AACL BIOFLUX

Aquaculture, Aquarium, Conservation & Legislation
International Journal of the Bioflux Society

Zeroes do matter: the tale of the missing fishes in Cabilao Island Bohol, Philippines

Erwin O. Brunio[1], Joeppette J. Hermosilla[1], Filipina B. Sotto[2], and Eiji Tanaka[3]

[1] Graduate School of Marine Science and Technology, Tokyo University of Marine Science and Technology, Konan, Minato, Tokyo, 108-8477, Japan; [2] Biology Department, University of San Carlos, Nasipit, Talamban Cebu City, 6000, Philippines; [3] Resources Analysis Laboratory, Tokyo University of Marine Science and Technology, Konan, Minato, Tokyo, 108-8477, Japan. Corresponding author: E. Tanaka, hermit@kaiyodai.ac.jp

Abstract. Data from plankton surveys typically contain many zeroes. Pomacentrid larvae collected from the reef of Cabilao Island, Bohol, Philippines was typical of such data wherein 294 out of 354 tows had 0 larval counts. This data is traditionally analyzed using the Poisson distribution in linear model. However, the Poisson model predicted far less zeroes and more missing pomacentrids than what is observed. To account for this kind of data, this paper examined other statistical distributions that can account for the extra zeroes and the missing fishes; these are negative binomial, zero-inflated poisson, zero-inflated negative binomial, zero-altered poisson and zero-altered negative binomial. We fitted these distributions and Poisson with location and moon phase as covariates to identify which of the model is suited for the data. The best model was selected based on having the lowest Akaike`s Information Criteria (AIC) score. The AIC result showed that the zero-altered negative binomial distribution (ZANB) with moon phase as covariate was suitable for the current data. The ZANB model showed that zeroes contribute a significant effect on abundance estimate; higher number of zeroes accounted for the significantly lower fishes observed in the first quarter than in the last quarter but not on the other moon phases. Excluding the zeroes, this difference is not statistically significant. Moreover, the ZANB model also was able to predict the actual number of zero and non-zero fishes better than the Poisson model, thus the tale of the missing fishes is explained by the inappropriateness of Poisson to fit data with many zeroes.

Key Words: Zero count, zero-inflated, zero-altered negative binomial, hurdle, Philippines.

Introduction. Count data from plankton surveys typically contain many zeroes. This data is referred to as “zero inflated” because the number of zeroes is so large (Martin et al 2005; Tu 2002; Heilbron 1994). The presence of many zeroes in the count data is a major challenge among plankton ecologists in estimating plankton abundance primarily because high zero counts may violate the assumption from the probability distribution used thereby resulting to incorrect interpretation of data. Additionally, the large number of zeroes created uncertainty in parameter estimation because these zeroes most likely decrease the effective sample size.

The zeroes in the data can be categorized as “true” zeroes or “false” zeroes depending on the source (Martin et al 2005). The “true” zeroes happen when species were not observed either because the site was not suitable or the site was not entirely saturated with the said species. Conversely, the “false” zeroes occur when species inhabit the site but were not present during the survey or was present during the survey but was not observed.

Count data from plankton survey, particularly ichthyoplankton, were most often analyzed using the Poisson distribution in linear model. However, the use of such approach may lead to incorrect inference considering the high number of zero counts in the data. In Figure 1a, the Poisson model predicted far less zeroes and more missing fishes than what was observed in the data. In this situation, the probability distribution base on the Poisson model may be inappropriate if the zeroes are coming from both sources (Hilbe 2011).

