Survey Designs for Distance Sampling: A Study of Zebra Mussels

Alana Danieu, Nick Fredrickson,
Emily Kaegi, Clara Livingston
Advisor: Katie St. Clair

Carleton College

April 3, 2018

Agenda

- Statistical Reasoning

Agenda

- Statistical Reasoning
- Lake Burgan Data

Agenda

- Statistical Reasoning
- Lake Burgan Data
- Simulations

Agenda

- Statistical Reasoning
- Lake Burgan Data
- Simulations
- Time Analysis

Agenda

- Statistical Reasoning
- Lake Burgan Data
- Simulations
- Time Analysis
- Further Research

Estimating Mussel Abundance and Density

2 Step Approach:

1. Fit a detection function, $g(x)$, to our data

Estimating Mussel Abundance and Density

2 Step Approach:

1. Fit a detection function, $g(x)$, to our data
   - x = distance perpendicular to transect

Estimating Mussel Abundance and Density

2 Step Approach:

1. Fit a detection function, $g(x)$, to our data
   - $x =$ distance perpendicular to transect
2. Use information from $g(x)$ to estimate abundance using Horvitz-Thompson estimators

Estimating Mussel Abundance and Density

2 Step Approach:

1. Fit a detection function, $g(x)$, to our data
   - $x =$ distance perpendicular to transect
2. Use information from $g(x)$ to estimate abundance using Horvitz-Thompson estimators
   - Use for simulations

Estimating Mussel Abundance and Density

2 Step Approach:

1. Fit a detection function, $g(x)$, to our data
   - $x =$ distance perpendicular to transect
2. Use information from $g(x)$ to estimate abundance using Horvitz-Thompson estimators
   - Use for simulations

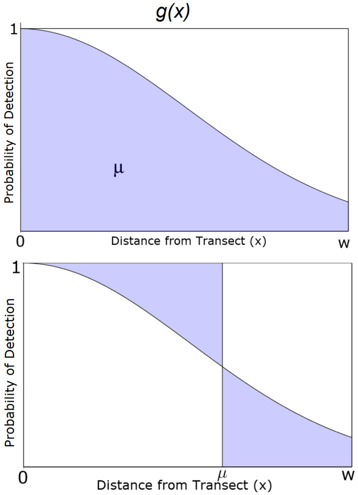
- We used a half-normal distribution for our models where

$$g(x) = exp\left[\frac{-x^2}{2\sigma^2}\right]$$

Estimating Detection Parameters

- Need proper probability density function that integrates to 1 for MLE

$$f(x) = \frac{g(x)}{\mu}$$



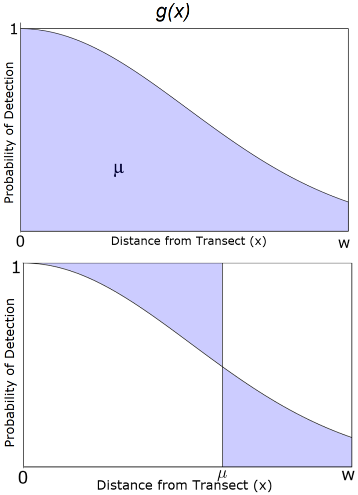
$\sigma = 0.5$, $\mu = 0.6$, and $w = 1$

Estimating Detection Parameters

- Need proper probability density function that integrates to 1 for MLE

$$f(x) = \frac{g(x)}{\mu}$$

- Normalizing Constant μ



$\sigma = 0.5$, $\mu = 0.6$, and $w = 1$

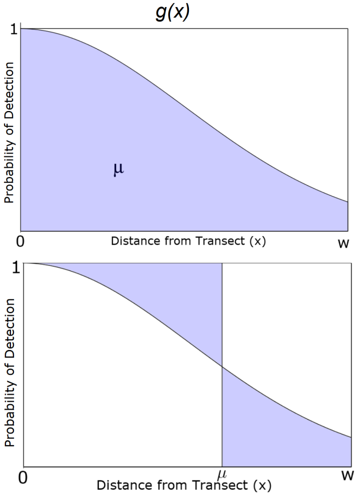
Estimating Detection Parameters

- Need proper probability density function that integrates to 1 for MLE

$$f(x) = \frac{g(x)}{\mu}$$

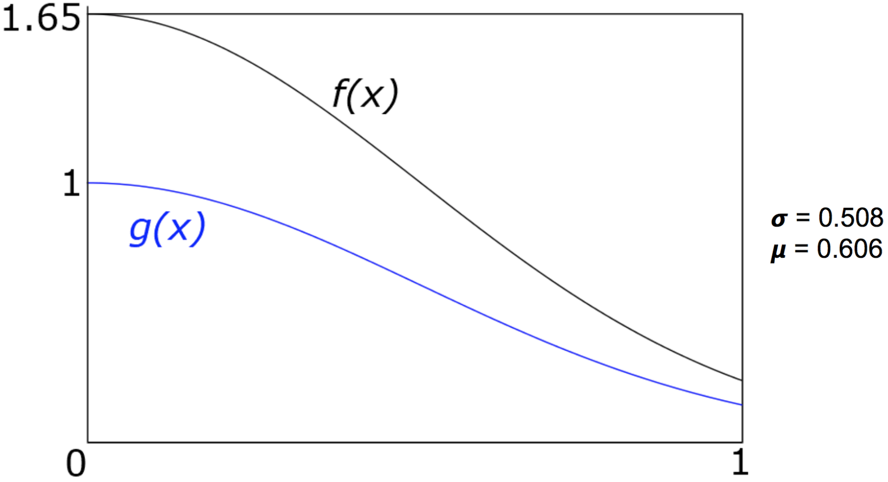
- Normalizing Constant μ
  - Effective Half-Width

$$\mu = \int_0^w g(x)dx$$



$\sigma = 0.5$, $\mu = 0.6$, and $w = 1$

Estimating Detection Parameters



Estimating Detection Parameters

- Maximum Likelihood Estimation

Estimating Detection Parameters

- Maximum Likelihood Estimation
  - Likelihood Function

$$L_x = \Pi_{i=1}^{n} f(x_i) = \frac{\Pi_{i=1}^{n} g(x_i)}{\mu^n} = \mu^{-n} exp\left[\frac{-\sum_{i=1}^{n} x_i^2}{2\sigma^2}\right]$$

Estimating Detection Parameters

- Maximum Likelihood Estimation
  - Likelihood Function

$$L_x = \Pi_{i=1}^{n} f(x_i) = \frac{\Pi_{i=1}^{n} g(x_i)}{\mu^n} = \mu^{-n} exp\left[\frac{-\sum_{i=1}^{n} x_i^2}{2\sigma^2}\right]$$

