Using GIS to Assess Species Distributions in the Pacific Northwest

Leslie Perry

Follow this and additional works at: https://digitalcommons.du.edu/geog_ms_capstone

Part of the Geographic Information Sciences Commons

Recommended Citation
https://digitalcommons.du.edu/geog_ms_capstone/67

This Capstone is brought to you for free and open access by the Department of Geography and the Environment at Digital Commons @ DU. It has been accepted for inclusion in Geography and the Environment: Graduate Student Capstones by an authorized administrator of Digital Commons @ DU. For more information, please contact jennifer.cox@du.edu,dig-commons@du.edu.
Using GIS to Assess Species Distributions in the Pacific Northwest

Leslie Perry

University of Denver Department of Geography

Capstone Project

for

Master of Science in Geographic Information Science

November 20, 2015
A B S T R A C T

A species distribution application was developed using ArcGIS (Model Builder) and Python to support management decisions and species reviews for the BLM and Forest Service in the Pacific Northwest. The application integrates observation data for flora and fauna species from agency geodatabases and available data on various environmental factors, such as ownership, land uses, and forest types. Results of the application include species-specific geodatabases containing observation data and combined observation-environmental data with tables presenting distribution information based on the environmental factors. These results are intended to provide baseline information about species distribution patterns and can be used for a variety of purposes. The application can also be adapted with different input data and be modified to incorporate other geoprocessing tools.
TABLE OF CONTENTS

Abstract ...................................................................................................................... ii
Introduction .............................................................................................................. 1
Background ............................................................................................................. 2
Purpose and Goals ................................................................................................. 5
Benefits of Project ................................................................................................. 7
Literature Review .................................................................................................. 8
Design and Implementation ..................................................................................... 16
Agency Input .......................................................................................................... 16
Literature Review .................................................................................................. 18
Application Concept and Design ........................................................................... 19
Data Sources .......................................................................................................... 22
Software/Hardware/Network Requirements ........................................................... 25
Application Operators ........................................................................................... 26
Challenges .............................................................................................................. 26
Results .................................................................................................................... 30
Discussion ............................................................................................................. 33
Areas for Further Research .................................................................................... 34
References Cited .................................................................................................... 37
Appendix A. Species Distribution Application Guide

LIST OF FIGURES AND TABLES

Figure 1. Northwest Forest Plan Area ................................................................. 4
Figure 2. Application Window for Flora Model ................................................... 21
Figure 3. Overall Application Design ................................................................. 22
Table 1. Input Data for Species Distribution Application .................................. 22
Figure 4. Forest Data ........................................................................................... 24
Figure 5. Sample Observation Data by Land Ownership for Boletus pulcherrimus (a
Fungus) .................................................................................................................. 31
Figure 6. Sample Frequency Tables for Ownership and Land Use Allocations ...... 32
Figure 7. Sample Output Geodatabase Files ....................................................... 32
Figure 8. Sample Hot Spot Analysis Data ............................................................ 33
INTRODUCTION

The U.S. Department of Agriculture (USDA) Forest Service (Forest Service) and U.S. Department of the Interior (USDI) Bureau of Land Management (BLM) are responsible for managing more than 42 million acres of lands across the Pacific Northwest, which encompasses Washington, Oregon, and northern California. As one of their management responsibilities, the agencies are tasked with managing species diversity and identifying species at risk due to declining populations. In addition to species listed as threatened or endangered under the Endangered Species Act (ESA), the agencies have developed targeted lists of species to manage, including Sensitive, Strategic, and Survey and Manage species. Due to a variety of factors, such as limited funding and the inability to conduct surveys across all of their lands, the agencies have few resources to gather information on species distributions and life history requirements, particularly for rare species.

In collaboration with the Forest Service and BLM staff in Portland, this project was completed to establish a new geographic information system (GIS) application that compiles and analyzes the distribution of recorded observations of species across the Pacific Northwest using available GIS data (referred to herein as the "species distribution application" or "application"). This application was designed to improve efficiency of species assessments and support agency management decisions in the Pacific Northwest. It
incorporates user input for the input data, file locations, and species to analyze and can be adapted to meet the user's needs.

**Background**

The Forest Service and BLM manage federal lands in accordance with resource management plans developed for each National Forest or BLM District. These plans provide management guidelines and policies for various resource topics and establish land use allocations for agency-managed lands that dictate the allowed uses of the lands. National-level guidance from Forest Service and BLM Manuals also requires that the agencies establish lists of "Sensitive" and "Strategic" species. Sensitive and Strategic species are plant and animal species identified by the agencies in coordination with other federal and state agencies that demonstrate a population viability concern, as evidenced by significant current or predicted downward trends in population numbers or density and habitat capability such that a species' existing distribution could be reduced. These species are not listed under the ESA, but may be warranted for listing with additional research or may have previously been listed (i.e., delisted). The species may be listed on State lists of threatened and endangered species.

The agencies are tasked with conserving Sensitive and Strategic species and ecosystems they depend on. The agencies are directed to implement management actions in a manner that improves the condition of
the species such that they no longer warrant protection and ensures the actions do not contribute to the need to list the species under the ESA.

In the Pacific Northwest, the agencies also developed the Northwest Forest Plan (NWFP) in 1994 as an amendment to existing resource management plans and as guidance for future plans to establish a comprehensive ecosystem management strategy for federally managed lands in the range of the northern spotted owl (a threatened species under the ESA) (USDA and USDI 1994). This management strategy focused on the management of habitat for the owl (late-successional old-growth [LSOG] forests in portions of WA, OR, and CA) and related species found in LSOG forests. The plan specifically established standards and guidelines for rare species found in LSOG forests and referred to them as "Survey and Manage" species. The agencies established strategies to survey for the species and manage known sites (hence the name "Survey and Manage"). These standards and guidelines were updated in 2001 (USDA and USDI 2001).

Survey and Manage species include certain amphibians, mollusks, arthropods, mammals, birds, bryophytes, lichens, fungi, and vascular plants that meet characteristics defined in the plan. These characteristics include their range (known to occur or have habitat in the range of the northern spotted owl, see Figure 1), habitat requirements (closely associated with LSOG forests), and status as a rare or endemic species (little may be known about the species or they may be known to have limited distributions).
Figure 1: Northwest Forest Plan Area

Plan Area Boundary:
BLM- and Forest Service-Managed Lands
in the Range of the Northern Spotted Owl

Legend
- Range of Northern Spotted Owl
- State
- Federal Land Manager
  - BLM
  - USFS

Data source: USFS/BLM data, Census data for states
Data projection/datum: NAD83, UTM Zone 10N
Prepared by: Leslie Perry, GEOG 3150, Fall 2014
Also as part of their responsibilities, the agencies are tasked with periodically reviewing the status of Sensitive and Strategic species to determine if they continue to warrant being on their lists and reviewing new information available on Survey and Manage species to determine if the species still meet the requirements of Survey and Manage standards and guidelines. During these reviews, the agencies may look at changes to federal or state listing status, global and state rankings, name changes, recent studies about the species, survey data, and other factors. The need to review survey data served as the basis for this project and was used to develop the project goals, as discussed below. Key spatial factors include: distribution patterns across multiple scales (range of the northern spotted owl, physiographic provinces, species' known range in the region, BLM/NFS lands, etc.), number of known sites or observations on federal lands and in various land allocations (e.g., late successional reserves, matrix), and proportion of known sites or observations in reserves (these are protected areas) and in LSOG forests.

