MURI hake sample selection

A. O. Shelton

2024-01-09

Table of contents

Initial Sample Selection (Feb. 2023) **1**
Samples from 2019 1
Identifying samples 3
Outstanding issues and notes 5

Second Sample Selection (Oct. 2023) **9**
New Scenarios (11/13/2023) 12
Sample Selection (1/9/2024) 12

Initial Sample Selection (Feb. 2023)

Samples from 2019

This is a short document identifying the samples identified for use in the first phase of the MURI project. These samples will be used in the initial application of the fish (MiFish) and marine mammal (D-LOOP) metabarcoding primers. We decided to begin with samples near the surface (3m and 50m depth) and to shoot for 45 individual samples at each depth as an exploratory set.

We can present all of the sample locations for 2019, 2021 and 2022 (Fig. 1).

Let's focus in on 2019 and focus in on sites north of 46.2 degrees latitude (off Washington state). There are 50 stations in this area (with 46 stations at the surface and 50 at 50m; Fig. 2). In total there are 192 water samples from these sites at 0 and 50m. Many of the surface samples were diluted for use with the hake qPCR (Fig. 2) with some stations having one of two replicates diluted and others having both replicates diluted.

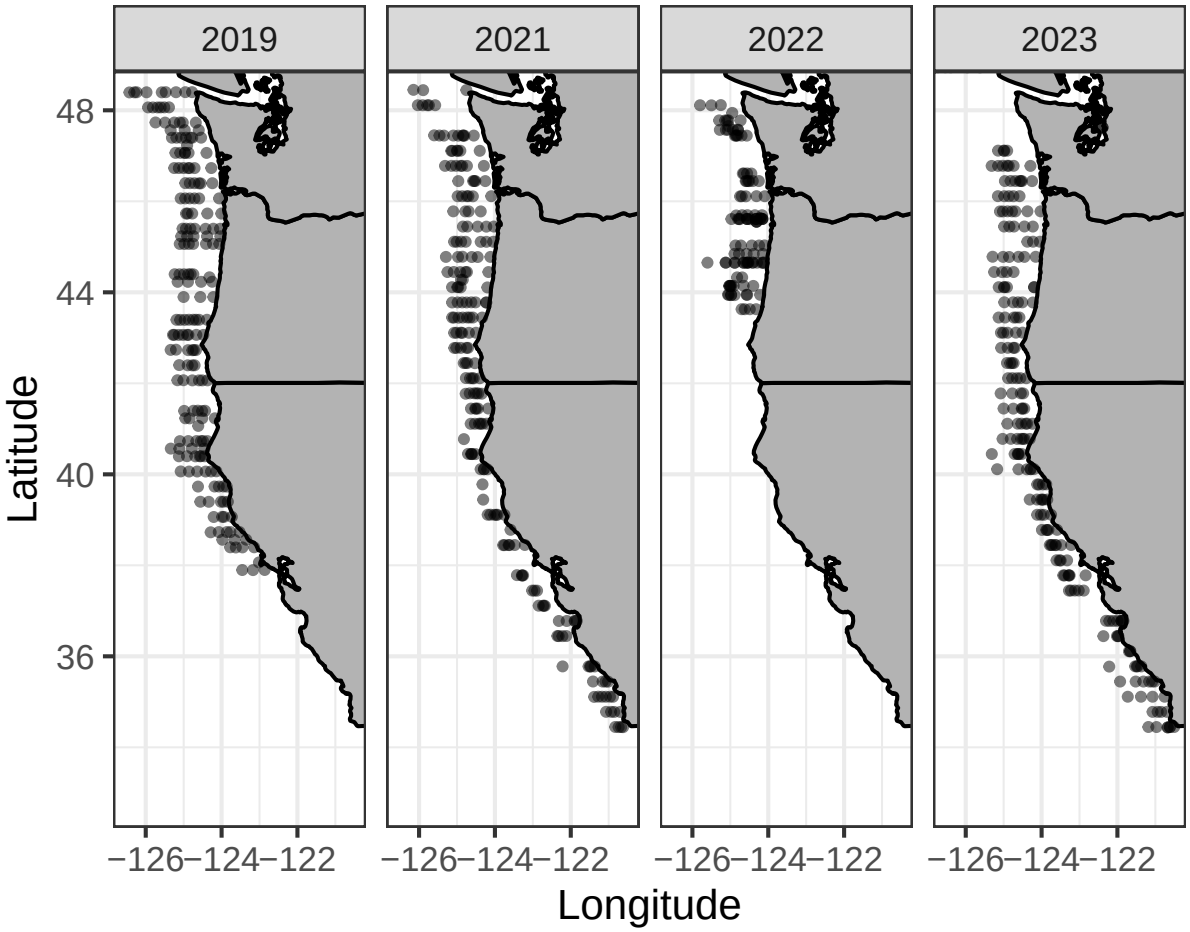

Figure 1: The locations of eDNA samples from three years of CTD collections from the hake acoustic-trawl survey.

From this set of samples, we are looking to include 90-ish total samples for preliminary analysis or approximately 45 from each depth. This leaves some space on a single plate for controls. To trim down from 192 individual water samples we will select a single water sample station which gets us to 96 samples, and leaves us to remove 6 samples to get to 90. Dropping one station from the middle of each line will get us there.

