1. Introduction

The numerical abundance of many species sharing the same ecosystem very different levels of the organism and are in constant change, depending on many factors. Due to the heterogeneous structure of the life cycles of organisms and abiotic resources in the environment based on census population densities derived from overdispersion (variance is higher than means in Poisson distribution) (Cox, 1983; Cameron and Trivedi, 1998) and a large number of zero values (zero-inflated data) is observed (Yeşilova et al, 2011). In such a case, zero-inflated Poisson (ZIP) regression model is a appropriate approach for analyzing a dependent variable having excess zero observations (Lambert, 1992; Böhning, 1998; Böhning et al, 1999; Yau and Lee, 2001; Lee et al, 2001; Khoshgoftaar et al, 2005; Yeşilova et al, 2010). Zero-inflation is also likely in data sets, excess zero observations. In such cases, a zero-inflated negative binomial (ZINB) regression model is an alternative method (Ridout et al, 2001; Yau, 2001; Cheung, 2002; Jansakul, 2005; Long and Frese, 2006; Hilbe, 2007; Yeşilova et al, 2009; Yeşilova et al, 2010). Moreover, The Poisson hurdle model and negative binomial hurdle model (Rose and Martin, 2006; Long and Frese, 2006; Hilbe, 2007; Yeşilova et al, 2009; Yeşilova et al, 2010), and zero-inflated generalized Poisson (ZIGP) model (Consul, 1989, Consul and Famoye, 1992; Czado et al., 2007) are widely used in the analysis of zero-inflated data.

In this part, the analysis of data with many zeros for Notonecta viridis Delcourt (Heteroptera: Notonectidae) and Chironomidae species (Diptera) were carried out by means of using the models of Poisson Regression (PR), negative binomial (NB) regression, zero-inflated Poisson (ZIP) regression, zero-inflated negative binomial (ZINB) regression and negative binomial hurdle (NBH) model.

Samplings

The study was based on periodical samplings of the coastal band of Van Lake, conducted between July-September 2005 and May-September 2006. Samples were taken at totally twenty sampling points as streams entrance (6 points), settlement coastlines (7 points) and naspeciesal coastlines (7 points). Samples were taken according to Hansen et al. (2000). The
invertebrates were collected with a standard sweep net (30 cm width, 1 mm mesh) (Southwood, 1978; Rosenberg, 1997; Hansen et. al, 2000; Yeşilova et al., 2011). Notonectid identification was made by Dmitry A. Gapon (Zoological Institute RAS, Universitetskaya nab., 1, St. Petersburg, Russia).

2. Methods

2.1 Poisson regression

The logarithm of mean of Poisson distribution ($\mu$) is supposed to be a linear function of independent variables ($x_i$) is,

$$\log(\mu_i) = (x_i, \beta)$$

Poisson Regression Model can be written as

$$\Pr(y_i / \mu_i, x_i) = \exp(-\mu_i)\mu_i^{y_i} / y_i! \text{, } y_i = 0, 1, \ldots$$ (1)

In equation 1, $y_i$ denotes dependent variable having Poisson distribution. Likelihood function for PR model is, (Böhning, 1998)

$$LL(\beta / y_i, x_i) = \sum_{i=1}^{n} \left[ y_i x_i \beta - \exp(x_i \beta) - \ln y_i! \right]$$ (2)

In equation 2, $\beta$ are unknown parameters. $\beta$ can be estimated by maximizing log likelihood function according to ML (Khoshgoftaar et al, 2005; Yau, 2006).

