Acta Scientiarum
http://periodicos.uem.br/ojs/acta
ISSN on-line: 1807-8664
Doi: 10.4025/actascitechnol.v41i1.41646

Generalized Mixed Models - an application to longitudinal data of citrus canker

João Pedro Serenini, Terezinha Aparecida Guedes, Juliana Glória Franco Roberto and William Mário de Carvalho Nunes

Universidade Estadual de Maringá, Avenida Colombo, 5790, 87020-900, Maringá, Paraná, Brazil. *Author for correspondence. E-mail: jp_serenini@hotmail.com

ABSTRACT. There are several techniques available for longitudinal data analysis. In the last decade, much emphasis has been placed on generalized mixed models. The present work is dedicated to give an overview of this technique, with emphasis on the formulation, interpretation and inference of the model. The guidelines are given for statistical practice in general. This form of modeling was applied to data from an experiment to evaluate the resistance of 12 varieties of sweet orange to citrus canker. The experiment consisted of provoking lesions on the leaves of orange trees and monitoring the diameter of the lesion over time. The adjustment of the observed data to the proposed model provided reliable results, since the assumptions necessary for the validity of the model were satisfied. Therefore, it can be said that this methodology is adequate to model the data, since it allowed the detection of the varieties more susceptible to citrus canker.

Keywords: orange; repeated measurements/measures; mixed models.

Received on February 8, 2018.
Accepted on April 26, 2018

Introduction

The study of longitudinal data or repeated measurements involves the performance of two or more observations in the same experimental units. This type of study may present a hierarchical structure, since repeated measures are nested within the individual. Therefore, it can be assumed that observations between individuals are independent and that nesting in them has dependence correlated with errors.

Longitudinal studies are more appropriate for investigating individual changes over time and for studies of age effects and other factors influenced by the passage of time (Molenberghs & Verbeke, 2001).

In longitudinal data analysis, correct inferences can be obtained only by taking into account the correlation between the repeated measures in each subject, which also depends on the time interval between these measures (Molenberghs & Verbeke, 2001).

One of the main statistical models for analyzing longitudinal data capable of incorporating the dependency and correlation structure of the errors are the mixed models (Wallace & Green, 2002). However, in order to resolve the non-independence of the data, it is necessary to stipulate correctly its fixed and random effects (Winter, 2013).

A mixed model is an extension of a linear model. This model is divided into two parts, one of fixed effects (already present in the linear models) and another of random effects, where the latter has the function of identifying the variation between the individuals of the problem in question. In this sense while the fixed coefficients are constant in all individuals, the random ones vary for each one, which in turn incorporates the differences between them. Therefore, this causes the model to have two parametrization components, one intra-individual and one between-individual.

A random-effect variable is one that in the study in question represents only a fraction of all levels that this variable can assume, or where the effect of each individual level is not of interest in the results of the study, but it is desired to know about its variability in the general population (Littell, 2002). In general, the type of effect that each variable will assume depends, in part, on the research objective.

The residual errors are a random-effect variable in every experiment, in addition, certain aspects in the experiment and data collection create random effects for the model (Schabenberger & Pierce, 2001).

According to Littell, Pendergast, and Natarajan (2000), the mixed models were developed by geneticists to assess the genetic potential of bulls. With the advancement of computational technology, the application of such models has spread through several areas of research.

The purpose of this article is to provide a discussion on mixed generalized linear models, summarizing from a current perspective its formulation, interpretation and implementation in research and new developments, especially in the agricultural sciences.

Material and methods

Material

The methods discussed will be applied in the evaluation of the resistance of 12 types of orange genotypes (Citrus sinensis) to the bacteria *Xanthomonas citri* subsp. *citri*, in vegetation house/greenhouse, by measuring the diameter of the lesions, Valência, Valência 2, Valência Puka, Valência Paloma, Perâ Oriçanga, Perâ Irradiada, Perâ Itapetininga, Perâ Maringá, Perâ IAC, Sanguine Mombuca, Hamlin and Precoce Oriçanga.

The samples were obtained from the inoculation of the bacteria in 6 leaves per plant, totaling 5 plants per variety, which received periodic fertilizations of N, P, K. The diameter of the lesions provoked was evaluated at 18, 24, 33, 39, 46, 61, 68, and 75 days after inoculation (DAI), where the presence of yellow halo around necrosis was disregarded. The experiment was carried out from June to August 2016. For the measurement of the diameters, a pachymeter was used, using three different grafts, Swingle, Sunki and Lemon clove/Rangpur Lime.

The data set is of the longitudinal type, consisting of a total of 2880 observations containing 4 variables, these being the varieties of oranges used in the experiment (variety), the number of plants in each variety (leaf), the days after inoculation (DAI) and the diameter of the lesion (Diam), the latter being the response variable, which is of the continuous type, measured in centimeters.

