



## **Analysis of Filtration Performance for the Environmental Sample Processor**

**Samarth Jayadev, California State University, Monterey Bay**

*Mentor: Jim Birch*

*Summer 2023*

**Keywords: Environmental Sample Processor, Ecogenomics, Filtration.**

### **ABSTRACT**

The Environmental Sample Processor (ESP) extracts particulates from water via filtration to detect target species of microbial life in harmful algal blooms (HABs) and when conducting in situ environmental DNA (eDNA) analysis. These samples must be collected in a time-sensitive manner to accurately represent the desired location of a study and to prevent degradation of eDNA. As particles are collected, the device lacks appropriate parameters to move between the onboard filters since clogging is often unpredictable in heterogeneous waters. To help address this issue, code was written in R to examine filtration rate reduction points in deployment data for both the second generation (2G) and third generation (3G) ESPs. Also, water samples with varying turbidities were passed through filters in lab experiments to assess clogging points for the new ESP-Sampler. The data collected and analyzed aims to refine the methodology used for future ESP deployments when defining sample conclusions to increase collection efficiency.

## INTRODUCTION

Prior to the development of the ESP, determining the presence of microscopic organisms and biological toxins in bodies of water required manual identification of species presence through enumeration, morphological recognition or inefficient DNA probes in the laboratory (Scholin et. al 2017). Sudden exponential growth of potentially toxic photosynthetic plankton called harmful algal blooms influenced the development of quicker in-situ detection methods, enhancing the ability to identify the presence of species. Using an rRNA targeted sandwich hybridization assay (SHA) that uses oligonucleotide probes on board the ESP, detection and monitoring of HABs became more efficient (Jones et. al 2008; Scholin et. al 2017). Today, ESPs are able to capture and analyze or preserve samples for a multitude of eDNA analysis.

Environmental DNA or eDNA is the genetic material shed from microbes or multicellular organisms found in water, sediment or soil. It is used to assess biodiversity of communities by providing data on species abundance and distribution using PCR primers and generation of sequencing data (Port et. al 2016). The ESP performs this eDNA analysis by using filters to extract particles from the water and applying a lysis buffer. The buffer is then heated to create a homogenate that is added to a puck (metal chamber that encloses the filter) with nucleic acid probes which attach to complementary nucleic acids if they are present (Scholin et. al 2017). When these filters clog with material, the device will continue to sample until a predetermined volume of 1000mL or minimum rate threshold of 0.2mL/sec is reached for both 2G and 3G-ESPs. If these periods accumulate over the course of multiple filters then the overall duration of a deployment may become longer than necessary.

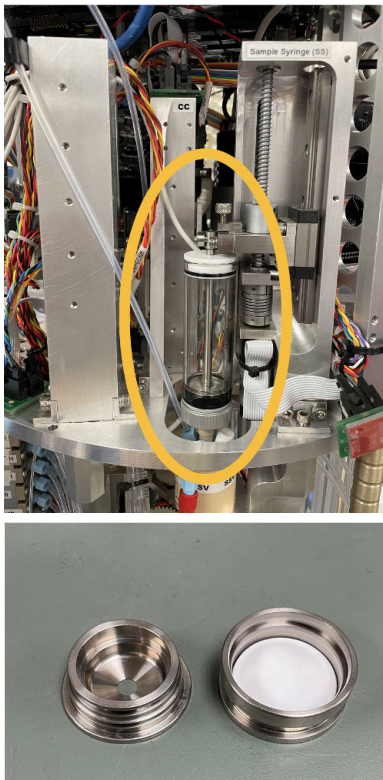
The primary issue with overextending sample duration past initial collection is dispersal and degradation of eDNA particles across the water body. Flowing water may cause concentrations of eDNA to lack correlation with local species as the sampling area becomes larger leading to inaccurate inferences of low or absent abundances of target species (Pilliod et. al 2013; Goldberg et. al 2016). Immediately after shedding from the

target species eDNA degradation begins and is further catalyzed by microbial activity, oxygenation, and mechanical forces (Goldberg et. al 2016). This adds to the necessity of efficient collection for accurate inferences to be made during any study.

By using data from previous deployments to visualize when the rate of filtration drastically reduces it is possible to increase the efficiency of future 2G and 3G-ESP deployments. Scientists can reference this data when setting sampling rate and volume limits before collection. Environmental factors that influence the quantity of particles experienced by the ESP are subject to change with every subsequent deployment. Thus previously collected data can be considered for setting approximate time and rate parameters and would be helpful in development of methodologies to predict sample clogging in situ. Collecting an adequate sample without taking longer than necessary is essential to accurately understand the ocean and freshwater systems that encompass the earth on an ecogenomic level.

