

Zero-Inflated Discrete Statistical Models for Fecundity Data Analysis in Channel Catfish, *Ictalurus punctatus*

HERBERT E. QUINTERO

*Department of Fisheries and Allied Aquacultures, Auburn University,
Auburn, Alabama 36849-5419 USA*

ASHEBER ABEBE

*Department of Mathematics and Statistics, Auburn University,
Auburn, Alabama 36849-5310 USA*

D. ALLEN DAVIS¹

*Department of Fisheries and Allied Aquacultures, Auburn University,
Auburn, Alabama 36849-5419 USA*

Abstract

Traditional fecundity analysis, either as number of eggs per body weight of female or number of eggs per gram of egg mass, pay little attention to females that do not spawn. These fecundity variables contain a high proportion of zeroes either because of the absence of eggs or the inability to recover the eggs. Zero-inflated discrete generalized linear models are an alternative method that can be developed to take into account females that do not spawn. In this case study, we propose discrete generalized linear models that use specially constructed mixture models to handle the excess zeroes such as the zero-inflated Poisson and zero-inflated negative binomial models. These models have the advantage of modeling fecundity simultaneously with the probability of spawning. The results show that age was the most significant factor influencing the number of eggs per gram of egg mass, while period of spawning was the most significant factor influencing the number of eggs per female body weight. These were also the most important variables that significantly affected the probability of successful spawning. Model residual diagnostics show that zero-inflated models exhibit superior performance compared to the traditional models like analysis of covariance, Poisson regression, and negative binomial regression models.

Reproduction in captivity has been the key that has opened the door to successful early rearing, metamorphosis, and grow out to market size of economically important fin fish (Donaldson 1996). However, unpredictable and variable reproductive performance is an important limiting factor for the successful mass production of juveniles (Kjørsvik et al. 1990; Fernández-Palacios et al. 1995; Izquierdo et al. 2001). Poor egg quality is one of the major constraints in the expansion of aquaculture of both marine and many freshwater fish species (Brooks et al. 1997). Several methods have been developed to assess the egg quality of fish. One of these is fecundity, which is the total number of eggs produced by each fish expressed either in terms of eggs/spawn or eggs/body weight (Izquierdo

et al. 2001). Fecundity for channel catfish, *Ictalurus punctatus*, ranges from 677 to 14,360 eggs/kg of female. Maternal effects have been considered to be the source of the variation seen in egg enumeration estimates (Walser and Phelps 1993). On the other hand, stripping of females may underestimate realized fecundity because manipulation during stripping and/or the timing of stripping in relation to ovulation time may result in the incomplete recovery of ovulated eggs (Lambert and Thorsen 2003). This effect may be more pronounced in channel catfish considering that they are sequential spawners (Smitherman et al. 1978).

Fecundity analysis either as number of eggs per body weight of female or number of eggs per gram of egg mass has typically considered positive integer numbers with little attention to those females that did not spawn. Usually the

¹ Corresponding author.

only account for this is a summary figure in terms of spawning success and/or total number of eggs or fry produced per total body weight per unit of area. The spawning process is probabilistic in nature (Bernardo 1993). Age and size are not the only factors that affect sexual maturity, as there is a complex process that involves body reserves, water quality, handling stress, etc. These are variables that lead to ovulation and successful stripping or, on the other hand, to egg retention in the body causing overripening. The latter indicates failure in the spawning process. A similar probabilistic condition could be inferred for fecundity data because they are directly related. Thus, careful attention must be given to those females that did not spawn because of stress caused by an excess of manipulation or environmental conditions that prevent recovery of ovulated eggs, either of which produces a high proportion of zeros in egg counts.

Analysis of fecundity is traditionally performed using ANOVA (Walser and Phelps 1993; Lambert 1998; Kristanto 2004), where it is assumed that the errors are independent and identically distributed (IID) according to the normal distribution with a mean of 0 and constant variance. Covariate analysis, simple linear, and multiple regression are also performed to establish relationships between fecundity and total length or body weight. However, discrete data are usually not in line with these considerations and as a result other analytic methods should be considered (Byers et al. 2003). Response variables that result from a process of counting, such as number of eggs, are better modeled using discrete probability distributions such as the Poisson distribution whenever possible. However, one problem with using the Poisson distribution is that it assumes that the mean and variance are equal. This restriction does not hold in many studies where the variance of the data is greater than the mean, a condition referred to as overdispersion. In this case, a disturbance or error term must be included in the model giving rise to the negative binomial (NB) regression model (Zelterman 1999, 2002; Byers et al. 2003). Accounting for overdispersion in count data is necessary to draw correct inferences about the mean (or probability) pa-

rameters (Lindsey 1999). One way to measure overdispersion is to take the square root of the ratio of deviance to its degrees of freedom, where a value greater than one indicates the existence of overdispersion (Zelterman 2002).

Zero-inflated modeling has recently been suggested as an approach, which gives a better fit to count data with excessive zero counts (Lambert 1992). These type of data have been reported in medical and public health studies, ecological studies, and fisheries, among others (Welsh et al. 2000; Lewsey and Thomson 2004; Mwalili et al. 2004; Fletcher et al. 2005; Martin et al. 2005). One of the objectives of this case study is to develop zero-inflated models that give better description of fecundity that can be used for inferences and predictions of future events.

Materials and Methods

Data were obtained from a project performed to improve reproductive efficiency to produce hybrids from female channel catfish, *I. punctatus*, with male blue catfish, *Ictalurus furcatus*, through nutrient manipulations. The study evaluated the influence and interaction of dietary protein level and feeding rate on egg production of two separate strains of catfish (each one with three ages). Two diets containing 42 and 32% protein level, and two feed frequencies (six times/week and three times/week) were considered, defining the following four treatment combinations (TC): TC 1 – 42% protein level, six feedings per week; TC 2 – 42% protein level, three feedings per week; TC 3 – 32% protein level, six feedings per week; TC 4 – 32% protein level, three feedings per week. Females were spawned during three periods (early, middle, and late season). Evaluation of the dietary protein level and feeding rate treatments were performed through spawning success, total number of eggs, and fertility (number of eggs per kilogram female).