Generalized Linear Mixed Model (GLMM)

A generalized mixed model is a statistical model that seeks to describe the behavior of a random variable in relation to one or more explanatory variables, where these can be both of fixed and random effects. The great difference of these models in relation to a mixed linear model is that in the generalized ones the response variable does not necessarily have to be associated with the normal distribution, what widens the scope of such models.

To specify a generalized mixed model, the response variable conditioned to random effects ($y_i \vee b$) must follow a distribution belonging to the exponential family, according Equation 1 and 2:

$$y_i \vee b f_{(Y_i|b)}(y|i \vee \theta) \quad (1)$$

$$f_{(Y_i|b)}(y_i|\theta) = b(y_i)exp[\eta(\theta)^T T(y_i) - a(\theta)] \quad (2)$$

Typically, is desired to relate the distribution parameters to the predictors of the model, since this is make a transformation of the mean μ_i, so that they relate to such predictors, according Equation 3 and 4:

$$E(y_i|b)=\mu_i, \quad (3)$$

$$g\big(E(y_i|b)\big) = g(\mu_i) = \eta_i = X_i\beta + Z_i b, \quad (4)$$

where:

$g(.)$ is a known function, called link function, responsible for relating the mean of y_i to the linear predictors, X_i is the i-th line of the fixed-effect matrix, β is the vector of the fixed-effect parameters, Z_i is the i-th line of the random-effect model matrix, and b is the vector of random effects assuming that they are mutually independent, normally distributed with vector means equal to 0 and variance matrix D.

As can be noticed in GLMMs, as in mixed linear models, there are two types of random effects. Simply put, if a random effect is an element of b, it is taken as an effect D, that is, in such a model the covariance structure of matrix D is being modeled, otherwise the covariance structure of the errors R is being modeled. Non-effect D models are also known as marginal models. The same model can contain both types of random effects, when this occurs the covariance matrix of Y is written as: $cov\ (Y) = V = ZDZ + R$ (SAS Institute, 2008).

The variance of $y_i \vee b$ is given by $Var(y_i \vee b) = \phi^2 v(\mu_i)$, where v is a known variance function given by the distribution of y_i and ϕ is the scale parameter of the distribution, which must be estimated.

In this context the mean, variance, covariance and correlation can be written as Equation 5 at 7:

$$E[y_i] = E[g^{-1}(X_i\beta + Z_ib)] \tag{5}$$

$$\begin{aligned} Var(y_i) &= Var\big(g^{-1}(X_i\beta + Z_ib)\big) + E\big[\phi^2 v\big(g^{-1}(X_i\beta + Z_ib)\big)\big] \\ Cov\big(y_i, y_j\big) &= cov\left(g^{-1}(X_i\beta + Z_ib), g^{-1}\big(X_j\beta + Z_jb\big)\right) \end{aligned} \tag{6}$$

$$Corr\big(y_i, y_j\big) = \frac{cov\left(g^{-1}(X_i\beta + Z_ib), g^{-1}\big(X_j\beta + Z_jb\big)\right)}{\sqrt{Var(y_i)}\sqrt{Var\big(y_j\big)}}. \tag{7}$$

In generalized mixed models, such as the linear ones, the parameters of fixed effects β, the variances and covariance of the random effects *D* and *R* are estimated. However, unlike linear models in GLMMs, the likelihood function does not usually have a closed-form expression, which generates integrals that cannot be solved analytically. Therefore, several approaches can be used in order to overcome such difficulties. One is done by maximizing the penalized likelihood function l_p (Rigby & Stasinopoulos, 2005), given by Equation 8:

$$l_p(\beta, b) = l - \frac{1}{2}\sum_{i=1}^{N} b_{i\mu}^T\big(D_{i\mu}^{-1}\big)b_{i\mu} + b_{i\sigma}^T(D_{i\sigma}^{-1})b_{i\sigma} \tag{8}$$

where:

$l = \sum_{i=1}^{n} ln\big(f(y_i \vee \mu, \sigma^2)\big)$ is the log-likelihood function of the data, b are the parameters of random effects and D_i^{-1} is the inverse matrix of D_i, the matrix of diagonal blocks (q x q) of the variances and covariances of the random effects b_i.