## **MATERIALS AND METHODS**

### **2G & 3G FILTRATION PROCESS**



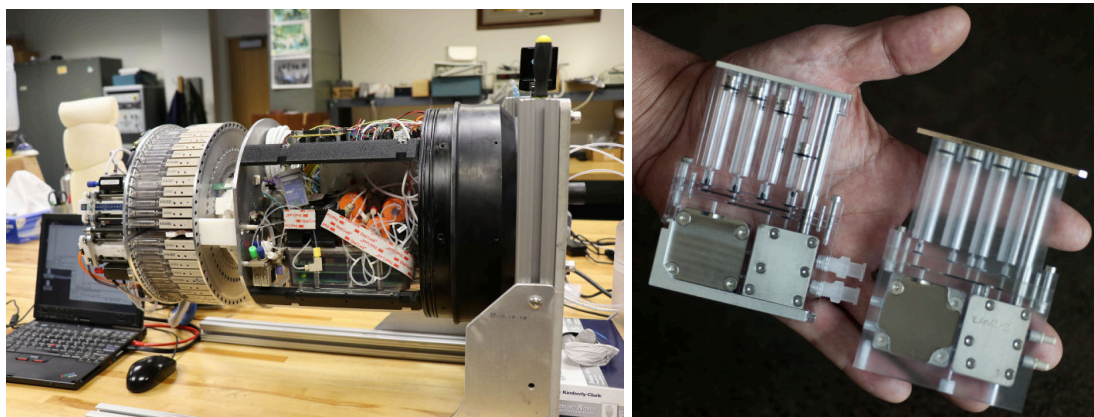
The 2G ESPs use 25mm diameter filters of either 5.0 $\mu$ m, 1.22 $\mu$ m or 0.22 $\mu$ m porosities enclosed in cylindrical metal chambers called pucks. There are two different clamps that can hold these pucks. The first is called the ‘collection’ clamp where the puck is held while the sampling syringe draws water through the filter and small solid particles aggregate. When either 1 liter has passed or clogging occurs (whichever comes first) the hand grasps the puck, the clamp opens, and the puck is moved to the second ‘processing’ clamp which presses down on it. A lysis buffer is then applied and the puck is heated to create a lysate. To preserve the sample for eDNA analysis, a preservative is added to the sample puck. Once completed



the puck is placed in an empty tube in the carousel where it remains until the device is recovered. This process gets repeated with a fresh puck for each sample collected. With every 25 milliliters that is passed the time in epoch seconds is recorded in a log. Files containing time (in epoch seconds) and volume (milliliters) were organized using the R programming language to display trends in the filtration rates of all pucks during a deployment in the Sandusky River of Ohio.

**Figure 1.** A syringe mechanism (top) in a 2G-ESP is used to draw water through the filters enclosed in metal chambers called pucks (middle). Once the lysate and preservative are added to the pucks, they are placed inside the tube (bottom).

The 3G-ESPs, like the 2G's, use 25mm diameter filters of the same porosities. Water is pushed (instead of pulled) through tubes and into cartridges that house the filters using a small three-head piston pump. 60 cartridges are rotated around a central axis that secures each cartridge in order using actuators to the output pipes from the pump. Time in epoch seconds is logged every 60 seconds and the corresponding volume is recorded.



**Figure 2.** A 3G-ESP (left) in preparation for a deployment. Water is pushed through a rotating ring, which holds cartridges (right) that are arranged radially around a central axis

The porosity of the filter is chosen based on the expected water turbidity and the target organisms of interest (i.e. viruses, bacteria, protozoans) ultimately determining the volume of water filtered. Since the collection of rare targets using eDNA is dependent on

sampled volume, filtering more water is usually preferred over less. It is important to note that eDNA is extremely small and will easily pass through the smallest 0.22 $\mu$ m filter. The filters are expected to capture the particles or cells the eDNA is attached to which are in the range of 1-30 $\mu$ m. Particle dense water clogs filters very quickly after only a small volume has been filtered thus turbid waters usually require the larger pore 5.0 $\mu$ m filter to prevent premature clogging. Clear water may need a 0.22 $\mu$ m filter to capture the finer more microscopic particles that may slip through a larger 1.22 $\mu$ m or 5.0 $\mu$ m filter. To visualize these differences deployments from the 2G-ESP in the Sandusky River of Ohio (turbid freshwater) and a 3G-ESP on a San Francisco, CA to Honolulu, HI Saildrone cruise (clear freshwater) can show how filtration rates change with different filters in different water types.

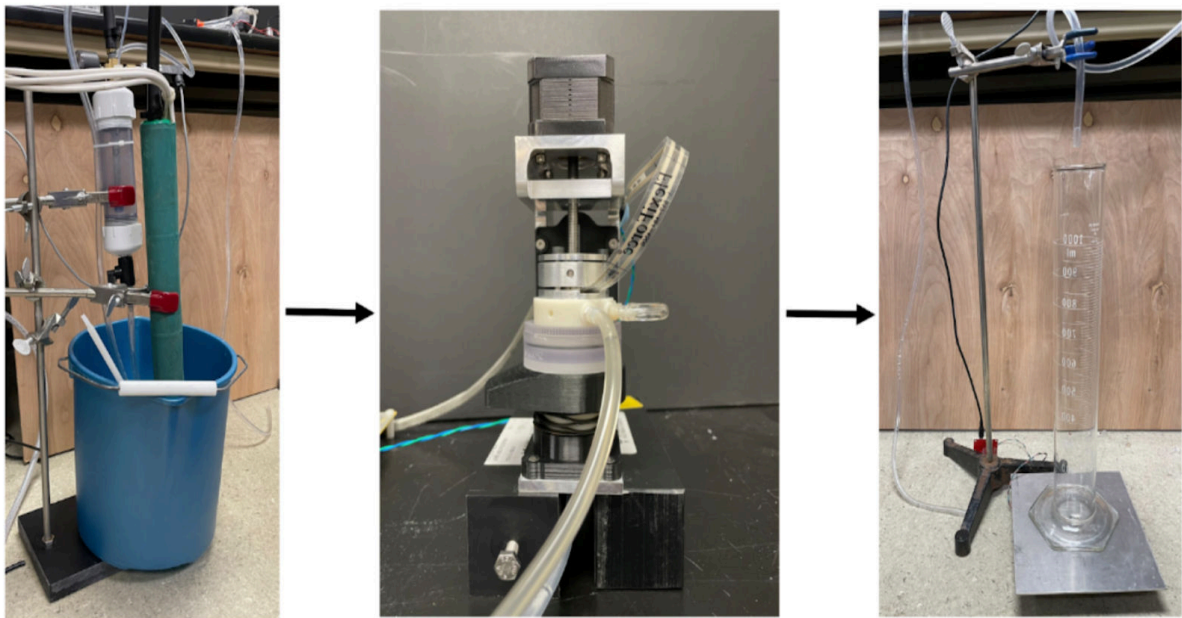
When plotting the 2G-ESP Sandusky River deployment (Figure 5 & 6) a local polynomial regression fitting or loess model was used. This nonparametric model for the curve was chosen because the relationship between rate and time amongst all pucks displayed a curved trend rather than a simple linear relationship. In order to capture points when rate begins to reduce unexpectedly, the model is able to apply multiple polynomial functions to better fit local points in the data. This helps with identifying where most pucks get clogged rather than assuming filtration rate is linear from start to finish.

