Tradeoffs in detections efficiency and area covered in the design of optimal survey

Jake M. Ferguson, Aislyn Keyes, Michael McCartney, Katie St. Clair, Douglas Johnson, John Fieberg

Introduction

- impacts and costs of surveying invasive species: Past work has focused primarily on the early detection (e.g., Ferguson et al. 2014, @Holden2016) and eradication (**???**) of invasive species or detecting rare and endangered species (**???**), however the complementary problem of designing efficient surveys to determine population distribution and density of animals post-detection is often a larger component of the management efforts. Determining how to optimize these efforts has received much less attention.
- Something about usefulness of Unionids as a study system for studying detection methods [Green1993, Smith2006]…
- There likely exists an empirical tradeoff between survey efficiency and survey coverage in most study designs. Surveys performed deliberately (*word choice*) will likely be time-consuming and thus cover a limited area, but have a higher probability of detecting indidivuals. If a survey is performed quickly it can cover more area at the expense of lower detection rates.
- Here we explored two types of surveys for zebra mussels. We used transect surveys with distance sampling, which covers a larger area but has imperfect detection, and quadrat surveys, which represent a low efficiency type of survey with high detection probability. We examined how each design performed under a range of densities. Finally, we used the empirical studies to parameterize a general model that allowed us to examine the determine the best survey approach for a given problem based on **system size and density**.

Methods

Surveys

We conducted two different types of surveys in three Minnesota lakes in 2018. We decided which lakes to survey based on initial visits to six different lakes throughout central Minnesota (Christmas Lake, East Lake Sylvia, Lake Burgan, Lake Florida, Little Birch Lake, Sylvia Lake) that have had confirmed recent zebra mussel infestations (as determined by Minnesota Department of Natural Resources). At each lake we visited 15 different sites distributed evenly around the lake. Each site was placed in 3 to 8 m of water and determined using a bathymetry shapefile in ArcMap. We located each point in the field using a GPS unit (Garmin GPSMAP 64s). At each site our team of two divers spent 15 minutes underwater counting zebra mussels. We used these counts to determine the three lakes to conduct full surveys on, selecting lakes that displayed a range of apparent densities. **how were quadrat/distance sites differented**

Based on our initial 15 minute exploratory dives we conducted full surveys on Lake Florida in Kandiyohi County, Lake Burgan in Douglas County, and Little Birch Lake in Todd County. Lake Florida covers an area of 273 hectares and has a maximum depth of 12 m, Lake Burgan covers an area of 74 hectares and has a maximum depth of 13 m, Little Birch Lake covers 339 hectares and has a maximum depth of 27 m. We surveyed each of the 15 previously selected sites in each lake using two type of surveys; quadrat and distance sampling with removal.

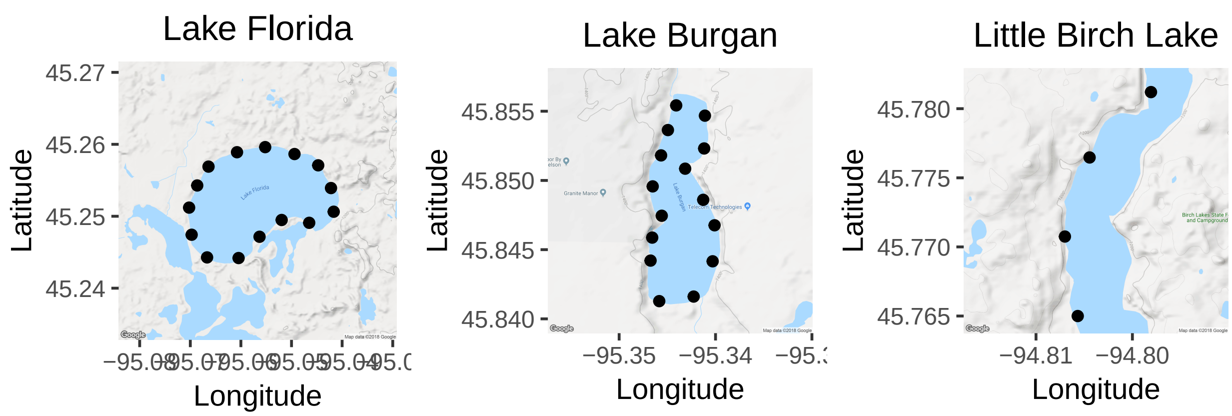


Figure 1: Map of survey sites, each point indicates the starting location of a transect. Remove place name labels. Fix x-axis on Florida.

Quadrat surveys

At each site we used the previously defined transect locations to determine the start of a transect. We ran out parallel 30 m transect lines 1 meter apart perpendicular to the shoreline, though transects were stopped earlier than 30 m if divers ran into the thermocline. Our team of two divers each took one of the transects, placing a 0.5×0.5 square meter quadrat every 2 meters along the transect. In each quadrat the diver counted all the mussels within the quadrat.

Distance sampling surveys

At each site we used the previously defined transect locations to determine the start of each survey transect. We then ran out a 30 m transect in a direction perpendicular to the shoreline, though transects were stopped earlier than 30 m if divers ran into the thermocline. Divers surveyed 1 m on either side of the transect for a transect belt that is 2 m wide.

We conducted removal surveys, which require two divers. In the removal survey, the first diver swam over the transect line marking detections for the second diver. The second diver then tried to detect animals missed by the first diver. We implemented the distance removal survey by having the primary diver swim over the transect line. Whenever the diver detected a zebra mussel or cluster of mussels, they marked the location with a survey flag then recorded the number of mussels in the cluster, the distance from the transect start to the detection (hereafter transect distance), and the perpendicular distance from the location of the detection to the transect line (hereafter detection distance). The secondary diver then looked for zebra mussels that were missed by the primary diver. Divers rotated through the primary and secondary observer roles in order to *average out* potentially innate differences between observers (Cook and Jacobson 1979).

- remove habitat survey information?

How were survey transects separated?

Statistical analysis

- Distance sampling: the counts for each transect are denoted as x_i with the total counts in the lake with n transects is denoted as $X = \sum_i^n x_i$. The length of each transect is denoted as l_i and the total length of $L = \sum_i^n l_i$. The encounter rate x_i/l_i is used to weight the survey effort. The estimated detection probability on a transect is denoted as $\hat{P}$ and the cluster size of each detection event is denoted as s. The estimated density and variance in density (Buckland et al. 2001) are given by

$$\hat{D} = \frac{\sum_i^n x_i}{\hat{P}A} \tag{1}$$

$$\text{var}(\hat{D}) = \hat{D}^2 \left(\frac{\text{var}(X)}{X^2} + \frac{\text{var}(\text{E}(s))}{\text{E}(s)^2} + \frac{\text{var}(A\hat{P})}{A^2\hat{P}^2} \right), \tag{2}$$

with the total survey area $A = \sum_i^n a_i$, variance in total counts $\text{var}(X) = \left(L \sum_i^n l_i(x_i/l_i - X/L)^2\right)/(n-1)$, and the expected value and variance in the expected cluster size (denoted as $\text{E}(s)$ and $\text{var}(\text{E}(s))$), respectively. The variance in the detectability $\text{var}(A\hat{P}$ is **XX**. Below we describe how to determine the detection probability component using distance sampling.

In distance sampling the transect distance is used to estimate how detectability changes with distance from the transect line. We modeled detection probabilities using two model subcomponents. The distance component of the detection function, describes how distance (y) leads to changes in the probability of detecting a zebra mussel. Based on our initial inspection of the detections (Figure **??**) we used the hazard-rate detectability model, which incorporates a shoulder in the detection function. A major assumption in the conventional distance sampling framework is that detection is perfect on the transect line, an assumption that has been shown to not hold for zebra mussel surveys (**???**). The mark-recapture component of the model is used to relax this assumption by using the removal surveys to estimate detectability on the transect line. In the mark-recapture model we assumed that each dive team collected data independently; thus, we used the point independence assumption described by Borchers et al. (2006). Point independence accounts for the effects of unmodeled covariates that can induce unexpected correlations between observers. We also included the effect of cluster size as a covariate influencing detection in the mark-recapture component of the model. We estimated detection using the mrds package (Laake et al. 2018), which accounts for removal survey and for the effect of distance on detectability. We obtain the transect-level detectability $\hat{P}$ of detecting an zebra mussel cluster by integrating the detectability function over the transect strip-width (Buckland et al. 2001).