Log-normal distribution

The log-normal distribution is an appropriate distribution for continuous, positive and asymmetric data, whose asymmetry occurs frequently when the means are small, the variances are large and the values are not negative (Limpert, Stahel, & Abbt, 2001), where the logarithm of the random variable is normally distributed. The probability density function can be defined according to Equation 9 (Stasinopoulos, Rigby, & Akantziliotou, 2008):

$$f(y|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\frac{1}{y}exp\left\{\frac{-(log(y) - \mu)^2}{2\sigma^2}\right\}, \tag{9}$$

where:

for $y > 0$, where $\mu > 0$ and $\sigma > 0$. $E(Y) = \omega^{\frac{1}{2}}e^{\mu}$ and $Var(Y) = \omega(\omega - 1)e^{2\mu}$, with $\omega = exp(\sigma^2)$, whose support is given by $supp(f) = x \in (0, \infty)$.

Gamma distribution

The Gamma distribution is an appropriate distribution for continuous, positive and asymmetric data, where several distributions such as exponential and chi-square are particular cases of it. When parameterized as a function of its mean μ, its probability density function denoted by $GA(\mu, \sigma^2)$ can be defined as Equation 10:

$$f(y|\mu, \sigma^2) = \frac{1}{(\sigma^2\mu)^{\frac{1}{\sigma^2}}} \cdot \frac{y^{\left(\frac{1}{\sigma^2} - 1\right)}e^{-y/(\sigma^2\mu)}}{\Gamma(1/\sigma^2)}, \tag{10}$$

where:

for $y > 0$, where $\mu > 0$ and $\sigma > 0$. $E(Y) = \mu$ and $Var(Y) = \sigma^2\mu^2$. Such reparameterization is presented by Johnson, Kotz, and Balakrishnan (1995), obtained by determining $\sigma^2 = \frac{1}{\alpha}$ and $\mu = \alpha\beta$. Additionally, μ is the location parameter, where as σ is the scale parameter. The support of the function is given by $supp(f) = x \in (0, \infty)$. Also, for any positive real α, $\Gamma(\alpha)$ is defined as: $\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1}e^{-x}dx$.

Generalized mixed model for gamma and log-normal distributions

In a context of repeated measures, a mixed-gamma and log-normal model with normal random effects is defined, with Y denoting random and independent observations, where $Y_i = (Y_{i1}, \ldots, Y_{in_i})$, with i = 1, ..., N observations and j = 1, ..., n_i moments. Being two random effects, $b_{i\mu}$ and $b_{i\sigma^2}$, which as seen are independent and with Normal distribution with vector means equal to 0 and covariance matrices D_μ and D_σ^2. Therefore, the regression model can be written as Equation 11 at 13 (Wallace & Green, 2002):

$$Y_{ij} \vee b_{i\mu} b_{i\sigma^2} \sim GA(\mu_{ij}, \sigma_{ij}^2), \quad (11)$$

$$Y_{ij} \vee b_{i\mu} b_{i\sigma^2} \sim ln(\mu_{ij}, \sigma_{ij}^2), \quad (12)$$

$$i = 1, \ldots, N, j = 1, \ldots, n_i \; b_{i\mu} \sim N(0, D_\mu), b_{i\sigma^2} N(0, D_{\sigma^2}), \quad (13)$$

The parameters μ_{ij} and σ_{ij} satisfies the relations Equation 14 and 15:

$$g_\mu = \eta_{ij\mu} = x_{ij\mu}^T \beta_\mu + Z_{ij\mu}^T b_{i\mu} \quad (14)$$

$$g_{\sigma^2} = \eta_{ij\sigma^2} = x_{ij\sigma^2}^T \beta_{\sigma^2} + Z_{ij\sigma^2}^T b_{i\sigma^2} \quad (15)$$

where:

- $x_{ij\mu} = \left(x_{ij\mu 1}, \ldots, x_{ij\mu p_\mu}\right)$ and $x_{ij\sigma^2} = \left(x_{ij\sigma^2 1}, \ldots, x_{ij\sigma^2 p_{\sigma^2}}\right)$ are matrices with the observations of the predictor variables, associated with the i-th individual in the j-th moment of the fixed effects of μ and σ^2;
- $Z_{ij\mu} = \left(Z_{ij1\mu, \ldots, Z_{ijq_\mu\mu}}\right)^T$ and $Z_{ij\sigma^2} = \left(Z_{ij1\sigma^2, \ldots, Z_{ijq_{\sigma^2}\sigma^2}}\right)^T$ are matrices with the observations of the predictor variables, associated with the i-th individual in the j-th moment of the random effects of μ and σ^2;
- $\beta_\mu = \left(\beta_0, \ldots, \beta_{(p-1)_\mu \mu}\right)^T$ and $\beta_{(\sigma^2)} = \left(\beta_0, \ldots, \beta_{(p-1)_{\sigma^2}\sigma^2}\right)^T$ are vectors of the fixed effects associated with each of the parameters of the distributions μ and σ^2;
- $b_{i\mu} = \left(b_{i1\mu}, \ldots, b_{iq_\mu\mu}\right)^T$ and $b_{i\sigma^2} = \left(b_{i1\sigma^2}, \ldots, b_{iq_{\sigma^2}\sigma^2}\right)^T$ are vectors of the random effects associated with the i-th individual in the j-th moment of each of the parameters of the distributions μ and σ^2 ;
- D_μ and D_{σ^2} are matrices $q_\mu \times q_\mu$, $q_{\sigma^2} \times q_{\sigma^2}$ of diagonal blocks, which represent the variance of $\gamma_{i\mu}$ and $\gamma_{i\sigma^2}$, respectively.

