	High	Moderate	High	High	High	Moderate	Moderate	Infested
Greers Ferry Lake	High	Moderate	High	High	High	Moderate	Moderate	Infested
Holt Lake	High	Moderate	High	High	High	Moderate	Moderate	Infested
Lake Ashtabula	High	High	High	High	Moderate	Moderate	High	Infested
Lake Cumberland	High	High	High	High	Moderate	Moderate	High	Infested
Oliver Lake	High	Moderate	High	High	High	Moderate	Moderate	Infested
Ozark Lake	High	Moderate	High	High	High	Moderate	Moderate	Infested
Rathbun Lake	High	High	High	High	Low	Low	High	Infested
Warrior Lake	High	Moderate	High	High	High	Moderate	Moderate	Infested
Waurika Lake	High	High	High	High	Moderate	Moderate	High	Infested
Wilson Lake	High	High	High	High	Moderate	Moderate	High	Infested

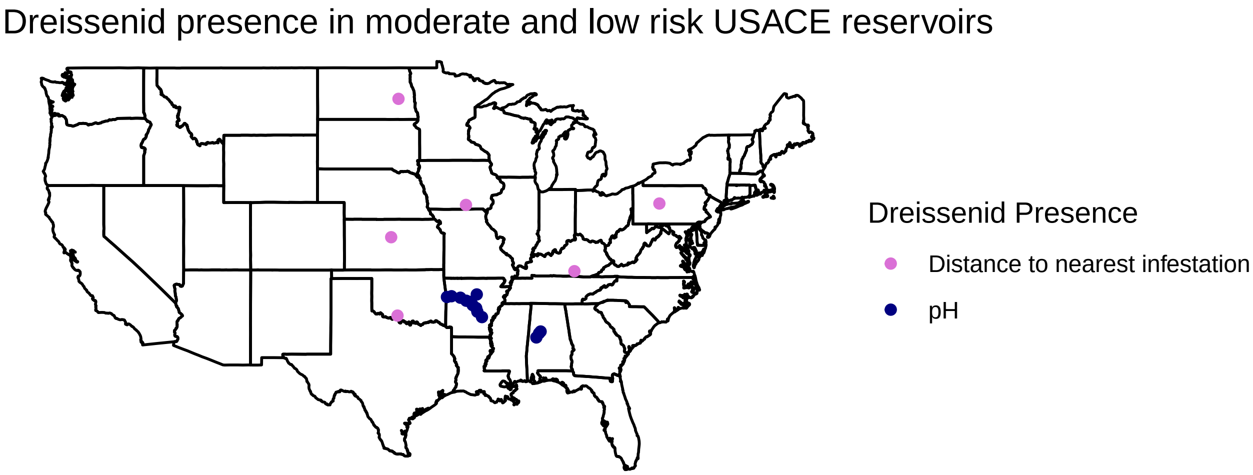
Dreissenid presence in moderate and low risk USACE reservoirs
Dreissenid Presence
● Distance to nearest infestation
● pH

Figure 6. Location of USACE reservoirs with moderate invasion risk based on pH or distance to nearest infestation that have positive Dreissenid occurrences according to NAS (Benson et al. 2024).

Data analysis

Next, we used the data available on a national scale to try to better understand how different environmental and social factors may be related to Dreissenid presence in USACE reservoirs. We explored several analysis methods to investigate our questions, which are outlined below:

Correlation Matrices

Before we used any predictive models to evaluate the risk of dreissenid mussel invasion, we wanted to understand how correlated different variables are. When variables are correlated they may not be suited for certain types of modeling, like logistic regression. In the below figures, bigger, darker blue dots indicate that two variables are positively correlated, while bigger, darker red dots indicate a negative correlation.

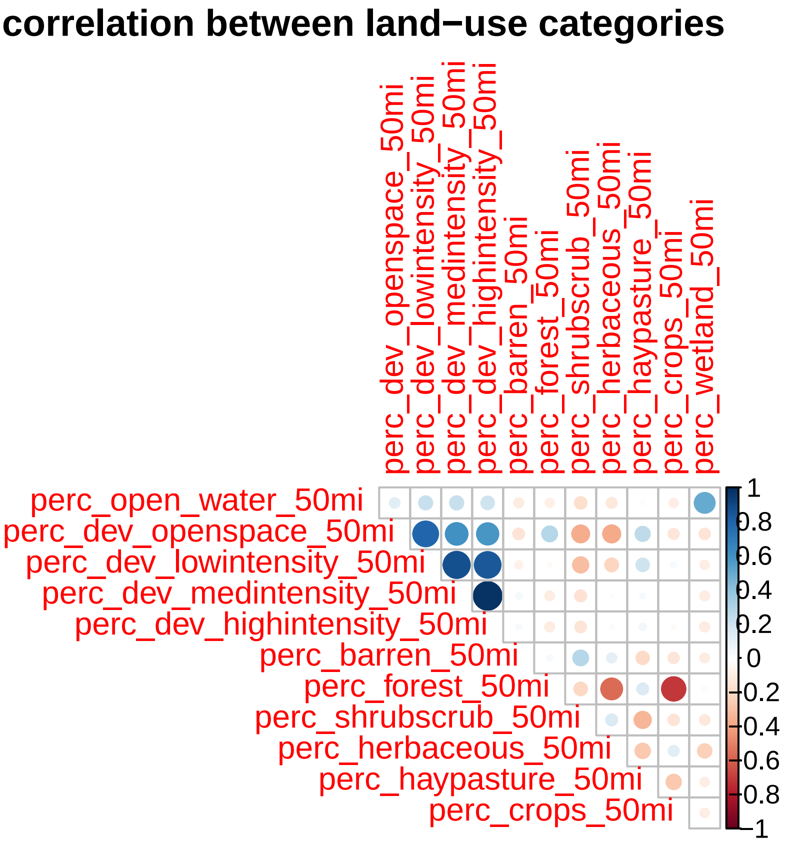

Figure 7. A correlation plot showing how correlated different land-use variables are. All urban land-use categories are positively correlated, while forested land is negatively correlated with herbaceous cover and crop cover.

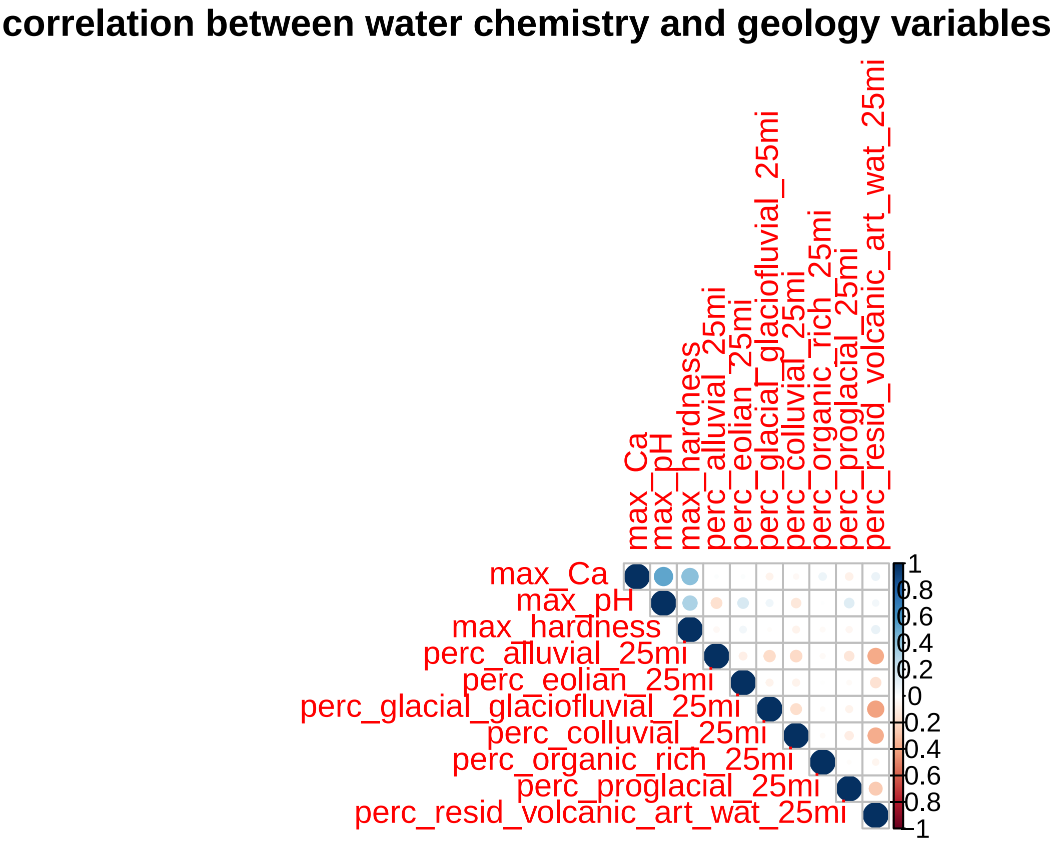

Figure 8. A correlation plot showing how correlated different water chemistry and geology variables are. There are no strong correlations among these variables.

