Abstract

The aim of this study is to determine the effectiveness of a natural tropical tree resin in controlling termites thus providing protection from their destruction. Tree resin from the bark of tropical trees offers the potential for this protection. Termites were fed by filter paper soaked in tropical tree resin dissolved in a solvent at different concentrations for 15 days. The number of termites still alive on each day was observed and recorded. In this paper, we use four types of statistical models: Partially linear model, piecewise linear model, cubic smooth spline and mix effect model to analyze the termite data. The results show that tropical tree resin, particularly at a higher concentration of 10mg is significantly more effective in killing termites. We show that two dishes under 10mg of tree resin were mistaken since the data for these two dishes are shown insignificantly different from data under 5mg. The partially linear model shows that there is non-linear (piecewise linear) time effect. Both piecewise linear model and cubic spline smoothing show that the most effective period is the first week. The nonparametric smoothing, cubic spline and piecewise linear model are not significantly different. Mixed effect model is consistent with partial linear model and piecewise linear model. The estimated treatment effect is time varying with a change point at day 7. Therefore, we suggest the piecewise linear model as the final simplest one for prediction. This model fits the data with adjusted $R^2=93.7\%$ and shows that on average, 10mg is 68.9% more efficient than 5mg in killing termites during the first week.

1. Introduction and Termite Data Collection

The aim of this study is to determine the effectiveness of a natural tropical tree resin in controlling termites thus providing protection from their destruction. Termite destruction in Florida is a serious problem. Each year wood termites bore into thousands of homes and businesses causing millions of dollars of damage. Current chemical pesticides that are used in the control of termites and for the protection from their damage are potentially harmful.
to Florida’s delicate environment. Natural tropical tree resins show the potential to be effective in the control of termites. If this potential could be harnessed, it could help provide a more environmentally sensitive way to resolving Florida’s termite problem. Such a solution could potentially be exportable, to other parts of the United States and the world that also have the termite problems.

The experiment is designed to measure the effects of the resin on the survival of termites. The resin was derived from the bark of tropical trees and was dissolved in a solvent and was placed on filter paper in two different levels of concentration, either 5mg or 10mg dosage. There are eight dishes for each dosage. Only 25 living termites were placed in each dish. Each dish was observed on 13 specific days. The number of living termites was counted for each dish. No observation was made on day 3 and day 9. Since the measurements of days 3 and 9 were never intended to be taken, the termite data can be classified as unbalanced data (Diggle, Liang and Zeger, 1994). The data originally are from the software package, Data Desk, Ithaca, NY: Data Description, Inc. (1993). For convenience, we show that data set in Table 1 below that is obtained from [http://lib.stat.cmu.edu/DASL/Datafiles/Termites.html](http://lib.stat.cmu.edu/DASL/Datafiles/Termites.html).

The purpose of this paper is to find the best model for the termite data and to quantify the effectiveness of two different concentrations of tropical tree resin dissolved in a solvent in providing protection from termites. It contains three steps: First to visualize the different behavior of termites according to two different dosages, second to explore the degree of association in the data set, and finally to develop a statistical model for the survival pattern of termites in response to varying dosage of resin. To achieve first two goals, we conduct exploratory data analysis (EDA) and correlation analysis in the section 2. Section 3 gives a guided direction for the analysis based on the information that we derived from the exploratory data analysis. We develop a final statistical model that fits the data well. Final conclusion is given in section 4. All analyses are conducted by R language. The codes for the analysis are available from the first author.

<table>
<thead>
<tr>
<th>dish</th>
<th>dose</th>
<th>day1</th>
<th>day2</th>
<th>day3</th>
<th>day4</th>
<th>day5</th>
<th>day6</th>
<th>day7</th>
<th>day8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>25</td>
<td>24</td>
<td>NA</td>
<td>22</td>
<td>18</td>
<td>17</td>
<td>15</td>
<td>14</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>25</td>
<td>25</td>
<td>NA</td>
<td>25</td>
<td>21</td>
<td>20</td>
<td>17</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>25</td>
<td>25</td>
<td>NA</td>
<td>24</td>
<td>23</td>
<td>21</td>
<td>20</td>
<td>19</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>25</td>
<td>25</td>
<td>NA</td>
<td>25</td>
<td>24</td>
<td>22</td>
<td>22</td>
<td>21</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>25</td>
<td>25</td>
<td>NA</td>
<td>22</td>
<td>19</td>
<td>18</td>
<td>17</td>
<td>14</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>25</td>
<td>25</td>
<td>NA</td>
<td>23</td>
<td>20</td>
<td>17</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>25</td>
<td>25</td>
<td>NA</td>
<td>24</td>
<td>23</td>
<td>22</td>
<td>20</td>
<td>19</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>25</td>
<td>25</td>
<td>NA</td>
<td>23</td>
<td>19</td>
<td>17</td>
<td>16</td>
<td>14</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>25</td>
<td>24</td>
<td>NA</td>
<td>23</td>
<td>22</td>
<td>21</td>
<td>19</td>
<td>18</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>25</td>
<td>25</td>
<td>NA</td>
<td>23</td>
<td>20</td>
<td>19</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>25</td>
<td>24</td>
<td>NA</td>
<td>12</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>25</td>
<td>24</td>
<td>NA</td>
<td>14</td>
<td>10</td>
<td>7</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>
2. Exploratory Data Analysis

