Developing a low-cost and long term population abundance measure to support emerging deep-learning tools: stochastic models for abundance estimation for remote underwater video (RUV) stations.

Background and Purpose of research

For marine population management and conservation, appropriate volumes of population or spatial abundance data is essential (Williams, et al., 2002). In marine environments, methods to obtain this data can be split broadly into three categories: Under Water Visual Census (UVC) techniques, fishing methods (Samoilys & Carlos, 2000) or Remote Underwater Video (RUV) stations (Stobart, et al., 2007).

First, the broad category of UVC techniques typically involve SCUBA divers present in the water to conduct surveys of the population present. UVC is often used as it enables many rapid measurements at once, like habitat structure, population distribution, species composition and size structure (Stobart, et al., 2007). However, UVC data incurs a large bias from the divers themselves. Due to variable skill, or unfavourable conditions, detection is often imperfect, leading to an underestimation of animal counts (Samoilys & Carlos, 2000).

Second, the category of fishing techniques often relies on stochastic models to estimate species abundance, often used in fisheries (Samoilys & Carlos, 2000). Though there is a wide variety of methods, most abundance models arise from either removal or capture-recapture techniques, or some combination of the two. Removal techniques are based on the principle that when more effort is put into catching animals from the ecosystem, the larger the fraction of the population is removed. The effect of this removal on subsequent catch intensity is used to estimate animal abundance (Katsanevakis, et al., 2012). In mark-recapture techniques, animals are caught and tagged. The distribution of tagged and untagged animals caught is then used to estimate species abundance (Katsanevakis, et al., 2012). One of the most robust of these models is known as the Jolly-Seber model (Schwarz & Arnason, 1996) which has been modified in this investigation.

Both UVC and fishing methods have the following in common: they both require extensive manpower in the field and high costs of equipment like renting research vessels or SCUBA gear when compared to the alternative – RUV stations.

RUV stations typically involve underwater cameras with variable deployments: either stationary or moving; and either floating or benthic (King, et al., 2018). Not only does this method require less resources, but recent developments in the field of machine learning, a subset of artificial intelligence, have opened the doors to even more cost and time efficient methods to support RUV studies (Beyan & Browman, 2020). This is achieved by providing automated video or image analyses to achieve species identifications, along with their corresponding detection probabilities to account for imperfect detection and incalculable human detection bias (Samoilys & Carlos, 2000). This allows new RUV methods to overcome the key obstacles in UVC techniques, while simultaneously requiring less time and incurring lower costs.

However, unlike fishing methods, RUV techniques are still in the early stages of developing general and robust stochastic models to obtain species abundance or population density estimates, failing to take full advantage of the supporting deep learning tools. Studies now still frequently use relative density rather than absolute density estimates, through count metrics like MaxN, the maximum species count identified in a series of frames, or MeanCount, the average species counts per frame (Williams, et al., 2018). These relative indices limit data from being compared between survey methods or survey sites, or for the direct assessment of populations (Williams, et al., 2018).

Most unmarked camera based population estimation models have currently been developed for terrestrial environments, like the Time to Event (TTE) and Space to Event (SE) models. However, they often are either closed population models or require extensive on-site auxiliary data (Moeller, et al., 2018).

Unlike terrestrial density estimates which are measured in unit areas, there is also a need for marine species density estimates to be obtained per unit volume, for extrapolating absolute abundance. This is due to the extensive three dimensional movement of marine species, compared to terrestrial species. This also leads to the heterogenous distributions of abundance not only on the two dimensional plane, but especially between water columns. Usually, species density tends to be higher in the benthic zone in shallow water regions like coral reefs or seagrass beds (Walker, 2012). On the other hand in pelagic environments, many animals tend towards the photic zone, thus having an inherent bias toward the surface. These factors, cause marine populations to be extremely open, causing most unmarked camera estimation models developed for terrestrial animals extremely inapplicable in marine environments.

Therefore, this investigation aims to develop a model that enables absolute density estimates of species complementing the use of modern RUV methods supported by deep learning tools, requiring as little auxiliary data as possible.

The model proposed in this investigation is referred to as the Open Time to Event (OTTE) model as it is an extension of the terrestrial Time to Event (TTE) model (Moeller, et al., 2018), and the open-population Jolly Seber Model (Schwarz & Arnason, 1996). To minimise video processing time and costs associated with deep learning tools, this model aims to use the timing of the first arrival, and first departure of a target species at each camera to accurately estimate its absolute abundance in a defined sampling region, within an open population.

