Montana State University
Department of Mathematical Sciences
Writing Project

TITLE

Author:
Meaghan Winder

Supervisor:
Dr. Andrew Hoegh

Spring 2020

A writing project submitted in partial fulfillment
of the requirements for the degree

Master's of Science in Statistics

APPROVAL

of a writing project submitted by

Meaghan Winder

This writing project has been read by the writing project advisor and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the Statistics Faculty.

_____   _____
Date   Andrew Hoegh
                                                       Writing Project Advisor

_____   _____
Date   Mark C. Greenwood
                                                       Writing Projects Coordinator

Contents

Abstract

abstract text here

free write... this is very messy, basically just word vomit

This is all information extracted from WILD 502 course material or things I learned from the class

Occupancy is the presence of a particular species on a given site, this may not be the first choice of state variables to ecologists but occupancy studies are useful when there is a large spatial scale or the study is conducted over many years, when abundance or vital rates are hard to measure. Occupancy studies are also useful over capture-recapture methods when individuals cannot be marked or uniquely identified. However, sometimes patterns of species occurrence are of interest, this happens when researchers are interested in the range of a species or the spread of invasion.

The sampling units for occupancy studies are called 'sites'. We can learn about detection probabilities when multiple site visits are used. Also, when using occupancy models we need to account for imperfect detection because it is possible that the researchers could miss the species even if it is present at the site. ψ represents the occupancy probability, p_i represents the probability of detecting the species on survey i given that the species occupies the site, and $p^* = 1 - \prod_{i=1}^{t}(1 - p_i)$ is the probability of detecting the species at least one time given the species occupies the site.

The assumptions are:

- The occupancy state of sites is constant during a single season.

2

- The occupancy probability is constant across sites, or is modeled appropriately using site-level covariates.

- The probability of detection given occupancy status is constant across sites, or modeled appropriately using site-level covariates.

- The species is not misidentified, no false positives.

As suggested above, site-level covariates can be used to model the occupancy probabilities and the detection probabilities.

In WILD 502 when talking about multi-season occupancy models, we talked about extirpation and colonization rates, but I think that these could be modeled with a latent variable(s)? I don't think they are of particular interest here.

In this case, if we consider site to be the lake then we have replication at the site level, but if each site is the sample location within the lake then there is replication for some (very few) on different dates.

This is what I used for my project proposal... Mark said this could make it's way into our abstract

Zebra mussels are a highly invasive species that have several negative impacts, both ecologically and economically, as they suffocate native mussels and cost millions of dollars to remove from man-made structures such as power plants (USGS). Zebra mussels have been established in the United States since the 1980s, and since then "they have spread rapidly throughout the Great Lakes region and into the large rivers of the eastern Mississippi

drainage. They have also been found in Texas, Colorado, Utah, Nevada, and California" (USGS). As the threat of zebra mussels spreading into the northwestern United States grows, one concern is that the detection method used to find zebra mussels is not highly effective.

I have also been working on fitting models to simulated occupancy data (see code at `https://github.com/meaghanwinder/Writing-Project/blob/master/Writing%20Project%20Code.Rmd`)

- constant $\psi = 0.6$ and constant $p = 0.6$ with constant $J = 8$ samples per $M = 10$ sites – DONE

- constant $\psi = 0.6$ and constant $p = 0.6$ with j_i samples per $M = 10$ sites – DONE

- ψ and p both depend on a single site level covariate, x, with j_i samples per $M = 10$ sites

$y_{i,j} = \{0, 1\}$ is the binary response for the jth sample within site i

If the site is occupied:

- $y_{i,j} \sim Bernoulli(p)$

- $y_i \sim Binomial(J, p)$ where y_i is the total number of detections from the J samples in the ith site

If the site is unoccupied:

- $y_{i,j} = 0$ with probability 1, since we assume there are no false detections

Things to think about for the simulation or questions I have:

- think about how changing the detection probabilities and occupancy probabilities impact the results:

    - High occupancy, high detection

    - High occupancy, low detection

    - Low occupancy, high detection

    - Low occupancy, low detection

- How do we/can we include sample level covariates to account for sampling effort (number of tows, if available)?