Applications

The log-normal $ln(\mu, \sigma^2)$ and gamma distributions $GA(\mu, \sigma^2)$ were used to analyze the database obtained by Franco (2016) with the objective of evaluating the resistance of 12 varieties of sweet orange.

The data set contains 2880 observations of the mean diameter of the lesions (12 Genotypes × 8 Days × 60 Leaves × 6 Repetitions), considering all co-variables and the effect of time. To obtain the parameter estimates, it was used the software *R* with the package *gamlss*, which allows the expansion of the systematic part of the model, allowing to model not only the average parameter, but also other parameters of the distribution. The penalized likelihood function l_p (8) was used to estimate the parameters of the models.

This function may not have an analytical solution, or if it exists, is very difficult to obtain. To overcome this difficulty, three algorithms have been implemented in the package *gamlss* to obtain the estimators, RS, CG and an algorithm called Mixed (Stasinopoulos et al., 2008). In the present work we used only the algorithm RS, because for many probability functions, whose expected values of the second-order derivative matrix of the likelihood function are zeros, except for the main diagonal, RS is used because it is simpler, in addition, this algorithm is faster for large data set, while the CG is more suitable for distributions whose parameters are highly correlated.

Description of basic statistics

This assay contains 2880 observations of the average diameter of the lesions (12 varieties × 8 days × 60 leaves × 6 repetitions), considering all covariates and time effect.

Table 1 presents a summary of the variable response of the diameter of the lesions in relation to each variety.

From Table 1 it can be seen that the means of the varieties vary relatively greatly, ranging from 1.71 cm for the variety Valencia to 2.45 cm for the variety Pera Ori, with standard deviations equal to 0.25 and 0.47 cm, respectively. In addition, a great variability is observed for the maximum and minimum values for each variety. It is also noted that the values of the coefficients of variation vary from 0.13 to 0.19, representing how much, on average, the deviations reach the value of the sample mean.

Figure 1 presents the box-plot for the behavior of the diameter size of the response variable by variety type.

From Figure 1 apparently the Irradiada and the two species of Valencia have lower average levels of diameter size in relation to the other varieties, whereas the PeraOri. has the highest average level. The other varieties have seemingly close average levels. There are no candidate measures for outliers.

Figure 2 shows the average profiles of the evaluated varieties. It is noted that, over time, apparently, the diameter of the lesion increases in all varieties, in addition, it is noted again that the two varieties of Valencia and Irradiada have the smallest measurements for this, and PeraOri, the largest. No significant major decays are observed over time, as observed in the first assay.

Adjustment

In the adjustment of the generalized mixed models based on the log-normal and gamma distributions we attempted to analyze the behavior of citrus canker in sweet orange varieties in order to analyze the relationship between the variable Diameter of the Lesion and the possible factors of influence, variety and days after inoculation, with these belonging to the fixed part of the model and the leaves belonging to the random part.

The models were adjusted in response variable y_{ijk}, i = 1, 2, ..., 12 (number of varieties), j = 1, 2, ..., 8 (days after inoculation), k = 1, 2, ..., 60 (number of leaves). The selection of the model was based on the Akaike Information Criterion (AIC), Global Deviance (GD) and Schwarz Bayesian Criterion (BIC) (Rigby & Stasinopoulos, 2010). The adjusted models are described in Table 2.

In such models α_i represents the effects due to the i-th variety, θ_j the effects due to the j-th day after inoculation, u_k is the effect of the random part on the k-th leaf.

Table 1. Statistical summary of the diameters of the lesions of citrus canker for each of the12 varieties.