Methodology

Proposed general method for data collection supporting the OTTE model

First, a sampling region needs to be determined prior to the study. It is defined as the volume of interest for the population survey, for which the absolute abundance estimate ($\widehat{N}$) is obtained.

Similar to the TTE model (Moeller, et al., 2018), a fixed number of cameras ($c = 1,2,\ldots K$) are randomly placed throughout the desired sampling region, along the same water column height. It must be ensured that the cameras do not have overlapping viewsheds, which can achieved by ensuring a large enough sampling region. Additionally, the maximum amount of time each camera records video footage, or takes a series of photographs, needs to be fixed. This maximum time will be referred to as T_{max}.

The following RUV camera placement options will be analysed separately in this investigation: stationary benthic cameras for shallow water abundance estimation, and unmoving floating cameras for pelagic population estimation (Heagney, et al., 2007).

Deep learning tools can then be used to analyse these videos to determine the timings when;

1. one animal first arrives into the viewshed of a camera
2. the animal first leaves the viewshed of that camera

This paper proposes that these two data-points for each camera is sufficient to obtain an average $\widehat{N}$ of the sampling region, valid for the time 0 to T_{max}. More cameras will lead to more robust estimates, however a minimum of three may suffice depending on the ecological characteristics of the animal, as long as the size of the sampling region is appropriately manipulated (ref Results and Discussion).

The OTTE Model Framework

Given that in an open population, animals may move in and out of the sampling region or the viewshed of the camera, $\widehat{N}$ will be dependent on the probability distribution the animal staying within a unit volume (Schwarz & Arnason, 1996), at a given time, $P(S)$. This fundamental concept has been adapted from the Jolly Seber Model (Schwarz & Arnason, 1996).

This probability can be derived from the arrival time, and departure time from the viewshed of the animals from each viewshed of the camera from time 0 to T_{max}. In accordance to the TTE model in which it assumes the independent movement of animals, these two events, arrival (A) and departure given arrival ($D|A$) are Poisson processes (Moeller, et al., 2018). Therefore, the time till a fish first arrives ($T_{A,c}$) and a fish first departs ($T_{D|A,c}$) at each camera is exponentially distributed

$$T_{c(A)} \sim Exponential(\lambda_A)$$

$$T_{c(D|A)} \sim Exponential(\lambda_{D|A})$$

where λ_A is the average count of event A occurrences, $\lambda_{D|A}$ is the average count of event $D|A$ occurrences, and $c = 1,2..K$ denotes the camera number. The estimates for λ_A and $\lambda_{D|A}$ can thus be found through the following likelihood functions modified from Moeller, et al. 2018:

$$L(\lambda_A|T_{c(A)}) = \prod_{c=1}^{K}(I_{(T_{c(A)}\leq T_{Max})}\lambda_A e^{-\lambda_A T_{c(A)}} + (1 - I_{(T_{c(A)}\leq T_{Max})}))e^{-\lambda_A T_{max}})$$

$$L(\lambda_{D|A}|T_{c(D|A)}) = \prod_{c=1}^{K}(I_{(T_{c(D|A)}\leq T_{Max})}\lambda_{D|A} e^{-\lambda_{D|A} T_{c(D|A)}} + (1 - I_{(T_{c(D|A)}\leq T_{Max})}))e^{-\lambda_{D|A} T_{max}})$$

In the above expressions, the indicator function $(1 - I_{(T_{c(D|A)}\leq T_{Max})}$ is used to right censor data when $T_{c(A)}$ or $T_{c(D|A)}$ values exceed T_{max}, especially due to low population densities (Moeller, et al., 2018). Using the estimates of λ_A and $\lambda_{D|A}$ obtained through the above likelihood functions, the following general exponential distributions of time, T, can be obtained:

$$f_A(T) = \lambda_A e^{-\lambda_A T}$$

$$f_{D|A}(T) = \lambda_{D|A} e^{-\lambda_{D|A} T}$$

The above two expressions can be used to obtain the time probability distribution of event S, where the animals stays within the unit volume.