- If we are getting more data:

    - How do we account for multiple sampling seasons (years)?

    - How do the assumptions of occupancy models change when we have several sampling years versus only 1

Additional questions, not directly related to the simulation:

- How do we account for this multilevel (for lack of a better word) testing process? For example, sometimes (?) when they fail to find them in the microscope, then they test them using polymerase chain reaction (PCR), and sometimes gene sequencing (? – looks like it was only used on one observation in the sample data), or scanning electron microscopy (SEM).

- One assumption of occupancy models is that there is a non-zero probability of misidentification. I read an article (Denise Holser) that suggests that there might be an issue with that assumption in this situation because there are similar looking organisms that may be present in the waters; she then suggest that the Bureau of Reclamation has attempted to mediate this issue with improved microscopic methods and improved PCR methods. I will look into this some more.

1 Introduction

In early 2020, the City of Austin, Texas approved the spending of four million dollars over the next five years in an attempt to remove zebra mussels from the city's source of drinking water with a liquid copper sulfate pentahydrate released into the water intake pipes (Bontke, 2020). This example is one of many pursuits to remove dreissenid mussels[1] from water bodies across the United States, and the four million dollars is only a small fraction of what is spent annually on control and mitigation efforts (Sepulveda et al., 2019).

Zebra mussels are native to the Caspian and Black Seas, but have become widespread in both Europe and the United States; zebra mussels were discovered in the Great Lakes in the late 1980s and have since spread rapidly across the United States (U.S. National Park Service, 2017). The United States National Park Service stated that "[o]nce a population of zebra mus-

---

[1]Zebra mussels (*Dreissena polymorpha*) and quagga mussels (*Dreissena rostriformis bugensis*) collectively.

sels has become established in a water body, there is very little to be done to remove them. Prevention, therefore, is the best way to keep a water body clean of zebra mussels" (2017); hence, early detection of invasive species, such as dreissenid mussels, has become a priority, so that organizations can plan, budget, and install necessary technologies before colonization has occured (Holser, 2017).

Zebra mussels live between two and five years; they start as microscopic veligers but mature to thumbnail sized adults (U.S. National Park Service, 2017). Zebra mussels begin reproduction at two years of age, after which, females can release up to one million eggs per year (U.S. National Park Service, 2017). Dreissenid mussel veligers free-swim in the water; often, they travel to uninfested waters on boats or through other aquatic recreational activies, however, sometimes they are moved by nature and travel downstream to uninfested waters (Bureau of Reclamation, 2019, Holser, 2017, U.S. National Park Service, 2017). Adult dreissenids attach and colonize hard surfaces in the water, which results in economic issues (Holser, 2011). This process of accumulation of adult zebra mussels on rocks, native mussels, docks, boats, or other hard surfaces is referred to as "biofouling," and objects that are in the water for long periods of time become difficult and costly to clean (U.S. National Park Service, 2017). Water supply and delivery facilities, water recreation sites, and other water dependent economies in dreissenid mussel infested waters become much more expensive to maintain and operate (Bureau of Reclamation, 2019). Dreissenid mussel infestations result not only

in economic impacts, but in environmental ones as well. Dreissenid mussels are filter feeders and siphon plankton from the water, which can lead to changes the water body ecosystem by increasing water clarity (U.S. National Park Service, 2017); a single adult dreissenid can filter about a liter of water per day, which reduces the availability of algae for native mussels and bottom feeding fish (Bureau of Reclamation, 2019). Additionally, "biofouling" can prevent native mussels from moving, feeding, reproducing, or regulating water (U.S. National Park Service, 2017). Several actions, such as the 2017 initiative, *Safeguarding the West from Invasive Species*, by the Department of the Interior, have been taken to protect water bodies in the western United States from the economic and ecological threats posed by the invasive dreissenid mussels (Bureau of Reclamation, 2019). Early detection of dreissenid mussel species can reduce the economic and ecological reprocussions of dreissenid mussel infestations, but there are issues with the available early detection methods.