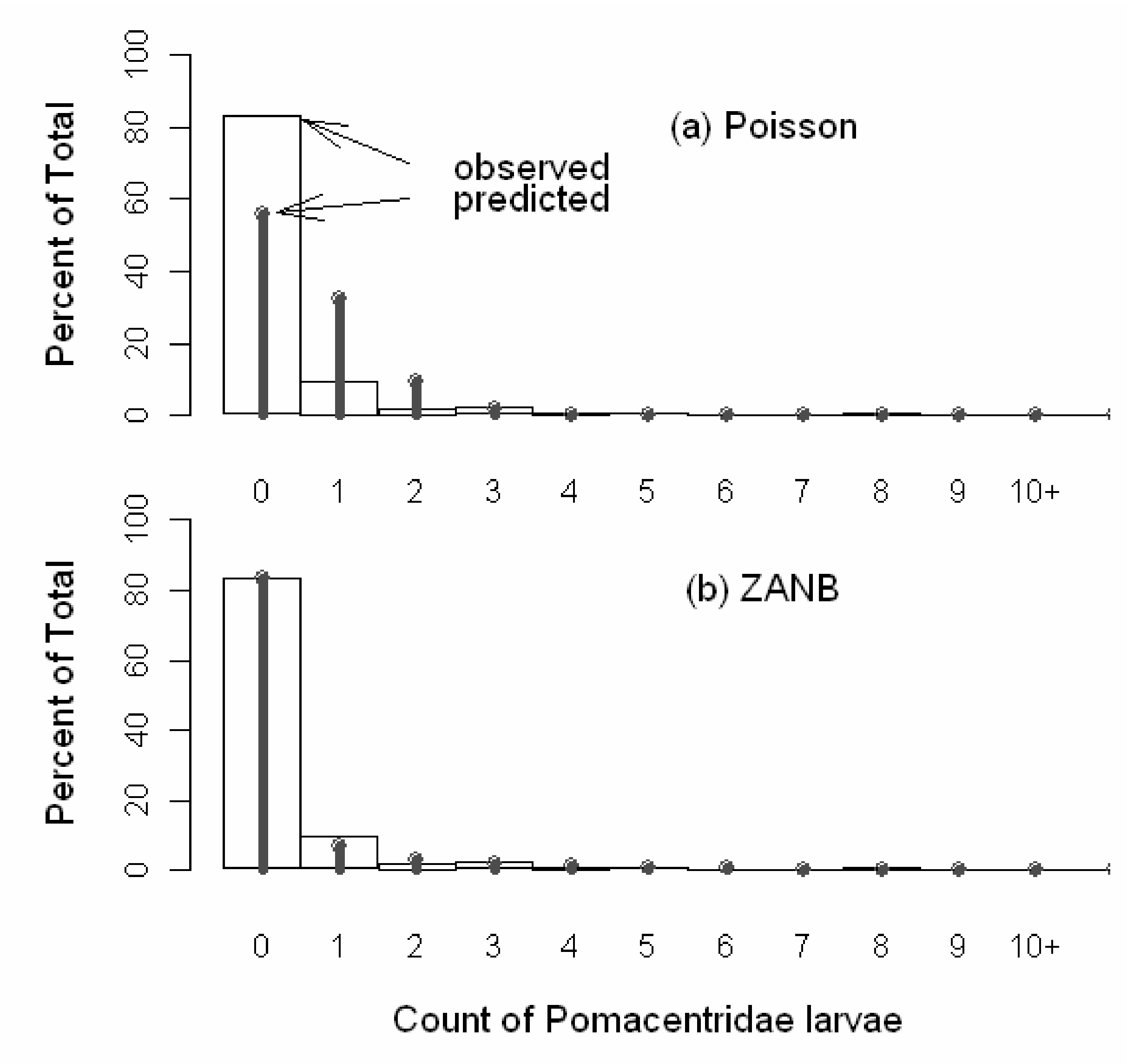

Figure 1. Histograph of Pomacentridae larvae, with predicted count based on a) Poisson and b) Zero-Altered Negative Binomial (ZANB) models. Model fitting using the Poisson distribution resulted to an underestimation of the percentages of zeroes as well as overestimation of some non-zero counts as compared with the zero-altered negative binomial distribution. Note that fish abundance data greater than 10 (max=67) were excluded in the figure for clarity.

This study attempted to describe the abundance of ichthyoplankton, particularly the pelagic larvae of pomacentrids in Cabilao Island, Bohol, Philippines with substantial number of zero counts in the data. The objective is to introduce alternative statistical approaches suitable for ichthyoplankton studies with inherent “zero-inflated” nature in their count data. Since abundance estimates of ichthyoplankton is beneficial in the assessment of larva supply and recruitment episodes for stock assessment and management, inference demands high level of confidence from the results of estimation.

Materials and Methods

Data collection. Planktonic larvae of pomacentrids from Cabilao Island, Loon, Bohol, Philippines were collected with a ring-net (500-micron nylon mesh, 0.74m^2 mouth area, 3.2 meters total length) that was towed parallel to the reef crest for eight minutes from a vessel traveling at a constant speed. The volume of water filtered by the net was measured using flowmeter. Samples were preserved in 4% borax-buffered formaldehyde and later identified, sorted and counted in the laboratory using a binocular dissecting microscope.

Sampling was conducted in three locations along the coasts of Cabilao Island particularly in Cambaquez fishing ground, Cabacongan Marine Protected Area, and Pantudlan Marine Protected Area (Figure 2). For each location, sampling was conducted at different moon phases (e.g., new moon, first quarter, full moon and last quarter). Although the current study also collected ichthyoplankton samples at night, nighttime samples were excluded in the analysis (n=23) considering that these were performed only during the full moon phase of the lunar cycle. A total of 354 plankton tows were conducted. In the actual data analysis, emphasis was on the ichthyoplankton data from family Pomacentridae considering that this group is an important component of the coral reef fish community of the island.

