To develop an eDNA sampling effort that is robust to detecting low abundance pike populations, we relied on the estimated mean detection probabilities of northern pike eDNA reported in Dunker (2016). The detection probabilities were estimated from results of replicate samples collected at 1, 10, and 40 meters from a single, caged, live northern pike and were estimated to be .89, .57, and .27 respectively. Samples will be collected in duplicate to account for lower detection probabilities using a Biomeme two3 device vs. traditional benchtop sampling.

The following calculations will be used to estimate how many eDNA samples are needed to detect a small northern pike population with a desired probability of detection provided lake acreage is known and no gillnet sampling occurs. Calculations will be based on the assumptions that (1) fish are randomly distributed throughout the sampling area, (2) there are no false detections, and (3) the probability of detection beyond 40 meters is zero, since no estimates are available for this region. We will decompose the 40-meter circle around each sample site into three distinct sub-regions to account for differences in the probability of detection observed in Dunker (2016). These sub-regions will be the circular areas less than 1 meter, between 1 and 10 meters, and between 10 and 40 meters from the sample site. Since the estimated probabilities of detection from Dunker (2016) were for the upper end of these intervals, they will serve as conservative estimates of the probability of detection for the entire respective sub-regions. Because a pike can only be present in one sub-region, for a single site, we note by the law of total probability:

|  |  |
| --- | --- |
| P(D) = P(D and ) + P(D and ) + P(D and ) | (1) |

Where D is the event a pike is detected and is the event that a single pike is present in sub-region *i*. By the definition of conditional probabilities, equation (1) is equivalent to:

|  |  |
| --- | --- |
| P(D | )\* P() + P(D | )\*P() + P(D | )\* P() | (2) |

The probabilities of detection given a pike is present in the region (P(D | )) are taken as the estimates from Dunker (2016). By assumption (1), the probability a pike is present in a sub-region (P()) is the proportion of total area represented by that sub region. Finally, for S sample sites, we note that the total proportion of area represented by a sub-region is now S\*P(), and thus:

|  |  |
| --- | --- |
| P(DS) = P(D | ))\*S\* P()) + P(D | ))\*S\*P()) + P(D | ))\*S\* P()) = S\*P(D) | (3) |

Where DS is the event a pike is detected at one of S sites and is the event that a single pike is present in one of the S sub-region *i’s* for *i* = 1, 2, 3. We also assume the conditional probabilities in equations (2) and (3) are identical, meaning the conditional probability of detection given a pike is in a region is the same as assuming the pike is in one of several regions. Since the N pike are randomly distributed, the number of successful detections follows a Bin(N, S\*P(D)) distribution. The probability of at least one detection at S sites is . We then set this expression to the desired probability of detection and solve for S.