The established standard for early detection of dreissenids in the western United States is plankton tow sampling for mussel veligers (Sepulveda et al., 2019). Using a fine mesh net, water and debris are collected at multiple sampling sites within each water body; the debris from each net collected at the same sampling site on the same day is aggregated and examined, using cross-polarized light microscopy, for the free-swimming veligers. Following the microscopic examination, positive species identification is confirmed with a polymerase chain reaction (PCR) test (Sepulveda et al., 2019). This early

detection method requires a breeding population, so is limited to the weeks immediately following a spawning event (Nichols, 1996); spawning begins at water temperatures above 10° C for quagga mussels and above 12° C for zebra mussels (McMahon, 1996, Mills et al., 1996). This suggests that veliger availability in northern latitude water bodies is typically limited to warmer months (Sepulveda et al., 2019).

An alternative method for detection of rare, endangered, or invasive aquatic organisms, one that is growing in popularity, are environmental DNA (eDNA) surveys (Schmelzle and Kinziger, 2016). As opposed to plankton tow sampling, eDNA is a molecular approach that can detect DNA diffused from target organisms into a water body (Sepulveda et al., 2019). It is suggested that the use of eDNA surveys may widen the seasonal sampling window over plankton tow methods, since eDNA does not rely on a breeding population (Sepulveda et al., 2019). This method is more time and cost effective than traditional sampling methods for species of low abundance (Rees et al., 2014); "eDNA is a highly sensitive tool for dreissenid mussel surveillance in newly invaded waters and in seasons when veligers are likely to be rare" (Sepulveda et al., 2019). However, a positive eDNA result does not necessarily mean that the target species is present or alive at the site; positive eDNA results can be obtained from "a failed introduction, from external sources, or from field contamination, rather than fresh DNA from mussel colonization" (Sepulveda et al., 2019). One criticism of the detection of dreissenid mussels using eDNA is that there is a possibility of obtaining a false-positive result;

since control efforts for invasive species are costly, there is hesitation in using eDNA surveys as decision-making tools for invasive species (Sepulveda et al., 2019).

When these two methods result in conflicting answers, decision making can be even more complicated, since a positive eDNA result only suggests that the DNA of the target species is present, regardless of whether the target species is alive or even present at all, but when veligers are detected, positive eDNA results indicate a potential colonization, which is useful to managers (Holser, 2017).

I wrote all this, and then realized I should probably include at least some stuff about occupancy models in order to discuss the research questions. I will work on incorporating this more when we nail down the data sets for sure.

Research question(s): How does the detection probability of zebra mussels using microscopy compare to the detection probability of zebra mussels using eDNA? What is the false negative rate for both methods? Given that, how many samples should we collect at a given water body?

2   Data

This could maybe be included in the introduction section... unless I do EDA here, then probably leave this as its own section

3   Modeling Background

This whole section might be able to be merged with the methods section

3.1   Occupancy Models

3.2   Bayesian Modeling Background

4   Methods

Package options for Multi-season single-species occupancy models:

- nimble.dynamic.occ

- STAN

- JAGS

- wiquid package??

- Frequentist Options:

  – unmarked

  – Program MARK

- write my own package?

5   Analysis

Could do EDA here

6 Conclusion

6.1 Further Investigations

7 References

Bontke, J. (2020). City spends $4 million on liquid compound to stop spread of zebra mussels. *CBS—Austin*.

Bureau of Reclamation (2019). Invasive mussels. `https://www.usbr.gov/mussels/index.html`.

Holser, D. M. (2011). Early detection of dreissenid species: Zebra/Quagga mussels in water systems. *Aquatic Invasions*, 8(2):217–222.

Holser, D. M. (2017). Where is the body? Dreissenid mussels, raw water testing, and the real value of environmental DNA. *Management of Biological Invasions*, 8(3):335–341.

McMahon, R. F. (1996). The physiological ecology of the zebra mussel, *Dreissena polymorpha*, in North America and Europe. *American Zoologist*, 36(3):339–363.

Mills, E. L., Rosenberg, G., Spidle, A. P., Ludyanskiy, M., Pligin, Y., and May, B. (1996). A review of the biology and ecology of the quagga mussel (*Dreissena bugensis*), a second species of freshwater dreissenid introduced to North America. *American Zoologist*, 36(3):271–286.

