Tobin Haefele

uNIVERSITY OF mONTANA

Russian Olive Report

rEPORT ON THE CURRENT AND FUTURE STATUS OF RUSSIAN OLIVE IN MISSOULA COUNTY

Table of Contents

[1. Introduction 2](#_Toc193663221)

[2. Background 2](#_Toc193663222)

[3. Methodology 4](#_Toc193663223)

[3.1 Model and Variable Selection 4](#_Toc193663224)

[3.2 Data Sources 6](#_Toc193663225)

[3.3 Data Cleaning and Integration 6](#_Toc193663226)

[3.4 Model Preparation 7](#_Toc193663227)

[4. Analysis 9](#_Toc193663228)

[4.1 Initial Analysis 9](#_Toc193663229)

[4.2 Random Forest Results 12](#_Toc193663230)

[5. Recommendations 15](#_Toc193663231)

[6. Conclusion 16](#_Toc193663232)

[7. Appendices 18](#_Toc193663233)

[8. References 19](#_Toc193663234)

# Introduction

The Russian Olive project originated from my interest in native plants and their ecological importance in Missoula and the surrounding area. This interest was sparked by personal experiences, such as planting native species in my yard, and grew into a desire to contribute to local conservation efforts. While I initially sought a project directly related to native plants, I realized that supporting native ecosystems also involves addressing the spread of invasive species. This led me to focus on invasive species management, specifically the Russian Olive, which was recently listed as a new invasive species in Missoula County in 2024.

To narrow the scope of the project, I chose to concentrate on Missoula County, as statewide efforts, such as those led by the Montana Natural Heritage Program (MTNHP) and Bryce Maxwell, have already made significant progress in this area. After reaching out to the Missoula County Ecology Extension, I learned about their ongoing field surveys and concerns regarding the Russian Olive. They provided valuable data and context, which became the foundation for this project. This report aims to support their efforts by identifying areas most susceptible to Russian Olive invasion and providing actionable insights for mitigation and removal strategies.

This report is intended for the staff of the Missoula County Ecology Extension, the Missoula County Weed Board, and other interested parties, including the public. Its goal is to inform decision-making around resource allocation and mitigation efforts, ensuring that limited resources are used efficiently to combat the spread of Russian Olive.

The primary question this project addresses is: Which areas in Missoula County are most susceptible to invasion by Russian Olive? Through answering this question, the report provides critical information about the current distribution of Russian Olive and identifies potential areas for future management efforts.

# Background

Russian Olive (*Elaeagnus angustifolia*) is a small tree native to southern Europe and western Asia. Introduced to North America during colonial times, it was initially planted for practical purposes, such as windbreaks, and for its ornamental appeal. However, it has since escaped cultivation and is now considered an invasive species, particularly in riparian zones, areas along riverbanks where it thrives and spreads rapidly. In Montana, Russian Olive was first planted as a windbreak as early as 1953, but its unchecked spread has led to significant ecological concerns. As of 2010, the Russian Olive is listed as State Regulated by the Montana Department of Agriculture, which means it is illegal to intentionally spread or sell.

One of the main issues stemming from the spread of Russian Olive, is the overcrowding and eventual overtaking of native species within the ecosystem. Native species such as the cottonwood and willow occur in the same environment as the Russian Olive, causing competition between the species. This is an issue due to certain characteristics that give Russian Olive an advantage in this competition. Unlike the native Cottonwood, the Russian Olive can reproduce in the shade[[1]](#endnote-1), meaning the Russian Olive begins to take over as the dominant species.

Another advantage is the aversion of Beavers to Russian Olive. Researchers found that Beavers tended to damage 57 to 78 percent of cottonwood trees, while only damaging a mere 15 to 18 percent of Russian Olive Trees[[2]](#endnote-2). Furthermore, the damage to Russian Olive tended to be primarily the limbs, while damage to the Cottonwood tended to be at the trunk or base of the tree.

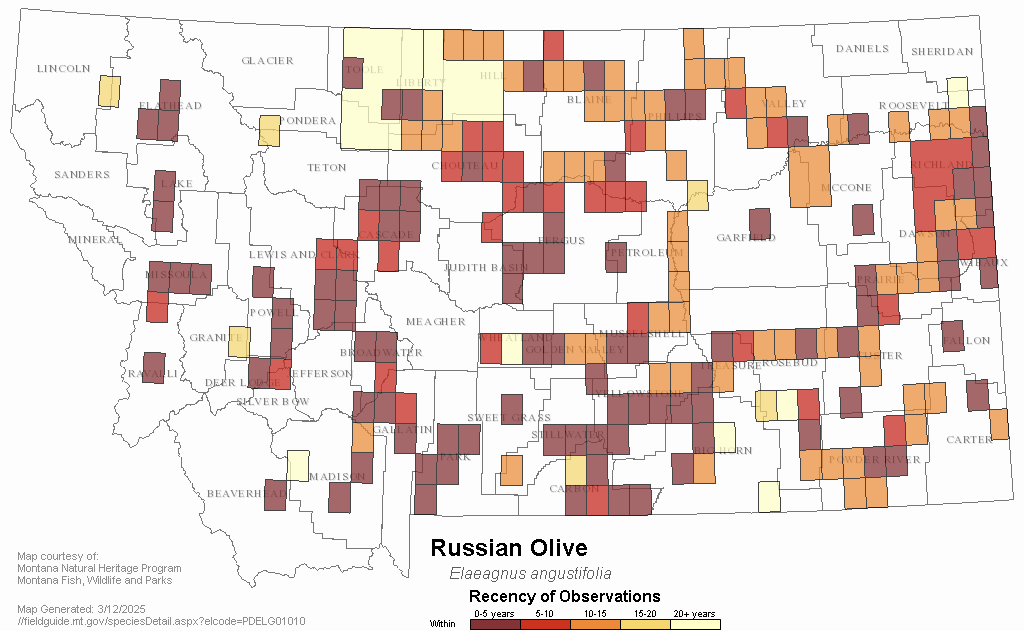
Additionally, Russian Olive thrives in areas with regulated river flows, such as those impacted by dams or irrigation systems. These human-altered environments create ideal conditions for their growth, allowing them to spread more aggressively. As a result, Russian Olive not only disrupts natural ecosystems but also exacerbates the challenges of managing riparian areas in the face of human activity.