2.2 Negative binomial regression

NB regression model is,

$$\Pr(Y = y_i / x_i) = \frac{\Gamma(y_i + 1)}{y_i! \Gamma(1 + \alpha \mu_i)} \left(\frac{1}{\alpha} \right)^{\alpha y_i + \frac{1}{\alpha}} - 1 \alpha > 0$$ (3)

In equation 3, $\alpha$ is a arbitrary parameter and indicates overdispersion level. Log likelihood function for NB regression model is (Hilbe, 2007; Yau, 2006),

$$\text{LL}(\beta, \alpha, y) = \sum_{i=1}^{n} \left[ \frac{1}{\alpha} \log(1 + \alpha \mu_i) - y_i \log \left(1 + \frac{1}{\alpha \mu_i}\right) \right] + \log \Gamma \left(y_i + \frac{1}{\alpha}\right) - \log \Gamma \left(\frac{1}{\alpha}\right) - \log y_i!$$
2.3 Zero inflated poisson regression
ZIP regression is [13],

\[ \Pr(y_i/x_i) = \begin{cases} \pi_i + (1-\pi_i)\exp(-\mu_i), & y_i = 0 \\ (1-\pi_i)\exp(-\mu_i)\mu_i^{y_i}/y_i!, & y_i > 0 \end{cases} \]  

(4)

In equation (4), \( \pi_i \) represents the possibility of extra zeros' existence. Log likelihood function for ZIP model is (Yau, 2006),

\[ LL = \sum_{i=1}^{n} \left\{ \begin{array}{l} I_{y_i=0} \log \left( \pi_i + (1-\pi_i)e^{-\mu_i} \right) \\ +I_{y_i>0} \log \left( (1-\pi_i)e^{-\mu_i} \right) \end{array} \right. \]

(5)

\[ LL = \sum_{i=1}^{n} \left\{ I_{y_i=0} \log \left( \pi_i + (1-\pi_i)e^{-\mu_i} \right) \\ +I_{y_i>0} \left( \log(1-\pi_i) + y_i \log \mu_i - \mu_i - \log y_i! \right) \right. \]

\( I(.) \), given in equation (5) is the indicator function for the specified event. Then \( \mu_i \) and \( \pi_i \) parameters can be obtained following link functions,

\[ \log(\mu) = B\beta \]  

(6)

and

\[ \log \left( \frac{\pi}{1-\pi} \right) = G\gamma \]  

(7)

In equations 6 and 7, B(nxp) and G(nxq) are covariate matrixes. \( \beta \) and \( \gamma \) are respectively unknown parameter vectors with px1 and qx1 dimension (Yau, 2006).

2.4 Zero inflated negative binomial regression
ZINB regression model is [18],

\[ \Pr(y_i/x_i) = \begin{cases} \pi_i + (1-\pi_i)(1+\alpha\mu_i)^{-\alpha}^{-1}, & y_i = 0 \\ \Gamma \left( y_i + \frac{1}{\alpha} \right) (\alpha\mu_i)^{y_i} \left( 1+\alpha\mu_i \right)^{-\alpha}^{-1}, & y_i > 0 \end{cases} \]  

(8)
In equation (8), \( \alpha \geq 0 \) indicates an overdispersion parameter. Log likelihood function for ZINB model is (Yau, 2006),

\[
LL(\mu, \alpha, \pi; y) = \sum_i \left( I_{y_i = 0} \log(\pi_i) + (1 - \pi_i) \left(1 + \alpha \mu_i\right)^{-\alpha - 1} \right) + I_{y_i > 0} \log \left( \frac{\Gamma(y_i + \frac{1}{\alpha})}{y_i! \Gamma\left(\frac{1}{\alpha}\right)} \right) \frac{\left(\alpha \mu_i\right)^{y_i}}{(1 + \alpha \mu_i)^{y_i + \frac{1}{\alpha}}} 
\]

\[= \sum_i \left( I_{y_i = 0} \log(\pi_i + (1 - \pi_i) \left(1 + \alpha \mu_i\right)^{-\alpha - 1} \right) + I_{y_i > 0} \left( \log(1 - \pi_i) - \frac{1}{\alpha} \log(1 + \alpha \mu_i) \right) - y_i \log \left(1 + \frac{1}{\alpha \mu_i}\right) + \log \Gamma\left(y_i + \frac{1}{k}\right) \right) - \log \Gamma\left(\frac{1}{\alpha}\right) - \log y_i! \]

(9)

\( I(.) \), given in equation 9 is the indicator function for the specified event. The model described by Lambert (1992) can be given as,

\[
\log(\mu) = X\beta \quad \text{and} \quad \log\left(\frac{\pi}{1 - \pi}\right) = G\gamma
\]

Here, \( X(nxp) \) and \( G(nxq) \) covariate matrices, \( \beta \) and \( \gamma \) are respectively unknown parameter vectors with px1 and qx1 dimension. Maximum likelihood estimations for \( \beta, \alpha \) and \( \gamma \) can be obtained by using EM algorithm.

