Estimating the current extent of beaver-created wetlands in the North Carolina Piedmont

Baines, Blake; Montgomery, Kellyn; Hahn, Rebecca; Yarnell, Patrick
North Carolina State University Raleigh, NC 27695, USA

Abstract

As ecosystem engineers, beavers and the dams they build have powerful hydrologic and geomorphic impacts on riparian landscapes. Beaver dams slow the drainage of streams and create wetlands that provide a multitude of ecological functions including providing valuable habitat for fish and wildlife, improving water quality through nutrient and sediment removal, increasing drought tolerance by maintaining base groundwater flows, and controlling floods by slowing runoff. Beaver populations, however, are often seen as pests and are heavily controlled by state agencies and landowners to prevent property damage. To better understand the trade-offs between beaver nuisance and the ecological services they can provide, an inventory of the current extent of dams and the wetlands they have created is needed. Such information can be used to estimate the opportunity cost of beaver pest control programs through lost wetland potential. This research uses machine learning and geospatial analysis to identify beaver dam locations and their associated wetlands in a watershed in the North Carolina Piedmont.

Introduction

The North American Beaver, Castor canadensis, has been a constant resident across the United States for several hundred years. Pre-European settlement, beavers were estimated to have spanned a geographic range of 15 million km² (Naiman et al.,1988). While modern beaver populations have dwindled, the geomorphic role of beavers in shaping the landscape is still widely evident. Beavers are considered ecosystem engineers, due to the hydrologic and geomorphic effects of their dams on both the landscape and the stream biology (MacFarlane et al., 2017 & Rosell et al., 2005). Beaver dams contribute to the impoundment of water within a stream channel (Naiman et al.,1988). This leads to increased lateral stream connectivity, increased floodplain and terrace development and an elevated water table (MacFarlane et al., 2017). Beaver dams also contribute to the formation of complex wetland and riparian habitats, which increases habitat availability for aquatic and terrestrial plants and animals (Butler & Malanson, 2005 & Rosell et al., 2005). Additionally, beaver dams have been proposed as a method used to restore incised streams. Stream and floodplain restoration can be expensive and is an intrusive process on the local ecosystem. Introducing the beaver to these systems can provide a natural and cost effective solution. Beaver dams have the ability to trap sediment and raise streambed elevations, which ultimately increases stream bank stability (DeVries et al., 2012 & Pollock et al.,2014).

Beavers play a major role in local stream ecosystems and potentially play a role in stream restoration. Therefore, it is beneficial to understand the extent and availability of beaver habitat.

Previous studies have been conducted using Geographic Information Systems (GIS) to identify ideal locations for beaver impoundments along a stream by modeling the habitat quality in combination with local geomorphology (Jakes et al., 2007 & MacFarlane et al., 2017). The goal of this project is to identify current beaver dam locations and estimate the total wetland area created by beaver activity in the Lynch Creek watershed in the Piedmont region of North Carolina using GIS and machine learning techniques.

Study Site

This research is focused on the Lynch Creek watershed (Figure 1), a tributary of the Tar River, located in Franklin County, NC (N 36.137132°, W 78.336621°). Lynch Creek is a 4th order stream located in the Piedmont physiographic region of NC, which is characterized by low, rolling hills. The watershed encompasses an area of approximately 91 km². Due to time constraints, an approximately 10.4 km² section in the southernmost part of the watershed with known beaver dam locations was selected to test our approach.