No.	Varieties	N	Mean	Standard Deviation	Min.	Max.	Coef. of Var.
1	PeraOri	240	2.45	0.47	1.61	3.48	0.19
2	Valen	240	1.71	0.25	1.13	2.23	0.15
3	PrecOri	240	2.31	0.43	1.73	3.13	0.19
4	Sanguine	240	1.94	0.25	1.38	2.50	0.13
5	Paloma	240	2.02	0.32	1.38	2.76	0.16
6	PeraIAC	240	2.11	0.34	1.42	2.82	0.16
7	Puka	240	1.99	0.29	1.35	2.55	0.15
8	VALEN	240	1.80	0.24	1.21	2.37	0.13
9	Itapetinig	240	2.32	0.43	1.65	3.23	0.18
10	PeraMga	240	2.16	0.34	1.50	2.99	0.16
11	Hamlin	240	2.25	0.43	1.45	3.17	0.19
12	Irradiada	240	1.84	0.24	1.30	2.39	0.13

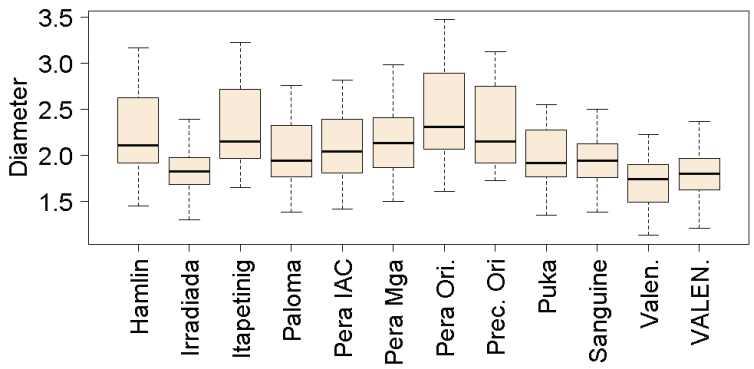

Figure 1. Box-plot of the data by variety for 2880 observations.

The values of all the decision measures of each model are presented in Table 3.

Through such decision measures and the GAIC stepwise algorithm (Stasinopoulos & Rigby, 2007), implemented in the Gamlss package, the models LN_3 and GA_3 were chosen as the best models for the Log-Normal and Gamma distributions, since among those that best explain the data, these are the most parsimonious.

The analysis of randomized quantile residuals are presented in Figure 3 and 5) for the Log-Normal and Gamma models (Dunn & Smyth, 1996).

Figure 3a and b, of the adjusted values *vs* residual quantiles, show that the residuals are distributed in a randomly around zero, which shows no violation of homoscedasticity. Figure 3c and d, of the theoretical quantiles *vs* sample quantiles, indicate that the residual quantiles follow a standard normal distribution, which is a requirement for proper adjustment.

In the adjustment of such model, the logarithmic function was used as the link function for the Log-Normal, as well for the Gamma model. In Table 4, the fixed-parameter estimates of the Log-normal and Gamma models are presented.

The results presented in Table 4 show how much the diameter of the lesions caused by *Xanthomonas* varies in average in relation to the Hamlin variety, and how much the diameter of the lesions varies in average with respect to the time compared to day 18. It was observed that the two varieties of Valencia and Irradiada were the ones that presented the lowest average diameter growth over time, while Perâ Oriçanga was the one with the largest diameter, followed by Itapetinig and Precoce Oriçanga. For the days after inoculation, it is observed that over time the diameter tends to increase, where day 75 presented the largest diameter measurements. These results confirmed what the previous descriptive analysis showed.

Table 2. Models proposed for Log-normal and Gamma distribution.

No.	Log-normal	Gamma	IDModel
1	LN_0	GA_0	$\eta = \beta_0$
2	LN_1	GA_1	$\eta_i = \beta_0 + \alpha_i Variety$
3	LN_2	GA_2	$\eta_{ij} = \beta_0 + \alpha_i Variety + \theta_j DAI$
4	LN_3	GA_3	$\eta_{ijk} = \beta_0 + \alpha_i Variety + \theta_j DAI + u_{0k}$
5		GA_4	$\eta_{ijk} = \beta_0 + \alpha_i Variety + \theta_j DAI + u_{1k} Variety$

Table 3. Measures of AIC, DG and BIC for Log-normal and Gamma models.

Distribution	No.	Model	AIC	GD	BIC
Log-normal	1	LN_0	2758.685	2754.685	2770.616
	2	LN_1	1751.555	1725.555	1829.107
	3	LN_2	−3848.887	−3888.887	−3729.577
	4	LN_3	−3917.803	−4036.911	−3562.531
Gamma	1	GA_0	2820.178	2816.178	2832.109
	2	GA_1	1768.072	1742.072	1845.624
	3	GA_2	−3845.892	−3885.892	−3726.581
	4	GA_3	−3913.806	−4032.41	−3560.038
	5	GA_4	−3937.298	−4020.624	−3688.754