2.5 Negative binomial hurdle model

Log-likelihood for negative binomial hurdle model (Hilbe, 2007),

\[
L = \ln(f(0)) + \left\{ \ln\left[1 - f(0)\right] + \ln P(j) \right\} 
\]

(10)

In equation (10), \( f(0) \) indicates the probability of the binary part and \( p(j) \) indicates the probability of positive count. The probability of zero for logit model is,

\[
f(0) = P(y = 0; x) = 1/(1 + \exp(xb1))
\]
and

1- \( f(0) \) is,

\[
\exp(xb1)/(1 + \exp(xb1))
\]

The log likelihood function for both parts of negative binomial hurdle Model is,

\[
L = \text{cond} \{ y = 0, \ln(1/1 - \exp(xb1)) \},
\]

\[
\ln(\exp(xb1)/(1 + \exp(xb1)))
\]

\[+ y \cdot \ln(\exp(xb)/(1 + \exp(xb)))\]

\[-\ln(1 + \exp(xb))/\alpha + \ln\Gamma(y + 1/\alpha)\]

\[-\ln\Gamma(y + 1) - \ln\Gamma(1/\alpha)\]

\[-\ln(1 - (1 + \exp(xb))(-1/\alpha))\]

### 2.6 Model selection

Akaiki Information Criteria (AIC) is goodness of criteria used for model selection. AIC,

\[
AIC = -2LL + 2r 
\]

In equations, LL indicates log likelihood, \( r \) indicates parameter number and \( n \) indicates sample size.

### 3. Results

In this study, R statistical software program was used. Insect densities were included to the model as dependent variable. Besides years, months, species and station are included as independent variables to the model. The 66 (20.63\%) of the 320 dependent variable were zero valued. The distribution of the insect densities was skewed to right because of excess zeros.

| Model | AIC     |
|-------|---------|
| PR    | 57846.00|
| ZIP   | 47791.71|
| NB    | 3176.40 |
| ZINB  | 2819.80 |
| PH    | 47791.71|
| NBH   | 2803.206|

Table 1. Model selection criteria for PR, NB, ZIP, ZINB, PH and NBH.
In PR analyses, deviance and Pearson Chi-square goodness of statistics higher than one (831.417 and 650.213, respectively). Thus, goodness of statistics represents that there is an overdispersion in insect densities. AIC model selection criteria for the models of PR, NB, ZIP, ZINB, PH, and NBH were given in Table 1. The model with the smallest AIC was NBH regression.

Maximum likelihood (ML) parameter estimations and standard errors for PR were given in Table 2.

|             | Estimate | Std. Error | z value | Pr(>|z|)       | $e^\beta$ |
|-------------|----------|------------|---------|---------------|------------|
| (Intercept) | 6.179499 | 0.054470   | 113.449 | <2e-16 ***    | 482.992    |
| year        | 0.118847 | 0.013069   | 9.094   | <2e-16 ***    | 1.125244   |
| month       | 0.175298 | 0.005066   | 34.604  | <2e-16 ***    | 1.191246   |
| Station     | -0.081353| 0.001124   | -72.357 | <2e-16 ***    | 0.921917   |
| species     | -1.943212| 0.018356   | 105.863 | <2e-16 ***    | 0.1432735  |

*p<0.05, **p<0.01, ***p<0.001

Table 2. Parameter estimations and standard errors for Poisson regression.