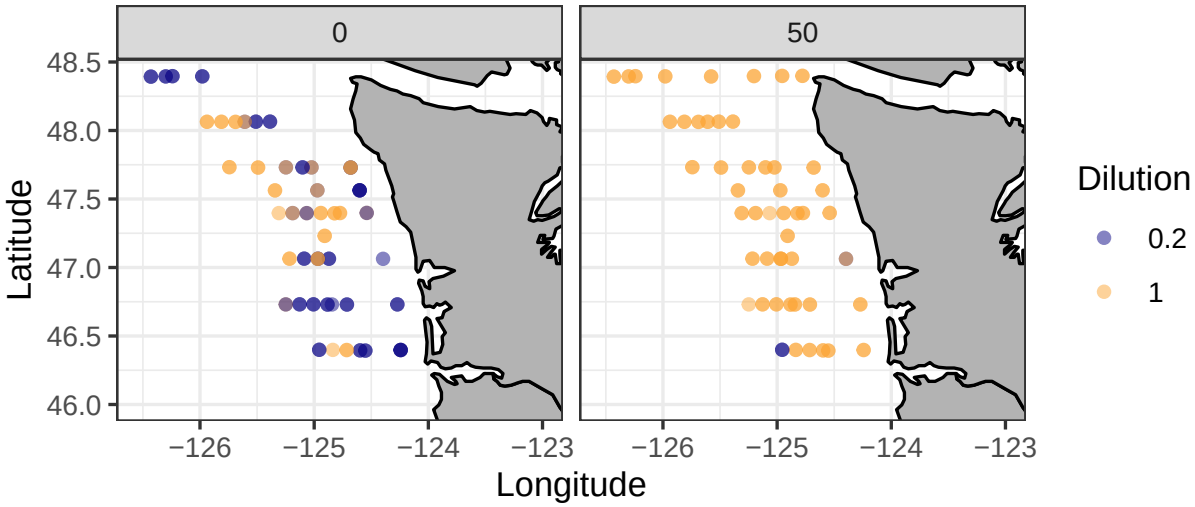

Figure 2: Samples from the 2019 hake cruise in Washington state waters at the surface and 50m depth. Colors indicate whether those samples were undiluted (1) or 1:5 diluted (0.2) in the qPCR included in hake analysis in 2019. Intermediate colors indicate that one of the replicate Niskin sample was diluted but the other was not.

The dropped lines are in the middle of each transect and most fall approximately along the shelf break, generally at either 300m or 500m bottom depth. Taking one sample from each station at each depth in Fig. 3 results in 82 samples (39 at the surface and 43 at 50m). Here is a plot of the chosen stations by depth (Fig. 4). The only other person that has used substantial numbers of the 2019 samples is Ramon. I have outline in red the samples that I believe Ramon used in his metabarcoding work. There is relatively limited overlap with the proposed set of stations and depths (Fig. 4).

Among samples that were diluted, there should be three levels of dilution - 1:2, 1:5, and 1:10. Only 1:5 dilution data were used in the 2021 hake paper, but the other dilutions should be available.

Identifying samples

Now that we have the stations identified (Fig. 4) we need to identify the sample to use at each station. I have been using two criteria to identify these samples: 1) We will use samples that have a lot of sample remaining, so that multiple kinds of analyses can be run on a single

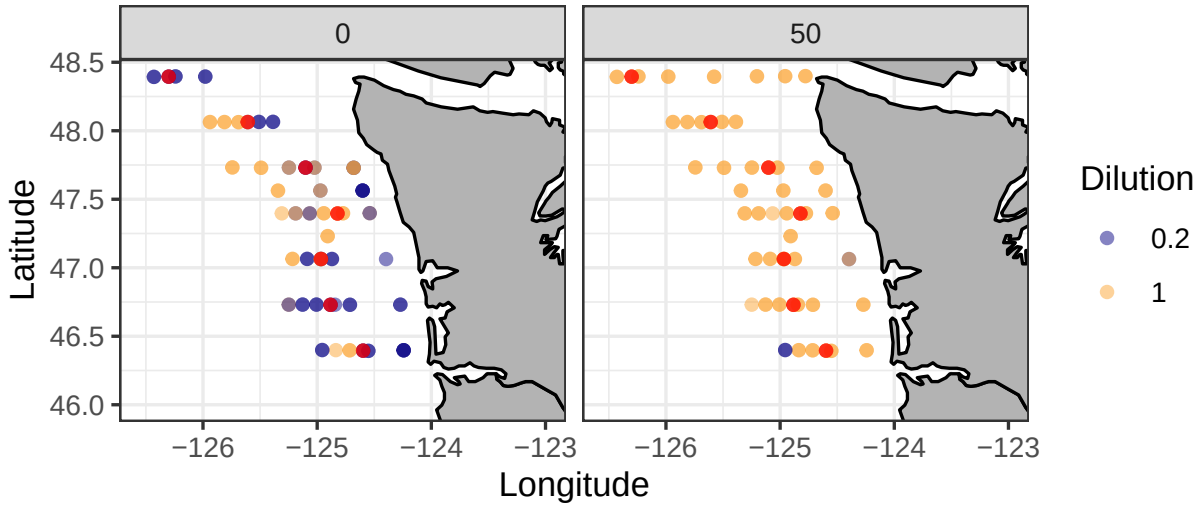

Figure 3: Samples from the 2019 hake cruise in Washington state waters at the surface and 50m depth. Colors indicate whether those samples were undiluted (1) or 1:5 diluted (0.2) in the qPCR included in hake analysis in 2019. Intermediate colors indicate that one of the replicate Niskin sample was diluted but the other was not. Red points identify stations proposed to be dropped