We obtained estimates of density from quadrat sampling by modifying equations 1 and 2 to assume that detection is perfect ($\hat{P} = 0$ and $\text{var}\left(\hat{P}\right) = 0$) and to account for counting individuals, not clusters ($\text{var}\left(\text{E}(S)\right) = 0$).

Finally, we looked at whether informally surveying an area can accurately represent the relative density of zebra mussels, Here, we used the initial surveys conducted at the start point of each transect for a fixed amount of time over an unknown area. We calculated the correlation between these count rates and the estimated densities on each transect.

Time budget analysis

- In order to determine the efficiency of each method, we quantified how much time each method required to complete a transect and the time required to move between transects.
- We modeled the time to conduct a transect survey by breaking up the total survey time into two components, the time to setup the transect and the time to perform the transect survey **I am not including time to do habitat survey because it is same across all types of survey, and we don't use habitat data in the analysis**. The first component, the time to setup the transect survey (hereafter referred to as the setup time), was the time required to place the transect line(s). We modeled the setup time using linear mixed-effects regression analysis with fixed-effect covariates of survey type (distance or quadrat, included as a qualitative variable), and the transect length (continuous variable). We also included lake as a random intercept term. The second component of the time budget was the time to perform the survey on each transect (hereafter referred to as the search time *word choice*). The model structure of this component included fixed effects of transect length, an interaction between the type of survey and the detection rate (number of detections per meter of transect length) and a random slope term allowing detection time to vary with lake to account for heterogeneity that occurs due to lake conditions.
- We used the fitted models to predict the setup time and search time for a transect of length 30 m under using the estimated from the three lakes in this study.
- We modeled the time to move between consecutive transects as a function of distance between transects (hereafter, referred to as the movement time). We built the movement time model using a linear mixed-effects model to relate the time spent relocating between consecutively sampled transects to the distance between those transects fore each lake in the initial density survey (for a total of six lakes). We included lake name as a random intercept term. All random effects models in this study were fit using lmer in the lme4 package (*citation*).
- We investigated the movement time for surveys with fewer transects by sequentially removing every ith transect and calculating the empirical distance between the new transect start locations. We iteratively removed every ith transect from the original 15 transects (for $i = 2$ to 8) in each of our lakes and calculated the distance between those sites. We used the travel time model described above to predict the the average time travel time for a study with n transects conducted on each lake. We modeled the travel time to conduct a survey on each lake with n transects.

Table 1: Summary of survey results.

	Lake Florida			Lake Burgan			Little Birch La	
Design	Transects	Area surveyed	Detections	Transects	Area surveyed	Detections	Transects	Area surveyed
Quadrat	15	112.5	8	15	71.5	40	4	21.5
Distance	15	900.0	10	15	602.0	79	4	124.0
Double	15	450.0	5	15	277.0	47	4	84.0

- We combined the the setup time model (denoted at τ_{S}), search time model (τ_{E}), and movement time model (τ_{M}) to determine the expected number of transects that can be under a specific amount of time. We then determined the maximum number of transects that can be completed when a fixed total amount of time is available to complete the surveys (τ_{T}). The maximum number of transects that can be completed can be determined by minimimizing the quantity $C(n)$ with respect to the number of transects surveyed, n, in the constrained optimization

$$C(n) = \tau_{\text{T}} - n\left(\tau_{\text{S}} + \tau_{\text{E}}(D,\, \hat{P}) + \tau_{\text{M}}(n)\right)$$
$$C(n) \geq 0.$$

remove equation?

- We determined the maximum number of surveys conducted, n_{max}, obtained from solving the above optimization problem. We used n_{max} and the empirical estimates of the standard error in lake density that were rescaled to a sample with n_{max} transects, $\hat{\sigma}\sqrt{\frac{n}{n_{\text{max}}}}$, where n is the number of transects from the original sample (given in Table 1). This rescaled standard error was used to compare the efficiency of the two different survey designs in each lake.

Simulation study

Look at tradeoffs in density and system size (travel time)… Options: * use empirical model to figure out time budget. * Use variance formula with Poisson/Tweedie/NB density function and ignore uncertainty in detection?. Plug into variance model.

Results

Surveys

- Quadrat surveys: The amount of area surveyed and number of zebra mussels counted in each lake are given in (Table 1). Estimates of density along with the corresponding standard errors are give in (Figure **??**). We estimated that Lake Florida has the lowest density, Little Birch Lake had the highest density, and Lake Burgan had an intermediate density.

- Distance survey summary: The amount of area surveyed and number of zebra mussel detections made in each lake are given in (Table 1). In Lake Florida all detections were of single zebra mussels, in Lake Burgan the average cluster size was 1.18 (standard deviation = 0.13), and Little Birch Lake had the largest average cluster size with 7.83 (4.02).
- Our estimates of the detection functions indicated that the detectability of a cluster were similar between lakes with the estimated transect detectability in Lake Florida (reported as mean (standard error)) 0.31 (0.11), Lake Burgan 0.53 (0.07), and in Little Birch Lake 0.47 (0.07). Unfortunately, a lack of detections by observer 2 in Lake Florida likely leads us to be overconfident of our detection

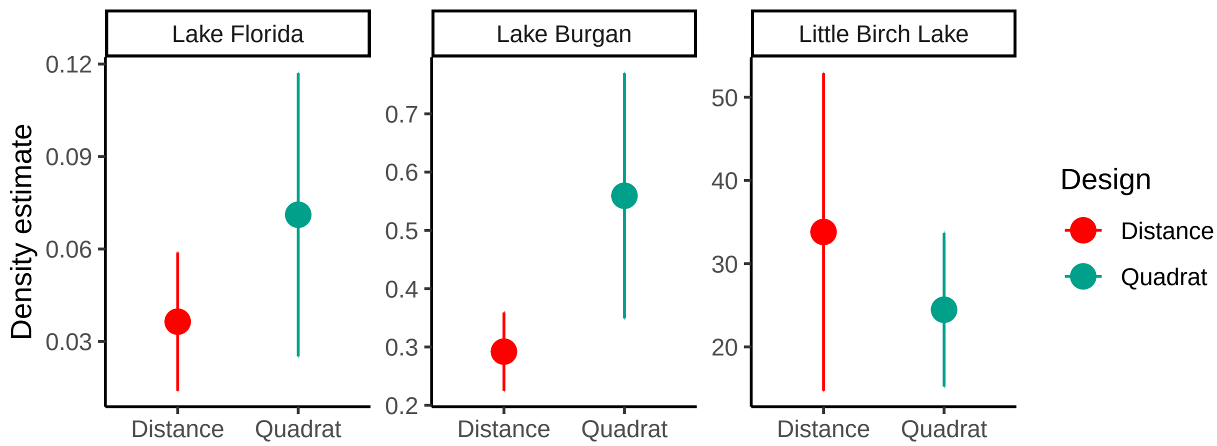


Figure 2: Density estimates and standard errors.

function. Despite the similarities in the overall estimates of detectability there was some differences in the shoulder width of the detection functions, potentially due to the higher vegetation levels in Little Birch Lake leading to a shorter shoulder (Figure 4).

- Estimated detection functions par estimates,
- Distance surveys: amount of area surveyed and number of detections in each lake, are given in Table 1. The average cluster size of detections in Lake Florida was , density estimate and standard error. Estimates of density along with the corresponding standard errors are give in (Figure **??**). We determined that Lake Florida has the lowest density, Little Birch Lake had the highest density, and Lake Burgan had an intermediate density.