**Purpose and Goals**

The species distribution application was developed based on BLM and Forest Service management plans and guidelines to improve efficiency of agency reviews of species they manage and provide useful information about the species to support agency decision-making processes. The specific purposes of the project are to:
- Create an application for the agencies to learn more about the species they manage, focusing on distribution patterns across different scales using available GIS data.

- Integrate appropriate spatial and statistical tools to assess possible patterns of observations and relationships with various environmental criteria, recognizing the limiting factor of the observation data (they are not presence/absence data).

- Create an application that can be run automatically for each species managed by the agencies, be easily updated with new data, and automatically create files specific to each species.

- Establish a reporting process that compiles information on species' distributions based on the output of the application.

The application includes geoprocessing tools to extract observation data for individual species and analyze that data to compile information on the species' known distributions. The primary goal of the application is to provide information on the species to address research questions the agencies may have, such as:

- How many sites or observations of the species are found on BLM or National Forest System land, on each National Forest in the Pacific Northwest, or in each BLM District in the Pacific Northwest?

- How many sites or observations of the species are found in each land use allocation? What proportion of the sites is in a protected allocation (i.e., reserves)?

- How often is the species found in each forest type and in LSOG forests? Is this pattern statistically significant?

- Where do hot spots (high density areas) of the species occur, recognizing that this may be attributed to more surveys of some areas?

- Is the pattern of each species' distribution random, uniform/dispersed, or clustered across regional and local scales (e.g., physiographic provinces, counties)?

- Is the species' distribution related to other species in the same genus or with similar characteristics?
Benefits of Project

The agencies have compiled and maintained GIS data for the species in agency geodatabases (known as Natural Resource Information System [NRIS] and Geographic Biotic Observations [GeoBOB]) and for other environmental characteristics, such as regional forest cover, land ownership, and land use allocation data. These data provide an opportunity for the agencies to obtain valuable information about the species they manage and use spatial analysis processes to assess spatial relationships and trends that could help the agencies manage their lands. Limited staff resources have resulted in limited efforts to analyze the data they have.

An application that efficiently and quickly processes and analyzes regional data on species observations and other factors will be useful to the Oregon and Washington offices of the BLM and Forest Service by filling information gaps and reducing staff time to obtain spatial distribution information for species. For example, the application will provide information on the numbers of species observations in different land use allocations that can be used when the agencies review the current statuses of species and determine the appropriate level of management guidance. If few observations are in protected land use allocations, land use changes may be warranted to ensure protection of the species, for example. For many species, an information gap exists that makes landscape-level management difficult—little is known about many species' habitat requirements and
preferences. Agency staff could also use the application results, specifically the numbers of observations in different forest types and in LSOG forests or statistical relationships between the observations and forest types, to support research and management of individual species or analyses of project-level impacts.

Evident patterns in species observations may be attributed to survey locations, but they may also provide information on areas that could be the focus of additional survey efforts. Integrating survey boundary data with the results could help identify a pattern in the observations. The flexibility of the application to incorporate other data will allow the analyses to be conducted across different scales. The focus of the application is on regional scale data, but local or project-level data, if available, could also be used as input data if desired. For project-level analyses, the application can be tailored to intersect observation data with project data to assess impacts, and the ability to run the application for multiple species with the click of a button would improve the efficiency of doing such an analysis.

**Literature Review**

Prior to development of the species distribution application, a variety of reports and articles were reviewed to assess methods used on similar projects that evaluated species distribution patterns and to help identify appropriate techniques and GIS-based tools for the application. As noted below, some sources provided useful information, whereas others were more
advanced or involved different types of data (e.g., raster-based analyses). Some previous studies may be relevant to future distribution analyses conducted by the BLM and Forest Service, as discussed in the section "Areas for Further Research" below.

In addition to the planning documents prepared by the BLM and Forest Service that were used as the primary basis for the project, other studies conducted by the agencies provide background information and data for the application. In 2011, the agencies published a report on the "Status and Trends of Late-Successional and Old-Growth Forests." This report compared the mapped extent of forests in the NWFP area (range of the northern spotted owl) between about 1994 and 2008 (Moeur, et al. 2011). It was developed as part of a 15-year effectiveness monitoring program for the NWFP. The authors developed maps of forest composition and structure for the beginning and end years of the monitoring period and developed annual maps of forest disturbance. These maps were used to assess the distribution and trends of LSOG forests on lands in the NWFP area. This regional data are also anticipated to be used for future monitoring efforts as the agencies continue to track the status of land management efforts with implementation of the NWFP management guidelines. The resulting LSOG data and the forest composition and structure data were used as input datasets for the species distribution application.
The agencies also conduct studies using GIS to manage species and habitat on their lands. On the Willamette National Forest and in the Salem District of the BLM in Oregon and Washington, Bacheller (2005) developed a GIS model of high-probability habitat for a rare vascular plant. The model combined species observations with habitat criteria (e.g., vegetation, elevation, streams) to identify habitat areas that have a high probability of supporting the plant. This model could be useful in combination with the species distribution application to help the agencies with species management on their lands.

Similarly, Larson, et al. (2003) developed habitat suitability index models using GIS for 12 terrestrial vertebrate species that represent different management concerns to assist the Forest Service with landscape-level management in southern Missouri. The models incorporated tree and other ecological land type data, road and other developed area data, slope, elevation, and other relevant data to assess habitat suitability for the species. The models were intended to also be useful for other species in other locations, assuming local data are available to incorporate into the models, and may be useful in combination with the species distribution application to support management efforts.

Wulff, et al. (2013) evaluated spatial distributions and correlations of various plant species in New Caledonia, an island off the coast of Australia, to identify narrow endemic species and assist with future conservation
efforts of native plants in the country. Pearson’s correlation coefficient was used to compare distributions of non-ranked species with those of known high risk species. A grid was overlaid to count the number of species in each 4-square-kilometer cell. The authors also modeled habitat suitability based on known ecological preferences of the species and then grouped the species by ecological similarities. Other data (e.g., mine locations, climate conditions) were used to evaluate habitat conditions and possible threats to species. Statistical results were used to identify correlations between the species being assessed and other species that have been assigned a conservation status in order to determine if any species without a current status might warrant being identified as a species with conservation concerns. Several hot spots of species were identified based on the higher densities of species in some areas, although this may partially be attributed to prime sampling localities. Correlations between species distributions were identified and were considered statistically significant for some non-ranked species and others that were considered critically endangered or endangered, indicating that those species should be prioritized for ranking and further assessment. The statistical processes were evaluated for incorporation into the species distribution application, but the other techniques used in this study relied on raster-based data and served a different purpose than the application.
Crain, et al. (2011) evaluated species richness and hot spots of globally and locally rare plants in Napa County, California using spatial analysis tools in ArcGIS. They used a 1-square-kilometer grid overlay to convert observation data to raster format and used the Raster Calculator to identify hot spots based on cell values. The analysis resulted in the identification of several hot spots of both globally and locally rare plants and of gaps between hot spots that might also provide habitat for the species. In addition, the authors compared the extent of protected land with the locations of rare plants and the hot spots to determine to what extent the plants might be protected. Only some of the hot spots fell on protected lands, and the authors concluded that about 4 percent of the county would benefit from being protected to preserve the rare plants. Similar to Wulff, et al. (2013), the statistical processes were useful, but the techniques used in this study relied on raster-based data.