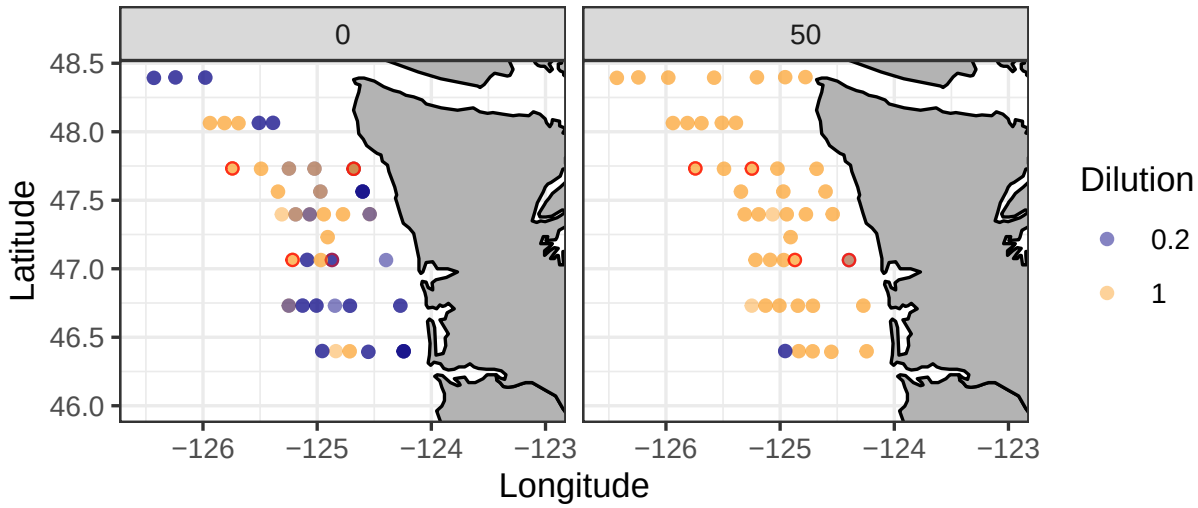

Figure 4: Stations from the 2019 hake cruise in Washington state waters at the surface and 50m depth. Colors indicate whether those samples were undiluted (1) or 1:5 diluted (0.2) in the qPCR included in hake analysis in 2019. Intermediate colors indicate that one of the replicate Niskin sample was diluted but the other was not. Excluded stations not shown. Samples used by Ramon are outlined in red.

sample; 2) We will prefer using undiluted to diluted samples from a given station because we are concerned about finding rare taxa.

This identification of samples that were used by Ramon is still tentative, but I plan on updating it as soon as I get the necessary info. The list is presented in Table 1. All of these samples putatitvely have more the $60\mu L$ available, this does include the amount of sample used during Ramon's analysis.

Abi has informed me that there are 6 additional samples from 2022 that were collected during times in which marine mammals were observed from the ship. I would suggest including those samples during the metabarcoding run as well.

Many of the surface samples were inhibited and diluted for use in the hake qPCR assay. It is not clear whether the diluted or undiluted make more sense to use in the metabarcoding analysis, so I propose using a subset of samples in these pilot runs to explore the value of running diluted vs. undiluted DNA extract. There are currently 18 samples in our set that are diluted. I propose using the 12 northern-most of these samples to run at both undiluted and 1:5 dilution in our first round of metabarcoding.

With these twelve samples included twice (once undiluted, once diluted 1:5), we are at 94 samples in total. Which is pretty close to the initial target of 90 samples. I print the samples of interest below, and in the csv ("2019_hake_samples_MURI_pilot_2023-02.csv"). Unfortunately, I do not have the location of the samples on the storage plates in this data frame, but they can be found in the file "Hake eDNA 2019 qPCR and results" in the "plates layout" tab in the google drive.

Outstanding issues and notes

1. Need to add information about other analyses that has used these samples to date, if any.

Table 1: This is a table of samples that seem like samples to target for the pilot application of the D-LOOP and MiFish primers. This includes 12 samples used twice (once undiluted, once diluted 1:5).

year	sample	station	depth_cat	dilution	vol_remain	ramon_16S
2019	1395	73-1	0	0.2	73.5	0
2019	1397	73-1	50	1.0	88.5	0
2019	1403	73-2	0	0.2	73.5	0
2019	1409	73-2	50	1.0	88.5	0
2019	1423	73-5	0	1.0	88.5	0
2019	1433	73-5	50	1.0	88.5	0
2019	1436	73-6	0	1.0	88.5	0
2019	1445	73-6	50	1.0	88.5	0
2019	1447	73-7	0	0.2	73.5	0
2019	1457	73-7	50	0.2	73.5	0
2019	1528	75-1	0	0.2	67.0	0
2019	1530	75-1	50	1.0	88.5	0
2019	1519	75-2	0	0.2	67.0	0
2019	1525	75-2	50	1.0	88.5	0
2019	1510	75-3	0	0.2	72.0	0
2019	1510	75-3	0	1.0	72.0	0
2019	1517	75-3	50	1.0	88.5	0
2019	1485	75-5	0	0.2	72.0	0
2019	1485	75-5	0	1.0	72.0	0
2019	1495	75-5	50	1.0	88.5	0
2019	1472	75-6	0	0.2	72.0	0
2019	1472	75-6	0	1.0	72.0	0
2019	1482	75-6	50	1.0	88.5	0
2019	1460	75-7	0	1.0	88.5	0
2019	1470	75-7	50	1.0	88.5	0
2019	1536	77-1	0	0.2	55.0	0
2019	1536	77-1	0	1.0	55.0	0
2019	1538	77-1	50	1.0	76.5	1
2019	1542	77-2	0	0.2	49.5	0
2019	1542	77-2	0	1.0	49.5	0
2019	1547	77-2	50	1.0	78.5	1
2019	1561	77-5	0	0.2	72.0	0
2019	1561	77-5	0	1.0	72.0	0
2019	1571	77-5	50	1.0	88.5	0
2019	1573	77-6	0	1.0	78.5	1
2019	1583	77-6	50	1.0	78.5	0
2019	1706	77-MT505	0	1.0	88.5	0