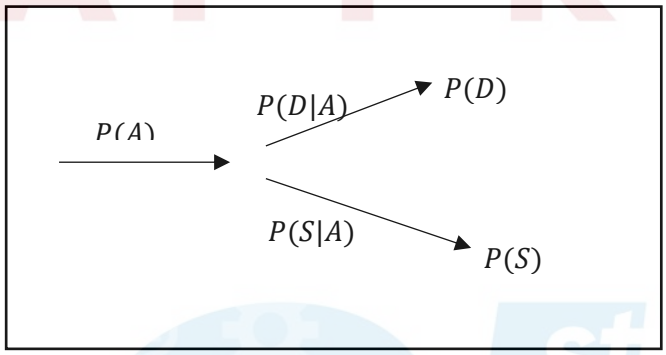

Figure 1

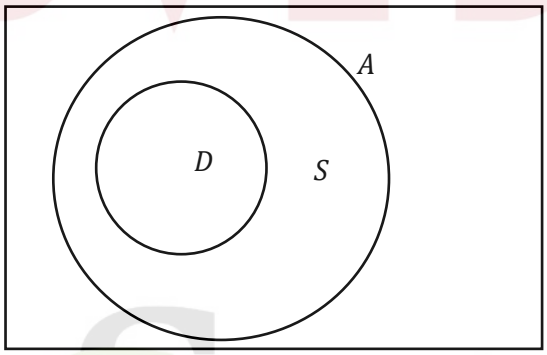

Figure 2

As shown in Figure 2, event D is a complete subset of event A as an animal may only move out of the camera viewshed if it was present in the viewshed in the first place. An animal in the viewshed also has only two mutually exclusive options, either leave (D), or stay (S), which is also a complete subset of event A. Thus, the probability distribution of S can be computed as shown in the tree diagram in Figure 1 where $P(S) = P(A)(1 - P(D|A))$.

$$f_s(T) = f_A(T) - f_A(T)f_{D|A}(T)$$

Next, since the expected value of $f_s(T)$ would give the expected amount of time an animal will stay within the viewshed, its reciprocal will give the average number of animals that stay within the viewshed, in a given time. This was computed to be:

$$\lambda_S = \frac{1}{E(f_S)} = \frac{\lambda_A(\lambda_A + \lambda_{D|A})^2}{(\lambda_A + \lambda_{D|A})^2 - \lambda_{D|A}\lambda_A^2}$$

Therefore, the absolute abundance estimate can be obtained:

$$\widehat{N} = \frac{V_{sampling\ reigon}\lambda_S}{V_{viewshed}}$$

Where V denotes volume.

Computer Simulations

A series of agent based simulations were conducted to evaluate abundance estimates of the OTTE model in open shallow water and open pelagic environments.

For both environments, the simulation of an open population was achieved through the use of a sampling region of 100x100x20 units (length, breadth, height) specified within a large universal volume. Animals are able to randomly move in and out of this sampling region, and the count of the number of animals within the sampling region was monitored over the period of the simulation (250 steps). The height of the sampling region was intentionally smaller than the length or the breadth due to heterogeneity between water columns. For all simulations, 10 cameras were used, each with a 5x5x5 viewshed.

In the benthic environment, the universal volume was cuboidal (500x500x100) to simulate shallow water conditions, and the animals had a downwards bias. The sampling region was located in the middle of the floor of the universal volume to simulate the use of benthic RUV stations. In the pelagic environment, the universal volume was cubic, while the animals had a surface bias. The sampling region was located in the middle and on the surface of the universal volume to simulate the use of surface cameras. These biases in the distribution of fishes throughout the water columns enables an realistic simulation of open populations as depicted in Figure 3.

