

Bighorn Sheep Movement and Disease Dynamics

Nathan Justice

Introduction

Bighorn sheep populations have experienced significant decline over the past 150 years. In recent decades, pneumonia outbreaks have proved to be one of the most perplexing challenges facing their restoration and conservation in the Northwestern United States. Pneumonia is introduced into bighorn populations through contact with domestic goats and sheep grazing in shared habitat. Once the pathogen has been transmitted to a herd member, the group mortality rate can range from 30% to 90%. Although some adult sheep survive from immunity, their nursing lambs typically die. Better understanding of outbreak dynamics can facilitate management efforts aimed at alleviating the low annual recruitment of new lambs into populations.

The scope of this project was to investigate the spatio-temporal dynamics of pneumonia spread and develop a classification model that can illuminate potential contributing factors for pneumonia infection in a herd over time. The goal was to address whether or not current management interventions of augmenting herd populations by translocating sheep from other areas is an effective practice or whether this can exasperate the already increasing morbidity rate. The project focused on the Hells Canyon study system in Idaho and utilized data collected by the Idaho Department of Fish and Game from 1970 to 2015. Specifically, the analysis itself concentrated on information collected in 1995 to 2015.

Summary of Datasets Used

Spatial Population Layer - This dataset provided information about the biologically defined herds that the Hells Canyon sheep are organized into (n=15). The spatial layer defines polygons that represent the geographic boundary of each population.

Pneumonia Status - This dataset provided the pneumonia status of each herd from 1995 to 2015. The categories are: healthy (no pneumonia), invasion (new outbreak), infection in lambs, and infection in both lambs and adults.

Population Summaries - This dataset provided information about the population counts of each herd from 1970 to 2015 by the number of ewes, rams, and lambs.

Translocation Events - This dataset provided information about the translocation efforts done between 1997 to 2016 (n=16). The records included the number, age-group, and sex of the sheep relocated and whether the relocation was done into a vacant habitat or into an existing population.

Study Sheep - This dataset provided information on individual sheep that were uniquely identified and monitored throughout their life (n=712) and spans from 1995 to 2015. Important information used includes the sheep's native herd, their sex, and whether they are a resident or translocated individual.

VHF Locations - This dataset represented the geographic coordinates of study sheep that have been radio collared and relocated periodically from 1997 to 2012. Each record represents a relocation instance of an individual (n=75,450).

Methodology

Model Input Table

The model input table was the final representation of the data used in the modeling and analysis. The components of which were derived from the outputs of the preprocessing steps described below. Each preprocessing step contributed multiple features to the input table. The dataset has a record for each herd in each year for which there is data, ranging from 1995 to 2015 depending on the scope of the information gained from the preprocessing step. The meaning of each column are as follows:

- *Herd* – names of distinct bighorn populations
- *Year* – ranges from 1995 to 2015
- *Status* – broad classification of pneumonia status
- *StatusClass* – binary classification of *Status*, where 0 = healthy (no pneumonia detected) and 1 = otherwise (pneumonia infection detected)
- *Area* – the geographic area of the *Herd* in squared meters
- *Perimeter* – the geographic perimeter of the *Herd* in meters
- *PopTot* – total population count of bighorn sheep (sum of *Ewes*, *Rams*, and *Lambs*)
- *PopAdults* – total count of adult bighorn sheep (sum of *Ewes* and *Rams*)

- *Ewes* – total count of ewes
- *Rams* – total count of rams
- *Lambs* – total count of lambs
- *NonResident* – Boolean value where 0 = no translocated individuals were members of the population and 1 = otherwise (value must be 1 if either *RT* or *T* is 1 and 0 if both *RT* and *T* are 0)
- *RT* – Boolean value where 0 = no resident-translocated (moved from one herd in Hell's Canyon to another) individuals were members of the population and 1 otherwise
- *T* – Boolean value where 0 = no translocated individuals were members of the population and 1 = otherwise
- *Translocation* – Boolean value where 0 = no translocation event occurred in the given year and 1 = otherwise
- *TypeBoth* – Boolean value where 0 = neither a translocation into an existing population or a vacant habitat occurred in the given year and 1 = otherwise (value must be 1 if both *TypeSupp* and *TypeVac* are 1)
- *TypeSupp* – Boolean value where 0 = no translocation event into an existing population occurred in the given year and 1 = otherwise
- *TypeVac* – Boolean value where 0 = no translocation event into a vacant habitat occurred in the given year and 1 = otherwise
- *Visitors* – the number of individuals that were non-members of the herd detected in the population boundary
- *VisEwes* – the number of individual ewes that were non-members of the herd detected in the population boundary
- *VisRams* – the number of individual rams that were non-members of the herd detected in the population boundary
- *VisR* – the number of non-translocated individuals that were non-members of the herd detected in the population boundary
- *VisRT* – the number of resident-translocated individuals that were non-members of the herd detected in the population boundary
- *VisT* – the number of translocated individuals that were non-members of the herd detected in the population boundary