year	sample	station	depth_cat	dilution	vol_remain	ramon_16S
2019	1716	77-MT505	50	1.0	88.5	0
2019	1694	78-MT506	0	1.0	88.5	0
2019	1704	78-MT506	50	1.0	88.5	0
2019	1655	79-1	0	1.0	88.5	0
2019	1657	79-1	50	1.0	88.5	0
2019	1647	79-2	0	1.0	88.5	0
2019	1653	79-2	50	1.0	88.5	0
2019	1623	79-5	0	1.0	88.5	0
2019	1633	79-5	50	1.0	88.5	0
2019	1611	79-6	0	1.0	NA	0
2019	1621	79-6	50	1.0	NA	0
2019	1598	79-7	0	1.0	88.5	0
2019	1607	79-7	50	1.0	88.5	0
2019	1585	79-8	0	1.0	88.5	0
2019	1595	79-8	50	1.0	88.5	0
2019	1662	80-1	0	0.2	72.0	0
2019	1662	80-1	0	1.0	72.0	0
2019	1664	80-1	50	1.0	88.5	0
2019	1671	80-2	0	1.0	88.5	0
2019	1678	80-2	50	1.0	88.5	0
2019	1680	80-3	0	1.0	88.5	0
2019	1690	80-3	50	1.0	88.5	0
2019	1721	81-1	50	1.0	78.5	0
2019	1725	81-1	0	1.0	78.5	1
2019	1728	81-2	0	1.0	88.5	0
2019	1733	81-2	50	1.0	88.5	0
2019	1746	81-5	0	1.0	78.5	0
2019	1755	81-5	50	1.0	78.5	1
2019	1757	81-7	0	1.0	88.5	0
2019	1767	81-7	50	1.0	88.5	0
2019	1769	81-9	0	1.0	78.5	1
2019	1779	81-9	50	1.0	71.0	1
2019	1782	83-10	0	1.0	88.5	0
2019	1792	83-10	50	1.0	88.5	0
2019	1836	83-5	0	0.2	70.0	0
2019	1836	83-5	0	1.0	70.0	0
2019	1840	83-5	50	1.0	88.5	0
2019	1828	83-6	0	0.2	70.0	0
2019	1828	83-6	0	1.0	70.0	0
2019	1834	83-6	50	1.0	88.5	0
2019	1806	83-8	0	1.0	88.5	0

year	sample	station	depth_cat	dilution	vol_remain	ramon_16S
2019	1816	83-8	50	1.0	88.5	0
2019	1794	83-9	0	1.0	88.5	0
2019	1804	83-9	50	1.0	88.5	0
2019	1892	85-1	50	1.0	88.5	0
2019	1879	85-11	0	0.2	67.0	0
2019	1879	85-11	0	1.0	67.0	0
2019	1887	85-11	50	1.0	88.5	0
2019	1867	85-13	0	0.2	67.0	0
2019	1867	85-13	0	1.0	67.0	0
2019	1877	85-13	50	1.0	88.5	0
2019	1843	85-15	0	0.2	65.0	0
2019	1843	85-15	0	1.0	65.0	0
2019	1853	85-15	50	1.0	88.5	0
2019	1898	85-3	50	1.0	88.5	0
2019	1904	85-5	50	1.0	88.5	0
2019	1908	85-8	50	1.0	88.5	0

Second Sample Selection (Oct. 2023)

The above described the initial samples selected to examine with metabarcoding. They were run with the D-Loop, MiFish, and Ceph primers. MiFish (targeting fish) appeared to be work relatively well. D-Loop (targeting cetaceans) detected some cetaceans but also amplified a bunch of off-target species (predominantly bacteria). I haven't heard much about how the cephalopod primer worked.

Anyway, we are ready to start identifying a broader set of samples with the aim of immediately applying the MiFish primer to these samples and, after some fine tuning, to apply the D-LOOP primer to these samples as well. Based on a discussion with Kim and Megan, we proposed to identify further samples for analysis using the following criteria.

1. Focus on samples at a couple depths across spatial extent in a single year.
2. For a subset of stations, run samples across the range of available depths, just to have some depth profiles.
3. Use water from only one Niskin replicate at each station.

These are just the criteria talked about. If you have think other samples should be examined, great. Just let us know.

Based on the above criteria, and the fact that we have already run some of the 2019 samples, we're going to focus on just the 2019 samples. Here is a map of the 2019 samples by depth with the samples that have already been run highlighted.

And here are the sample locations that have not had any samples run as yet

Table 2: This is the number of distinct stations that remain unsampled at each depth for 2019.

Depth Category (m)	Distinct Stations
0	139
50	138
100	154
150	149
300	120
500	105

If we take one sample from each unsampled station at the surface, 50m, and 150m, that is 426 samples or about 4.5 plates (95ish wells per plate) of samples. That leaves about 50 spaces from other depths to get to 5 plates. If we did surface, 50m, and 300m, that is 397 samples, leaving about 80 spaces for other depths. As predominantly a fish person, I'd lean toward using the 150m depth, but I know there is some idea that beaked whales loiter in the

