

# Analyzing weight evolution in mice infected by *Trypanosoma cruzi*

Breno Gabriel da Silva<sup>1\*</sup>, Terezinha Aparecida Guedes<sup>2</sup>, Vanderly Janeiro<sup>2</sup>, Érika Cristina Ferreira<sup>2</sup>, Silvana Marques de Araújo<sup>2</sup> and Larissa Ciupa<sup>3</sup>

<sup>1</sup>Programa de Pós-Graduação em Estatística e Experimentação Agronômica, Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Avenida Pádua Dias, 11, 13418-900, Piracicaba, São Paulo, Brazil. <sup>2</sup>Universidade Estadual de Maringá, Maringá, Paraná, Brazil. <sup>3</sup>Centro Universitário Ingá, Maringá, Paraná, Brazil. \*Author for correspondence. E-mail: brenogabriel\_silva95@outlook.com

**ABSTRACT.** Concerning the specificities of a longitudinal study, the trajectories of a subject's mean responses not always present a linear behavior, which calls for tools that take into account the non-linearity of individual trajectories and that describe them towards associating possible random effects with each individual. Generalized additive mixed models (GAMMs) have come to solve this problem, since, in this class of models, it is possible to assign specific random effects to individuals, in addition to rewriting the linear term by summing unknown smooth functions, not parametrically specified, then using the P-splines smoothing technique. Thus, this article aims to introduce this methodology applied to a dataset referring to an experiment involving 57 Swiss mice infected by *Trypanosoma cruzi*, which had their weights monitored for 12 weeks. The analyses showed significant differences in the weight trajectory of the individuals by treatment group; besides, the assumptions required to validate the model were met. Therefore, it is possible to conclude that this methodology is satisfactory in modeling data of longitudinal sort, because, with this approach, in addition to the possibility of including fixed and random effects, these models allow adding complex correlation structures to residuals.

**Keywords:** longitudinal data; mixed models; generalized additive mixed models; p-splines

Received on December 13, 2019.

Accepted on April 2, 2020.

## Introduction

Regression analysis is a statistical technique that allows investigating and modeling the relationship between two or more variables of interest, with the regression being linear if one considers that the relationship of the response variable (dependent variable) with the regressor variables (independent variables) is a linear function of some parameters, and non-linear if at least one of the partial derivatives in relation to the parameter is a function of parameters that are not known and that cannot be linearized by means of transformations.

Paulson (2006) reports that the literature contains methodologies focused on non-linear modeling, such as Gompertz, Von Bertalanffy, Brody, and others; according to Paro de Paz et al. (2004), they can be applied to describing growth curves, besides being parameters of easy biological interpretation. Moreover, segmented regressions, called spline functions, can be used as well.

According to Keele (2008), splines are segmented polynomial functions united by cutoff points known as "nodes", which are used for fitting a curve in a dataset. Spline functions have been widely used in the analysis of longitudinal data – for instance, in the fitting of lactation and growth curves, as seen in Caldeira (2011), Pinheiro et al. (2013), Domingues, Gonçalves, Braz, & Pereira (2014), and Santos et al. (2015). The literature provides several techniques to represent this fitting by parts, such as linear spline, cubic spline, B-spline, etc.

Authors Boor (1978), Eilers and Marx (1996), and Keele (2008) define that a B-spline can be expressed as a set of polynomial functions of the  $m$  order for each node-limited interval; also, the connections in the nodes are smoothed, which does not happen in linear splines, thus becoming one single continuous curve. In order to make the model more flexible and prevent a possible "overfitting", Eilers and Marx (1996) proposed a new technique, which consists of combining B-splines with discrete penalty (to control the smoothness of the fit) in the log-likelihood function, and named it P-spline.

Because it is not always possible to describe a phenomenon through a linear regression model, new classes of regression models have emerged, such as generalized linear models, generalized additive models, etc. When it comes to non-linear modeling techniques, Generalized Additive Mixed Models stand out; in them, the linear term is replaced by unknown smooth functions, giving more flexibility to the hypothesis of a linear relationship between the response variable and the regressor variables on the scale of a link function. Currie and Durbán (2002) highlight that this class of models allows estimating individual-difference curves using non-parametric smoothing functions, considering the presence of random effects associated with subjects assigned to treatment groups.

Thus, this research aims to use smoothing via P-spline associated with a generalized additive mixed model in order to model the weight evolution of 57 male Swiss mice, with 56 days of age, subjected to homeopathic treatments with different dilutions and observed for 12 weeks, all infected by *Trypanosoma cruzi* and sourced from the State University of Maringá's central vivarium.

## Material and methods

### Experiment description

The dataset for this research was collected by Ferreira et al. (2018) with the aim of assessing the biotherapeutic effect of chicken serum on rodents experimentally infected by *Trypanosoma cruzi*, considering clinical, immunological and parasitological parameters. The project for the execution of the experiments was approved by the Ethics Committee on Research Involving Animals of the State University of Maringá, Paraná, Brazil, under legal opinion CEUA 2401220716/2016.