Time budget analysis

- In both the setup time and search time models we found that transect length and detection rate had postive effects on the time budget (Figure **??**).
- We found that the transit time varied by transect distance, however "
- We found that the transit time varied by transect distance, however

Determining the optimal design

Optimization criterion

We want to know the optimal design under a fixed amount of time to sample as a function of the underlying density of zebra mussels. We start with the density variance model in equation (2), making a few simplifying assumptions. We will start assume that all transects are of the same length and area, a. Our goal is to find a function that can be optimized to determine the best design given the detectability P of targets, the underlying density of the population D, and a certain amount of time available to conduct the survey. We will use the estimated density precision $\text{cv}(\hat{D})$ as the objective function. Thus, we need to determine how the

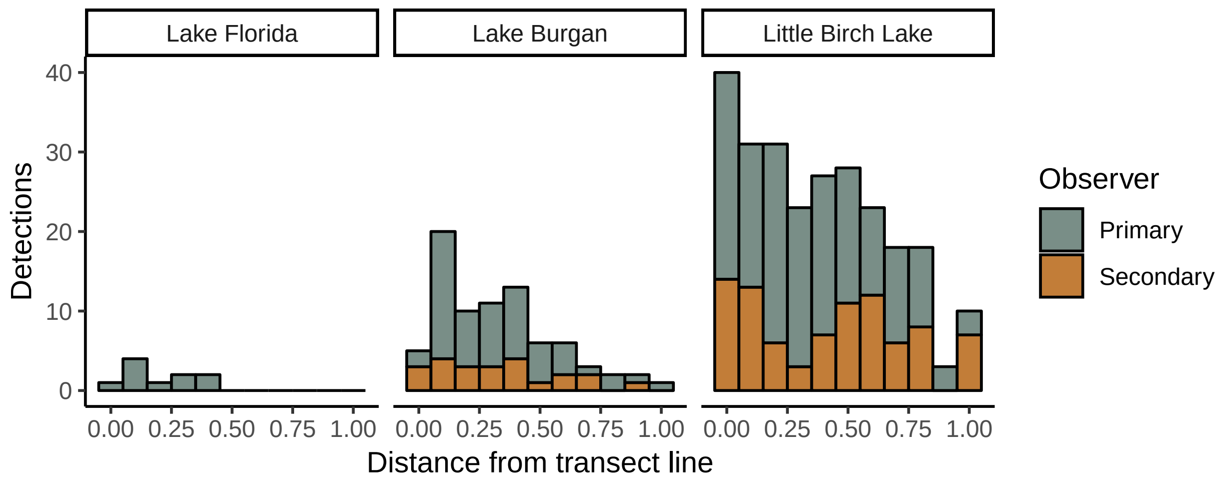


Figure 3: Stacked histogram showing the total number zebra mussel detections made by the primary and secondary diver in each lake. Distance bin widths are 0.1 m.

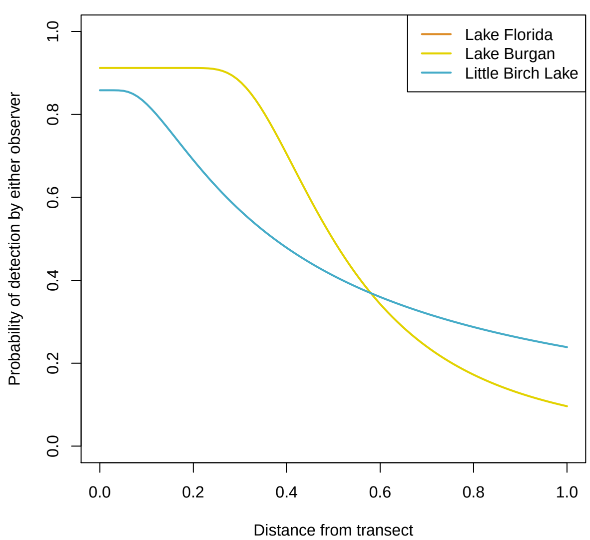


Figure 4: Average detection probability as a function of distance estimated using distance sampling for each lake. Consider removing figure...

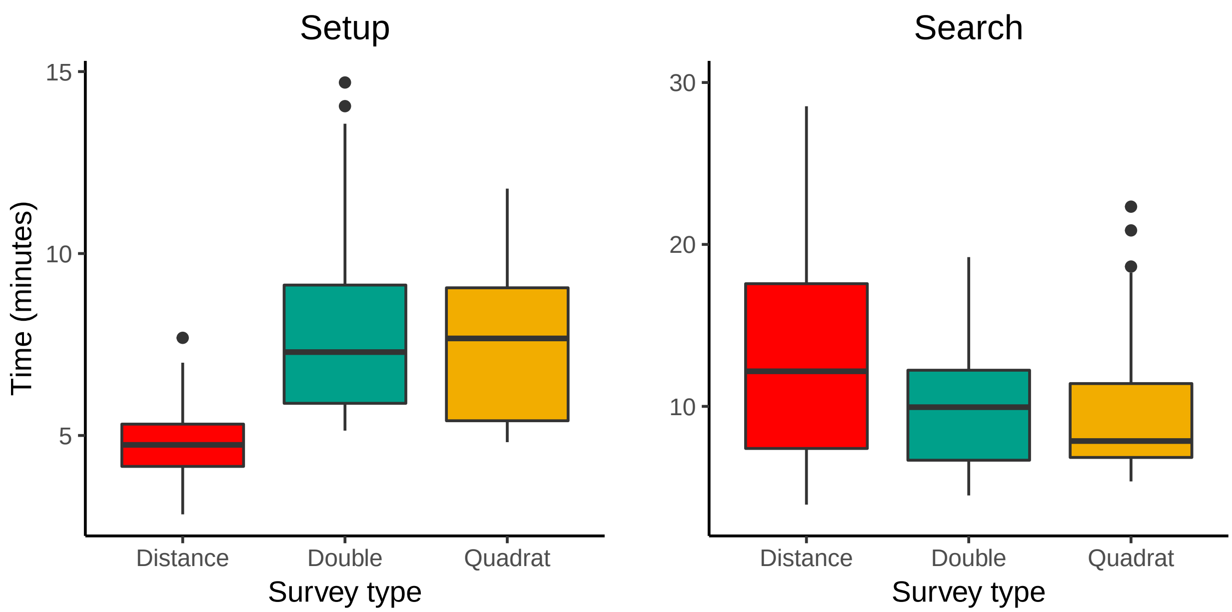


Figure 5: Empirical times for the setup and search. Consider removing figure...

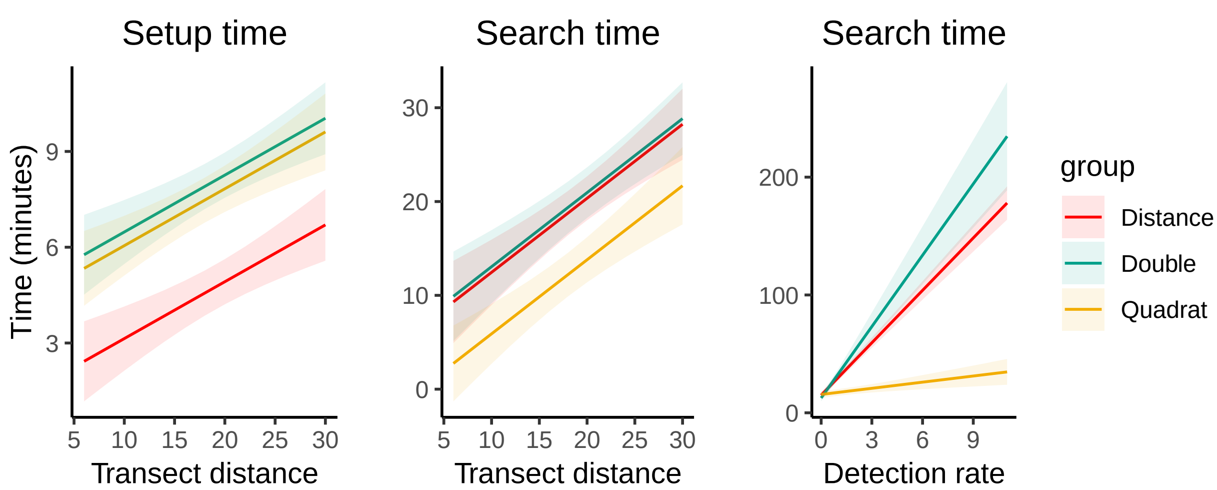


Figure 6: The impact of transect distance and detection rate on on the setup and search time. Predictinos are of the fixed effects. Bands represent 95% confidence intervals of the predicted values.