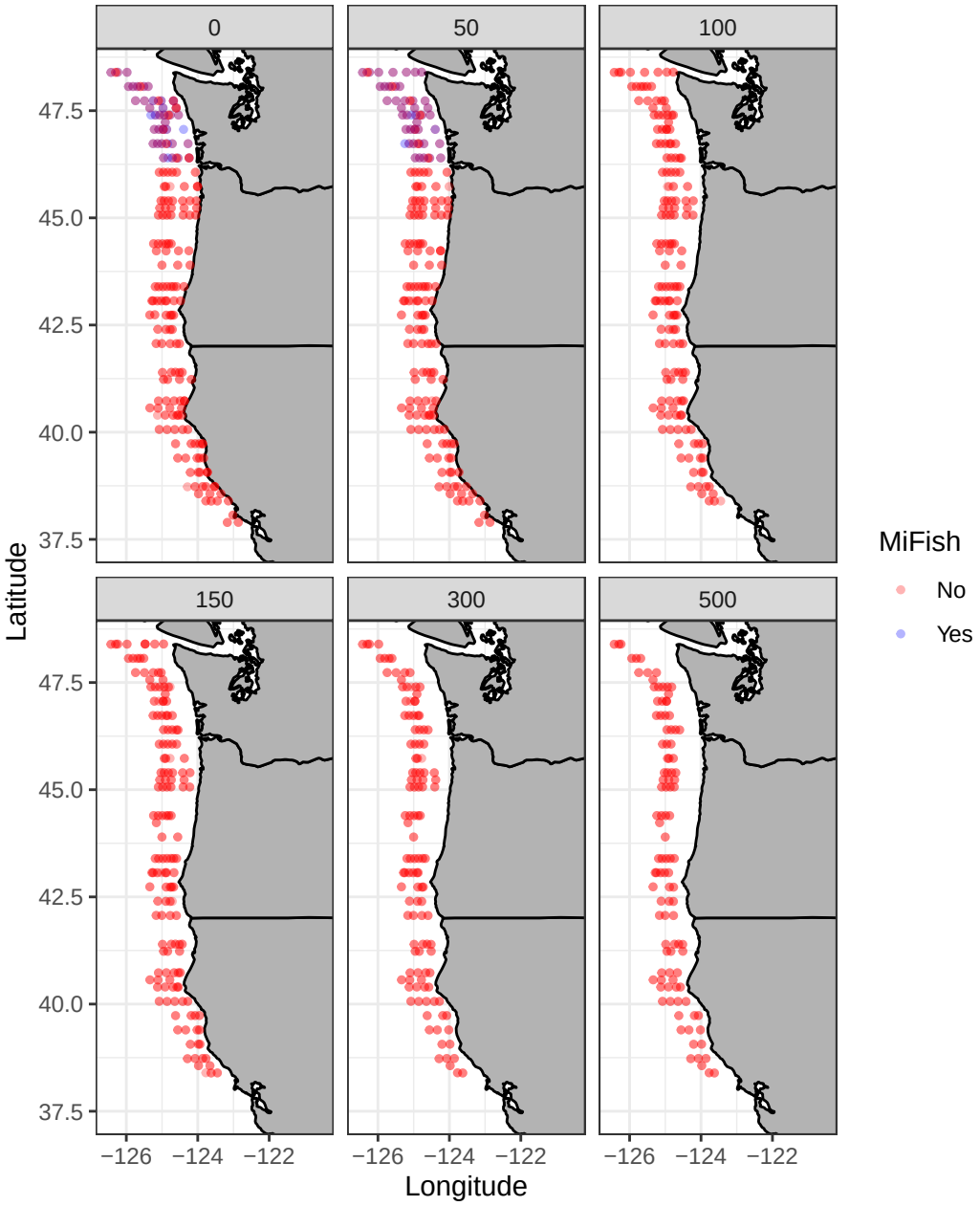

Figure 5: Location of samples faceted by sample depth and if a sample has already been run using the MiFish primer.

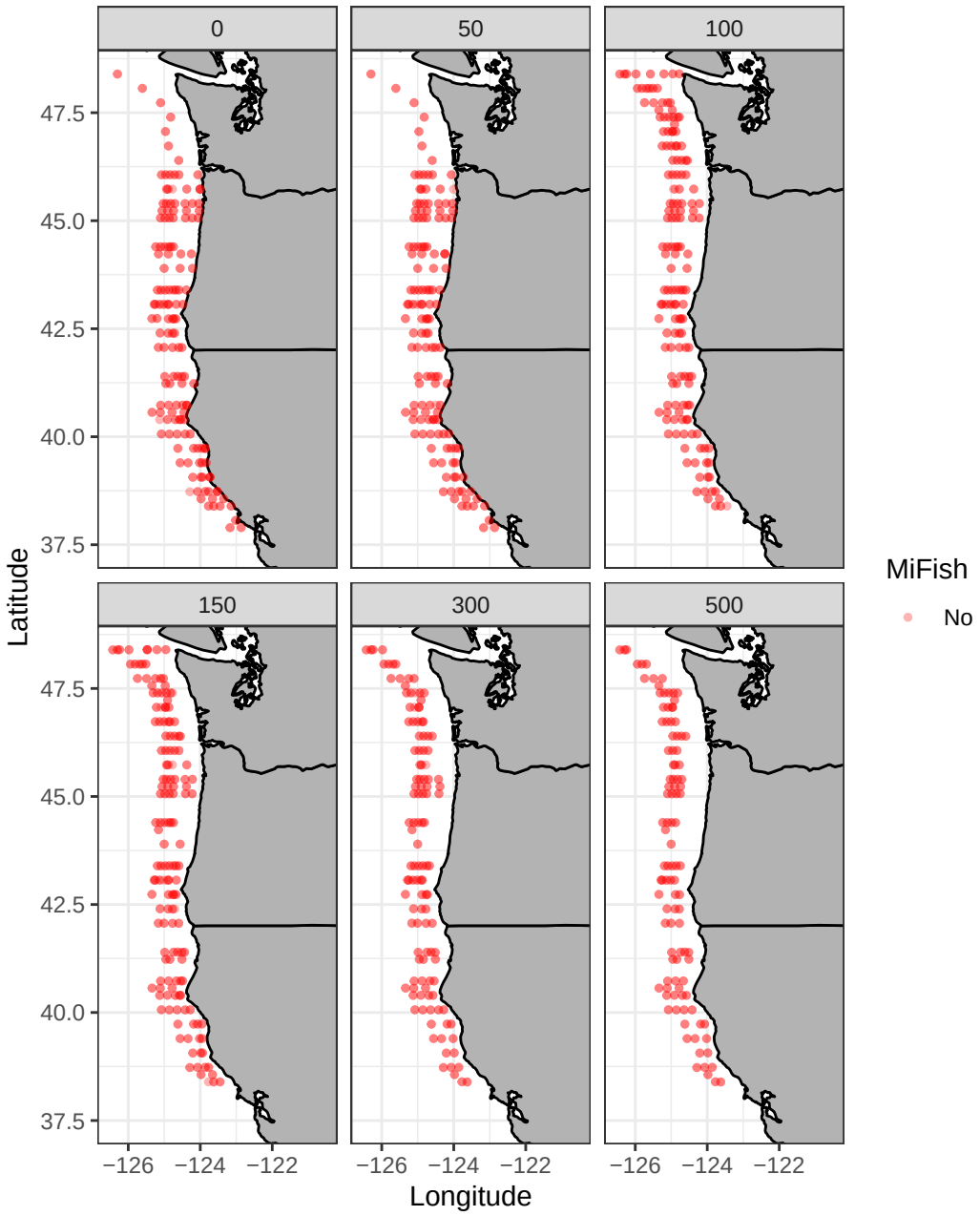

Figure 6: Location of samples faceted by sample depth. Only stations that have not been run shown.