Based on Ferreira et al. (2018), the experiments were run through blind, controlled and randomized assays. They used 57 male Swiss mice, aged 56 days, sourced from the central vivarium of the State University of Maringá. The mice were divided into treatment groups in a way that the initial mean weight of the animals in each group was approximately equal, and kept in cages with a maximum of 5 individuals; all groups were maintained under the same experimental conditions. With  $n$  being the number of individuals per group, and  $g$  the number of cages, the experimental groups (treatments) were characterized as follows:

- G1: (Non-infected control). Non-infected and non-treated animals ( $n=5$ )( $g=1$ )
- G2: Animals treated with biotherapeutic chicken serum with 13cH dilution ( $n=13$ )<sup>1</sup>
- G3: Animals treated with biotherapeutic chicken serum with 6cH dilution ( $n=13$ )( $g=3$ )
- G4: (Infection control). The animals were infected and received no treatment ( $n=13$ )( $g=3$ ).
- G5: Animals treated with biotherapeutic chicken serum with 3cH dilution ( $n=13$ )( $g=4$ ).

Their body weight was monitored over 12 weeks, measured on a semi-analytical scale (Balance BEL®)

### P-splines

According to Currie and Durbán (2002), a P-spline is a combination of B-splines and a difference penalty function of the  $\Delta$  order associated with the estimated coefficients of the B-splines, and the use of this tool helps reduce the flexibility of the B-splines, thus preventing an overfitting of the curve.

Authors Eilers and Marx (1996) present in their article an interesting result concerning modeling through the combination of B-splines; they state that the integral of the second derivative squared can be expressed as a quadratic function in the coefficients associated with the sum. Thus, the roughness of a curve is expressed by taking a base with several B-splines associated with a penalty, as follows:

$$R = \int_a^b [f''(x)]^2 dx, \quad (1)$$

$$S(f) = \sum_{i=1}^m \{f(x_i) - \sum_{j=1}^n \gamma_j B_j(x_i)\}^2 + \lambda \sum_{j=i+1}^n (\Delta^d \gamma_j)^2, \quad (2)$$

in which  $[a,b]$  is the study interval,  $d$  is the order of the  $\Delta$  difference operator for ,  $B$  is the regression basis function,  $\lambda$  is the control parameter of the smoothing, and  $()$  is a function that composes B-spline sums.

### Generalized additive mixed models using P-splines

Over the last decade, advancements have been observed in the application of P-splines to the context of generalized additive models. This is due to the possibility of rewriting the linear term by summing unknown

<sup>1</sup> Nomenclature used in homeopathy to refer to the dilution scale of medicines, with the following scales being used: decimal (DH), centesimal (CH), and fifty-millesimal (LM). On the centesimal scale, the dilution ratio is 1:100 (one part of active ingredient to 100 parts of inert ingredient, that is, one part of solute to 99 parts of solvent).

smooth functions, not parametrically specified. In this sphere, it is possible to assign peculiar random effects to subjects, generating a new class, called Generalized Additive Mixed Models.

Currie and Durbán (2002), Durbán, Harezlak, Wand, and Carroll (2005) and Durbán (2014) describe that the appropriate model for this situation is that which allows estimating the curves of specific differences between individuals using non-parametric smooth terms. In this model, a smoothing non-parametric  $g_i$  function is incorporated, thus enabling, according to Ruppert, Wand, and Carroll (2003), an interaction between treatment groups with the continuous predictor. This model can be described as:

$$y_{ij} = \sum_{k=2}^K \gamma_k tr_{ik} + f_k(x_{ij}) + g_i(x_{ij}) + \epsilon_{ij}, \quad (3)$$

in which  $y_{ij}$  indicates the response variable,  $() = \gamma + g$ , in which  $\gamma$ ,  $g$  denote the random effects, and  $f_k()$  is a smoothing function that explains the trajectory of each individual, and  $i$  indicates the  $i$ -th sampling unit observed at the  $j$ -th instant, and the  $k$  index indicates the  $k$ -th treatment group. Besides, the  $f_k$  function is estimated using P-splines, which can be rewritten in the context of mixed model; besides:

$$(a_{i1}, a_{i2})' \sim N(\mathbf{0}, \Sigma)$$

$$\epsilon_{ij} \sim N(0, \sigma_\epsilon)$$

$$tr_{ik} = \begin{cases} 1, & \text{if the } i\text{-th receives the } k\text{-th treatment} \\ 0, & \text{otherwise} \end{cases}$$

$$v_k \sim N(0, \sigma_v^2)$$

in which the specific curves of the  $(.)$  sampling units have parametric components  $\gamma + g$  and non-parametric components  $f_k$ , with both being random.

Thus, model 3 can be rewritten in its matrixial form as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}, \quad (4)$$

The  $\mathbf{X}$  matrix of fixed effects is expressed by:

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{T}_1 \\ \vdots & \vdots \\ \mathbf{X}_m & \mathbf{T}_m \end{pmatrix}; \mathbf{X}_i = \begin{pmatrix} 1 & x_{i1} \\ \vdots & \vdots \\ 1 & x_{in_i} \end{pmatrix}, \quad (5)$$

in which  $\mathbf{T}_i$  is a matrix that indicates if the  $i$ -th subject receives the  $k$ -th treatment.