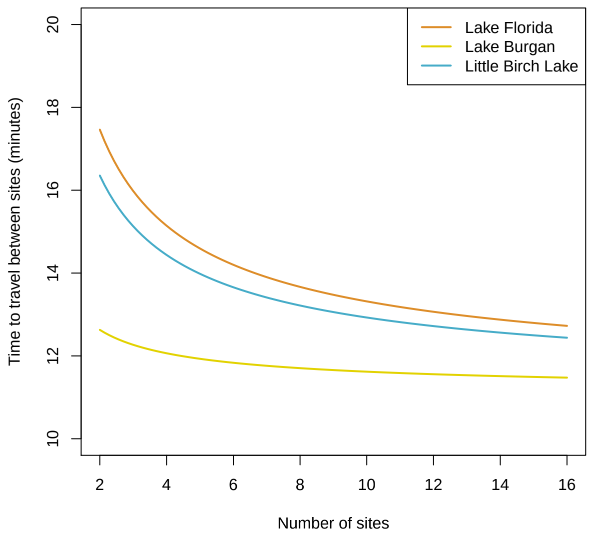


Figure 7: Predicted travel time at each lake.

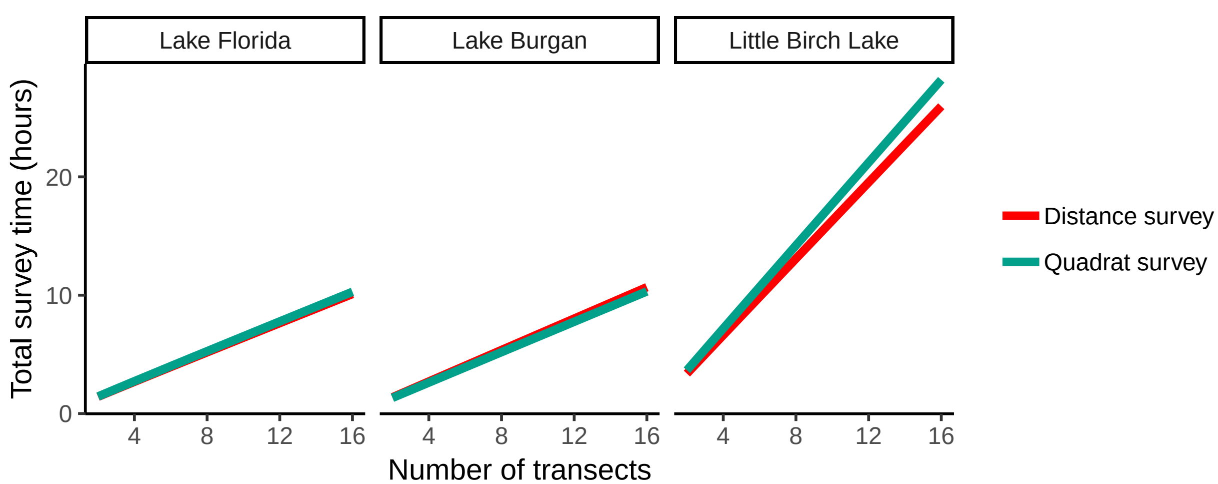


Figure 8: Maximum number of transects surveyed as a function of the total time available to conduct surveys.

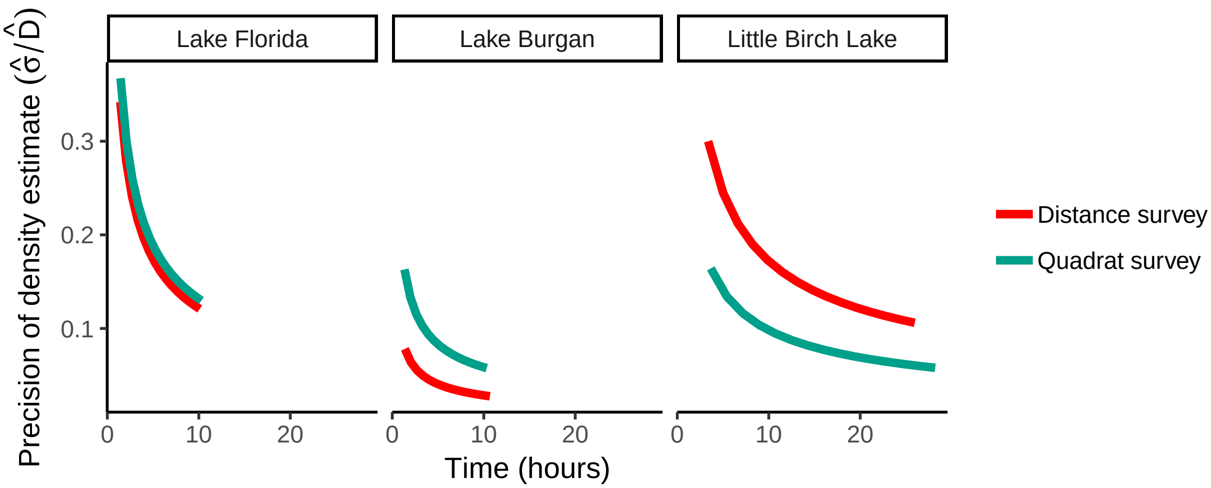


Figure 9: The predicted coefficient of variation in each lake. Fix x-axis.

variance in counts ($\text{var}(X)$) and the variance in the estimated detection ($\text{var}(\hat{P})$) drive variation in density.

$$\frac{\text{var}(\hat{D})}{\hat{D}^2} = \frac{\text{var}(X)}{X^2} + \frac{\text{var}(\hat{P})}{\hat{P}^2}$$
$$\text{cv}(\hat{D})^2 = \text{cv}(X)^2 + \text{cv}(\hat{P})^2$$

We can consider the total counts, X as a random variable arising from the Poisson (negative binomial) distribution with estimated rate $\lambda = nDaP$, then $\text{cv}(X)^2 = 1/\lambda$ ($1/\lambda + \alpha$, where α is the overdispersion). Here we are treating X as a random variable

We will assume that the detections can be modeled as binomial samples (following Moran 1951 and Zippen 1956) with equal detection probabilities when multiple observers are used. We start with a single observer, whose detection ability is calibrated through a designed trial where the total number of objects is known. In this case $\text{var}(\hat{P}) = \frac{\hat{P}(1-\hat{P})}{m}$, where m is the number of objects used in the trial.

In a multiple observer setting, detection can be estimated directly from the survey data, eliminating the need for the independent trials used in the single observer case. If the individual detection probability of a zebra mussel is p **notation** (and the probability of failing to detect is $q = 1 - p$) then the probability of a detection by either observer is $P = 1 - q^2$. We can get $\text{var}(\hat{P})$ using the delta method, $\text{var}(\hat{P}) = 4q^2\text{var}(\hat{p})$. We use the large sample results of Zippin (1956) that gives $\text{var}(\hat{p}) = \frac{(1-\hat{p})(\hat{p}-2)^2}{X}$. We combine these to get

$$\begin{aligned}
\text{var}(\hat{P}) &= \frac{4(1-\hat{P})^{3/2}\left(2-\hat{P}+2\sqrt{1-\hat{P}}\right)}{X} \\
&\approx \frac{1}{X}\left(16 - 32\hat{P} + 17\hat{P}^2\right) && \text{for } \hat{P} < 1 \\
&\approx \frac{16}{X} && \text{for } \hat{P} \ll 1
\end{aligned}$$

So in the case that $\hat{P}$ is much less than 1, $\text{cv}(\hat{P})^2 = \frac{16}{Dan\hat{P}^3}$.