The loss of native species like the Cottonwood also means the loss of habitat for native animal species, for example cavity-nesting birds that rely on the Cottonwood to reproduce do not appear to use Russian Olive as a replacement. Ungulates such as the White-Tailed Deer forage near cottonwood trees at a much higher rate than near Russian Olive. Preserving these fragile ecosystems is an important step in combatting climate change at the local level.

Several studies have examined the distribution of Russian Olive in Montana, including a notable study by Lesica and Miles, which tracked its spread along the Marias and lower Yellowstone rivers in eastern Montana. More recently, researchers used aerial images from the National Agriculture Imagery Program (NAIP) and a type of computer model called a random forest to map the types of plants and land cover found along the valley bottoms of ten eastern Montana rivers. Russian Olive was included as one of the plant types in this mapping. These studies have provided valuable insights into the species’ behavior and impact in Eastern Montana.

However, fewer studies have focused on the western part of the state, including Missoula County. According to data from the Montana Natural Heritage Program (MTNHP), observations of Russian Olive in Missoula County have increased in recent years, as shown in Figure 1 below. This uptick in observations underscores the need for localized research to understand the current distribution of Russian Olive and predict areas where it may spread in the future.

Figure : Russian Olive Observations in Recent Years via MTNHP

One caveat to this statement is the small amount of data in the MTNHP database for Missoula County (around 53 entries). The other issue is the lack of data prior to 2023 from Missoula County. There are some observations starting in 2012, but most of them come from the 2023 field survey.

Given that Russian Olive was recently added to Missoula County’s watch list of invasive species, this project seeks to fill critical gaps in knowledge about its spread in the region. By identifying areas most susceptible to invasion, the findings can inform targeted mitigation efforts and help allocate resources more effectively to protect native ecosystems.

# Methodology

## Model and Variable Selection

I chose to use a method called Habitat Suitability Modeling (HSM) for this project, as it is a well-established approach for predicting how well an area can support a given species. This method is commonly used in conservation science, ecology, and land management to guide decisions about where to focus monitoring, restoration, or control efforts.

Habitat Suitability Modeling works by identifying the key environmental conditions that a species needs to survive and thrive, such as soil type, moisture levels, temperature, land cover, and proximity to water. Once these factors are known, the model evaluates a geographic area to determine how suitable each part of it is for the species in question. The result is typically a map that highlights areas with higher or lower habitat suitability.

For this project, I applied HSM to assess the potential spread and establishment of Russian Olive in Missoula County. I built on prior work, including a 2017 study that modeled Russian Olive habitat in parts of eastern Montana, as well as the Montana Natural Heritage Program’s (MTNHP) statewide habitat suitability model. These sources helped inform which variables to include and guided my understanding of how Russian Olive responds to different environmental conditions in Montana.

Selecting the appropriate environmental variables requires careful consideration. I drew on insights from Lesica (2012), who described the typical habitat of Russian Olive in Montana as including woodlands, thickets, riparian forests, and moist meadows around wetlands in plains and valleys. Russian Olive also tends to grow in soils with low to moderate soluble salt concentrations and exhibits some tolerance for saline conditions[[3]](#endnote-3). Additionally, I referenced the MTNHP’s Maxent model, which used 22 statewide biotic and abiotic environmental layers. Their results highlighted the importance of several key variables, including:

* **Land Cover:** Wetland Riparian, Introduced Vegetation, and Conifer Forests
* **Climate:** Frost-free days, Degree Days, and Maximum Summer Temperature
* **Soil:** Soil pH and Bulk Density.

To gather the necessary data, I used the Montana Spatial Data Infrastructure (MSDI) and ArcGIS. Below is an overview of the environmental variables I incorporated into my model:

**1. Land Cover**

**Land cover level** refers to the amount of detail used to classify the types of land and vegetation across a landscape. Land cover data is typically organized into multiple levels, with each level representing a different degree of classification detail. In this project, land cover is mapped at a 30-meter resolution, meaning each pixel on the map represents a 30x30 meter area. The classification system is broken down into three levels:

* **Level 1:** Broad categories that group land into major ecological or human use types (8 classes total). Examples include Forest and Woodland Systems, Grassland Systems, Human Land Use, and Open Water/Wetlands.
* **Level 2:** A more detailed breakdown, with 27 classes. This level adds ecological nuance while still maintaining manageability for modeling purposes.
* **Level 3:** The most detailed level, where each pixel is assigned a highly specific classification, often down to particular vegetation communities or land uses.

For this project, I used Level 2 land cover classifications. This level strikes a good balance by providing enough ecological detail to support accurate modeling, while avoiding the complexity and potential overfitting that can come with Level 3 data. This approach also aligns with the methodology used in the MTNHP statewide habitat suitability model, ensuring consistency with previous research.

**2. Climate Data**

Climate data was sourced from the **Montana Climate Office**, and the key variables I included are:

* **Frost-Free Days:** Estimated number of days without frost (daily minimum temperature > 32°F).
* **Relative Effective Annual Precipitation (REAP):** 30-year precipitation data adjusted for slope and aspect.
* **Precipitation:** Mean annual precipitation (mm) for the 1991–2020 period.
* **Maximum/Minimum Temperature:** Mean maximum and minimum temperatures (°C) in July and January, respectively.

**3. Soil Data**

Soil variables were selected based on their relevance to Russian Olive’s growth preferences:

* **Soil pH:** pH of the topsoil layer (0–5 cm depth).
* **Bulk Density:** Mass of the topsoil layer (0–5 cm depth).