- Find σ that maximizes L_x

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}$$

Estimating Detection Parameters

- Maximum Likelihood Estimation
  - Likelihood Function

$$L_x = \Pi_{i=1}^{n} f(x_i) = \frac{\Pi_{i=1}^{n} g(x_i)}{\mu^n} = \mu^{-n} exp\left[\frac{-\sum_{i=1}^{n} x_i^2}{2\sigma^2}\right]$$

- Find σ that maximizes L_x

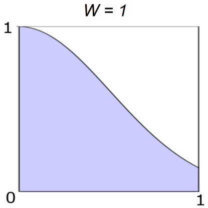
$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}$$

  - Only when we assume $w = \infty$

Average Probability of Detection

- Average Detectability

$$P_a = \frac{2\mu L}{2wL} = \frac{\mu}{w}$$

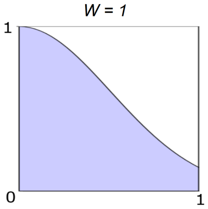


- $P_a = \mu = 0.606$
- $\sigma = 0.508$

Average Probability of Detection

- Average Detectability

$$P_a = \frac{2\mu L}{2wL} = \frac{\mu}{w}$$

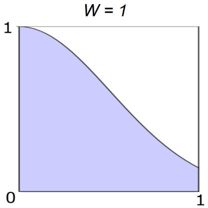


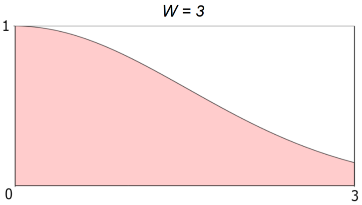
- $P_a = \mu = 0.606$
- $\sigma = 0.508$

Average Probability of Detection

- Average Detectability

$$P_a = \frac{2\mu L}{2wL} = \frac{\mu}{w}$$





- $P_a = \mu = 0.606$
- $\sigma = 0.508$

- $P_a = 0.606$
- $\mu = 1.818$
- $\sigma = 1.525$

Estimating Abundance

- Horvitz-Thompson Estimator

$$\hat{N} = \sum_{i=1}^{n} \frac{1}{p_i}$$

Estimating Abundance

- Horvitz-Thompson Estimator

$$\hat{N} = \sum_{i=1}^{n} \frac{1}{p_i}$$

- Where p_i is the probability that a detected mussel was found, thus

$$\hat{p}_i = \frac{a\hat{P}_a}{A}$$

Estimating Abundance

- Horvitz-Thompson Estimator

$$\hat{N} = \sum_{i=1}^{n} \frac{1}{p_i}$$

- Where p_i is the probability that a detected mussel was found, thus

$$\hat{p}_i = \frac{a\hat{P}_a}{A}$$

- And plugging back in we have

$$\hat{N} = \frac{nA}{a\hat{P}_a}$$

Calculating Standard Error of Density

- Coefficient of variation

Calculating Standard Error of Density

- Coefficient of variation
- Let $\hat{D}$ be our value of interest

$$CV(\hat{D}) = \frac{SE(\hat{D})}{\hat{D}}$$

Calculating Standard Error of Density

- Coefficient of variation
- Let $\hat{D}$ be our value of interest

$$CV(\hat{D}) = \frac{SE(\hat{D})}{\hat{D}}$$

- Thus, rewritten

$$SE(\hat{D}) = \hat{D} * CV(\hat{D})$$

Calculating Standard Error of Density

- Thus, we write

$$CV(\hat{D}) = \sqrt{\frac{\frac{K}{L^2(K-1)}\sum_{k=1}^{K} l_k^2(\frac{n_k}{l_k} - \frac{n}{L})^2}{(n/L)^2} + \frac{1}{2n}}$$

Calculating Standard Error of Density

- Thus, we write

$$CV(\hat{D}) = \sqrt{\frac{\frac{K}{L^2(K-1)} \sum_{k=1}^{K} l_k^2 (\frac{n_k}{l_k} - \frac{n}{L})^2}{(n/L)^2} + \frac{1}{2n}}$$

- L = total length of transects in survey

Calculating Standard Error of Density

- Thus, we write

$$CV(\hat{D}) = \sqrt{\frac{\frac{K}{L^2(K-1)} \sum_{k=1}^{K} l_k^2 (\frac{n_k}{l_k} - \frac{n}{L})^2}{(n/L)^2} + \frac{1}{2n}}$$

- L = total length of transects in survey
- K = total number of transects

Calculating Standard Error of Density

- Thus, we write

$$CV(\hat{D}) = \sqrt{\frac{\frac{K}{L^2(K-1)} \sum_{k=1}^{K} l_k^2 (\frac{n_k}{l_k} - \frac{n}{L})^2}{(n/L)^2} + \frac{1}{2n}}$$

- L = total length of transects in survey
- K = total number of transects
- n = total number of mussels found

Calculating Standard Error of Density

- Thus, we write

$$CV(\hat{D}) = \sqrt{\frac{\frac{K}{L^2(K-1)} \sum_{k=1}^{K} l_k^2 (\frac{n_k}{l_k} - \frac{n}{L})^2}{(n/L)^2} + \frac{1}{2n}}$$

- L = total length of transects in survey
- K = total number of transects
- n = total number of mussels found
- n_k = number of mussels found on the kth transect