The  $\boldsymbol{\beta}$  parameter vector is defined by:

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \gamma_1 \\ \vdots \\ \gamma_k \end{pmatrix} \quad (6)$$

The  $\mathbf{Z}$  matrix of random effects is defined by:

$$\mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 & \mathbf{X}_1 & \mathbf{0} & \dots & \mathbf{0} & \mathbf{Z}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{Z}_2 & \mathbf{0} & \mathbf{X}_2 & \dots & \mathbf{0} & \mathbf{0} & \mathbf{Z}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{Z}_m & \mathbf{0} & \mathbf{0} & \dots & \mathbf{X}_m & \mathbf{0} & \mathbf{0} & \dots & \mathbf{Z}_m \end{pmatrix} \quad (7)$$

The vector of random effects is expressed by:

$$\mathbf{u} = (\mathbf{u}_1, \dots, \mathbf{u}_k, \mathbf{a}_{11}, \mathbf{a}_{12}, \dots, \mathbf{a}_{m1}, \mathbf{a}_{m2}, \mathbf{v}_1, \dots, \mathbf{v}_k)' \quad (8)$$

And the covariance matrix is written as:

$$\mathbf{G} = \text{Cov}(\mathbf{u}) = \begin{pmatrix} \sigma_u^2 \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \text{blockdiagonal}_{1 \leq i \leq m} \Sigma & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_v^2 \mathbf{I} \end{pmatrix} \quad (9)$$

According to Diggle, Heagerty, Liang, and Zeger (2002), correlation structures are used for modeling covariance matrices, referring to the dependence between observations; when it comes to mixed-effect models, they are used for modeling the dependence between errors within the group. Because the variance is related to components  $\mathbf{Z}_j \mathbf{G} \mathbf{Z}_i^T$  and  $\mathbf{R}_i$ , one can describe the structures of  $\mathbf{G}$  or  $\mathbf{R}$ , or both. Littell,

Pendergast, and Natarajan (2000) and Wolfinger (1993) mention that the covariance between two observations for the same individual depends on the length of the time interval between measurements, and that the variance is considered constant over time. Covariance can be described in terms of the variance, and correlations can be expressed as a function of the time interval between measurements.

### Model selection

Choosing or selecting the model that best explains data is undoubtedly fundamental in any statistical analysis in the modeling context. There are numerous tests and criteria for model selection in the literature, with one of them being the Akaike information criterion (AIC). According to Akaike (1974), the idea behind this criterion is to compare models, expressed by the following equation:

$$AIC = -2l(\hat{\theta}) + 2p, \quad (10)$$

with  $l()$  being the logarithm of the likelihood function applied to the estimator of maximum likelihood of  $\theta$ ,  $()$  and  $p$  the number of explanatory variables. Thus, the satisfactory model will be the one with the lowest AIC statistical value.

### Computational aspects

*Software* R, version 3.6.2 R Core Team (2019), was used, for both data treatment and the fitting of the models employed. The `gamm` function in the `mgcv` package by Wood (2017) was used for the modeling involving generalized additive mixed models. The estimation method selected was REML, by means of the `method` function, and the `lme` function in the `nlme` package by Pinheiro, Bates, Debroy, Sarkar, and R Core Team (2019) for the fitting of the mixed models. For the B-splines bases to be selected, the `bs` function in the `splines` package was used. To build the charts, the following packages were used: `lattice` by Sarkar (2008), `lattice Extra` by Sarkar and Andrews (2019), and `ggplot2` by Wickham (2016).

## Results

### Description of basic statistics

To conduct the statistical analyses, only the 'Treatment groups' and 'Experiment follow-up time' variables were considered, that is, the 'Cage' variable was not treated due to the study objective. Table 1 contains summarized statistical data on the 'weight' response variable for each one of the treatment types:

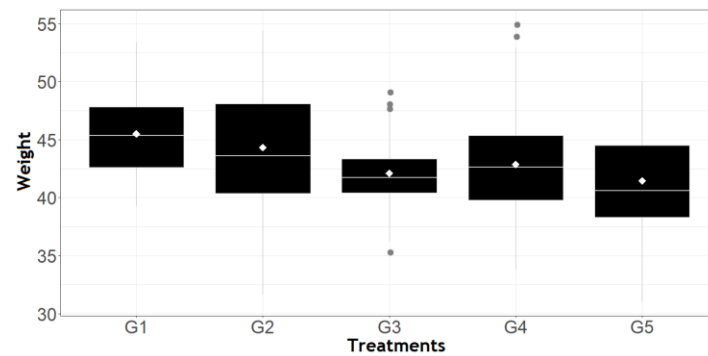
**Table 1.** Summarized measures for each treatment.

	Treatment groups	n	Mean	SD	Min.	Max.	CV
G1	Control	5	45.49	3.48	39.20	53.40	0.076
G2	Treated (13cH)	13	44.34	4.71	31.60	54.40	0.106
G3	Treated (6cH)	13	42.09	2.83	35.40	49.20	0.067
G4	Infection control	13	42.85	4.52	33.80	55.00	0.105
G5	Treated (3cH)	13	41.44	3.84	31.00	49.90	0.092

Observing Table 1 allows concluding that the mean of the weights in relation to each one of the treatments is 45.49 g for the 'Non-infected control' treatment group, and 41.44 g for the individuals allocated in the homeopathic treatment with 3cH dilution.