By incorporating these variables, I aimed to create a robust model that reflects the environmental conditions most conducive to Russian Olive’s spread in Missoula County. A detailed table summarizing these variables and the corresponding sources can be found in Appendix A.

## Data Sources

I used two distinct datasets for this model, both representing locations in Missoula County where Russian Olive has been observed, commonly referred to as **presence points**. The first dataset, provided by the Missoula County Ecology Extension, consists of field survey data collected during the past year (2023–2024). I consider this dataset to be the more dependable of the two, as each observation has been verified through direct fieldwork.

In addition to latitude and longitude coordinates, this dataset includes valuable contextual information, such as "Woody Growth" (indicating the plant’s growth stage) and "Woody Setting" (describing the surrounding environment). These additional attributes provide insight into the ecological conditions where Russian Olive is currently growing, which enhances the quality and interpretability of the habitat suitability model.

The second dataset comes from the Montana Natural Heritage Program (MTNHP), which aggregates data from surveys, iNaturalist users, and other sources. While this dataset is less dependable due to the number of unverified observations, it provides broader spatial coverage. In addition to location data, the MTNHP dataset includes details such as the observer’s name, the observation date, and any additional comments. It also contains a spatial precision value, which estimates how closely the mapped location matches the real-world position, with lower values indicating higher accuracy.

## Data Cleaning and Integration

There are several ways to combine datasets like these, either programmatically in R or Python, or using GIS software such as ArcGIS. For this project, I chose to use ArcGIS because it offered a more streamlined and efficient workflow for preparing and integrating spatial data. ArcGIS made it easy to combine and manipulate both presence point datasets using built-in geoprocessing tools, without needing to write custom code.

Additionally, ArcGIS simplified the process of importing environmental raster layers and extracting values at each presence point location. This avoided the need to switch between platforms or reformat data, helping to maintain consistency and reduce the potential for errors during data transfer. Overall, using ArcGIS allowed me to complete the data preparation phase more efficiently and with greater confidence in the spatial accuracy of the results.

To begin, I combined the data points from the Missoula County Ecology Extension with those from the Montana Natural Heritage Program (MTNHP). To simplify this process, I first clipped the state dataset to include only points within Missoula County. Next, I filtered the state dataset to retain only points with a spatial precision of less than 800 meters, as recommended in the MTNHP model.

To preserve the integrity of both datasets, I used the **Append** tool in ArcGIS, which allowed me to merge the two while retaining all relevant attribute fields from the MTNHP dataset. I also addressed overlapping points between the datasets by using the **Near** tool in ArcGIS. This tool helped identify points located within a user-defined distance of one another, allowing me to randomly select one point from each overlapping pair to avoid duplication.

Next, the combined dataset only includes points where Russian Olive is present. While some modeling techniques, like Maxent, are specifically designed to work with presence-only data, my random forest model requires both presence and absence data to function effectively. Because true absence data, confirmed locations where Russian Olive does *not* occur, can be hard to come by, I used **pseudo-absence points** instead.

Pseudo-absence points are randomly generated locations that are assumed not to contain the species based on current knowledge and available observations. These points act as stand-ins for true absences and are essential for training machine learning models like random forests, which need both positive and negative examples to learn patterns and make predictions.

Including pseudo-absences allows the model to better distinguish between suitable and unsuitable habitat by contrasting the environmental conditions of known presence locations with those of areas where the species is unlikely to occur. This helps improve the model’s accuracy and its ability to generalize across the landscape.

To ensure reliability, I took care to generate pseudo-absence points in areas far enough from known presence points to reduce the risk of misclassifying locations where Russian Olive might be present but undetected.

To generate pseudo-absence points, I used ArcGIS to randomly distribute them across Missoula County. It is crucial to maintain an **equal ratio of pseudo-absence points to presence points** to create a balanced dataset. Having an equal number of presence and pseudo-absence points helps ensure that the model can learn to differentiate between the environmental conditions where Russian Olive is found and those where it is absent.

If the ratio is skewed—such as having far more presence points than pseudo-absence points—the model may become biased, giving too much weight to the presence data and underestimating the importance of identifying areas where Russian Olive does not occur. This can lead to overfitting, where the model becomes too focused on the presence locations and struggles to generalize to new, unobserved areas.

By aiming for a balanced ratio, I ensured the model receives an equal opportunity to learn about both suitable and unsuitable habitats, improving its predictive accuracy and overall performance.

Before combining the pseudo-absence points and presence points, I created a new field called PA (Presence-Absence). This column contains a Boolean value that indicates whether a given data point (row) represents a location with Russian Olive or not (0 = no Russian Olive, 1 = Russian Olive present). This is a crucial component for Random Forest modeling, as it serves as the dependent variable—the target the model is trying to predict. Later in the paper, I will go into more detail about how the random forest model uses this column, along with the environmental variables, to classify areas based on their suitability for Russian Olive.

Once I have the combined dataset, including Missoula County points, state data points, and pseudo-absence points, I added the environmental variables, overlaying each variable layer or raster (Table 1 Below) onto the point dataset.

| **Category** | **Variable** | **Description** | **Source** |
| --- | --- | --- | --- |
| **Land Cover** | Level 2 Classification | Intermediate detail with 27 land cover categories. Provides sufficient detail without complexity. | Montana Spatial Data Infrastructure (MSDI) |
|  | Level 1 Classification | Broad categorizations with 8 land cover classes. | MSDI |
|  | Level 3 Classification | Highly granular, with each 30-meter pixel assigned a unique value. | MSDI |
| **Climate** | Frost-Free Days | Estimated number of days without frost (daily minimum temperature > 32°F). | Montana Climate Office |
|  | Relative Effective Annual Precipitation (REAP) | 30-year precipitation data adjusted for slope and aspect. | Montana Climate Office |
|  | Precipitation | Mean annual precipitation (mm) for the 1991–2020 period. | Montana Climate Office |
|  | Maximum/Minimum Temperature | Mean maximum and minimum temperatures (°C) in July and January, respectively. | Montana Climate Office |
| **Soil** | Soil pH | pH of the topsoil layer (0–5 cm depth). | MSDI |
|  | Bulk Density | Mass of the topsoil layer (0–5 cm depth). | MSDI |

Table 1: Environmental Variables Used

After this step, I extracted the environmental data for each presence and pseudo-absence point into a single table. This table included each point’s geographic coordinates (latitude and longitude) along with all corresponding environmental variables, such as temperature, precipitation, land cover type, and proximity to water.