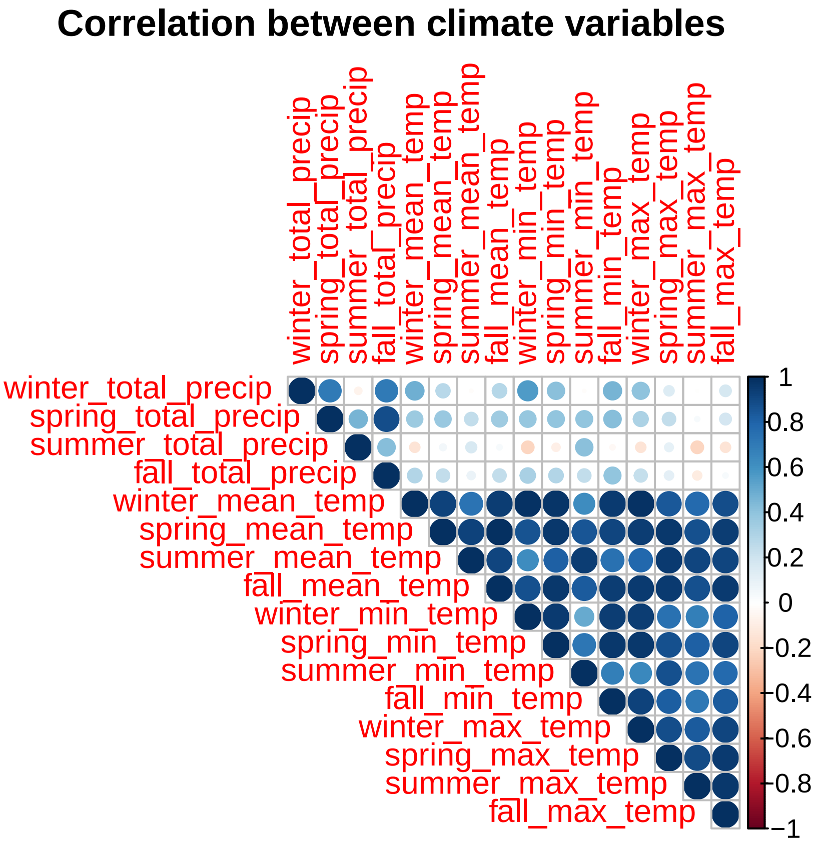

Figure 9. A correlation plot showing how correlated different climate variables are. Temperatures are all very correlated.

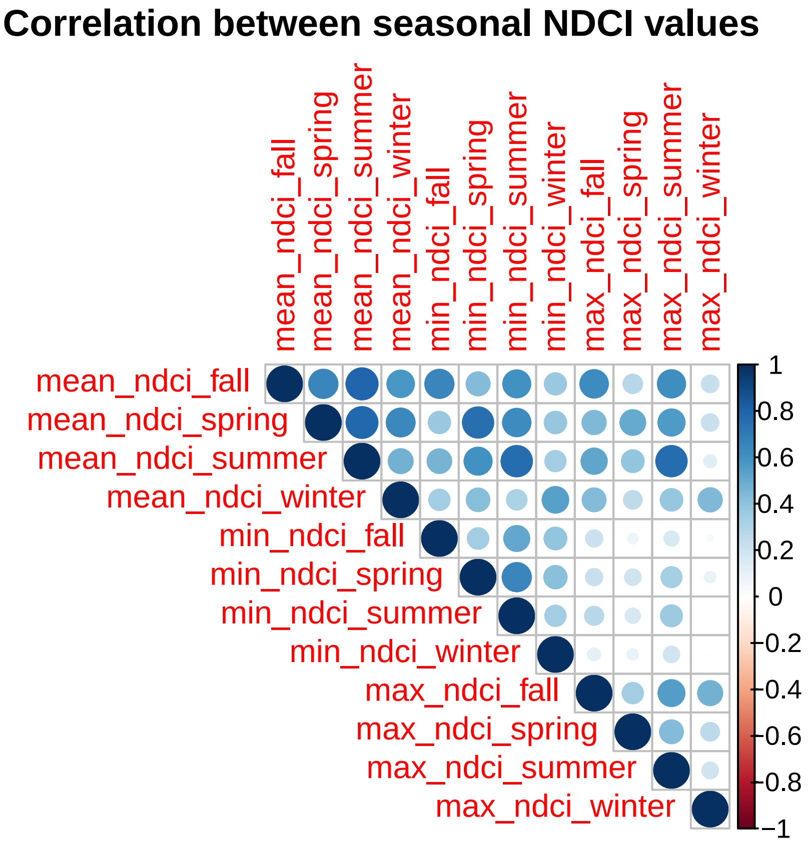

Figure 10. A correlation plot showing how correlated seasonal NDCI variables are. Seasonal mean NDCI values are somewhat correlated with all NDCI variables.

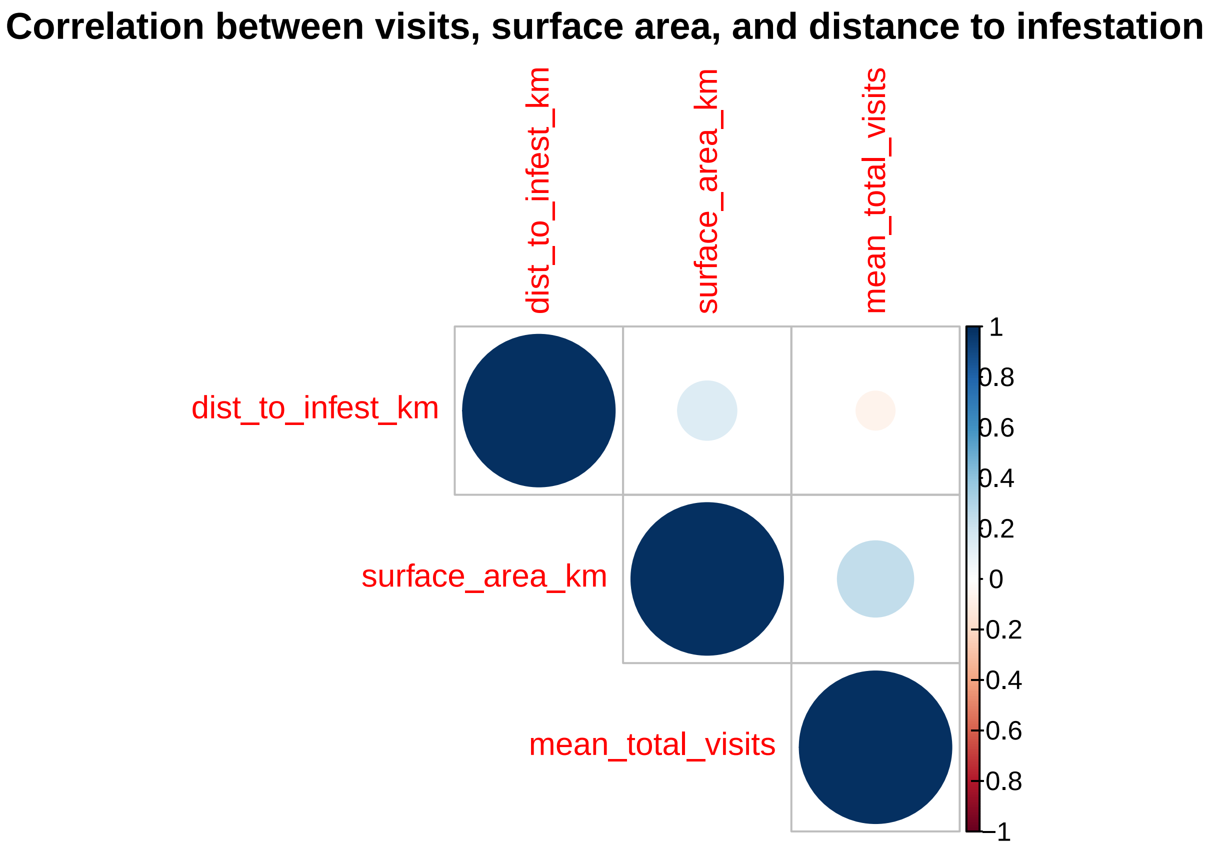

Figure 11. A correlation plot showing how correlated surface area, distance to infestation, and visitation variables are. There are no strong correlations between the three variables.

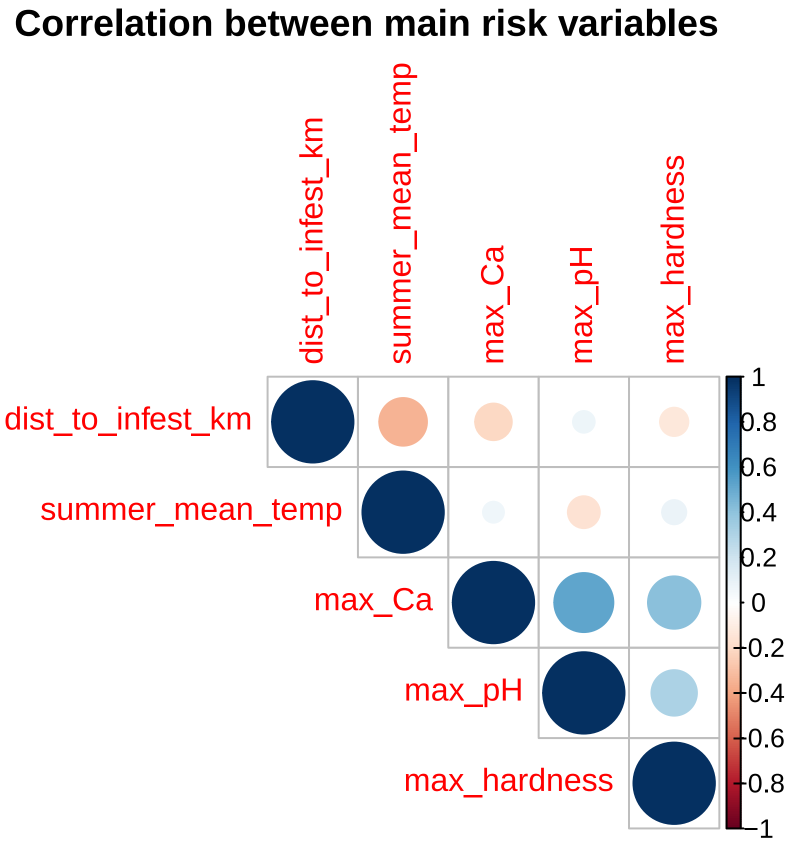

Figure 12. A correlation plot showing how correlated the variables that determine our risk categories are. There are no strong correlations among these variables

Logistic regressions

We ran several exploratory logistic regression models using binary infested (1) versus not infested (0) as the outcome variable. We used different input variables in each logistic regression. Our goal was not to predict Dreissenid infestation, but rather to identify significant environmental predictors and the directionality of their influence.

We used the Akaike Information Criterion or AIC value to evaluate the quality of each model. The AIC value looks at the predictive power of the model as well as the number of inputs in each model. Generally, fewer inputs are better, because it reduces the chance of model over-fitting. There is no goal AIC value, rather comparatively lower AIC values indicate a better quality model compared to models with higher AIC values.