Experimental Fish

A total of 414 female channel catfish were maintained at the North Auburn Experimental Station, Auburn University. All broodstock were stocked in 16 ponds, using four ponds per

treatment. The females were divided into two strains on the basis of prior spawning behavior (high spawning strain [HSS] – Strain 1 and low spawning strain [LSS] – Strain 2), and based on that characteristic, they were assigned proportionally in a randomized manner to each pond (Table 1). Female body weights ranged from 0.4 to 3.4 kg for HSS and from 0.6 to 3.1 kg for LSS with a combined mean weight of 1.74 kg. The fish were stocked in February 2004, in 0.04-ha ponds at a density of ~1130 kg/ha, giving an acclimation period of ~1 mo. The trial period was 70–90 d depending on the spawning period. Feeding was done during the warmest part of the day between 1500 and 1700 h, at a rate of 1.7% of total biomass of brood fish stocked per pond. Water quality parameters were taken daily for dissolved oxygen and temperature and twice weekly for pH, ammonia-N, and nitrite-N. Alkalinity and hardness were recorded at stocking and just before harvesting.

For the first spawning period (early), two ponds of each treatment were drained, and 16 females (out of 32) were selected based on external characteristics (abdominal fullness, softness and palpability of the ovaries, redness or swollen appearance of the genitals). Second spawning period (middle) was performed using one pond of each treatment, selecting 16 females (out of 32). The last spawning period (late) selected all the remaining females in all ponds.

Hormone Injections

Selected females were transferred to holding tanks (per treatment) supplied with continuous flow-through water, placed individually in soft mesh bags. Total length, body weight, and girth were recorded. Hormone injections were admin-

istered in two doses, a priming injection of 20 µg/kg luteinizing hormone-releasing hormone analogues (LHRHa), followed 12 h later by resolving dose of 100 µg/kg.

Collection and Fertilization of Gametes

Twenty-four hours after the second injection, females were monitored for ovulation. Females with released eggs were removed from holding tanks and anesthetized in 250 mg/L tricaine methane sulfonate (MS-222). Then females were stripped, and eggs were collected in aluminum containers lubricated previously with vegetable shortening. Those females that did not express eggs were returned then rechecked later. Stripping of gametes ceased when all females had been stripped or attempts to strip them had been made.

Egg Enumeration

Total weight of the egg mass was determined gravimetrically and then number of eggs per gram of egg mass was recorded. Egg number per gram of female body weight was estimated from total number of eggs in the egg mass divided by total female weight.

Statistical Analysis

Fecundity, either as eggs per gram of egg mass or eggs per gram of female body weight, was analyzed using five different approaches: analysis of covariance (ANCOVA), Poisson regression, NB regression, zero-inflated Poisson (ZIP) regression, and zero-inflated negative binomial (ZINB) regression. The explanatory variables examined were protein level, feed frequency, and their interactions, using strain, age of fish, and period of spawning as covariates. The Genmod and NLMixed procedures from SAS® (SAS Institute Inc., Cary, NC, USA) were used.

The statistical model for the ANCOVA was given by:

$$y_{ijklm} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_k + \tau_l + \eta x_{ijklm} + \varepsilon_{ijklm}$$

where μ represents the average response; α_i , $i = 1, 2$, represents the effect of protein i ; β_j , $j = 1, 2$, represents the effect of feeding frequency j ; $(\alpha\beta)_{ij}$, $i, j = 1, 2$, represents the interactive effect of protein i and feeding frequency j ;

TABLE 1. Females by strain per treatment.

Treatment combination	Protein level (%)	Feed frequency	Strain		
			High	Low	Total
1	42	6	44	55	99
2	42	3	55	57	112
3	32	6	44	56	100
4	32	3	44	59	103
Total			187	227	414

$\gamma_k, k = 1, 2, 3$, is the effect of period k ; $\tau_l, l = 1, 2$, represents the effect of strain l ; x_{ijklm} represents the age of fish m in the cell (i, j, k, l) , and ϵ_{ijklm} represents the random error associated with fish m in cell (i, j, k, l) . It is commonly assumed that ϵ_{ijklm} are IID random variables following the normal distribution with mean zero and constant variance σ^2 , that is, $\epsilon_{ijklm} \sim \text{IID } N(0, \sigma^2)$. In the notation of generalized linear models (McCullagh and Nelder 1989), this ANCOVA model can be written as $y_{ijklm} \sim N(\mu_{ijklm}, \sigma^2)$, where μ_{ijklm} is linked to the explanatory variables as follows:

$$\mu_{ijklm} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_k + \tau_l + \eta x_{ijklm},$$

and σ^2 is left unrestricted.

Although this model is widely applied in the literature, assuming normality of the response variable under consideration is not very realistic, it is a discrete-count type random variable. More appropriate models involve discrete probability distributions like the Poisson and NB distributions.

The models we consider are members of the family of models known as generalized linear models (McCullagh and Nelder 1989). In the Poisson regression model, we assume that $y_{ijklm} \sim \text{Poisson}(\lambda_{ijklm})$. Equivalently, the probability mass function (pmf) of the response variable is given by:

$$P(y_{ijklm} = t) = \frac{\exp(-\lambda_{ijklm})\lambda_{ijklm}^t}{t!}, \quad t = 0, 1, 2, \dots,$$

where $\lambda_{ijklm} > 0$ is a parameter that is unknown. Note that the expected value and variance of y_{ijklm} are both equal to λ_{ijklm} , a restriction that may be questionable for the response variables that are considered in this paper. The next task is to link the parameter λ_{ijklm} to the explanatory variables similar to the manner in which the normal distribution parameter μ_{ijklm} was linked to the explanatory variables in the ANCOVA model. However, in this case, the link function must guarantee that $\lambda_{ijklm} > 0$. Although there are several link functions that satisfy this requirement, we shall use the simplest of them all given by:

$$\log(\lambda_{ijklm}) = \mu_{ijklm},$$

where μ_{ijklm} is as defined in the ANCOVA model above.