In this section, we carry out various exploratory data analyses. First we use exploratory data analysis to visualize the data. Figure 1 shows the spaghetti plot of the number of live termites for each dish under the two tree resin concentrations along time in days. It suggests that there are time effect and treatment effect on the number of live termites.

Next we construct a plot of the dose means with standard deviation. The resulting diagram is shown in Figure 2. On average, the numbers of live termites for dishes under 10 mg dosage are lower than those under 5 mg dosage on all post visit days. Both Figures 1 and 2 show that termites die faster under 10mg dosage than under 5mg dosage. However, we can detect some strange behaviors in Figure 1. Termites in some dishes under 10mg dosage behave similarly to those under 5mg dosage. We will refer these dishes as problem dishes or dishes 1 and 2 under 10mg dosage in the paper.
Figure 1: Spaghetti plot of the number of live termites against days.

Figure 2: Plot of mean profiles and standard deviations for each dosage

Due to the two problem dishes, the standard deviation is very large for the number of live termites under 10mg dosage even after classified by time effect and dose effect.
Figure 3 graphically displays the data using the box-plot. The (blue) dots represent the average number of live termites each day. From Figure 3, the median number of live termites under 10mg dosage declines more rapidly than the one under 5mg dosage. However, the average numbers of live termites are much higher than the median numbers of live termites under 10mg dosage from Figure 3(b). Moreover, the variation under 10mg dosage is larger than the one under 5mg dosage. It is suspected that the bigger difference between mean and median, and larger variation under 10mg dosage than the one under 5mg dosage are caused by the unusual behaviors of the two problem dishes under 10mg dosage.

Now, we turn to conduct correlation analysis. To explore the degree of association in a longitudinal data graphically, we first use linear model to remove the effect of time. Figure 4 shows that the treatment effect is obvious after adjusting for time effect.

To observe correlation structure, partially autocorrelation function (PACF) and autocorrelation function (ACF) can be used. Due to similarity of the PACF and ACF graphs for all 16 dishes, only three randomly selected dishes are shown in Figure 5. Using $ARMA(p,q)$ model, the significantly large value of PACF at lag 1 from Figure 5(a) indicates that $p=1$ while the significantly large value of ACF up to lag 2 from Figure 5(b) shows that $q=2$. Therefore, we characterize $r_{ij}$ as an $ARMA(1,2)$ process.
Figure 4: Residual plot of the number of live termites after removing time effect

Figure 5(a). PACF

Dish1 (5mg)  Dish2 (5mg)  Dish5 (5mg)
In section 2, it is discovered that dish 1 and dish 2 do not behave like the other dishes under 10mg dosage. Therefore, we suspect that the doses of these dishes are not truly under 10mg. If our suspect about the dose of those dishes is true, then we can take two different actions. One is to remove the data for dish 1 and dish 2 from 10mg dosage. The other choice is to consider these dish 1 and dish 2 treated by different dosage from 10mg. So we add an additional unknown dosage level for dish 1 and dish 2. Further, we test for significant difference between 5mg and unknown dosage.

Base on these two cases, we construct a partially linear model, a nonparametric model, a piecewise linear model and a cubic regression spline. For a partially linear model, we consider the original data set with same time effect. For a nonparametric model, a piecewise linear model, and a cubic regression spline, it is assumed that the data from the different doses over the study period have the different time effects. To implement the model fitting, we apply nonparametric functions, piecewise regression or cubic regression splines separately for each dose group.