Preprocessing Spatial Population Layer

The cleaning process for the spatial population layer began by using ArcGIS Pro software to change the spatial projection of the data from UTM to latitude and longitude coordinates (WGS 84). This had to be done outside of Python because the computer used for the project could not

properly configure the necessary spatial Python packages. Next, a crosswalk was developed using the proper naming conventions contained in this dataset. Building the crosswalk involved going through all of the other datasets and recording all possible naming conventions for each herd by hand. Lastly, the area and perimeter of each herd was extracted and put aside to be later added to the model input table.

Preprocessing the Pneumonia Status Data

The pneumonia status data for each herd was in the form of one file for 1995 to 2011 and individual spatial files for years 2012 to 2015. The 2012 to 2015 files had to first be converted from their Esri format into a CSV using ArcGIS Pro. Using Python, the 2012 to 2015 datasets were then added to the 1995 to 2011 dataset. The value naming conventions between the aggregate dataset and the individual files differed and were therefore standardized to the same meaning. Additionally, the herd names were standardized using the crosswalk. The full pneumonia status dataset was then put aside to be later added to the model input table.

Preprocessing the Population Summaries Data

The cleaning process for this dataset included standardizing herd names, adding features summarizing the total population across all age and sex groups for each herd in each year, as well as summarizing the total number of adults. The population count-related features were extracted and put aside to be added to the model input table.

Preprocessing the Translocation Events Data

After formatting the translocation event dates, new features were added summarizing the total population across adult and lamb age groups. The event type was encoded to include Boolean indicators for translocation into a vacant habitat, supplementing an existing herd, or both types. Next, the records were summed by year. The translocation counts and type features were extracted and put aside to be added to the model input table.

Preprocessing the Study Sheep Data

The cleaning process for this dataset included standardizing herd names and encoding the source (describing the sheep's residency within Hells Canyon) into Boolean indicators for resident-translocate, translocated, and one for either category. The residency features were extracted and put aside to be added to the model input table.

Preprocessing the VHF Location Data

The cleaning process for this dataset included standardizing the herd names, removing records where coordinate values could not be interpreted as numeric values, removing records where the coordinates exceed the bounds of the UTM projection, and where coordinates were too far from the study system to be considered plausible entries. Lastly, the UTME and UTMN coordinates were converted to latitude and longitude values.

Preprocessing the VHF Polygon-in-Point Location Data

The first cleaning step for this dataset was to bring the cleaned VHF location data into ArcGIS Pro and construct a spatial data file from the original CSV. The same latitude and longitude projection as the herd layer (WGS 84) was then added. Next, an intersection tool was applied to perform a polygon-in-point analysis for the VHF locations and the population polygon layer. This method added a feature indicating which population boundary the data point was detected in. The VHF polygon-in-point dataset was brought into Python where the sex and residency source were appended to each record by matching the sheep identification number to the IDs in the cleaned study sheep dataset. Next, the number of visitors broken down by sex and residency source was added to each herd. A visitor was defined as a VHF location instance where the assigned biological herd was different from the polygon the VHF was recorded in. Lastly, the visitor information was summed across each distinct year. The visitor features were extracted and put aside to be added to the model input table.

Building the Model Input Table

Building the final dataset to be used in the modeling process involved aggregating all of the features derived from the outputs of the preprocessing stages described above. Additionally, the pneumonia status was categorized into binary class labels representing healthy or infected. Missing bighorn count data from the compiled population dataset were imputed using the mean values of the years on record. Lastly, the records were trimmed to include only those with pneumonia status values for all years from 1995 to 2015, thereby excluding herds that formed after 1995.