One possible approach to alleviate the rather stringent restriction imposed by the equality of the mean and variance of y_{ijklm} is to use the NB regression model assumes that the pmf of y_{ijklm} is given by:

$$P(y_{ijklm} = t) = \frac{\Gamma(t + 1/r)}{\Gamma(t + 1)\Gamma(1/r)} \times \frac{(r\lambda_{ijklm})^t}{(1 + r\lambda_{ijklm})^{t+1/r}}, \quad t = 0, 1, 2, \dots,$$

where $\lambda_{ijklm} > 0$ and $r > 0$ are unknown parameters and $\Gamma(\cdot)$ is the gamma function defined by the integral $\Gamma(z) = \int_0^\infty w^{z-1} \exp(-w) dw$. The expected value and variance of y_{ijklm} are λ_{ijklm} and $\lambda_{ijklm}(1 + r\lambda_{ijklm})$, respectively. Note that the variance is always greater than the expected value; thus, this distribution is appropriate to model response variables that may have overdispersion, a situation that cannot be adequately modeled using the Poisson distribution. Once again we shall employ the link function $\log(\lambda_{ijklm}) = \mu_{ijklm}$ and estimate r from the data in an unrestricted manner.

In some cases, the number of zero responses is in excess of that consistent with the Poisson and NB distributions. A relatively easy way to build a model for this zero-inflated data is to attach a positive probability, say $1 - p$ for $0 < p < 1$, to the event that a zero value is not consistent with the Poisson or NB distribution (Lambert 1992). This gives rise to the ZIP and ZINB distributions that sample the respective distribution a proportion p times and zero a proportion $1 - p$ times.

Under the ZIP regression model, pmf of y_{ijklm} is given by:

$$P(y_{ijklm} = t) = \begin{cases} (1 - p) + p \exp(-\lambda_{ijklm}) & \text{if } t = 0 \\ \frac{p \exp(-\lambda_{ijklm}) \lambda_{ijklm}^t}{t!} & \text{if } t > 0. \end{cases}$$

The mean and variance of y_{ijklm} are now given by $E(y_{ijklm}) = p\lambda_{ijklm}$ and $V(y_{ijklm}) = \lambda_{ijklm}(1 + \lambda_{ijklm}(1 - p))$, respectively (Simonoff 2003). The function $\log(\lambda_{ijklm}) = \mu_{ijklm}$ is used to link the parameter λ_{ijklm} to the explanatory variables.

Similarly, the ZINB regression model uses a mixture distribution that assigns a mass $1 - p$ to the “extra” zeroes and a mass p to a NB distribution, where $0 < p < 1$. The pmf of y_{ijklm} in the ZINB model is given by:

$$P(y_{ijklm} = t) = \begin{cases} (1-p) + \frac{p}{(1+r\lambda_{ijklm})^{1/r}} & \text{if } t=0 \\ p \frac{\Gamma(t+1/r)}{\Gamma(t+1)\Gamma(1/r)} \frac{(r\lambda_{ijklm})^t}{(1+r\lambda_{ijklm})^{t+1/r}} & \text{if } t>0. \end{cases}$$

The mean and variance of $E(y_{ijklm}) = p(\lambda_{ijklm})$ and $V(y_{ijklm}) = p\lambda_{ijklm}(1 + \lambda_{ijklm}(r + 1 - p))$, respectively. It can be observed that this distribution approaches the ZIP distribution as $r \rightarrow 0$ and the NB distribution as $p \rightarrow 1$ (Mwalili et al. 2004).

Assessment of fit was based on Akaike’s information criterion (AIC) (Akaike 1987), which is a criterion for selecting the best model among a number of candidate models. The estimated AIC value is given by two times the number of estimated parameters minus two times the achieved log-likelihood. One is rewarded for larger values of the log-likelihood but penalized for larger number of parameters; thus, lower AIC values correspond to better models. Because the models we are considering are not all properly nested, to make the AIC values comparable, we utilize the full log-likelihood in the computation of the AIC. Plots of residuals versus predicted values are used to perform graphical diagnosis of the fitted models. A visible pattern in a residual plot is indicative of misspecification problems with the corresponding model.

The Genmod and NLMixed Procedures from SAS® version 9.1 (SAS Institute Inc.) were used. SAS codes for these analyses are presented in Appendix.

Results

The frequency histogram of number of eggs per gram of egg mass indicates the presence of

a relatively high number of zeroes (Fig. 1). A similar observation is made regarding the distribution of number of eggs per gram of female body weight (Fig. 2). Evaluation of fecundity as number of eggs per gram of egg mass detected overdispersion using the ANCOVA, with an overdispersion value of 24.8, whereas the overdispersion values were reduced to 5.2 and 2.4 for the Poisson and NB models, respectively (Table 2). The AIC value on the other hand was extremely high for the Poisson model with a value of 6898.6, followed by the ANCOVA model with 2148.6, the NB model with 2042.9, the ZIP model with 1619.7, and the ZINB model with 1510.5 (Table 2). Analysis of fecundity as number of eggs per gram of female body weight exhibited a similar pattern however to a lower degree. The overdispersion values were closer to each other and they drop from 3.9 to 0.96 as we move from the ANCOVA model to the NB model (Table 2). The AIC values were 1549.4 for the Poisson model, 1289.4 for ANCOVA, 1260.8 for NB, 1016.2 for ZIP, and 1018.2 for ZINB (Table 2). These AIC values indicate that zero-inflated models provide fits that are superior to the traditional models for these data sets. These results are not unexpected, and they are consistent with our observations from the frequency histograms.

