title: "Response to reviewer's statistical comments on the paper: GLR-D-15-00174 - Effect of net size on estimates of abundance, size, age and sex ratio of Mysis diluviana Journal of Great Lakes Research" author: "John Tipton" date: "October 12, 2015" ---

Associate Editor: This is a useful analysis of mysid sampling methods that has implications to Great Lakes and beyond. The paper has received two careful reviews. Both reviewers were generally positive, but have some suggestions for improvements.

### Both reviewers wanted to know why you did not use a paired t test when comparing the two net types, which given the paired approach to sampling would appear to be the most appropriate method to analyze the data. If you have reasons not to use that approach, please elaborate in your response letter and add a sentence or two to the methods.

We begin the analysis by loading the data used for measuring total count. At each sampling event we recorded the number of mysis shrimp caught and important covariates including date, site, net size, etc.

load("~/mysis/data/mysisCountData.RData") ## load count data

By looking at the distributions of counts across time, we see that they are similar between net sizes, but the normality of the counts is questionable.

ggplot(data=mysisCountData, aes(count)) + geom\_histogram() + facet\_grid(net ~ date) +   
 ggtitle("Histogram of normalized counts by net size and date")

Because counts cannot be less that zero and often are right skewed, the assumption of Normality in the data is questionable. Hence, we don't want to use Gaussian methods like linear regression and ANOVA. But, for completeness, let's look a a paired t-test and see what the results look like with the knowledge that this analysis is less than satisfactory.

y1 <- mysisCountData$count[mysisCountData$net == "Small Net"]  
y2 <- mysisCountData$count[mysisCountData$net == "Large Net"]  
pairttest <- t.test(y1, y2, paired=TRUE)

From the paired t-test, we find that there is no significant difference ( = 0.069, = 39, -value = 0.945 ) in expected counts.

We can also look at a paired t-test for the mean length of the nets, although this loses alot of information because we are only testing the sample level means without accounting for the differences in uncertainties about the site level means due to different sample sizes between nets

load("~/mysis/data/mysisLengthData.RData")

First lets look at some histograms and get an idea of what the data look like.

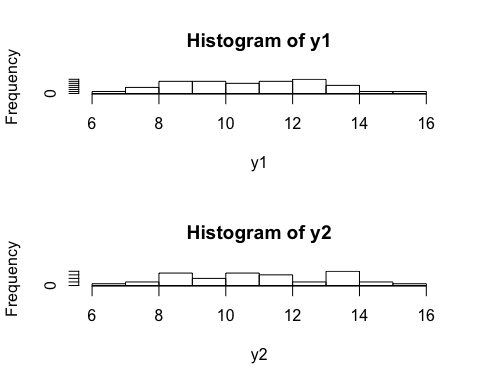
ggplot(data=mysisLengthData, aes(y)) + geom\_histogram() + facet\_grid(net ~ date) +  
 ggtitle("Lengths by net size")

When looking at the distribution of lengths between the two net sizes, we see that the histograms look quite similar across the different dates. What stands out is that the small net catches about a quarter of the number of shrimp, as expected by the difference in net sizes, whereas the general shapes of the histograms appear quite similar across net sizes, suggesting there is not a difference in shrimp length distribution between nets across time.

ggplot(data=mysisLengthData, aes(y)) + geom\_histogram() + facet\_grid(net ~ gender) +   
 ggtitle("Lengths by sex and net size")

When we plot the distribution of lengths between the net sizes with respect to sex class, the distributions look quite similar as well. This suggests that there is not a lot of difference in length distribution of shrimp between the two net sizes when broken down by sex class.

y1 <- by(mysisLengthData$y[mysisLengthData$net == "Small Net"],   
 mysisLengthData$label[mysisLengthData$net == "Small Net"], mean)  
y1 <- y1[!is.na(y1)]  
y2 <- by(mysisLengthData$y[mysisLengthData$net == "Large Net"],   
 mysisLengthData$label[mysisLengthData$net == "Large Net"], mean)  
y2 <- y2[!is.na(y2)]  
pairttestlength <- t.test(y1, y2, paired=TRUE)  
layout(matrix(1:2,2,1))  
hist(y1)  
hist(y2)



From the paired t-test, we find that there is a significant difference ( = 0.069, = 39, -value = 0.945 ) in expected counts, although the -value is close to the 0.05 level. Despite the statistically significant difference, the practical difference is quite small as judged by the histograms and small difference in shrimp size.