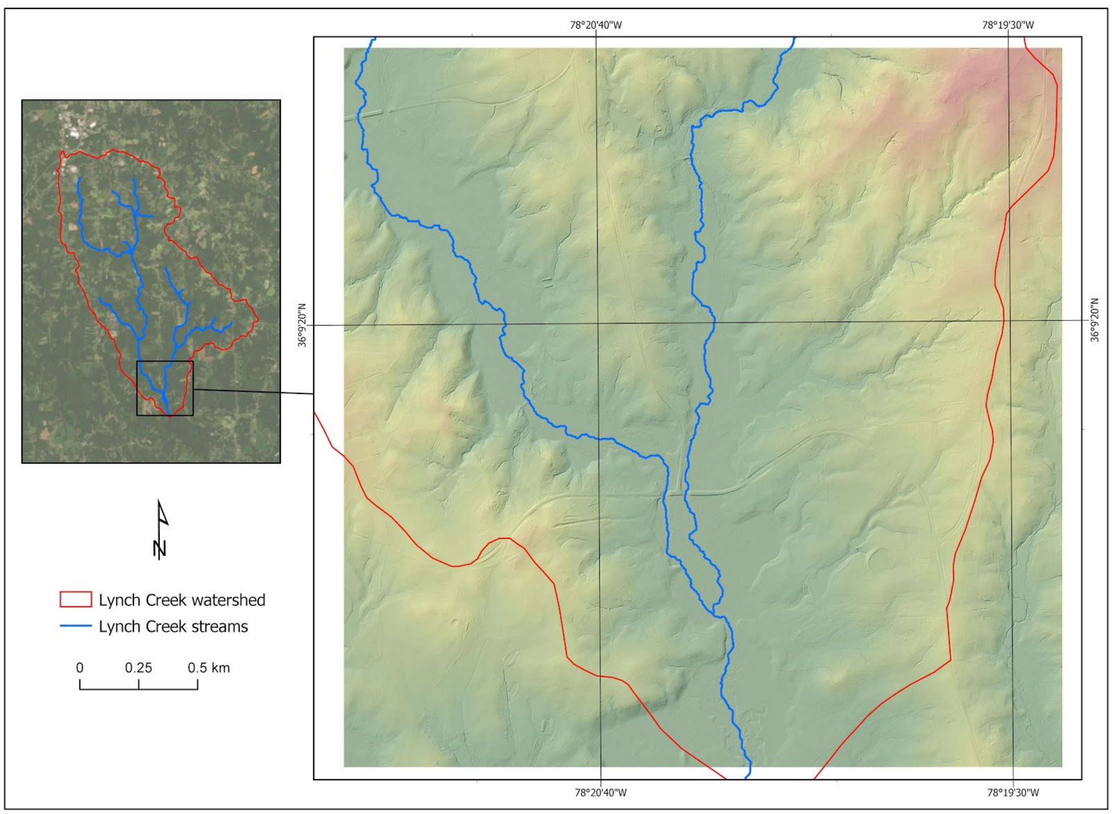

Figure 1: Map of the Lynch Creek watershed study region in Franklin County, NC.

Methods

Data Preparation and Processing

Lidar

Lidar data collected in March 2015 with a points spacing of 8 points per meter was obtained from North Carolina Flood Mapping Program Spatial Data Download portal (https://sdd.nc.gov/) for the 10.4 km² study area. GRASS GIS was used to compute the following rasters from the lidar point data. A bare earth digital elevation model (DEM) with 2 m resolution was generated by interpolating the lidar points classified as ground. The DEM was used to compute a topographic wetness index raster of the area. A digital surface model (DSM) was created by binning all points at 2 m resolution using the maximum point height as the cell value. A slope map was computed using the DSM. Gaps in the DSM (cells with zero binned points) were filled using interpolation. Rasters representing lidar point density and intensity were also created by binning all points at 2 m resolution using the number of points and average point intensity, respectively, as the cell values. Gaps were filled using interpolation.

Stream Network

A stream network vector layer was computed from a 3 m DEM obtained from the USGS National Elevation Dataset using GRASS GIS. First, a flow accumulation raster was calculated using a least cost path algorithm, which effectively routes water through any depressions in the DEM. Streams lines were then extracted from the flow accumulation raster using a minimum accumulation threshold of 20,000.

Sentinel NDWI

Sentinel-2A Multispectral Instrument data collected during December 2015 at 10 m resolution was obtained for the Lynch Creek watershed. The Normalized Difference Water Index (NDWI), used to map water bodies, was calculated using Bands 3 (Green) and 8 (Near-Infrared) according to the following formula:

$$Index = \frac{Band\ 3 - Band\ 8}{Band\ 3 + Band\ 8}$$

The resulting raster was used to identify areas with high NDWI values which correspond to high concentrations of water. An Index threshold of 0.5 was used to extract areas with significant amounts of water. These areas may correlate to portions of Lynch Creek where beaver dams are present.