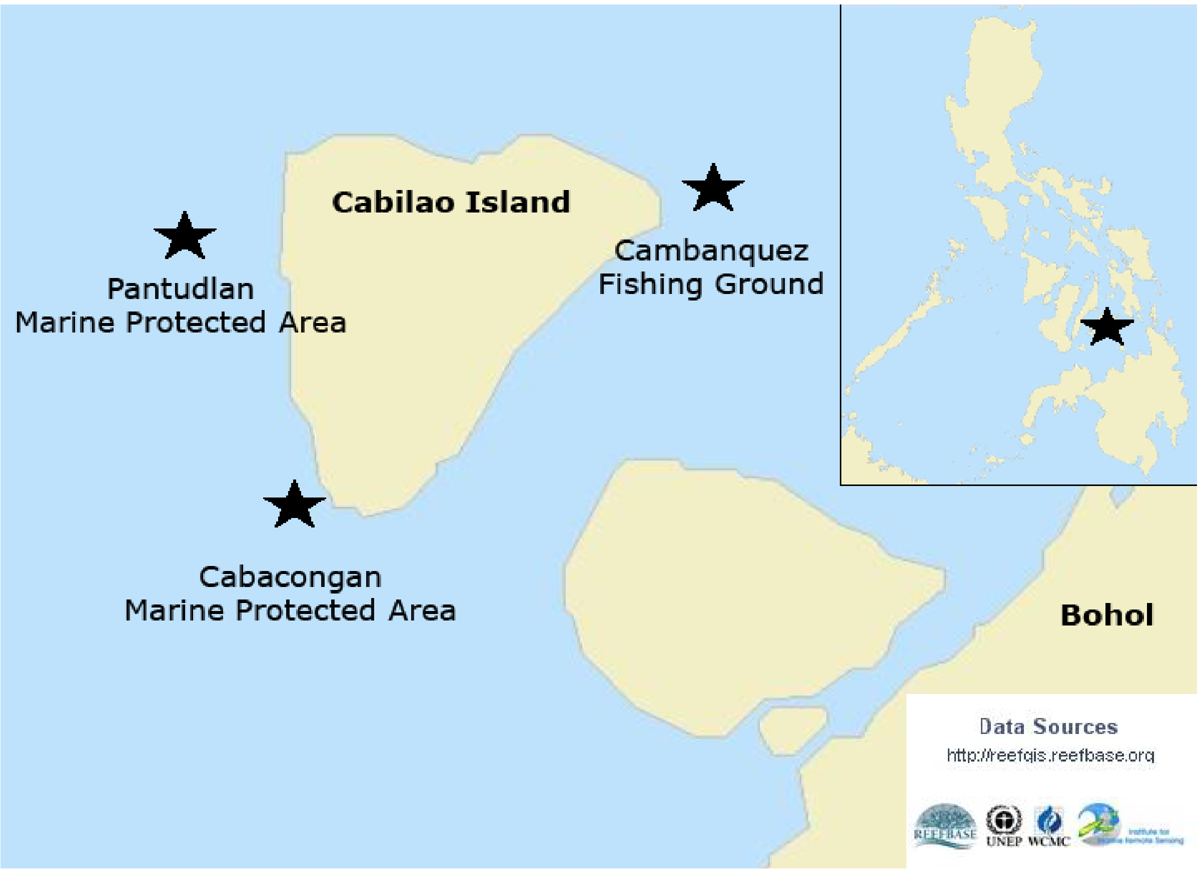

Figure 2. Map of sampling locations in Cabilao Island, Bohol, Philippines (redrawn from map provided by ReefBase).

Selection of probability distribution. Pomacentrid larval data was fitted using the maximum likelihood approach with the Poisson and 5 statistical distributions that can account for extra zeroes particularly the negative binomial, zero-inflated Poisson, zero-inflated negative binomial, zero-altered Poisson and zero-altered negative binomial. The Poisson distribution is the standard approach to analyze discrete count data such as the data from plankton tows. The Poisson is described by the probability density function (PDF):

$$\Pr(Y = y) = \frac{e^{-\mu}\mu^{y}}{y!} \qquad \text{eq. 1}$$

where μ is the mean and y is the count of fish per plankton tow. The Poisson distribution assumes that the variance is equal to the mean. This assumption is commonly often violated in ecological data where variance usually exceeds the mean (over-dispersion). Over-dispersion occurs especially if the data are zero-inflated because single distribution such as Poisson assumes that the zero values are only coming from one source, the true zeroes (Cameron & Trivedi 1998).

An alternative statistical model for count data is the Negative Binomial. The Negative Binomial assumes that zeroes are only true zeroes, but could accommodate higher number of zeroes than the Poisson distribution. The PDF is given by:

$$\Pr(Y=y)=\frac{\Gamma(y+\theta)}{\Gamma(\theta)y!}\cdot\frac{\mu^{\theta}\theta^{\theta}}{(\mu+\theta)^{y+\theta}} \qquad \text{eq. 2}$$

where μ is the mean and θ is a scaling parameter to accommodate extra dispersion in the data (Zeileis et al 2008; White & Bennetts 1996). The negative binomial conforms to empirical plankton studies that show the variance exceeding the mean (Power & Moser 1999; Cyr et al 1992).