Since 3G-ESPs output data in frequent intervals of 1 second, the dots on a scatter plot between pucks may merge or overlap on a larger time scale of over one-thousand seconds causing difficulties in trend distinctions between pucks. Instead of plotting all points from a deployment in one graph, they were divided into groups of 10 pucks for both filtration rate vs time and percent of total volume vs time (Figure 7 & 8).

## ESP SAMPLER DATA COLLECTION

The ESP-Sampler is an ESP being designed to collect samples and preserve them for lab DNA analysis. It is necessary to collect new filtration rate data because of the larger surface area of the new 47mm diameter filters being used on this device. An experimental

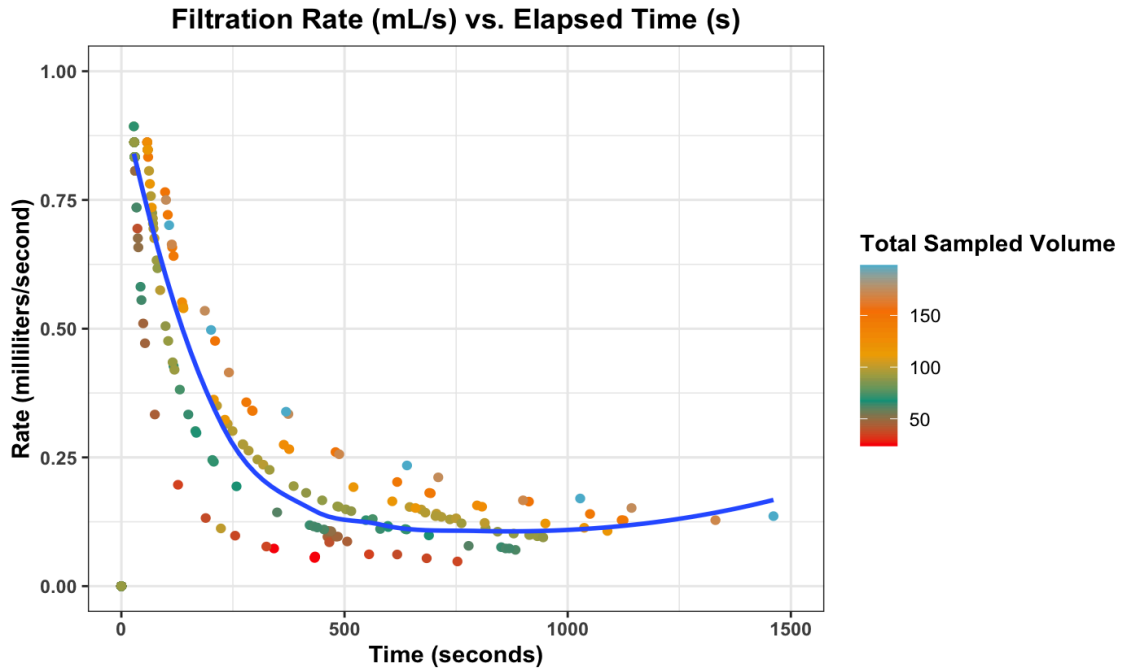
setup was created to draw water past the filters and measure rate of water collection. Samples of water from the local tap, Monterey Bay, and the California Delta were pumped at 20 psi through 5.0 $\mu\text{m}$ , 1.22 $\mu\text{m}$ , and 0.22 $\mu\text{m}$  filters clamped in pucks. The water was collected in a graduated cylinder atop a weighing scale where weight (grams) and time (milliseconds) were collected in TeraTerm. A flowmeter was also used at the end of the water passage to verify the accuracy of volume being collected. This data was processed in excel to generate plots of rate vs time and proportion of total volume vs time similar to the 2G and 3G plots (Figure 9 & Table 1).