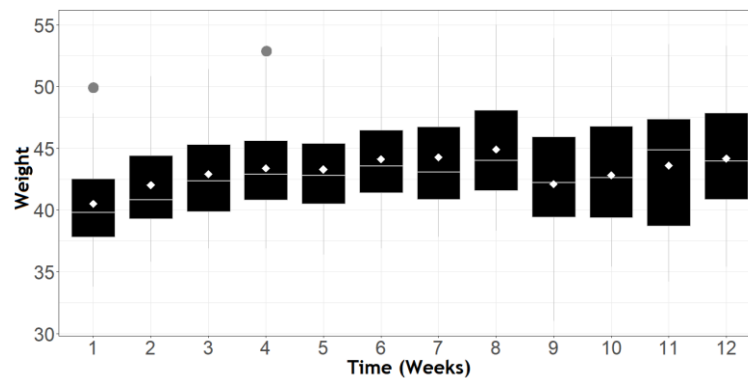
It is also possible to see a difference as to the maximum and minimum values for the weights in each one of the treatment groups; the minimum weights of the individuals allocated in the homeopathic treatment group with 3cH dilution (G5) is 31.00 g, whereas the minimum weight of the individuals in the Control treatment group (G1) is 39.20 g; as for the maximum weight, it ranges from 55.00 g, representing the 'Infection control' group of individuals (G4), to 49.20 g, representing the group under homeopathic treatment with 6cH dilution (G3). Finally, about the coefficient of variation, there is a small oscillation, with values ranging from 0.076 g and 0.106 g, approximately.

Figure 1 displays the behavior of the response variable for each one of the treatments; the group under homeopathic treatment with 3cH dilution (G5) had an average weight level lower than that of the other treatment groups, while the non-infected control group had the highest average weight level compared to the other treatment groups.



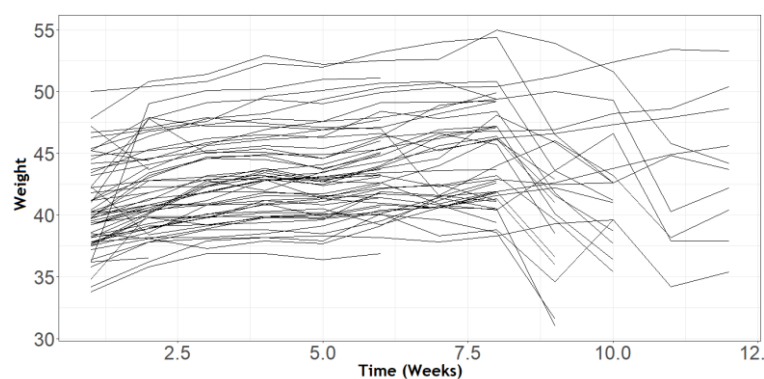
**Figure 1.** Boxplot for the response variable by treatment.

Figure 2 presents the behavior of the response variable as a function of time, with the weight showing an increase behavior over time and oscillating as of the sixth week. Another remarkable fact was to find potentially outlier measures in the first week and in the fourth week of monitoring.



**Figure 2.** Boxplot for the response variable over time.

About the profile chart, Figure 3 shows an increase behavior as of the second week as for weight over time. An oscillating behavior is seen for the weights soon in the first week, a factor that can be explained by this being the beginning of the treatment; besides, the response variable seems to behave linearly from the second to the sixth week for each subject in relation to the treatment groups as time varies.



**Figure 3.** Average profile for each subject observed over time.

Figure 4 shows dispersion charts by treatment; there may be some variability between these observations, justifying the possibility of using some type of smoothing that takes into account the variability found.

It is also possible to notice that the weight gain of the mice in each treatment group varies greatly between the intercept and the slope of the lines that have been fitted; this allows considering mixed models with random effects, be they independent or correlated between intercept and slope in relation to time – in this case, for the Time covariable.

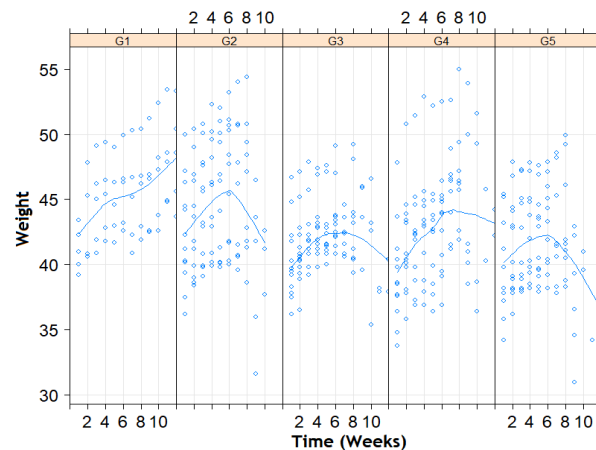


Figure 4. Dispersion chart by treatment.