Data that are over-dispersed and with extra zeroes that cannot be explained by the Poisson or Negative Binomial distributions can be modeled by mixture models. Mixture models are two-component models combining a point mass at zero and a count distribution to account for extra zeroes in the observations (Zeileis et al 2008; Lambert 1992). While zeroes from the Poisson and Negative Binomial are true zeroes, zeroes in the mixture models come both from the true zeroes and the false zeroes. The false zeroes is fitted with a Binomial distribution as the first component in the data analysis. The second component fits both the true zero and the non-zero data either with a Poisson or Negative Binomial distribution. The general form of the mixture model is:

$$\Pr(Y=y)=\begin{cases} p+(1-p)\cdot g(y) & y=0 \\ (1-p)\cdot g(y) & \text{otherwise} \end{cases} \qquad \text{eq. 3}$$

where p is the probability that a species is absent, and $g(y)$ is a known PDF describing the non-zero observation (Maunder & Punt 2004). Mixture models are also known as zero-inflated models and specifically referred to as Zero-Inflated Poisson (ZIP) or Zero-Inflated Negative Binomial (ZINB) if $g(y)$ is described by the Poisson or negative binomial distribution, respectively.

Another approach in modeling over-dispersed data with extra zeroes is the use of two-part models. Two-part models are two-component models that treat the modeling of the zeroes separately from the non-zero values. The first component models the probability of the data being zero (absence) with a Binomial distribution. The second component models the non-zero values only using a zero-truncated probability distribution. The distribution is given by:

$$\Pr(y=y)=\begin{cases} p & y=0 \\ (1-p)\cdot g(y) & \text{otherwise} \end{cases} \qquad \text{eq. 4}$$

where p is the probability that zero fish will be observed (absence), y is the density of fish (number/200m^3) and $g(y)$ is the truncated probability distribution of a positive variable given that at least 1 fish was observed (Maunder & Punt 2004; O'Neill & Faddy 2003). The truncated probability function is given in equation 5a and 5b for the Poisson and negative binomial, respectively:

$$g(y)=\begin{cases} \dfrac{e^{-v}v^{y}}{(1-e^{-v})y!} & a \\[2ex] \dfrac{\Gamma(y+\theta^{-1})}{\Gamma(\theta^{-1})y!}\cdot\dfrac{(\theta v)^{y}(1+\theta v)^{-(y+\theta^{-1})}}{1-(1+\theta v)^{-\theta^{-1}}} & b \end{cases} \qquad \text{eq. 5}$$

In the equation above, y is the non-zero density of fish (number/200m^3), v is the conditional mean value given that at least 1 fish was observed, and θ is a dispersion parameter. Two-part model is also known as zero-altered model and referred to as Zero-

Altered Poisson (ZAP) or Zero-Altered Negative Binomial (ZANB) if the count distribution comes from the truncated Poisson or the truncated negative binomial respectively. Zero-altered model is also referred as a hurdle model in the literature because a hurdle must be overcome for positive non-zero values to occur (Maunder & Punt 2004; Mullahy 1986).

In this study, the ichthyoplankton data was modeled based on the four approaches with six statistical distributions: Poisson (P), Negative Binomial (NB), Zero-Inflated Poisson (ZIP), Zero-Inflated Negative Binomial (ZINB), Zero-Altered Poisson (ZAP) and Zero-Altered Negative Binomial (ZANB). Modeling with the said probability distributions was implemented together with the covariates.

Selection of covariates. To find any variation in the abundance of pomacentrid larvae in Cabilao Island in relation to the lunar cycle (moon) and location (area), these covariates were included as explanatory variables in the model-fitting process. Moon was categorized into 4 levels (e.g., new moon, first quarter, full moon, and last quarter) while area was categorized into 3 levels (e.g., Pantudlan, Cabacongan and Cambanquez). These covariates were structured as additive models (Ellison 2004) and fitted to the data with number of fish larvae observed per 200m^3 as the response variable. The compound combination of covariates referred to as the maximal model is given by:

$$\log\left(\frac{\text{Count}}{\text{Volume}}\right) = \text{Intercept} + \text{Area} + \text{Moon} \qquad \text{eq. 6}$$

where Volume is included as an offset variable to account for differences in the amount of water filtered by the plankton net (min=43.66 m^3, max=302.55 m^3, mean=200.75 m^3). The area and moon interaction was excluded due to unbalanced design. For two-component models (e.g., ZIP, ZINB, ZAP, ZANB), the 1st component (binomial component) is also fitted with the same covariates in the same manner with the maximal model given by:

$$\log\left(\frac{p}{1-p}\right) = \text{Intercept} + \text{Area} + \text{Moon} \qquad \text{eq. 7}$$

There were 4 covariate combinations for single distributions (e.g., P and NB) – area and moon, moon only, area only, and no covariate. There were 16 possible combinations of covariates for two-component models (e.g., ZIP, ZINB, ZAP and ZANB) with 4 combinations in each component.