To assess the significance of each input we used the p value. The smaller the p value, the lower the chance that the input variable has no influence on the response variable. A p value of under 0.05 is indicated by a period, and anything smaller than that is indicated by one or more asterisks.

To evaluate the directionality of influence, we examined the coefficient value. The coefficient value indicates the change in log odds that the response variable will be 1 if the input variable increases by one unit. In our case, it indicates the log odds that a reservoir will be infested if the predictor variable is increased by one unit. Coefficients should only be evaluated if the model is of good quality and if the p value for the input value is very low.

Landcover

We ran a logistic regression with the percent of land dedicated to each land-use category within a 25 and 50 mile buffer of each reservoir as input variables. Data was from the NLCD.

First we ran the model with a 50 mile buffer for each land cover category.

Note that the model did not converge with percent perennial snow and ice included in the logistic regression.

Below is the model summary. The coefficient name (or input variable name) is in the first column on the left. The next column over, called `Estimate`, contains the coefficient value with tells you the directionality of influence that the coefficient has on the outcome variable. The column with the p value is on the far right. the AIC value is stated below the table output.

Logistic Regression 1

```
##
## Call:
## glm(formula = infest_status ~ ., family = "binomial", data = landcover_50)
##
## Coefficients:
##                              Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  -28.6349    23.9705  -1.195    0.232
## perc_open_water_50mi           0.4004     0.2559   1.565    0.118
## perc_dev_openspace_50mi        0.2799     0.2912   0.961    0.336
## perc_dev_lowintensity_50mi     0.4439     0.3335   1.331    0.183
## perc_dev_medintensity_50mi    -0.5109     0.7684  -0.665    0.506
## perc_dev_highintensity_50mi    1.8672     1.3501   1.383    0.167
## perc_barren_50mi              -0.1246     0.5992  -0.208    0.835
## perc_forest_50mi               0.2531     0.2388   1.060    0.289
## perc_shrubscrub_50mi           0.2536     0.2405   1.055    0.292
## perc_herbaceous_50mi           0.2866     0.2393   1.198    0.231
## perc_haypasture_50mi           0.3022     0.2403   1.258    0.208
## perc_crops_50mi                0.2729     0.2394   1.140    0.254
## perc_wetland_50mi              0.2894     0.2441   1.185    0.236
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 386.85  on 351  degrees of freedom
## Residual deviance: 343.14  on 339  degrees of freedom
## AIC: 369.14
##
## Number of Fisher Scoring iterations: 7
```

Notes: In this model, there are no significant predictors of infestation.

We then ran the model with a 25 mile buffer. Note that the model also did not converge with percent open water or percent snow and ice variables included.

Logistic Regression 2

```
##
## Call:
## glm(formula = infest_status ~ ., family = "binomial", data = landcover_25)
##
## Coefficients:
```

```
##                             Estimate Std. Error z value Pr(>|z|)    
## (Intercept)                 16.35570    5.95379   2.747 0.006012 ** 
## perc_dev_openspace_25mi     -0.11981    0.12985  -0.923 0.356156    
## perc_dev_lowintensity_25mi  -0.05621    0.17323  -0.325 0.745558    
## perc_dev_medintensity_25mi  -0.40367    0.37247  -1.084 0.278465    
## perc_dev_highintensity_25mi  0.12076    0.62306   0.194 0.846317    
## perc_barren_25mi             0.82453    0.47925   1.720 0.085350 .  
## perc_forest_25mi            -0.20381    0.06132  -3.324 0.000889 ***
## perc_shrubscrub_25mi        -0.22602    0.06811  -3.318 0.000905 ***
## perc_herbaceous_25mi        -0.16481    0.06109  -2.698 0.006978 ** 
## perc_haypasture_25mi        -0.15367    0.06211  -2.474 0.013359 *  
## perc_crops_25mi             -0.18519    0.06086  -3.043 0.002342 ** 
## perc_wetland_25mi           -0.18851    0.07237  -2.605 0.009191 ** 
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
## 
## (Dispersion parameter for binomial family taken to be 1)
## 
##     Null deviance: 386.85  on 351  degrees of freedom
## Residual deviance: 335.87  on 340  degrees of freedom
## AIC: 359.87
## 
## Number of Fisher Scoring iterations: 6
```

Notes: When we ran the model with a 25 mile buffer, it showed barren land, forest cover, shrub cover, herbaceous cover, hay pasture cover, crop cover, and wetland cover to be significant predictors of infestation. Barren land has a slight positive influence on infestation, while all other significant variables are negatively associated with infestation. The model did not show any development categories to be significant. While this model does have a lower AIC value than the previous model, it still is not very low, indicating that this model is not performing particularly well.

In the next model, we aggregate all development classifications into a single category, again using the 25 mile buffer.

Logistic Regression 3

```
## 
## Call:
## glm(formula = infest_status ~ ., family = "binomial", data = dev_sum)
## 
## Coefficients:
##                      Estimate Std. Error z value Pr(>|z|)    
## (Intercept)          15.79985    5.87562   2.689 0.007166 ** 
## dev_sum_25mi         -0.14177    0.06223  -2.278 0.022724 *  
## perc_barren_25mi      0.77166    0.46505   1.659 0.097054 .  
## perc_forest_25mi     -0.19608    0.05988  -3.275 0.001058 ** 
## perc_shrubscrub_25mi -0.22305    0.06769  -3.295 0.000983 ***
## perc_herbaceous_25mi -0.16006    0.06032  -2.654 0.007966 ** 
## perc_haypasture_25mi -0.14483    0.06087  -2.379 0.017354 *  
## perc_crops_25mi      -0.17793    0.05971  -2.980 0.002881 ** 
## perc_wetland_25mi    -0.18231    0.07143  -2.552 0.010702 *  
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
## 
## (Dispersion parameter for binomial family taken to be 1)
```

```
##
##     Null deviance: 386.85  on 351  degrees of freedom
## Residual deviance: 336.80  on 343  degrees of freedom
## AIC: 354.8
##
## Number of Fisher Scoring iterations: 6
```

Notes: The model showed all variables to be significant. Percent barren land has a postive relationship with infestation, while all other variables have a negative relationship with infestation. Of all land use logistic regressions, this one has the lowest AIC score.

Water chemistry and geology

Next we ran a logistic regression with various water chemistry and geology characteristics as predictors. Note that many water chemistry variables are likely correlated, which is not ideal for logistic regression. This model is for exploratory purposes only. Below is the model summary.

Logistic Regression 4

```
##
## Call:
## glm(formula = infest_status ~ ., family = "binomial", data = water_chem)
##
## Coefficients:
##                                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)                     -7.847e+02  5.125e+04  -0.015   0.9878
## max_Ca                           4.953e-03  2.998e-03   1.652   0.0985 .
## max_pH                           9.714e-01  3.795e-01   2.560   0.0105 *
## max_hardness                    -3.759e-05  6.256e-04  -0.060   0.9521
## perc_alluvial_25mi               7.764e+00  5.125e+02   0.015   0.9879
## perc_eolian_25mi                 7.754e+00  5.125e+02   0.015   0.9879
## perc_glacial_glaciofluvial_25mi  7.755e+00  5.125e+02   0.015   0.9879
## perc_colluvial_25mi              7.753e+00  5.125e+02   0.015   0.9879
## perc_organic_rich_25mi           7.755e+00  5.126e+02   0.015   0.9879
## perc_proglacial_25mi             7.745e+00  5.125e+02   0.015   0.9879
## perc_resid_volcanic_art_wat_25mi  7.759e+00  5.125e+02   0.015   0.9879
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 386.85  on 351  degrees of freedom
## Residual deviance: 356.49  on 341  degrees of freedom
## AIC: 378.49
##
## Number of Fisher Scoring iterations: 13
```

Notes: This model seems to indicate that pH and calcium content are statistically significant predictors of infestation. Increases in pH and calcium are associated with increases in the risk of infestation. This model has a comparatively higher AIC score, indicating poor model performance.

Climate

We also ran a logistic regression with various climate variables as predictors. Note that climate variables are likely all correlated, which is not ideal for logistic regression. This model is for exploratory purposes only.

Logistic Regression 5

```
##
## Call:
## glm(formula = infest_status ~ ., family = "binomial", data = climate)
##
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)         -4.471e+00  9.365e+00  -0.477 0.633050
## winter_total_precip  9.624e-03  5.576e-03   1.726 0.084373 .
## spring_total_precip  2.279e-05  8.894e-03   0.003 0.997955
## summer_total_precip -1.341e-02  6.011e-03  -2.231 0.025685 *
## fall_total_precip   -1.767e-02  9.292e-03  -1.902 0.057151 .
## winter_mean_temp    -6.034e+00  2.448e+00  -2.464 0.013722 *
## spring_mean_temp     6.235e+00  2.011e+00   3.101 0.001931 **
## summer_mean_temp    -7.187e-01  1.917e+00  -0.375 0.707684
## fall_mean_temp       4.972e+00  2.370e+00   2.098 0.035876 *
## winter_min_temp      2.918e+00  1.193e+00   2.446 0.014461 *
## spring_min_temp     -4.115e+00  1.161e+00  -3.544 0.000393 ***
## summer_min_temp      2.272e-01  9.379e-01   0.242 0.808576
## fall_min_temp       -1.211e-01  8.714e-01  -0.139 0.889505
## winter_max_temp      1.570e+00  9.311e-01   1.686 0.091739 .
## spring_max_temp     -3.186e+00  1.027e+00  -3.103 0.001915 **
## summer_max_temp      6.417e-01  1.211e+00   0.530 0.596288
## fall_max_temp       -1.927e+00  1.041e+00  -1.851 0.064148 .
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 386.85  on 351  degrees of freedom
## Residual deviance: 287.22  on 335  degrees of freedom
## AIC: 321.22
##
## Number of Fisher Scoring iterations: 7
```

Notes: This model shows many temperature variables as being significant, and also summer and winter precipitation. The AIC value for this model is comparatively lower than models we examined previously.

