CSU Research Competition

Author: Jessie Doyle

PI: Dr. John Olson

Title: Predicting Fish Distributions in Remote Areas using Environmental DNA, Satellites, and Models

Abstract

The lack of location data of threatened fish species can make the conservation of biodiversity difficult for land managers. This is especially true in remote places such as the North Slope of Alaska. Species Distribution Models (SDMs) are one way to predict fish distributions. To apply SDMs across landscapes we need environmental data characterizing the environmental spatial and temporal variation that could be related to species locations. As data cannot be effectively collected on the ground in the North Slope, remote sensing offers a way of characterizing the environment for these models. We characterized watershed environments using Earth Observations from a variety of platforms (i.e., measurements collected using aerial Synthetic Aperture Radar, MODIS, and LandSat satellites). Because river environments are controlled by up-stream conditions, we adapted a process of accumulating watershed environmental data for the contiguous US known as StreamCat (Hill et al. 2016) to the North Slope. The remote sensing data and the StreamCat process allowed us to measure spatial and temporal environmental variability for every stream segment across the entire North Slope. We saw several interesting patterns of inter-year & spatial trends. This includes noting that land surface temperature was warmer at lower latitudes and higher elevation than at higher latitudes. This approach helps us understand the arctic landscape and minimize the effects of oil and gas development on biodiversity across the North Slope.

Introduction

Biodiversity has always had an important role in shaping natural ecological processes and providing ecosystem services (Ronald Sandler, 2012). By taking biodiversity into account, land managers can make more comprehensive and long-term decisions (Alho, 2008).

Given time and space dependent variables and their complex interactions, it can be difficult to predict and calculate biodiversity over large landscapes due to the time, effort, and research necessary. Biodiversity can be measured by species richness of a region (species count) and can be used to focus management efforts. The richness of all species within an ecosystem are important, however if resources are limited, focusing on a few keystone species in an ecosystem can optimize biodiversity predictions.

Biodiversity can even more difficult to measure in remote areas, but it is valuable information to have where data is already limited. The remote area of the North Slope of Alaska is important to identify biodiversity hotspots as recent legislations has opened large areas of the North Slope (the National Petroleum Reserve - Alaska and the Arctic Wildlife National Refuge) for natural gas and oil extraction. There are several fish species in the North Slope of Alaska that hold cultural and commercial importance (i.e. Cultural: Least Cisco & Commercial: Salmon) and calculating these fish' richness can provide insight of the biodiversity on a larger scale.

Species distribution models are an established method for predicting occurrences of endangered or threatened species of concern. A system of mapping and predicting species with species distribution models has already successfully predicted occurrences of invasive algae in the Northern Rocky Mountains. In this area and throughout the continental United States there are many species datasets. Land managers have used the results of these models to currently make decisions regarding human access and invasive species management.

Commented [JO1]: I'd delete – doesn't really help argument

We employed methodology similar to that used in Northern Rocky Mountains to a remote area, the North Slope of Alaska, using remotely sensed data to locate areas with high fish biodiversity. To help land manager efforts, 1) we will first calculate distribution of individual fish species using models 2) then we will combine these distributions to calculate species richness for stream segments throughout the North Slope.

Methods

The North Slope of Alaska includes the National Petroleum Reserve – Alaska (NPRA) and the Arctic National Wildlife Refuge (ANWR). The North Slope is difficult to obtain data for as time, resources and research for these areas have been limited. For example, a limiting factor to collecting data is there are no roads in the NPRA, meaning field crews must be flown in by helicopter which is very expensive. In addition, the sheer size of the NPRA is about 36,800 sq. miles which is approximately 4 times the size of the state of Maryland makes it hard to obtain for all rivers and landscapes within the area. However, remote sensing technology allows us to characterize environments at large spatial and temporal scales, thus enabling the development of species distribution models that can predict fish distributions across the landscape.

To calculate the distribution of species throughout the North Slope of Alaska we needed data on fish locations to train our models. Environmental Data (e-DNA) and traditional fish count survey data was used for this training dataset as our response variable. Remote sensing data such as climate, vegetation, water and landscape (i.e. climate: fire, vegetation: gross primary production, water: national hydrology, landscape: ecoregions) were taken from a variety of satellites (i.e. NASA Modis, Syntenic Aperture Radar, and Landsat) and summarized at watershed scales using a python process known as StreamCat (Hill et al., 2016). StreamCat is a dataset provided by the U.S. Environmental Protection Agency for the continental United States, however, we used the python process that makes those datasets to produce similar data for Alaska. We delineated local catchment areas for each stream segments in our river dataset,calculated zonal statistics for each independent variable, and used the StreamCat process to calculate watershed averages for every stream segment across the North Slope (Appendix A: Figure 1).