To better account for the different sample sizes and other ....

X <- model.matrix(~ net + date + gender, data=mysisLengthData)  
idx <- attr(uniquecombs(X), "index")  
summary(lmmod <- lm(y ~ net + date + gender + station, data=mysisLengthData))

##   
## Call:  
## lm(formula = y ~ net + date + gender + station, data = mysisLengthData)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -15.14 -1.08 -0.06 1.02 10.84   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 15.7599 0.0877 179.68 < 2e-16 \*\*\*  
## netLarge Net 0.1188 0.0428 2.77 0.0056 \*\*   
## dateAugust 0.9915 0.0458 21.64 < 2e-16 \*\*\*  
## dateSeptember 1.0257 0.0484 21.21 < 2e-16 \*\*\*  
## genderJ -8.8020 0.0670 -131.40 < 2e-16 \*\*\*  
## genderM -1.5957 0.0719 -22.19 < 2e-16 \*\*\*  
## genderU -4.2663 0.0710 -60.08 < 2e-16 \*\*\*  
## stationDIL-10 -0.1645 0.0824 -2.00 0.0459 \*   
## stationDIL-2 -0.0105 0.0713 -0.15 0.8833   
## stationDIL-3 -0.5019 0.0674 -7.44 1.1e-13 \*\*\*  
## stationDIL-4 -0.1939 0.0600 -3.23 0.0012 \*\*   
## stationDIL-5 0.1631 0.0790 2.06 0.0391 \*   
## stationDIL-6 0.0950 0.0808 1.18 0.2398   
## stationDIL-7 0.3619 0.0870 4.16 3.2e-05 \*\*\*  
## stationDIL-8 -0.0300 0.0900 -0.33 0.7389   
## stationDIL-9 0.2367 0.0825 2.87 0.0041 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.63 on 9111 degrees of freedom  
## Multiple R-squared: 0.83, Adjusted R-squared: 0.83   
## F-statistic: 2.96e+03 on 15 and 9111 DF, p-value: <2e-16

library(lme4)  
# summary(lmemod <- lmer(y ~ net \* date + net \* gender +  
# date \* gender + (1|station), data=mysisLengthData))  
# summary(lmemod <- lmer(y ~ net + date \* gender + (1|station),  
# data=mysisLengthData))  
# summary(rlmmod <- rlm(y ~ net \* date + net \* gender +   
# date \* gender, data=mysisLengthData))))  
summary(rlmmod <- rlm(y ~ net + date + gender + station, data=mysisLengthData))

##   
## Call: rlm(formula = y ~ net + date + gender + station, data = mysisLengthData)  
## Residuals:  
## Min 1Q Median 3Q Max   
## -15.2223 -0.9932 -0.0204 1.0178 10.8570   
##   
## Coefficients:  
## Value Std. Error t value   
## (Intercept) 15.784 0.082 192.345  
## netLarge Net 0.086 0.040 2.147  
## dateAugust 1.144 0.043 26.683  
## dateSeptember 1.235 0.045 27.307  
## genderJ -8.807 0.063 -140.531  
## genderM -1.508 0.067 -22.412  
## genderU -4.639 0.066 -69.837  
## stationDIL-10 -0.193 0.077 -2.504  
## stationDIL-2 -0.089 0.067 -1.328  
## stationDIL-3 -0.500 0.063 -7.926  
## stationDIL-4 -0.275 0.056 -4.904  
## stationDIL-5 0.070 0.074 0.941  
## stationDIL-6 0.029 0.076 0.377  
## stationDIL-7 0.300 0.081 3.692  
## stationDIL-8 -0.070 0.084 -0.833  
## stationDIL-9 0.089 0.077 1.154  
##   
## Residual standard error: 1.49 on 9111 degrees of freedom