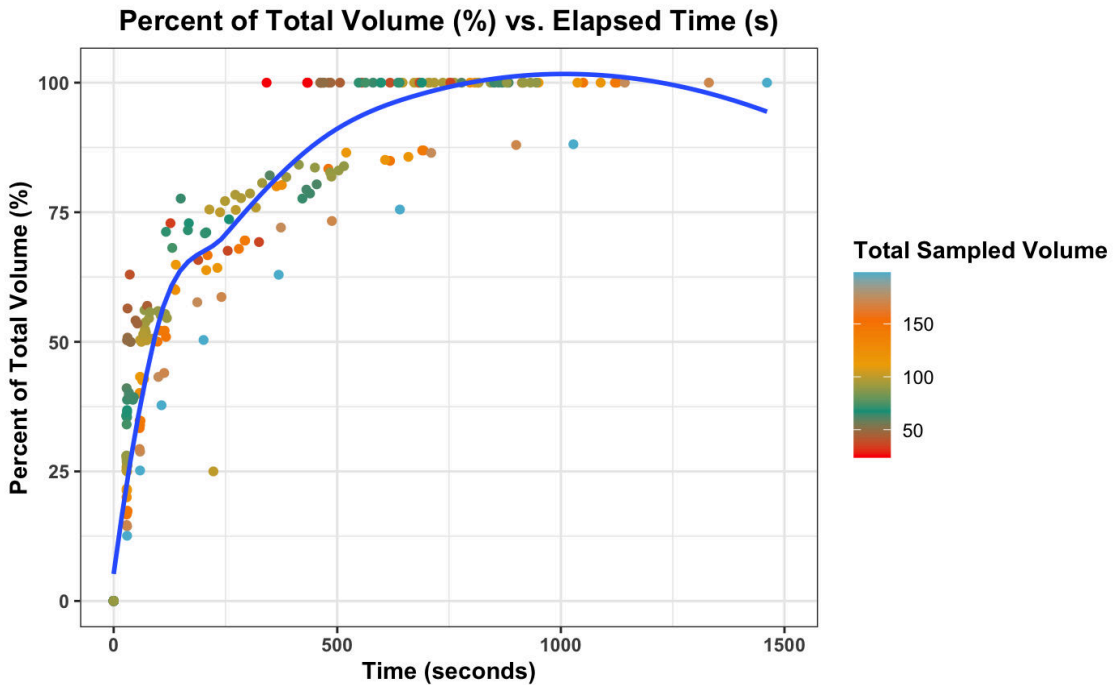
**Figure 3.** ESP-Sampler filtration experiment with pump/gas phase separator (left), filter enclosed in clamped puck (middle), and graduated cylinder atop weighing scale (right).

## RESULTS

### 2G FILTRATION DATA



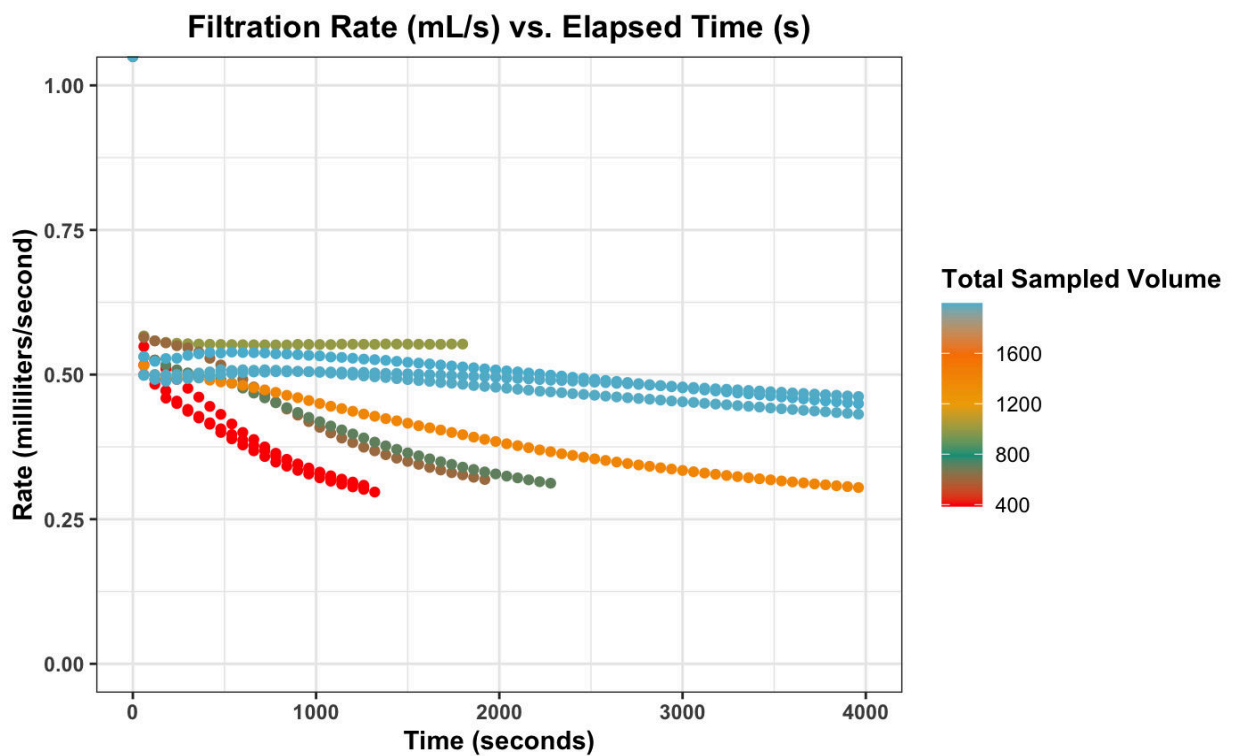
**Figure 5.** Scatter plot of rate (milliliters/second) vs time (seconds) for 61 pucks in a 2G-ESP collected for a deployment in the Sandusky River, Ohio. The blue line of best fit follows a local polynomial regression fitting or loess model and the color legend (right) represents the total sampled volume each puck reaches.



**Figure 6.** Scatter plot of percent of total volume (%) vs time (seconds) for the same deployment in the Sandusky River.