We matched these watershed statistics with fish observation data from the same location to predict fish occurrences in each stream segment for the North Slope (Holder et al., manuscript in preparation). Predictions were made using two different machine learning algorithms, MaxEnt & Random Forest. These results were combined into an ensemble prediction which was used to estimate the relative number of species we may see in each stream segment. These predictions were then added together to calculate a relative richness of all fish species in each stream segment.

Results

Local catchment scale statistics were successfully combined into watershed scale statistics using the Stream Cat process resulted in some interesting trends on the North Slope (Appendix A: Figure 2). One problem that was seen with the remote sensing data before calculating zonal statistics was large data gaps (Appendix A: Figure 3). These large data gaps were noted and corrected for before local catchment statistics were summarized in StreamCat.

Models were successfully created and validated for accuracy (Holder et al., In preparation) for all the species of concern. Maps were made to represent the modeling results and presented to the Bureau of Land Management for their database (Appendix A: Figure 4). In addition, relative species richness was calculated for each stream segment on the North Slope of Alaska (Appendix A: Figure 5).

Conclusion

Summarizing landscape variables like remotely sensed tempertures at watershed scales provides insight into how landscape variation influences river environments from which fish distributions can be predicted. In some cases, such as land surface temperature, we see that the scale is an important factor to consider when predicting species as more specific temperatures values were represented in smaller, catchment scales (Appendix A: Figure 2). Also, we saw that fires may have a more localized impact compared to the watershed scale.

Remotely sensed data is invaluable to characterizing the environment in places like the North Slope, but given the large data gaps even in some of the best data available may require some refinement before use (Appendix A: Figure 3). .

We observed that species richness was greatest in headwater rivers in the North Slope (Appendix A: Figure 5), similar to other watersheds throughout the United States. This is important information to have as recent legislation has allowed for increased natural gas and oil extraction from the North Slope and knowing that these headwater rivers are where there is more fish biodiversity may inform land managers decisions on locations to avoid drilling.

We can conclude that this is a successful way of mapping species of concern especially in remote areas. Remotely sensed data allowed us to accuratly predicted fish biodiversity across entire landscapes. We have helped the Bureau of land management in their effects to calculate fish distribution models and have identified areas for biodiversity conservation.

References

Alho, C. (2008). The value of biodiversity. Retrieved February 25, 2019, from http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1519-69842008000500018

Hill, R. A., Weber, M. H., Leibowitz, S. G., Olsen, A. R., & Thornbrugh, D. J. (2016). The Stream-Catchment (StreamCat) Dataset: A Database of Watershed Metrics for the Conterminous United States. *JAWRA Journal of the American Water Resources Association,52*(1), 120-128. doi:10.1111/1752-1688.12372

Holder, A. (in press). Manuscript submitted for publication.

Sandler, R. (2012). Intrinsic Value, Ecology, and Conservation. Retrieved February 25, 2019, from https://www.nature.com/scitable/knowledge/library/intrinsic-value-ecology-and-conservation-25815400

Appendix A.

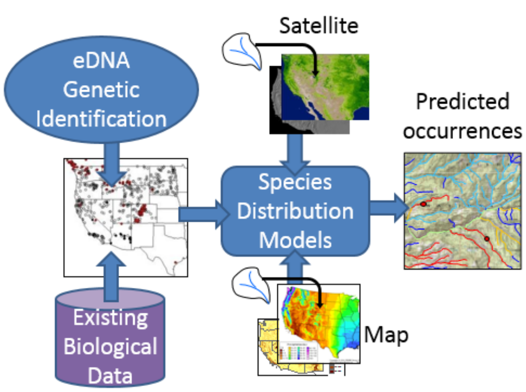

Figure 1. Flowchart of system of mapping species of concern.

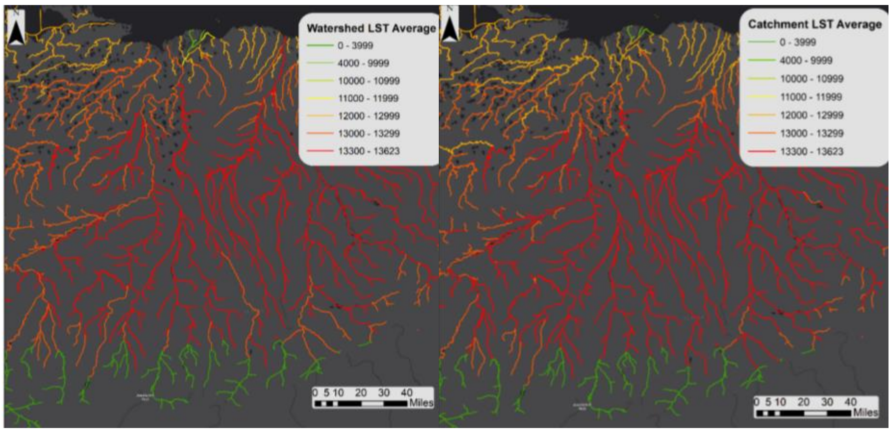

Figure 2. Land Surface Temperature at the Watershed and Catchment zonal statistics level. Results indicate that some variables may have more of an impact at different scales.