### Adjustment

Considering the smoothing function over time, the model considered is described by:

$$Weight_{ij} = \sum_{k=1}^K \gamma_k G_{ik} + f_k(Time_{ij}) + a_{i1} + a_{i2}Time_{ij} + \epsilon_{ij}, \quad (11)$$

$$G_{ik} = \begin{cases} 1, & \text{if the } i\text{-th subject receives the } k\text{-th treatment} \\ 0, & \text{otherwise} \end{cases}$$

in which  $Weight_{ij}$  means the weight of the  $i$ -th subject in the  $j$ -th week, for  $i = 1, \dots, 57$ ,  $j = 1, \dots, 12$ . The homeopathic treatment groups range from  $k = 1, \dots, 5$ . Moreover,  $f_1$  refers to the mean curve for the individuals allocated in the Non-infected control group,  $f_2$  refers to mean curve for the subjects that received the homeopathic treatment with 13cH dilution,  $f_3$  refers to the mean curve for the individuals that received the treatment with 6cH dilution,  $f_4$  refers to the mean curve for the individuals allocated in the Infection control group, and  $f_5$  refers to the mean curve for the individuals that received the homeopathic treatment with 3cH dilution, while  $a_{i1}$  and  $a_{i2}$  refer to the random effects corresponding to the intercepts and slopes of the curves.

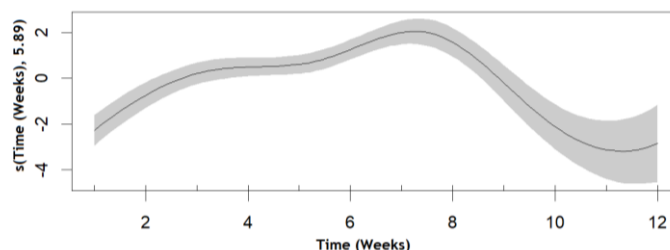
Also,  $f(Time)$  is the unknown smoothing function that describes the trajectory of this variable in relation to Weight evolution, in which  $f(\cdot)$  is estimated by means of P-splines, with third-polynomial-degree B-splines basis function, and a second-order discrete penalty. The variability between subjects is determined through the  $a_i$  random effect and a random effect in the slope for the Time variable, in which  $(a_{i1}, a_{i2})' \sim N(\mathbf{0}, \Sigma)$  and  $\epsilon_{ij} \sim N(0, \sigma_\epsilon)$ , assuming the dependence of the random terms.

Table 2 displayed the fitted models, considering different structures of variances and covariances, defined by: a model without considering a correlation structure in the errors, and three other models using the AR(1), CS and ARMA(1,1) correlation structures. For model selection, the Akaike criterion was used.

Table 2. Comparing models with different correlation structures.

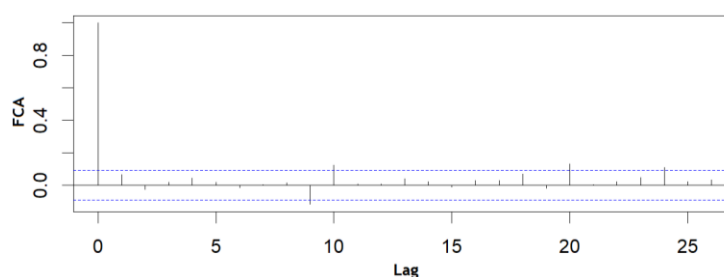
Model	AIC
Non-structured	2666.09
AR(1)	1999.07
CS structure	2045.82
ARMA (1,1) structure	2002.60

Observing Table 2 allows concluding that the model to be selected is the one with the AR(1) correlation structure, since it provides the lowest value for the AIC statistics; in addition, this structure allows improving the fit of the curves both for the independent variables and the dependent variable. Figure 5 contains the chart for the smoothed curve of the Time covariable after the determination of the correlation structure; it is possible to see that the Weight evolution presented an increase behavior until the 7<sup>th</sup> week approximately and, after this period, it decreased until the 12<sup>th</sup> week; as a consequence, the effect of this covariable follows a non-linear pattern.



**Figure 5.** Smooth curve with 5.89 effective degrees of freedom (e.d.f), considering the AR(1) correlation structure.

Figure (6) displays the correlogram with the AR(1) structure in the residuals; it is possible to see a significant portion of the autocorrelation values contained within the significance levels, that is, there is a decrease in the issue found as to the assumption of independence in the residuals of the model without considering a correlation structure.



**Figure 6.** Correlogram considering the AR(1) correlation structure.

Table 3 contains the estimates for the  $\lambda$  parameter, with their respective effective degrees of freedom (e.d.f), F test, and p-value. It can be concluded that the smoothing of the Time covariable for each homeopathic treatment group was highly significant in relation to Weight evolution.