All the Poisson regression model parameters were found to be significant at the 5% level of significance, while the ANCOVA model found feed frequency and age of fish as significant, and early period of spawning as highly significant ($P < .0001$) in determining the number of eggs per gram of egg mass (Table 3). The NB regression model detected only age and early period of spawning as highly significant. The ZIP and ZINB models found age to be the only variable that is highly significant ($P < .0001$) for the same response variable (Table 3). The variables that significantly affect zero inflation, and hence the probability of successful spawning, are found to be feed frequency, age, and early period of spawning in both the ZIP and ZINB models for number of eggs per gram of egg mass. Residual plots for number of eggs per gram of egg mass, with zeroes dropped for better visual inspection, are presented in

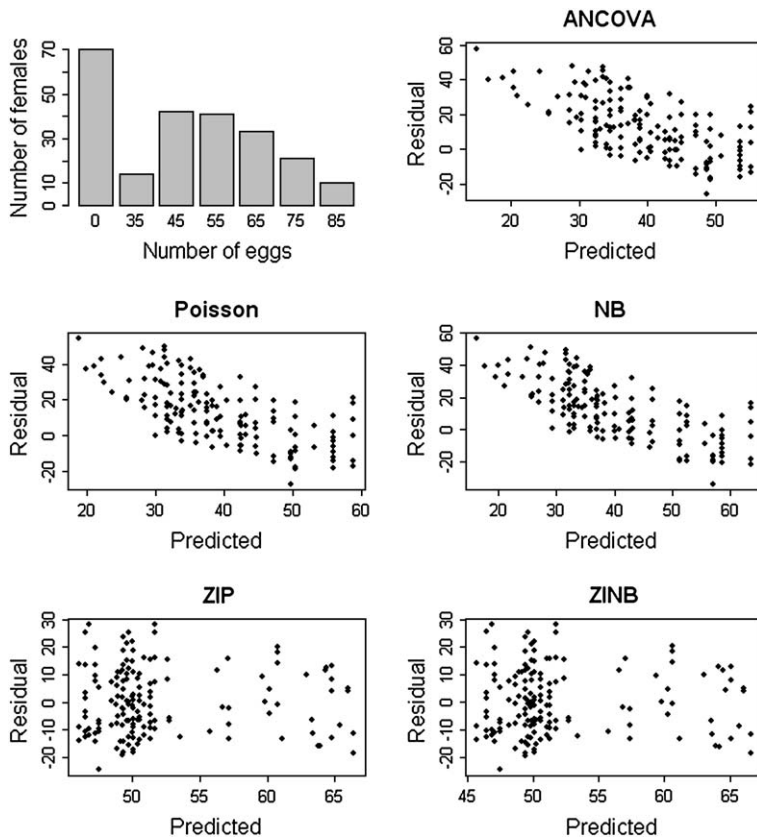


FIGURE 1. Frequency histogram and residual plots for number of eggs per gram of egg mass.

Figure 1. ANCOVA, Poisson, and NB models exhibited a pattern that is generally decreasing, while residuals of the zero-inflated models were evenly distributed above and below the zero line with no visible pattern. This suggests that the ZIP and ZINB models are more appropriate for this type of data.

When analyzing the number of eggs per gram of female body weight, the Poisson model found feed frequency, age of fish, and spawning period to be variables that have a significant effect on this response (Table 4). On the other hand, ANCOVA and NB models found only age and spawning period as significant (Table 4). Finally ZIP and ZINB models found a significant effect of the middle spawning period. In both the ZIP and ZINB models, the variables that significantly affected zero inflation were found to be feed frequency, age of fish, and early period of spawning. Residual plots show once again that

ZIP and ZINB models provide superior fits when compared to all the other models (Fig. 2).

Discussion

Fecundity data analysis enables hatchery producers to estimate their broodstock requirement. As a result, a model for fecundity is considered to be a critical tool to optimize fry production. We approach fecundity analysis using two indexes: number of eggs per gram and number of eggs per gram of female body weight. Fecundity is a trait that is a function of the size or age of fish. According to some authors, the ability to reproduce does not depend on longevity but rather on the attainment of an adequate body size (Sokołowska and Skóra 2001). Thus, most of the fecundity analysis has mainly considered female body weight (Froese and Luna 2004). On the other hand, the effect of age in fecundity indices of freshwater fishes remained

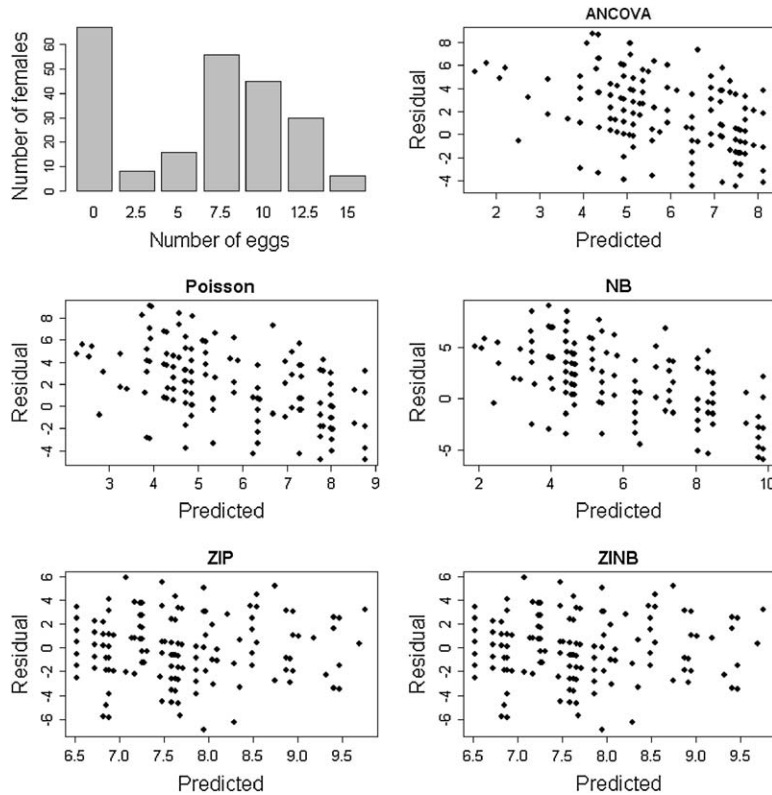


FIGURE 2. Frequency histogram and residual plots for number of eggs per gram of female body weight.