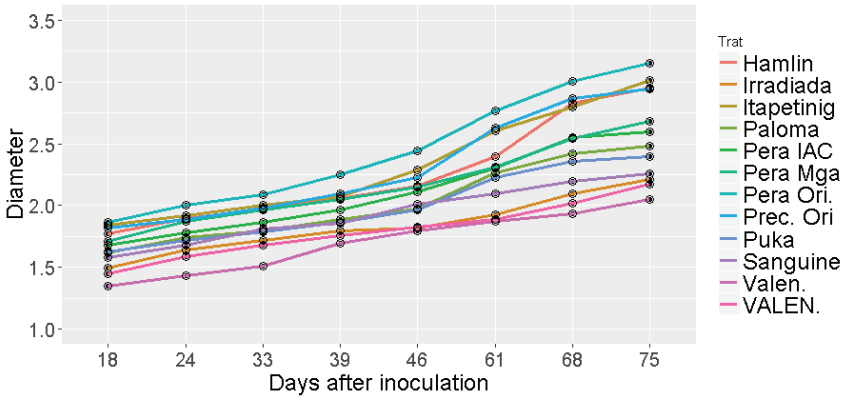

Figure 2. Average profile for each variety over time.

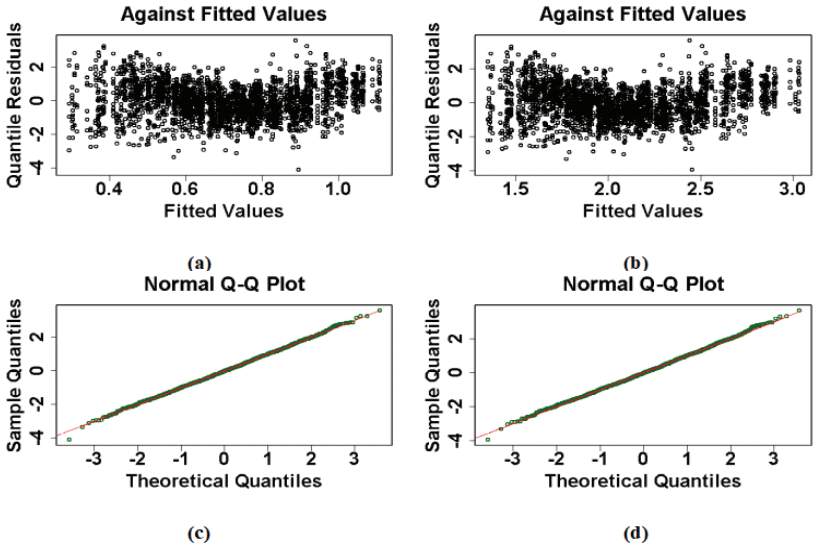

Figure 3. (a), (b) Adjusted Values vs Residual Quantiles for Log-Normal and Gamma models; (c), (d) Theoretical Quantiles vs Sample Quantiles for Log-Normal and Gamma models.

Table 4. Estimates of Log-Normal and Gamma models parameters for μ.

a) Log-normal model					
Variable	Parameter	Estimate	Standard Error	T value	Pr (> |t|)
Intercept	β_0	0.580	0.005	115.920	< 2e-16
Irradiada	α_1	-0.194	0.006	-35.232	< 2e-16
Itapetinig	α_2	0.027	0.005	5.263	1.52e-07
Paloma	α_3	-0.104	0.005	-22.867	< 2e-16
PeraIAC	α_4	-0.061	0.006	-13.559	< 2e-16
PeraMga	α_5	-0.037	0.005	-7.611	3.72e-14
PeraOri	α_6	0.084	0.005	16.940	< 2e-16
PrecOri	α_7	0.022	0.005	4.387	1.19e-05
Puka	α_8	-0.116	0.005	-24.558	< 2e-16
Sanguine	α_9	-0.139	0.006	-25.456	< 2e-16
Valen	α_{10}	-0.266	0.006	-45.835	< 2e-16
VALEN2	α_{11}	-0.214	0.006	-36.443	< 2e-16
DAI 24	θ_1	0.066	0.005	14.154	< 2e-16
DAI 33	θ_2	0.112	0.005	24.102	< 2e-16
DAI 39	θ_3	0.161	0.004	36.570	< 2e-16
DAI 46	θ_4	0.220	0.004	50.508	< 2e-16
DAI 61	θ_5	0.320	0.005	68.757	< 2e-16
DAI 68	θ_6	0.400	0.005	83.994	< 2e-16
DAI 75	θ_7	0.439	0.005	90.106	< 2e-16
b) Gamma model					
Variable	Parameter	Estimate	Standard Error	T value	Pr (> |t|)
Intercept	β_0	0.479	0.017	28.973	< 2e-16
Irradiada	α_1	-0.294	0.018	-16.516	< 2e-16
Itapetinig	α_2	0.027	0.016	1.738	0.0823
Paloma	α_3	-0.027	0.016	-1.723	0.0849
PeraIAC	α_4	-0.034	0.015	-2.296	0.0217
PeraMga	α_5	-0.022	0.015	-1.411	0.1585
PeraOri	α_6	0.064	0.016	3.909	9.47e-05
PrecOri	α_7	0.010	0.015	0.680	0.4964
Puka	α_8	-0.067	0.015	-4.404	1.10e-05
Sanguine	α_9	-0.142	0.016	-8.957	<2e-16
Valen	α_{10}	-0.341	0.018	-19.395	<2e-16
VALEN2	α_{11}	-0.345	0.018	-19.521	< 2e-16
DAI 24	θ_1	0.112	0.015	7.528	< 2e-16
DAI 33	θ_2	0.29303	0.016	18.328	< 2e-16
DAI 39	θ_3	0.25854	0.015	17.233	< 2e-16
DAI 46	θ_4	0.26140	0.015	17.856	< 2e-16
DAI 61	θ_5	0.42012	0.016	26.313	< 2e-16
DAI 68	θ_6	0.53002	0.017	31.719	< 2e-16
DAI 75	θ_7	0.47733	0.016	29.501	< 2e-16