tvalues <- summary(rlmmod)$coefficients[, 3]  
df <- summary(rlmmod)$df[2]  
pvalues <- pt(abs(tvalues), df, lower.tail=FALSE) \* 2

tableData <- cbind(summary(rlmmod)$coeff, pvalues)  
colnames(tableData) <- c("Estimate", "Std. Error", "z value", "Pr(>|z|)")  
rownames(tableData) <- c("Intercept", "Net Size", "August", "September",   
 "Juveniles", "Males", "Unknown Gender",  
 "Station 10", "Station 2", "Station 3", "Station 4",  
 "Station 5", "Station 6", "Station 7", "Station 8",  
 "Station 9")  
  
kable(tableData, digits = 4, format="html")

Although the statistical test is significant, we have very large sample sizes and and the regression coefficients are small relative to the other sources of variation in the model. Of greater interest is whether the observed effect is of practical significance.

SD <- sqrt(((sum(mysisLengthData$net == "Large Net") - 1) \*  
 sd(mysisLengthData$y[mysisLengthData$net == "Large Net"])^2 +   
 (sum(mysisLengthData$net == "Small Net") - 1) \*  
 sd(mysisLengthData$y[mysisLengthData$net == "Small Net"])^2) /   
 (sum(mysisLengthData$net == "Large Net") +   
 sum(mysisLengthData$net == "Small Net")))  
  
d <- (mean(mysisLengthData$y[mysisLengthData$net == "Large Net"]) -   
 mean(mysisLengthData$y[mysisLengthData$net == "Small Net"])) / SD

We see that there is a significant effect of net size on mean length caught (p = 0.032) after controlling for date, gender class, and sampling location. Although this is statisticallly significant, the effect size is small (0.086mm) and the sample size is large (n = 9127). Given a large sample size, a hypothesis test will show staistical significance unless the population effect size is exactly zero (this explains why all of the p-values in the table above are less than 0.05). Therefore, the practical effect of a difference in mean length of 0.086mm on a species with a mean length of 10.351mm is small (this is the smallest effect of all the effects estimated by almost a factor of two) and a difference in means of this size is not of practical interest. A measure of effect size is Cohen's which measures the difference in means relative to a pooled standard deviation. For our data, Cohen's =0.075, which implies that the effect of net size is quite small in terms of practical significance.

## Another issue to consider is the catch of a few large and small individuals in the larger nets. You should be able to test if that is simply due to the larger sample size obtained in the larger net. You suggest that may be the case.

### Finally, the length analysis is using every measured mysids as an independent sampling unit. I think it would be more appropriate to use a paired t test on the mean lengths for that comparison as well, using each net instead of each mysid as the sampling unit. There is always the possibility that different size groups occur in different areas of the reservoir, and therefore that each mysids measured are not independent from the other mysids measured in a given net haul. Please consider that in your response.

Lars

## Reviewers' comments:

## Reviewer #1: Review of GLR-D-15-00174

### 8) Line 145, 150 I'm not sure you controlled for these factors, rather it seems like they were factors in the analysis?

### 9) Line 149 I don't think the mean length is the correct variable of interest. The size distributions could be very different and provide similar mean lengths. Rather, I think an analysis of the size structure is appropriate, perhaps using some sort of contingency table to evaluate the homogeneity of proportions. I also think the analysis of population structure (male-female, juvenile) and male:female ratio would be more appropriately done by testing for homegenity of proportions and then looking at standardized deviation to see where the nets differed (if at all).