200-300m depth range and so maybe we should do the 300m and about half (~75) of the 150m deep samples to get to 5 plates.

Below, I made two maps of potential samples & depths to run. These are just mock-ups, so if someone wants to propose other criteria for running these samples, please let me know. Both of the scenarios below have about 475 samples.

Questions:

1. Is there any reason we want to do fewer than 5 plates.
2. Are there economies of scale such that we should get all of the samples out at once (i.e. choose samples for each station-depth at one time even if we aren't going to run them immediately)?
3. Right now I haven't included any 500m deep samples. But we could add some if there was interest.

New Scenarios (11/13/2023)

After talking through the above maps with Kim, Ryan, Megan, and others, we decided to focus on the surface samples plus a set of samples where we can look at the full depth profile. So here are new maps based on the following criteria:

1. One biological replicate from each surface station. (139 samples)
2. One biological replicate from each 50m deep station EXCEPT those where the bottom depth is 50m (i.e. only include more offshore samples). (115 samples)
3. One biological replicate from a subset of station where there are samples from 150m, 300m and 500m bottom depths. This excludes samples from on the continental shelf / slope from water depths of between 150m and 300m. (~74 stations means 3×74 samples = 222 samples).
4. TOTAL SAMPLES = 476

In practice this means that we are sampling from about 70/% of the deep water sampling stations (74 of 105) Here's what that looks like on a map. I right now I have just randomly selected stations to include. If you think an area of spatial interest has been missed, lemme know and we can manually select different locations that seem important (Fig. 9).

Sample Selection (1/9/2024)

OK. In looking through the samples, Megan found that some of the identified samples did not have enough fluid left to run for the MiFish and DLoop primers. Here we note the missing samples and then find the pair for these samples that are worth using in their stead

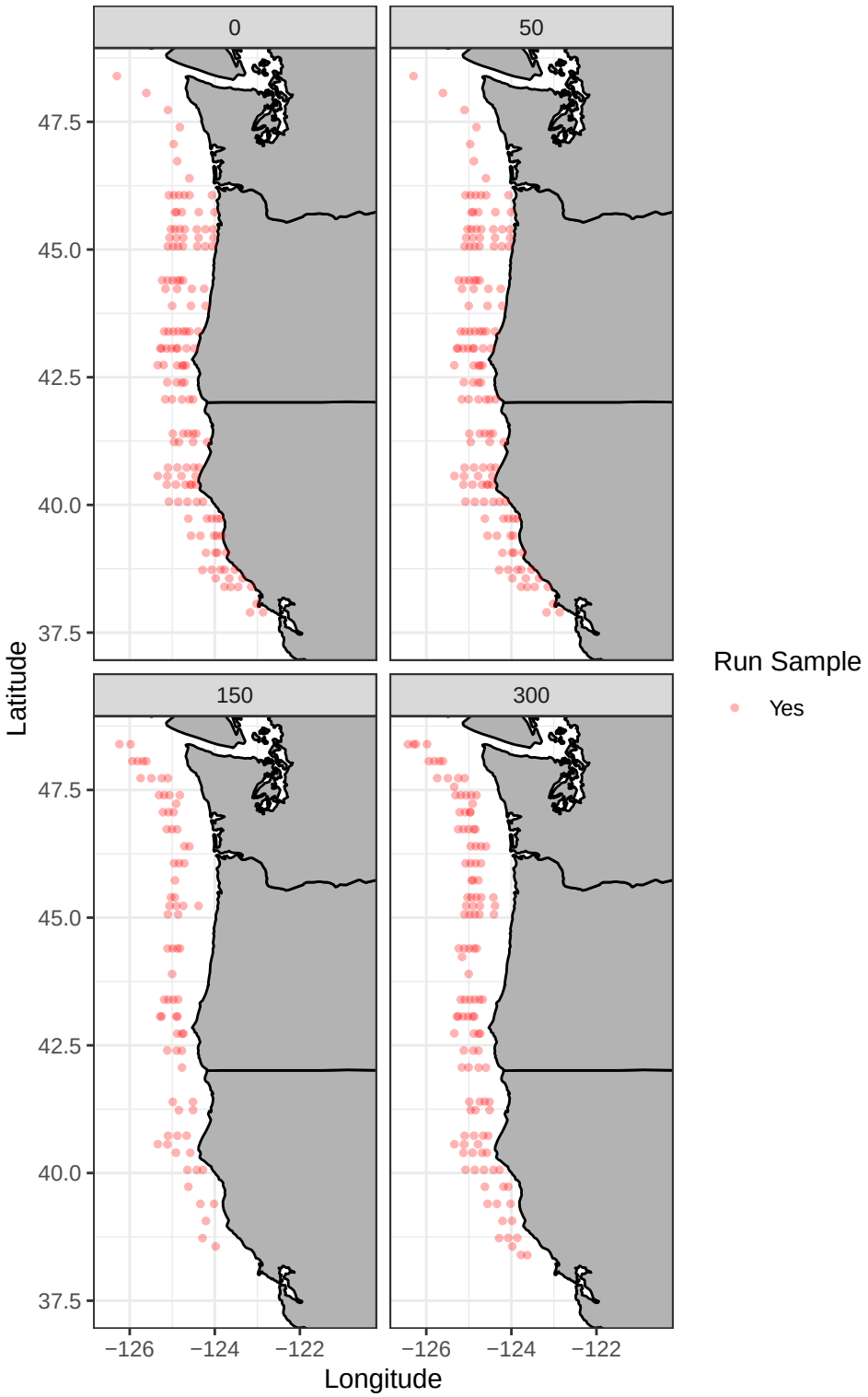

Figure 7: Scenario 1: Sample all locations at the surface, 50m and 300m, about half (75) of the stations at 150m. The samples at 150 have been arbitrarily selected in this map. We could certainly select sample on other criteria (e.g. only take samples occuring in 300m+ deep water) if people have strong opinions