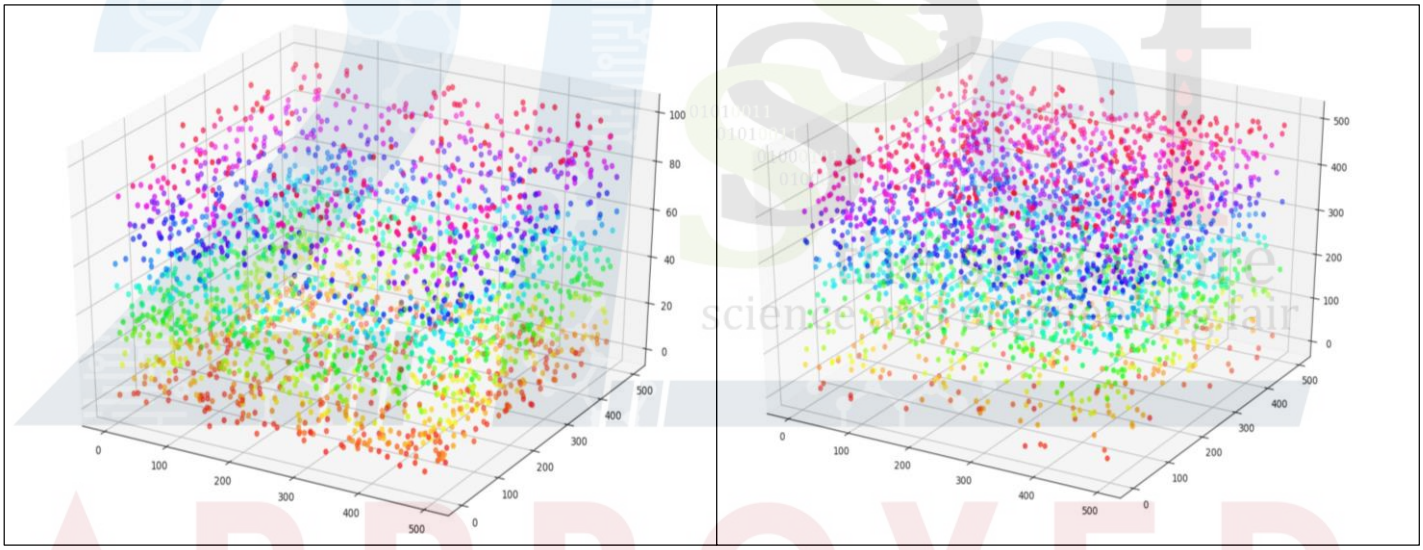

Figure 3: Benthic simulation (left) has a higher density at the bottom of the universal volume than the top; Pelagic simulation (right) has a higher density at the top of the universal volume than the bottom.

The simulations were conducted at different animal maximum speeds (1,5,10,25 and 50 step length), and at varying general densities: 25000 animals in the universal volume for the high density simulation, and 2500 animals in the universal volume for the low density simulation. These two universal populations were kept the same for the pelagic environment despite a larger universal volume so as to simulate the lower population density in Pelagic environments compared to shallow water environments like coral reefs. Each simulation was repeated for 10 replicates.

The accuracy of the $\widehat{N}$ using only the first arrival time and first departure time of animals at cameras through the OTTE model was evaluated using the t-test with a 95% significance level where the following null hypothesis (H_0), alternative hypothesis (H_1) pair was used:

H_0: $\widehat{N}$ estimate and actual N have no significant difference ($p > 0.05$)

H_1: $\widehat{N}$ estimate and actual N exhibit a significant difference ($p < 0.05$)

When H_0 is accepted, the OTTE model is considered to be a robust estimation method for those conditions.