underexplored in contrast to marine species (Shatunovskii 2006). In a catfish aquaculture setting, female age has been found responsible for spawning performance rather than its size (Santiago 1979). In fact, fecundity has been found to be poorly correlated with female body weight (Bice 1981; Walser and Phelps 1993; Argue 1996; Lambert 1998). It is also not uncommon, especially in commercial hatcheries,

to have several females that do not spawn. We provide statistical support to the fact that fish age has significant effect in fecundity even when there were several females that did not spawn, and which likely were younger females, because the odds of spawning of 5-yr-old females were higher than the odds of spawning of 3-yr-old females (Quintero et al. submitted). In this particular analysis, female size was highly variable.

TABLE 2. Summary of the dispersion and AIC of the statistical models for number of eggs per gram of egg mass and eggs per gram of female body weight (0.001 kg).

Estimate	Analysis of covariance	Poisson	Negative binomial model	Zero-inflated poisson distribution	Zero-inflated negative binomial distribution
Eggs per gram of egg mass					
Overdispersion	24.8	5.2	2.4	—	—
AIC	2148.6	6898.6	2042.9	1619.7	1510.5
Eggs per gram of female body weight (0.001 kg)					
Overdispersion	3.9	2.0	0.96	—	—
AIC	1289.4	1549.4	1260.8	1016.2	1018.2

AIC = Akaike's information criterion.

TABLE 3. Summary of estimates of model parameters for number of eggs per gram of egg mass.

	Analysis of covariance	Poisson	Negative binomial model	Zero-inflated Poisson distribution	Zero-inflated negative binomial distribution
Parameters					
Protein level					
*	-4.3697	-0.1030	-0.1354	-0.0245	-0.0231
+	(-13.29, 4.55)	(-0.16, -0.04)	(-0.71, 0.44)	(-0.08, 0.03)	(-0.12, 0.07)
o	0.3355	0.0006	0.6429	0.4098	0.6384
Feed frequency					
*	-10.2593	-0.2766	-0.3113	-0.0323	-0.0328
+	(-19.74, -0.78)	(-0.34, -0.21)	(-0.92, 0.30)	(-0.10, 0.03)	(-0.14, 0.07)
o	0.0340	<0.0001	0.3159	0.3274	0.5453
Protein level × feed frequency					
*	8.1034	0.2123	0.3364	-0.0260	-0.0298
+	(-4.63, 20.83)	(0.12, 0.30)	(-0.49, 1.17)	(-0.11, 0.06)	(-0.17, 0.12)
o	0.2110	<0.0001	0.4249	0.5608	0.6858
Strain					
*	-1.6655	-0.0519	-0.0813	-0.0064	-0.0088
+	(-8.20, 4.86)	(-0.10, -0.01)	(-0.50, 0.34)	(-0.05, 0.04)	(-0.08, 0.06)
o	0.6158	0.0224	0.7021	0.7781	0.8125
Age					
*	6.8203	0.2061	0.2520	-0.1255	-0.1263
+	(3.06, 10.58)	(0.18, 0.23)	(0.01, 0.49)	(-0.15, -0.10)	(-0.17, -0.08)
o	0.0004	<0.0001	0.0420	<0.0001	<0.0001
Early season					
*	14.7158	0.3983	0.4623	0.0156	0.0140
+	(7.65, 21.78)	(0.35, 0.45)	(0.01, 0.92)	(-0.03, 0.06)	(-0.06, 0.09)
o	<0.0001	<0.0001	0.0464	0.5226	0.7272
Middle season					
*	3.7879	0.1234	0.1407	-0.0096	-0.00153
+	(-5.12, 12.70)	(0.06, 0.19)	(-0.44, 0.72)	(-0.07, 0.05)	(-0.12, 0.09)
o	0.4031	0.0002	0.6307	0.7681	0.7723
Inflation					
Protein level					
*				0.5764	0.4363
+				(-0.39, 1.54)	(-0.52, 1.39)
o				0.2409	0.3697
Feed frequency					
*				1.1902	1.0105
+				(0.18, 2.20)	(0.01, 2.01)
o				0.0211	0.0478
Protein level × feed frequency					
*				-1.1846	-0.9455
+				(-2.50, 0.14)	(-2.26, 0.37)
o				0.0789	0.1582
Strain					
*				0.1695	0.1350
+				(-0.50, 0.84)	(-0.53, 0.80)
o				0.6184	0.6914
Age					
*				-1.0327	-1.0448
+				(-1.41, -0.66)	(-1.42, -0.67)
o				<0.0001	<0.0001

TABLE 3. *Continued.*

	Analysis of covariance	Poisson	Negative binomial model	Zero-inflated Poisson distribution	Zero-inflated negative binomial distribution
Early season					
*				-1.5307	-1.5849
+				(-2.33, -0.73)	(-2.39, -0.78)
o				0.0002	0.0001
Middle season					
*				-0.2699	-0.3329
+				(-1.11, 0.57)	(-1.18, 0.51)
o				0.5282	0.4382

* = parameter estimate; + = 95% confidence interval for the parameter; o = *P* value of the parameter estimate.

