

Use of Generalized Linear Mixed Model for Pest Density in Repeated Measurement Data

Heungsun Park* and Kijong Cho¹

ABSTRACT

The estimation of pest density is a prime concern of Integrated Pest Management (IPM) because the success of artificial intervention such as spraying pesticides or natural enemies depends on pest density. Also, the spatial pattern of pest population within plants or plots has been studied in various ways. In this study, we applied generalized linear mixed model to *Tetranychus urticae* Koch, two-spotted spider mite count in glasshouse grown roses. For this analysis, the subject-specific as well as population-averaged approaches are used.

KEY WORDS: GEE, GLIMMIX, random effect

1 INTRODUCTION

The two-spotted spider mites (TSSM), *Tetranychus urticae* Koch (Acari: Tetranychidae), are among the most serious pests of many ornamental plants, including the glasshouse rose which is the most important cut flower plant. This species causes damage to glasshouse roses not only by destruction of chloroplasts, but also by their webbing which leads to aesthetic injury. Adult two-spotted spider mites are the sizes of 0.3-0.5 millimeters and are able to live for 30 days in 25 degrees Celsius.

To take a control of this pest, we need to know the distributional pattern or spatial pattern within plants, so that the relevant pesticides can be applied at the right spot. For this study, we randomly select 24 rose plants within a glass house and observed the mite counts once a month from June 1997 to November 1997. From a selected plant, we collected 3 leaves (high, medium, low) from each vertical position (down, mid, up), and the number of mites were counted from each collected leaf surface with a microscope.

The prime purpose of this study is to investigate the significance of the effects such as time, plant, position and leaves. To do this, we used Generalized Estimating Equation (GEE) for repeated data analysis, which Liang and Zeger (1986) presented as an extension of generalized linear models (McCullagh and Nelder, 1989) to the repeated data analysis, which is a population-averaged approach (Zeger, Liang and Albert, 1988). Besides this, as a subject-specific approach, we used penalized quaslikelihood (PQL) estimation method which Breslow and Clayton (1993) developed to handle random effects within the generalized

¹Heungsun Park is Associate Professor, Department of Statistics, Hankuk University of Foreign Studies, 449-791, Yongin. hspark@hufs.ac.kr; Kijong Cho is Associate Professor, Department of Agricultural Biology, Korea University, 136-701, Seoul. kjcho@korea.ac.kr

linear model framework, and used the algorithm Wolfinger and O'connell (1993) presented in a SAS macro program (SAS, 1992), GLIMMIX.

2 GEE and PQL

When $Y_i = (y_{i1}, y_{i2}, \dots, y_{it_i})^T$ denotes the $t_i \times 1$ ($t_i \leq t$) vector of outcome and $X_i = (x_{i1}^T, x_{i2}^T, \dots, x_{in}^T)$ the $t_i \times p$ matrix of covariates, where x_{ij} is a $1 \times p$ covariate vector for $i = 1, \dots, n$. The marginal density of y_{ij} is

$$f(y_{ij}) = \exp \left[\frac{y_{ij}\theta_{ij} - b(\theta_{ij})}{a(\phi)} + c(y_{ij}, \phi) \right]. \quad (1)$$

The corresponding mean and variance of y_{ij} are derived as

$$E(y_{ij}) = \mu_{ij} = b'(\theta_{ij}), \quad \text{var}(y_{ij}) = V(\mu_{ij})a(\phi), \quad (2)$$

where $V(\mu_{ij}) = b''(\theta_{ij})$ is the variance function and ϕ is an unknown scale parameter. When we let $\eta_i = X_i\beta$, where $\eta_i = (\eta_{i1}, \dots, \eta_{it_i})^T$ with $\eta_{ij} = g(\mu_{ij})$ and $\eta_{ij} = x_{ij}\beta$. Here $\beta = (\beta_1, \beta_2, \dots, \beta_p)^T$ is the $p \times 1$ vector of unknown parameters and $g(\cdot)$ is a link function.