We also ran a logistic regression with just total precipitation and summer mean temperature, to simplify the number of variables and eliminate highly correlated variables.

Logistic Regression 6

```
##
## Call:
## glm(formula = infest_status ~ ., family = "binomial", data = mean_temp)
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -4.0045472  1.0938661  -3.661 0.000251 ***
## total_precip     -0.0006516  0.0003993  -1.632 0.102715
## summer_mean_temp  0.1465910  0.0449926   3.258 0.001122 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 386.85  on 351  degrees of freedom
## Residual deviance: 374.23  on 349  degrees of freedom
## AIC: 380.23
##
## Number of Fisher Scoring iterations: 4
```

Notes: When looking at total precipitation and mean summer temperature, only mean summer temperature is significant. This model has a higher AIC value compared to the previous climate model.

Size, recreation, and connectivity

In our next logistic regression, we looked at reservoir size (surface area), connectivity to other reservoirs, and the number of visitors.

Logistic Regression 7

```
##
## Call:
## glm(formula = infest_status ~ ., family = "binomial", data = connectivity)
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)              5.536e-01  2.856e-01   1.938   0.0526 .
## dist_to_infest_km       -5.588e-02  8.144e-03  -6.861 6.82e-12 ***
## surface_area_km          3.274e-03  1.541e-03   2.125   0.0336 *
## connectivityDam         -1.165e+00  9.439e-01  -1.235   0.2170
## connectivityLock and Dam  8.350e-01  4.958e-01   1.684   0.0921 .
## mean_total_visits        9.962e-08  2.066e-07   0.482   0.6296
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 367.75  on 333  degrees of freedom
## Residual deviance: 232.76  on 328  degrees of freedom
##   (18 observations deleted due to missingness)
## AIC: 244.76
##
## Number of Fisher Scoring iterations: 9
```

Notes: distance to infestation is the most significant predictor, and surface area is also significant. Increasing distance has a negative relationship with infestation while increasing surface area has a slight positive relationship with infestation. Connectivity is also significant. Reservoirs with a lock and dam are more likely to be infested than those with just a dam or no connection. This model has the lowest AIC so far, indicating better predictive power when compared with the previous logistic regression models.

Previously identified risk variables

Next we will run a logistic regression using all the variables we used to classify risk. These are variables that have been identified as potentially limiting to Dreissenid mussels in the literature. The variables are distance to nearest infestation, mean hardness, average calcium content, average pH, and mean summer temperature.

Logistic Regression 8

```
##
## Call:
## glm(formula = infest_status ~ ., family = "binomial", data = risk_vars)
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -9.911954   4.328969  -2.290  0.02204 *
## dist_to_infest_km -0.055244   0.008089  -6.829 8.53e-12 ***
## summer_mean_temp   0.183406   0.060080   3.053  0.00227 **
## max_Ca             0.004862   0.003623   1.342  0.17959
## max_pH             0.729067   0.500918   1.455  0.14554
## max_hardness       0.001348   0.001494   0.902  0.36697
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 386.85  on 351  degrees of freedom
## Residual deviance: 233.97  on 346  degrees of freedom
## AIC: 245.97
##
## Number of Fisher Scoring iterations: 9
```

Notes: Distance to infestation is significant and negatively associated with infestation. Mean summer temperature is significant and positively associated with infestation. All other input variables are not significant. This model has a low AIC, but it is slightly higher than the size, recreation, and connectivity logistic model.

Gradient Boosted Model

A gradient boosted model (GBM) is a flexible machine learning algorithm that is immune to issues caused by multicollinearity. It works by combining multiple weak models (think small decision trees) into one strong model.

In this instance, we split the data into training data to train the model and testing data to test the model. Since there is so little data, taking even a few observations away to later test the model may impact the model training.

We made the following modifications to the data to reduce the number of inputs.

- We used percent land-use within 25 miles and dropped percent land-use within 50 miles. We also combined all development categories into one category.
- We calculated total annual precipitation and used that and summer mean temperature from the climate data.
- We used only summer mean NDCI

- We used max calcium, max pH, and max hardness as water chemistry inputs

Below is a summary of the model, trained using the training data.

Gradient Boosted Model 1

```
## gbm(formula = infest_status ~ ., distribution = "bernoulli",
##     data = train_split, n.trees = 1000, interaction.depth = 4,
##     shrinkage = 0.01)
## A gradient boosted model with bernoulli loss function.
## 1000 iterations were performed.
## There were 28 predictors of which 26 had non-zero influence.
```

Below is a list of the most influential variables, according to the gradient boosted model. Variables with a higher `rel.inf`are more influential in the model outcome. Note that the influence value is meant to be interpreted relative to other influence values. Even though predictions generated using GBM models are not affected by multicollinearity, if two variables are highly correlated, the influence will be split between them.

```
##                                                                var      rel.inf
## dist_to_infest_km                                dist_to_infest_km 33.394228414
## max_Ca                                                      max_Ca  9.035128418
## surface_area_km                                    surface_area_km  7.210721680
## summer_mean_temp                                  summer_mean_temp  4.923495140
## perc_barren_25mi                                  perc_barren_25mi  4.674058332
## mean_total_visits                                mean_total_visits  4.142491482
## perc_herbaceous_25mi                          perc_herbaceous_25mi  3.625223924
## perc_open_water_25mi                          perc_open_water_25mi  3.216159585
## perc_haypasture_25mi                          perc_haypasture_25mi  3.202106787
## mean_ndci_summer                                  mean_ndci_summer  3.044001260
## perc_shrubscrub_25mi                          perc_shrubscrub_25mi  2.945639873
## max_pH                                                      max_pH  2.581286958
## perc_crops_25mi                                    perc_crops_25mi  2.330582737
## mean_slope                                              mean_slope  2.300751314
## total_precip                                          total_precip  2.053904295
## mean_elev_m                                            mean_elev_m  2.016660057
## perc_resid_volcanic_art_wat_25mi perc_resid_volcanic_art_wat_25mi  1.909936609
## dev_sum_25mi                                          dev_sum_25mi  1.754394237
## perc_forest_25mi                                  perc_forest_25mi  1.684790511
## perc_alluvial_25mi                              perc_alluvial_25mi  1.175634198
## perc_glacial_glaciofluvial_25mi   perc_glacial_glaciofluvial_25mi  1.038232303
## perc_colluvial_25mi                            perc_colluvial_25mi  0.971640049
## num_connections                                    num_connections  0.375657763
## max_hardness                                          max_hardness  0.353293086
## connectivity                                          connectivity  0.033091476
## perc_proglacial_25mi                          perc_proglacial_25mi  0.006889512
## perc_eolian_25mi                                  perc_eolian_25mi  0.000000000
## perc_organic_rich_25mi                      perc_organic_rich_25mi  0.000000000
```

Note: Distance to infestation is vastly more influential than other predictors.

It is also possible to see the general relationship between the predictor variables and the outcome variable. Below are figures showing the relationship between infestation on the Y axis (recall 0 indicates not infested and 1 indicates infested) and the predictor variable on the X axis. Note the different axis values between plots.

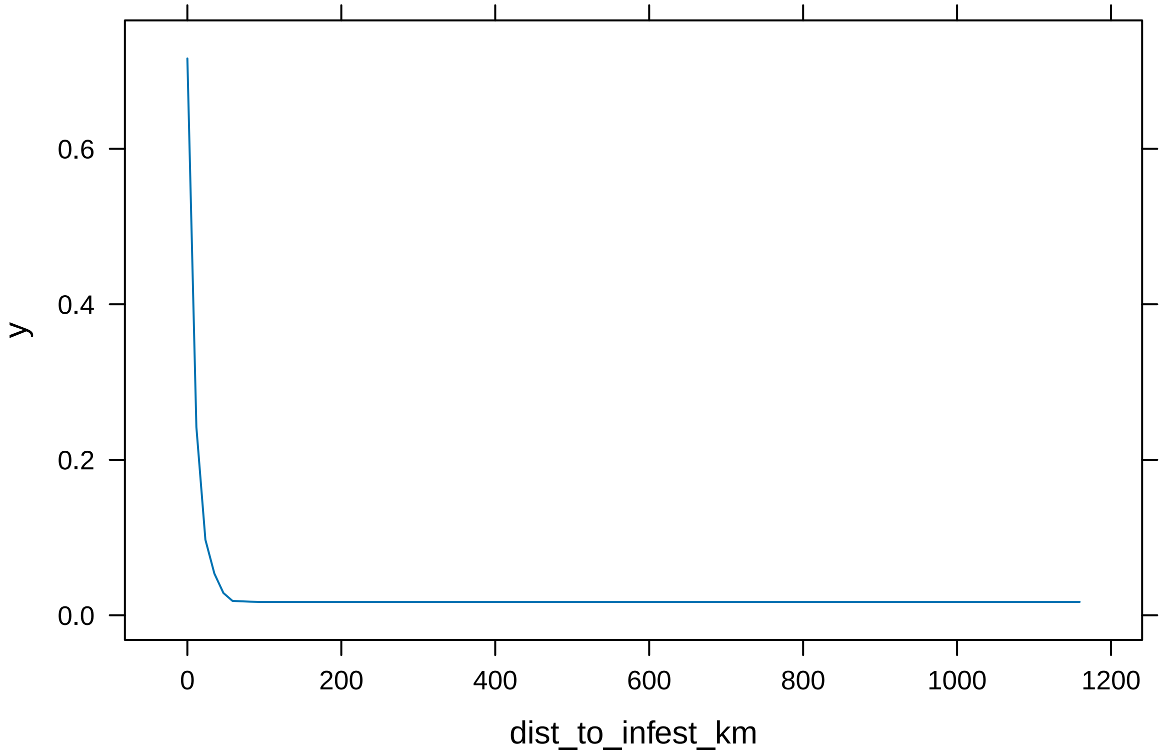

Figure 13. There is a negative relationship between infestation and the distance to the nearest infested body of water. Infestation is much more likely when other infestations are close.