Aspinall, et al. (1998) looked at species distributions based on recorded observations and environmental conditions to assess distribution patterns of two species in two different landscapes (Scotland and Wyoming). Their work shows how GIS can benefit species management and support environmental decision making. Concepts presented by Aspinall, et al. were considered for the species distribution application.

Belongie (2008) used GIS to develop a model to identify prime habitat for the gray wolf in the western upper peninsula of Michigan based on
written literature. Raster datasets for population density, road density, prey density, and land use/land cover were ranked and combined to develop the model. The final habitat suitability model was compared against observations of wolf packs, human population densities, available prey density (white-tailed deer), and road density. The model was developed using ArcToolBox to be adapted to other locations with other datasets. This study served a different purpose than the species distribution application, but the use of tools in ArcToolBox was relevant to see how other studies use ArcGIS.

Carpenedo (2011) developed a habitat suitability model to identify potential stream segments that could support the relocation of nuisance beavers in Montana. The model incorporated vector and raster datasets and was set up as a script to be adaptable to other datasets in other locations. Logistic regression analysis was used to characterize the relationship between presence/absence of the beaver along stream segments and identify independent variables that influence this relationship. The result of the model was a map that identifies potential beaver translocation sites to be further evaluated and utilized by natural resource managers. Similar to Belongie (2008), this study was not directly relevant, but it demonstrates the use of GIS and spatial statistics to support agency management decisions.

Danzinger (2011) used GIS to develop a habitat suitability model for the wolverine to assess the feasibility of reintroduction efforts in Colorado.
Using previous models, Danziger identified three significant parameters: elevation, human population densities, and road densities. A combination of vector and raster data were used in the model, with the output being raster data that were analyzed for suitability based on the significant parameters.

The Landscape Fragmentation Tool from University of Connecticut was used to further analyze the results of the model, then a wildlife corridor model was developed to identify the least cost path between main habitat areas. Similar to Carpenedo (2011), this study was not directly relevant, but it demonstrates the use of GIS to support agency management decisions.

Pereira and Itami (1991) conducted a study using GIS and statistics to analyze habitat suitability for the Mt. Graham red squirrel on the Coronado National Forest in Arizona. The study was designed to integrate observation data from surveys with statistical analyses using GIS and identify habitat where the squirrel is most likely to be found. Various environmental factors were processed through logistic multiple regression models to assess habitat suitability, and the results were integrated with observation data using Bayesian statistics inference techniques. The results of the study were used to conduct a project-level habitat impact analysis and could be used by the Forest Service to manage the species and its habitat. Although newer GIS tools and techniques are available now than were in the early 1990s, the concepts of this study could be applied to species in the Pacific Northwest to
assess habitat suitability and aid the BLM and Forest Service in management of the species.

Brown (2014) developed a set of GIS-based tools to support species distribution modeling. These tools evaluate species richness, landscape connectivity, spatial segregation of known occurrences, climate-based distribution change predictions, statistical analyses, and various basic data processing tools to support the more advance tools. Most of the tools require Spatial Analyst and involve raster-based analyses, but they are adaptable to different input data and parameters that can be specified by the user. These tools may be helpful to the agencies in subsequent analyses of species distributions to support their management efforts, but they were more advanced or specific than what was intended for the application.

Etherington (2011) developed Python scripts to support spatial analysis of landscape genetics. The scripts are designed to work in ArcGIS, but are adaptable to work with other GIS programs. They provide tools to evaluate connectivity between points, such as species observations, and look at Euclidean distances, cost-distances, and least-cost paths between the points. The results of these analyses are intended to support research on landscape genetics and incorporate a spatial analysis component into the research. For rare species in the Pacific Northwest, these tools could be useful for understanding distribution patterns of the species and possibly
determining where the species are most likely to be found, with adaptations of the tools to incorporate certain environmental conditions.

**Design and Implementation**

Development of the species distribution application involved a review of relevant literature and studies, coordination with BLM and Forest Service staff, identification of data and geoprocessing tools to incorporate, review of Python programming language and ArcGIS Model Builder functions, coordination with a Python specialist, and compilation of tools and scripts to produce the application. Initial coordination with the agencies began in fall 2014 to determine if the application was of interest to them and to discuss how it might be developed. At that time, a draft implementation plan was prepared to discuss components of the application and its intended use. In spring 2015, a draft proposal was prepared and submitted to the agencies to get feedback on the project and discuss the implementation steps. Agency input was incorporated into the proposal to produce a final proposal in fall 2015. Data gathering and literature review took place between fall 2014 and spring 2015 and were completed in fall 2015. Application development began in late summer 2015.

**Agency Input**

The BLM and Forest Service expressed interest in the application as a tool to support status reviews of species and other management actions. They provided insight on data, GIS tools, and software available to them and
how the agencies use GIS for management decisions. The main contacts who helped with this project were Carol Hughes, an Interagency Special-Status/Sensitive Species Specialist with the Forest Service Region 6 and Oregon State Office of the BLM in Portland, and Kelli Van Norman, the Inventory Coordinator of the Interagency Special-Status/Sensitive Species Program with the Forest Service Region 6 and Oregon/Washington Office of the BLM.

The agency contacts expressed some concerns regarding the limitations of the observation data and species naming conventions. The observations of species only represent where surveys have been conducted. They are not considered to be an actual representation of species distributions and do not incorporate negative results (i.e., confirmed surveyed area with no detections of a species). Also, surveys have primarily been conducted where projects have been implemented, although some additional surveys were conducted prior to 2006 to obtain information on species locations (e.g., through the Random Multi-Species Surveys). Some species are known by multiple names, and the species data may not reflect all the names or provide a linkage between alternate names. The agencies have attempted to establish crosswalk databases to link alternate names for species, but have had difficulty maintaining them and integrating them with their species geodatabases. These limitations may affect the accuracy of the
results of the application and will need to be considered prior to use of the results for agency management actions.

Literature Review

The literature review focused on other studies that involved GIS modeling for species distributions or assessments. Most literature seemed to focus on habitat suitability models developed using GIS, with less literature focused on species distribution models. The review entailed browsing articles and reports for the author’s use of GIS to determine how and to what purpose GIS was used. If the study involved GIS modeling or programming (e.g., with Python), the article or report was more closely reviewed to determine if the GIS tools and study purpose were relevant to this project. For studies involving statistical methods using GIS, the tools and processes used for the studies were reviewed and assessed as to their relevance for this project. Online blogs and ArcGIS help forums were also reviewed as needed to develop scripts and tools for the application.

Literature developed by the agencies was also reviewed to gather background information on their management processes and direction and to assess available data used for other studies in the Pacific Northwest. Agency management documents served as the basis for the application, and species evaluation criteria (e.g., for Survey and Manage species) were reviewed and incorporated into the application processes. Agency reports and websites
were also helpful for better understanding how the agencies operate and what tools they have available to them.