NAIP Orthoimagery

Visible band (RGB) orthoimagery from the National Agriculture Imagery Program (NAIP) collected in summer 2016 at 1 m resolution was obtained for the 10.4 km² study area. A grey level raster was calculated by averaging the visible band reflectance values and recoding the results into 16 quantiles. The R package glcm was used to compute the grey level co-occurrence matrix for the recoded raster and Haralick texture features, including angular second moment and contrast.

Identification of dams for model training and validation

The data given to train the identification model were identified using Google Earth and transferred as a KMZ file containing line data. Probable beaver dams were located in Google Earth by following the stream channel and searching for either: 1) areas where the stream significantly widens and forms a small pond and then thins further downstream, referred to as “ponding”, or 2) locations where physical dam-like obstructions could be identified across the width of the stream. In general, the identification of ponded areas was used as evidence to support the probability that the structure identified in Google Earth was a beaver dam.

The most reliable visual evidence of beaver dams (Figure 2) was found to concentrate mostly near the pour point of the Lynch Creek watershed, where it branches off from the larger Tar River.

Figure 2: Regions A and B along Lynch Creek, reveal the possible beaver dams and subsequent ponding. The dams are identified by the red arrow. In both figures, the probable dam is seen as a structure across the river resulting in the formation of a pond behind the structure.

Assessing dam location likelihood with Maxent

In an effort to identify areas within the Lynch Creek watershed with high probability of beaver dam occurrence, Maxent, a species distribution machine learning algorithm was used. Maxent takes as inputs coordinates of the occurrence of the event of interest, in this case a beaver dam existence, and environmental layers. These environmental layers are raster files of the same extent and resolution, that each represent a variable that influences the location of beaver dams. In order to generate coordinates for beaver dam occurrence, a KML file was created in Google Maps that contained line features coincident with dams. Dams were identified visually with the help of Google Maps imagery. For use in Maxent, these line features were converted into point features using a Python script for GRASS GIS (see Appendix A), and the coordinates were written to a CSV file formatted properly for Maxent. Environmental layers used are listed below:

- DSM Elevation (2m)
- DSM Slope (2m)
- Point Density (2m)
- Distance to Ponding Feature (2m)
- TWI (2m)
- NDWI (2m)
- Return Intensity (2m)

Environmental layers were restricted to a 150m buffer around the main channel of the Lynch Creek stream network to reduce the possibility of prediction of beaver dams in non-wetland areas. These metrics were derived for LiDAR data for the lower Lynch Creek watershed. Maxent was run using a 0 percent test percentage (all points were used for training) and a regularization parameter of 0.25 (Phillips, 2005). This process created an output raster whose cell values are the probability of a beaver dam existing in that cell. The 'Distance to Ponding Feature' raster was created by first extracting cells from the point density raster with a value of 0, as these cells are likely to represent water. The resulting raster was converted to vector and filtered for features with an area of >= 40sq ft. These were considering ponding features. Then a raster was created that represented distance from each cell to the closest ponding feature boundary. Areas with high probability of beaver dam occurrence were most successfully predicted by the 'Distance to Ponding Feature' raster, so it was decided that using the probability raster as a proxy for distance to ponding/ wetlands was not a gross oversimplification.

Dam locations along stream network

Once the Maxent probability surface was created, cells with a probability of >= 50% were isolated using raster algebra, and then vectorized. Again these features were filtered for an area of >= 100sq ft. to eliminate small, discontiguous polygons. Areas were then spatially generalized by creating convex hulls for each grouping of polygons, using 25 ft. as the distance threshold below which polygons were aggregated. These areas were treated as areas close to ponding features, although the fact that beaver dams are most likely to occur downstream of these ponding features needed to be accounted for. To accomplish this, convex hull polygons were then intersected with a stream network line vector created from a 3m USGS DEM. What this yielded was areas (polygons) with a high probability (50-100%) of beaver dam occurrence as predicted by Maxent. These high probability zones were then intersected with the steam vector and for each discrete polygon, the point of intersection with the lowest elevation was selected as a predicted dam location. This was done to ensure that we were selecting points that were downstream of a ponding feature. These predicted dam location can be seen in Figure x.