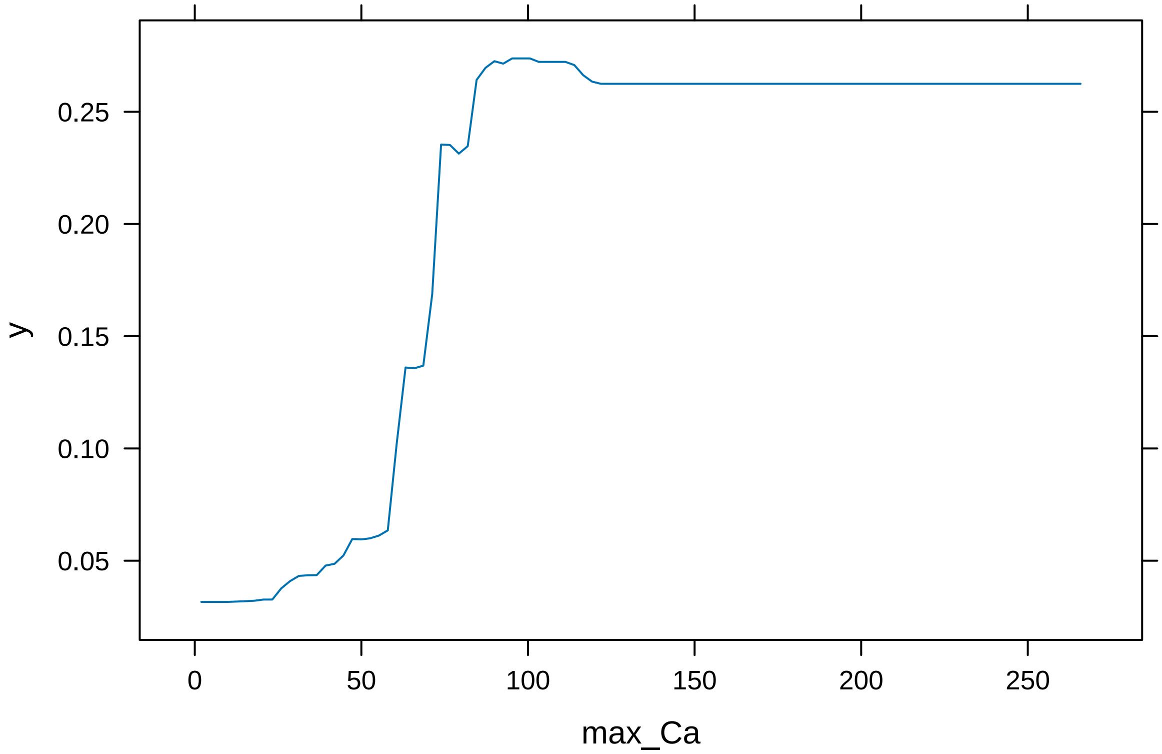

Figure 14. There is a positive relationship between observed maximum calcium levels and infestation status.

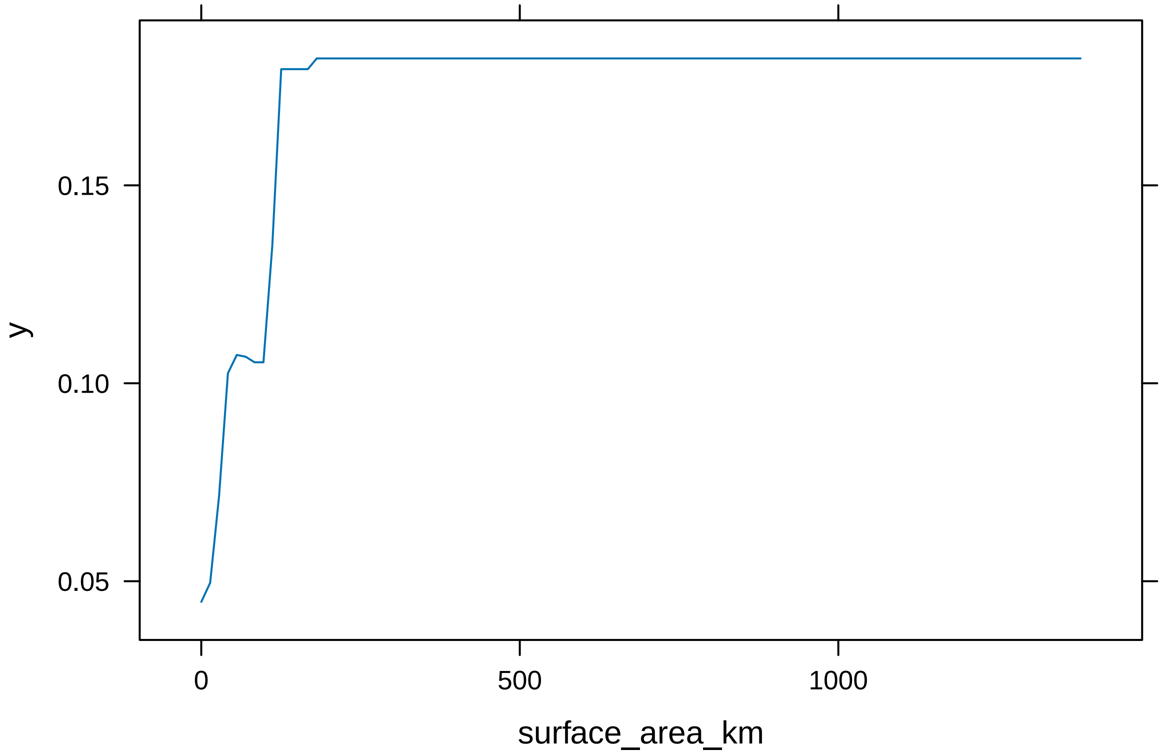

Figure 15. There is a positive relationship between infestation and the size of the reservoir. Larger reservoirs are more likely to be infested.

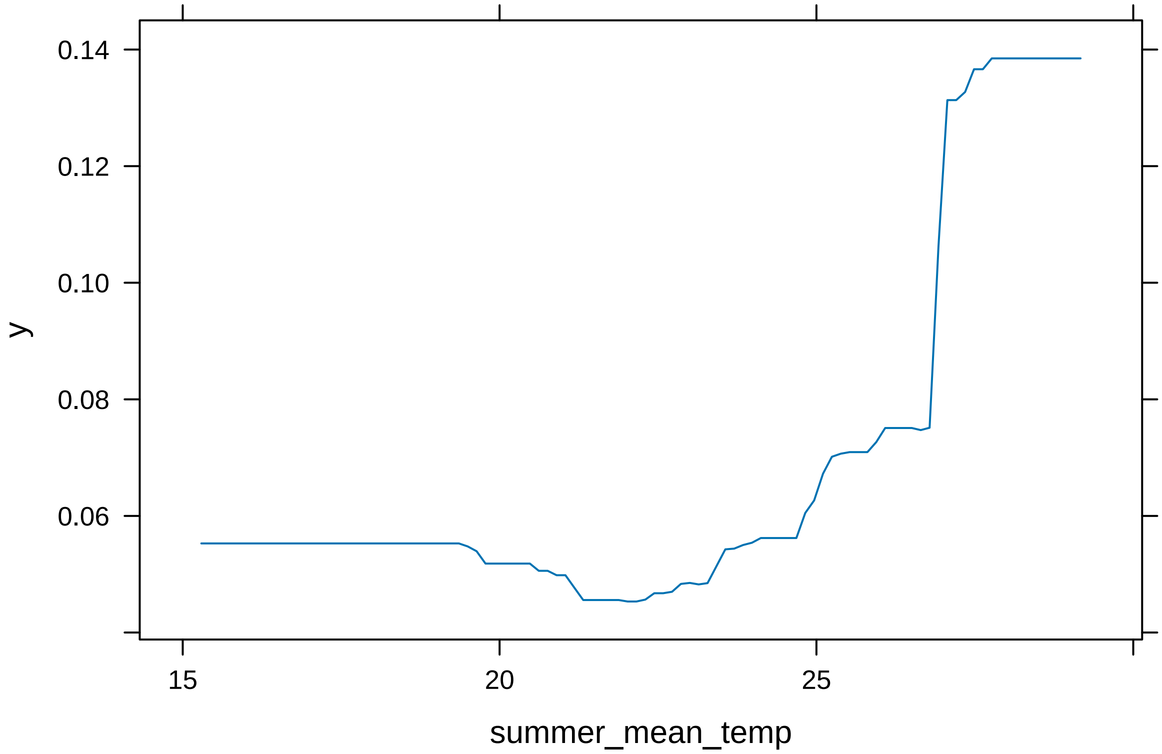

Figure 16. There is a positive relationship between summer mean temperatures and infestation status.

Next we use the model to generate predictions. We will then create a confusion matrix and calculate RMSE. A confusion matrix shows the number of true negatives (top left value), false negatives (top right), false positives (bottom left), and true positives (bottom right). RMSE stands for Root Mean Squared Error. A smaller RMSE indicates the model is more accurate.

Below is a confusion matrix for the training data.

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   0   1
##          0 213   2
##          1   1  65
##
##                Accuracy : 0.9893
##                  95% CI : (0.9691, 0.9978)
##     No Information Rate : 0.7616
##     P-Value [Acc > NIR] : <2e-16
##
##                   Kappa : 0.9705
##
##  Mcnemar's Test P-Value : 1
##
##             Sensitivity : 0.9701
##             Specificity : 0.9953
##          Pos Pred Value : 0.9848
##          Neg Pred Value : 0.9907
##              Prevalence : 0.2384
##          Detection Rate : 0.2313
##    Detection Prevalence : 0.2349
##       Balanced Accuracy : 0.9827
##
##        'Positive' Class : 1
##
```

Below is the RMSE for the training data

```
## [1] 0.1134874
```

Below is a confusion matrix for the test data.

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  0  1
##          0 47  5
##          1  7 12
##
##                Accuracy : 0.831
##                  95% CI : (0.7234, 0.9095)
##     No Information Rate : 0.7606
##     P-Value [Acc > NIR] : 0.1023
##
##                   Kappa : 0.5539
##
##  Mcnemar's Test P-Value : 0.7728
##
##             Sensitivity : 0.7059
##             Specificity : 0.8704
##          Pos Pred Value : 0.6316
##          Neg Pred Value : 0.9038
##              Prevalence : 0.2394
##          Detection Rate : 0.1690
##    Detection Prevalence : 0.2676
##       Balanced Accuracy : 0.7881
##
##        'Positive' Class : 1
##
```

Below is RMSE for the test data.

```
## [1] 0.3349775
```

The shows some signs of overfitting. It "predicts" infestation with perfect accuracy when given the training data, but it is less accurate when predicting infestation with new data. Still, the model is reasonably effective and may be further improved with model tuning.