This extraction process was critical because it transformed the environmental data, which was typically stored in separate raster or vector layers, into a structured format that could be easily analyzed. By combining all the relevant data—both spatial and environmental—into one table, I created a comprehensive dataset that was ready for analysis and modeling.

Having all the data in one table simplified the import process into R. With this organized dataset, I was able to efficiently apply statistical or machine learning techniques, ensuring that each point had all the necessary information for model training. It also reduced the risk of errors or mismatches that might have arisen from managing separate datasets during analysis. Overall, this step was key to streamlining the workflow and ensuring the model received the correct input for accurate predictions.

## Model Preparation

Before diving into the data preparation steps, it is important to briefly explain why I chose the Random Forest algorithm for this project. Random Forest is a machine learning technique used for classification and regression tasks. It works by constructing multiple decision trees based on random subsets of the data and then combining their predictions to improve accuracy and reduce overfitting. A **decision tree** is a flowchart-like structure that splits the data into branches based on the values of input variables, leading to a prediction at the end of each branch. While a single tree can be sensitive to noise or bias in the data, Random Forest overcomes this by averaging the results of many trees, resulting in a more stable and dependable model.

One of the key advantages of Random Forest is its ability to manage large datasets with complex interactions between variables, making it well-suited for ecological modeling where numerous environmental factors influence species distribution.

In this project, I used Random Forest to predict the suitability of habitat for Russian Olive based on various environmental variables. It is particularly useful in this context because it can handle both categorical and continuous data, and it provides insights into the relative importance of different environmental factors in determining species’ presence. The model’s ability to handle non-linear relationships and interactions between variables made it a strong choice for this analysis.

Using the final combined table, I imported the data into R for further preparation and modeling. Several critical steps were involved in preparing the data for analysis. First, I converted the presence-absence column to a factor, as this is required for classification tasks in random forest modeling. Similarly, I converted all text-based columns, such as land cover, into factors to ensure they were treated as categorical variables rather than numeric ones.

Next, I reviewed the dataset for irrelevant or redundant columns. I dropped several columns that served only as identifiers (e.g., unique IDs) or contained too many categories to be meaningful for the model. For example, columns with highly granular or sparse data were removed to simplify the dataset and improve model performance.

Once the data was cleaned and formatted, I needed to split it into training and testing sets. However, standard random splitting methods are not suitable for spatial data due to spatial autocorrelation, the tendency for nearby locations to have similar environmental conditions. In ecological studies, this means that locations that are geographically close to each other often share similar characteristics, which can lead to overfitting if data points from the same region are used in both the training and testing sets.

To address this issue, I used the **blockCV** package in R, which is specifically designed for spatial data. This package divides the dataset into **spatially separated folds**, subsets of the data that are geographically distinct from each other. Each fold can be thought of as a spatial “chunk” or region of the study area, and each chunk is treated as a self-contained group of data points.

By splitting the data in this way, I ensured that the training and testing data were independent of each other, meaning that data points from the same geographic area were not present in both sets. This helps prevent spatial leakage, where the model might inadvertently "learn" spatial patterns that are not generalizable to other areas.

*A map of numbers and lines

AI-generated content may be incorrect.*Each fold in the **blockCV** process contains a subset of the data that is spatially distinct from the others, creating a more realistic scenario for model validation. When the model is trained on one-fold and tested on another, it simulates how the model would perform when applied to new, unseen areas of the landscape. This approach ensures that the model’s performance is robust and more representative of how it will generalize to different locations.

Figure : Folds overlayed on Missoula County

To evaluate the performance of the random forest model, I used several standard metrics: OOB Accuracy, AUC-ROC curve, and a confusion matrix. These metrics provide a comprehensive assessment of the model’s predictive accuracy and reliability.

* **OOB Accuracy:** Estimate of model’s performance in predicting presence.
* **AUC-ROC:** Evaluates the model’s ability to distinguish between presence and absence points, with values closer to one indicating better performance.
* **Confusion matrix:** Provides a detailed breakdown of true positives, true negatives, false positives, and false negatives, allowing for a deeper understanding of the model’s classification performance.

In addition to these metrics, I assessed the importance of each environmental variable using an importance plot. This plot ranks variables based on their contribution to the model’s predictive power, helping to identify which factors (e.g., land cover, climate variables) are most influential in determining Russian Olive’s habitat suitability.

# Analysis

## Initial Analysis

Before discussing the model output, I wanted to address several key questions that provide context for the current distribution of Russian Olive in Missoula County. These insights help clarify where Russian Olive is found, how it is spreading, and potential challenges for remediation efforts.

Looking at the distribution of data points, Russian Olive tends to be concentrated in areas near rivers or streams, which aligns with its preference for riparian habitats. However, there are also several locations—such as the large concentration near the Fish and Wildlife Department on Spurgin Road —where no visible water source is nearby. This suggests that Russian Olive may be spreading beyond its typical riparian zones, possibly due to human activity or other environmental factors.