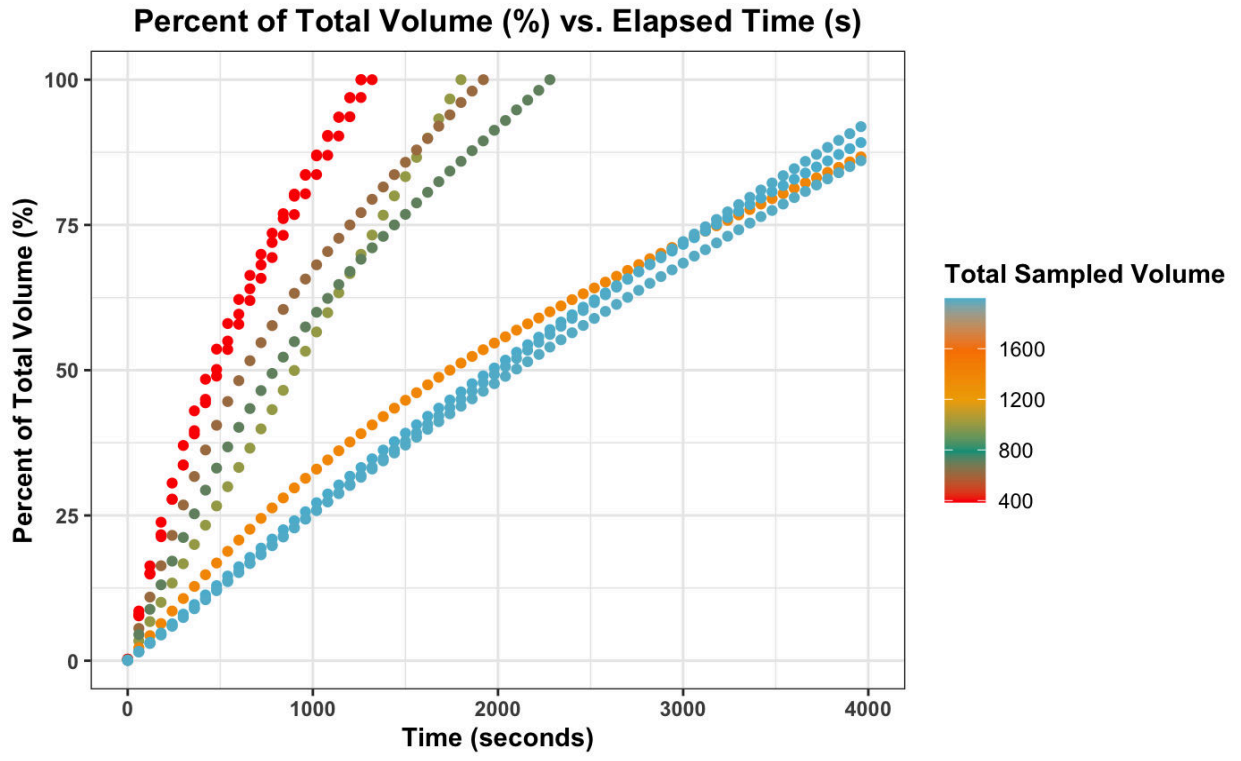
Before a deployment the filtration rate vs time graph can be used to trace the trendline or an individual puck's rate to the clogging point and match the corresponding time. The rate of clogging and its corresponding volume can help to set the threshold for sample conclusion if the current sampling rate falls below or if the maximum volume has been reached. Percent of total volume plots along with the table can be used to understand how much water was able to pass through the filters before clogging occurred.

### 3G FILTRATION DATA



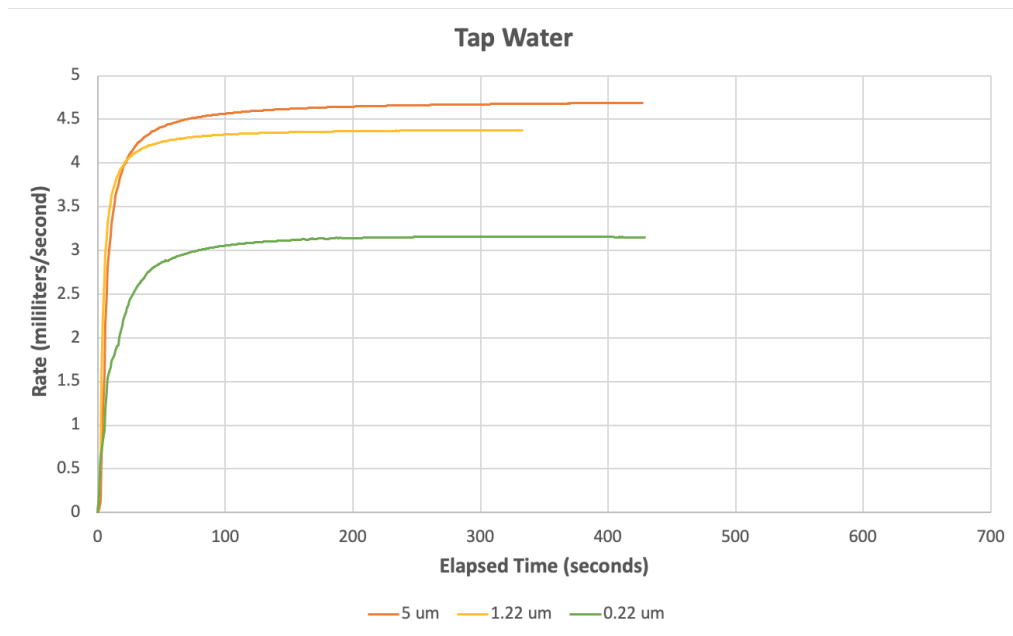
**Figure 7.** Scatter plot of 10 pucks in a 3G-ESP deployment from San Francisco, CA to Honolulu, HI. Rate changes are observed over time in one second intervals and total sampled volume is depicted by the color of the puck. Line of best fit was omitted as fewer pucks were used and individual puck trends are easily distinguishable.

