Search model

The criterion to optimize is the variance of the estimated density using standard form this is:

$$\frac{\mathrm{var}(\hat{D})}{\hat{D}^2} = \frac{\mathrm{var}(X|\hat{P})}{E[X]^2} + \frac{\mathrm{var}(\hat{P}|X)}{E[\hat{P}]^2}$$

The quantities important ot know are X, the total number of zebra mussels counted and P, the detection probability of each zebra mussel. We conduct n transects so $X = \sum_{i=1}^{n} X_i$, where n is actually a function of X, as described below. This dependence is because the n depends on the amount of time available, which depends on how many detections have been made. This is described in more detail in the following section but this means that n is a random variable and that the variance of the total counts is $\mathrm{var}(X|\hat{P}) = E[n]\mathrm{var}(X_i|\hat{P}) + E[X_i|P]^2\mathrm{var}(n)$. **John do you agree with this expression (that we must account for variation in n)?** The transect-level counts and detection probability are related by $X_i \sim NegBin(\mu = DaP, \alpha)$, where D is the true zebra mussel density and a is the surveyed area in a transect. Below we discuss how to calculate $\mathrm{var}(n)$.

Incorporating a time-budget into the design

We want to model the search behavior of a diver looking for zebra mussels. We will follow the approach of Holling and partition the survey time into the time spent searching τ_S, the time spent recording each observation, τ_H, also called the handling time, and an additional quantity that determines the time needed to setup the transect, τ_0. Then, if X_i objects are detected on a transect the time-budget equation for a single transect is

$$\tau_0 + \tau_S + \tau_H X_i. \qquad (1)$$

This equation says that the the total time spent surveying a transect depends on the number of detections made. We will assume that τ_H is fixed but that τ_S can be varied by changing the swim velocity but at the cost of influencing the detection probability. The total expected time to conduct a transect is then $E[\tau_T] = \tau_0 + \tau_S + \tau_H E[X_i]$ and $\mathrm{var}(\tau_T) = \tau_H^2 \mathrm{var}[X_i]$. The total number of transects that can be completed in the fixed amount of time, T hours, is $n = \lfloor T/\tau_T \rfloor \approx T/\tau_T$ with $E[n] = \frac{T}{\tau_0+\tau_S+\tau_H E[X_i]}$ and $\mathrm{var}[n] = \frac{T^2\tau_H^2\mathrm{var}(X_i)}{(\tau_0+\tau_S+\tau_H E[X_i])^2}$.

For a model of detection by a single observer we use

$$p = \frac{\tau_S}{\beta + \tau_S}, \tag{2}$$

where β is a measure of the searcher inefficiency where higher values correspond to searchers taking longer to have the same detection probability. An alternative is to use $p = 1 - e^{-\beta\tau_S}$, a model that has been used in optimal foraging theory. However the qualitative behavior of the two models is nearly identical and our choice is slightly easier to work with. Both functions fit the marshmallow search data about the same (*insert image here*).

We can use the time-budget to determine the number of transects that can be completed in a given chunk of time, τ_T, as $n = \lfloor \tau_T/(\tau_0 + \tau_S + \tau_H E[X_i]) \rfloor \approx \tau_T/(\tau_0 + \tau_S + \tau_H E[X_i])$. Because the counts vary, the number of transects also varies. This quantity can be approximated with the delta method as $\mathrm{var}(n) \approx \mathrm{var}(X_i)\left(\frac{\tau_T\tau_H}{(\tau_0+\tau_S+\tau_H E[X_i])^2}\right)^2 = \mathrm{var}(X_i)E(n)^4\tau_H^2/\tau_T^2$.

We will consider the total counts in each transect, X_i as a random variable arising from the negative binomial distribution with estimated rate $\lambda = DaP$, then $\mathrm{cv}(X)^2 = \frac{E(n)\mathrm{var}(X_i)+E(X_i)^2\mathrm{var}(X_i)E(n)^4\tau_H^2/\tau_T^2}{E(n)^2E(X_i)^2} = \frac{\mathrm{var}(X_i)\left(1+E(X_i)^2E(n)^3\tau_H^2/\tau_T^2\right)}{E(n)E(X_i)^2}$, where α is the overdispersion. We will assume that the detections can be modeled as binomial random variables with equal detection probabilities when multiple observers are used.