Experimental Design

Given the research question and discrepancies in the time scale across the datasets, the experimental analysis was divided into two mirror processes using different subsets of the model input table. The divisions were:

- *Category A* - The data used spanned from 1995 to 2015 but did not include visitor information derived from the VHF location data because the visitor data spans from only 1997 to 2012. The final, curated table had 235 records and 15 features (including the target).
- *Category B* – The data used spanned from 1997 to 2012 and included visitor information derived from the VHF location data. However, all records with a year value before 1997 and after 2012 were excluded. The final, curated table had 188 records and 21 features (including the target).

Analysis Protocol

The following protocol was applied to each of the division categories:

- 1) Target and feature data were balanced with random oversampling.
- 2) A decision tree was fit to get a broad view of the data structure.
- 3) A random forest was fit with 100 estimators.
- 4) Important features were manually chosen based on random forest ranking. Deliberate effort was placed in the selection process to avoid colinear features that represent the same type of information.
- 5) A logistic regression was fit with the selected features and scored with a 70%-30% training and testing split.
- 6) A logistic regression was fit with all of the features and scored with a 70%-30% training and testing split.

Results

The decision tree created by the *Category A* trial indicated population counts, specifically the number of lambs, contained the most information for classifying the state of herds likely to have pneumonia infection (Figure 1).

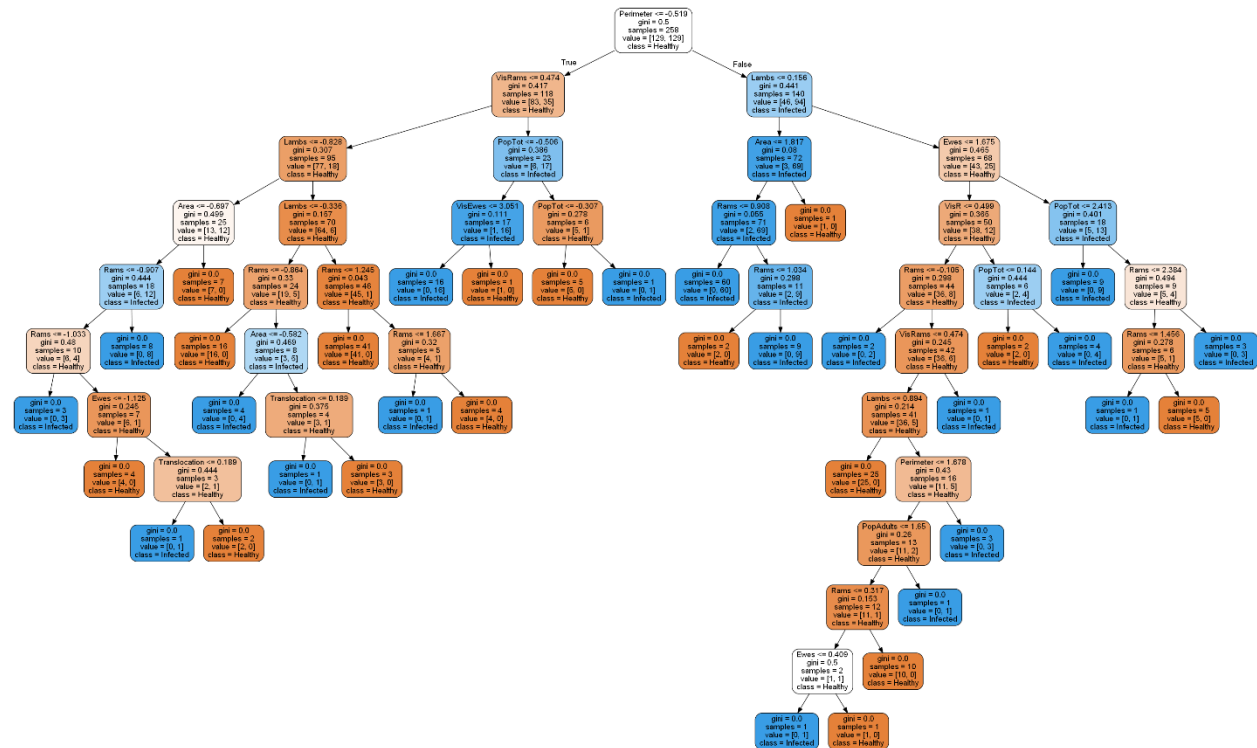


Figure 1. Results of decision tree fit with Category A data.