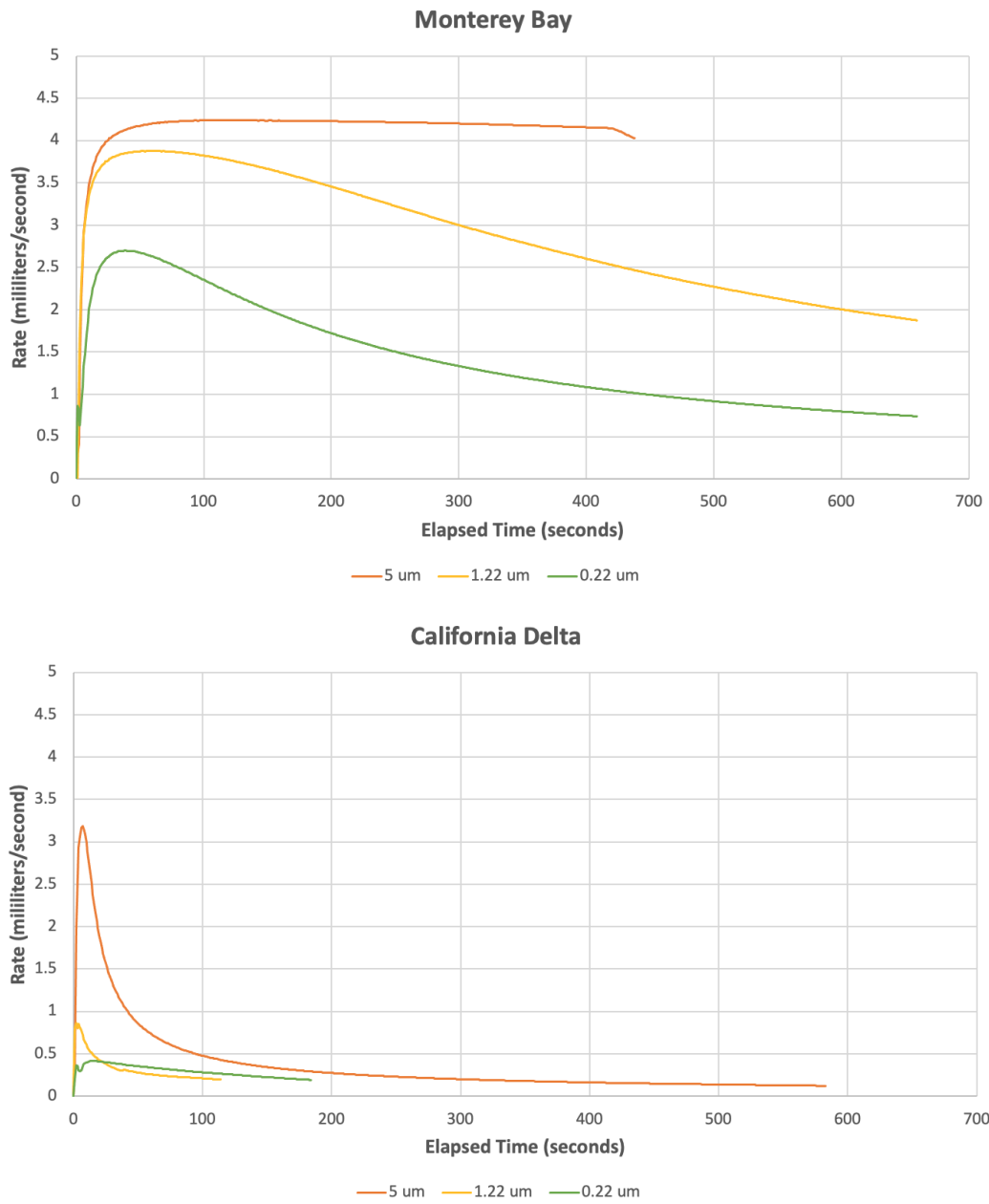




**Figure 8.** Scatter plot of proportion of total volume vs time for the same 3G-ESP deployment from San Francisco, CA to Honolulu, HI.

### ESP-SAMPLER FILTRATION DATA





**Figure 9.** Plot of rate (milliliters/second) vs. time (seconds) for the three filter porosities (5.0µm, 1.22µm or 0.22µm) in each of the three water samples (Tap, Monterey Bay, and CA Delta) used for the ESP Sampler lab experiment.

<b>California Delta (5 um)</b>			
<b>Seconds</b>	<b>Volume</b>	<b>Rate</b>	<b>Proportion</b>
0	0.00	0.00	0.00
1	0.30	0.30	0.00
2	3.90	1.95	0.05
3	7.77	2.59	0.10
4	11.73	2.93	0.16
5	15.30	3.06	0.20
6	19.00	3.17	0.25
7	22.30	3.19	0.30
8	25.27	3.16	0.33
9	27.80	3.09	0.37
10	29.93	2.99	0.39
11	31.63	2.88	0.42
12	33.03	2.75	0.43
13	34.20	2.63	0.45
14	35.03	2.50	0.46
15	35.73	2.38	0.47
16	36.30	2.27	0.47
17	36.80	2.16	0.48
18	37.17	2.06	0.49
19	37.57	1.98	0.49
20	37.93	1.90	0.50

**Table 1.** Example reference table of time (seconds), volume (mL), rate (mL/sec), and proportion of total volume for the average of three test runs with California Delta water through a 5.0µm pore size filter.

## DISCUSSION

It is important to note that the high variability of particles in freshwater and ocean environments can make it difficult to know how much of a sample is adequate. Clogging can represent when a majority of particles have been captured from the sampled volume of water. Sometimes filter clogging may never occur as demonstrated by the clearer ocean water deployment from San Francisco to Honolulu. In other instances, clogging can occur as quickly as 500 seconds or a little over 8 minutes in the murky freshwater environments of the Sandusky River (Figure 5). As more data from ESPs gets recorded and examined in different water types the parameters of rate and total volume sampled can become better defined.