Piecewise linear model and cubic spline fitting are parametric models that are good on making predictions. For the piecewise linear model, we use both continuous and discontinuous segment models. For discontinuous case, Koul & Qian (2001) and Koul, Qian and Surgailis (2003) showed that the asymptotic theory for maximum likelihood and M-estimators.

However, if our suspect about dose of dish 1 and dish 2 recorded under 10mg dosage is not true, we can think that all dishes are not identical, but the effects vary with dishes. That is, dishes are regarded as a random sample of a larger population of subjects and hence any
effects that are not constant for all subjects are regarded as random. For this case, mixed effect model is used to analyze the dish effects since we consider dishes are randomly selected from available dishes. Figure 6 describes the direction of the inferential statistical analysis.

### 3.1 Partially Linear Model

In this section, under the assumption that the data have the same time effect for each dose, we will use a partially linear model to fit the data. A partially linear model is a semi-parametric model [Haerdle and Liang]. It has many benefits.

- It is more efficient than the standard linear regression model, when the response variable depends on some variables in linear relationship, but is nonlinearly related to other covariates.
- It can provide a parsimonious description of relationship between the response variable and explanatory variable.
- It has the flexibility of the nonparametric model.

![Figure 6. The direction for the inferential statistics analysis: subject=dish](image)
In addition, a partially linear regression model is good at fitting the unbalanced data since it only requires that the explanatory variables to be either independent and identically distributed random designed points or fixed designed points. In our case, the explanatory variables are fixed designed points. A partially linear regression model used in this section is defined by

\[ y_{ij} = x_i^T \beta + g(t_{ij}) + \varepsilon_{ij}, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n_i, \]  

where \( n_i \) is the number of observations for \( i \)th subject, and \( m \) is the number of subjects. Since all dishes are observed 13 times, we can write \( n_i \) as \( n \) where \( n=13 \). The vector of covariate \( x_i = (x_{i0}, \ldots, x_{ik})^T \) is the number of different doses, and \( x_i \) is a vector of explanatory dummy variable with \( k=1 \); \( \beta = (\beta_0, \ldots, \beta_k)^T \) is a vector of coefficients and represent dose effects; the measurements are taken at \( t_{ij} \) for \( i \)th subject at \( j \)th time and \( t_{ij} \) is the fixed designed points in termite data set since all dishes were observed on the same specific days. Thus, \( t_{ij} \) can be written as \( t_j \). And \( g \) is an unknown nonparametric real function; \( \varepsilon_{ij} \)'s are additive errors with a mean zero. Thus for termite data, the equation (3.1) can be rewritten as the following:

\[ y_{ij} = x_i^T \hat{\beta} + g(t_j) + \varepsilon_{ij}, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n. \]  

The model (3.2) is an additive model. For the time effect term \( g(t) \), we use the smooth spline \texttt{smooth.spline} in R language. In order to estimate the parametric and nonparametric parts of model, the back-fitting algorithm is used [Hastie, p. 118; also Diggle, Liang and Zeger, p. 111].

Note that due to correlated observations, \( \text{var}(\varepsilon) = \sigma^2 V \) where some of off-diagonal elements of \( V \) are nonzero. Therefore, to estimate \( \beta \), we use the generalized least square instead of the ordinary least squares since the data are serially correlation. If the ordinary least square is used, then the estimated \( \beta \) has no longer the minimum variance, which may cause the erroneous conclusion of conventional tests (\( t \) test, \( F \) test), even though \( \beta \) estimate is still unbiased.

The non-parametric estimate of \( g(.) \) using smooth spline is combined iteratively with the generalized least squares calculation for \( \hat{\beta} \). There are several ways to decide the bandwidth (or \( \lambda \)) in estimating the nonparametric function. The visualization is one of ways, but it can be very subjective. So we use the generalized cross-validation, which is a default option in R function \texttt{smooth.spline}. (For the more detail, refer to Diggle, p.47).
The generalized least squares estimator of $\beta$ is obtained by using \texttt{gls} in \texttt{library(nlme)} in R language. For the correlation structure option in \texttt{gls} function, we use \texttt{ARMA(1,2)} compound symmetry based on the correlation analysis.

Figure 7 displays the partially linear model with the original data set. As one can observe, the curve that represents 10mg dose is shifted up because of dish 1 and dish 2 under 10mg dosage. There is a difference between the curves for 5mg and for 10mg. However, the curve for 10mg dosage doesn't fit the points well. This is the result that we expected since in EDA (Exploratory Data Analysis), we already discovered the strange behavior of dish 1 and dish 2 under 10mg dosage.