Nichols, S. J. (1996). Variations in the reproductive cycle of *Dreissena polymorpha* in Europe, Russia, and North America. *American Zoologist*, 36(3):311–325.

Rees, H. C., Maddison, B. C., Middleditch, D. J., Patmore, J. R., and Gough, K. C. (2014). The detection of aquatic animal species using environmental dna – a review of edna as a survey tool in ecology. *Journal of Applied Ecology*, 51:1450–1459.

Schmelzle, M. C. and Kinziger, A. P. (2016). Using occupancy modelling to compare environmental DNA to traditional field methods for regional-scale monitoring of an endangered aquatic species. 16:1–14.

Sepulveda, A. J., Amberg, J. J., and Hanson, E. (2019). Using environmental DNA to extend the window of early detection for dreissenid mussels. *Managment of Biological Invasions*, 10(2):342–358.

U.S. National Park Service (2017). Invasive Zebra Mussels. `https://www.nps.gov/articles/zebra-mussels.htm`.

8   Appendix - R Code

A script containing all code used for this analysis is available at

github link here(?)... either that or include all code here

```r
# SIMULATED DATA 1
set.seed(1202020)
p <- 0.6 # constant detection probability... could change to depend on covariate
psi <- 0.6 # constant occupancy probability... could change to depend on covaria
M <- 10 # number of sites
J <- 8 # constant number of samples per site... does not need to be constant

z <- rbinom(M, 1, psi) # site occupancy

y1 <- matrix(NA, nrow = M*J, ncol = 4)
y1[, 1] <- rep(1:M, each  = J) # column indicating site
y1[, 2] <- rep(1:J, M) # column indicating sample within site
y1[, 3] <- rep(z, each = J) # column indicating true site occupancy

# creates a column of 1's and 0's indicating whether the species was detected
for(i in 1:(M*J)){
    y1[i, 4] <- rbinom(1, 1 , p*y1[i, 3]) # if the site is not occupied the spec
}

colnames(y1) <- c("Site", "Sample", "True Occupancy", "Detected")

library("car")
some(y1) # view 10 sample rows of the simulated data

# or

y2 <- rbinom(M, J, p*z) # total number of detections for the J samples within ea
# SIMULATED DATA 2
set.seed(1222020)
p <- 0.6 # constant detection probability... could change to depend on covariate
psi <- 0.6 # constant occupancy probability... could change to depend on covaria
```

```r
M <- 10 # number of sites
J <- sample(1:10, M, replace = T) # number of times each of the sites were sample

z <- rbinom(M, 1, psi) # site occupancy

y <- rep(NA, M)

for(i in 1:M){
  y[i] <- rbinom(1, J[i], p*z[i]) # total number of detections for the J samples
}
#SIMULATED DATA 3
set.seed(1222020)
M <- 10 # number of sites
x <- runif(10, 0, 10)
beta1.true <- 2
beta2.true <- 0.5
beta3.true <- 0.5
beta4.true <- 0.2
p <- exp(beta1.true - beta2.true*x)/(1 + exp(beta1.true - beta2.true*x)) # detect
psi <- exp(beta3.true + beta4.true*x)/(1 + exp(beta3.true + beta4.true*x)) # occu
J <- sample(1:10, M, replace = T) # number of times each of the sites were sample

z <- rep(NA, M)
y <- rep(NA, M)

for(i in 1:M){
  z[i] <- rbinom(1, 1, psi[i]) #site occupancy
  y[i] <- rbinom(1, J[i], p[i]*z[i]) # total number of detections for the J sampl
}
library("readxl")
USBR.login <- read_excel("C:/Users/mwind/OneDrive/Writing-Project/USBR_2018_data_s
USBR.sampleprep <- read_excel("C:/Users/mwind/OneDrive/Writing-Project/USBR_2018_d
USBR.microscopy <- read_excel("C:/Users/mwind/OneDrive/Writing-Project/USBR_2018_d
USBR.PCR <- read_excel("C:/Users/mwind/OneDrive/Writing-Project/USBR_2018_data_sam

unique(USBR.sampleprep$`Water Name`) # number of lakes (sites) in the 2018 sample
```