Random forest analysis

Random forest classification is a type of machine learning that functions similarly to classification and regression tree analysis (CART). Random forest analysis does not assume linearity, normality, or homoscedasticity and is less sensitive to spatial autocorrelation and multicollinearity.

Random forest classification works by generating a series of bootstrapped trees that have low correlation with one another and averaging the results of these individual trees across a "forest" of many trees to prevent overfitting. Individual trees are created by randomly selecting a subset of all possible variables and using them to build and test a classification scheme using randomly selected training (64%) and test data (36%). Since each tree uses a random subset of variables and different training and test data, they should be relatively different from one another.

We can evaluate the performance of the random forest classification by obtaining an estimate of out-of-bag (OOB) error, which describes the overall percentage of incorrectly categorized data (in this case presence versus absence), averaged across trees.

We used the 'randomForest' package in R to run this analysis (Liaw and Wiener, 2002).

We obtained several outputs from the random forest classification, including:

1) Accuracy: overall accuracy of classification (average % of time random forest correctly classifies records)
2) Confidence interval: The 95% confidence interval for the accuracy rate
3) No information rate: The accuracy of the model if we were to assign all reservoirs a value of zero without running a classification
4) P-value (Acc > NIR): Indicates whether your model accuracy is significantly higher than your no information rate, which is an indication that your model provides valuable information
5) McNemar's test p-value: This test statistic serves to test whether the counts of false positives in the model are significantly different from the counts of false negatives.If the p-value of the test is significant, than we can say relatively confidently that the model has a different proportion of false positives and false negatives. McNemar's test compares the difference in the relative proportion of error between the two rather than the difference in error itself
6) Sensitivity: The probability that an infested reservoir will be correctly classified as infested
7) Specificity: The probability that an uninfested reservoir will be correctly classified as uninfested
8) Pos pred value: The proportion of true positives captured by the total number of predicted positives
9) Neg pred value: The proportion of true negatives captured by the total number of predicted negatives
10) Prevalence: The rate of all true positives in the whole population (The actual prevalence of infestations across reservoirs, regardless of classification)
11) Detection rate: The rate of detected true positives in the whole population (how many infested reservoirs are categorized as infested)
12) Detection prevalence: The rate of predicted positives in the whole population (how many reservoirs are predicted to be infested based on the classification)
13) Balanced accuracy: The average of sensitivity and specificity scores
14) Positive class: The class identified as indicating a 'positive' record (i.e., Dreissenid presence)
15) Importance: The average decrease in model accuracy if a given variable were dropped from the analysis

To run our random forest classification, we created 1000 trees using the default settings, including:

mtry: number of variables randomly sampled as candidates at each split (sqrt(p) by default, where p is the number of variables).

nodesize: minimum size of terminal nodes (1 by default)

To account for highly correlated variables, which can affect variable importance scores in the random forest model, we identified any variables with a correlation coefficient higher than 0.8 and retained only one variable from each of these groups. For LULC, we retained data for 25 mile buffers over 50 mile buffers given their better performance in the logistic regressions. For water quality data, we retained maximum values over average values because they performed better when classifying reservoir invasion risk. For developed land cover and seasonal precipitation, we summed categories based on logistic regression performance.

Random forest with test and training datasets

Below is a random forest model trained with the same training data given to the gradient boosted model. First the variable importance is displayed:

Random Forest Model 1

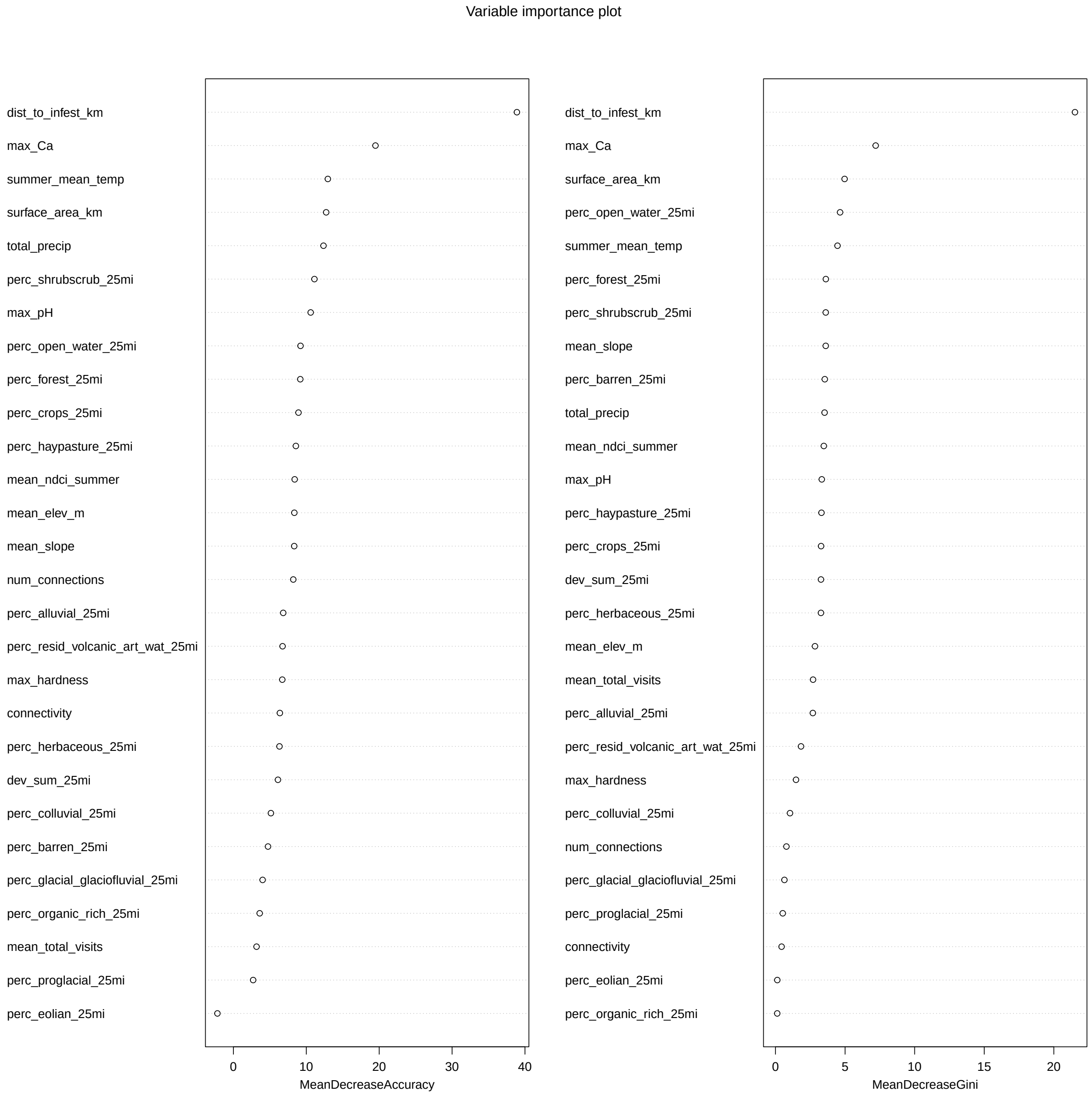

Figure 17. Variable importance plot for variables in the random forest model. 'MeanDecreaseAccuracy' indicates the decrease in accuracy of the model should a given variable be randomly permuted (i.e., its effect removed). 'MeanDecreaseGini' is a measure of the average gain in purity of splits for a given variable.

Then we will create a confusion matrix for the training data and test its performance for classification:

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   0   1
##          0 202   0
##          1   0  63
##
##                Accuracy : 1
##                  95% CI : (0.9862, 1)
##     No Information Rate : 0.7623
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 1
##
##  Mcnemar's Test P-Value : NA
##
##             Sensitivity : 1.0000
##             Specificity : 1.0000
##          Pos Pred Value : 1.0000
##          Neg Pred Value : 1.0000
##              Prevalence : 0.2377
##          Detection Rate : 0.2377
##    Detection Prevalence : 0.2377
##       Balanced Accuracy : 1.0000
##
##        'Positive' Class : 1
##
```

Above, we see that RF model is 100% accurate and the confusion matrix for the training data indicates that all Dreissenid presences and absences are correctly classified. The accuracy rate is statistically significantly higher than the no information rate, and the prevalence and detection rates mirror that of the actual population (~24% of reservoirs have positive records). Next, we will test our model with the testing dataset:

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  0  1
##          0 46  7
##          1  5 10
##
##                Accuracy : 0.8235
##                  95% CI : (0.712, 0.9053)
##     No Information Rate : 0.75
##     P-Value [Acc > NIR] : 0.1008
##
##                   Kappa : 0.5102
##
##  Mcnemar's Test P-Value : 0.7728
##
##             Sensitivity : 0.5882
##             Specificity : 0.9020
##          Pos Pred Value : 0.6667
##          Neg Pred Value : 0.8679
##              Prevalence : 0.2500
##          Detection Rate : 0.1471
##    Detection Prevalence : 0.2206
##       Balanced Accuracy : 0.7451
##
##        'Positive' Class : 1
##
```

Here, our RF model correctly classifies reservoirs based on Dreissenid presence approximately 82% of the time, although the balanced accuracy is lower (around 74%) because the model's sensitivity is poor. In other words, the model is not very good at accurately recognizing when Dreissenids are present, even though it is fairly good at predicting when Dreissenids are absent (as indicated by the high specificity rate). Furthermore, the accuracy rate is not significantly better than the no information rate, indicating that we could just assign all reservoirs as "absent" and still perform about as well as our model does now. The prevalence of Dreissenids in the test dataset is similar to the whole dataset, but the detection rate and detection prevalence are lower than the true prevalence. Our model may be limited by the "noise" of having so many variables included in it, so we decided to try reducing the number of variables by looking at only the 10 most important variables according to the mean decrease in accuracy metric.