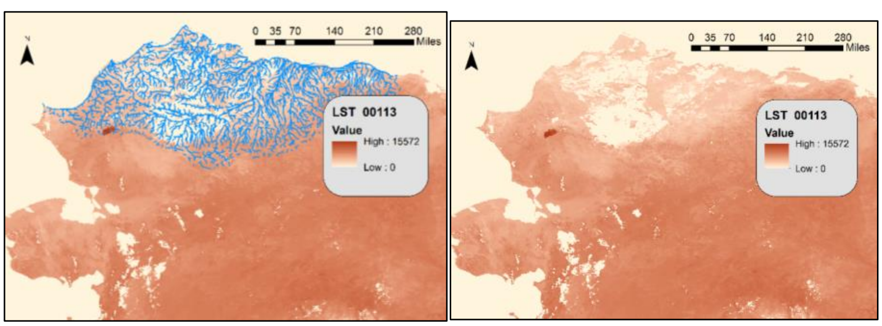

Figure 3. One of the problems that was seen with remote sensing data was large data gaps such as the for-Land Surface Temperature (LST) in 2000 on day 113. These data needed to be corrected for before zonal statistics were run.

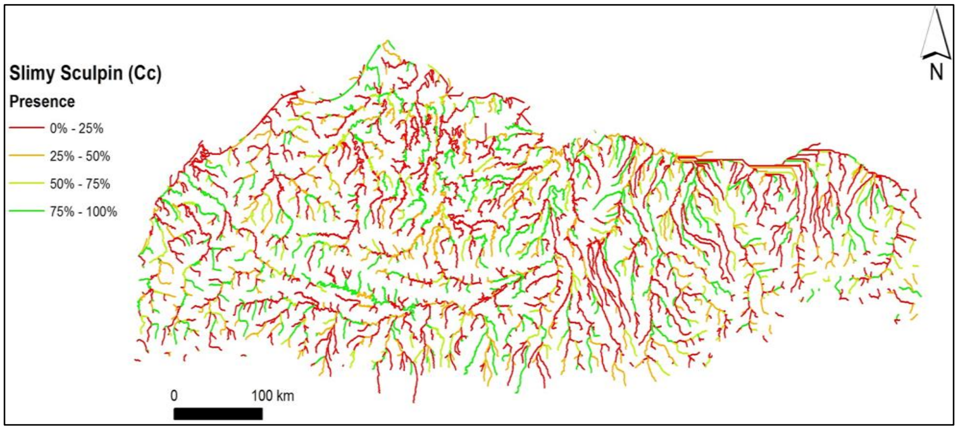

Figure 4. Example of species distribution model results for a species of concern – Slimy Sculpin. Likelihood of occurrence is represented with red being least likely river segments to dark green which is most likely.

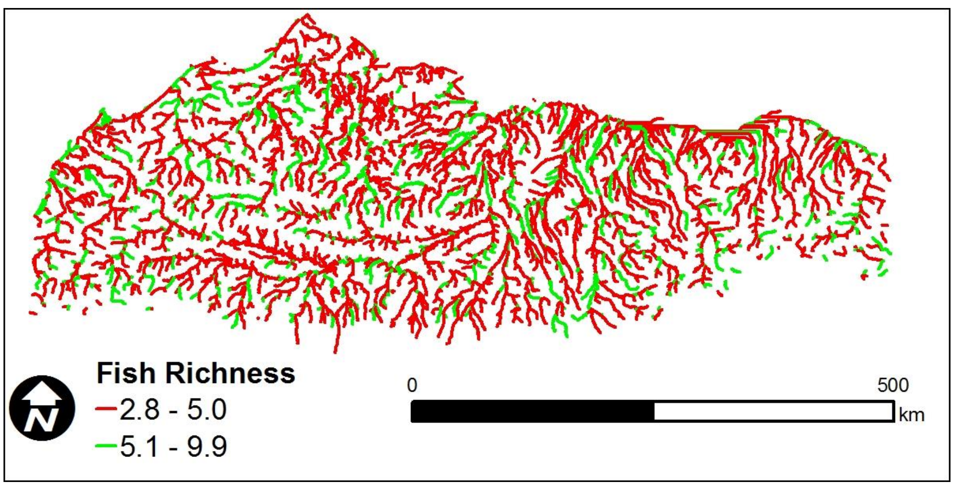

Figure 5. Richness indicated for each stream segment. Red represents lower species richness and green represents higher species richness.