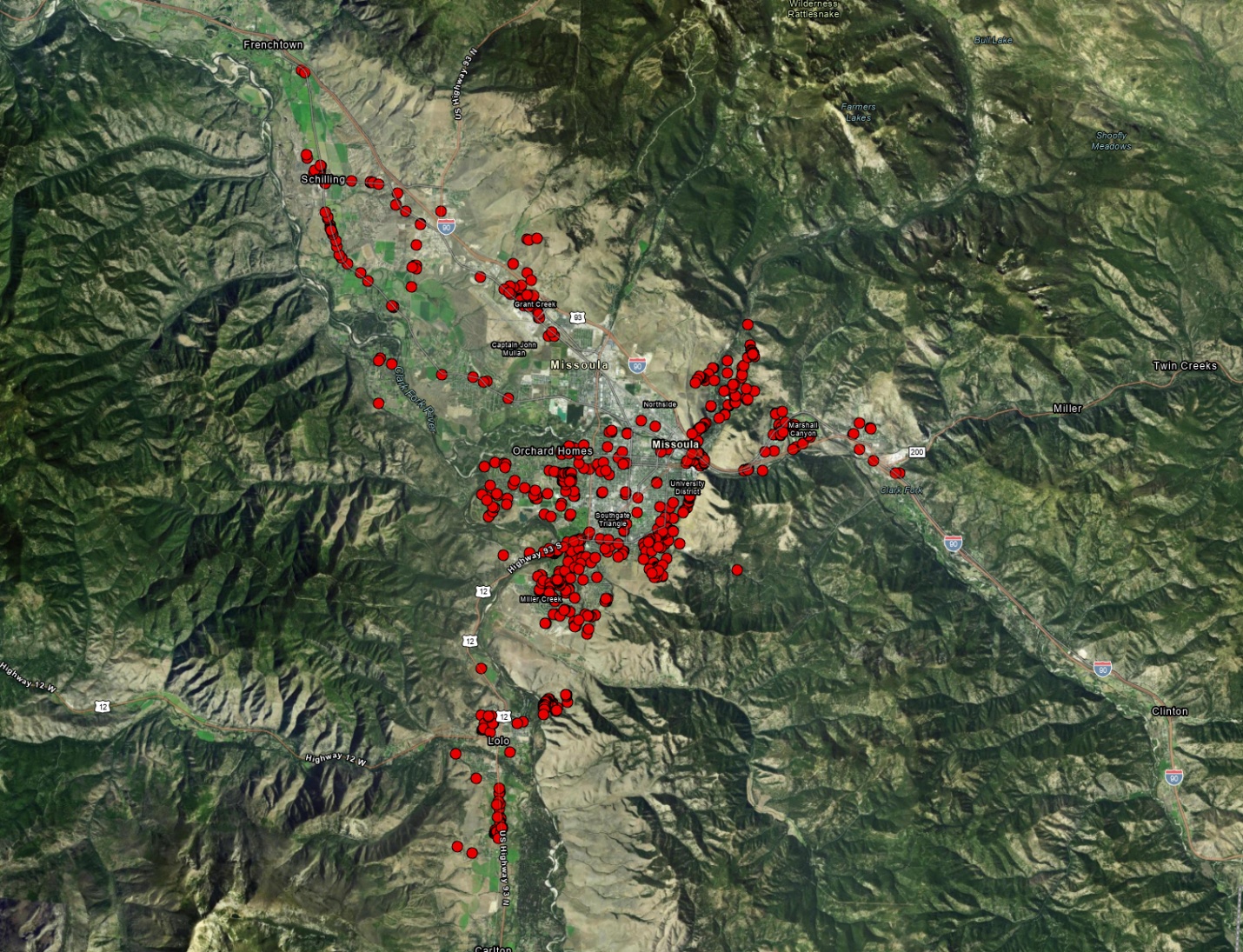


Figure : Russian Olive in Missoula

The dataset categorizes Russian Olive observations into four settings, which provide insight into how the species is established in the area:

* **Ornamental (Blue)**: Planted intentionally by individuals (e.g., for landscaping).
* **Escaped (Red):** Established from seeds dispersed from other plants.
* **Windbreak (Green):** Planted as part of a windbreak.
* **Other:** Includes cases where the setting is unknown or does not fit into the above categories.

A satellite image of a mountain range

AI-generated content may be incorrect.

Figure : Setting of Russian Olive in Missoula

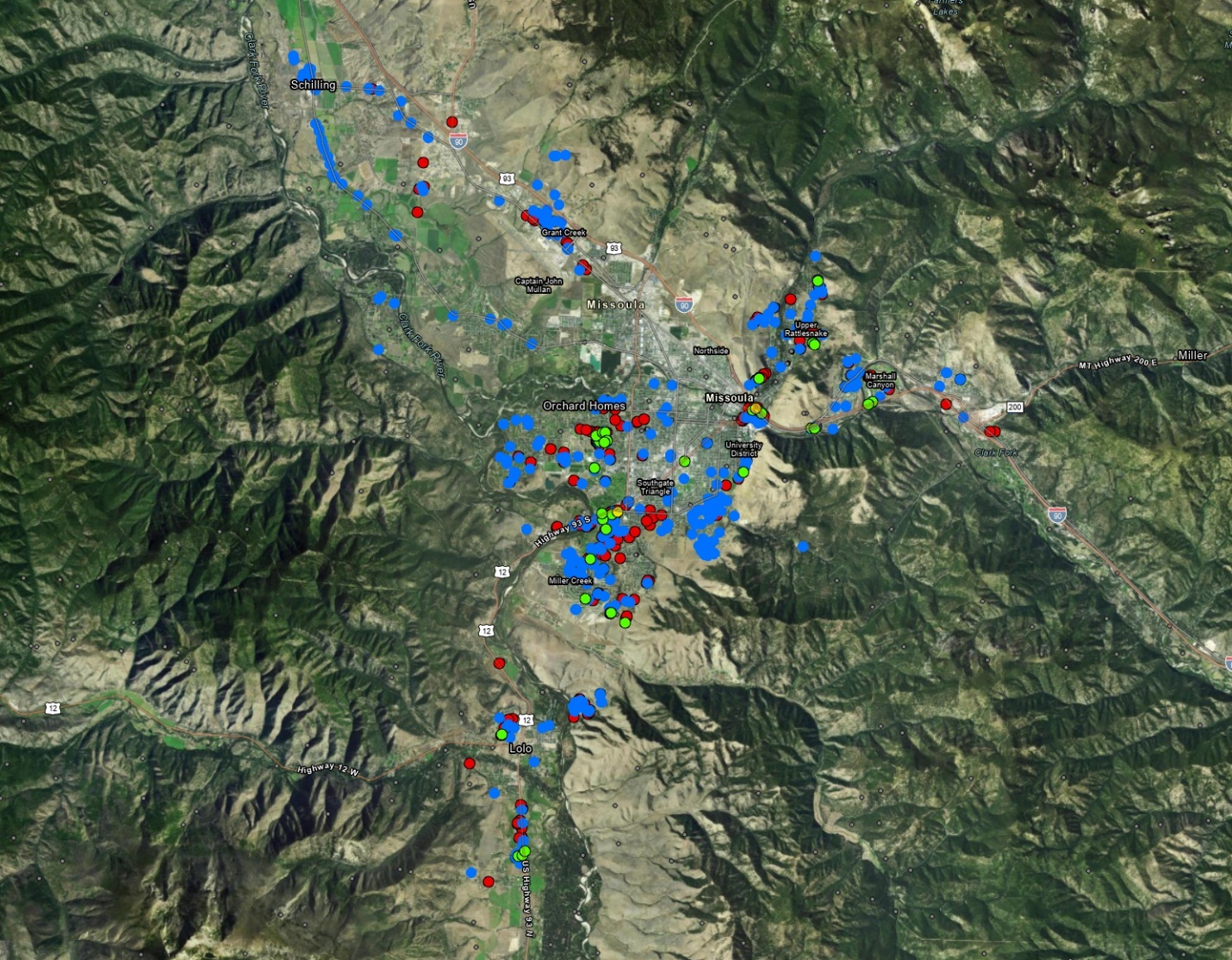
Last, I want to know what growth stage the observed plant is at. Like the woody setting, there are several distinct stages that are observed, including:

* **Immature (Red):** Not yet fully grown.
* **Mature (Blue):** Fully grown and reproducing.
* **Seedling (Green):** Recently sprouted.
* **Senescent (Yellow):** Older plants that are in decline.
* **Other:** Includes cases where the growth stage is unknown or not specified.

One limitation of these maps is several observations categorized as "Other" or "NA." This is primarily due to the state dataset, which does not consistently record setting or growth stage information. Despite this limitation, the available data still provides valuable insights into the distribution and characteristics of Russian Olive in Missoula County.

## Interpreting Model Results

Figure : Growth Stage of Russian Olive in Missoula



The Random Forest model identified several environmental variables as having the most significant impact on the likelihood of Russian Olive presence in each location. All of the variables are listed in Table 2 below. Key factors include REAP, Frost Free Days, LEVEL2 land cover, and soil\_pH which align with known ecological preferences of Russian Olive.

Each column in Table 2 provides different insights into how the model used these variables:

* **0 and 1**: These columns represent the average values of the variable when the target class is 0 (absence) or 1 (presence). This helps illustrate how the variable differs between areas where Russian Olive is and is not found.
* **MeanDecreaseAccuracy**: This value reflects how much the model’s overall accuracy would decrease if that variable were removed. Higher values indicate that the variable plays a more critical role in making accurate predictions.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **0** | **1** | **MeanDecreaseAccuracy** | **MeanDecreaseGini** |
| REAP | 9.03 | 26.94 | 29.12 | 91.12 |
| FROST\_FREE\_DAYS | 1.39 | 20.11 | 21.02 | 45.87 |
| ANNUAL\_PRECIP | 7.64 | 11.85 | 13.36 | 15.06 |
| SUMMER\_AVG\_MAXTEMP | 7.98 | 12.03 | 14.89 | 27.60 |
| WINTER\_AVG\_MINTEMP | 3.67 | 21.10 | 22.03 | 16.61 |
| SOIL\_PH | 15.37 | 35.21 | 35.08 | 197.90 |
| SOIL\_BULK\_DENSITY | 4.64 | 15.81 | 16.96 | 60.92 |
| LEVEL1 | 12.75 | 13.72 | 16.49 | 90.67 |
| LEVEL2 | 19.56 | 18.45 | 24.90 | 102.47 |
| NEAR\_WATER\_150m | 1.66 | 7.21 | 7.01 | 1.70 |

Table 2: Variable Importance Table

* **MeanDecreaseGini**: This is a measure of how important the variable is in splitting the data within the trees of the model. A higher value means the variable contributed more frequently and effectively to decision-making within the random forest.

Key environmental factors included REAP and Frost-Free Days, both of which relate to temperature and moisture conditions, capturing the semi-arid climate that Russian Olive favors. The LEVEL2 land cover classification also played a major role, reinforcing the species’ tendency to establish in specific vegetation types such as riparian areas and disturbed grasslands. Finally, soil pH stood out as the most important predictor overall, suggesting a strong ecological preference for certain soil chemistry, which aligns with previous research indicating Russian Olive's tolerance for alkaline soils.

These results demonstrate that the model not only performed well in predicting Russian Olive presence but also identified ecologically meaningful variables that reflect the species’ known environmental preferences.

Digging into these results a little further, I wanted to understand which specific land cover types within the **LEVEL2** category contributed most to the model’s performance. To do this, I generated a **partial dependence plot (PDP)** for the **LEVEL2** variable. This technique allows us to explore the relationship between each individual land cover type and the predicted probability of Russian Olive presence, while holding all other variables constant.

Table 3 presents the values from the PDP, highlighting how different land cover classes influence the model’s predictions. Notable land cover types include **Conifer Forest**, **Deciduous Grassland and Shrubland**, and **Montane Grassland and Shrubland**, which showed particularly strong associations with the likelihood of Russian Olive presence or absence.