In Table 5 the estimates of variances and covariances of the random part of the Log-Normal and Gamma models are presented for the parameter μ.

The variance estimates presented in Table 5, represent how much of the variance presented in the measurement of the diameter is due to the leaves and how much of the variance cannot be explained by the model (Residual).

The models Log-Normal and Gamma present Nagelkerke R^2 values equal to 0.9157485 and 0.9175203, respectively, that is, such models account for just over 90% of the variability of the diameter of the lesion (Nagelkerke, 1991).

Figure 4 shows the scatter plots with equality line and the Bland Altman plot for the Log-normal and Gama models.

In Figure 4a and b the values measured for the diameter are green and the values adjusted by the models are in red, in such graphs it is expected that such points will be equal measures, overlapping the blue line as much as possible, which is actually occurring. In Figure 4c and d a good adjustment is again noted, the line of the mean of the differences being very close to 0 and having 95% of the differences within the confidence intervals. In addition, the scales for the differences are close to 0.

Table 5. Estimates of the variances of the random part (leaf) for the Log-normal and Gamma μ model.

Distribution	Variable	Parameter	Variance	SD
Log-normal	Leaf	u_0	0.000187	0.01367
	Residual	e	0.003536	0.05946
Gamma	Leaf	u_0	0.000184	0.01356
	Residual	e	0.003549	0.05957

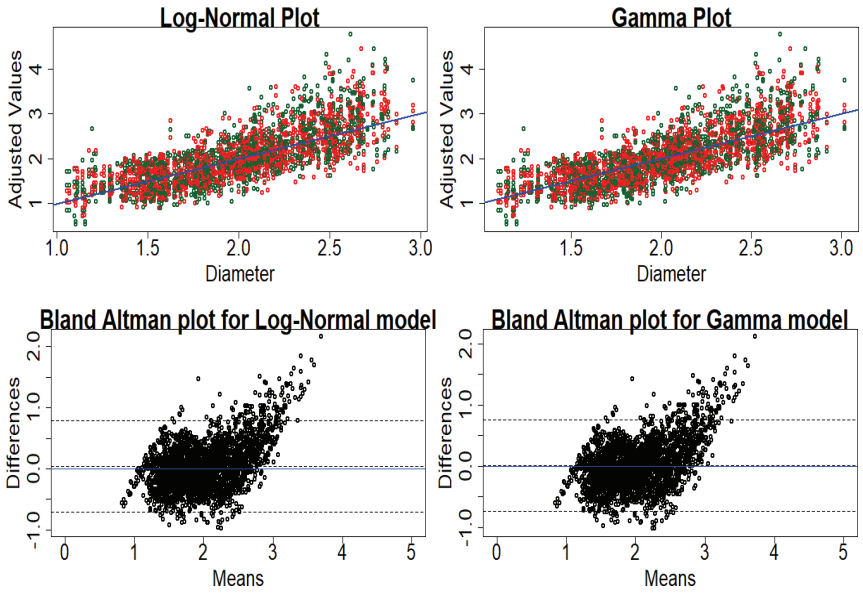

Figure 4. Scatter plots with equality line and Bland Altman plots for log-normal and gamma models.

Discussion

In this study, the methodology of mixed generalized linear models for random variables with log-normal and gamma distributions was discussed. These models were applied to data whose responses showed adherence to such distributions. The link between the predictors and the mean of the response variable was made by the logarithmic function. Generalized linear mixed models allow us to introduce a source of dependence that is not taken into account in the usual regression models. This dependence is quantified by the correlation between observations that occurs for data observed over time. The introduction of random-effect factors into the model allows this correlation to be modeled (McCulloch, Searle, & Neuhaus, 2001). In this way, the model will present a better fit, since it will be able to capture the variability within the subjects whose measurements were performed over time. These models were adequate to describe the behavior of citrus canker evaluated in varieties of sweet orange.