Random forest with reduced variables

Random Forest Model 2

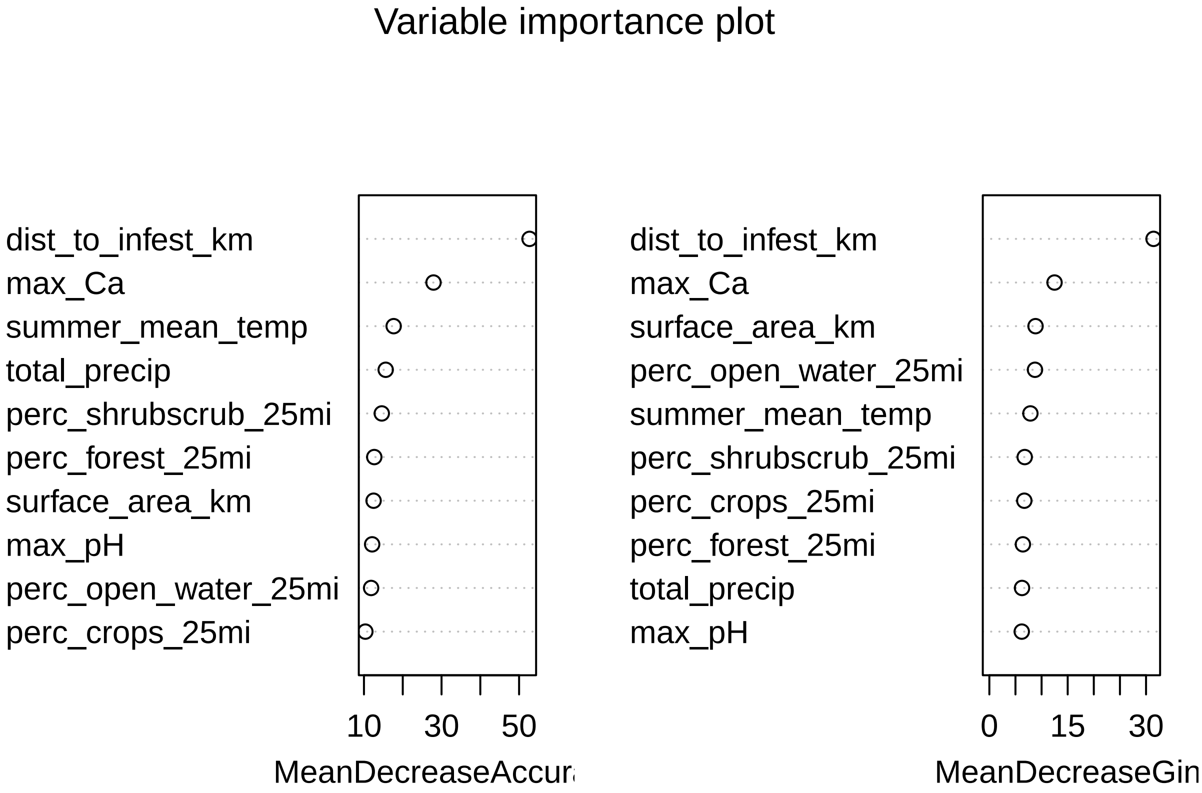

Figure 18. Variable importance plot for the top 10 variables in the random forest model. 'MeanDecreaseAccuracy' indicates the decrease in accuracy of the model should a given variable be randomly permuted (i.e., its effect removed). 'MeanDecreaseGini' is a measure of the average gain in purity of splits for a given variable. Next, we will create a confusion matrix for the training data.

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   0   1
##          0 214   0
##          1   0  67
##
##                Accuracy : 1
##                  95% CI : (0.987, 1)
##     No Information Rate : 0.7616
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 1
##
##  Mcnemar’s Test P-Value : NA
##
##             Sensitivity : 1.0000
##             Specificity : 1.0000
##          Pos Pred Value : 1.0000
##          Neg Pred Value : 1.0000
##              Prevalence : 0.2384
##          Detection Rate : 0.2384
##    Detection Prevalence : 0.2384
##       Balanced Accuracy : 1.0000
##
##        ’Positive’ Class : 1
##
```

Similarly to our RF model with all variables, this RF model is 100% accurate for the training data. Next, we’ll see how it performs with the test data:

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  0  1
##          0 47  7
##          1  7 10
##
##                Accuracy : 0.8028
##                  95% CI : (0.6914, 0.8878)
##     No Information Rate : 0.7606
##     P-Value [Acc > NIR] : 0.2476
##
##                   Kappa : 0.4586
##
##  Mcnemar’s Test P-Value : 1.0000
##
##             Sensitivity : 0.5882
##             Specificity : 0.8704
##          Pos Pred Value : 0.5882
##          Neg Pred Value : 0.8704
##              Prevalence : 0.2394
##          Detection Rate : 0.1408
##    Detection Prevalence : 0.2394
##       Balanced Accuracy : 0.7293
##
##        'Positive' Class : 1
##
```

Removing all but the 10 most influential variables actually decreases the performance of the model slightly, and the model accuracy is still not significantly better than the no information rate. We will also try a random forest model containing just the variables used in our risk assessment to see if that improves the performance.

Random forest with risk variables

Next, we will try the random forest using only the variables identified for risk thresholds to see if we can successfully classify reservoirs based on only a few environmental variables. Below, we have the variable importance plot:

Random Forest Model 3

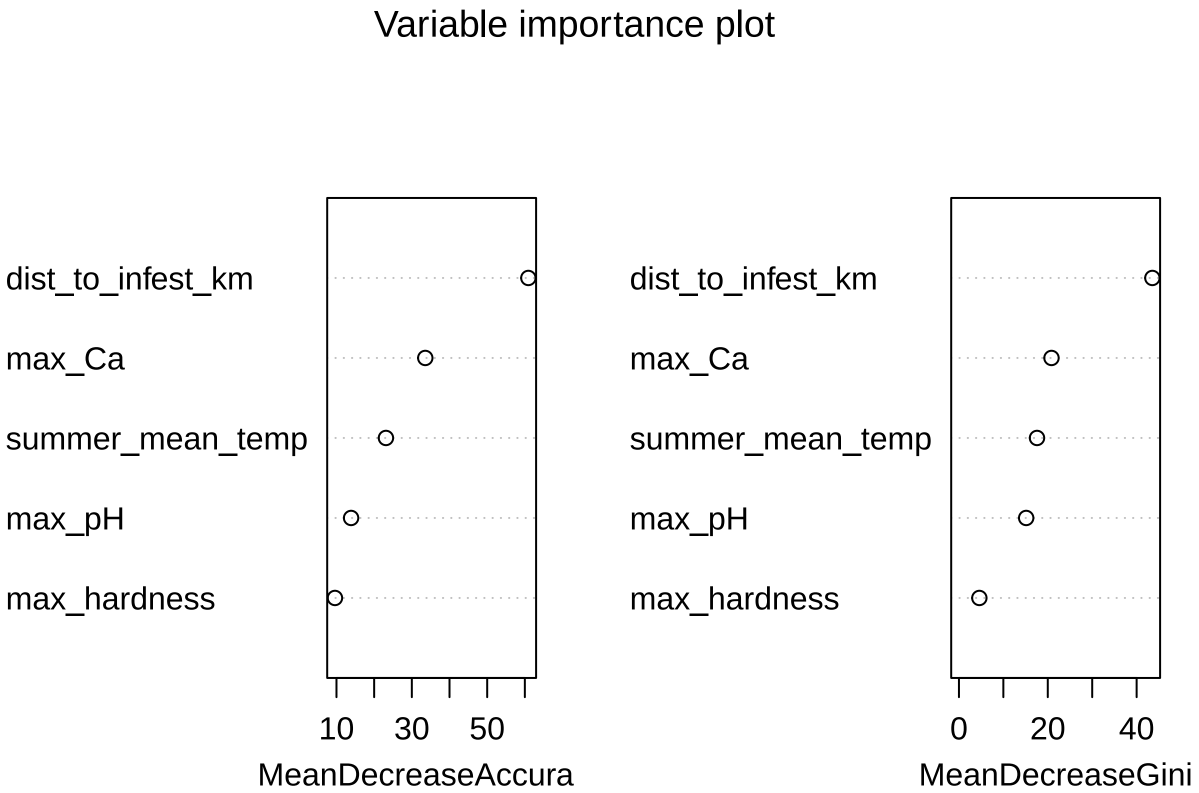

Figure 19. Variable importance plot for risk variables in the random forest model. 'MeanDecreaseAccuracy' indicates the decrease in accuracy of the model should a given variable be randomly permuted (i.e., its effect removed). 'MeanDecreaseGini' is a measure of the average gain in purity of splits for a given variable. Next, we will create a confusion matrix for the training data.