Model selection. To identify which of the six distribution best fit the pomacentrid data, the Akaike`s Information Criteria (Akaike 1974; Burnham & Anderson 2004) was used. Akaike's Information Criterion (AIC) is given by:

$$\text{AIC} = -2\ln(L(\hat{\theta})) + 2K \qquad \text{eq. 8}$$

where $\ln(L(\hat{\boldsymbol{\theta}}))$ is the natural logarithm of the likelihood function evaluated at the maximum likelihood, and *K* is the number of parameters estimated in the model. Model with the minimal AIC score was selected as the "best" approximating model (Franklin et al 2000; Hobbs & Hilborn 2006; Burnham & Anderson 2004; Ellison 2004). AIC have been used to discriminate among probability distribution for identical set of regressors (Dick 2004) and in covariate selection (Burnham & Anderson 2004; Rooper & Martin 2009; Ellison 2004). These approaches were combined by simultaneously discriminating between probability distribution and covariate together. A total of 72 models were examined from these combinations.

Calculation of AIC scores and the corresponding maximum likelihood estimate were performed using the statistical computing language R (R Development Core Team 2011). Fitting for Poisson model was implemented using the stat package (R Development Core Team 2011), Negative Binomial with the MASS package (Venables & Ripley 2002), and two-component models with the PSCL package (Jackman 2011; Zeileis et al 2008). Confidence interval for the fitted values was calculated from bias-corrected accelerated percentile interval (BCa) non-parametric bootstrap with 1999 replicates using the boot package (Fox 2002; Davison & Hinkley 1997; Canty & Ripley 2011).

Results and Discussion. AIC result for different combinations of covariates for each statistical distribution (Table 1) showed that the ZANB distribution had the lowest score of 532.9 (ZANB model 15), indicating that the said distribution model was more suitable with the pomacentrid data. Conversely, the Poisson model performed poorly among the six statistical distribution models considering that it attained an AIC score of 1199.4 (Poisson model 3). The pomacentrid larvae collected by plankton tows near the coast of Cabilao Island is typical of ichthyoplankton data with substantially high zero count in addition to the fact that the count data itself is inherently positively skewed. Of the 354 plankton tows, about 72-89% had zero pomacentrid larvae. It is expected that the Poisson model is inadequate in dealing with this kind of data as it cannot accommodate over-dispersed or zero-inflated data (Lewin et al 2010; Potts & Elith 2006). Our result confirm such observation and found that the ZANB was a much better model in dealing with zero-inflated data as compared to the Poisson model (Table 1).

Table 1

AIC scores for different combinations of covariates for each statistical models (P=Poisson, NB=Negative Binomial, ZIP=Zero-Inflated Poisson, ZINB=Zero-Inflated Negative Binomial, ZAP=Zero-Altered Poisson, ZANB=Zero-Altered Negative Binomial)

Model	*Covariates*		*Probability distributions*					
	Zero vs Non Zero	*Count*	*P*	*NB*	*ZIP*	*ZINB*	*ZAP*	*ZANB*
1		(None)	1319.9	549.6				
2		Area	1302.0	549.6				
3		Moon	1199.4	534.3				
4		Moon + Area	1181.8	537.8				
5	(None)	(None)			901.6	551.6	901.5	546.7
6	(None)	Area			859.8	551.6	859.2	542.6
7	(None)	Moon			833.7	536.3	839.4	537.2
8	(None)	Moon + Area			819.5	539.8	821.5	539.5
9	Area	(None)			904.8	555.1	904.6	549.9
10	Area	Area			861.9	550.9	862.3	545.7
11	Area	Moon			836.9	540.0	842.5	540.3
12	Area	Moon + Area			821.1	543.4	824.6	542.6
13	Moon	(None)			897.5	549.9	897.3	542.5
14	Moon	Area			855.6	549.4	855.0	538.4
15	Moon	Moon			836.1	538.9	835.2	532.9
16	Moon	Moon + Area			821.1	541.3	817.3	535.3
17	Moon + Area	(None)			900.7	550.9	900.4	545.7
18	Moon + Area	Area			857.6	544.4	858.1	541.6
19	Moon + Area	Moon			839.3	537.1	838.3	536.1
20	Moon + Area	Moon + Area			823.0	538.8	820.4	538.4

The ZANB model is not popularly employed in statistical analysis dealing with ichthyoplankton due to its complicated formula. However, the recent developments and availability of statistical programs makes it feasible to fit the ZANB distribution model to count data with many zeroes. The ZANB model had been shown to better explain the abundance of recreational catch data (O'Neill & Faddy 2003) and in temperate river fishes (Lewin et al 2010). Correspondingly, the result of the current study also suggested that the use of ZANB model was suitable for count data of ichthyoplankton with many zeroes.