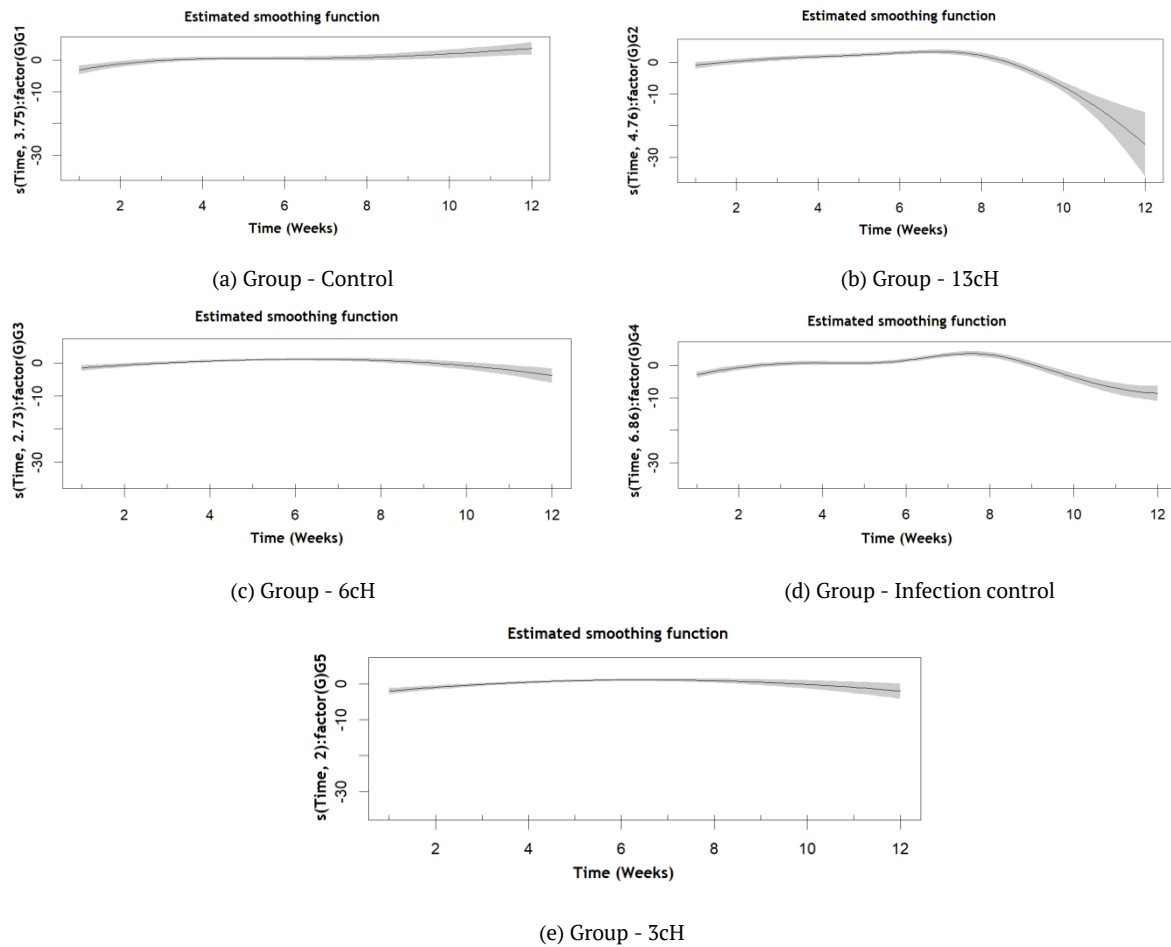
**Table 3.** Summarized measures for each treatment.

Smoother	e.d.f	Parameter $\lambda$	F	p-value
s(time): Group 1 - Control	3.751	$2.334 \times 10^{-4}$	7.764	<0.001
s(time): Group 2 - 13cH	4.763	$2.214 \times 10^{-3}$	43.248	<0.001
s(time): Group 3 - 6cH	2.726	$3.011 \times 10^{-5}$	15.681	<0.001
s(time): Group 4 - Infection control	6.864	$4.662 \times 10^{-2}$	41.414	<0.001
s(time): Group 5 - 3cH	2.000	$7.176 \times 10^{-10}$	19.867	<0.001

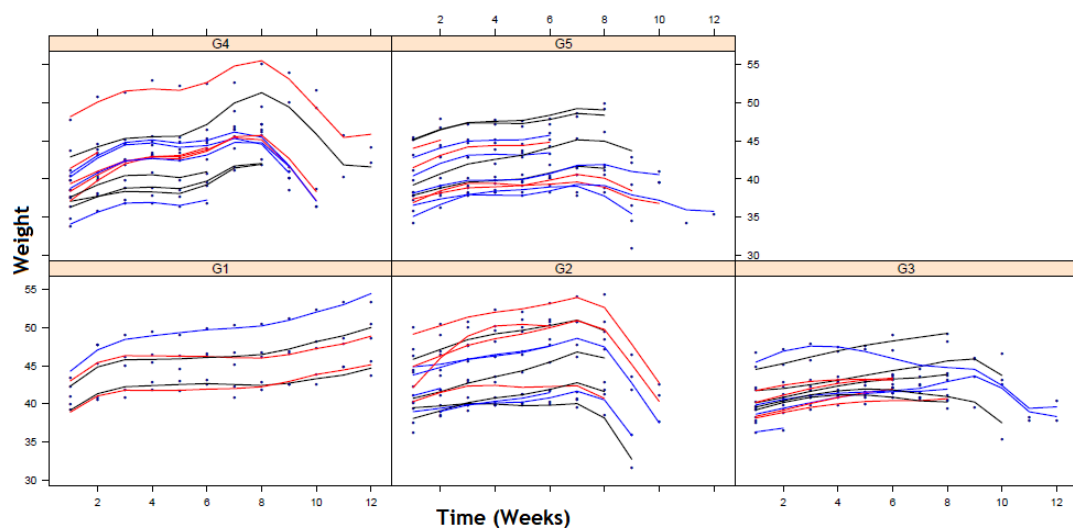
Figure 7 shows the behavior of the estimated smoothing non-parametric functions of model 11 for each treatment group; the continuous line is the estimated smoothing function, and the gray line represents the ranges with 95% confidence. Because the effect of the Time variable for each treatment group is not linear, the estimated smoothing functions are significant with 3.751, 4.763, 2.726, 6.864 and 2.000 effective degrees of freedom. Additionally, the evolution of the weights over time has a period of increase until the 7th week approximately in the groups under homeopathic treatment with 13cH dilution (Figure 7b), 6cH dilution (Figure 7c), and 3cH dilution (Figure 7e) with the passing of time; as for the Non-infected control group (Figure 7a) and the Infection control group (Figure 7d), there is a slight rise in the beginning of the treatment, with an oscillation as of the 2nd week. Figures 7b and 7d show a significant decrease, thus evidencing that the animals lost weight after said period.