ML parameter estimations and standard errors for negative binomial regression were given in Table 2.

|             | Estimate | Std. Error | z value | Pr(>|z|)       | $e^\beta$ |
|-------------|----------|------------|---------|---------------|------------|
| (Intercept) | 8.52318  | 0.99249    | 8.588   | 4.16e-16 ***  | 5029.119   |
| year        | -0.15794 | 0.24824    | -0.636  | 0.525         | 0.853901   |
| month       | -0.08205 | 0.09168    | -0.895  | 0.372         | 0.9212259  |
| Station     | -0.08031 | 0.01949    | -4.121  | 4.82e-05 ***  | 0.9228302  |
| species     | -1.92518 | 0.22452    | -8.575  | 4.56e-16 ***  | 0.1458495  |

*p<0.05, **p<0.01, ***p<0.001

Table 3. Parameter estimations and standard errors for negative binomial regression.

ML parameter estimations and standard errors for zero-inflated Poisson regression both count model and logit model were given in Table 4 and Table 5, respectively.

|             | Estimate | Std. Error | z value | Pr(>|z|)       | $e^\beta$ |
|-------------|----------|------------|---------|---------------|------------|
| (Intercept) | 6.017745 | 0.056073   | 107.32  | <2e-16 ***    | 410.6515   |
| year        | 0.271101 | 0.013047   | 20.78   | <2e-16 ***    | 1.311408   |
| month       | 0.162333 | 0.005271   | 30.80   | <2e-16 ***    | 1.176252   |
| station     | -0.046859| 0.001122   | -41.76  | <2e-16 ***    | 0.954222   |
| species     | -2.002676| 0.018382   | -108.94 | <2e-16 ***    | 0.1349736  |

*p<0.05, **p<0.01, ***p<0.001

Table 4. Parameter estimations and standard errors for ZIP count model.
Table 5. Parameter estimations and standard errors for ZIP logit model.

ML parameter estimations and standard errors for zero-inflated negative binomial regression both count model and logit model were given in Table 6 and Table 7, respectively.

Table 6. Parameter estimations and standard errors for ZINB count model.

Table 7. Parameter estimations and standard errors for ZINB logit model.

ML parameter estimations and standard errors for Poisson hurdle both count model and logit model were given in Table 8 and Table 9, respectively.

Table 8. Parameter estimations and standard errors for PH count model.
ML parameter estimations and standard errors obtained for the NBH count model was given in Table 8. While stations and species were significant on the insect densities, the effect of years and the effect of months were not significant on the insect densities.

ML parameter estimations and standard errors obtained for the NBH logit model was given in Table 9. The effects months, years and species were not significant on the insect densities. However, the effect of station was significant on the insect densities.

|                  | Estimate | Std. Error | z value | Pr(>|z|)   | $e^\beta$ |
|------------------|----------|------------|---------|------------|----------|
| (Intercept)      | 3.92991  | 1.33906    | 2.935   | 0.00334 ** | 50.9024  |
| year             | -0.50266 | 0.33705    | -1.491  | 0.13587    | 0.6049194|
| month            | -0.04405 | 0.11930    | -0.369  | 0.71197    | 0.9569061|
| station          | 0.17250  | 0.02994    | -5.761  | 8.36e-09 ***| 1.188272 |
| species          | 0.44380  | 0.30013    | 1.479   | 0.13923    | 1.558619 |

*p<0.05, **p<0.01, ***p<0.001

Table 9. Parameter estimations and standard errors for PH logit model.

ML parameter estimations and standard errors obtained for negative binomial hurdle both count model and logit model were given in Table 10 and Table 11, respectively.

ML parameter estimations and standard errors obtained for the NBH count model was given in Table 10. While stations and species were significant on the insect densities, the effect of years and the effect of months were not significant on the insect densities.

|                  | Estimate | Std. Error | z value | Pr(>|z|)   | $e^\beta$ |
|------------------|----------|------------|---------|------------|----------|
| (Intercept)      | 9.43372  | 1.26292    | 7.470   | 8.03e-14 ***| 1250.925 |
| year             | -0.19128 | 0.24381    | -0.785  | 0.4327     | 0.8259013|
| month            | -0.17020 | 0.11124    | -1.530  | 0.1260     | 0.8434961|
| station          | -0.04587 | 0.02096    | -2.188  | 0.0287 *   | 0.9551661|
| species          | -2.33333 | 0.25071    | -9.307  | <2e-16 *** | 0.0969723 |

*p<0.05, **p<0.01, ***p<0.001

Table 10. Parameter estimations and standard errors for NBH count model.