TABLE 4. *Summary of estimates of model parameters for number of eggs per gram of female body weight (0.001 kg).*

	Analysis of covariance	Poisson	Negative binomial model	Zero-inflated Poisson distribution	Zero-inflated negative binomial distribution
Parameters					
Protein level					
*	-0.2174	-0.0286	-0.0499	0.0535	0.0535
+	(-1.61, 1.17)	(-0.18, 0.13)	(-0.45, 0.35)	(-0.10, 0.21)	(-0.15, 0.26)
o	0.7581	0.7146	0.8039	0.4946	0.6095
Feed frequency					
*	-1.2111	-0.2312	-0.2931	0.0445	0.0445
+	(-2.69, 0.26)	(-0.40, -0.06)	(-0.72, 0.13)	(-0.13, 0.22)	(-0.16, 0.25)
o	0.1072	0.0081	0.1758	0.6103	0.6686
Protein level × feed frequency					
*	0.9036	0.1694	0.3281	-0.0674	-0.0674
+	(-1.08, 2.89)	(-0.06, 0.40)	(-0.25, 0.91)	(-0.30, 0.16)	(-0.39, 0.25)
o	0.3700	0.1462	0.2653	0.5679	0.6802
Strain					
*	-0.4302	-0.0935	-0.1529	-0.0495	-0.0495
+	(-1.45, 0.59)	(-0.21, 0.02)	(-0.45, 0.14)	(-0.17, 0.07)	(-0.20, 0.10)
o	0.4054	0.1150	0.3060	0.4063	0.5125
Age					
*	1.4213	0.3106	0.3744	-0.0300	-0.0300
+	(0.84, 2.01)	(0.23, 0.39)	(0.20, 0.55)	(-0.11, 0.05)	(-0.14, 0.08)
o	<0.0001	<0.0001	<0.0001	0.4528	0.5856
Early season					
*	2.5629	0.4975	0.5998	0.1062	0.1062
+	(1.46, 3.66)	(0.37, 0.63)	(0.28, 0.92)	(-0.02, 0.24)	(-0.02, 0.24)
o	<0.0001	<0.0001	0.0002	0.1080	0.1107
Middle season					
*	1.9993	0.4109	0.4467	0.2626	0.2626
+	(0.61, 3.39)	(0.25, 0.57)	(0.05, 0.84)	(0.10, 0.42)	(0.10, 0.42)
o	0.0049	<0.0001	0.0268	0.0013	0.0016
Inflation					
Protein level					
*				0.4345	0.4350
+				(-0.53, 1.40)	(-0.53, 1.40)
o				0.3767	0.3762
Feed frequency					
*				1.1424	1.1421
+				(0.13, 2.15)	(0.13, 2.15)
o				0.0268	0.0269

TABLE 4. *Continued.*

Analysis of covariance	Poisson	Negative binomial model	Zero-inflated Poisson distribution	Zero-inflated negative binomial distribution
Protein level \times feed frequency				
*			-0.9762	-0.9762
+			(-2.30, 0.35)	(-2.30, 0.35)
o			0.1477	0.1478
Strain				
*			0.1131	0.1129
+			(-0.56, 0.79)	(-0.56, 0.79)
o			0.7406	0.7409
Age				
*			-1.0715	-1.0715
+			(-1.45, -0.69)	(-1.45, -0.69)
o			<0.0001	<0.0001
Early season				
*			-1.7153	-1.7156
+			(-2.53, -0.90)	(-2.52, -0.90)
o			<0.0001	<0.0001
Middle season				
*			-0.4285	-0.4285
+			(-1.28, 0.42)	(-1.28, 0.42)
o			0.3215	0.3215

* = parameter estimate; + = 95% confidence interval for the parameter; o = *P* value of the parameter estimate.

For instance, the range for 3-yr-old females was 1.16–3.46 kg, 4-yr-old females was 0.74–2.8 kg, and 5-yr-old females was 0.94–4.04 kg, which explains the poor correlation between fecundity and female body weight.

Fecundity as number of eggs per gram of egg mass showed a significant effect of age on this trait. We believe that this is because of the fact that the reference point is the egg mass, which is a result of female condition that includes age, size, length, nutrition, etc. Thus, older females had significantly lower number of eggs than younger females, and in other words, they produce bigger eggs. According to Shatunovskii (2006), this phenomenon can be attributed to an increased reproductive function in ontogeny, which is realized as a more active synthesis of ovovitellin in the liver and its storage in oocytes as well as to an elongated period of trophoplasmatic growth of oocytes. This situation is also seen in walleye, where female age accounted for a greater amount of variation in egg mass than fork length or size (Johnston 1997).

Fecundity as number of eggs per gram of female body weight did not show significant effect of age on this trait. This is very likely

a result of confounding, because the reference point is size, and as a result effect from age is eliminated. In fact, period of spawning came up as a variable with a significant effect on this trait. This is not a surprise because fecundity increased from the beginning of the season and then declined as the spawning season is ending (Lambert 1998).

For this data set, zero-inflated models performed better than ANCOVA, Poisson, and NB models. This was evident from the investigation of AIC values and residual plots. Residual plots for number of eggs per gram of egg mass display a clear systematic pattern for the ANCOVA, Poisson, and NB models (Fig. 1). This points out the violation of the constant variance assumption for these models. Residual plots for the zero-inflated models appear to be evenly distributed around the zero line and do not have any visually discernible pattern (Fig. 1). This indicates that there is no evidence of violation of the constant variance assumption for the ZIP and ZINB models. Similar patterns are observed for the residual plots corresponding to the number of eggs per gram of female body weight (Fig. 2).

The ZINB regression model predicted the mean values of the number of eggs per gram per egg mass for age 3, 4, and 5 females to be 63, 55, and 49, respectively. These mean values were 53, 52, and 51 for early, middle, and late spawning periods, respectively. When analyzing number of eggs per gram of female body weight using the ZINB regression model, the predicted means for females of age 3, 4, and 5, were 8.2, 8.3, and 7.5, respectively. For the same response variable, the mean values were 7.8 for early season, 8.2 for middle season, and 6.9 for late-season periods of spawning.

