Neural, Parallel, and Scientific Computations 17 (2009) 339-350

READ RATE ANALYSIS OF RFID SYSTEMS USING MIXED MODELS

GOKARNA R. ARYAL[1], LASH MAPA[2], AND CHRIS P. TSOKOS [3]

[1]Department of Mathematics, CS and Statistics
[2] Department of Engineering Technology
Purdue University Calumet
Hammond, IN 46323 USA
[3]Department of Mathematics and Statistics
University of South Florida
Tampa, FL 33620 USA

ABSTRACT: Linear mixed models (LMM) correctly models correlated errors, whereas procedures in the general linear model (GLM) family usually do not. One important question we need to answer is how to select the covariance structure for the given data analysis because it will be used as a starting point in the iterative restricted maximum likelihood(REML) or maximum likelihood(ML) algorithms for estimating parameters. In this study we investigate the choice of the covariance structure for the Radio Frequency Identification (RFID) data considering the effect of the distance and the material on which the tag is being placed. RFID is an automatic identification system used to create an automated tracking solution to provide real-time inventory visibility, operational status, movement of assets through a supply chain, and safety and security for people. Placing the RFID tag on an object with respect to the antenna position is a major contributor to the variation in output tag readability. However, multiple sources of variation and lack of an adequate model (that considers all the variables) results in unpredictability of tag readability.

Key Words: RFID, longitudinal data, repeated measure, mixed model, covariance structure

1. INTRODUCTION

Radio Frequency Identification (RFID) is an automatic identification wireless communication technology that integrates physical objects with the digital data. RFID is a form of automatic identification technology that–much like bar codes and magnetic stripes–can be used to carry data about an object and transfer it to a computer, reducing the time and labor needed for manual data entry. While most automatic identification technologies require at least some labor (scanning, swiping, etc.), an RFID system can be truly automatic[6].

A basic RFID system includes an RFID tag (also known as transponder), an RFID reader (or interrogator) and a host computer. The tag is the basic building

Received June 29, 2009 1061-5369 $15.00

block of RFID. When a tag is energized, the data stored in the tag's memory is transmitted to the reader via radio waves. The reader then communicates the necessary data to the host computer so the computer's software can act on the data. This entire process can be completed with no human intervention. Hence, when the tag is in the vicinity of the RFID antennas the data stored in the tag such as the identity, time, data, and picture of the object can be detected. There are two types of RFID systems, "active" and "passive" that vary in their mode of operation and operating performance. In "active" RFID systems, the tag contains a small battery that enables it to control communication with the reader continuously whereas a "passive" RFID tag has no battery but it derives its power for its operation from the reader's radio communication signal. RFID has the potential to revolutionize warehouse management systems. Dramatic benefits can be achieved through the ability of RFID to more effectively track parts and inventory throughout the supply chain. Figure 1 below is an illustration of the components of RFID systems connected in order to read the data.

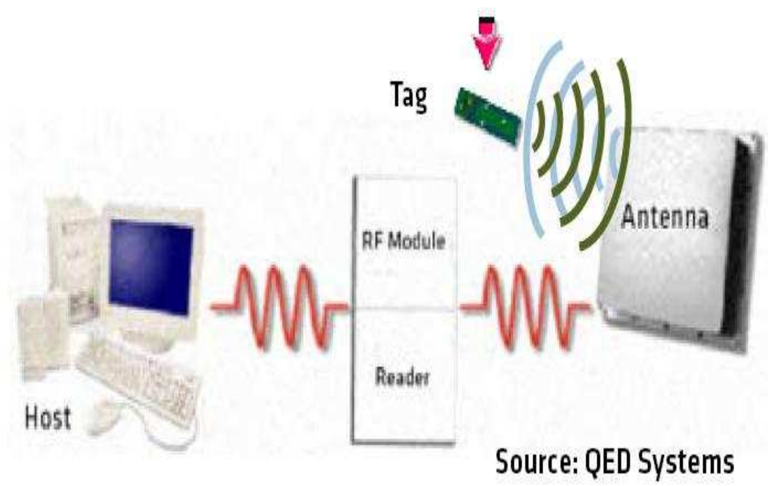

FIGURE 1. Components of RFID system

RFID systems can be used just about anywhere, from clothing tags to missiles to pet tags to food – anywhere that a unique identification system is needed. The tag can carry information as simple as a pet owners name and address or the cleaning instruction on a sweater to as complex as instructions on how to assemble a car. Some auto manufacturers use RFID systems to move cars through an assembly line. There are hundreds of millions of RFID tags currently deployed; it is believed that this number will reach tens of billions within few years [7]. RFID is also called dedicated short range communication (DSRC).

The main aim of data analysis using the linear mixed model is to define an adequate error covariance structure in order to obtain efficient estimates of the regression parameters. Because of the nature of the data for the read rate analysis of RFID system it is our interest to identify the appropriate covariance structure associated with the subject data. First we will introduce the linear mixed model and then perform the analysis of the data.