The scatter plots of 2G data (Figure 5 & 6) showed that the parameters of 1000 mL maximum sampled volume or a rate threshold of 0.06mL/sec for a 2G-ESP can prolong sampling duration by letting slow drops of water pass through until the volume or rate parameter is reached. This prevents the transition to the following puck for capture of

more material from the water body. The local polynomial regression of loess model line of best fit displays the clogging point to be at about 0.1 to 0.125 mL/sec. If sampling continues beyond this filtration rate, the rate will be so slow (i.e., the filter can be considered ‘clogged’), that spurious results may be introduced, such as eDNA degradation or inaccurate representation of the intended sampling location. There are several pucks that fall well below this line and thus must also be accounted for in some way to prevent premature collection stoppage.

3G-ESPs present challenges in scatter plot interpretation due to the high frequency of data points recorded and the clear water bodies examined in this experiment (Figure 7 & 8). The controls of pucks were fed a known target (positive control) or purified water (negative control) and the low particle density in the clear ocean water made trends between collection samples and controls difficult to distinguish. For the sake of this analysis the deployment from San Francisco, CA to Honolulu, HI which consisted of 60 samples (52 were successfully preserved) were split into sets of 10 samples to distinguish individual puck trends (Preston et. al 2023). Four pucks displayed sampling conclusions between 1000 and 2000 seconds whereas the other six filtered 1000mL without clogging (Figure 8). This shows a variability in frequency of suspended particles captured over a single deployment by different or never-reaching clogging points amongst just 10 samples.

The code written in R for both 2G and 3G-ESPs are applicable for any previous deployment. Puck trends can be combined, individually examined or used to make inferences on sampling rate and volume filtered. The issue in making such inferences however lies in the premise that different pucks clog at different times and thus filter different volumes of water. To seamlessly move from puck to puck and capture the material essential for eDNA analysis in a time-sensitive manner is essential towards making accurate inferences about species populations and biodiversity.

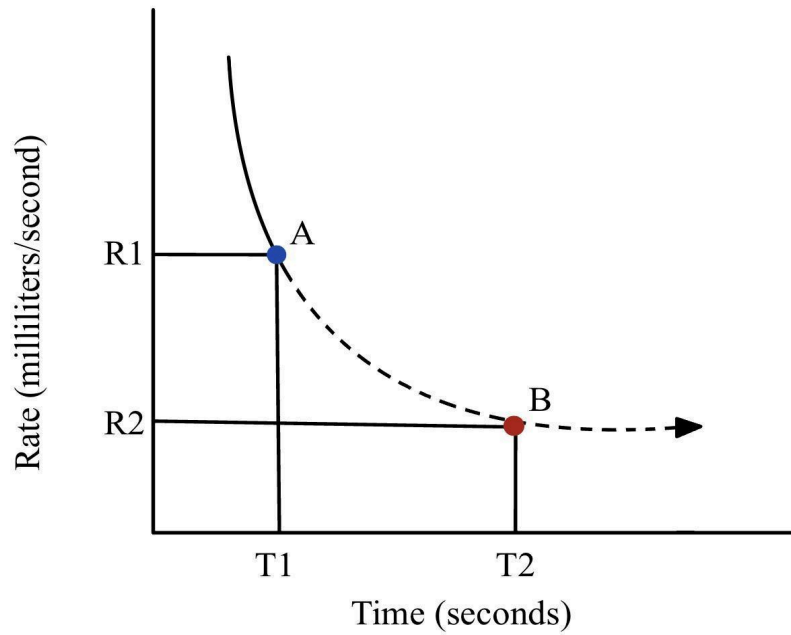
The ESP Sampler experiment provides a small sample data set of filter performance in extremely clear (Tap Water), moderately turbid (Monterey Bay), and extremely turbid

(California Delta) water types. Each line on the graph displays the averages of the three trials for the respective pore size. In clear water all pores held a constant rate at different values which is attributed to the pore size of the filter (Figure 9). In other words 5.0 $\mu\text{m}$  remained constant at  $\sim 4.68\text{mL}/\text{sec}$ , 1.22 $\mu\text{m}$  at  $\sim 4.37\text{mL}/\text{sec}$ , and 0.22 $\mu\text{m}$  at  $\sim 3.14\text{mL}/\text{sec}$ . The same properties amongst filter porosities are observed in the Monterey Bay and California Delta water samples but there is a decrease in the rate over time due to the accumulation of particles clogging the pores. In all samples the rate initially increases as the filters are yet to clog and the scale begins collecting fluid after starting at 0 grams.

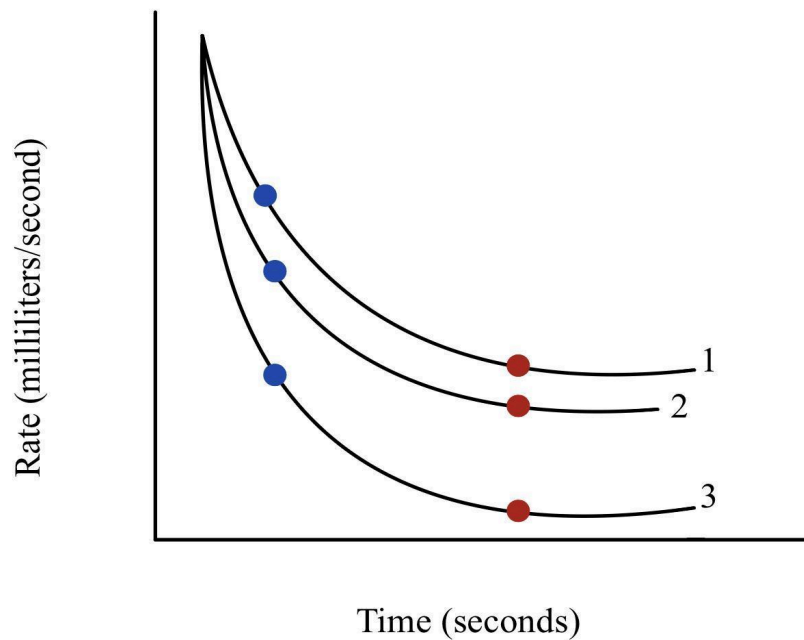
The reference table can be used when deploying the ESP Sampler device in similar water types. Researchers can set a desired sample volume or rate parameter and the table will help them understand the duration of the deployment as well as what possible proportion of the sample they might get from that period (Table 1).