The decision tree created by the *Category B* trial indicated the size of the herd, as well as the population counts, and the number of visiting rams contain the most information for classifying the state of herds likely to have pneumonia infection (Figure 2).

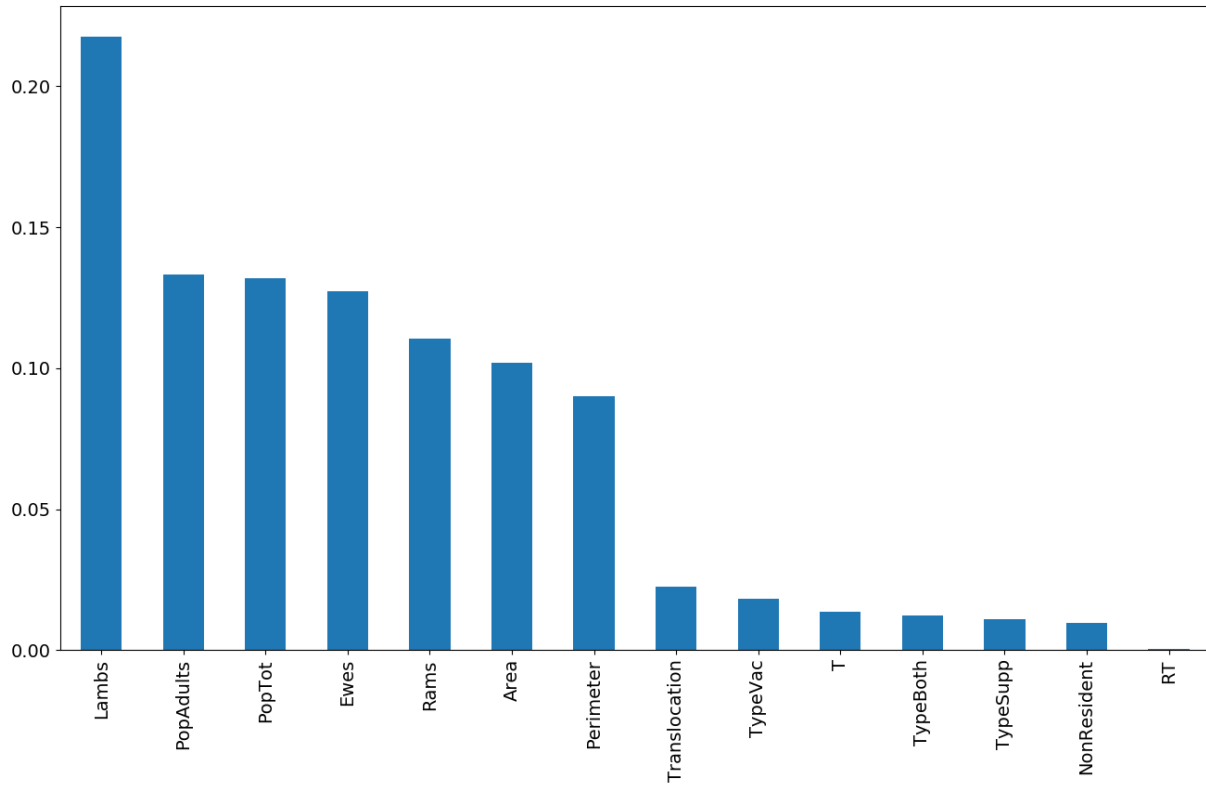


Figure 3. Results of random forest fit with Category A data.

Based on the features of importance illuminated by the random forest ensemble classifier using *Category B* data (Figure 4), the following features were selected as covariates for the constrained logistic regression with *Category B* data: the lamb population count (*Lambs*), the population count of adults (*PopAdults*), the area of the herd (*Area*), whether or not a translocation event occurred in a given year (*Translocation*), and the number of visiting rams (*VisRams*).