Application Concept and Design

The overall concept of the species distribution application is to overlay recorded observations of species of management concern with various data available for the NWFP area to assess distribution patterns and relationships. The application is accessed through a toolbox in ArcGIS and runs automatically using Model Builder (see Species toolbox image at right). Data used as inputs for the models during development are packaged with the application. These data are easily interchangeable in the event newer or more appropriate data become available. A model is included with the application to process species observation data prior to running the main model ("Species Input Data" model). Two main models are included with the application to analyze the processed species observation data for fauna and flora separately ("Main Model for Species (Fauna)" and "Main Model for Species (Flora)"). The input and output species data are compiled in geodatabases, and output results are saved in the output species geodatabase, depending on the file type, or as separate files in a designated folder. The application uses species codes for flora, based on the input species data and established by the agencies or U.S. Department of
Agriculture, and abbreviated codes for fauna, also based on the scientific name, for naming the output files to keep track of the results.

Some pre-processing of the non-species input data used for the analyses was necessary prior to developing the application, specifically to merge and standardize datasets from multiple sources. These data were processed manually to incorporate specific attributes based on the parameters and tool requirements in the application, and updated or different data will need to match the tool requirements. More information on the data is presented in the next section and in Appendix A.

The application toolbox, named "Species," contains five models and two Python scripts:

- **Species Input Data:** The model establishes the input species geodatabase that contains the combined versions of flora and fauna observations and a dissolved version of the combined data for the analyses.

- **Main Model for Species (Flora) and Main Model for Species (Fauna):** The models prompt the user to select the input data, species of interest (one or more may be selected), and output folder. Separate flora and fauna models are provided because of the differences in file naming for each type of species.

- **Species Analysis Processes (Flora) and Species Analysis Processes (Fauna):** Submodels contained within the Main Models for Species that run the analysis processes and repeat the analyses if multiple species are selected.

- **Choose Multi Value Script:** A Python script that is used in the Main Model for Species to establish a list of species to pick from based on user-selected data and the field to pull values (e.g., scientific names) from for the list.

- **Fauna Name:** A Python script that converts the scientific name of fauna species, based on the input species data, into an abbreviation
for use with file naming. The abbreviation is set at the first three letters of the first name and first three letters of the second name.

Figure 2 displays the application window for selecting one or more species and running the application (shown for Flora). Figure 3 displays the overall processes contained in the Species Analysis Processes model. With the models being in Model Builder, the tools used in them can easily be updated through ArcGIS, and the user would only need a basic understanding of ArcGIS and Model Builder to modify the application. If desired, the models could be exported to Python scripts and modified through Python by users familiar with this programming language. Additional details on the models and scripts are presented in Appendix A.

Figure 2. Application Window for Flora Model
Data Sources

All input data used for the application are from secondary sources; no primary data collection or surveys were conducted. The input data used for the application development process can be updated with user-provided data when the application is run. This section discusses the data used for application development. Table 1 lists the key input data for the application. No data were purchased for this project.

Table 1. Input Data for Species Distribution Application

<table>
<thead>
<tr>
<th>Data</th>
<th>Feature Type</th>
<th>Use in Application</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined Flora Observations; Combined Fauna Observations</td>
<td>Geodatabase (polygon)</td>
<td>Species table to select species to analyze; main data of observations combined from NRIS and GeoBOB into one layer for flora and one layer for fauna</td>
<td>Processed from NRIS and GeoBOB extracts</td>
</tr>
<tr>
<td>Master Flora; Master Fauna</td>
<td>Geodatabase (polygon)</td>
<td>Input data for species analyses; dissolved to avoid duplication of observations</td>
<td>Dissolved from Combined Observations data</td>
</tr>
</tbody>
</table>
# Table 1. Input Data for Species Distribution Application

<table>
<thead>
<tr>
<th>Data</th>
<th>Feature Type</th>
<th>Use in Application</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>NWFP Area</td>
<td>Shapefile</td>
<td>Plan area boundary for Survey and Manage species analyses</td>
<td>BLM/Forest Service</td>
</tr>
<tr>
<td>NWFP Provinces</td>
<td>Shapefile</td>
<td>Physiographic province boundaries for Survey and Manage analyses (&quot;Name&quot; field used for analysis); distribution across provinces</td>
<td>BLM/Forest Service</td>
</tr>
<tr>
<td>Land Ownership</td>
<td>Shapefile</td>
<td>Distribution across land ownerships (&quot;Owner&quot; field used for analysis)</td>
<td>BLM (WA, OR), State (CA); data combined into one shapefile with refined attributes</td>
</tr>
<tr>
<td>BLM/Forest Service Land Owners</td>
<td>Shapefile</td>
<td>Distribution across BLM- and Forest Service-managed lands</td>
<td>Extracted from Land Ownership data</td>
</tr>
<tr>
<td>National Forests</td>
<td>Shapefile</td>
<td>Distribution across National Forests</td>
<td>Forest Service</td>
</tr>
<tr>
<td>BLM Districts</td>
<td>Shapefile</td>
<td>Distribution across BLM Districts</td>
<td>BLM</td>
</tr>
<tr>
<td>NWFP Land Use Allocations</td>
<td>Shapefile</td>
<td>Distribution across land use allocations in NWFP area for Survey and Manage analyses (&quot;LUA&quot; field used for analysis)</td>
<td>BLM/Forest Service, USGS for &quot;Riparian Reserves&quot;</td>
</tr>
<tr>
<td>Counties</td>
<td>Shapefile</td>
<td>Distributions across counties for WA, OR, and CA</td>
<td>U.S. Census TIGER, States</td>
</tr>
<tr>
<td>Forest Types</td>
<td>Shapefile</td>
<td>Distribution across forest types in NWFP area for Survey and Manage analyses (&quot;GRID CODE&quot; used for analysis)</td>
<td>Processed from forest raster data obtained from LEMMA</td>
</tr>
<tr>
<td>LSOG Forests</td>
<td>Shapefile</td>
<td>Distribution in LSOG forests in NWFP area</td>
<td>LEMMA</td>
</tr>
</tbody>
</table>

Note: Additional details on the data and attributes can be found in Appendix A.

The Forest Service and BLM maintain NRIS and GeoBOB, respectively. These geodatabases contain the input data for species observations and provide the best representation of species observations on BLM- and Forest Service-managed lands in the Pacific Northwest. Because the geodatabases
are updated on a regular basis, the agencies provided an extract of the geodatabases for the project (the data were current as of December 2014), which should be updated each time the application is run to incorporate the latest observation data. NRIS and GeoBOB data are not considered public domain data and have restricted use by the agencies due to the sensitivity of species location information.

The Forest Service teamed with Oregon State University to compile digital gradient nearest neighbor imputation maps of portions of Washington, Oregon, and California as part of the Landscape Ecology, Modeling, Mapping & Analysis program. Data are provided in raster format as 30-meter resolution ArcGIS grids and only cover forest land (defined as areas with at least 10% tree cover). The raster data were reclassified and generalized to create a shapefile containing forest types (coniferous, mixed hardwood-coniferous, and hardwood forests) in the NWFP area (see Figure 4 at the right). The program also developed LSOG forest data for the NWFP area that are used in the application.

Other data were available from the agencies and other public domain sources, as noted in Table 1 above. Most data were in shapefile format and
were projected (or re-projected) in NAD 83, UTM Zone 10N. Non-spatial information was also used, such as tables of the species’ names and regulatory statuses.