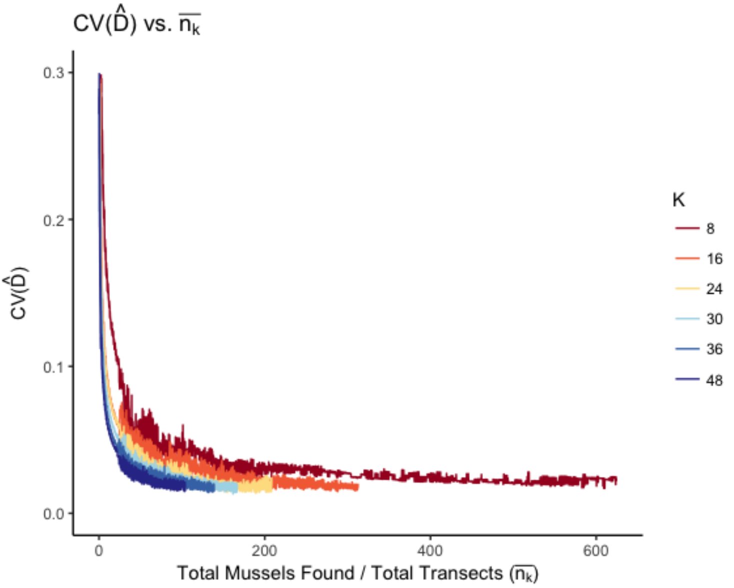
Calculating Standard Error of Density

- Thus, we write

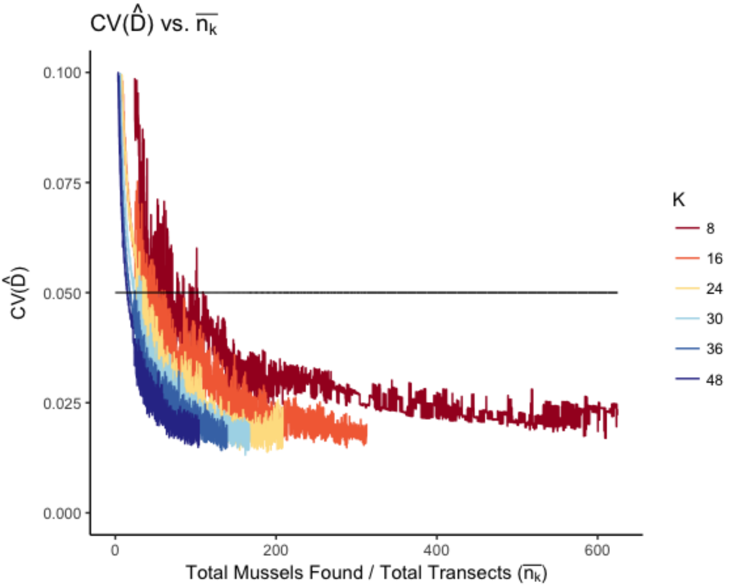
$$CV(\hat{D}) = \sqrt{\frac{\frac{K}{L^2(K-1)} \sum_{k=1}^{K} l_k^2 (\frac{n_k}{l_k} - \frac{n}{L})^2}{(n/L)^2} + \frac{1}{2n}}$$

- L = total length of transects in survey
- K = total number of transects
- n = total number of mussels found
- n_k = number of mussels found on the kth transect
- l_k = length of of the kth transect (30 meters for all)

Effects of Changing K and n



Effects of Changing K and n



Cause of Variability in cv($\hat{D}$)

$$CV(\hat{D}) = \sqrt{\frac{\frac{K}{L^2(K-1)} \sum_{k=1}^{K} l_k^2 \left(\frac{n_k}{l_k} - \frac{n}{L}\right)^2}{(n/L)^2} + \frac{1}{2n}}$$

- Multinomial randomization for variation in transects

Cause of Variability in cv($\hat{D}$)

$$CV(\hat{D}) = \sqrt{\frac{\frac{K}{L^2(K-1)} \sum_{k=1}^{K} l_k^2 \left(\frac{n_k}{l_k} - \frac{n}{L}\right)^2}{(n/L)^2} + \frac{1}{2n}}$$

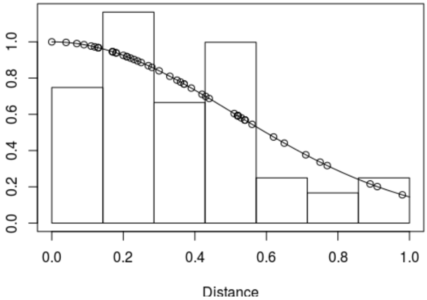
- Multinomial randomization for variation in transects
- Assume equal probability

Lake Burgan


Google
Imagery ©2018 , DigitalGlobe, USDA Farm Service Agency

Fitting Lake Burgan Data to a Model

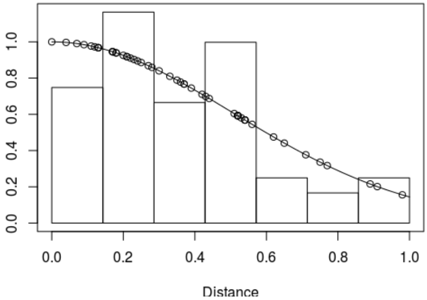
- $n = 52$ mussels



Parameter	Estimate	Std. Error	CV
$\hat{\sigma}$	0.508	0.084	0.165
$\hat{\mu} = \hat{P}_a$	0.606	0.075	0.123
$\hat{D}$	0.090	0.0199	0.222
$\hat{N}_a$	89.584	19.921	0.222
$\hat{N}_A$	10,760	2,392	0.222

Fitting Lake Burgan Data to a Model

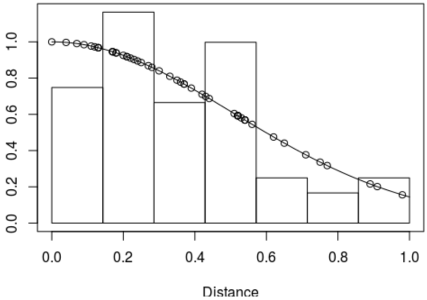
- $n = 52$ mussels
- $a = 999$ meters^2



Parameter	Estimate	Std. Error	CV
$\hat{\sigma}$	0.508	0.084	0.165
$\hat{\mu} = \hat{P}_a$	0.606	0.075	0.123
$\hat{D}$	0.090	0.0199	0.222
$\hat{N}_a$	89.584	19.921	0.222
$\hat{N}_A$	10,760	2,392	0.222

Fitting Lake Burgan Data to a Model

- $n = 52$ mussels
- $a = 999$ meters^2
- $A = 120,000$ meters^2



Parameter	Estimate	Std. Error	CV
$\hat{\sigma}$	0.508	0.084	0.165
$\hat{\mu} = \hat{P}_a$	0.606	0.075	0.123
$\hat{D}$	0.090	0.0199	0.222
$\hat{N}_a$	89.584	19.921	0.222
$\hat{N}_A$	10,760	2,392	0.222

Lake Burgan


Google
Imagery ©2018 , DigitalGlobe, USDA Farm Service Agency

Simulations

The variables we controlled in our simulations were:

- Region Size: 4000×30 $meters^2$

Simulations

The variables we controlled in our simulations were:

- Region Size: 4000×30 $meters^2$
- Population Size N

Simulations

The variables we controlled in our simulations were:

- Region Size: 4000×30 $meters^2$
- Population Size N
- Number of Transects K

Simulations

The variables we controlled in our simulations were:

- Region Size: 4000×30 $meters^2$
- Population Size N
- Number of Transects K
- Detection Scale Parameter σ

Simulations

The variables we controlled in our simulations were:

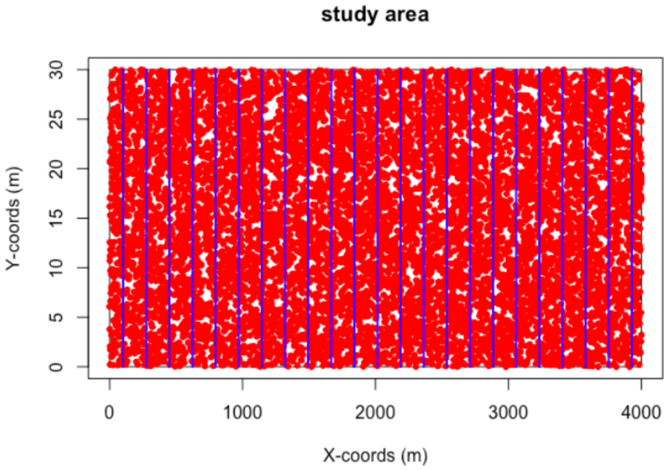
- Region Size: 4000×30 $meters^2$
- Population Size N
- Number of Transects K
- Detection Scale Parameter σ
- Number of Strata