Lunar cycles are thought to regulate the spawning episodes as well as the subsequent dispersal of planktonic larvae of some marine fishes because of the tide and current regimes associated with it. As for the pelagic fish eggs and larvae, most often the pelagic dispersal is properly timed to assist its dispersal away from reef habitats to avoid predators and to colonize wider areas. For example, the spawning of *Centropyge potteri* (Jordan & Metz, 1912) (Pomacanthidae) occurred at dusk during the week preceding full moon where the difference in the height of the tide from high to low was greatest (Lobel

1978). Fishes under families Leptocephalidae, Bothidae and Labridae also showed positive correlation between hours of dark flood over new moon (Thorrold et al 1994). In *Thalassoma bifasciatum* (Bloch, 1791) (Labridae), larval settlement occurred around new moon (Victor 1986). In the family Pomacentridae, spawning and settlement also followed lunar cycles, as observed in *Stegastes partitus* (Poey, 1868), according to Robertson et al (1988) and *Amphiprion melanopus* Bleeker, 1852; peak in hatching of *A. melanopus* occurred 1.5 hours after sunset near full moon and new moon when the tide was high and the current was strong (Ross 1978).

The result of the current study confirmed the differential effect of the moon on the abundance of pomacentrid larvae in the neritic waters of Cabilao Island. In the best model (ZANB model 15, Table 1), the covariate moon has a significant effect in both the absence-presence ($\chi^2=10.22$, $df=3$, $p<0.01$) and non-zero count ($\chi^2=15.58$, $df=3$, $p<0.001$) components of the ZANB.

In the absence-presence component, the percentage of zero count was higher during the first quarter (89.36%, CI=82.01–95.25) than in the first last quarter (72.37%, CI=61.95–81.82) and considering that their confidence interval did not intersect, such result was statistically significant (Figure 3a). On the other hand, the percentage of zero in other moon phases was similar.

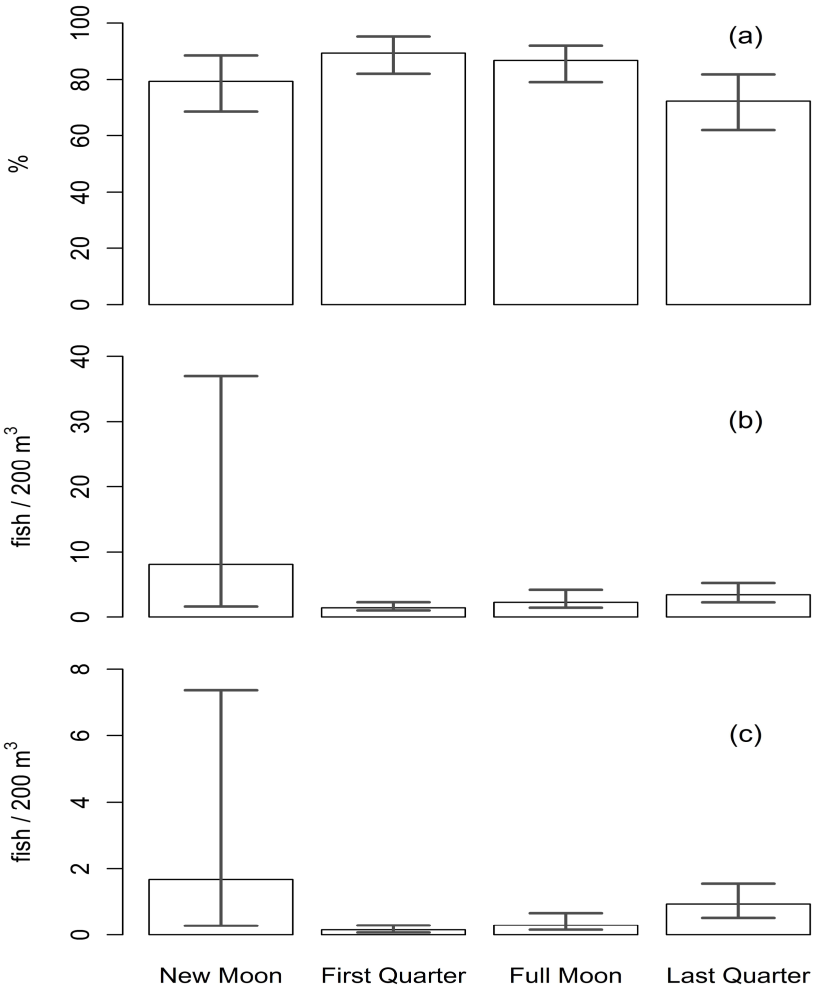

Figure 3. Predicted abundance of pomacentrid larvae with bootstrapped 95% confidence interval for a) percentage of zero count, b) non-zero density by excluding the zeroes, and c) over-all density including the zeroes.