A single-observer with known detection

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 $\mathrm{var}(\hat{P}) = \frac{\hat{P}(1-\hat{P})}{m}$, where m is the number of objects used in the trial to estimate P. The overall detection probability of a target is the detection probability of a single observer ($P = p$), determined using the detectability that was independently esitmated using a sample size of m.

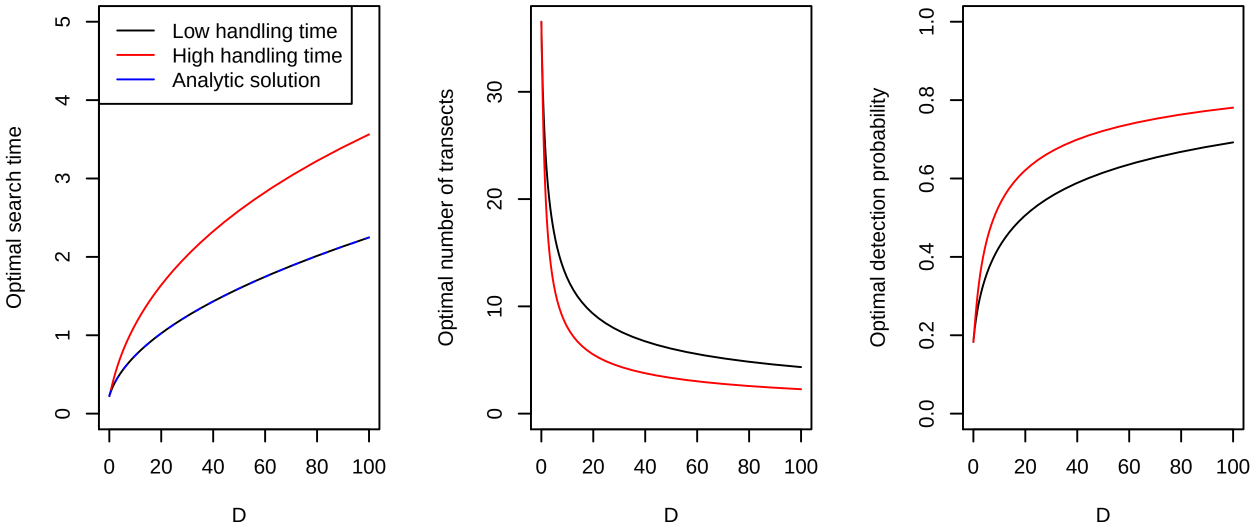

Figure 1: Comparison of numerical estimates and an analytical approximation for the optimal solution. $T = 10$, tau_0=0.05, beta=1, alpha=0, m=100. The high handling time is th=0.01.

Analytical approximations

Case 1: $\tau_H \ll T$, $\alpha = 0$

Now we take a look at what happens when the handling time is much less than the total survey time, T, for example $\tau_H \approx 0.001T$. In a 10 hour survey period, this corresponds to each recording taking about 30 seconds. In this case, the contribution of $\text{Var}(N)$ will drop out. In addition, we will assume that there is nospatial heteregenity, $\alpha = 0$. For readability, I will drop the $\hat{P}$ notation, and just call the estimate P. In this case we expect that the contribution due to We can solve for the optimal search time, τ_S^* as:

$$
\begin{aligned}
\frac{d}{d\tau_S}\text{cv}(\hat{D})^2 &= \frac{d}{d\tau_S}\left[\frac{\text{Var}(X_i)}{E[n]E[X_i^2]} + \frac{1-P}{Pm}\right] \\
0 &= \frac{n'P + P'n}{Dan^2P^2} + \frac{P'}{Pm} - \frac{P'(1-P)}{P^2m} \\
\tau_S^* &= \sqrt{Da\beta T/m + \beta\tau_0}
\end{aligned}
$$

This gives the same solution at $D = 0$, and then increases with the root of D thereafter. There is no dependence on the handling time, τ_H, so the optimal strategy is indepedent of the ‘cost’ of making detections nor is there any dependence on the degree of clustering, α. Thus, the optimal strategy