Software/Hardware/Network Requirements

The Forest Service and BLM currently use ArcGIS; therefore, the application was developed using ArcGIS and is intended to run with ArcGIS 10.2.2 or compatible versions. The application requires an advanced license to run because of the use of the Frequency tool (adaptations to some processes and scripts in the application may be possible to allow it to run with the basic license). Data conversion and processing of the forest type data required the Spatial Analyst extension, but this was done outside the application and the application does not require this extension. Other software that was used during application development included Microsoft Word, Microsoft Excel, Adobe Acrobat, and Python (IDLE).

Hardware already available to the agencies to run GIS and other programs on computers at the agencies’ offices will be adequate to run the application. The application and associated data are packaged together in a single folder to be saved in a designated folder on the agencies’ servers. Depending on the specific network configurations, the data and application should be able to be shared via existing network connections, similar to how the NRIS and GeoBOB geodatabases are maintained.
Application Operators

The application was designed to be used as needed with minimal agency staff time and no maintenance requirements until it is used. Existing staff in the agencies' Interagency Special-Status/Sensitive Species Program or GIS group will likely oversee and run the application. User training can be provided at the onset once the agencies receive the application. An application guide was prepared as part of the application development (see Appendix A) to present information on how the application runs and what can be done to troubleshoot errors.

Challenges

Development of the application involved reviewing tools and scripts developed by others for similar purposes and adapting those for this application. Some troubleshooting was required until the models ran as they were intended. One of the biggest challenges was establishing a tool that incorporated user input and allowed the user to select more than one value that would then be used as the input into the analysis tools. Model Builder has an "iterator" tool that can be used in a single model to run through tools multiple times, but the selection of input values required integration of a Python script into the model. Python also has a "loop" script that functions similar to the iterator tool, but the use of this script for the overall analysis processes would have required exporting or creating the analyses using Python script instead of Model Builder. For functionality, keeping the analysis
tools in Model Builder was preferred over a Python script. Several different scripts were tested and altered until one was created that provided the desired function.

Another challenge with the application was the use of variables as file names and parameters and making sure the variables were in the right format and properly referenced throughout the model. For example, the naming of output files needed to match the species being analyzed, but required an abbreviated version of the species’ scientific name. The input species data contains a field for species codes, which all flora have but only some fauna have. This species code was established as a variable by extracting it from the input species data once the species was selected and using it as the name for output files. This step worked for flora, but required modification to work for fauna.

Because of the differences in the available attributes for flora compared with fauna, the initial models worked great for flora, but the file naming system (using a variable for species code) did not work for all fauna. With the input species data being separate for flora and fauna, the application was set up to have two sets of models, one for flora and one for fauna. The submodel for fauna incorporates the use of Python script for abbreviating the species’ scientific name in order to use it for file naming in the model.
As noted by the agencies, the need to cross-reference alternate species names will benefit the application by providing a means to incorporate the latest accepted scientific names and any alternate names and ensure all names of a species are selected when extracting observations from the species data. This is more of a concern for flora, particularly fungi and lichens, because the taxonomic status of many of these species is still under review. Although the agency geodatabases are regularly maintained, frequent changes to scientific names may not be fully integrated into the geodatabases due to staff limitations and other priorities. The accepted and alternate names would need to be updated in a table that can be related to the species input data, then queried during the selection process. This is expected to require the use of a Python script and still needs to be researched.

Some level of spatial statistics analysis was desired for the application, but the relevance and use of the results for agency purposes was uncertain. The basic distribution-based statistical analyses were incorporated as geoprocesses in the application to present information on types of distributions (e.g., clustered, dispersed) and attempt to identify hot spots of observations. However, these analyses will only be useful for species with a large quantity of observations. Without knowing how many observations a species has before selecting it for analysis, the application may run into an error when it gets to the statistics tools. This could pose a problem with the
application completing its processes and will need to be troubleshooting as this error occurs. The statistical results will also require user interpretation, and the output files of the statistical tools will need to be reviewed and saved in a usable format.

Limitations of the input data were also a concern, but the ability to replace data with more appropriate or newer data will ensure the user is using the best available data when running the application. One of the main data limitations is the lack of regional data (i.e., across the NWFP area) that accurately represents land ownership and land use allocations. GIS data are available for these topics, but they do not necessarily represent the current on-the-ground conditions, which is primarily a result of agency and staff limitations to maintain up-to-date GIS data for these topics. For project- or local-level analyses, this should not be a concern if the user replaces the regional data with current local data. However, with the regional data, some errors may occur, such as the Forest Service land ownership not completely matching the National Forest boundaries or land use allocations not reflecting the latest agency-designated use. The user will need to review the results and determine the extent of possible errors to identify if the results are acceptable for use or if the errors should be corrected, to the extent practicable, by using different data.
RESULTS

This project was not a standard research project to answer specific questions; instead, it provides a tool to support other research and answer questions about species distributions, focused on species managed by the BLM and Forest Service in the Pacific Northwest. The outcome of this project is an application based on models and scripts developed using ArcGIS’ Model Builder and Python programming language. The application integrates GIS-based data for species observations with environmental data, such as ownership, land uses, and forest types. The spatial analyses included in the application provide information on how species are distributed based on recorded observations (i.e., not necessarily a full representation of their distribution) and how these observations are related to various environmental factors.

Throughout the application development process, the models and geoprocessing tools were tested on a number of species to assess the functionality of the application and conduct a preliminary review of the results. Sample results are presented in this section for a fungus, Boletus pulcherrimus, which is a Survey and Manage species. Examples of the output results include:

- species-specific geodatabases named according to the species code (flora) or abbreviated species name (fauna) (for example, "BOPU4" for Boletus pulcherrimus);
- an extraction of species observations from the master species data;
- intersected species observations with physiographic provinces, land ownership (see Figure 5 for sample output data), BLM and Forest Service ownership, BLM Districts and National Forests, and land use allocations;
- tables depicting the counts of species observations from each intersection (from data listed previously) and spatial join (for forest types, LSOG forests, provinces, and counties) (see Figure 6 for sample tables); and
- statistical data, tables, and/or graphs of statistical results.

Figure 5. Sample Observation Data by Land Ownership for Boletus pulcherrimus (a Fungus)
Compatible data and tables are saved in the species-specific geodatabase (see Figure 7 at right), and other results from the statistical processes are saved in a generic output folder. The tables of counts may include double-counted observations for observations that overlap multiple environmental factors; for example, one observation could be on both BLM and Forest Service land and would be counted twice. If the totals in the tables do not match with the total number of observations, this is likely the reason.

Interpretation of these results is not the focus of this project and is expected to be done by the agencies when they are ready to conduct species analyses. However, the information and data generated by the application

**Figure 6. Sample Frequency Tables for Ownership and Land Use Allocations**

**Figure 7. Sample Output Geodatabase Files**
can be expected to quickly provide counts of species observations across multiple scales and in different environmental conditions that can be used to support agency decision-making processes. The statistical results (see Figure 8 at the right for sample hot spot analysis) may also be useful for supporting further analyses or conducting habitat-level analyses to assess future areas to survey or where to focus management efforts for the species. For species with few observations, the results may not be as informative, although they will inform the agencies that the species are known to occur in few locations and may warrant continued protection.