The significance of period of spawning and age in determining zero inflation coincided with the results of a previous study of spawning success using exact logistic regression analysis, where these same variables were found to significantly affect the probability of successful catfish spawning (Quintero et al. submitted).

Conclusions

Reproductive performance is critical in hatchery management, and quantitative models that describe variation in fecundity can be used to optimize resources in this field. Predictions related to number of eggs that can be spawned for a group of fish can help us choose the best spawners. In this regard, zero-inflated models gave better performance than the traditional ANCOVA, Poisson, and NB models. This was evident from investigations of the AIC values and residual plots. Given that results from the zero-inflated models provided a statistical resolution that corresponded with expected biological responses, it appears that this analysis is an optimal choice for these types of data.

Acknowledgments

The authors would like to extend their thanks to those who have taken the time to critically review this manuscript as well as those who helped in supporting this research. The Southern Regional Aquaculture Center SRAC, grants number 2001-38500-10307 and 2002-38500-11805 supported this research. Mention of a trademark or proprietary product does not constitute an endorsement of the product by Auburn University and does not imply its approval to the

exclusion of other products that may also be suitable.

Literature cited

- Akaike, H.** 1987. Factor analysis and AIC. *Psychometrika* 52(3):317–332.
- Argue, B. J.** 1996. Performance of channel catfish, *Ictalurus punctatus*, blue catfish, *I. furcatus*, and their F1, F2, F3 and backcross hybrids. PhD dissertation. Department of Fisheries and Allied Aquacultures, Auburn University, Auburn, Alabama, USA.
- Bernardo, J.** 1993. Determinants of maturation in animals. *Trends in Ecology and Evolution* 8:166–173.
- Bice, T. O.** 1981. Spawning success, fecundity, hatchability, and fry survival in strain and reciprocal pairings for Marion and Kansas channel catfish. Master's thesis. Department of Fisheries and Allied Aquacultures, Auburn University, Auburn, Alabama, USA.
- Brooks, S., C. R. Tyler, and J. P. Sumpter.** 1997. Egg quality in fish: what makes a good egg? *Reviews in Fish Biology and Fisheries* 7:387–416.
- Byers, A. L., H. Allore, T. M. Gill, and P. N. Peduzzi.** 2003. Application of negative binomial modeling for discrete outcomes: a case study in aging research. *Journal of Clinical Epidemiology* 56:559–564.
- Donaldson, E. M.** 1996. Manipulation of reproduction in farmed fish. *Animal Reproduction Science* 42: 381–392.
- Fernández-Palacios, H., M. S. Izquierdo, L. Robaina, A. Valencia, M. Salhi, and J. M Vergara.** 1995. Effect of n-3 HUFA level in broodstock diets on egg quality of gilthead sea bream, *Sparus aurata* L. *Aquaculture* 132:325–337.
- Fletcher, D., D. I. Mackenzie, and E. Villouta.** 2005. Modelling skewed data with many zeros: a simple approach combining ordinary and logistic regression. *Environmental and Ecological Statistics* 12:45–54.
- Froese, R. and S. Luna.** 2004. No relationship between fecundity and annual reproductive rate in bony fish. *Acta Ichthyologica et Piscatoria* 34(1):11–20
- Izquierdo, M. S., H. Fernández-Palacios, and A. G. J. Tacon.** 2001. Effect of broodstock nutrition on reproductive performance of fish. *Aquaculture* 197:25–42.
- Johnston, T. A.** 1997. Within-population variability in egg characteristics of walleye, *Stizostedion vitreum*, and white sucker, *Catostomus commersoni*. *Canadian Journal of Fisheries and Aquatic Sciences* 54:1006–1014.
- Kjørsvik, E., A. Mangor-Jensen, and I. Holmeffjord.** 1990. Egg quality in fishes. Pages 71–113 in J. H. S. Blaxter, F. S. Russell, and M. Yonge, editors. *Advances in marine biology*, volume 26. Academic Press, London, England, UK.
- Kristanto, A. H.** 2004. Evaluation of various factors to increase the efficiency of channel-blue hybrid catfish embryo production. PhD dissertation. Department of Fisheries and Allied Aquacultures, Auburn University, Auburn, Alabama, USA.

- Lambert, D.** 1992. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* 34:1–14.
- Lambert, D. M.** 1998. Comparison of techniques for artificial production of channel catfish, *Ictalurus punctatus*, females x blue catfish, *I. furcatus* males, hybrid embryos. Master's thesis. Department of Fisheries and Allied Aquacultures, Auburn University, Auburn, Alabama, USA.
- Lambert, Y. and A. Thorsen.** 2003. Integration of captive and wild studies to estimate egg and larval production of fish stocks. *Journal of Northwest Atlantic fishery science* 33:71–79.
- Lewsey, J. D. and W. M. Thomson.** 2004. The utility of the zero-inflated Poisson and zero-inflated negative binomial models: a case study of cross-sectional and longitudinal DMF data examining the effect of socioeconomic status. *Community Dentistry and Oral Epidemiology* 32:183–189.
- Lindsey, J. K.** 1999. Responses surfaces for overdispersion in the study of the conditions for fish eggs hatching. *Biometrics* 55(1):149–155.
- Martin, T. G., B. A. Wintle, J. R. Rhodes, P. M. Kuhnert, S. A. Field, S. J. Low-Choy, A. J. Tyre, and H. P. Possingham.** 2005. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. *Ecology Letters* 8:1235–1246.
- McCullagh, P. and J. A. Nelder.** 1989. Generalized linear models. Monographs on statistics and applied probability, 37. Chapman and Hall, New York.
- Mwalili, S. M., E. Lesaffre, and D. Declercq.** 2004. The zero-inflated negative binomial regression model with correction for misclassification: an example in caries research. Technical report 0462. IAP Statistics Network. Interuniversity Attraction Pole. <http://www.stat.ucl.ac.be/ISpub/tr/2004/TR0462.pdf>. Accessed 16 November 2005.
- Santiago, A. C.** 1979. Effects of feeding regime on reproductive performance of female channel catfish in ponds. PhD dissertation. Department of Fisheries and Allied Aquacultures, Auburn University, Auburn, Alabama, USA.
- Shatunovskii, M. I.** 2006. Some patterns of age and geographical variation in fish fecundity. *Biology Bulletin* 33(2):195–198.
- Simonoff, J. S.** 2003. Analyzing categorical data. Springer texts in Statistics. Springer, New York, USA.
- Smitherman, R. O., E. Hussein, and R. E. Reagan.** 1978. Genetics and breeding of channel catfish. Southern Coop. Series Bulletin 223. Alabama Agricultural Experiment Station. Auburn University, Auburn, Alabama, USA.
- Sokolowska, E. and K. E. Skóra.** 2001. Fecundity of ninespine stickleback, *Pungitius pungitius* L., 1758, in the Puck Bay. *Acta Ichthyologica et Piscatoria* 31(1):45–60.
- Walser, C. A. and R. P. Phelps.** 1993. Factors influencing the enumeration of channel catfish eggs. *Progressive Fish-Culturist* 55:195–198.
- Welsh, A. H., R. B. Cunningham, and R. L. Chambers.** 2000. Methodology for estimating the abundance of rare animals: seabird nesting on North East Herald Cay. *Biometrics* 56:22–30.
- Zelterman, D.** 1999. Models for discrete data. Oxford University Press, New York, New York, USA.
- Zelterman, D.** 2002. Advanced Log-linear models using SAS®. SAS Institute Inc., Cary, North Carolina, USA.