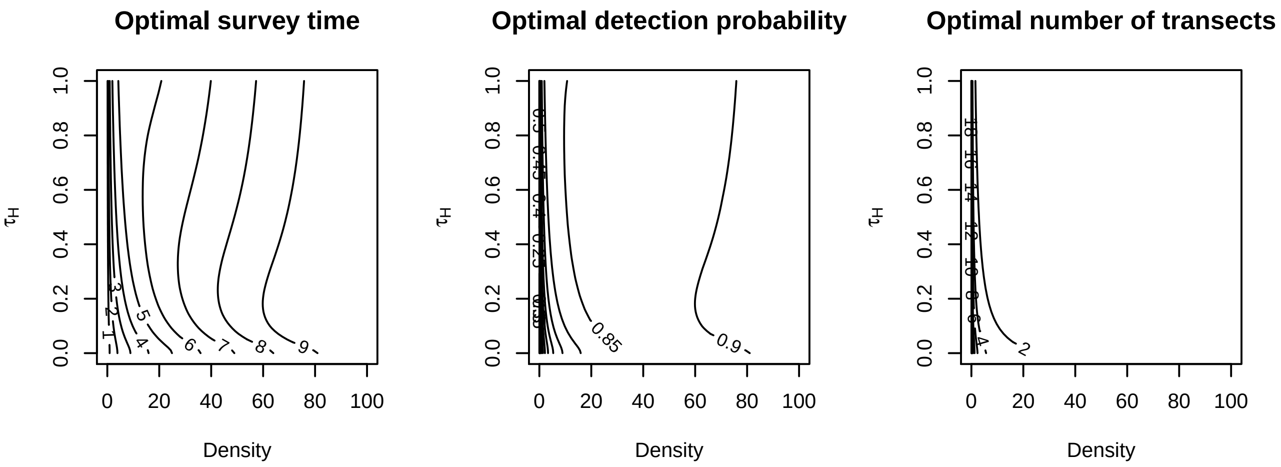

Figure 2: Exploring solutions for the single observer model, T=10, t0=0.1, D=0.1, m=10, alpha=0.

depends on the overall density but not on the spatial variation in that density. It is also important to note that the search strategy depends on the total amount of time spent on the survey, with search time increasing with the square-root of T. We also see that as precision in detection increases, the optimal strategy become a constant, $\sqrt{\beta\tau_0}$, that increases with the geometric mean of the search inefficiency and setup time. The solution is Figure **??**.

Case 2: τ_H is larger and $\alpha = 0$ ugly polynomials... no insights.

Case 3: $\alpha \neq 0$ ugly polynomials.. no insights. from simulations, α reduces the optimal search time.

Numerical solutions

Here are some simulations for when we can't get an analytical solutions. In Figure 2 we see that density and handling time interact in a complex way, though as the sample size m increases, the behavior becomes monotonic.

The non-monotonic behavior in Figure 2 arises at small sample sizes and very high densities. At larger sample sizes the solutions become monotonic.

There is also some bizarre behavior in Figure 3 at high densities. The survey time increases with density up to a point, then starts to decline. It's not clear to me if this behavior is due to any approximations we made (e.g., using the delta method), should explore by testing a higher order expansion. This only