DISCUSSION

This project was conducted to establish an application with GIS-based tools and data that can improve efficiencies of species distribution analyses for the BLM and Forest Service. The application was also designed to provide flexibility with the input data and accommodate easy modification of the parameters and variables to adapt it for different purposes and with different datasets. This functionality provides an opportunity for the agencies and
others to incorporate the application into their GIS-based analyses. The overall concepts of the application—to select one or more values and run those values through a series of geoprocessing analyses—can be used for a variety of purposes.

This application was specifically designed with species observation data provided by the BLM and Forest Service and is intended to provide information on species distributions across a portion of the Pacific Northwest in order to support BLM and Forest Service management actions and planning for species they manage. It can be used by agency staff or consultants to the agencies on an as needed basis.

**Areas for Further Research**

With further research, the application may be improved to include other geoprocessing tools and spatial analysis processes and to export results of the model in a report format, such as in Microsoft Word. Other analyses may be desired depending on the use of the application, and as the need arises, additional tools may be added to the application to enhance its use. The statistical analyses may also be modified or adapted to provide different types of statistics to assess relationships between the species observations and different data. For project-level analyses, these adaptations may include incorporating site-specific data and creating maps to display the results.
Other refinements to the application to add functionality may include adding tools to join or relate species data with information from other tables, such as regulatory status. This type of refinement could allow the user to then select a group of species for analysis based on their regulatory status (for example, all Survey and Manage species).

Python has the ability to further enhance the functionality of the application through the development and incorporation of scripts into the application. For example, Python has a dictionary that allows the user to export information from one source into a Microsoft Word document. A search and replace-type script would enable the application to export results of the geoprocessing tools into a Word template that is set up with tables and text to report the results. This option would provide species-based reports that could be used as supporting documentation for agency decisions and be provided in a format that anyone can read.

Some of the literature reviewed for this project should also be further reviewed to determine if the types of tools and processes used for the studies could be adapted to conduct further research on agency-managed species in the Pacific Northwest. Tools developed by Brown (2014) and Etherington (2011), for example, may be adapted with data from the agencies to conduct focused research on rare species or model habitat for species with limited life history information. Results of the species
distribution application may also help the agencies determine which species should be the subject of more focused research.

As new tools and functions become available in newer versions of ArcGIS, the application may also need to be adapted to work best in the latest version. The application was developed using ArcGIS 10.2.2, which has more tools and functions than older versions, such as ArcGIS 9.3. ArcGIS is constantly being updated and adapted to improve its operations, and further research on these improvements should be conducted to enhance the application accordingly.
REFERENCES CITED


Crain, Benjamin J., Jeffrey W. White, and Steven J. Steinberg. 2011. "Geographic Discrepancies Between Global and Local Rarity Richness
Patterns and the Implications for Conservation." Biodiversity Conservation 20: 3489-3500.


Overview of Application

The species distribution application was developed using ArcGIS 10.2.2 (ArcInfo version) and is accessible via ArcToolbox. The application includes various models developed using Model Builder and scripts developed using Python 2.7. The input data default to data used during application development, but they can be modified to any dataset meeting the same parameters, as described in "Data Requirements."

The primary intended purpose of this application is to assess distribution patterns of Survey and Manage species to support species reviews and analyses conducted by the U.S. Department of the Interior, Bureau of Land Management (BLM) and U.S. Department of Agriculture, Forest Service (Forest Service). However, the application can also be used for other species with recorded observations in agency geodatabases with some modifications to the input data and parameters. To support the primary purpose, the application contains the tools/models listed below and described further in "Geoprocessing Models":

- Select species observations by attribute (species name) from master geodatabase and import into a new species-specific geodatabase
- Generate a count of total species observations in the region (range of the northern spotted owl)
- Conduct statistical analyses of species distributions based on the observations (nearest neighbor and multi-distance spatial cluster analysis)
- Intersect species observations with physiographic province data, generate a count of observations per province, conduct a hot spot
analysis across the provinces using a spatial join, and produce graphs of the statistical results

- Perform a spatial join of species observations and counties, conduct a hot spot analysis across the counties, and produce graphs of the statistical results

- Intersect species observations with land ownership data and generate a count of observations per owner type

- Intersect species observations with BLM and Forest Service ownership data, intersect results with BLM Districts and National Forests data, and generate a count of observations in each District and on each Forest

- Intersect species observations with land use allocation data, generate a count of observations per allocation type, and calculate the proportion of observations in reserve land allocations [not currently part of application]

- Perform a spatial join of species observations with forest types and generate a count of observations per forest type

- Select species observations that fall within regionally mapped late-successional old-growth (LSOG) forests, get a count of the total, and calculate the proportion of observations in LSOG forests [not currently part of application]

Data Requirements

Multiple datasets are necessary to run the application. All input and output data are projected in NAD83, UTM Zone 10N or have a geographic coordinate system of North American 1983. The default input data include:

- Agency-provided extracts of GeoBOB and NRIS species geodatabases

- Master geodatabase of species observations, combined and modified from GeoBOB and NRIS

- Boundary data for the range of the northern spotted owl (Northwest Forest Plan area), physiographic provinces, states, counties, BLM Districts, and National Forests from BLM, Forest Service, or public domain sources (e.g., U.S. Census TIGER Line Files)

- Land ownership data for California and Washington/Oregon from State of California and BLM
- Land use allocation data for BLM and National Forest System lands in the range of the northern spotted owl from BLM and Forest Service
- Forest cover raster data for the range of the northern spotted owl from BLM and Forest Service, reclassified by forest type (coniferous, mixed hardwood-coniferous, and hardwood) and converted to shapefile format
- Late-successional old-growth forest cover for the range of the northern spotted owl from BLM and Forest Service

Some of the data have limitations because of their regional scale. Land ownership data were created by merging data available from different sources (BLM and State of California), which contained different attributes and were developed at different times. These data may also not represent the latest information on ownership across the region, but they do represent the best available regional-scale data for land ownership and were determined to be the best data for this application. Similarly, the land use allocation data from the agencies does not incorporate changes to these designations since the data were developed (2002), and the data do not include riparian reserves, which are designated at a local level (by National Forest or BLM District). The forest type data are a reclassification of forest cover data, which represent one interpretation of the forest cover data. The original data could be reclassified using other factors to generate different results.

Each of the input datasets can be replaced with a newer version of the data or a different dataset based on user input, but the data need to be in the same format (i.e., only replace polygon data with polygon data or
geodatabase data with geodatabase data) with attributes that match the
criteria in the models to avoid complications with the models. The application
presents an option for the user to select the master species dataset to use;
the default is set with the data that were available at the time the
application was developed. Other data can be updated, as well, but may
need to be changed within the models themselves (see "Troubleshooting"
section).

The application generates output data based on user-specified values
for the species to evaluate and the geoprocesses to run. For a typical
application run, the following output data and files will be generated:

- Species geodatabase, named based on the species code (for flora)
or abbreviated species name (for fauna) and used to save all
  subsequent data
- Species observations feature layer (named based on geodatabase
  name, which is used for all subsequent data)
- Species observations intersected and joined with physiographic
  province data
- Species observations intersected with ownership data
- Species observations intersected with BLM and National Forest
  System lands data
- Species observations intersected with BLM Districts and National
  Forests boundaries data
- Species observations intersected with land use allocations
- Species observations joined with forest types
- Species observations joined with late-successional old-growth
  forests
- Species observations joined with county boundaries
- Files depicting the results of counts generated from the
  geoprocessing steps and results of statistical analyses
**Geoprocessing Models**

As noted above, the application contains various geoprocessing tools as models to run with the default or user-provided input data. This section discusses the models included in the application and the parameters established to provide flexibility with the application.