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   0   1
##          0 214   0
##          1   0  67
##
##                Accuracy : 1
##                  95% CI : (0.987, 1)
##     No Information Rate : 0.7616
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 1
##
##  Mcnemar's Test P-Value : NA
##
##             Sensitivity : 1.0000
##             Specificity : 1.0000
##          Pos Pred Value : 1.0000
##          Neg Pred Value : 1.0000
##              Prevalence : 0.2384
##          Detection Rate : 0.2384
##    Detection Prevalence : 0.2384
##       Balanced Accuracy : 1.0000
##
##        'Positive' Class : 1
##
```

Similarly to our RF model with all variables, this RF model is 100% accurate for the training data. Next, we'll see how it performs with the test data:

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  0  1
##          0 48  7
##          1  6 10
##
##                Accuracy : 0.8169
##                  95% CI : (0.7073, 0.8987)
##     No Information Rate : 0.7606
##     P-Value [Acc > NIR] : 0.1654
##
##                   Kappa : 0.4869
##
##  Mcnemar's Test P-Value : 1.0000
##
##             Sensitivity : 0.5882
##             Specificity : 0.8889
##          Pos Pred Value : 0.6250
##          Neg Pred Value : 0.8727
##              Prevalence : 0.2394
##          Detection Rate : 0.1408
##    Detection Prevalence : 0.2254
##       Balanced Accuracy : 0.7386
##
##        'Positive' Class : 1
##
```

Here, we can see that reducing the number of variables in our model resulted in little decrease in model accuracy (~82%) and the balanced accuracy (~74%) of our RF model. However, the model's sensitivity is still relatively poor and our accuracy is not significantly better than the no information rate. One factor that could be contributing to this is that the average temperatures of the reservoirs may not be particularly useful for distinguishing between reservoirs that are at risk of invasion. This is pretty clear in our risk assessment, because only a single reservoir is classified below high risk based on our thresholds. Hence, we decided to also try a Random Forest model without mean summer temperature.

Random forest with reduced variables no temperature

Since we know we have limited information about duration of temperature stressors, we will also try the random forest using only the variables identified for risk thresholds, but excluding temperature risk thresholds.

Random Forest Model 4

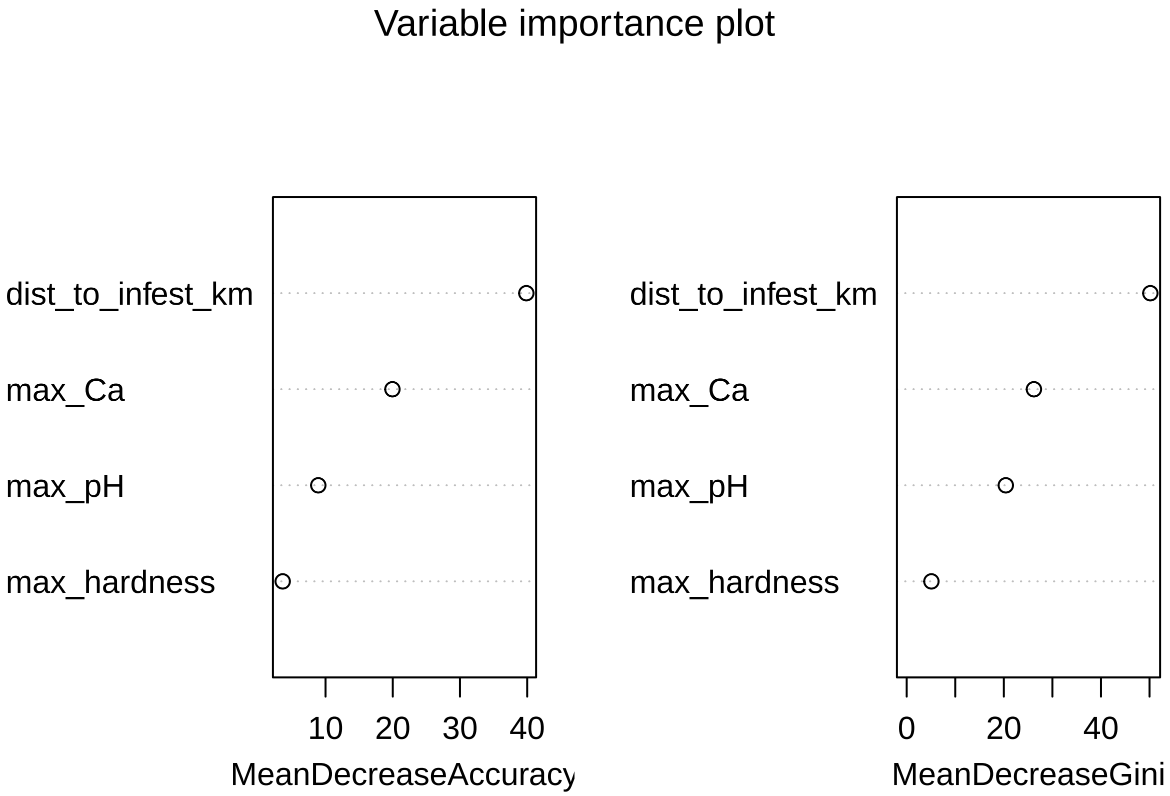

Figure 20. Variable importance plot for risk variables in the random forest model, excluding mean summer temperature. 'MeanDecreaseAccuracy' indicates the decrease in accuracy of the model should a given variable be randomly permuted (i.e., its effect removed). 'MeanDecreaseGini' is a measure of the average gain in purity of splits for a given variable. Then we will create a confusion matrix for the training data.

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   0   1
##          0 214   0
##          1   0  67
##
```

```
##                Accuracy : 1
##                  95% CI : (0.987, 1)
##     No Information Rate : 0.7616
##     P-Value [Acc > NIR] : < 2.2e-16
##
##                   Kappa : 1
##
##  Mcnemar's Test P-Value : NA
##
##             Sensitivity : 1.0000
##             Specificity : 1.0000
##          Pos Pred Value : 1.0000
##          Neg Pred Value : 1.0000
##              Prevalence : 0.2384
##          Detection Rate : 0.2384
##    Detection Prevalence : 0.2384
##       Balanced Accuracy : 1.0000
##
##        'Positive' Class : 1
##
```

Like all the other models, the RF model performs perfectly on the training data.

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  0  1
##          0 51  7
##          1  3 10
##
##                Accuracy : 0.8592
##                  95% CI : (0.7562, 0.9303)
##     No Information Rate : 0.7606
##     P-Value [Acc > NIR] : 0.03019
##
##                   Kappa : 0.5794
##
##  Mcnemar's Test P-Value : 0.34278
##
##             Sensitivity : 0.5882
##             Specificity : 0.9444
##          Pos Pred Value : 0.7692
##          Neg Pred Value : 0.8793
##              Prevalence : 0.2394
##          Detection Rate : 0.1408
##    Detection Prevalence : 0.1831
##       Balanced Accuracy : 0.7663
##
##        'Positive' Class : 1
##
```

When we apply the new RF model to the test data, our accuracy improves to around 86%, with a balanced accuracy around 77%. Although the model's sensitivity is still relatively low, here we see that our RF model is significantly better than the no information rate, meaning it contains information that is actually useful.

Conclusions

Based on our analysis, the most limiting factor related to Dreissenid presence in USACE reservoirs seems to be the distance to the closest current Dreissenid infestation. This is particularly concerning because it suggests that Dreissenid spread may be more limited by dispersal than by environmental conditions, meaning that efforts to control the spread of Dreissenids are tantamount for controlling continued expansion of populations in the U.S. Beyond distance, maximum calcium was consistently an important variable, which aligns with past research on environmental requirements of Dreissenids. Our risk analysis suggests that just over 50% of USACE reservoirs may be considered high risk for Dreissenid infestation, with an additional 21% considered moderate risk. Efforts to prevent the continued spread of Dreissenids could benefit from additional focus on preventing transport of veligers or mussels between nearby lakes via recreational activities. Furthermore, collection of more fine scale water quality data could help refine our risk estimates, particularly for lakes that exhibit higher intra-reservoir variation in waer quality due to geology or other factors.

References

Benson, A. J., Raikow, D., Larson, J., Fusaro, A., Bogdanoff, A. K., and A. Elgin. (2024). *Dreissena polymorpha* (Pallas, 1771): U.S. Geological Survey, Nonindigenous Aquatic Species Database, Gainesville, Florida. https://nas.er.usgs.gov/queries/FactSheet.aspx?speciesID=5, Revision Date: 12/21/2023, Access Date: 7/29/2024

Brownlee, J. (2019). How to calculate McNemar's test to compare two machine learning classifiers. https://machinelearningmastery.com/mcnemars-test-for-machine-learning/. Access Date: 12/3/2024

Brownlee, J. (2019). Feature selection with the Caret R Package. https://machinelearningmastery.com/feature-selection-with-the-caret-r-package/. Access Date: 12/03/2024.

Cohen, A. N. (2005). A review of Zebra Mussels' environmental requirements: a report for the California Department of Water Resources. San Francisco Estuary Institute, Oakland, California. 33 pp.

Cohen, A. N., and A. Weinstein. (1998). The potential distribution and abundance of Zebra Mussels in California. San Francisco Estuary Institute, Richmond, California.

Creamer, D. A., Rogosch, J. S., Patiño, R., & McGarrity, M. E. (2025). Identifying lakes critical to the westward spread and establishment of zebra mussels. Biological Conservation, 302, 110931. https://doi.org/10.1016/j.biocon.2024.110931

Doll, B. (1997). Zebra Mussel colonization: North Carolina's risks. Sea Grant North Carolina, Raleigh, North Carolina (UNC SG-97-01).

Kislik, C., Dronova, I., Grantham, T. E., Kelly, M. (2022) Mapping algal bloom dynamics in small reservoirs using Sentinel-2 imagery in Google Earth Engine. Ecological Indicators, Volume 140. https://doi.org/10.1016/j.ecolind.2022.109041.

Liaw A, and M. Wiener. (2002). Classification and regression by Random Forest. R News. 2(3):18–22. Available from: https://CRAN.R-project.org/doc/Rnews/.

Mishra, Sachidananda & Mishra, Deepak. (2012). Normalized difference chlorophyll index: A novel model for remote estimation of chlorophyll-a concentration in turbid productive waters. Remote Sensing of Environment. 117. 394-406. 10.1016/j.rse.2011.10.016.

Sorba, E. A., and D. A. Williamson. (1997). Zebra Mussel colonization potential in Manitoba, Canada. Water Quality Management Section, Manitoba Environment, Report No. 97-07.

White, J. D., S. K. Hamilton, and O. Sarnelle. (2015). Heat-induced mass mortality of invasive Zebra Mussels (Dreissena polymorpha) at sublethal temperatures. Canadian Journal of Fisheries and Aquatic Sciences. 72(8):1221-1229. https://doi.org/10.1139/cjfas-2015-0064.