Wetland classification using Random Forest

The R package randomForest was used to classify pixels into four categories - wetland, forest, grass, and bare earth. One-hundred and seventy-one polygons were digitized and used to train the classification algorithm (108 wetland, 49 forest, 5 grass, and 9 bare earth). Variables used for the classification include DSM elevation, DSM slope, topographic wetness index, red reflectance, green reflectance, blue reflectance, texture features - angular second moment and contrast, normalized difference wetness index, lidar point density, and lidar intensity.

The dam locations identified in the Maxent likelihood and stream network intersection analysis were overlaid with the map of classified wetlands. Since the dams were concentrated in areas classified as wetlands, these wetlands were assumed to be related to beaver activity. Since a separate class for man made ponds was not used, a polygon delineating the area of interest and excluding man made ponds was drawn and used as a mask for the wetland area computation.The total area classified as wetlands was computed by multiplying the number of wetland pixels in the area of interest by the cell size (4 m^2).

Results

Dam location likelihood

Figure 3 shows the dam probability map created with Maxent. The highest dam probability areas occurred in ring patterns around ponds along the stream network. Several dams identified in Google Earth that were set aside for model validation were correctly identified as having high dam probability.

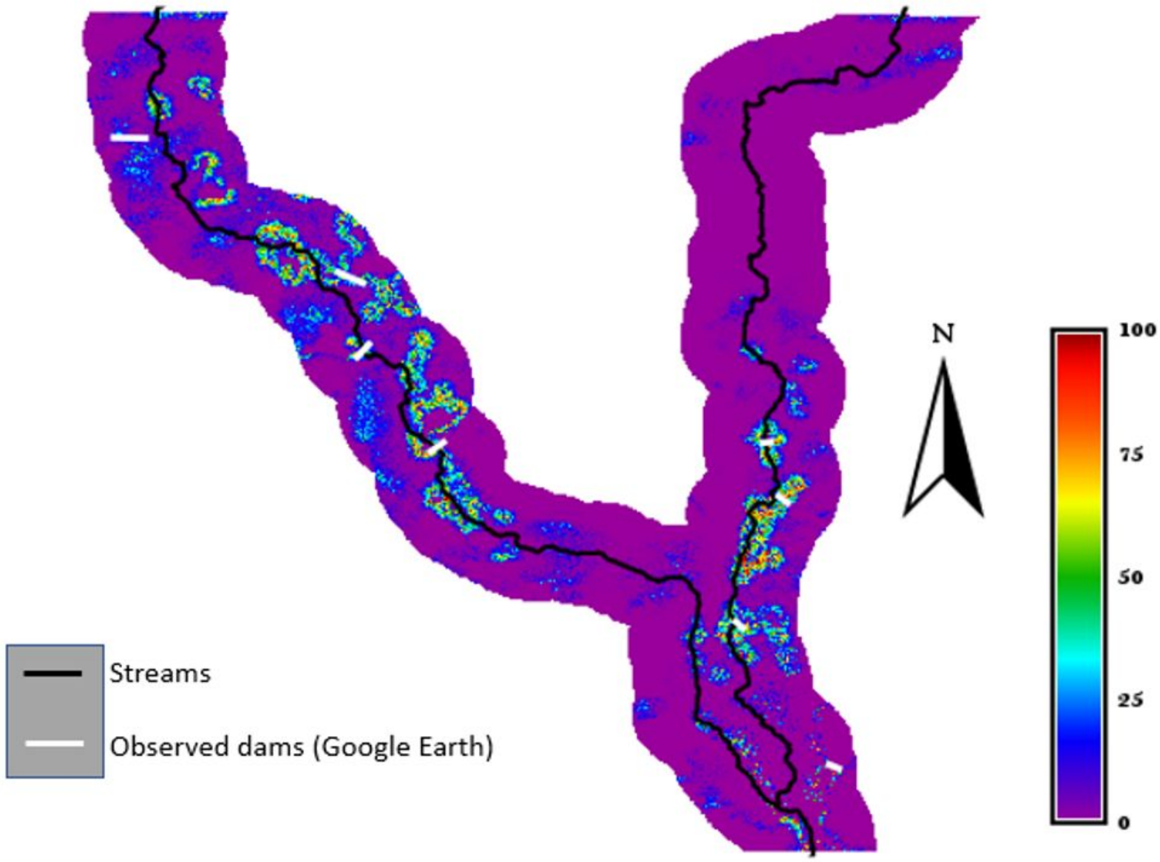

Figure 3: The image above shows the dam probability map output from Maxent, with the probability of beaver dam existence legend shown on the right. Lynch Creek streams and observed dams (used for training Maxent) are included for context.