Simulations

The variables we controlled in our simulations were:

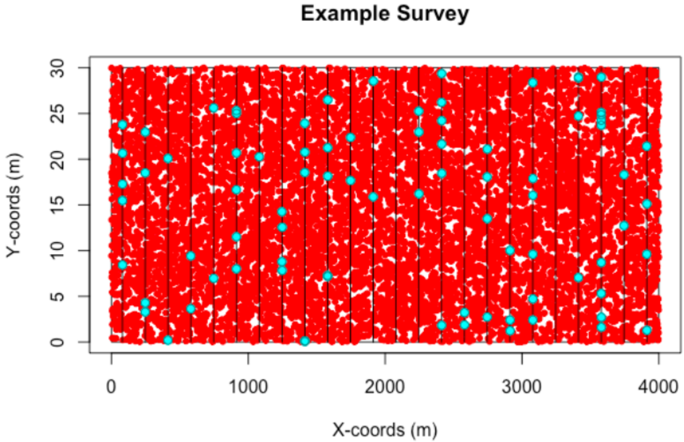
- Region Size: 4000×30 $meters^2$
- Population Size N
- Number of Transects K
- Detection Scale Parameter σ
- Number of Strata
- Addition of Hotspots (areas of elevated density)

Basic Simulation



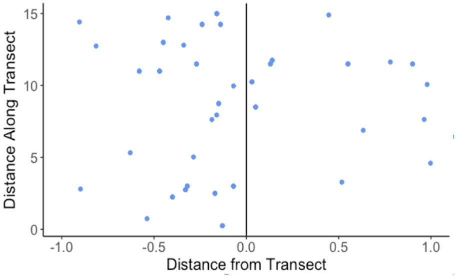
N = 10,000
K = 24

Basic Simulation



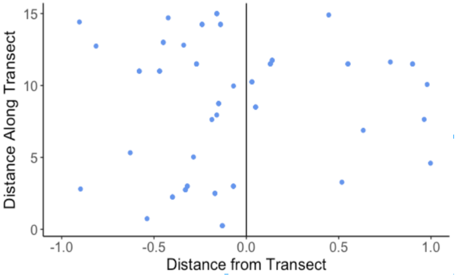
N = 10,000
K = 24
σ = .7

How the Simulation Works



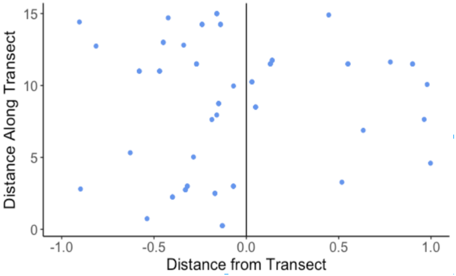
- x = distance from transect

How the Simulation Works



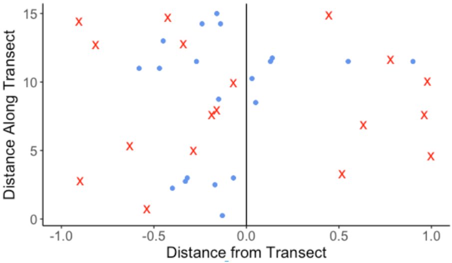
- x = distance from transect
- Each mussel is assigned a probability p_i

How the Simulation Works



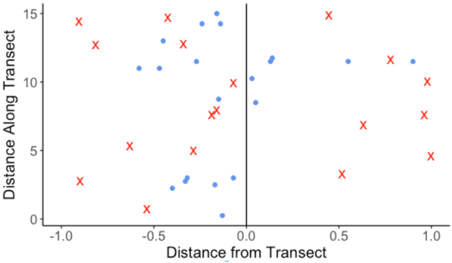
- x = distance from transect
- Each mussel is assigned a probability p_i
- $p_i = g(x_i)$

How the Simulation Works



- Detection $\sim$ Bernoulli(p_i)

How the Simulation Works



- Detection $\sim$ Bernoulli(p_i)
- Assigned a 1 if found, 0 if not found (red X)

Comparing Simulation Results

There are two results we use to quantify the difference between sampling designs:

- Percent Bias (Accuracy)

$$\% \hat{Bias}_{\hat{N}} = \frac{\hat{\hat{N}} - N}{N} \times 100\%$$

Comparing Simulation Results

There are two results we use to quantify the difference between sampling designs:

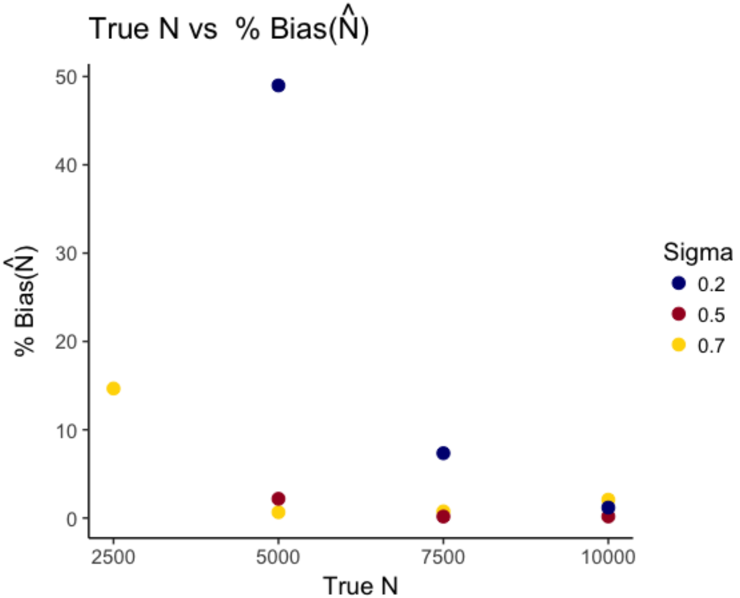
- Percent Bias (Accuracy)