Results and Discussion

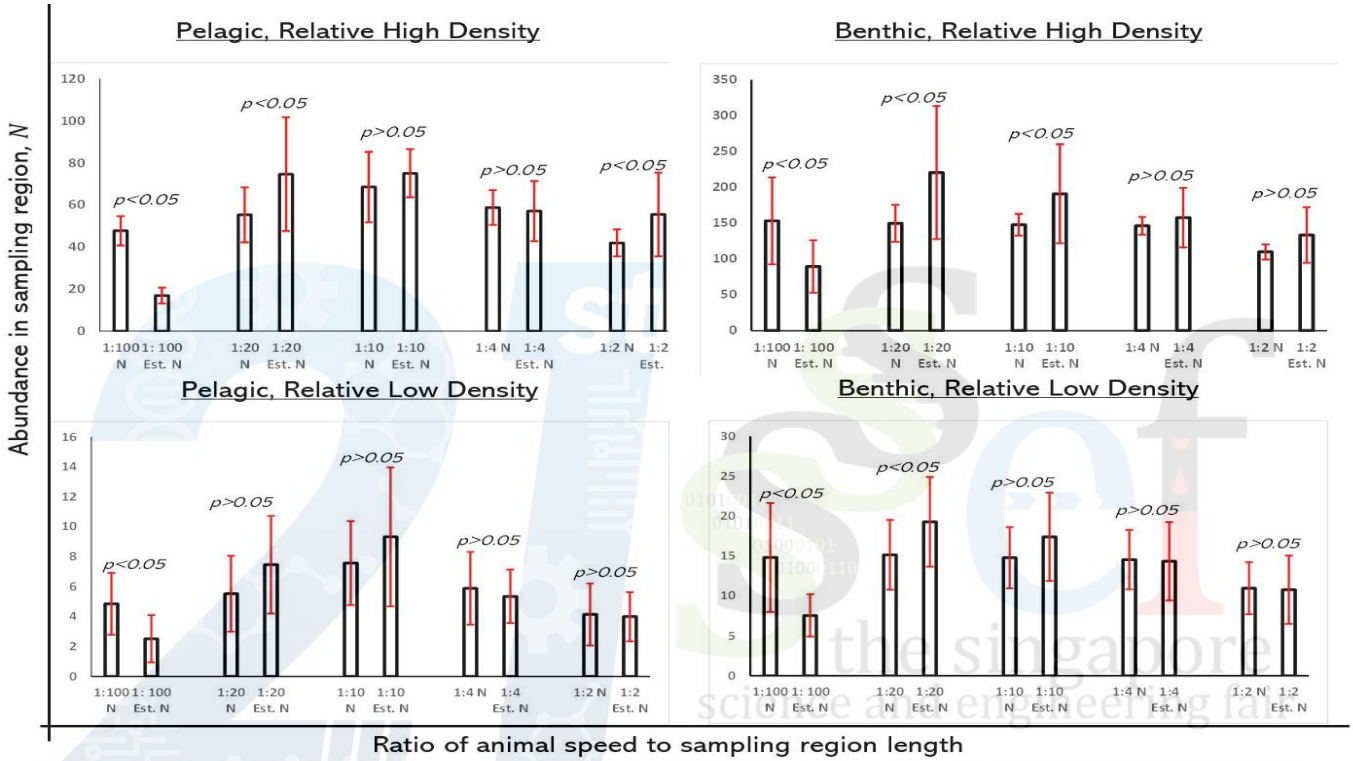

Figure 4: Simulation data summary. Full data available in appendix.

Figure 4 is obtained from data in the summary table displayed in the Appendix. It shows the average of 10 replicates in the 4 different conditions, at 5 maximum speeds (1,5,10,25 and 50 step length) which were converted to ratios of maximum speed and the length of the sampling reigon. This was done to observe the spatial significance of the speed, in relation to the sampling reigon.

Under all conditions, when the step length was at its slowest, casuing the speed ratio to become extremely small (1:100), the OTTE model incurred an inaccurate estimate ($p < 0.05$) of absolute abundance, underestimaiting $\widehat{N}$ (Est. N in Fig 4) by a significant margin. However, when maximum speed increased to 5, there was an immediately visbile improvement in the estimate as the difference in height between the $\widehat{N}$ (Est. N in Figure 4) bar and N bar becomes notably smaller. This pattern where maximum animal speed has a predictable positive relationship with accuracy of $\widehat{N}$ continues for all conditions, up till the speed ratio of 1:4. For both benthic and pelagic conditions, the best estimates occurred at 1:4 speed ratio at high density and 1:5 speed ratio at low density repectively.

An increase in maximum speed likley causes and increase in OTTE estimate accuracy due to the improvement of the speed to length ratio. When an animal moves faster, it is able to move through a larger volume in a shorter period of time, increasing the chances of it being captured by a camera before time T_{max} is reached. This reduces the number of data points which have to be right censored, thus causing an increase in accuracy. One way to address this issues is to increase T_{max} to minimise the amount of data that is right censored.

At high density, the accuracy of the $\widehat{N}$ may be compromised by the higher varinace in the data for arrival times and departure times between replicates. This was observed by the larger error bars at speed ratio 1:2 under high density conditions. This occurs due to the random movement of the animals, coupled with higher speeds, and a higher density causeing a magnifying effect on the volatility of the arrival and departure times between replicates. This can be addressed through the use of more cameras while increasing the size of the sampling reigon. This way, the speed ratio is reduced to a level that is less volatile, while not reducing to the point of compromising accuracy. This way, variance between replicates should decrease, enabling more accurate estimates. Nevertheless, this solution reuires further investigation.

Conclusion

The presence of $p > 0.05$ events reveals that the low cost OTTE method can succesfully make accurate estimates of open and unmarked populations. The accuracy of the OTTE model is not affected by the environment (ie. benthic or pelagic), but it is sensitive to an animal's ecological chracteristics like its swimming speed or population distribution. For extensive applicatinons in fisheries managements or species conservation, the results of this study can be used to structure future implemntations of the OTTE model developed in this paper for the most accurate results. First, if the target species has a low maxium speed, the length (and thus volume) of the sampling reigon should be considerably reduced. This maintains a high speed to sampling reigon length ratio, maintaining the accuracy of the $\widehat{N}$ estimate. Second, if the species of intrest has a known high population density and a high speed, increase the size of the sampling reigon through the use of more cameras will also enable a more accurate $\widehat{N}$ estimate.