Dam locations along streams

Figure 4 shows the minimum bounding geometry (convex hulls) of areas with 50% or greater probability of beaver dams in grey and the stream intersection point with the lowest elevation value in red. A total of twelve dam locations were identified. Several smaller high probability polygons that did not intersect with the stream network, and therefore did not result in a dam siting, are also visible.

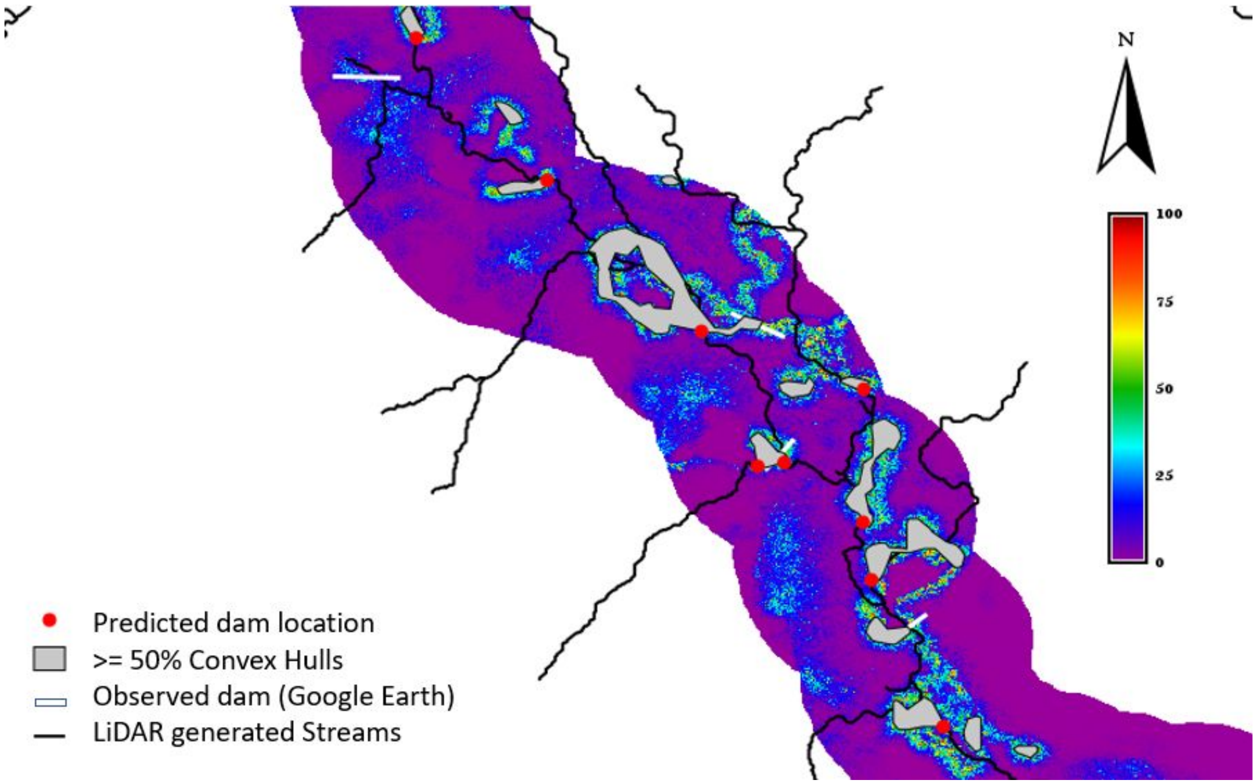

Figure 4: The image above shows a section of the Lower Lynch Creek watershed, with streams created from LiDAR data, convex hull polygons, and predicted dam location from the convex hull and stream network intersection described in the methods section. These are overlaid on the Maxent output raster to provide context.