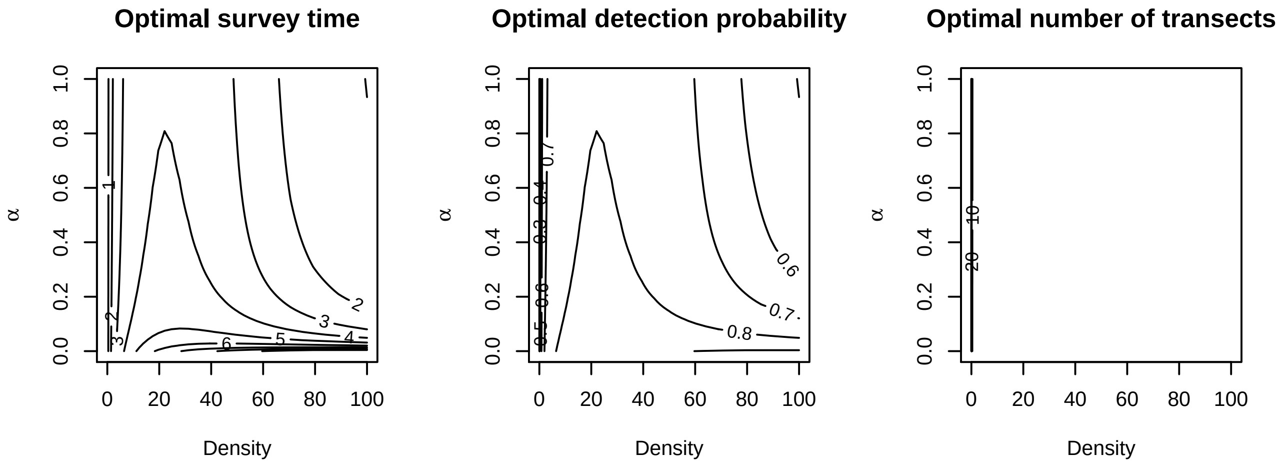

Figure 3: Exploring solutions for the single-observer model, T=10, t0=0.01, tH=0.2, beta=1, m=100

happens when the handling time is pretty high, otherwise the solutions look monotonic with density. Interestingly the number of transects is monotonic (Figure 4), even though the searchers optimal behavior is not. This is interesting, but probably a mathematical curiosity for now. It also looks like for intermediate densities, the optimal number of transects is a power-law, it's worth trying to figure out what this power depends on somehow, not sure how to do that yet...

Double observer survey

Here we consider two divers on each transect, this allows us to jointly estimate density and uncertainty in detection. The first diver proceeds as if they were a single diver but the second diver proceeds by only looking for zebra mussels that were missed by the first observer. If the detection probability for each diver 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, $\text{var}(\hat{P}) \approx 4q^2 \text{var}(\hat{p}) = 4(1-p)^3 p/X$. The coefficient of variation is then

$$\text{cv}(\hat{P}) \approx \frac{4(1-p)^3 p}{DanP^3} \tag{3}$$

This solution to the optimal search time is a nasty 9th order polynomial. Not very informative.

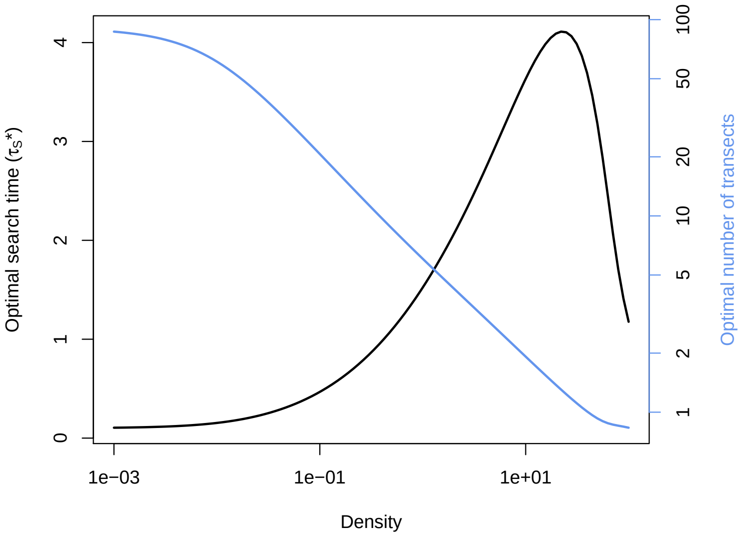

Figure 4: Slice from the above contour panel illustrating the monotonicity of the optimal search time, and the mononotic function of the optimal number of transects

Numerical simulations

The solutions for the double observer survey in Figures 5 and 6 indicate a complex relationship between the handling time and density. The relationship between density and overdispersion (Figure 6) is much simpler than the single-observer relationship, though this may depend on the parameterization.