When the zero count of the data were deliberately excluded in the analysis (e.g. non-zero count component of ZANB), the abundance of pomacentrid larvae among the four moon phases was statistically the same (Figure 3b). Conversely, by including these zeroes (e.g. by combining the result of the two components), the result was fairly different with cumulative abundance of pomacentrid being significantly higher in the last quarter (0.94 fish/200m^3, CI=0.52–1.55) than in the first quarter (0.15 fish/200m^3, CI=0.07–0.29). Abundance of pomacentrids larvae during full moon (0.30 fish/200m^3, CI=0.16-0.66) and new moon (1.67 fish/200m^3, CI=0.28-7.37) were not statistically different from the other moon phases because of high variability in the data (Figure 3c).

Most of the research conducted in the Philippines that deal with plankton focused on ecological characterization, such as in identification and description of species (Armada 1997; Blanco & Villadolid 1951; Wade 1951). Estimation of abundance are few, and if available is limited to over-all abundance, with no specific estimate for particular species or families. For example, Armada (1997) estimated that over-all plankton concentration ranges from 4 to 1,420 larvae per 100m^3 in the Sulu Sea, while Dolar & Alcala (1993) reported density range from 0 to 56 per 1m^3. This may be attributed to the difficulty in analyzing plankton count data with many zeroes. If researchers wanted to analyze abundance at the species or family level, the number of zeroes increases and estimation become complicated.

The result of this study suggests that the use of statistical models such as zero-altered negative binomial distribution can account for plankton count data with many zeroes. The selected ZANB model was able to predict the actual number of zero and non-zero fishes better than the Poisson model, thus the tale of the missing fishes is explained by the inappropriateness of the later to fit data with many zeroes. Moreover, the zero-altered models do not require the separation of the sources of the zeroes prior to analysis. Such separation is often difficult to meet in ecological studies dealing with ichthyoplankton due to limitations brought about by field conditions regardless of how careful the researchers are in collecting samples from the field. Finally, by separately modeling the zeroes and non-zeroes, it allows researcher to identify covariates that influence whether a species will be present or not as well as covariates that affect its abundance given it was present as demonstrated from the pomacentrid data.

Conclusions. Our result showed that zeroes did matter in the abundance estimation of ichthyoplankton. When zeroes are removed in the statistical analysis, lunar variability of pomacentrid larvae was not statistically different. However, when zeroes are included, these differences are significant.

Acknowledgements. The authors would like to thank C. M. Caipang, the editor and referees for their helpful suggestions that improved this manuscript. Map provided by ReefBase (http://www.reefbase.org).

References

Akaike H., 1974 A new look at the statistical model identification. IEEE Transactions on Automatic Control **19**(6):716–723.

Armada N., 1997 Larval and early juvenile fishes of the Sulu Sea and adjacent waters. UPV J Natural Science **2**(2):102-137.

Blanco G., Villadolid D., 1951 The young of some fishes of Luzon. Philippine J Fish **1**(1): 79-104.

Burnham K. P., Anderson D. R., 2004 Multimodel inference: understanding AIC and BIC in model selection. Soc Meth & Res **33**(2):261-304

Cameron A. C., Trivedi P. K., 1998 Regression analysis of count data. Econometric Society Monograph No. 30 Cambridge University Press, Cambridge, 432 p.

Canty A., Ripley B. D., 2011 Boot: Bootstrap R (S-Plus) Functions. R package version 1.3-2.

Cyr H., Downing J. A., Lalonde S., Baines S. B., Pace M. L., 1992 Sampling larval fish populations: choice of sample number and size. Trans Am Fish Soc **121**(3):356–368.

Davison A. C., Hinkley D. V., Canty A. J., 1997 Bootstrap methods and their applications. Cambridge University Press, Cambridge. ISBN 0-521-57391-2. 592 p.

Dick E. J., 2004 Beyond `lognormal versus gamma': discrimination among error distributions for generalized linear models. Fish Res **70**(2-3):351–366.

Dolar L. L., Alcala A. C., 1993 Notes on the ichthyoplankton and marine mammals of the Sulu Sea, Philippines. Silliman J **2**:43–57.

Ellison A. M., 2004 Bayesian inference in ecology. Ecol Let **7**(6):509-520.

Fox J., 2002 Bootstrapping regression models. Appendix to an index to An R and S-PLUS Companion to Applied Regression, pp. 1-14.

Franklin A. B., Anderson D. R., Gutierrez R. J., Burnham K. P., 2000 Climate, habitat quality, and fitness in northern spotted owl populations in northwestern California. Ecol Monographs **70**(4):539–590.

Heilbron D. C., 1994 Zero-Altered and other regression models for count data with added zeros. Biometrical Journal **36**(5):531–547.