### 10) The analysis of abundance (and other factors) was done basically to see if this study was done with one net or the other, would the results over the reservoir differ. This is an interesting question and an important one to address. But the net tows were done in paired combinations, so there is more information here. I would like to see something like a paired t-test (using each individual set of net tows as a sample), or a regression of values from one net plotted against the other from each sample to see if there was any bias based on the paired tows.

### 11) Line 163 and 166 I think you mean density and not number?

### 12) Please provide F values, df so the stats can be evaluated

### 13) Line 177 also caught more of the smallest individuals

## Reviewer #2: This manuscript has one straightforward message that will be useful to researchers interested in the collection of mysids, in that it supports the use of smaller diameter nets, which will be convenient from small vessels often used in lakes these animals have spread into. Overall this manuscript is acceptable for publication with some revisions. The details of the statistical tests need to be expanded upon and some language that downplays some of the size-distribution differences with different net sizes should be discussed in greater detail.

Major points:

### 2) The manuscript is short so you have plenty of space to expand upon your statistical methodology which isn't made clear, and your presentation of only p values is insufficient to assess your analysis.

### 4) Downplaying the differences in size-distribution associated with the net sizes doesn't make the paper more persuasive. There are plenty of reasons where researchers may target the tiny fraction of smallest or largest individuals which are being missed by the smaller nets. Discuss this difference and end with the point you make that if possible, use the larger nets, but for most work, the smaller nets are an acceptable substitute.

Detailed comments: L38 Only 3 dates were sampled.

### L40 Size structure was slightly different (more at tails of size-distribution for large nets). Without this fact the last sentence of the abstract may not be clear.

### L122 use ind or N rather than lowercase n

### L130 use s or seconds

### L141 Be consistent N/m2, number/m2

### L142-151. You can be more concise here for some content and have to expand on other. You always have the same effects so you don't have to relist them. State the data isn't normal so a negative binomial transformation was used for regression on abundance, sex ratio…etc. See next note regarding statistical methods. You would be better off explaining the rationale for methods and the format in more detail rather than just the factors. For instance there is no detail given on the negative binomial model which can be constructed and assessed in a range of ways (likelihood format, AIC etc). Be descriptive here. If you used a specific stat software package, this would be the place to note it.

### L145-148 Why wouldn't you use a paired test? You collected each at the same time, so you could control much of the variability in catch (patchiness) by comparing them to each other (you comment L153-155 on this). Your size data shows multiple cohorts so it isn't surprising that you have non-normal data but the distributions are similar so should be easily discriminated using pairwise testing.

### L162-170. You give p values but no specifics as to method. The negative binomial model requires a range of output to assess the fit. The p values by themselves are not necessarily relevant so F (or z) values or r-mean squares should be included as well.

### L156. It is unusual for an author to cite unpublished data on their own paper. Ideally, you could easily add an additional graph that shows the series given the few number of tables and figures.

L161. Are you alluding to a drawback in your methods here given the lack of identification of sex increasing during the study? ### L167 Don't repeat yourself. "between net size and sampling date" redundant. Instead, "effect or interactions". ### L168 As above, you can include all of these as one statement: "density or juvenile proportion". L173 "likely not of practical interest". ### L175. So did you check for homogeneity of variance? A larger range but similar mode suggests that the larger nets had increased variance. This doesn't necessarily affect your results but should be noted. This might be relevant for any attempt at bioenergetics modelling for instance. The quartile plots in Fig. 2 suggest that the smaller net mean lengths are skewed downward (lower quartile larger than upper) which is reversed in the larger nets (except in July, which is more likely driven by the bimodal distribution in this month and suggests (there is no testing?) that the distributions are not normal and means are unlikely to be relevant for this month). You might be better off testing the bimodal distribution as two distributions (mode to mode). This may be a result of the different efficiencies of the two net sizes (you make no note of flowmeter readings and in your next section you calculate efficiency using areal ratio). It also might be worth more text describing the different distributions of %juveniles since the august and September distributions are very different even though their modes are similar (the distribution tail).