$$\% \hat{Bias}_{\hat{N}} = \frac{\hat{\hat{N}} - N}{N} \times 100\%$$

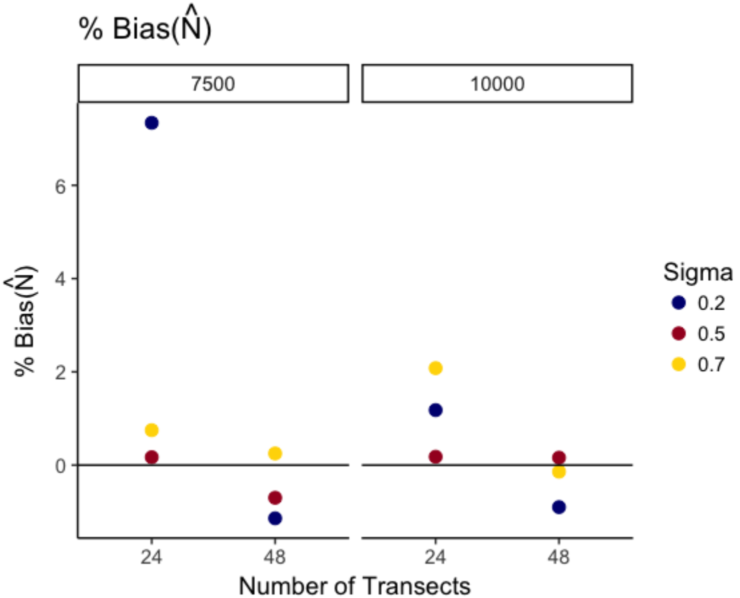
- Coefficient of Variation (Precision)