Figure 10 highlights how imprecision in the detection is much higher at low values of $\hat{P}$ while variation in counts doesn't change as fast. Overdispersion in counts will increase the amount of variation in counts, shifting the black curve in the above figure up. However, it would have to be very high to be comparable to the amount of variation in detection at low P.

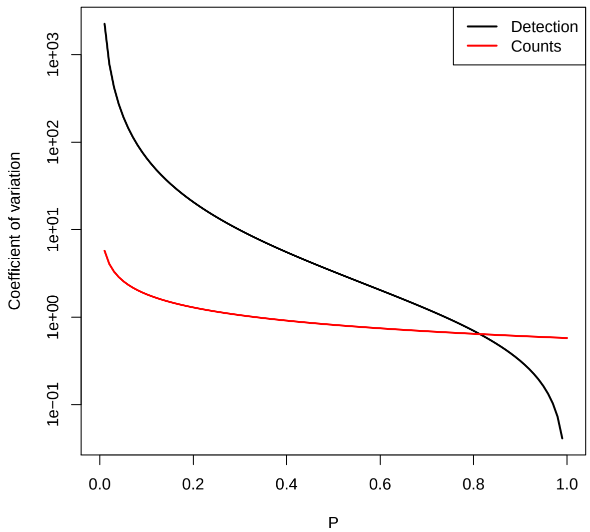


Figure 10: Coefficient(s) of variation.

The approximation for var(P) will likely be easier to work with than the full formula (Figure 11). In the range where our approximation to var(P) works, we end up with

$$\mathrm{cv}(\hat{D})^2 \approx \frac{1}{Dan\hat{P}} + \frac{1}{Dan}\left(\frac{16}{\hat{P}^3} - \frac{32}{\hat{P}^2} + \frac{17}{\hat{P}}\right).$$

At low values of detectability ($\hat{P} \ll 1$) we can get by with just approximating this as $16/Dan\hat{P}^3$, see figure below.

Incorporating time into the precision

From the section on the Time budget analysis, modeling the total time to conduct a single survey as $\tau = \tau_0 + \tau_S X_i$, seems reasonable. Here, τ_0 is the activities such as setting up the transect that doesn't vary with density or effort, τ_S is the stuff, such as search time, that does vary with density and effort, and X_i is a random variable of the counts in a transect. For now we will only condsider $E[X_i] = DaP$, the expected number of counts in a transect. We can use this equation to determine the number of transects that can be completed in a given chunk of time, τ_T, as $n = \lfloor \tau_T/(\tau_0 + \tau_S DaP) \rfloor \approx \tau_T/(\tau_0 + \tau_S DaP)$. **Here we are treating the travel time as a constant.**

Now we want to relate the search efficiency to the search time. We can model the detection probability of an observer as a monotonically increasing function of effort, a simple choice is the disc equation, $p = \tau_S/(\epsilon + \tau_S)$, where ϵ is the half-saturation constant. Larger values of ϵ correspond to less efficient search behavior. We want to find the value of τ_S that minimizes the estimator precision by plugging in our equations for $P(\tau_S)$ and $n(\tau_S)$ into cv$(\hat{D})$, then figuring out the value of τ_S that minimizes this objective function.Another option would be exponential saturation, $1 - e^{-\varepsilon\tau_S}$ but we won't look at this one.

A single-observer with calibrated detection

In the case of a single observer, the overall detection probability of a target is equal to the detection probability of a single observer ($P = p$). Under a Poisson distribution of targets, $\mathrm{cv}(\hat{D}) = \frac{1}{Dan\hat{P}} + \frac{(1-\hat{P})}{\hat{P}m}$.

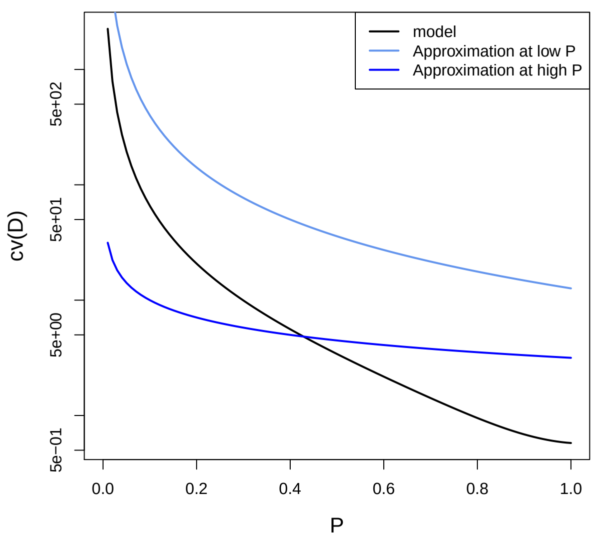


Figure 11: Testing approximations to the variance.

Case 1: high P In the case that $\mathrm{cv}(\hat{P})$ is very small (when $P \approx 1$) , the optimal search time τ_S^* is:

$$\frac{d}{d\tau_S}\mathrm{cv}(\hat{D})^2 = \frac{d}{d\tau_S}\frac{1}{DanP}$$

$$0 = \frac{d}{d\tau_S}\frac{\left(\tau_0 + Da\tau_S^2/(\epsilon + \tau_S)\right)(\tau_S + \varepsilon)}{Da\tau_T\tau_S}$$

$$0 = \frac{aD\tau_S^2 - \tau_0\epsilon}{Da\tau_T\tau_S^2}$$

$$\tau_S^* = \sqrt{\frac{\tau_0\epsilon}{aD}}$$

Thus, when the error in detection is low (corresponding to very high detection rates), the strategy should be to spend less time searching in higher density areas. The scaling constant $\tau_0\epsilon$ is an important factor that tells us as the time to do setup increases, or as search efficiency decreases we need to spend more time doing search but that these changes are not linear functions. Thus if we quadruple the time to do setup then we double to search time. Overdisperion will not affect these results because this affects the coefficient of variation as a constant term, that will dissapear when differentiating. We can also get the optimal number of transects to complete in transects with average density, $n^* = \frac{\tau_T}{\tau_0 + \tau_S^* aDP}$ **Get formula**. Plotting these quantities below for a transect strip of length 30 meters, with $\epsilon = 0.1414738$ (determined from $\hat{P} = 0.9513599$, averaged across lakes and $\tau_S \approx 0.5$ from analysis of search time for a 30 meter transet) and a total search time of 8 hours:

We see that the total number of transects is approximately indpendent of the setup time (surprising) and that the number that should be completed is high. The actual search times (Figure **??**) were much higher than what we would predict as optimal but those search times were also recorded for the double observer survey.

Case 2: the case of non-neglible detection Now, guess what, we let $\mathrm{cv}(\hat{P})$ be large!