![Figure 7. The partially linear model for the original data set](image)

Table 2: The estimated dose effects after accounting for time effect for the original data

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Std. Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(intercept)</td>
<td>17.1214</td>
<td>1.2646</td>
<td>13.54</td>
<td>0.0000</td>
</tr>
<tr>
<td>Dose 10mg</td>
<td>-4.7001</td>
<td>1.7884</td>
<td>-2.63</td>
<td>0.0092</td>
</tr>
</tbody>
</table>
Table 2 shows the result from generalized linear model for the original data set. The p-value of dose coefficient is less than 0.01, which indicates the significance of the dose effect.

Under the assumption that dishes 1 and 2 under 10mg dosage are outliers, we also model the data without the outliers (dishes 1 and 2 under 10mg). Again, the model (3.2) is used. The data set contains all the original data except the dish 1 and 2 from 10mg. The results are shown in Table 3.

Table 3. The estimated dose effects after accounting for time effect and removing outliers

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Std. Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(intercept)</td>
<td>17.9638</td>
<td>0.6471</td>
<td>27.76</td>
<td>0.0000</td>
</tr>
<tr>
<td>Dose 10mg</td>
<td>-8.7551</td>
<td>0.9885</td>
<td>-8.88</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Comparing Tables 2 and 3, one can observe that the dose effects are changed from 4.70 to 8.76 comparing 10mg to 5mg. Both data sets show significant dose effects with p-values 0.0092 and 0.0000, respectively.

Figure 8 displays the partial linear model with the modified data set. Since the outliers are removed, the curve for 10mg dose fit better.

Figure 8. The partially linear model after removing outliers
The plot of residuals in time sequence is shown in Figure 9(a). It is ideal that a horizontal band in this plot will enclose all of the residuals, and the residuals will fluctuate in a more or less random fashion within band. It shows that the model has room for improvement.

The plot of residuals against fitted value \( y \) shown in Figure 9(b) is narrow at the both ends. This double-bow pattern can occur when \( y \) is a proportion between zero and one. The variance of a binomial proportion near 0.5 is greater than one near zero or one. The suitable transformation to either regressor or the response variable or the method of weighted least squares might solve the inequality of variance.

(a) Time vs. Residual    (b) Fitted Y vs. Residual

Figure 9. Residual Plots

The normal probability plot shown in Figure 10 flattens at extreme. This indicates that a sample comes from a distribution with heavier tails than the normal. However, since we did not assume that error term is normal in (3.2), this is not a big concern.

It is known that the standard method of cross-validation (CV) tend to undersmooth serially correlated data because it tracks too closely the individual subject's trajectories. We used the generalized cross-validation (GCV) function, which is very similar to the cross-validation function. Therefore, it is possible that our GCV might undersmooth the data since our data is serially correlated. To solve this problem, we might be able to use the modified CV instead of GCV. In CV, we can compare observed \( y_{ij} \) to the predicted curve obtained by leaving out all observations of \( i \)th subject, instead of leaving out only \( j \)th observation of \( i \)th subject [Zeger and Diggle 1994].
Notice also that the partial linear model does not fit the data very well. For the rest of the paper, we assume that the time effect varies on the dose level. Also, notice that there are two dishes recorded as 10mg dosage, but the data shows very different from 10mg dosage. Thus, in order to identify whether it is a recording mistake or true difference for 10mg, we denote the dose level for these two dishes as a third difference level of dose. Denote $k$ be the number of dose levels detailed as follows:

\[
k = \begin{cases}
1, & \text{for 5mg}, \\
2, & \text{for 10mg}, \\
3, & \text{for unknown}.
\end{cases}
\]

### 3.2 Nonparametric Modeling

In this section we model the data set that has three different dose levels using a nonparametric model:

\[
y_{ij} = g_k(t_{ij}) + \epsilon_{ijk}, \quad i = 1, \ldots, m; \quad j = 1, \ldots, n_i \quad \text{and} \quad k=1,2,3, \tag{3.3}
\]

where $n_i$ is the number of observations for each subject, $m$ is the number of subjects and $t_{ij}$ is the time measurements taken for $i$th subject at $j$th time, and $g_k$ is a nonparametric function.

For a nonparametric function, we use the cubic smooth spline. To select $\lambda$ that minimizes the combined variance and bias, we use the generalized cross-validation, which is a default option in R function `smooth.spline`.  