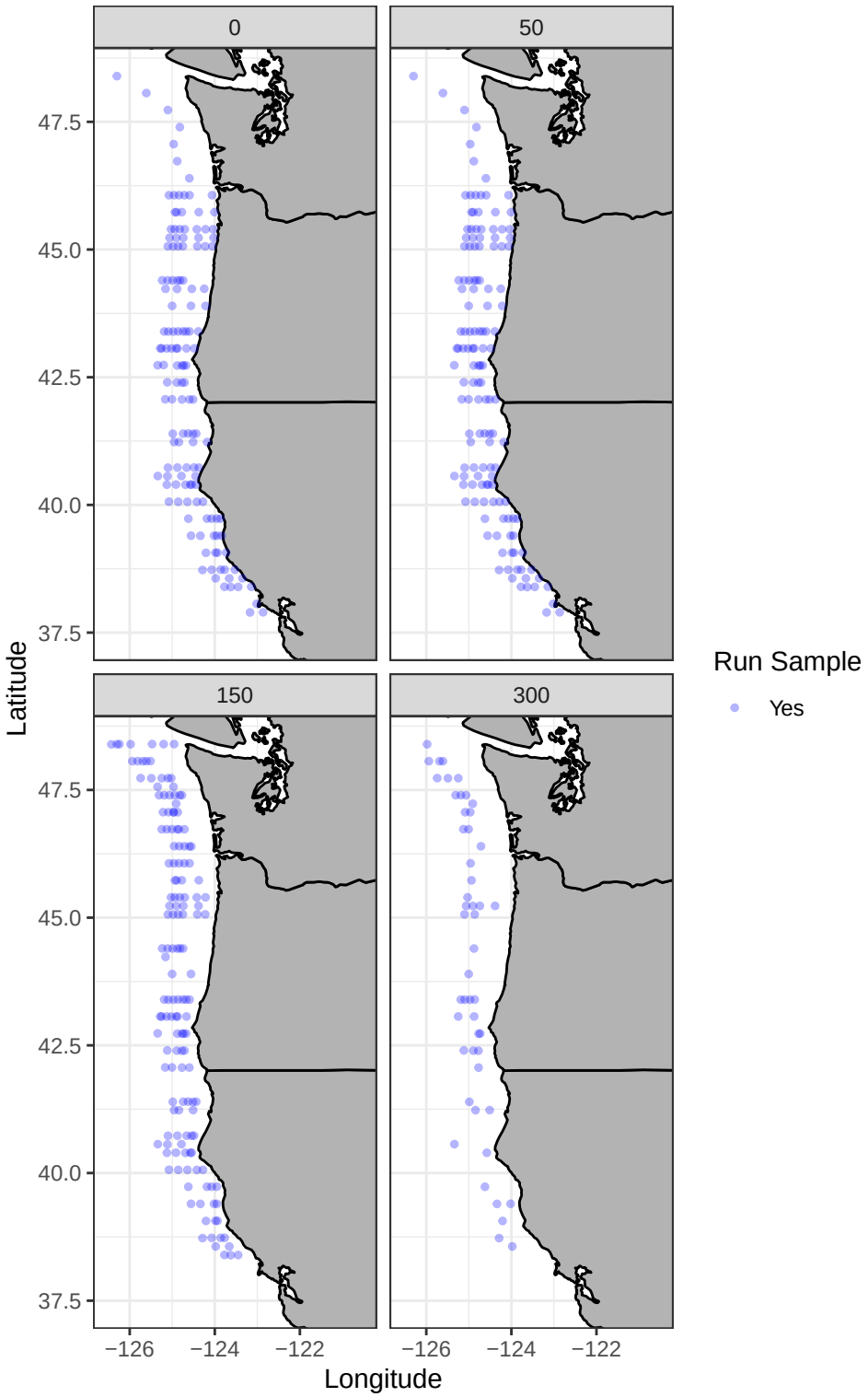

Figure 8: Scenario 2: Sample all locations at the surface, 50m and 150m, about one-third (50) of the stations at 300m. The samples at 300 have been arbitrarily selected in this map. We could certainly select samples using additional criteria.