*Species Input Data Model*

The species input data model combines input data from GeoBOB and NRIS and generates an output geodatabase of all species observations from the agency geodatabases (refer to Figure A-1 below). The input feature classes include observation points (NRIS flora points, GeoBOB fauna observations) and polygons or sites (GeoBOB and NRIS flora sites/polygons, NRIS fauna polygons). It includes steps to buffer point data, merge polygon data, and modify attributes to simplify and standardize the data for purposes of the application. This model may need to be modified to incorporate queries for the data prior to processing, such as to select reproductive sites of certain fauna. Alternatively, the input data could be queried with a definition query prior to running the model to only pull in those features that satisfy the query.
Figure A-1. Master Species Model Flow Diagram

This model also dissolves the observations to combine overlapping polygons with the same species code (flora) or scientific name (fauna); the resulting master species feature classes are used as the primary input into the rest of the application. Two separate processes are currently set up in the model to establish "Master_Flora" and "Master_Fauna" feature classes. Parameters set up for the model include the input datasets and buffer distances for point data; these parameters can be modified prior to running the model. Output data include a combined feature class and a dissolved
feature class for flora and fauna with the fields and attributes listed in Table A-1.

### Table A-1. Attributes of Species Data

<table>
<thead>
<tr>
<th>Field Name</th>
<th>Description of Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Combined Flora and Fauna Data</strong></td>
<td></td>
</tr>
<tr>
<td>OBJECTID</td>
<td>Default ID for polygon assigned by ArcGIS</td>
</tr>
<tr>
<td>Obs_Date</td>
<td>Date of observation from GeoBOB or NRIS (if available)</td>
</tr>
<tr>
<td>Sp_Code</td>
<td>Assigned species code for flora and fauna (if available) from GeoBOB or USDA</td>
</tr>
<tr>
<td>Sci_Name</td>
<td>Scientific name of species as reported in GeoBOB or NRIS</td>
</tr>
<tr>
<td>Source</td>
<td>Source of the observation (GeoBOB or NRIS)</td>
</tr>
<tr>
<td>Date_Run</td>
<td>Date data were created</td>
</tr>
<tr>
<td>Shape_Length</td>
<td>Automatic length field</td>
</tr>
<tr>
<td>Shape_Area</td>
<td>Automatic area field</td>
</tr>
<tr>
<td><strong>Master Flora and Fauna Data</strong></td>
<td></td>
</tr>
<tr>
<td>OBJECTID</td>
<td>Default ID for polygon assigned by ArcGIS</td>
</tr>
<tr>
<td>Sp_Code (Flora only)</td>
<td>Assigned species code for flora and fauna (if available) from GeoBOB or USDA; field used for dissolve of combined flora data</td>
</tr>
<tr>
<td>Sci_Name (Fauna only)</td>
<td>Scientific name of species as reported in GeoBOB or NRIS; field used for dissolve of combined fauna data</td>
</tr>
<tr>
<td>Shape_Length</td>
<td>Automatic length field</td>
</tr>
<tr>
<td>Shape_Area</td>
<td>Automatic area field</td>
</tr>
</tbody>
</table>

Once the species geodatabase was created, it was enhanced with the addition of tabular data containing lists of the species, alternate species names or codes, and regulatory statuses of the species. This tabular information was imported from a Microsoft Excel workbook, which can be updated as species names or regulatory statuses change. The relationships and tables can be viewed in ArcCatalog, and tables can be added to ArcGIS.
Species Geodatabase and Selection Geoprocesses

The first step of the main application (main model for species) is to identify the input data, field to select by, and output folder and select the species to evaluate (parameters for the model). Once the species are selected (one or more species can be selected to run through the model), the model creates a geodatabase specifically for each species with a feature class containing the extracted species observations from the master species data. Depending on the species selected (how many and which ones), the model may take about an hour or more to run through all the steps. A submodel runs the geoprocesses automatically as part of the master model based on user input (Figures A-2 and A-3), and the output data feed into the rest of the submodel to run geoprocessing tools. The default naming convention for the geodatabase and species feature class is the flora species code, as depicted in the master species data, or an abbreviation of the fauna species’ scientific name, as depicted in the master species data. [Tools to be developed: The selection model also pulls in observations of the species based on alternate species codes or names imported from the Excel workbook into the species geodatabase. If alternate species codes or names exist, the model updates the attribute field of the species data to use the current accepted code or name.]
Figure A-2. Main Model for Species Flow Diagram

The tools included in the main model for species from Figure A-2 are:

- **Choose Multi Value Script**: This is a Python-based script file that includes specific parameters for a feature class, field, and value. It includes a validation code that presents a list of values derived from the field attribute of the feature class. The list allows the user to check one or more of the values to run through the model. When the model is opened from ArcToolbox, a window pops up that presents the parameter options; when the "OK" button is clicked, the model automatically runs in its entirety for the selected values.

- **Species Analysis Processes Submodel**: This model pulls in the "Values to Iterate" based on the user selection and runs them through the processes outlined in the next section. The parameters of the submodel are the output folder location and species data, which can be selected by the user at the main window when selecting input data and the values to use. Other parameters can also be set, such as the specific ownership or land use allocation data to use as inputs for the model, but this requires a modification to the model and is not the current default.

---

1 The models appear mostly white because they are waiting for user input on the values to use. Default data and values are already set for the components of the model in color. The "P" indicates a parameter for user input.
The tools shown in Figure A-3 for a portion of the submodel for species analysis processes are:

- **Iterate Multivalue**: This tool pulls in all the selected values from the main model and tells the model to run all subsequent processes for each value. This is a Model Builder-only tool, and Model Builder only allows one “iterator” tool in a model. Because of this subsequent steps were set up as individual processes, although some additional efficiency may be realized if multiple iterations could be done (this would require use of Python scripts).

- **Make Feature Layer and Select Layer by Attributes**: These tools are creating a temporary selection of the species from the combined species feature class in order to pull in the species code for flora and use that code for naming and selection purposes later in the model. The species code of the selected species becomes a variable in the model and is cross-referenced later. For fauna, the selection is by scientific name and a Python script creates the file name from that name instead of these tools and the Get Field Value and Calculate Values tools described below.

---

2 The submodel was set up specifically for use with the flora data, which have set species codes already. A separate model is provided for fauna species that essentially runs the same but abbreviates species names for file naming.
- **Get Field Value and Calculate Value**: These tools extract the species code from the selected data and create the variable called "Name" for use with the subsequent processes.

- **Create File GDB**: This tool creates a file geodatabase in the designated output folder with the species code (Name). This geodatabase is used as the output location or workspace for subsequent processes.

- **Select Layer by Attribute and Feature Class to Feature Class**: These tools select the species from the master species data (note this is the dissolved version of the combined species data, which prevents duplicate features) and export the selected data to a new feature class in the file geodatabase set up in the previous step. The name of the new feature class is the "Name" variable.