Connecting to zebra mussel surveys

From the empirical study we estimated τ_H, τ_0, and τ_S for each survey type (Distance, double-observer, and quadrat). I am using the average values of the these estimates across lakes to look at when it would be best to switch sampling behaviors. We don't know β, the search inefficiency but we do have esimates of a divers detection probability, $\hat{p}$, so we will determine the β that is most consistent with this value by solveing $p = \tau_S/(\beta + \tau_S)$. Right now I am using the variance formula of the double-observer formula in the distance survey (equation 3) so will need to update that but I expect what we have is conservative.

We will not consider travel time here, but we previously found that travel time was only weakly related to distance in our surveys so it's probably ok to ignore it here.

Here are the parameters used in each survey type. In addition, we used the overdispersion estimate, α,

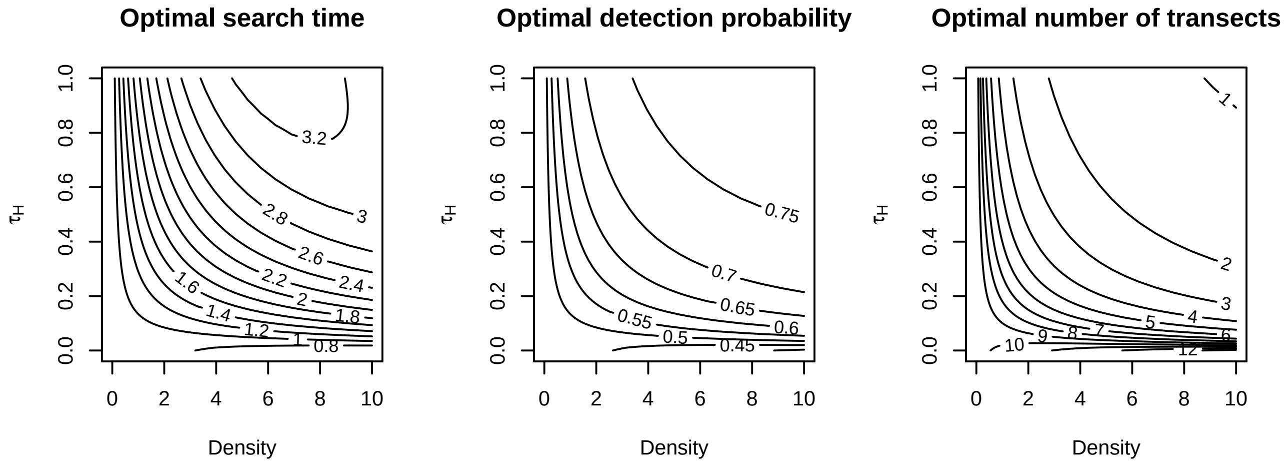

Figure 5: Numerical estimates for the optimal solution with the double observer methodology when alpha=0.1, tT=10, t0=0.1, beta=1