### L183. "probably not biologically relevant". Since you do not give these distributions (beyond the graphs), this isn't necessarily your call since some researchers may be focused on the very tails of the distribution that you dismiss. Instead, note that they are very small differences in numbers and that researchers interested in these larger individuals may wish to utilize the larger nets. Instead, move the sentence that is "Thus, these two nets…" up to be the second sentence (you state that there is essentially no difference in their distribution), then follow up with the caveat with respect to the larger size distribution afterwards.

### L187. The slight bias where the smaller nets are missing the largest individuals may by more relevant in late October or November when many programs sample to get egg production rates. Larger individuals appear to be able to avoid the smaller nets (they are certainly more powerful swimmers) and may also detect a difference in the net-mouth "head-wave" if your efficiencies are different.

### L203. As noted previously, be careful with your wording. Your equipment is similar to certain studies, but it can just as easily be said that your mesh size (for instance) is smaller than the majority that specifically are used to target mysids. Your comment on the size preventing loss of small mysids may be spurious. Given your distributions (and mean size of ~10mm) did you even see any mysids less than 1 mm in size (or even up to 3 mm)? Fig. 3 suggests you didn't, so you shouldn't expect net extrusion at a 3:1 ratio. If the mesh size was used to be consistent with other monitoring methods, then just state that, which is a perfectly fine reasoning and supported by your results.

L214 "Our findings…" ### L217 "While the larger net captured more…" - you have waited until here to note that there are two possible reasons (increased sampling area or reduced net avoidance). This should be in your first sentence of your discussion since it is the one factor which is clearly different between the sampling methods. Your second point about the large net also catching larger numbers of smaller individuals only suggests a sampling area effect (wording should be "did not support"). Both factors can be happening with the different size animals - larger mature adults are significantly more powerful swimmers than the smaller juveniles. Your last point in this paragraph should be expanded upon to note the comments listed above (that some researchers may be targeting the largest individuals to estimate population growth rates). L229. Remove "Conclusions" - it is the last paragraph anyway. L230. The term "biologically relevant" is loaded and likely to create disagreement even if there isn't any in the findings. You don't know what is biologically relevant to mysids or to other trophic levels that are affected by them (e.g. large individuals are known to have a significant impact on trophic dynamics beyond just their biomass (a la Fry and Quinones 1994, Ellis et al. 2011) and the presence of mysids will shift size structure of the entire plankton when present. Size is important. L236. This sentence essentially repeats the last in the previous paragraph. Combine them and expand on as noted in the L217 comment.

Table 1: You can save a huge amount of space in your table by just using "Conical" instead of "Conical plankton" and "Pyramid" instead of "Inverted pyramid" on two lines and note in the caption (you should also explain "framed net"). This table can easily be portrait with this change thus freeing up more space. This table is a major source of your extremely large reference section. This would be better suited for a supplementary appendix since it is not directly essential to the finding of this study.

Fig. 1. If you are going to note lat-long lines on your map, you should use a different line marking (e.g. dashed). Fig. 2 and Fig. 3 captions don't match your graphs because the next two figures pages are reversed. Fig. 2. The dots are means and the lines are modes? This isn't made clear. Axes titles may have to be larger. Fig. 3 The 1.0 m diameter net bars should be white to be consistent with the other graph. Font may have to be larger for the axes and the inset pie-graphs.

This document shows the analyses performed in Johnson et. al.

## Model for counts by sex

The next question we wish to explore is whether the total counts when grouped into sex classes of male, female, juvenile, and unknown vary with net size or other covariates. To do this we use () above. We begin by loading the data where counts are divided into sex classes.