These values differ from the variable importance scores in Table 2, which measure how much the overall **LEVEL2** variable contributed to the model’s performance. In contrast, the PDP allows us to isolate and interpret the influence of each individual land cover class, offering a more nuanced ecological interpretation of how Russian Olive interacts with its environment.

|  |  |
| --- | --- |
| **LEVEL2** | **yhat** |
| Blank | 0.05 |
| Agriculture | 2.71 |
| Alpine Sparse and Barren | 5.37 |
| Conifer-dominated forest and woodland (mesic-wet) | 5.44 |
| Conifer-dominated forest and woodland (xeric-mesic) | 6.96 |
| Deciduous Shrubland | 6.71 |
| Developed | -3.75 |
| Floodplain and Riparian | 3.61 |
| Harvested Forest | 0.36 |
| Mining and Resource Extraction | -3.7541597 |
| Mixed deciduous/coniferous forest and woodland | 3.72678161 |
| Montane Grassland | 6.28932458 |
| na | 5.55560784 |
| Open Water | 0.3317129 |
| Recently burned | 5.69602685 |
| Sagebrush Steppe | 0.54032671 |
| Wet meadow | 0.44689006 |

Table 3: Sub-variables of Land Cover

The yhat values in Table 3 represent the **predicted probability of Russian Olive presence** for each land cover type within the **LEVEL2** variable, based on the partial dependence plot (PDP). In this context, yhat is the model’s estimated response (i.e., likelihood of presence) when the given land cover type is present, while all other environmental variables are held constant.

For example:

* A yhat value of **6.96** for *Conifer-dominated forest and woodland (xeric-mesic)* suggests that this land cover type is associated with a **higher likelihood** of Russian Olive presence.
* In contrast, a yhat of **-3.75** for *Developed* areas indicates a **lower likelihood**, meaning Russian Olive is less likely to be found there.

These values help reveal how individual land cover types influence the model’s predictions, even though LEVEL2 was treated as a single categorical variable in the main model. The PDP approach breaks it down, offering insight into which specific habitats are suitable for Russian Olive based on the model’s behavior.

A map of mountains and valleys

AI-generated content may be incorrect.Using these results, I generated a habitat suitability map for Missoula County (see Figure 7). This map visualizes the predicted probability of Russian Olive presence, ranging from 0 (low suitability) to 1 (high suitability). To create this continuous surface, I applied a spatial interpolation method known as **Kriging**, using the **predicted probabilities** from the Random Forest model as the input values. Kriging leverages spatial autocorrelation, where nearby locations tend to have similar environmental conditions, to estimate probabilities for unsampled areas, creating a smooth and realistic representation of habitat suitability. The resulting raster highlights areas with high probability of containing Russian Olive with a darker shade and areas that do not in white.

Figure 6: Habitat Suitability Map

The habitat suitability map reveals several key patterns:

1. **Riparian Zones:** Russian Olive is most likely to be found near rivers and streams, consistent with its known preference for moist, disturbed habitats.
2. **Urban and Suburban Areas:** The model also identified high suitability in developed areas, likely due to ornamental plantings and human-mediated dispersal.
3. **Remote Areas:** Some high-suitability areas were located far from water sources, suggesting that Russian Olive could potentially spread outside of known areas.

**The model achieved an accuracy of approximately 98%, as indicated by an out-of-bag (OOB) error rate of about 2%. The OOB error is an unbiased estimate of the model's generalization performance, calculated using samples not included in the training of each tree. To further evaluate the performance, I produced a ROC curve (see Figure 8), which displays the model’s ability to distinguish between presence and absence points. The AUC (Area Under the Curve) value of .989 indicates it has strong performance of predicting the presence of Russian Olive given the environmental variables.

Figure : ROC Curve

Despite these results, the model has several limitations, many of which stem from uncertainties in the input data. First, the model’s accuracy is highly dependent on the quality and reliability of the presence points. It assumes that locations marked as having Russian Olive are accurate representations of the species' current distribution. However, this assumption may not always hold true. Trees may have been removed, newly established, or misidentified since the data was collected, especially in areas where land use changes rapidly.

Additionally, a significant portion of the presence data from the statewide dataset, consists of “provisional” points, meaning they have not been confirmed through field verification. These unverified observations introduce uncertainty into the model, as they may not accurately reflect real-world conditions.

The same applies to the pseudo-absence points, which were randomly generated across the study area and are assumed to represent locations without Russian Olive. However, since these areas have not been surveyed, it is possible that some of them do contain Russian Olive. This mislabeling could reduce the model’s ability to clearly distinguish between suitable and unsuitable habitat, ultimately affecting its predictive accuracy.

Another limitation is the amount of data, overall, there are around 800+ presence points for Russian Olive, this is a large enough set to provide some predictive power but still a relatively small sample size.

Finally, the model assumes that the environmental conditions driving Russian Olive distribution remain consistent over time, which may not account for future changes due to climate change or human activity. The model’s predictions are limited to the spatial and temporal scope of the data, meaning it may not fully capture rare or emerging patterns of invasion.

## 4.3 Identifying Areas of Interest

Beyond model evaluation, I wanted to explore spatial patterns in model predictions to identify potential hotspots or outlier areas for Russian Olive suitability. To do this, I calculated the Z-score for each data point based on the model's predictions. The Z-score measures how much a value deviates from the mean, helping to pinpoint unusual or standout locations that differ significantly from average suitability. Points with particularly high or low Z-scores may indicate especially suitable or unsuitable habitats, making them valuable targets for further investigation.

Once the Z-scores were calculated in R, I exported the point data to ArcGIS. This step streamlined the mapping process and improved integration with my final digital product. In ArcGIS, I first considered binning points into spatial units (e.g., grids or hexagons), which aggregates point values across defined areas. However, simple binning can obscure finer spatial trends.

To provide more detailed spatial context, I used a point interpolation method called Kriging. Kriging not only aggregates nearby points but also accounts for spatial autocorrelation—the principle that points closer together are more likely to have similar values. Using the Z-score as the input variable, I created a continuous raster surface representing variation in suitability across the landscape. This helped highlight clusters of outlier locations and provided a richer visualization of potential Russian Olive habitats that might not be obvious from raw point data alone.