$$CV(\hat{N}) = \frac{SE(\hat{N})}{\hat{\hat{N}}}$$

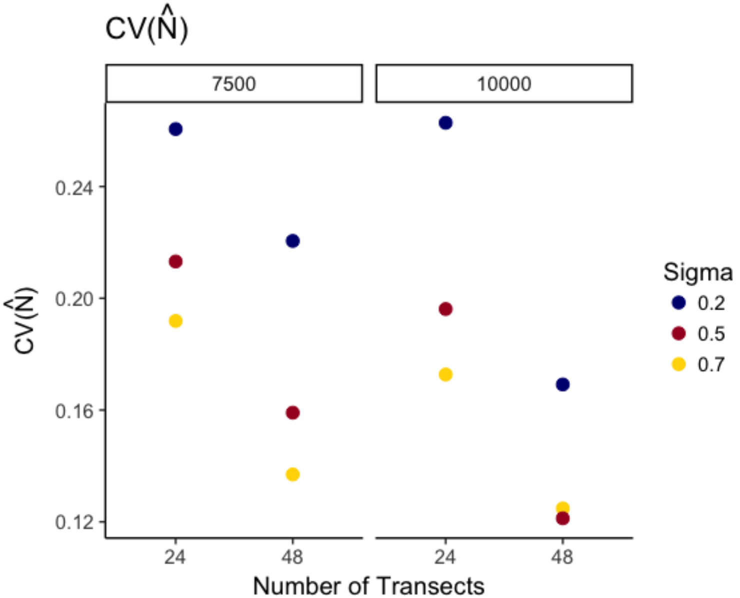
Varying N and σ



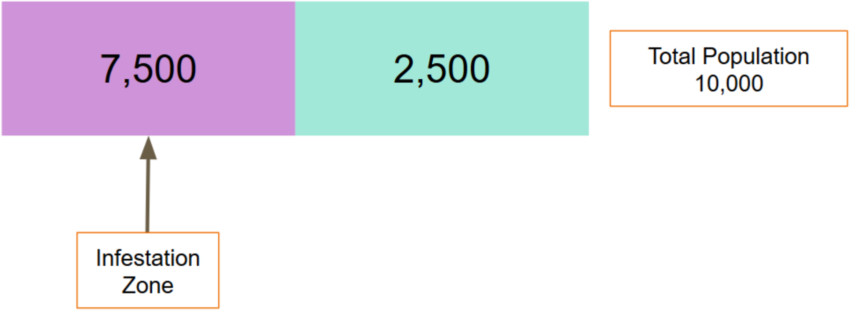
Varying number of transects, K



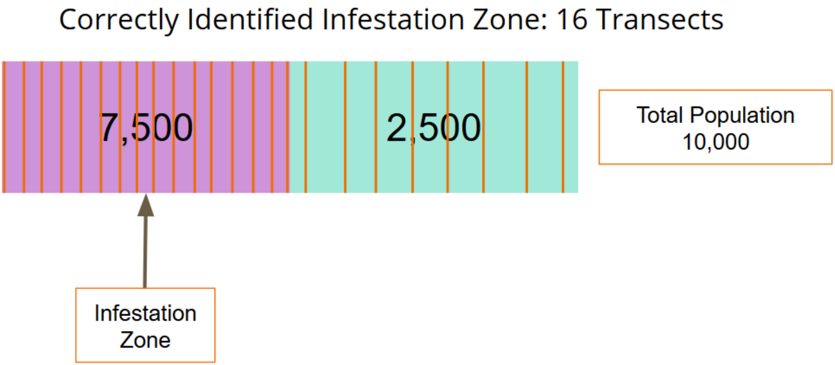
Varying number of transects, K



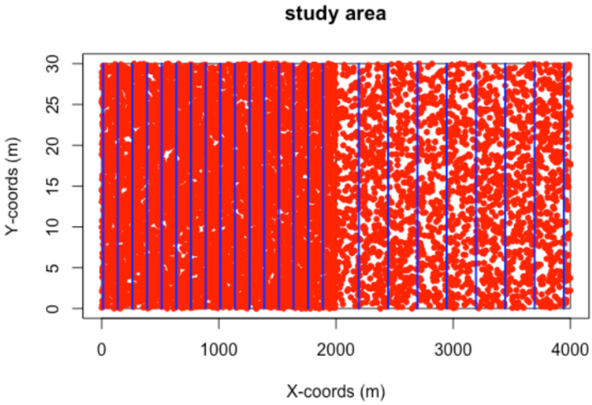
Stratified Design



Stratified Design: Correctly Identified

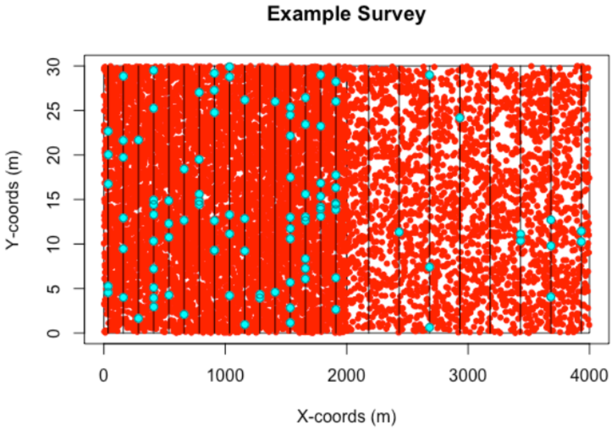


Stratified Design Simulation



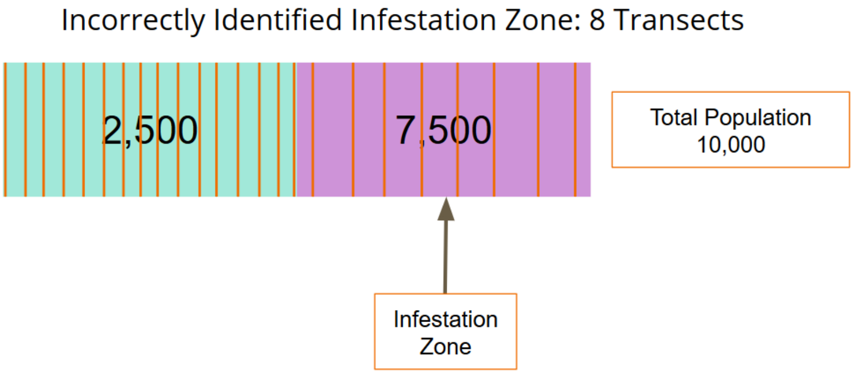
N = 7,500 & 2,500
K = 16, 8

Stratified Design Simulation



N = 7,500 & 2,500
K = 16, 8

Stratified Design: Incorrectly Identified



Stratified Design Results

Table: How Stratified Designs Effect Estimates

Design	$\bar{n}$	$\hat{\bar{N}}$	$\%Bias_{\hat{N}}$	$CV(\hat{N})$
Constant K	90	10,107	1.07%	.16
Correctly Identified	105	10,032	.32%	.16
Incorrectly Identified	75	9,985	−.15%	.17

Stratified Design Results

Table: How Stratified Designs Effect Estimates

Design	$\bar{n}$	$\hat{\hat{N}}$	$\%Bias_{\hat{N}}$	$CV(\hat{N})$
Constant K	90	$10,107$	1.07%	.16
Correctly Identified	105	$10,032$	.32%	.16
Incorrectly Identified	75	$9,985$	$-.15\%$	.17

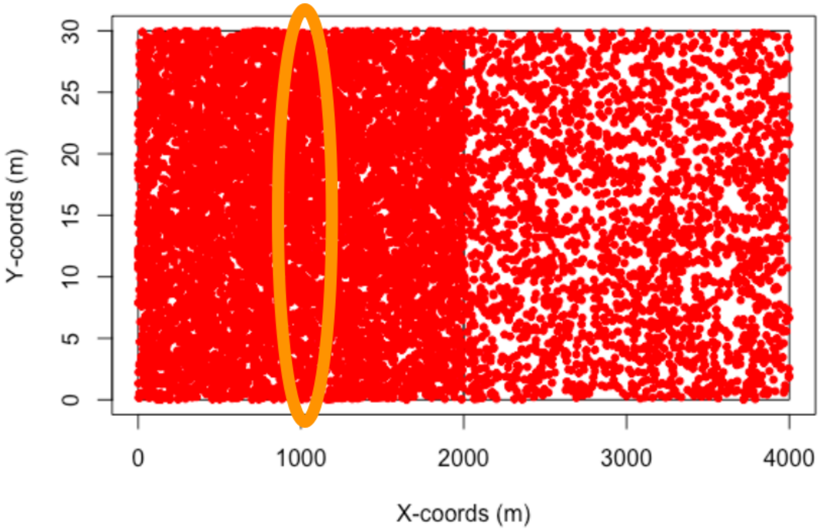
Stratified Design Results

Table: How Stratified Designs Effect Estimates

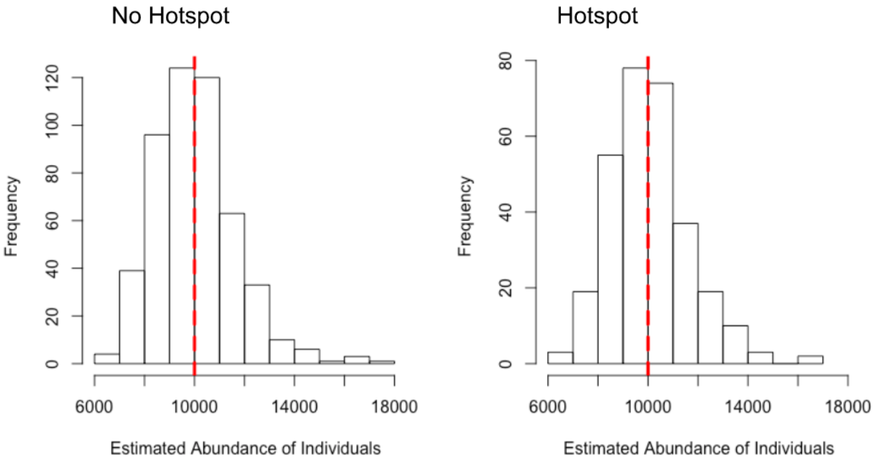
Design	$\bar{n}$	$\hat{\hat{N}}$	$\%Bias_{\hat{N}}$	$CV(\hat{N})$
Constant K	90	$10,107$	1.07%	.16
Correctly Identified	105	$10,032$	.32%	.16
Incorrectly Identified	75	$9,985$	$-.15\%$	.17

A .01 difference in $CV(\hat{N})$ is a difference in SE of $.01 * 10000 = 100$ mussels