Wetland classification

The results of the random forest classification and twelve probable dam locations along the stream network are shown in Figure 5. The overall classification accuracy was 99.76% with a Cohen's Kappa score of 0.9963. The user and producer accuracy for each class was greater than 99%. Most of the error that occurred was confusion between the wetland and forest classes. Relative predictor variable importance listed from most important to least important is as follows: red reflectance, DSM slope, DSM elevation, NDWI, lidar point intensity, texture contrast, lidar point density, topographic wetness index, texture angular second moment, green reflectance, and blue reflectance. The area of interest used for computing total wetland area is shown outlined in gray in Figure 5. The total wetland area was 0.37 km^2, which is approximately 3 ha per predicted beaver dam.

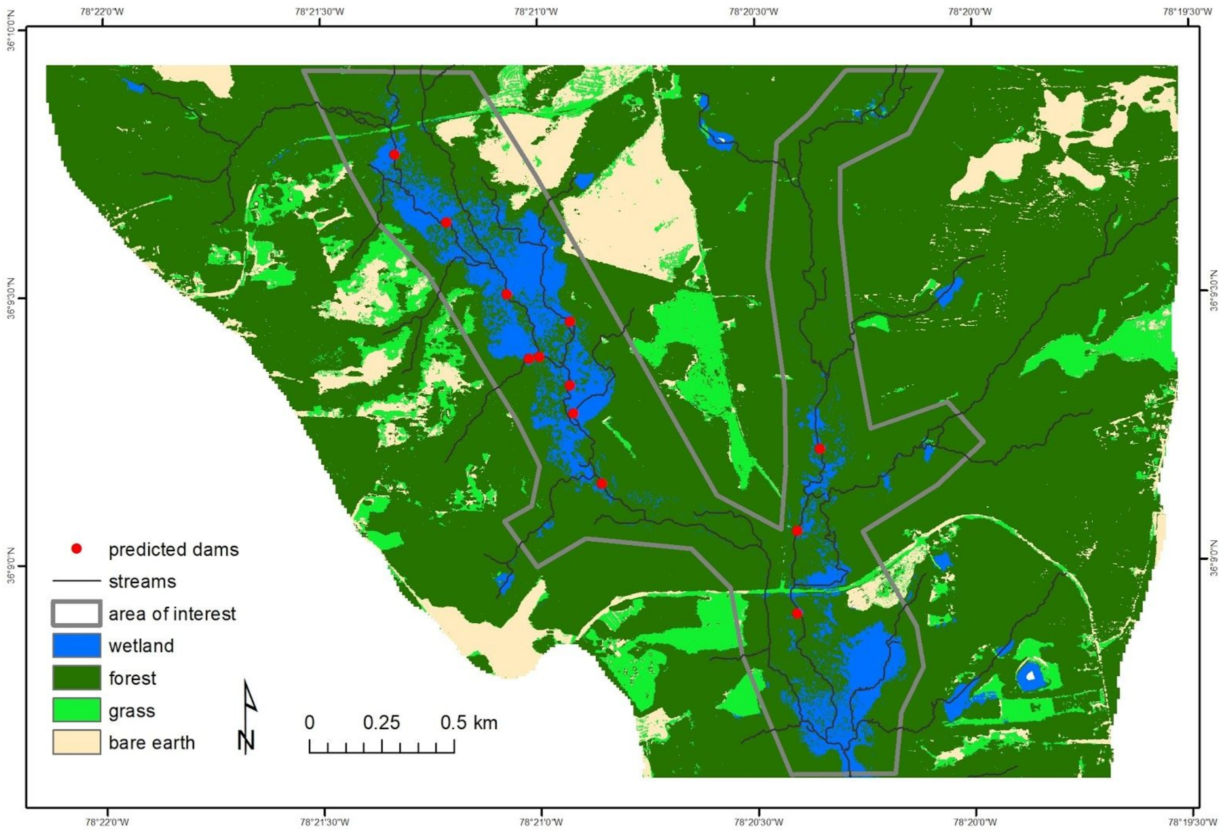

Figure 5: Wetland classification map showing the locations of predicted dams along the stream network and the area of interest used for the total wetland area calculation.