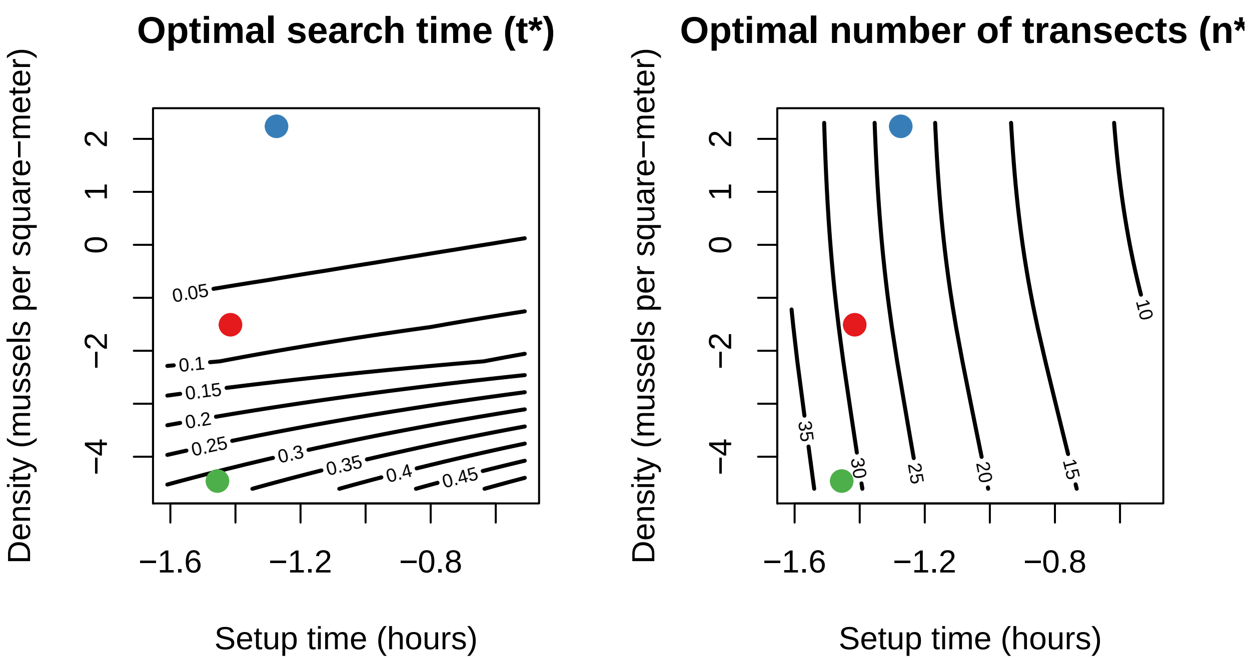


Figure 12: Optimal search time and number of transects in a single observer survey when the variability in detectability is negligible. fix axes.

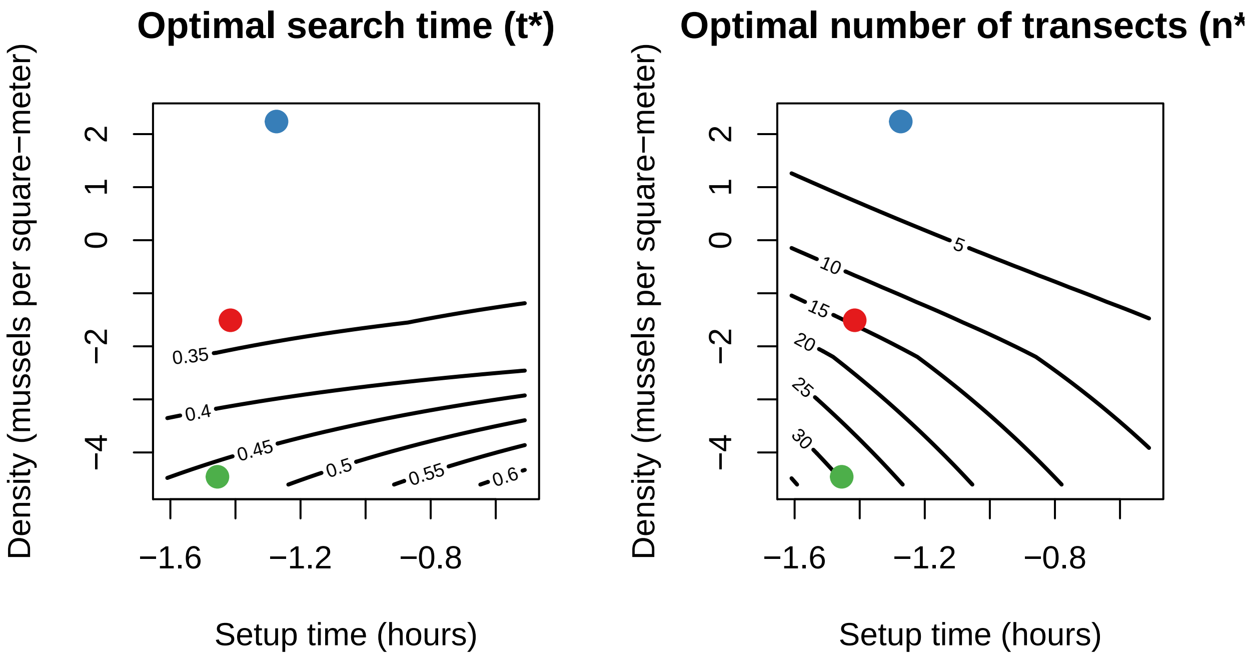


Figure 13: Optimal search time and number of transects in a single observer survey when the variability in detectability is not negligible. Single observer with the number of trial objects set to $m = 100$. Lower values of m lead shift the y-intercept of the contour lines down and change the curvature from concave to convex, maybe this is important…

$$
\begin{aligned}
\frac{d}{d\tau_S}\mathrm{cv}(\hat{D})^2 &= \frac{d}{d\tau_S}\frac{1}{DanP} + \frac{d}{d\tau_S}\frac{1-\hat{P}}{\hat{P}m} \\
0 &= \frac{d}{d\tau_S}\frac{(\tau_0 + Da\tau_S^2/(\epsilon+\tau_S))\,(\tau_S+\varepsilon)}{Da\tau_T\tau_S} + \frac{d}{d\tau_S}\frac{\epsilon}{m\tau_S} \\
0 &= \frac{aD\tau_S^2 - \tau_0\epsilon}{Da\tau_T\tau_S^2} - \frac{\epsilon}{m\tau_S^2} \\
\tau_S^* &= \sqrt{\frac{\tau_0\epsilon}{aD} + \frac{\tau_T\epsilon}{m}} \qquad (3)
\end{aligned}
$$

Figure **??** plots the solution.

This figure highlights that when there is uncertainty in the detection in detection we should go much slower. The optimal search time is much more sensitive to variation in density.

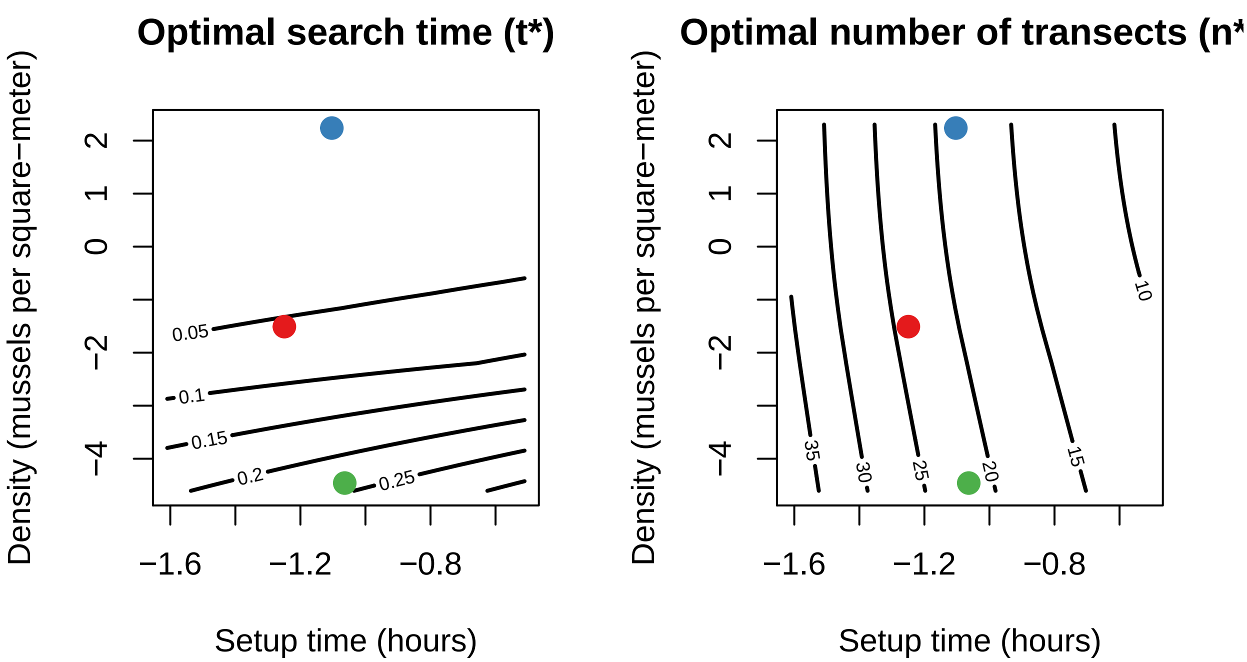
Double observer design

In the case where we have a double observer survey and estimate detection as a part of the design, the detection of a zebra mussel by either observer is now $P = 2p - p^2 = \frac{2\tau_S\epsilon+\tau_S^2}{(\tau_S+\epsilon)^2}$.