ML parameter estimations and standard errors obtained for the NBH logit model was given in Table 11. The effects months, years and species were not significant on the insect densities. However, the effect of station was significant on the insect densities.

|                  | Estimate | Std. Error | z value | Pr(>|z|)   | $e^\beta$ |
|------------------|----------|------------|---------|------------|----------|
| (Intercept)      | 3.92991  | 1.33906    | 2.935   | 0.00334 ** | 50.9024  |
| year             | -0.50266 | 0.33705    | -1.491  | 0.13587    | 0.6049194|
| month            | -0.04405 | 0.11930    | -0.369  | 0.71197    | 0.9569061|
| station          | -0.17250 | 0.02994    | -5.761  | 8.36e-09 ***| 0.8415583|
| species          | 0.44380  | 0.30013    | 1.479   | 0.13923    | 1.558619 |

*p<0.05, **p<0.01, ***p<0.001

Table 11. Parameter estimations and standard errors for NBH logit model.
Average insect density observed in the year 2005 has shown 17% decrease in reference to the year 2006. Insect densities observed at monthly sampling ranges depending on water temperature species were increased with the rise of temperature species, but specifically after the month of July such intensity was decreased at the rate of 16% \( (e^{0.19128} \sim 0.8434961) \) towards the month of September within the both years. It has been determined that insect intensities observed at different stations have shown differentiation at the rate of 5%. Chironomid larvae which are included in prey of notonectidae fed by different sources of food at aquatic environment have been found at rather lower density in reference to notonectid density. However, it is hard to guess that such decrement has been formed under the impact of notonectidae. Nevertheless notonectidae do not depend on a single host, their sources of food are rather wide range of variety. Small arthropods on the water surface, small crustaceans living in water, larvae of aquatic insects, snails, small fish or larvae of frog are among their preys (Bruce et al., 1990).

4. References

[1] Böhning, D. Zero-Inflated Poisson Models and C. A. MAN. (1998). A Tutorial Collection of Evidence. *Biometrical Journal*, 40(7), 833-843.

[2] Böhning, D., Dietz, E. and Schlattmann, P. (1999). The Zero-Inflated Poisson Model and the Decayed, Missing and Filled Teeth Index in Dental Epidemiology. *Journal of Royal Statistical Society, A*, 162, 195–209.

[3] Bruce, A.M., Pike, E.B. and Fisher, W.J. (1990). A review of treatment processes to meet the EC Sludge Directive. *J. Inst. Wat. Environ. Management*, 4, 1-13.

[4] Cameron, A.C. and Trivedi, P.K. (1998). Regression Analysis of Count Data. New York, Cambridge University Pres.

[5] Cheung, Y.B. (2002). Zero-Inflated Models for Regression Analysis of Count Data. *A Study of Growth and Development. Statistics in Medicine*, 21, 1461-1469.

[6] Consul, P. C. Generalized Poisson distributions, Volume 99 of Statistics: Textbooks and Monographs. New York: Marcel Dekker Inc. Properties and applications 1989.p.1-20.

[7] Consul, P. C. and F. Famoye. (1992). Generalized Poisson regression model. Comm. Statist. Theory Methods. 21(1), 89-109.

[8] Cox, R. (1983). Some Remarks on Overdispersion. *Biometrika*, 70, 269-274.

[9] Czado, C., Erhardt, V., Min, A., Wagner, S. (2007). Dispersion and zero-inflation level applied to patent outsourcing rates Zero-inflated generalized Poisson models with regression effects on the mean, *Statistical Modelling*. 7(2) : 125-153

[10] Hansen, J., Mki. Sato, R. Ruedy, A. Lacis & V. Oinas. (2000). Global warming in the twenty-first century: An alternative scenario,

[11] Hilbe, J.M. (2007). Negative Binomial Regression. Cambridge, UK.

[12] Jansakul, N. (2005). Fitting a Zero-inflated Negative Binomial Model via R. In Proceedings 20th International Workshop on Statistical Modelling. Sidney, Australia, 277-284.