Hilbe J. M., 2011 Negative binomial regression, Cambridge University Press, 553 p.

Hobbs N. T., Hilborn R., 2006 Alternatives to statistical hypothesis testing in ecology: a guide to self teaching. Ecological Applications **16**(1):5-19.

Jackman S., 2011 Pscl: Classes and Methods for R Developed in the Political Science Computational Laboratory, Stanford University. Department of Political Science, Stanford University. Stanford, California. R package version 1.03.10. URL http://pscl.stanford.edu/

Lambert D., 1992 Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics **34**(1):1–14.

Lewin W. C., Freyhof J., Huckstorf V., Mehner T., Wolter C., 2010 When no catches matter: coping with zeroes in environmental assessments. Ecological Indicators **10**(3):572–583.

Lobel P. S., 1978 Diel, lunar, and seasonal periodicity in the reproductive behavior of the pomacanthid fish, *Centropyge potteri*, and some other reef fishes in Hawaii. Pac Sci **32**:193-207.

Martin T. G., Wintle B. A., Rhodes J. R., Kuhnert P. M., Field S. A., Low-Choy S. J., Tyre A. J., Possingham H. P., 2005 Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. Ecol Let **8**(11):1235–1246.

Maunder M. N., Punt A. E., 2004 Standardizing catch and effort data: a review of recent approaches. Fish Res **70**(2-3):41–159.

Mullahy J., 1986 Specification and testing of some modified count data models. Journal of econometrics **33**(3):341–365.

O'Neill M. F., Faddy M. J., 2003 Use of binary and truncated negative binomial modelling in the analysis of recreational catch data. Fish Res **60**(2-3):471–477.

Potts J. M., Elith J., 2006 Comparing species abundance models. Ecol Mod **199**(2):153–163.

Power J. H., Moser E. B., 1999 Linear model analysis of net catch data using the negative binomial distribution. Can J Fish Aq Sci **56**(2):191–200.

R Development Core Team, 2011 R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Robertson D. R., Green D. G., Victor B. C., 1988 Temporal coupling of production and recruitment of larvae of a Caribbean reef fish. Ecology **69**(2):370–381.

Rooper C. N., Martin M. H., 2009 Predicting presence and abundance of demersal fishes: a model application to shortspine thornyhead *Sebastolobus alascanus*. MEPS **379**: 253–266.

Ross R. M., 1978 Reproductive behavior of the anemonefish *Amphiprion melanopus* on Guam. Copeia **1**:103–107.

Thorrold S. R., Shenker J. M., Maddox E. D, Mojica R., Wishinski E., 1994 Larval supply of shorefishes to nursery habitats around Lee Stocking Island, Bahamas. II. Lunar and oceanographic influences. Mar Bio **118**(4):567–578.

Tu W., 2002 Zero-Inflated Data. In: Encyclopedia of Environmetrics. El-Shaarawi A. H., Peiegorsch W. W. (eds), John Wiley and Sons, Chichester.

Venables W. N. Ripley B. D., 2002 Modern applied statistics with S. 4th ed, Springer, New York. ISBN 0-387-95457-0, 495 p.

Victor B. C., 1986 Larval settlement and juvenile mortality in a recruitment-limited coral reef fish population. Ecol Mon **56**(2):145–160.

Wade C. B., 1951 Larvae of tuna and tuna-like fishes from Philippine waters. Fish Bull **51**:445–485.

White G. C., Bennetts R. E., 1996 Analysis of frequency count data using the negative binomial distribution. Ecology **77**(8):2549-2557.

Zeileis A., Kleiber C., Jackman S., 2008 Regression models for count data in R. Journal of Statistical Software **27**(8):1-25.

Received: 01 October 2011. Accepted: 19 October 2011. Published online: 12 November 2011.

Authors:
Erwin O. Brunio, Graduate School of Marine Science and Technology, Tokyo University of Marine Science and Technology, Konan, Minato, Tokyo, 108-8477, Japan, e-mail: erwin@erwinbrunio.com.
Joeppette J. Hermosilla, Graduate School of Marine Science and Technology, Tokyo University of Marine Science and Technology, Konan, Minato, Tokyo, 108-8477, Japan, e-mail: joeppette@yahoo.com.
Filipina B. Sotto, Biology Department, University of San Carlos, Nasipit, Talamban Cebu City, 6000, Philippines, e-mail: arpee55@yahoo.com.
Eiji Tanaka, Resources Analysis Laboratory, Tokyo University of Marine Science and Technology, Konan, Minato, Tokyo, 108-8477, Japan, e-mail: hermit@kaiyodai.ac.jp.

How to cite this article:
Brunio E. O., Hermosilla J. J., Sotto F. B., Tanaka E., 2011 Zeroes do matter: the tale of the missing fishes in Cabilao Island Bohol, Philippines. AACL Bioflux **4**(5):611-620.