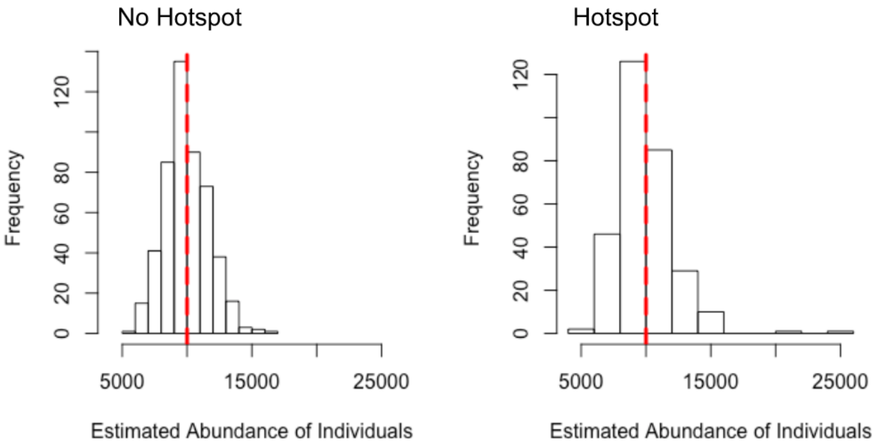
Addition of a Hotspot



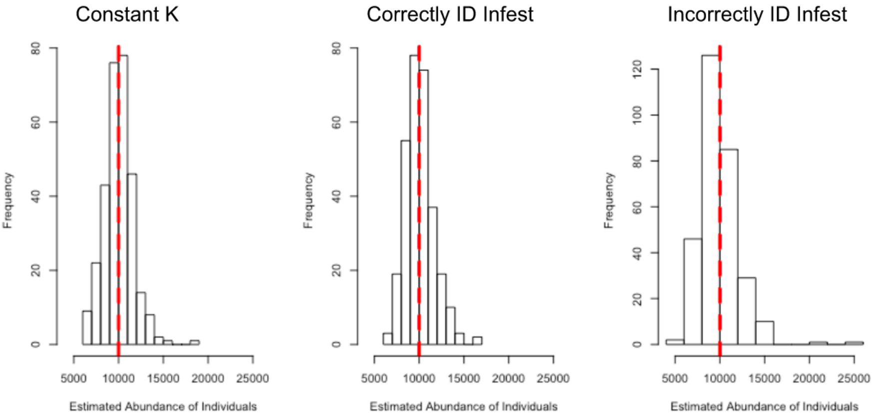
Hotspot Results:Correctly Identified Infestation Zone



Hotspot Results: Incorrectly Identified Infestation Zone



Simulation Results: Infestation Zone with Hotspot



$CV(\hat{N}) = 0.163$ $\quad$ $CV(\hat{N}) = .155$ $\quad$ $CV(\hat{N}) = .222$

Simulation Discussion

- Higher $\bar{n}$ meant more accurate and precise results

Simulation Discussion

- Higher $\bar{n}$ meant more accurate and precise results
  - Greater N and σ increase n

Simulation Discussion

- Higher $\bar{n}$ meant more accurate and precise results
  - Greater N and σ increase n
- Buckland suggests an n of at least 60-80

Simulation Discussion

- Higher $\bar{n}$ meant more accurate and precise results
  - Greater N and σ increase n
- Buckland suggests an n of at least 60-80
  - $\%Bias_{\hat{N}}$ was not significantly different than 0

Simulation Discussion

- Higher $\bar{n}$ meant more accurate and precise results
  - Greater N and σ increase n
- Buckland suggests an n of at least 60-80
  - $\%Bias_{\hat{N}}$ was not significantly different than 0
- Incorrectly identified hotspots can create large prediction errors

Simulation Discussion

- Higher $\bar{n}$ meant more accurate and precise results
  - Greater N and σ increase n
- Buckland suggests an n of at least 60-80
  - $\%Bias_{\hat{N}}$ was not significantly different than 0
- Incorrectly identified hotspots can create large prediction errors
- Predicted $SE(\hat{N})$ was smaller than the actual distribution of the $\hat{N}$ values from the 300-500 runs

Simulation Discussion

- Higher $\bar{n}$ meant more accurate and precise results
  - Greater N and σ increase n
- Buckland suggests an n of at least 60-80
  - $\%Bias_{\hat{N}}$ was not significantly different than 0
- Incorrectly identified hotspots can create large prediction errors
- Predicted $SE(\hat{N})$ was smaller than the actual distribution of the $\hat{N}$ values from the 300-500 runs
  - $SE(\hat{N})$ equation is biased

Experiment on Time

- Randomly placed 30 small marshmallows within transect

Experiment on Time

- Randomly placed 30 small marshmallows within transect
- $l = 24$ meters

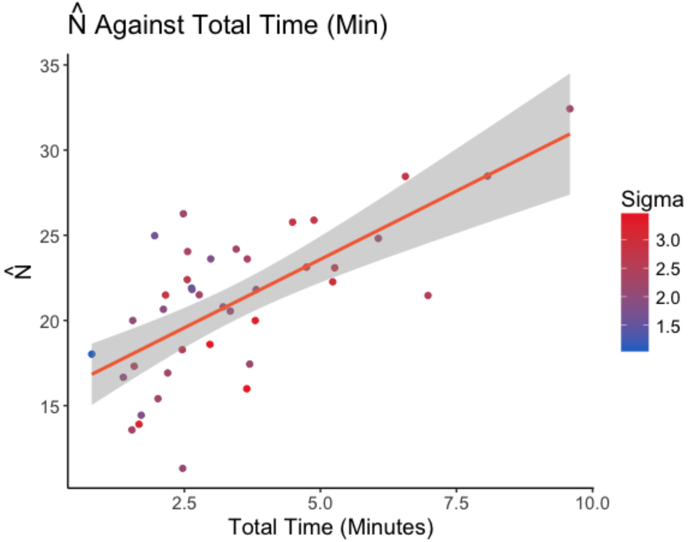
Experiment on Time

- Randomly placed 30 small marshmallows within transect
- $l = 24$ meters
- $w = 5$ meters

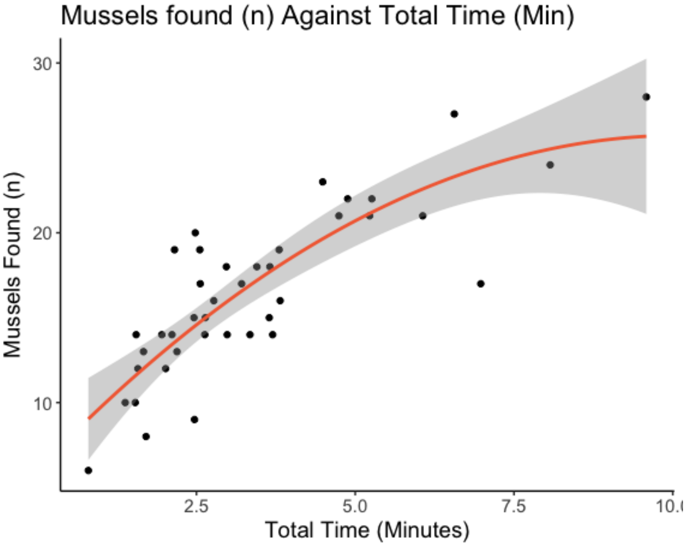
Experiment on Time

- Randomly placed 30 small marshmallows within transect
- $l = 24$ meters
- $w = 5$ meters
- Timed participants to see how time affects estimates