Figure 8 contains the specific estimated curves for the subjects allocated in their treatment groups; model 11 describe said curves, taking into account the non-linearity and the associated random effects. It is possible to observe the fitting capacity of the model using 10 nodes, which presents itself satisfactorily as to interpolation.

Thus, with the model defined in 11, it is possible to describe the mean trajectories by treatment group and verify the objective proposed in the beginning of this application, in which Figure 9 presents the estimated curves by treatment group:



**Figure 7.** Smoother for Weekly Time in each Treatment Group.



**Figure 8.** Individual curves for the subjects, fitted by treatment (G1: Group – Control; G2: Group – 13cH; G3: Group – 6cH; G4: Group – Infection Control, and G5: Group – 3cH).

Observing Figure 9, it is possible to verify that the animals allocated in the control group had, during a significant portion of the study, an average weight superior to that of the other treatment groups. When it comes to the group under homeopathic treatment with 13cH dilution, none of the animals reached the end of the study and, in said group, the greatest dilution of the biotherapeutic serum was used for treating the Chagas disease. Finally, one can further observe that, on average, the subjects allocated in the groups under homeopathic treatment with 13cH, 6cH and 3cH dilutions and the Infection control group had a decrease in weight as of the 7<sup>th</sup> week.



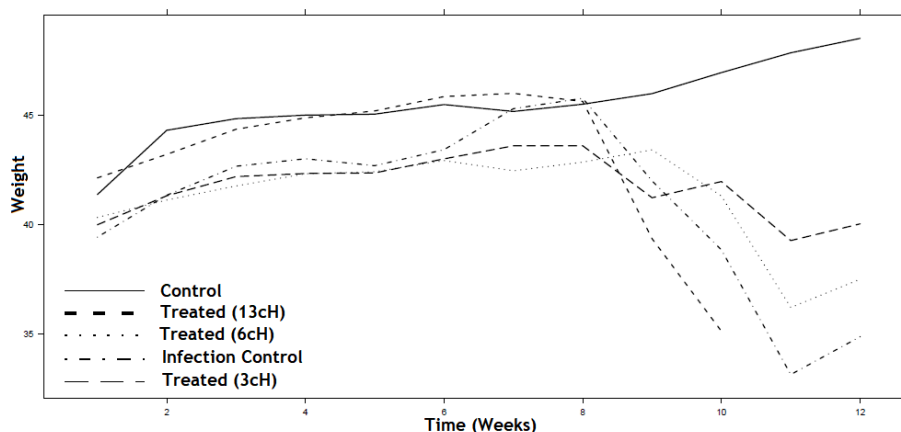


Figure 9. Estimated curves by treatment.

## Conclusion

Using the P-splines smoothing technique in the context of mixed models for analyzing longitudinal data provides researchers with greater flexibility in terms of modeling, since this methodology allows describing the individual trajectories of a certain situation as a function of time.

Furthermore, the introduction of random effects is taken into account for each one of the estimated curves by treatment, through the smoothing non-parametric function, thus supplying information on the average behavior of a certain object classified into observation groups.

As for the result, it is possible to conclude that the weight, until the 6<sup>th</sup> week, of the animals treated with a dilution level of 13cH of biotherapeutic serum is superior to that of the other treatment groups. Concerning the animals treated with a dilution level of 3cH of biotherapeutic serum, they reached the end of the study, with average weights greater than those of the animals treated with a dilution level of 6cH of biotherapeutic serum.