Case 1: high P In the case that $\mathrm{cv}(\hat{P})$ is very small (when $P \approx 1$) , the optimal search time τ_S^* is:

$$
\begin{aligned}
\frac{d}{d\tau_S}\mathrm{cv}(\hat{D})^2 =& \frac{d}{d\tau_S}\frac{1}{DanP} \\
0 =& \frac{d}{d\tau_S}\frac{(\tau_0(\tau_S+\epsilon)^2 + Da\tau_S(2\tau_S\epsilon+\tau_S^2)}{Da\tau_T(2\tau_S\epsilon+\tau_S^2)} \\
0 =& aD(\tau_S^2+2\tau_S\epsilon)(\tau_S^2+\tau_S\epsilon) + aD\tau_S(2\tau_S+2\epsilon)(\tau_S^2+\tau_S\epsilon) + 2\tau_0(\tau_S+\epsilon)(\tau_S^2+\tau_S\epsilon) - \\
& aD\tau_S^2(\tau_S+2\epsilon)(2\tau_S+\epsilon) - \tau_0(\tau_S+\epsilon)^2(2\tau_S+\epsilon) \\
0 =& \ aD\tau_S^2(\tau_S+2\epsilon)(\tau_S+\epsilon) + 2aD\tau_S^2(2\tau_S+2\epsilon)^2 + \tau_0(\tau_S+\epsilon)^2(2\tau_S+\epsilon)^2(1-2\tau_S+ \\
0 =& \ aD\tau_S^2(\tau_S^2+4\epsilon\tau_S^2+2\epsilon^2) + \epsilon\tau_0(\tau_S^2+2\epsilon\tau_S+\epsilon^2) \\
\tau_S^* =& -\epsilon/2 + 1/2\sqrt{-2\epsilon\frac{\epsilon^2-\epsilon\tau_0-2aD\epsilon^2}{\sqrt{(aD\epsilon+\tau_0)\epsilon aD}} - \epsilon\frac{2aD\epsilon-\tau_0}{aD}} + \\
& 1/2\sqrt{\epsilon\frac{aD\epsilon+\tau_0}{aD}}
\end{aligned}
$$

Quartics aren't easy to deal with. The above root is the only one that is positive and real but doesn't provide a whole lot of intuition on what is going on.As we found previously the solution is independent to the total time budget τ_T. At high densities the solution is $\tau_S \approx 1/2\sqrt{\epsilon\frac{aD\epsilon+\tau_0}{aD}}$.



The results for the double observer design show that they are qualitately similar to the single observer design (Figure **??**), with search times that are much slower than in the single observer design.

Case 2: the case of non-neglible detection, low $\hat{P}$

Here we use the approximation of low P in the double observer survey.

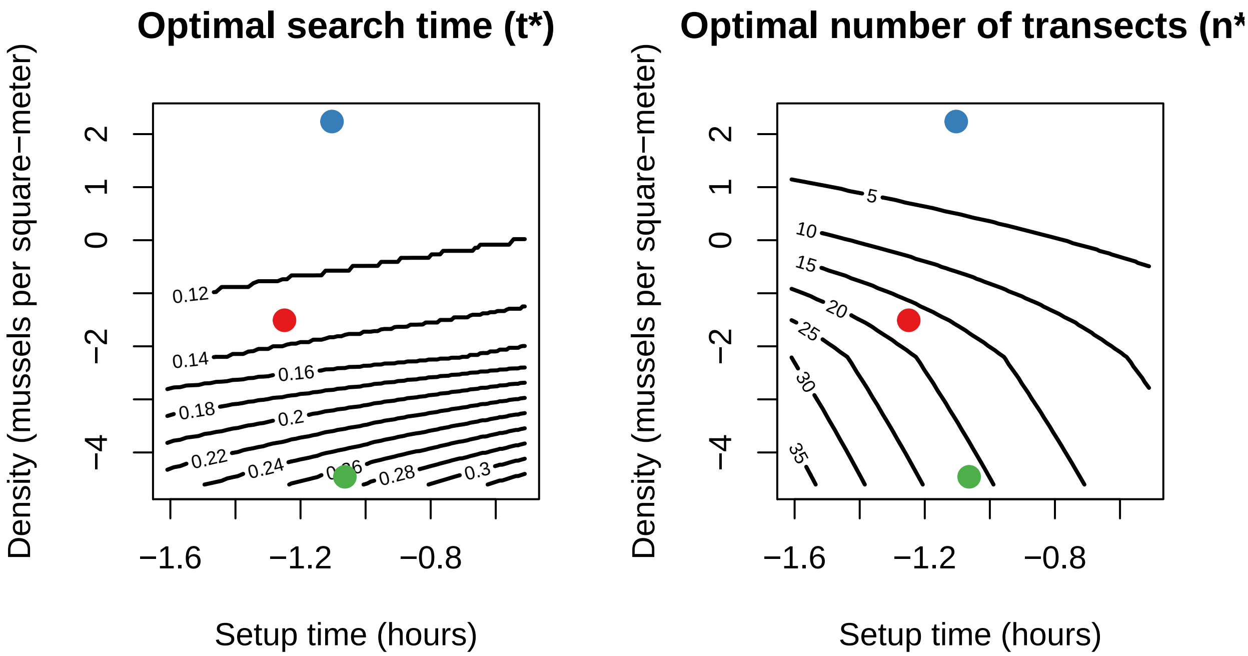


Figure 14: Numerical estimates for the optimal solution (using a simple grid search).

$$
\begin{aligned}
\frac{d}{d\tau_S}\mathrm{cv}(\hat{D})^2 &= \frac{d}{d\tau_S}\frac{16}{DanP^3} \\
&= \frac{d}{d\tau_S}\frac{16(\tau_0 + \tau_S DaP)}{Da\tau_T P^3} \\
&= \frac{d}{d\tau_S}\frac{16(\tau_0 + \tau_S Da(2\tau_S\epsilon + \tau_S^2))(\tau_S + \epsilon)^6}{Da\tau_T(\tau_S + \epsilon)^2(2\tau_S\epsilon + \tau_S^2)^3} \\
0 &= Da\tau_S^5 + 5Da\epsilon\tau_S^4 + 4Da\epsilon^2\tau_S^3 - 2(2Da\epsilon^3 + \tau_0)\tau_S^2 - 4\epsilon\tau_0\tau_S - 6\epsilon^2\tau_0
\end{aligned}
$$

This is a fifth order polynomial with no general closed form solution (i think). So we will get the computational solution for this one.

Qualitatively the numerical solution for the double observer survey looks like the single observer survey, though the optimal search times are slightly lower, likely due to the improved detection associated with having two observers.