## **CONCLUSIONS/RECOMMENDATIONS**

This solution to inefficient deployments can provide a ballpark estimate of sample conclusion based on water type, but as water particle turbidity differs between seasons, years, environmental conditions, and over the course of a single deployment there needs to be a way for the ESP to calculate sampling rate and volume thresholds during collection. Moreover the rate of filtration may not behave the same for all pucks in heterogeneous waters. It is therefore recommended that the ESPs use an onboard algorithm that can project when sampling will stop based on the current sampling rate, previously known sampling rates, and real-time turbidity measurements. The following figures can display what this predictive model may look like.



**Figure 10.** Graph of rate vs time to visually represent the sampling algorithm. At point A the rate data until  $R_1$  at time  $T_1$  will be used to make a projection of where to stop (point B). As sampling approaches point B the rate data approaching  $R_2$  until time  $T_2$  will be used to refine that projection.



**Figure 11.** A visual representation of rate vs time on a test data set from a previous deployment to verify the accuracy of the projection. The blue points are examples of times where rate data can be used to predict the red stopping point.

Researched algorithms and methods of prediction can be investigated to apply on board both 2G and 3G-ESPs. It is further recommended that lab experiments determining relationships between possible on-board turbidity sensors and filter porosities be conducted to improve appropriate sampling parameters for the predictive algorithm. The waters that encompass earth are constantly changing and with these additions the ESP can better adapt to these environments for improved studies conducted with the device.

## **ACKNOWLEDGEMENTS**

I am incredibly grateful for the opportunity to learn and work alongside the ESP team. This experience has provided me with invaluable insight towards development of my own low-cost tool for tracking harmful algal blooms. Thank you to my mentor Dr. Jim Birch for supporting my learning, growth and advising me during this project. Thank you to Chris Preston and Nadia Allaf for showing me the detailed functions of the ESP and providing me with the experience of working in a technology development lab. Thank you to Brent Roman for helping me work around many of the coding issues I ran into. Thank you to Scott Jensen and Kevan Yamaharra for setting up and helping me with the ESP Sampler experiment. Finally, thank you to George Matsumoto for organizing the MBARI summer internship and providing me and my fellow interns with an unforgettable experience. The MBARI internship is made possible through the Dean and Helen Witter Family Fund, the Rentschler Family Fund, the David and Lucile Packard Foundation, and the Maxwell/Hanrahan Foundation.

## **References:**

- Goldberg CS, Turner CR, Deiner K, Klymus KE, Thomsen PF, Murphy MA, Spear SF, McKee A, Oyler-McCance SJ, Cornman RS, et al. 2016. Critical considerations for the application of environmental DNA methods to detect aquatic species. Gilbert M, editor. *Methods in Ecology and Evolution*. 7(11):1299–1307. doi:<https://doi.org/10.1111/2041-210x.12595>.
- Jones WJ, Preston CM, Marin R, Scholin CA, Vrijenhoek RC. 2008. A robotic molecular method for *in situ* detection of marine invertebrate larvae. *Molecular Ecology Resources*. 8(3):540–550. doi:<https://doi.org/10.1111/j.1471-8286.2007.02021.x>.

- Port JA, O'Donnell JL, Romero-Maraccini OC, Leary PR, Litvin SY, Nickols KJ, Yamahara KM, Kelly RP. 2015. Assessing vertebrate biodiversity in a kelp forest ecosystem using environmental DNA. *Molecular Ecology*. 25(2):527–541. doi:<https://doi.org/10.1111/mec.13481>.
- Preston, C., Yamahara, K., Pargett, D., Weinstock, C., Birch, J., Roman, B., Jensen, S., Connon, B., Jenkins, R., Ryan, J., and C. Scholin. 2023. Autonomous eDNA collection using an uncrewed surface vessel over a 4200-km transect of the eastern Pacific Ocean. *Environmental DNA*, 00, 1–18. <https://doi.org/10.1002/edn3.468>
- Scholin, C.A., J. Birch, S. Jensen, R. Marin III, E. Massion, D. Pargett, C. Preston, B. Roman, and W. Ussler III. 2017. The quest to develop ecogenomic sensors: A 25-Year history of the Environmental Sample Processor (ESP) as a case study. *Oceanography* 30(4):100–113, <https://doi.org/10.5670/oceanog.2017.427>.
- Snyder ED, Tank JL, Brandão-Dias PFP, Bibby K, Shogren AJ, Bivins A, Peters B, Curtis EM, Bolster D, Egan SP, et al. 2023. Environmental DNA (eDNA) removal rates in streams differ by particle size under varying substrate and light conditions. *Science of The Total Environment*. 903:166469–166469. doi:<https://doi.org/10.1016/j.scitotenv.2023.166469>.
- Turner CR, Barnes MA, Xu CCY, Jones SE, Jerde CL, Lodge DM. 2014. Particle size distribution and optimal capture of aqueous microbial eDNA. Gilbert M, editor. *Methods in Ecology and Evolution*. 5(7):676–684. doi:<https://doi.org/10.1111/2041-210x.12206>.
- Sharing Advances in Performance and Affordability of the Second-Generation Environmental Sample Processor. (2012, July 11). NCCOS Coastal Science Website.<https://coastalscience.noaa.gov/news/sharing-advances-in-performance-and-affordability-of-the-second-generation-environmental-sample-processor/>