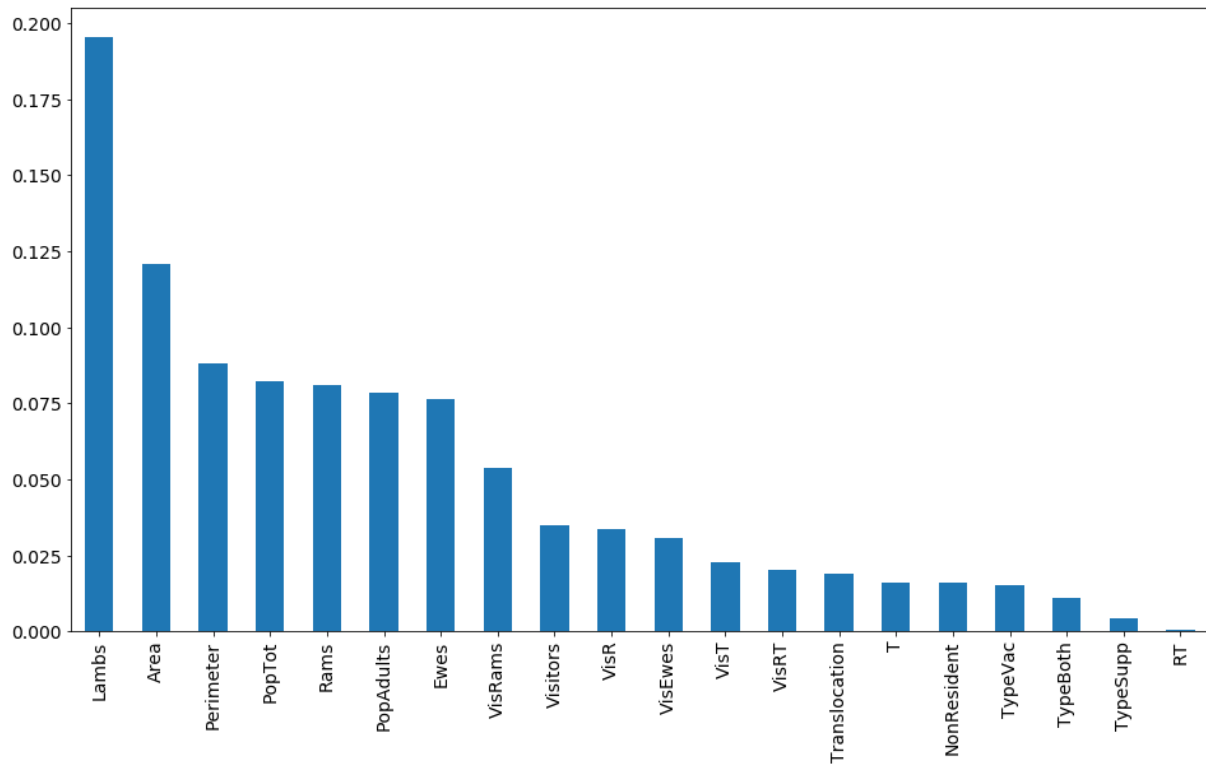
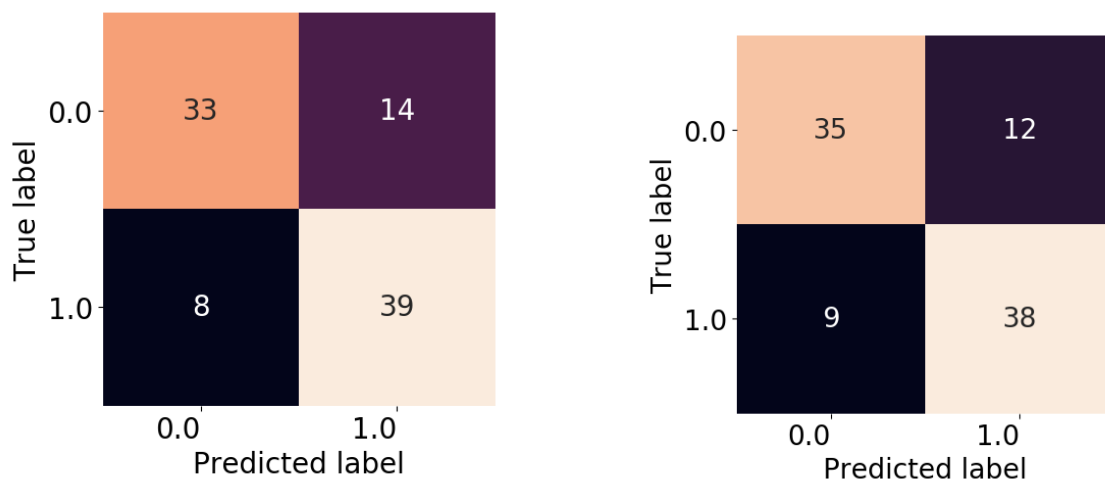