Figure 8 below displays this, with the darker red indicating areas that might be outliers in the model results.

A map of the state of colorado

AI-generated content may be incorrect.

Figure :Variance Map

# Recommendations

**Recommendations**

While I am not an expert in invasive species removal, my personal experience—gained through assisting my father, a longtime arborist—has taught me the importance of careful planning and consideration of ecological impacts. Russian Olive, despite being invasive, can provide habitat for native birds and insects. This complicates efforts to remove it without also considering the potential consequences for local wildlife.

To address this, I recommend the following strategies in the areas of risk:

**1. Community Engagement, Education, and Surveying**

Before large-scale removal, invest in early detection and public involvement to minimize spread and improve the accuracy of site data.

**Public Education and Outreach**  
Educate local landowners, farmers, and recreation groups about the ecological risks of Russian Olive and how to identify it. Some of the highest-risk areas identified in this model intersect with publicly accessible or privately owned lands. Public awareness can significantly increase early detection and reduce future spread.

**Volunteer Surveys and Citizen Science**  
Engage the community in citizen science programs to support field surveys, especially in areas marked as “provisional” or unconfirmed. Volunteers can help confirm presence-absence data, track regrowth, and foster community investment in restoration.

**Cost-Saving Benefits**  
Early detection through community involvement can save the department significant time and money by reducing the need for intensive treatment. Accurate surveys also improve modeling and help direct resources more effectively.

**2. Targeted Removal and Treatment**

In areas with confirmed high density, use a combination of removal and long-term treatment. According to Lesica and Miles (2001), mature trees respond well to herbicide every 10 years, while younger trees may only need treatment every 30 years. This controlled approach minimizes ecological disruption while reducing long-term reinfestation risk.

**3. Replacement with Native Species**

After removal, restore ecological balance by planting native species such as Cottonwood, commonly displaced by Russian Olive and well-suited to riparian zones. Reintroducing native vegetation can prevent reinvasion and stabilize habitat.

**4. Monitoring and Adaptive Management**

Implement long-term monitoring of treated and replanted sites. Regular surveys will help assess Russian Olive regrowth, native species establishment, and overall ecosystem recovery. Based on these insights, adjust strategies over time using adaptive management to ensure sustainable outcomes.

# Conclusion

This project significantly enhances our understanding of Russian Olive distribution and habitat suitability in Missoula County. By providing the Missoula County Ecology Extension and the Missoula County Weed Board with actionable insights, this report supports targeted management efforts to control the spread of Russian Olive. While the task of managing and eradicating invasive species in the county is substantial, I hope this project serves as a valuable resource and paves the way for future efforts to address new invasive species in the region.

A key goal of this project was to create a flexible, modular framework that can be adapted for other species of interest without the need for a complete rebuild of the modeling pipeline. By designing the workflow to accommodate new presence-absence data and environmental variables with minimal reconfiguration, this system becomes a scalable tool that supports ongoing ecological monitoring and management efforts.

This approach is inspired by existing systems such as the Montana Natural Heritage Program, which enables users to input data for various species and generate predictive habitat models. Similarly, the framework developed in this project can be repurposed to address future concerns, whether related to invasive plants, rare species, or shifting habitat conditions due to climate change.

From a management perspective, this adaptability is especially valuable. It allows county or agency staff to quickly respond to emerging threats, run updated models with new data, and make evidence-based decisions without investing significant time or technical overhead. Over time, this can lead to substantial savings in labor and analytical resources, enabling more efficient allocation of personnel and budget toward fieldwork, outreach, or restoration projects.

Moreover, by integrating spatial analysis and citizen science data, this tool can act as a bridge between technical research and community-driven conservation, encouraging collaboration between scientists, land managers, and the public. Ultimately, building this kind of adaptable infrastructure lays the foundation for a proactive and sustainable invasive species management strategy in Missoula County and beyond

# Appendices

# References

Barbet-Massin, M., Jiguet, F., Albert, C.H. and Thuiller, W. (2012), *Selecting pseudo-absences for species distribution models: how, where and how many?* Methods in Ecology and Evolution, 3: 327-338. <https://doi.org/10.1111/j.2041-210X.2011.00172.x>

Lesica, P. and S. Miles. 2001. *Natural history and invasion of Russian olive along eastern Montana rivers*. Western North American Naturalist 61(1):1-10.

Lesica, P., M.T. Lavin, and P.F. Stickney. 2012. *Manual of Montana Vascular Plants*. Fort Worth, TX: BRIT Press. viii + 771 p.

Montana Natural Heritage Program. 2023. *Elaeagnus angustifolia (Russian Olive) predicted suitable habitat model* created on March 14, 2023. Montana Natural Heritage Program, Helena, MT. 17 pp.

Program, M. N. H. (n.d.). *Russian olive*. Montana Field Guide. Retrieved March 23, 2025, from https://fieldguide.mt.gov/speciesDetail.aspx?elcode=PDELG01010

Tobalske, C., Vance, L. K., United States Environmental Protection Agency, & Montana Natural Heritage Program. (2017). *Predicting the distribution of Russian olive stands in eastern Montana valley bottoms using NAIP imagery*. Montana Natural Heritage Program.

1. Lesica, P. and S. Miles. 2001. *Natural history and invasion of Russian olive along eastern Montana rivers*. Western North American Naturalist 61(1):1-10. [↑](#endnote-ref-1)
2. Lesica, P. and S. Miles. 2001. *Natural history and invasion of Russian olive along eastern Montana rivers*. Western North American Naturalist 61(1):1-10. [↑](#endnote-ref-2)
3. Lesica, P. and S. Miles. 2001. *Natural history and invasion of Russian olive along eastern Montana rivers*. Western North American Naturalist 61(1):1-10. [↑](#endnote-ref-3)