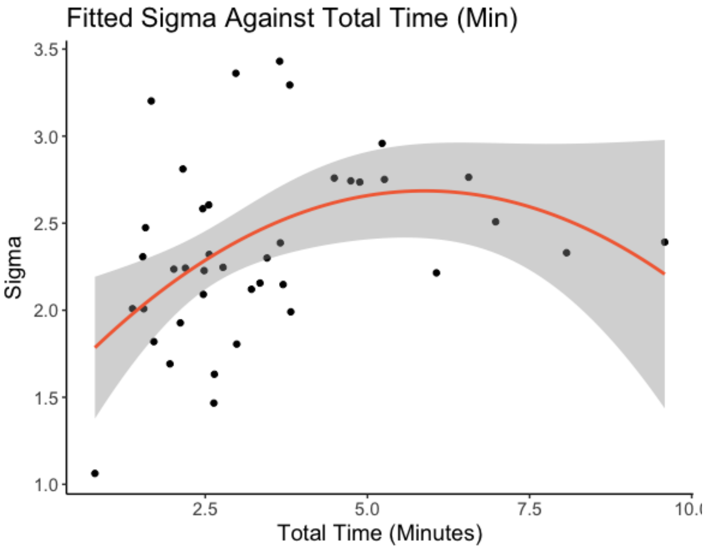
$\hat{N}$ Against Time



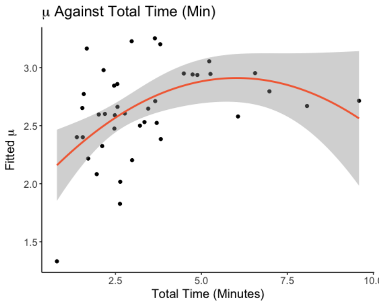
n Against Time

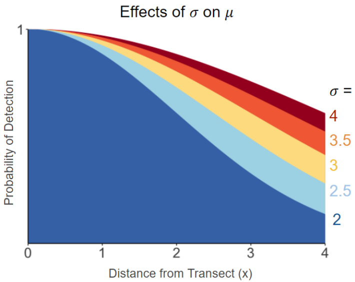


σ Against Time



Fitted μ


μ Against Total Time (Min)
Fitted μ
3.0
2.5
2.0
1.5
2.5
5.0
7.5
Total Time (Minutes)


Effects of σ on μ
Probability of Detection
1
0
1
2
3
4
Distance from Transect (x)
σ =
4
3.5
3
2.5
2

Relationship between σ, μ, n, and $\hat{N}$

$$\hat{N} = \frac{nA}{a\hat{P}_a}$$

Relationship between σ, μ, n, and $\hat{N}$

$$\hat{N} = \frac{nA}{a\hat{P}_a}$$

$$\hat{P}_a = \frac{\hat{\mu}}{w}$$

Relationship between σ, μ, n, and $\hat{N}$

$$\hat{N} = \frac{nA}{a\hat{P}_a}$$

$$\hat{P}_a = \frac{\hat{\mu}}{w}$$

$$\hat{N} = \frac{nA}{a(\hat{\mu}/w)}$$

Relationship between σ, μ, n, and $\hat{N}$

$$\hat{N} = \frac{nA}{a\hat{P}_a}$$

$$\hat{P}_a = \frac{\hat{\mu}}{w}$$

$$\hat{N} = \frac{nA}{a(\hat{\mu}/w)}$$

$\hat{N}$ is a function of n and μ, which depends on σ

Experiment Takeaways

- Time has a nonlinear relationship with σ, μ, and n

Experiment Takeaways

- Time has a nonlinear relationship with σ, μ, and n
- Time has a linear relationship with $\hat{N}$ as a result

Experiment Takeaways

- Time has a nonlinear relationship with σ, μ, and n
- Time has a linear relationship with $\hat{N}$ as a result
- Choose a time that maximizes detection

Experiment Takeaways

- Time has a nonlinear relationship with σ, μ, and n
- Time has a linear relationship with $\hat{N}$ as a result
- Choose a time that maximizes detection
- Choose a time that optimizes σ

Experiment Takeaways

- Time has a nonlinear relationship with σ, μ, and n
- Time has a linear relationship with $\hat{N}$ as a result
- Choose a time that maximizes detection
- Choose a time that optimizes σ
- Increased σ implies increased n

Experiment Takeaways

- Time has a nonlinear relationship with σ, μ, and n
- Time has a linear relationship with $\hat{N}$ as a result
- Choose a time that maximizes detection
- Choose a time that optimizes σ
- Increased σ implies increased n
- Supports the claim that we can control $CV(\hat{D})$ using n

Further Research

- Incorporating habitat covariates

Further Research

- Incorporating habitat covariates
- Realistic hotspot

Further Research

- Incorporating habitat covariates
- Realistic hotspot
- More thorough experiment on time

Further Research

- Incorporating habitat covariates
- Realistic hotspot
- More thorough experiment on time
- Data limitations

References

- Buckland, S.T., Rexstad, E.A., Marques, T.A., Oedekoven, C.S. 2015. Distance Sampling: Methods and Applications. Switzerland. Springer International Publishing.
- Hart, R.A., A.C. Miller, and M. Davis. 2001. Empirically Derived Survival Rates of a Native Mussel, Amblema plicata, in the Mississippi and Otter Tail Rivers, Minnesota. American Midland Naturalist 146: 254-263.
- Hebert, P. D. N., B. W. Muncaster, G. L. Mackie. 1989. Ecological and genetic studies on Dreissena polymorpha (Pallas): a new mollusk in the Great Lakes. Can. J. Fish. Aquat. Sci. 46: 1587-1591.
- Limburg, K. E., V. A. Luzadis, M. Ramsey, K. L. Schulz, and C. M. Mayer. 2010. The good, the bad, and the algae: perceiving ecosystem services and disservices generated by zebra and quagga mussels. Journal of Great Lakes Research 36:86-92.
- Marshall, Laura. 2017. DSsim: Distance Sampling Simulations. R package version 1.1.2. https://CRAN.R-project.org/package=DSsim
- Miller, David Lawrence. 2017. Distance: Distance Sampling Detection Function and Abundance Estimation. R package version 0.9.7. https://CRAN.R-project.org/package=Distance
- Miller, E. B., M. C. Watzin. 2007. The effects of zebra mussels on the lower planktonic foodweb in Lake Champlain. Journal of Great Lakes Research 33(2):407-420.
- Qualls, T. M., D. M. Dolan, T. Reed, M. E. Zorn, J. Kennedy. 2007. Analysis of the impacts of the zebra mussel, Dreissena polymorpha, on nutrients, water clarity, and the chlorophyll-phosphorus relationship in Lower Green Bay. Journal of Great Lakes Research 33(3):617-626.
- USGS Nonindigenous Aquatic Species. Dreissena polymorpha. https://nas.er.usgs.gov/queries/factsheet.aspx?speciesID=5
- Vanderploeg, H. A., J. R. Liebig, W. W. Carmichael, M. A. Agy, T. H. Johengen, G. L. Fahnenstiel, and T. F. Nalepa. 2001. Zebra mussel (Dreissena polymorpha) selective filtration promoted toxic Microcystis blooms in Saginaw Bay (Lake Huron) and Lake Erie. Can J. Fish. Aquat. Sci. 58: 1208-1221.
- Virginia Department of Game and Inland Fisheries. Zebra Mussels (Dreissena polymorpha). https://www.dgif.virginia.gov/wildlife/zebra-mussels/