[13] Khoshgoftaar, T.M., Gao, K. and Szabo, R.M. (2005). Comparing Software Fault Predictions of Pure and Zero-inflated Poisson Regression Models. *International Journal of Systems Science*, 36(11), 705-715.
[14] Lambert, D. (1992). Zero-Inflated Poisson Regression, with an Application to Defects in Manufacspeciesin. \textit{Technometrics}, 34(1), 1-13.

[15] Lee, A.H., Wang, K. and Yau, K. K. W. (2001). Analysis of Zero-Inflated Poisson Data Incorporating Extent of Exposure. \textit{Biometrical Journal}, 43(8), 963-975.

[16] Long, J.S. and Freese, J. (2006). Regression Models for Categorical Dependent Variable Using Stata. A Stata Pres Publication, USA.

[17] McCullagh, P. and Nelder, J. A. (1989). Generalized Linear Models. Second Edition, Chapman and Hall, London.

[18] Ridout, M., Hinde, J. and Demetrio, C.G.B. (2001). A Score Test for a Zero-Inflated Poisson Regression Model Against Zero-Inflated Negative Binomial Alteratves. \textit{Biometrics}, 57, 219-233.

[19] Rose, C.E, Martin, S.W., Wannemuehler, K.A. and Plikaytis, B.D. (2006). On the of Zero-inflated and Hurdle Models for Medelling Vaccine Adverse event Count Data. \textit{Journal of Biopharmaceutical Statistics}, 16, 463-481.

[20] Rosenberg, D. M., I. J. Davies, D.G. Cobb & A.P. Wiens. (1997). Protocols For Measuring Biodiversity: Benthic Macroinvertebrates in Fresh Waters, http://www.emanrese.ca/eman/ecotools/protocols/freshwater/benthics/intro.html, (Data accessed: 10.10.

[21] Southwood, T.R.E. (1978). Ecological Methods, with Particular Reverence to the Study of Insect Populations, 2nd ed., Chapman and Hall, London and New York, 524 pp.

[22] Yau, K.K.W. and Lee, A.H. (2001). Zero-Inflated Poisson Regression with Random Effects to Evaluate an Occupational Injury Prevention Programme. \textit{Statistics in Medicine}, 20, 2907-2920.

[23] Yau, Z. (2006). Score Tests for Generalization and Zore-Inflation in Count Data Modeling. Unpublished Ph. D. Dissertation, University of South Caroline, Columbia.

[24] Yeşilova, A., Kaydan, B. and Kaya, Y. (2010). Modelling Insect-Egg Data with Excess Zeros using Zero-inflated Regression Models. \textit{Hacettepe Journal of Mathematics and Statistics}. 39(2),273-282.

[25] Yeşilova, A. Y., Kaya, B., Kaki, I.Kasap (2010)Analysis of Plant Protection Studies with Excess Zeros Using Zero-Inflated and Negative Bi Binomial Hurdle Models \textit{G.U. Journal of Science}.

[26] Yeşilova, A., Özgökçe, M. S., Atlhan, R., Karaca, İ., Özgökçe, F., Yıldız, Ş. and Kaydan, B. and Kaya, Y. (2011). Investigation of the effects of physico-chemical environmental conditions on population fluctuations of \textit{Notonecta viridis} Delcourt, 1909 (Hemiptera: Notonectidae) in Van Lake by using zero-inflated generalized Poisson regression. \textit{Turkish Journal of Entomology}. 35(2).
It is our hope that this book will be of interest and use not only to scientists, but also to the food-producing industry, governments, politicians and consumers as well. If we are able to stimulate this interest, albeit in a small way, we have achieved our goal.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:

Abdullah Yeşilova, M. Salih Özgökçe and Yılmaz Kaya (2012). Zero-Inflated Regression Methods for Insecticides, Insecticides - Basic and Other Applications, Dr. Sonia Soloneski (Ed.), ISBN: 978-953-51-0007-2, InTech, Available from: http://www.intechopen.com/books/insecticides-basic-and-other-applications/statistical-analysis-for-insecticides