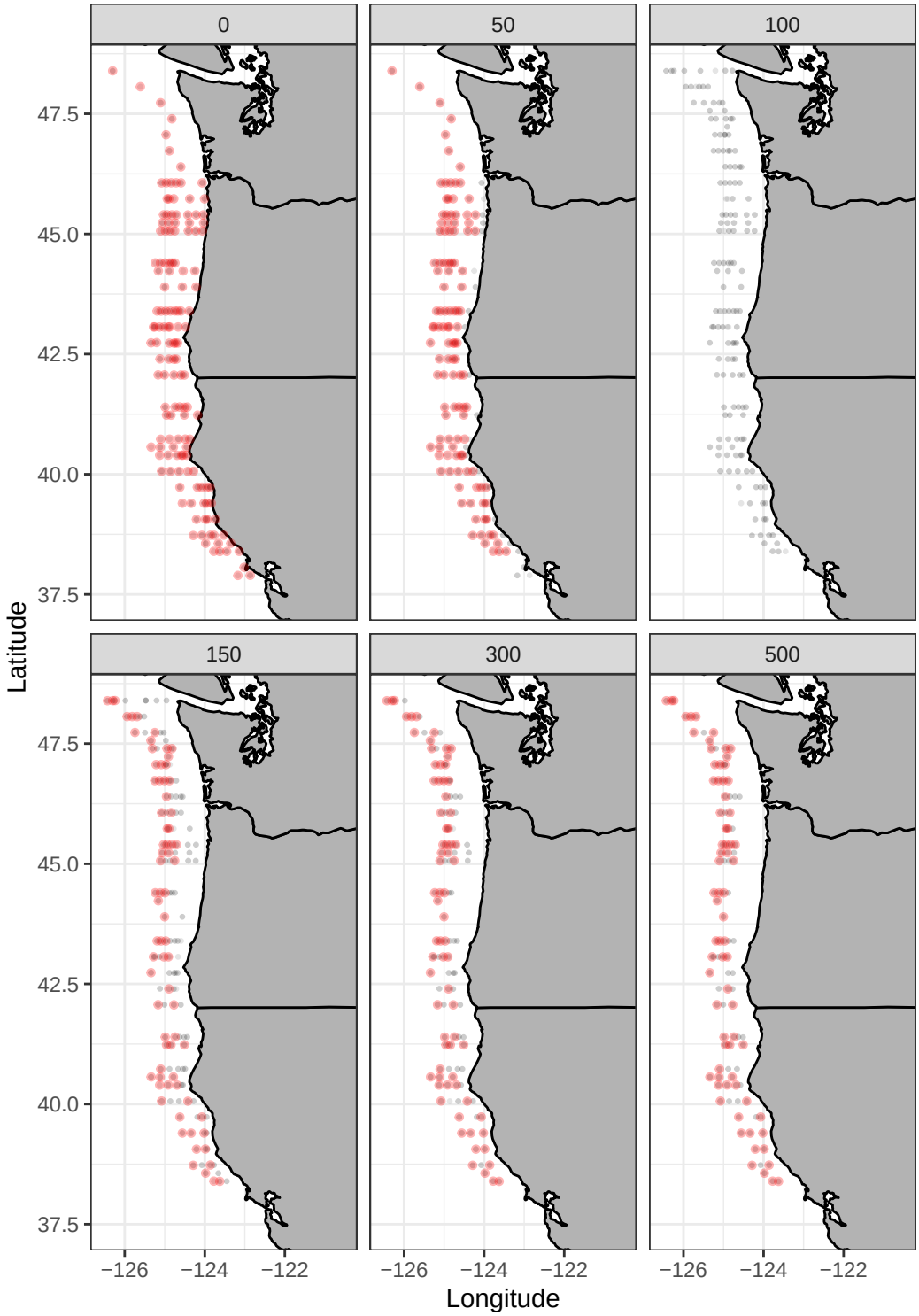

Figure 9: Scenario 3: Sample all locations at the surface, offshore samples at 50m, about 70% of the stations at 150, 300 and 500 including the most-offshore station on each transect (33 stations). The remaining 41 deep samples have been arbitrarily selected. Red show selected stations, small black dots show remaining, unselected stations.

Table 3: These are the samples that Megan identified as missing

year	sample	station	depth_cat	dilution	missing
2019	11	22-1	0	1.0	Yes
2019	17	23-1	0	1.0	Yes
2019	31	25-6	0	1.0	Yes
2019	29	25-6	50	1.0	Yes
2019	82	26-3	0	1.0	Yes
2019	135	27-1	0	1.0	Yes
2019	105	27-6	0	0.2	Yes
2019	94	27-8	0	1.0	Yes
2019	92	27-8	50	1.0	Yes
2019	162	29-4	0	0.2	Yes
2019	174	29-6	50	1.0	Yes
2019	225	31-1	0	1.0	Yes
2019	221	31-2	0	1.0	Yes
2019	212	31-4	0	1.0	Yes
2019	198	31-7	50	1.0	Yes
2019	186	31-9	50	0.2	Yes
2019	261	33-4	0	1.0	Yes
2019	249	33-5	0	1.0	Yes
2019	247	33-5	50	1.0	Yes
2019	237	33-9	0	1.0	Yes
2019	344	35-11	0	1.0	Yes
2019	296	35-3	0	1.0	Yes
2019	320	35-7	0	0.2	Yes
2019	318	35-7	50	1.0	Yes
2019	402	37-2	0	1.0	Yes
2019	382	37-5	0	0.2	Yes
2019	369	37-7	50	1.0	Yes
2019	362	37-7	300	1.0	Yes
2019	356	37-9	50	1.0	Yes
2019	417	38-1	0	1.0	Yes
2019	453	38-4	0	1.0	Yes
2019	460	39-1	0	0.2	Yes

Table 4: These are the replacement samples. Note that there are only 31 replacements for the 32 missing.

year	sample	station	depth_cat	dilution	missing
2019	12	22-1	0	1.0	NA
2019	18	23-1	0	1.0	NA
2019	30	25-6	0	1.0	NA
2019	28	25-6	50	1.0	NA
2019	83	26-3	0	1.0	NA
2019	132	27-1	0	1.0	NA
2019	106	27-6	0	0.2	NA
2019	93	27-8	50	1.0	NA
2019	163	29-4	0	0.2	NA
2019	175	29-6	50	1.0	NA
2019	226	31-1	0	1.0	NA
2019	222	31-2	0	1.0	NA
2019	213	31-4	0	1.0	NA
2019	199	31-7	50	1.0	NA
2019	187	31-9	50	0.2	NA
2019	262	33-4	0	1.0	NA
2019	250	33-5	0	1.0	NA
2019	248	33-5	50	1.0	NA
2019	238	33-9	0	1.0	NA
2019	345	35-11	0	1.0	NA
2019	297	35-3	0	1.0	NA
2019	321	35-7	0	0.2	NA
2019	319	35-7	50	1.0	NA
2019	403	37-2	0	1.0	NA
2019	383	37-5	0	0.2	NA
2019	368	37-7	50	1.0	NA
2019	363	37-7	300	1.0	NA
2019	357	37-9	50	1.0	NA
2019	418	38-1	0	1.0	NA
2019	454	38-4	0	1.0	NA
2019	461	39-1	0	0.2	NA

The samples are in “2024-01 sample selection.csv”.