Bibliography

Beyan, C. & Browman, H. I., 2020. Setting the stage for the machine intelligence era in marine science. *ICES Journal of Marine Science,* 77(4), pp. 1267-1273.

Heagney, E. C., Lynch, T. P., Babcock, R. C. & Suthers, I. M., 2007. Pelagic fish assemblages assessed using mid-water baited video: standardising fish counts using bait plume size. *Marine Ecoogy Progress Series (MEPS),* Volume 350, pp. 255-266.

Katsanevakis, S., Pipitone, C., Leopold, M. & Weber, A., 2012. Monitoring marine populations and communities: review of methods and tools dealing with imperfect detectability. *Aquatic Biology,* p. 16(1):31–52.

King, A. J. et al., 2018. Efficacy of remote underwater video cameras for monitoring tropical wetland fishes. *Hydrobiologia,* Volume 807, pp. 145-164.

Moeller, A. K., Lukacs, P. M. & Horne, J. S., 2018. Three novel methods to estimate abundance of unmarked animals using remote cameras. *Ecosphere,* Volume 9(8):e02331. 10.1002/ecs2.2331.

Samoilys, M. A. & Carlos, G., 2000. Determining Methods of Underwater Visual Census for Estimating the Abundance of Coral Reef Fishes. *Environmental Biology of Fishes,* Volume 57, p. 289–304.

Schwarz, C. J. & Arnason, A. N., 1996. A General Methodology for the Analysis of Capture-Recapture Experiments in Open Populations. *International Biometric Society,* 52(3), pp. 860-873.

Stobart, B. et al., 2007. A baited underwater video technique to assess shallow-water Mediterranean fish assemblages: Methodological evaluation. *Journal of Experimental Marine Biology and Ecology,* 345(2), pp. 158-174.

Walker, B. K., 2012. Spatial Analyses of Benthic Habitats to Define Coral Reef Ecosystem Regions and Potential Biogeographic Boundaries along a Latitudinal Gradient. *PlosOne.*

Williams, B. K., Nichols, J. D. & Conroy, M. J., 2002. *Analysis and Management of Animal Populations.* San Diego, California, USA; London, UK: Academic Press.

Williams, K. et al., 2018. A method for computing volumetric fish density using stereo cameras. *Journal of Experimental Marine Biology and Ecology,* Volume 508, pp. 21-26.

Acknowldegments

I would like to thank Dr Matthew Lauretta for inspiring me to embark on this project, and guiding me where nessecary. I would also like to thank Adi Merothra who aided me in setting up the computer simulations (Refer Resultsa and Dsicussioin) on Python veriosn 3.8.3.

Appendix

Maximum speed	Maximum speed to sampling region length ratio	Total population	Average, actual N	$SD(N)$	Average $\widehat{N}$ from OTTE model	$SD(\widehat{N})$	Null Hypothesis Accepted
pelagic							
1	1:100		47.74	7.00	16.89	3.77	No
5	1:20		55.33	13.10	74.68	27.06	No
10	1:10		68.57	16.76	75.11	11.48	Yes
25	1:4		58.82	8.30	57.10	14.32	Yes
50	1:2	25000	41.97	6.38	55.51	19.94	No
1	1:100		4.86	2.07	2.52	1.58	No
5	1:20		5.53	2.54	7.47	3.26	Yes
10	1:10		7.58	2.80	9.33	4.64	Yes
25	1:4		5.90	2.43	5.36	1.79	Yes
50	1:2	2500	4.14	2.07	4.00	1.64	Yes
benthic							
1	1:100		153.14	60.56	89.45	36.64	No
5	1:20		149.70	25.78	220.52	92.79	No
10	1:10		147.75	15.24	190.92	69.12	No
25	1:4		146.28	12.27	157.61	41.41	Yes
50	1:2	25000	109.86	10.61	133.39	38.82	Yes
1	1:100		14.84	6.82	7.57	2.66	No
5	1:20		15.15	4.36	19.28	5.62	No
10	1:10		14.78	3.83	17.41	5.52	Yes
25	1:4		14.55	3.72	14.35	4.90	Yes
50	1:2	2500	10.97	3.25	10.79	4.27	Yes