![Figure 10. The normal probability plot of the model after removing outliers](image)
Figure 11 shows that the slope of the curve under 10mg dosage is much steeper than the slopes of the curves under 5mg and unknown dosages. If one observes at only the behavior of slope before certain threshold (say day 7), then those from 5mg and unknown dosages behave similarly. In the other hand, after seventh day, termites start to die very slowly with a similar fashion for both 10mg and unknown dosages. Therefore, if one focuses on how steep the slope of the curve is, those from 5mg and unknown dosages have the same slope. Furthermore, the smooth fit is the same as a linear fit for 5mg.

These nonparametric graphs suggest that there is a change point in the underlying model and motivate to develop a parametric model that is useful for prediction.

Figure 11. Nonparametric function for the data set with three different dose levels

### 3.3 Piecewise Linear Regression Model

From Figure 11, one observes that the linear trend for the first week, then trend levels off. Thus, we consider piecewise linear model. Using M-estimation (Koul, Qian and Surgailis, 2003), we obtain the estimated change point is day 7 as one can see in Figure 11. However, the data for 5mg dosage does not have a change point. Therefore, the piecewise regression for 10mg and unknown dosages, can be expressed as the following:

\[
E(y|x) = \begin{cases} 
\beta_0 + \beta_1x, & \text{if } x < \tau, \\
\alpha_0 + \alpha_1x, & \text{otherwise.} 
\end{cases}
\] (3.4)
Model (3.4) has colorful names. To name a few for instance, it is called two-phase linear regression (Koul and Qian, 2002 and 2003) and segmented regression model if $\alpha_0 + \alpha_1 \tau \neq \beta_0 + \beta_1 \tau$, while piecewise regression model if $\alpha_0 + \alpha_1 \tau = \beta_0 + \beta_1 \tau$. Figure 12 displays the two-phase regression model fittings. As one can see, the segmented regression is discontinuous at change point (day 7) for 10mg dosage but is almost continuous for unknown dosage as shown in Figure 12(b). Both discontinuous and continuous nonlinear regression models show similar patterns for 10mg and unknown dosages with change point at day 7.

![Figure 12. Piecewise regression and Segmented regression](image)

The piecewise linear model is reported:

$$m(x) = \begin{cases} 
1.600 - 3.842(x - 7)I(x \leq 7) - 0.101(x - 7)I(x > 7), & \text{for dose 10mg;} \\
18.421 - 1.197(x - 7), & \text{for dose 5mg;} \\
18.927 - 1.090(x - 7)I(x \leq 7) - 0.488(x - 7)I(x > 7), & \text{for dose unknown.}
\end{cases}$$

It explains 93.9% of the variability in the response variable with the estimated standard deviation 2.1 termites. From this model, one can observe that there is a treatment and time interaction effect or time-varying dose effect. Before the change point, there is a strong evidence of significant difference between 10mg and 5mg. By the day 7, the effect is 16.821 termites between 10mg ad 5mg. This shows that on average, 10mg is 67.3% more efficient than 5mg in killing termites during the first week. Because of the limited termites in each dish, those who survive after the seventh day stay alive for a while in dishes under 10mg. This means the most effective time frame is about one week for 10mg and unknown dosages. But for the dose 5mg, the effect is still linear by the end of the study. For the
unknown dosage and before change point, it is not significantly different from 5mg, but it is different after the change point.

3.4: Cubic regression splines

In early section, we already talked about cubic smooth splines. Thus, we like to see whether the smooth splines will improve the modeling. The choices of knots are the same as in the previous section. But we have three as a degree of polynomial, so it has more flexibility than the piecewise model we used in the section 3.3. To implement this model, \texttt{bs(variable)} in \texttt{lm} function is used.

\begin{align*}
\text{(a) 5mg} & \\
\text{(b) 10mg} & \\
\text{(c) Unknown dose} & \\
\end{align*}

Figure 13. Cubic B-Spline with 95% confidence intervals.

The cubic spline of the regression function can be expressed for 10mg and unknown dosages as the following:

\begin{equation}
m(x) = \alpha + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 (x - 7)^3
\end{equation}

Figure 13 shows the cubic B-spline with 95% CI. The blue piecewise line represents the piecewise regression model and the straight line is the know location (x=7). From Figure 13, one can see that the piecewise linear model fits the data as well as the B-spline. The later one is more complicated (three parameters vs five parameters) assuming one change point (knot).