Discussion

The method used for identifying dam locations heavily depended on the probability map output from Maxent and the stream network used. For this study, we used streams derived from a 3 m USGS DEM. Since the predictor variables used in Maxent were derived from lidar data, better results may be achieved by using a stream network derived from lidar data as well. For example, in some cases, the high probability polygons created from the Maxent output did not intersect a stream, although they were located very near a stream. In those cases, a dam was not identified because the method we used only sited dams where high probability polygons intersected the stream network. In future iterations, it may be important to use streams derived from the same elevation data used to generate the Maxent dam probability map.

Leaf-on imagery (NAIP) was used to create training polygons for training the random forest classification. Much of the wetland areas appeared heavily covered in vegetation in the imagery, so it was important to including training data for these areas to ensure they were included and not misclassified as forest. Without ground truthing, however, it is possible that some areas were incorrectly assumed to be wetlands. Using leaf-off imagery for future iterations may improve certainty in distinguishing wetlands when selecting training areas. Also, since red reflectance was the most important predictor variable, using leaf-off red band imagery may improve the overall classification.

This approach can easily be applied to the full Lynch Creek watershed, as well as watersheds in the coastal and mountain physiographic regions of North Carolina. It is important to understand the differences in the amount of wetland area created by beaver dams in regions with more and less topographic relief. Also, to accurately quantify wetlands created by beaver activity, an approach for associating wetland areas with downstream dams is needed. For this research, all wetlands in the study area were assumed to be related to beaver dams because of the high density of dams predicted in the area.

Conclusions

This research presents an approach for estimating the current number of dams and wetlands created by beavers within a North Carolina watershed. Environmental predictors can effectively be used with machine learning techniques to approximate dam locations and quantify wetland area created by those dams. For the Piedmont region of North Carolina, beaver dams result in approximately 4 ha. of upstream wetlands. This is a useful metric for quantifying the ecosystem services and economic value provided by beaver activity. This analysis should be repeated for other physiographic regions to determine the area of wetlands created by beavers dams in areas with more or less topographic relief.

Acknowledgements

Karl Wegmann for providing the external harddrive used to obtain and store the lidar data used for this research.

References

Butler, D. R., & Malanson, G. P. (2005). The geomorphic influences of beaver dams and failures of beaver dams. *Geomorphology*, *71*(1-2), 48-60.

DeVries, P., Fetherston, K. L., Vitale, A., & Madsen, S. (2012). Emulating riverine landscape controls of beaver in stream restoration. *Fisheries*, *37*(6), 246-255.

Jakes, A. F., Snodgrass, J. W., & Burger, J. (2007). Castor canadensis (Beaver) impoundment associated with geomorphology of southeastern streams. *Southeastern Naturalist*, *6*(2), 271-282.

Macfarlane, W. W., Wheaton, J. M., Bouwes, N., Jensen, M. L., Gilbert, J. T., Hough-Snee, N., & Shivik, J. A. (2017). Modeling the capacity of riverscapes to support beaver dams. *Geomorphology*, *277*, 72-99.

Naiman, R. J., Johnston, C. A., & Kelley, J. C. (1988). Alteration of North American streams by beaver. *BioScience*, *38*(11), 753-762.

Pollock, M. M., Beechie, T. J., Wheaton, J. M., Jordan, C. E., Bouwes, N., Weber, N., & Volk, C. (2014). Using beaver dams to restore incised stream ecosystems. *Bioscience*, *64*(4), 279-290.

Rosell, F., Bozser, O., Collen, P., & Parker, H. (2005). Ecological impact of beavers Castor fiber and Castor canadensis and their ability to modify ecosystems. *Mammal review*, *35*(3-4), 248-276.

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2005, July 14). Maximum entropy modeling of species geographic distributions. Retrieved from https://www.sciencedirect.com/science/article/pii/S030438000500267X.

Appendix

A) The following is a link to the public Github repository with the Python script used to convert line feature to points. The script is meant to be called from the GRASS GIS console. The repository also includes the R script used for the random forest image classification.

https://github.com/bhbaines/beaverDams