In the experiment that gave rise to the data analyzed, the diameter of the lesion provoked in the leaves was measured over time in 12 varieties of sweet orange. The adjustment of the observed data to the proposed model provided reliable results since the assumptions necessary for the validity of the model were satisfactorily met. The two Valencia varieties showed greater resistance to citrus canker, followed by the variety Perâ Irradiada. On the other hand, the variety Perâ Oriçanga proved to be less resistant to the disease, followed by the varieties Perâ Itapetininga and Precoce Oriçanga. Therefore, it can be said that this methodology was adequate to model the data, since it allowed the detection of the varieties more susceptible to citrus canker, which in experiments of this type, is the objective of the researcher.

Conclusion

Generalized linear mixed models are in increasing use for the analysis of longitudinal data, they have the ability to predict the individual trajectories of the object in question over time, in addition to the fact that the introduction of random effects to the model counts a lot to model the existing correlation in repeated measures. In general, both the log-normal and gamma models were adequate to describe the behavior of citrus canker in the evaluated sweet orange genotypes presents in the study.

References

Dunn, P. K., & Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics, 5*(3), 236-244. doi: 10.2307/1390802.

Franco, J.G. (2017). *Avaliação da resitência de genótipos de laranja doce (Citrus sinensis) ao cancro cítrico em condições de casa de vegetação e em folhas destacada*. Maringá, PR: Universidade Estadual de Maringá.

Johnson, N. L., Kotz, S. I., & Balakrishnan, N. (1995) *Continuous univariate distributions*. New York, NY: John Wiley & Sons.

Limpert, E., Stahel, W. A., & Abbt, M. (2001). Log-normal distributions across the sciences: keys and clues: On the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize log-normal distributions, which can provide deeper insight into variability and probability normal or log-normal: That is the question. *AIBS Bulletin, 51*(5), 341-352. doi: 10.1641/0006-3568(2001)051[0341:LNDATS]2.0.CO;2

Littell, R. C. (2002). Analysis of unbalanced mixed model data: a case study comparison of ANOVA versus REML/GLS. *Journal of Agricultural, Biological, and Environmental Statistics, 7*(4), 472-490. doi: 10.1198/108571102816

Littell, R. C., Pendergast, J., & Natarajan, R. (2000). Tutorial in biostatistics: modelling covariance structure in the analysis of repeated measures data. *Statistics in Medicine, 19*(13), 1793-1819. doi: 10.1002/1097-0258(20000715)19:13<1793::AID-SIM482>3.0.CO;2-Q

McCulloch, C. E., Searle, S. R., & Neuhaus, J. M. (2001). *Generalized linear mixed models*. Bethesda, MD: John Wiley & Sons, Ltda.

Molenberghs, G., & Verbeke, G. (2001). A review on linear mixed models for longitudinal data, possibly subject to dropout. *Statistical Modelling, 1*(4), 235-269. doi: 10.1191/147108201128195

Nagelkerke, N. J. (1991). A note on a general definition of the coefficient of determination. *Biometrika, 78*(3), 691-692. doi: 10.1093/biomet/78.3.691

Rigby, R. A., & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 54*(3), 507-554.

Rigby, R. A., & Stasinopoulos, D. M. (2010). *A flexible regression approach using GAMLSS in R*. Lancaster, GB: University of Lancaster.

SAS Institute. (2008). *SAS/STAT 9.1 user's guide the reg procedure:(book excerpt)*. Cary, NC: SAS Institute.

Schabenberger, O., & Pierce, F. J. (2001). *Contemporary statistical models for the plant and soil sciences*. Washington, DC: CRC press.

Stasinopoulos, D. M., & Rigby, R. A. (2007). Generalized additive models for location scale and shape (GAMLSS) in R. *Journal of Statistical Software, 23*(7), 1-46. doi: 10.18637/jss.v023.i07

Stasinopoulos, M., Rigby, B., & Akantziliotou, C. (2008). *Instructions on how to use the gamlss package in R* (2nd ed.). Retrieved from http://www.gamlss.com/wp-content/uploads/2013/01/gamlss-manual.pdf

Wallace, D., & Green, S. B. (2002). Analysis of repeated measures designs with linear mixed models. In D. S. Moskowitz & S. L. Hershberger (Eds.)*, Multivariate applications book series. Modeling intraindividual variability with repeated measures data: Methods and applications* (p. 103-134). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Winter, B. (2013). *A very basic tutorial for performing linear mixed effects analyses*. Retrieved from http://www.bodowinter.com/tutorial/bw_LME_tutorial2.pdf