Therefore, we recommend the piecewise linear model. Thus, we rerun this model using all data (doses) together and keep all significant terms in the model, the final model is:

\begin{align*}
\hat{y} = \begin{cases}
1.30 - 3.93(Day - 7)I(Day < 7), & \text{for dose} = 10mg, \\
18.45 - 1.20(Day - 7), & \text{for dose} = 5mg, \\
18.45 - 1.20(Day - 7)I(Day < 7) - 0.79(Day - 7)I(Day \geq 7), & \text{for dose} = \text{unknown},
\end{cases}
\end{align*}

(3.6)
with adjusted $R^2=93.7\%$ and estimated standard deviation $=2.14$. This model shows that on average, 10mg is 68.9\% more efficient than 5mg in killing termites during the first week.

### 3.5 Random Effect Model with time covariate within each dish

It is possible that the dishes 1 and 2 from 10mg dosage are not mistaken, but each dish is not identical. Under this assumption, dishes can be considered as random, because dishes can be regarded as a random sample of a larger population of dishes, so dish effect is not constant for all dishes [McCulloch, p. 15]. We first consider random effect model:

$$y_{ijk} = (\beta_0 + \alpha_k) + \beta_1 \text{Day} + \beta_2 I(Dose = 10\text{mg}) + \beta_3 I(Dose = \text{unknown}) + \epsilon_{ijk},$$

where $y_{ijk}$ represents the number of live termites for the $k$th dish on the $j$th dose treatment (5mg, 10mg and unknown), $\beta_0$, $\beta_1$, $\beta_2$ and $\beta_3$ are regression coefficients, and $\epsilon_{ijk}$ is the error term with mean zero and variance $\sigma^2$. The random intercept $\alpha_k$ is the $k$th dish random effect that models the shift in intercept for each dish. It is assumed to be normally distributed with zero mean but variance $\sigma_a^2$. More details can be found in Pinheiro and Bates (2000).

Adjusting the correlation structure $ARMA(1,2)$ with compound symmetry $\rho=.99$, the results are shown in Table 4. It shows that the dose effects are $-8.18$ (10mg) and $2.40$ (unknown dosage) comparing to 5mg. Thus, 10mg is very significant from 5mg with p-value $0.000$, while the unknown dosage is not significant from 5mg with p-value $0.27$. Both dose 10mg and time effects are very significant (p-value=$0.0000$). The estimated standard deviation of the random intercept effect is $0.064$ and the corresponding value for the residual term is $3.894$.

Table 4. Fixed effect part of the random intercept model with time covariate within dish.

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Std.Error</th>
<th>DF</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>27.3936</td>
<td>1.1550</td>
<td>191</td>
<td>23.72</td>
<td>0.0000</td>
</tr>
<tr>
<td>Day</td>
<td>-1.2777</td>
<td>0.0953</td>
<td>191</td>
<td>-13.40</td>
<td>0.0000</td>
</tr>
<tr>
<td>10mg</td>
<td>-8.1813</td>
<td>1.3250</td>
<td>13</td>
<td>-6.17</td>
<td>0.0000</td>
</tr>
<tr>
<td>Unknown</td>
<td>2.2370</td>
<td>1.9395</td>
<td>13</td>
<td>1.15</td>
<td>0.2695</td>
</tr>
</tbody>
</table>

This random effect is consistent with the result from the partial linear model presented earlier and piecewise linear model. From Figure 8, one observes that the partial linear model does not fit the data perfectly. Therefore, we suggest once again the piecewise linear model as the final model since it is much easier to be understood.
4. Conclusions

In this paper, we use four types of statistical models: Partially linear model, piecewise linear model, cubic smooth spline and mix effect model to analyze the termite data. The results show that tropical tree resin, particularly at a higher 10mg concentration is significantly more effective in killing termites. We show that two dishes under 10mg concentration of tree resin were mistaken since the data show insignificantly different under these two dishes from 5mg. The partially linear model shows that there is non-linear (piecewise linear) time effect. Both piecewise linear model and cubic spline smoothing show that the most effective period is the first week. The nonparametric smoothing, cubic spline and piecewise linear model are not significantly different. Mixed effect model is consistent with partial linear and piecewise linear models. The estimated treatment effect is time varying with a change point at day 7. Therefore, we suggest the piecewise linear model (3.6) as the final simplest one for prediction. This final model fits the data with adjusted $R^2=93.7\%$ and the estimated standard deviation 2.14. The final model shows that on average, 10mg is 68.9\% more efficient than 5mg in killing termites during the first week.

Acknowledgement: The authors thank anonymous referees and the Editor-in-Chief for many helpful comments.

5. References