Note that for binary outcome variables, $g(\mu_{ij}) = \log(\mu_{ij}/(1 - \mu_{ij}))$, $V(\mu_{ij}) = \mu_{ij}(1 - \mu_{ij})$ and $a(\phi) = 1$. For Poisson count outcomes, $g(\mu_{ij}) = \log(\mu_{ij})$, $V(\mu_{ij}) = \mu_{ij}$ and $a(\phi) = 1$. If $S_i = Y_i - \mu_i$ with $\mu_i = (\mu_{i1}, \dots, \mu_{it_i})^T$ and $D_i = \partial\mu_i/\partial\beta$, the quaslikelihood estimation type equation for the i -th subject is

$$D_i^T V_i^{-1} S_i = 0 \quad (3)$$

where V_i is equal to $\text{cov}(Y_i)$. Therefore, the generalized estimating equation for the repeated data for n independent subjects is defined as

$$\sum_{i=1}^n D_i^T V_i^{-1} S_i = 0. \quad (4)$$

Then V_i , the $t_i \times t_i$ working covariance matrix can be partitioned into

$$V_i = (A_i \Phi_i)^{1/2} R_i(\alpha) (A_i \Phi_i)^{1/2} \quad (5)$$

where $A_i \Phi_i$ is a $t_i \times t_i$ diagonal matrix with $\text{var}(Y_i) = A_i \Phi_i$, Φ_i is defined as $\Phi = \text{diag}(\phi_1, \dots, \phi_t)$, and A_i is a $t_i \times t_i$ matrix defined as $A_i = \text{diag}(b''(\beta))$.

As for the penalized quaslikelihood estimation, we consider a linear (nonlinear) link function $g(\cdot)$ such as

$$g(\mu) = X\alpha + Z\beta \quad (6)$$

where α is a vector of unknown fixed effects with a covariate matrix X , and β is a vector of unknown random effects with a corresponding matrix Z . Assume $E(\beta) = 0$ and $\text{cov}(\beta) = D$, where D is unknown. Also, $e = y - \mu$ is a unknown error with $E(e|\mu) = 0$ and

$$\text{cov}(e|\mu) = R_\mu^{1/2} R R_\mu^{1/2} \quad (7)$$

Table 1: Parameter Estimates in population-averaged model (GEE) with Compound Symmetry structure for Two-Spotted Mite Count data

Parameter	DF	Estimate	S.E.
Intercept	1	2.1347	0.0423
down	1	-0.3030	0.0731
mid	1	-0.5463	0.0778
up	0	0.0000	0.0000
date 1	1	-0.4314	0.0684
date 2	1	-0.4925	0.0669
date 3	1	-1.2812	0.0892
date 4	1	-1.9570	0.1158
date 5	1	-1.6148	0.1039
date 6	0	0.0000	0.0000

where R_μ is a diagonal matrix containing evaluations at μ of a known variance function and R is unknown. To maximize l over D and R , we need a numerical method such as Newton-Raphson or EM algorithm. Once we get \hat{D} and \hat{R} , the estimates for α and β are obtained by solving the mixed model equation defined as

$$H \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} X^T W^{1/2} \hat{R}^{-1} W^{1/2} \nu \\ Z^T W^{1/2} \hat{R}^{-1} W^{1/2} \nu \end{bmatrix} \quad (8)$$

where

$$H = \begin{bmatrix} X^T W^{1/2} \hat{R}^{-1} W^{1/2} X & X^T W^{1/2} \hat{R}^{-1} W^{1/2} Z \\ Z^T W^{1/2} \hat{R}^{-1} W^{1/2} X & Z^T W^{1/2} \hat{R}^{-1} W^{1/2} Z + \hat{D}^{-1} \end{bmatrix}. \quad (9)$$

The results of solving the equations are followings:

$$\begin{aligned} \hat{\alpha} &= (X^T \hat{V}^{-1} X)^{-1} X^T \hat{V}^{-1} \nu \\ \hat{\beta} &= \hat{D} Z^T \hat{V}^{-1} \hat{r} \end{aligned} \quad (10)$$

3 SAS Programs

PROC GENMOD in SAS software provides a way to handle an algorithm of GEE for repeated data analysis. With mite data, we adopted GEE method as in Appendix.