Appendix. SAS codes.

```

data spawning;
input pro fre str age grp Y3 Y4;
/*Y3 is number of eggs per gram of egg mass,
and Y4 is number of eggs per gram of female
body weight*/
data spawning3;
set spawning;
if (fre eq 3) then fre = 0; if (fre eq 6) then
fre = 1;
if (pro eq 32) then pro = 0; if (pro eq 42) then
pro = 1;
if (str eq 1) then str = 0; if (str eq 2) then
str = 1;
if (age eq 3) then age3 = 1; else age3 = 0; if
(age eq 4) then age4 = 1; else age4 = 0;
if (grp eq 1) then grp1 = 1; else grp1 = 0; if
(grp eq 2) then grp2 = 1; else grp2 = 0;
run;

proc nlmixed data = spawning3;
title 'eggs per gram - Ancova';
parms b0 = 0 b1 = 0 b2 = 0 b12 = 0
b31 = 0 b32 = 0 b4 = 0 b5 = 0;
mean = b0 + b1*fre + b2*pro + b12*pro*fre
+ b31*grp1 + b32*grp2 + b4*str + b5*age;
model y3 ~ normal (mean, v);
predict mean out = est;
run;

proc nlmixed data = spawning3;
title 'Eggs per gram - Poisson using
nlmixed';
parms b0_nb = 0 b1_nb = 0 b2_nb = 0
b12_nb = 0 b31_nb = 0 b32_nb = 0
b4_nb = 0 b5_nb = 0 k = 20;
eta_nb = b0_nb + b1_nb*fre + b2_nb*pro +
b12_nb*pro*fre + b31_nb*grp1 + b32_nb*grp2 +
b4_nb*str + b5_nb*age;
mean = exp(eta_nb);
model y3 ~ poisson(mean);
predict mean out = est;
run;

```

```

proc nlmixed data = spawning3;
  title 'Eggs per gram - Negative Binomial
  using nlmixed';
  parms b0_nb = 0 b1_nb = 0 b2_nb = 0
  b12_nb = 0 b31_nb = 0 b32_nb = 0
  b4_nb = 0 b5_nb = 0;
  eta_nb = b0_nb + b1_nb*fre + b2_nb*pro +
  b12_nb*pro*fre + b31_nb*grp1 + b32_nb*grp2 +
  b4_nb*str + b5_nb*age;
  mean = exp(eta_nb);
  p_else = exp(lgamma(y3+(1/k)) - lgamma(y3+
  1) - lgamma(1/k) + y3*log(k*mean) - (y3+
  (1/k))*log(1+k*mean));
  loglike = log(p_else);
  model y3 ~ general(loglike);
  predict mean out = est;
  run;
proc nlmixed data = spawning3;
  title 'Eggs per gram - Zero-Inflated Poisson';
  parms b0_zip = 0 b1_zip = 0 b2_zip = 0
  b12_zip = 0 b31_zip = 0 b32_zip = 0
  b4_zip = 0 b5_zip = 0
  b0_nb = 0 b1_nb = 0 b2_nb = 0 b12_nb = 0
  b31_nb = 0 b32_nb = 0 b4_nb = 0 b5_nb = 0;
  eta_zip = b0_zip + b1_zip*fre +
  b2_zip*pro + b12_zip*pro*fre + b31_zip*grp1 +
  b32_zip*grp2 + b4_zip*str + b5_zip*age;
  p0_zip = 1/(1 + exp(-1*eta_zip));
  eta_nb = b0_nb + b1_nb*fre + b2_nb*pro +
  b12_nb*pro*fre + b31_nb*grp1 +
  b32_nb*grp2 + b4_nb*str + b5_nb*age;
  mean = exp(eta_nb);
  p0 = p0_zip + (1-p0_zip)* exp(lgamma(y3+
  (1/k)) - lgamma(y3 + 1) - lgamma(1/k) +
  y3*log(k*mean) - (y3+(1/k))*log(1+k*-
  mean));
  p_else = (1-p0_zip)* exp(lgamma(y3+(1/k)) -
  lgamma(y3 + 1) - lgamma(1/k) + y3*log
  (k*mean) - (y3+(1/k))*log(1+k*mean));
  if y3 = 0 then loglike = log(p0);
  else loglike = log(p_else);
  model y3 ~ general(loglike);
  predict mean out = est;
  run;
/*Similar codes for Y4*/

```