As above we see that the solutions are qualitively similar to the case of a single observer, but with slower search times. This suggests we can take the results for a single observer study but we need to figure out what the effective value of ϵ is for the double observer model. The interpretion of ϵ in the single observer model is where the detection probability $p = 1/2$. We can calculate the effective half-width, call it ε as the value where $P = 1/2$.

$$
\begin{aligned}
P = 1/2 &= \frac{2\varepsilon\epsilon + \varepsilon^2}{(\varepsilon + \epsilon)^2} \\
\varepsilon &= \frac{\epsilon}{1 + \sqrt{2}}.
\end{aligned}
$$

In figure **??** we compare the predictions. Fro this initial figure it looks as if plugging in the effective half-width,

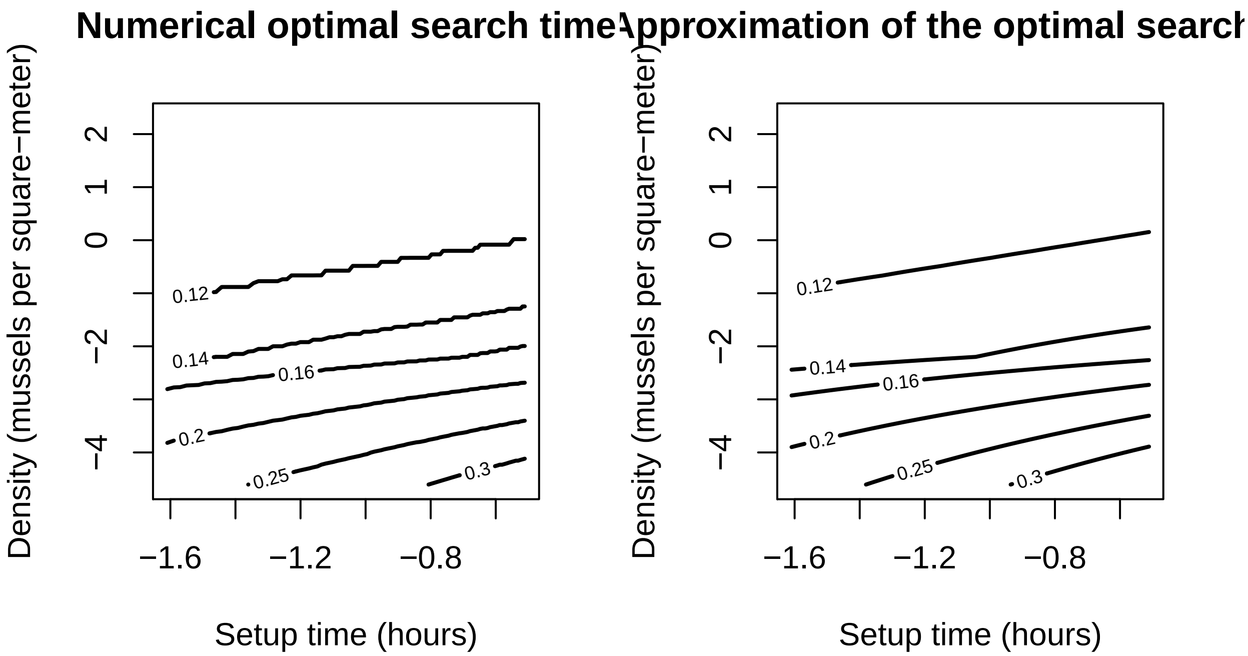


Figure 15: Approximation to double observer. Plugged in m=35 into equation

ε, into the solution for the single observer model equation (3) $\left(\tau_S^* = \sqrt{\frac{\tau_0 \epsilon}{aD} + \frac{\tau_T \epsilon}{m}}\right)$ qualitativly captures the actual solution.

Comparison to empirical search times

We can now compare the model estimates to the actual search times in each lake to determine how closely our divers came to achieving optimality.

Summary

- Theory tells that all else being equal, faster surveying methods are optimal as density increases.
- Our results for a single observer qualitavely match the double observer results in terms of optimal search times and optimal number of transects.
- Our searchers innate behavior is the opposite of what theory predicts.
-

Borchers, D. L., J. L. Laake, C. Southwell, and C. G. M. Paxton. 2006. "Accommodating unmodeled heterogeneity in double-observer distance sampling surveys." *Biometrics* 62 (2): 372–78. https://doi.org/10.1111/j.1541-0420.2005.00493.x.

Buckland, S. T., D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas. 2001. *Introduction to distance sampling: estimating abundance of biological populations.* 1st ed. Oxford: Oxford University Press.

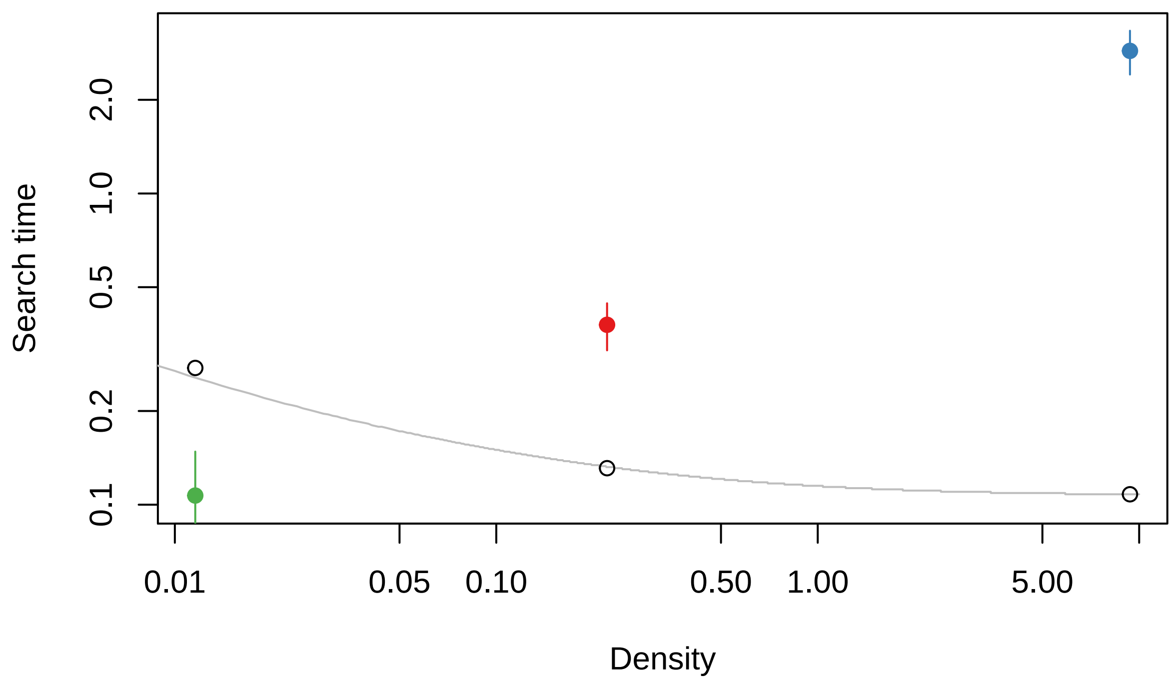


Figure 16: Colored dots are the predicted observer search times for a 30 m transect in the three lakes, open circles are predicted values for each lake, grey line is the prediction for the average setup time across all lakes.

Cook, R D, and J O Jacobson. 1979. “A design for estimating visibility bias in aerial surveys.” *Biometrics* 35 (4): 735–42. https://doi.org/10.2307/2530104.

Ferguson, Jake M., Jessica B. Langebrake, Vincent L. Cannataro, Andres J. Garcia, Elizabeth A. Hamman, Maia Martcheva, and Craig W. Osenberg. 2014. “Optimal Sampling Strategies for Detecting Zoonotic Disease Epidemics.” *PLoS Computational Biology* 10 (6): 1–26. https://doi.org/10.1371/journal.pcbi.1003668.

Holden, Matthew H., Jan P. Nyrop, and Stephen P. Ellner. 2016. “The economic benefit of time-varying surveillance effort for invasive species management.” *Journal of Applied Ecology* 53 (3): 712–21. https://doi.org/10.1111/1365-2664.12617.

Laake, Jeff, David Borchers, Len Thomas, David Miller, and Jon Bishop. 2018. “mrds: Mark-Recapture Distance Sampling.”