The estimation of fixed effects helps us to understand the spatial distribution of mites within plants. This spatial distribution is of prime interest for plant farmers because they want to spray the pesticides as little as possible and kill the pest efficiently. The Table 1 gives us a hint that the density of mites are distributed differently along the vertical position. Also, date effect is significant so that we can say mite density differs day by day.

In order to investigate more detail, we can use contrast statement as shown in previous code. The result shows that down, mid and up position have different density of mites. In

Table 2: Contrast Results for GEE Analysis

Contrast	DF	Square	Pr > ChiSq	Type
pos mid to up	1	10.35	0.0013	Score
pos down to up	1	8.89	0.0029	Score
pos down to mid	1	0.00	0.9801	Score

Table 3: Parameter Estimates in population-averaged model (GEE) with Compound Symmetry structure for Two-Spotted Mite Count data

Parameter	DF	Estimate	S.E.
Intercept	1	2.0872	0.1280
down	1	-0.3539	0.1702
mid	1	-0.5050	0.1806
up	0	0.0000	0.0000
date 1	1	-0.4630	0.1552
date 2	1	-0.5036	0.1518
date 3	1	-1.2978	0.2028
date 4	1	-1.9681	0.2635
date 5	1	-1.6448	0.2370
date 6	0	0.0000	0.0000

detail, there is significant difference between down and up or between mid and up, but not between mid and down.

The same mite data observed within 6 equally spaced period are analyzed in penalized quaslikelihood estimation method via GLIMMIX macro program. The relevant code is described as in Appendix.

As shown above, the main difference from GEE is to handle random effect. (i.e. I treat plant effect as random) As I select 24 plants among their population group randomly, I believe that this is a more reasonable approach than the previous GEE method. Our prime interest lies in searching for significant effects, so that I investigate the fixed effects first.

As you see, there exists the vertical spatial pattern as well as day to day difference. These results coincide with that of GEE even though the estimates are slight different.

Table 4: Contrast Results for GLIMMIX Analysis

Label	Estimate	Standard			
		Error	DF	t	Pr > t
pos mid to up	-0.5029	0.1338	1090	-3.76	0.0002
pos down to up	-0.5156	0.1406	1090	-3.67	0.0003
pos down to mid	-0.01263	0.1622	1090	-0.08	0.9380


```
class plant position lnum date;
model tssm = position date position*date/solution;
random plant;
repeated /type=cs sub=lnum(plant*position) rcorr;
estimate 'pos mid to up' position 0 1 -1;
estimate 'pos down to up' position 1 0 -1;
estimate 'pos down to mid' position 1 -1 0;
), error=poisson, link=log )
run;
```

REFERENCES

- Breslow, N.E., and Clayton, D.G. (1993), "Approximate Inference in Generalized Linear Mixed Models", *Journal of the American Statistical Association*, 88, 9-25.
- Liang, K.Y. and Zeger, S.L. (1986) "Longitudinal data analysis using generalized linear models", *Biometrika*, 73, 13-22.
- McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear Models*, 2nd ed. London: Chapman and Hall.
- SAS Institute Inc. (1992) SAS Technical Report P-229. SAS/STAT Software, Proc Mixed Procedure, Cary, NC.
- Wolfinger, R. and O'connell, M. (1993), "Generalized Linear Mixed Models: A Pseudo-Likelihood Approach", *Journal of Statistical Computation and Simulation*, 48, 233-243.
- Zeger, S.L., Liang, K.Y. and Albert, P. (1988) "Models for longitudinal data: A generalized estimating equation approach", *Biometrics*, 44, 1049-1060