**Intersection and Frequency Geoprocesses**

For each geoprocessing step that entails an intersection, the selected species data are intersected with another dataset (e.g., physiographic provinces, land ownership) to produce a new feature class in the selected species geodatabase with a name specific to the species and the intersected data. The output data are then queried to count the number of observations with specific attributes (e.g., owner as BLM, province as Cascades, LUA as OTHER), and the counts by attribute are exported to database tables in the geodatabase. The purpose of these tables with counts is to quickly obtain summaries of the numbers of observations in each unique attribute of the intersected data. The geoprocesses rely on specific fields in the input data to run the queries, as listed in Table A-2. The input data used to intersect or analyze the selected species data can be exchanged for other data with minimal effort, but the fields noted below should be incorporated into the input data to avoid complications with the model.
**Table A-2. Key Fields and Attributes of Model Data**

<table>
<thead>
<tr>
<th>Data</th>
<th>Field and Attributes</th>
<th>Use in Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Species Feature (output of first steps of submodel)</td>
<td>Sp_Code (unique code for flora species)</td>
<td>Data of species observations; file name derived from Name variable and used as name value for all output data and tables</td>
</tr>
<tr>
<td></td>
<td>Sci_Name (unique name for fauna species)</td>
<td></td>
</tr>
<tr>
<td>NWFP_Provinces (input physiographic provinces data)</td>
<td>NAME (name of physiographic provinces)</td>
<td>Field used to count observations that fall within each province</td>
</tr>
<tr>
<td>Owner_WA_OR_CA (input merged land ownership data)</td>
<td>Owner (name of owner; e.g., USFS, BLM, ST, PVT)</td>
<td>Field used to count observations that fall within each ownership status</td>
</tr>
<tr>
<td>Owner_USFS_BLM (input ownership data extracted for USFS and BLM owners)</td>
<td>Owner (USFS or BLM)</td>
<td>Field used to select species observations managed by each agency</td>
</tr>
<tr>
<td>ForestType (input forest cover data)</td>
<td>GRIDCODE (designated gridcodes established during conversion of raster data to shapefile; 1 = coniferous, 2 = mixed hardwood-coniferous, 3 = hardwood)</td>
<td>Field used to count observations that fall within each forest type</td>
</tr>
<tr>
<td>R5_R6_Forests (input National Forest boundaries data)</td>
<td>Mgmt_Unit (name of management unit, e.g., Willamette National Forest)</td>
<td>Field used to count observations that fall within each National Forest</td>
</tr>
<tr>
<td>BLM_Districts_WA_OR_CA (input BLM District boundaries data)</td>
<td>Mgmt_Unit (name of management unit, e.g., Coos Bay District, Redding Field Office)</td>
<td>Field used to count observations that fall within each BLM District or Field Office</td>
</tr>
<tr>
<td>NWFP_LUA_FS_BLM</td>
<td>LUA (code of land use allocation, e.g., CR, LSR, OTHER)</td>
<td>Field used to count observations that fall within each land use allocation on BLM or National Forest System land</td>
</tr>
<tr>
<td>Counties_WA_OR_CA</td>
<td>NAME (name of county)</td>
<td>Data used for spatial join and statistical analyses</td>
</tr>
</tbody>
</table>

Also as part of the model, the total number of species observations from the selected species data and the number of observations that fall within LSOG forests are counted, and the results are exported to database tables in the geodatabase.
The tools shown in Figure A-4 for a portion of the submodel for species analysis processes are:

- **Intersect**: The species feature class is intersected with multiple datasets (all separately) to produce a new feature class that incorporates attributes of the intersected data with the species data. For example, the intersect with the owner data results in owner values added to each species feature, where the data intersect.

- **Spatial Join**: For larger datasets (Forest Types and LSOG), the species observation data is spatially joined with the data to create a new feature class of the joined data. This tool was selected over the intersect tool to minimize processing time.

- **Frequency**: The frequency tool is used to obtain counts of species observations based on specific attributes (e.g., how many observations are in each owner type or in each province?). The result of this tool is a unique table imported to the geodatabase and named using the species name, intersected or joined data, and count.

![Figure A-4. Example Intersection, Spatial Join, and Frequency Flow Diagram](image)

**Spatial Statistics Geoprocesses**

The selected species data are also run through a series of spatial statistics geoprocessing steps to assess statistical significance of distribution.
patterns and relationships between datasets. The processes included with the models are:

- **Average Nearest Neighbor to determine the distribution pattern of the species observations** (clustered, random, dispersed) by evaluating the distance between the feature centroids and their nearest neighbors (Spatial Statistics Tools–Analyzing Patterns–Average Nearest Neighbor, Euclidean Distance);

- **Multi-Distance Spatial Cluster Analysis using Ripley's K-function** to assess how the spatial clustering or dispersion of feature centroids (species observations) changes when the distance band or threshold distance changes (Spatial Statistics Tools–Analyzing Patterns–Multi-Distance Spatial Cluster Analysis, 10 distance bands with 9 permutations as default);

- **Spatial Autocorrelation** to evaluate the statistical significance of the distribution pattern of the species observations joined with the province and county data (Spatial Statistics Tools–Analyzing Patterns–Spatial Autocorrelation, Count as input field, inverse distance with no distance value); and

- **Hot Spot Analysis with Rendering** to identify the locations of statistically significant spatial clusters of high values (hot spots) and low values (cold spots) of the species observations joined with the province and county data (Spatial Statistics Tools–Rendering–Hot Spot Analysis with Rendering, Count as input field, default distance band).

The output results of these processes are HTML files and tables. Default file names are used for HTML files (pending resolution of renaming issues with the statistics tools), so the output files will need to be reviewed to determine which species they belong with (the input data file name is noted in the HTML file to track to the species). The results window in ArcGIS should also be reviewed to determine if any errors or issues were encountered during the processes and obtain some of the results information. User interpretation of these results will be necessary to better
understand the species distributions and applicability of the statistical processes.

Troubleshooting

This section presents information on common errors associated with the application and possible solutions to fix the errors and maintain the application.

The main models for species are linked to the choose multiple script and species analysis processes. If these links are broken, the model will not run. All of the files are saved in a single toolbox, but they can easily be copied into a new toolbox to modify or use with other models or scripts. The files should be copied together to ensure operation of the model. The workspace for the models should default to the folder where the data are saved, but if not, this can easily be updated in the model properties dialogue box. The default workspace for ArcToolbox should be checked and set as the desired output folder prior to running the model. Also, if the Species toolbox is opened in a different map document than the one provided with the application, the data links may need to be verified before running the models.

The statistical geoprocesses may not run correctly for data with too few observations. This should not stop the model, and an error message should appear in the results window once the model is done running.
If multiple species are selected to analyze, the models may take a long time to run (e.g., more than an hour), particularly if the species have numerous observations (e.g., more than a couple hundred). To test the model, it is best to select one or two species to see how long the processing takes.

The application uses tools, specifically the Frequency tool, that require the user to have an Advanced ArcGIS license. The application will generate an error message if a user attempts to run it on a computer with a lower level license. Similarly, the application was set up in ArcGIS 10.2.2 and may not run correctly in lower versions of ArcGIS or in ArcGIS 10.3.

If data errors arise, any modifications to the input data should be verified to ensure they match the original input data used with the development of the application. The data should be in the same format and contain at least one field that matches the parameters used in the models.

If individual tools or scripts are not working correctly, a quick solution is to search GIS StackExchange (http://gis.stackexchange.com/) or ESRI help pages (http://resources.arcgis.com/en/help/) and blogs to determine if others have experienced similar errors.