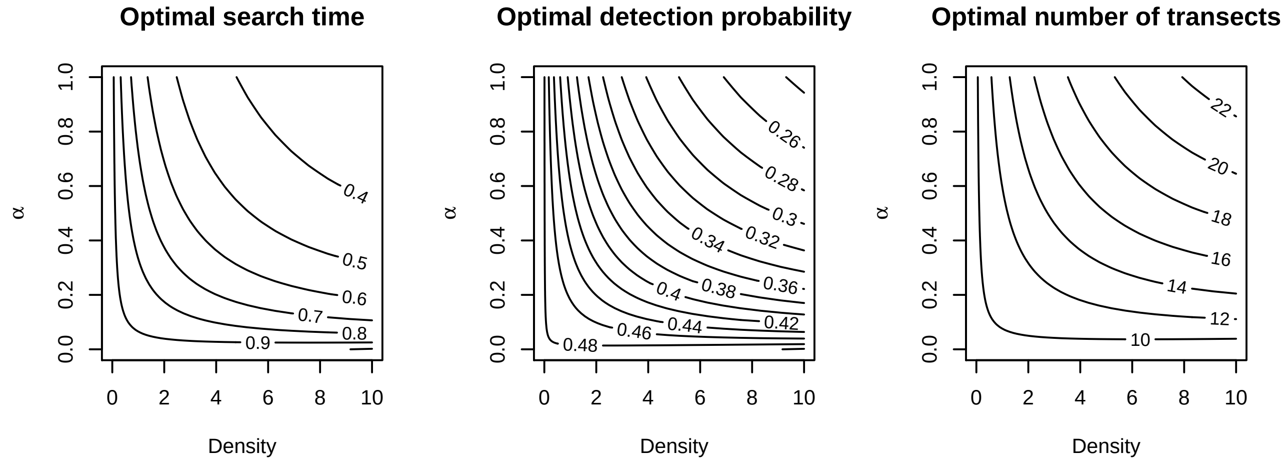

Figure 6: Numerical estimates for the optimal solution with the double observer methodology tH=0.01, tT=10, t0=0.1, beta=1

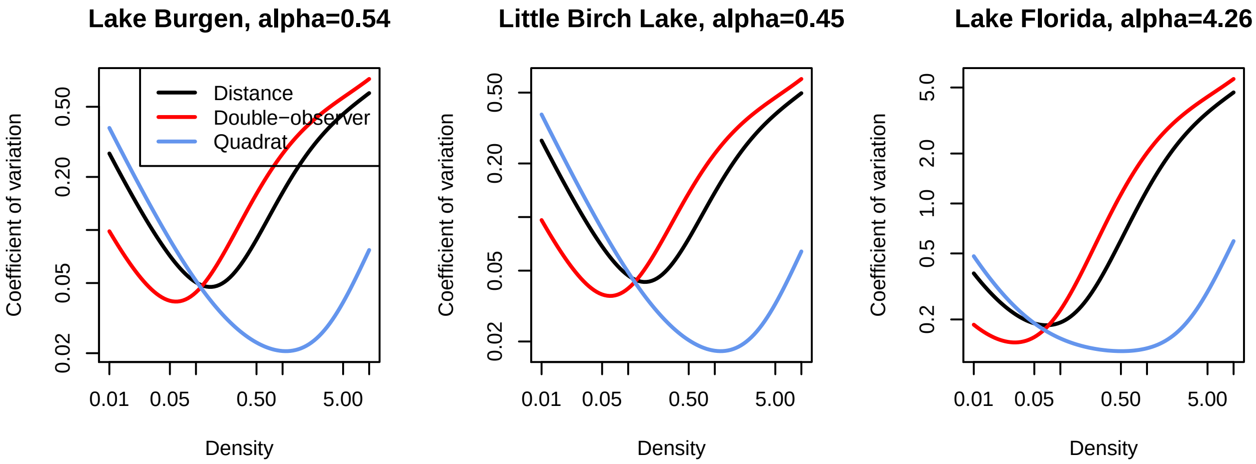

Figure 7: The predicted coefficient of varation, across a range of densities, for overdispersion lakes corresponding to each lake and empirical estimates of the setup time, handling time, and search time.

averaged across estimates of α from the data collected in each survey design.

Distance survey:

$a = 60$, $\hat{\tau}_H = 0.009$, $\hat{\tau}_0 = 0.08$, $\hat{p} = 0.25$. Estimated $\hat{\tau}_S = 0.19$, corresponding to $\hat{\beta} = 0.56$

We will assume that the variance of the distance survey can be modeled using the double observer formula… for now.

Double survey: $a = 30$, $\hat{\tau}_H = 0.014$, $\hat{\tau}_0 = 0.14$, $\hat{p} = 0.78$, $\beta = 0.04$. Estimated $\hat{\tau}_S = 0.13$,, correpsonding $\hat{\beta} = 0.04$

Quadrat survey: $a = 3.75$, $\hat{\tau}_H = 0.002$, $\hat{\tau}_0 = 0.12$, $\hat{p} = 1$, $\beta =$, $T = 10$. Estimated $\hat{\tau}_S = 0.17$

We see that the general pattern (Figure 7) is for double observer and distance sampling methods to be more efficent at low densities and quadrat surveys at high densities. This is broadly consistent with what we found empirically, though the coefficient of variation of the distance survey appears to be consistently higher than we found in the data, where we had variances in distance estimates as less than half of the quadrat surveys. Could be due to the large sample variance approximation of $\hat{P}$, or something else… but maybe not worth sweating over if the qualitative conclusions hold.