## load data  
load("~/mysis/data/mysisSexCountData.RData")   
## construct model matrix for covariates  
X <- model.matrix(~ net \* gender + net \* date + gender \* date-1,   
 data=mysisSexCountData)  
## fit model  
nbmod <- glm.nb(y ~ ., data = data.frame(y=mysisSexCountData$count, X))

To explore this data in more detail we examine two sets of histograms, the counts of each sex class broken down by date in figure1 and the counts of each sex class broken down by net size in figure2. From these we see...

ggplot(data=mysisSexCountData, aes(count)) +   
 geom\_histogram() + facet\_grid(net ~ gender) +   
 ggtitle("Histogram of normalized counts by sex and net size")  
ggplot(data=mysisSexCountData, aes(count)) +  
 geom\_histogram() + facet\_grid(gender ~ date) +   
 ggtitle("Histogram of normalized counts by sex and date")

To test for a change in counts broken down by gender category between the two net sizes, we construct a negative binomal regression model that examines the effects of net size, date of sampling, the interaction between net size and sampling date, and the interaction between sampling date and gender. The results shown below show that there is no evidence of an interaction between net size and sampling date and no evidence of an interaction between net size and gender.

tableData <- cbind(summary(nbmod)$coeff)  
colnames(tableData) <- c("Estimate", "Std. Error", "z value", "Pr(>|z|)")  
rownames(tableData) <- c("Intercept", "Net Size", "Females", "Juveniles",   
 "Unknown Gender", "August", "September",   
 "Net Size \* Females", "Net Size \* Juveniles",   
 "Net Size \* Unknowns", "Net Size \* August",  
 "Net Size \* September",  
 "Females \* August", "Juveniles \* August",   
 "Unknown Gender \* August",  
 "Females \* September", "Juveniles \* September",   
 "Unknown Gender \* September")  
kable(tableData, digits = 4, format="html")

Thus, we refit the model without the interaction between net size and sampling date. To do this we start by loading the new covariate matrix .

On this reduced model, the results shown below show that there is no evidence of an effect on the number of counts observed between the two net sizes (p=0.636).

tableData <- cbind(summary(nbmod)$coeff)  
colnames(tableData) <- c("Estimate", "Std. Error", "z value", "Pr(>|z|)")  
rownames(tableData) <- c("Intercept", "Net Size", "Females", "Juveniles",   
 "Unknown Gender", "August", "September",   
 "Net Size \* Females", "Net Size \* Juveniles",   
 "Net Size \* Unknowns",   
 "Females \* August", "Juveniles \* August",   
 "Unknown Gender \* August",  
 "Females \* September", "Juveniles \* September",   
 "Unknown Gender \* September")  
kable(tableData, digits = 4, format="html")

## Model for proportion of juveniles

ggplot(data=data, aes(prop)) + geom\_histogram() + facet\_grid(net ~ .) +   
 ggtitle("Histogram of proportion of juveniles by net size")

To test for a change in the proportion of juveniles caught between the two net sizes, we construct a beta regression model that examines the effects of net size, date of sampling and the interaction between net size and sampling date. The results shown below show that there is no evidence of an effect on the proportion of juveniles caught between the two net sizes (p=0.783).

tableData <- cbind(summary(brmod)$coeff$mean)  
colnames(tableData) <- c("Estimate", "Std. Error", "z value", "Pr(>|z|)")  
rownames(tableData) <- c("Intercept", "Net Size", "August", "September",   
 "Net Size \* August", "Net Size \* September")  
kable(tableData, digits = 4, format="html")

## Length model

ggplot(data=mysisLengthData, aes(y)) + geom\_histogram() + facet\_grid(net ~ date) +  
 ggtitle("Lengths by net size")  
ggplot(data=mysisLengthData, aes(y)) + geom\_histogram() + facet\_grid(net ~ gender) +   
 ggtitle("Lengths by sex and net size")  
ggplot(data=mysisLengthData, aes(y)) + geom\_histogram() + facet\_grid(date ~ gender) +   
 ggtitle("Lengths by sex and date")