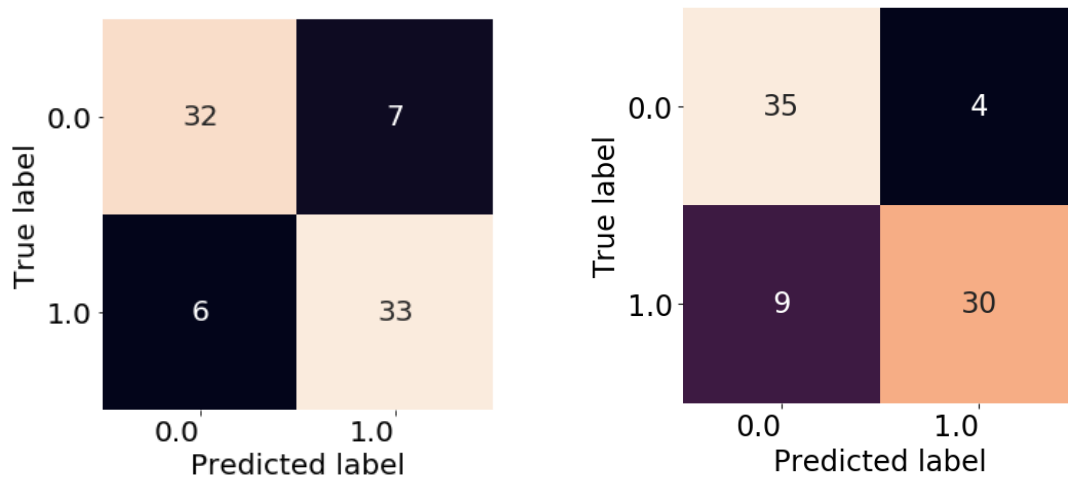
Figure 3. Results of random forest fit with Category A data.

The accuracy of the logistic regression with *Category A* data and with selected features was 77.66%, and the accuracy with all features included was 76.6%.



Confusion matrices for Category A data with selected features and all features, respectively.

The accuracy of the logistic regression with *Category B* data and with selected features was 83.33%, and the accuracy with all features included was 83.33%.



Confusion matrices for Category B data with selected features and all features, respectively.

Conclusion

Based on the results, the following conclusions can be drawn:

- 1) Using random forest to select features to be used in the logistic regression did not significantly impact the accuracy of the herd pneumonia classification.
- 2) The model outputs suggest the number of lambs is an important factor in predicting pneumonia infection. This fits the biological hypothesis that lambs are more susceptible to death from pneumonia because they haven't yet developed a fully functioning immune system. For conservation managers, it would be beneficial to allocate more resources to monitoring herds with high annual birthrates.
- 3) The model outputs suggest including information related to the spatial movement of individuals across the landscape is more beneficial in future analyses, even if it's at the cost of less data collected in other areas. This makes sense from an epidemiological standpoint because pneumonia is transmitted between herds through direct contact with infected individuals.
- 4) The model outputs suggest translocation events do have an impact on the spread of pneumonia across the populations, particularly in relation to translocated rams. This makes

sense biologically because rams are known to be more migratory than ewes, often forming isolated and mobile bachelor groups far from the ewes and lambs, rejoining only during the mating season. Given the evidence translocated rams seem to facilitate the spread of pneumonia, it would be beneficial to conservation managers to consider stopping the translocation of rams into Hells Canyon.

References

All background information comes from personal knowledge working with this study system under the direction of Dr. Raina Plowright at Montana State University. All modeling code was adapted from lecture material provided by Dr. Amir Jafari and Dr. Yuxiao Huang at George Washington University.

Resources

Software

- Python 3
- ArcGIS Pro

Python Packages

- math
- pandas
- geopandas
- utm
- numpy
- imblearn
- sklearn
- matplotlib
- seaborn