## References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723. DOI: 10.1109/TAC.1974.1100705
- Boor, C. D. (1978). *A practical guide to splines* (Applied Mathematical Sciences, Vol. 27). New York: Springer-Verlag.
- Caldeira, J. F. (2011). Estimacao da Estrutura a Termo da Curva de Juros no Brasil através de Modelos Paramétricos e Nao Paramétricos. *Análise Econômica*, 29(55), 95–122. DOI: 10.22456/2176-5456.13198
- Currie, I. D., & Durbán, M. (2002). Flexible smoothing with p-splines: a unified approach. *Statistical Modelling*, 2(4), 333–349. DOI: 10.1191/1471082x02st039ob
- Diggle, P., Heagerty, P. J., Liang, K.-Y., & Zeger, S. (2002). *Analysis of longitudinal data*. Oxford, UK: Oxford University Press.
- Domingues, J. S., Gonçalves, M. C., Braz, S. C., & Pereira, F. J. (2014). Usando splines cúbicas na modelagem matemática da evolução populacional de Pirapora/MG. *ForScience*, 2(1), 17–30. DOI: 10.29069/forscience.2014v2n1.e23
- Durbán, M., Harezlak, J., Wand, M., & Carroll, R. (2005). Simple fitting of subject-specific curves for longitudinal data. *Statistics in medicine*, 24(8), 1153–1167. DOI: 10.1002/sim.1991
- Durbán, M. (2014). *Splines con penalizaciones: Teoría y aplicaciones*. Madrid, ES: Universidad Carlos III de Madrid.
- Eilers, P. H., & Marx, B. D. (1996). Flexible smoothing with b-splines and penalties. *Statistical Science*, 11(2), 89–102. DOI: 10.1214/ss/1038425655
- Ferreira, É. C., Ciupa, L., Portocarrero, A. R., Brustolin, C. F., Massini, P. F., Aleixo, D. L., & Araújo, S. M. (2018). Phosphorus protects cardiac tissue by modifying the immune response in rats infected by *Trypanosoma cruzi*. *Cytokine*, 102, 102–106. DOI: 10.1016/j.cyto.2017.07.017

- Keele, L. J. (2008). *Semiparametric regression statistic for the science*. Chichester, EN: John Wiley and Sons LTD.
- Littell, R., Pendergast, J., & Natarajan, R. (2000). Modeling covariance structure in the analysis of repeated measure data. *Statistics in Medicine*, 19(13), 1793-819. DOI: 10.1002/1097-0258(20000715)19:13<1793::AID-SIM482>3.0.CO;2-Q
- Paro de Paz, C. C., Packer, I. U., Freitas, A. R., Tambasco-Talhari, D., Regitano, L. C. A., Alencar, M. M., & Cruz, G. M. (2004). Ajuste de modelos não-lineares em estudos de associação entre polimorfismos genéticos e crescimento em bovinos de corte. *Revista Brasileira de Zootecnia*, 33(6), 1416-1425. DOI: 10.1590/S1516-35982004000600008
- Paulson, D. S. (2006). *Handbook of regression and modeling: Applications for the clinical and pharmaceutical industries*. [S.l.]: Chapman and Hall/CRC.
- Pinheiro, J., Bates, D., Debroy, S., Sarkar, D., & R Core Team. (2019). *nlme: Linear and Nonlinear Mixed Effects Models* (R package version 3.1-139). Retrieved from <https://CRAN.R-project.org/package=nlme>
- Pinheiro, V. R., Silva, F. F., Guimarães, S. E. F., Resende, M. D. V., Lopes, P. S., Cruz, C. D., & Azevedo, C. F. (2013). Mapeamento de QTL para características de crescimento de suínos por meio de modelos de regressão aleatória. *Pesquisa Agropecuária Brasileira*, 48(2), 190-196. DOI: 10.1590/S0100-204X2013000200009
- R Core Team. (2019). R: a language and environment for statistical computing. Vienna, AT: R Foundation for Statistical Computing.
- Ruppert, D., Wand, M., & Carroll, R. (2003). *Semiparametric regression* (Cambridge Series in Statistical and Probabilistic Mathematics). Cambridge, EN: Cambridge University Press. Doi: 10.1017/CBO9780511755453
- Sarkar, D. (2008). *Lattice: Multivariate Data Visualization with R*. New York: Springer. Retrieved from <http://lmdvr.r-forge.r-project.org>
- Sarkar, D., & Andrews, F. (2019). *Extra Graphical Utilities Based on Lattice* (R package version 0.6-28). Retrieved from <https://CRAN.R-project.org/package=latticeExtra>
- Santos, D. C. C., Scalez, D. C. B., Passafaro, T. L., Turra, E. M., Fernandes, A. F. A., Alvarenga, E. R., ... Toral, F. L. B. (2015). Modelagem das trajetórias médias do peso vivo e do rendimento de carcaça de tilápias do Nilo (*Oreochromis niloticus*) por meio de funções B-spline. *Arquivo Brasileiro de Medicina Veterinária e Zootecnia*, 67(3), 899-908. DOI: 10.1590/1678-4162-6885
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag. DOI: 10.1007/978-3-319-24277-4
- Wolfinger, R. (1993). Covariance structure selection in general mixed models. *Communications in statistics-Simulation and computation*, 22(4), 1079-1106. DOI: 10.1080/03610919308813143
- Wood, S. N. (2017). *Generalized additive models: an introduction with R* (2nd ed.). Boca Raton, FL: CRC Press/Taylor & Francis Group.