2. LINEAR MIXED MODEL

The normal linear model with p parameters is given by

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \epsilon_i$$

where we usually take $x_{i1} = 1$ so that β_1 is a constant or intercept and we assume that $\epsilon_i \sim NID(0, \sigma^2)$. We express the linear model in matrix form as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$
$$\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$$

where $\mathbf{y} = (y_1, y_2, \ldots, y_n)'$ is the response vector, $\mathbf{X}$ is the design matrix, with typical row $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})$; $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)'$ is the vector of regression coefficients; $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)'$ is the vector of errors; N_n represents the n-variable multivariate-normal distribution; $\mathbf{0}$ is an $n \times 1$ vector of zeroes; and $\mathbf{I}_n$ is the order-n identity matrix. Note that the model has one random effect, the error term ϵ_i.

The mixed-effect models (or just mixed models) on the other hand include not only the fixed effect but an additional random-effect terms, and are often appropriate for representing clustered, and therefore dependent, data arising, for example, when data are collected hierarchically, when observations are taken on related individuals or when data are gathered over time on the same individuals.

The linear mixed model is in fact an application of the general linear model and is defined as

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i$$

where $\mathbf{y}_i$ is the $n_i \times 1$ response vector of repeated measures data for the i^{th} subject, $\mathbf{X}_i$ is the $n_i \times p$ known design matrix for the fixed effects for observations in the group i, $\boldsymbol{\beta}$ is the $p \times 1$ vector of fixed-effects parameters, $\mathbf{Z}_i$ is another known design matrix for the random effects for observations in group i, $\boldsymbol{\eta}_i$ is the vector of random effect parameters for group i, and $\boldsymbol{\epsilon}_i$ is the $n_i \times 1$ vector of errors for observations in group i. The fixed part of the model is specified by $\mathbf{X}_i\boldsymbol{\beta}$ and the random part by $\mathbf{Z}_i\boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i$. Note that the linear mixed model assumes that $\boldsymbol{\eta}_i$ and $\boldsymbol{\epsilon}_i$ have an independent multivariate normal distribution with a mean zero and covariance matrix say $\mathbf{G}_i$ and $\mathbf{R}_i$ respectively. Therefore, we have

$$E(\mathbf{y}_i) = \mathbf{X}_i\boldsymbol{\beta}$$
$$Var(\mathbf{y}_i) = \mathbf{Z}_i\mathbf{G}_i\mathbf{Z}_i' + \mathbf{R}_i$$

where $\mathbf{Z}_i\mathbf{G}_i\mathbf{Z}_i^{'}$ is the between-subjects component and $\mathbf{R}_i$ the within-subject component, such that the covariance matrix of observations is a function of $\mathbf{G}_i = Var(\boldsymbol{\eta}_i)$ and $\mathbf{R}_i = Var(\boldsymbol{\epsilon}_i)$. It should be noted that the decomposition of the $Var(\mathbf{y}_i)$ into $\mathbf{G}_i$ and $\mathbf{R}_i$ is not necessarily unique, which means there may be two or more structures for $\mathbf{G}_i$ and $\mathbf{R}_i$ that yield the same variance-covariance matrix for the i^{th} unit. Note that the fixed effects define the expected values of the observations, while the random effects represent the variances and covariances of the observations [8].

In longitudinal data, the repeated measures can be considered as dependent variables. For that reason, the multivariate analysis is a good alternative to the univariate analysis. However, the advantage of the linear mixed model over traditional analytic approaches to longitudinal data is that it models the covariance matrix. Thus, the fixed parameter estimates are more efficient and the model is more powerful in terms of testing the effects associated with the repeated measures. This approach is also more robust than traditional univariate and multivariate tests. However, when the covariance structure is not adequately fitted and sample sizes are small, a positive bias in type I error is produced [12]. Therefore, it is our interest to find which covariance structure best describe the RFID data. First we introduce some basic covariance structures.

2.1. **Simple (VC) or Uncorrelated.** The simple model assume that the observations are independent over repeated measurements and homogeneous variance. Consequently, we have $\mathbf{G}_i = \mathbf{0}$ and $\mathbf{R}_i = \sigma^2\mathbf{I}$.

$$\mathbf{VC} = \mathbf{R}_i = \sigma^2 \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

This model may not be realistic in most of the cases with repeated measures obtained from the same subject due to their non-independence.

2.2. **Unstructured(UN).** In this structure the variance at all time points and the correlation between any two measurements are all differen. This is the most heterogeneous structure and offers the best fit. It takes the following form:

$$\mathbf{UN} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \sigma_{n3} & \cdots & \sigma_n^2 \end{bmatrix}$$

Note that in this structure we have to estimate as many as $\frac{n(n+1)}{2}$ parameters, where n is the number of repeated measure or observations. The most general form assumes that the homogeneous variance which means $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2$ and in this case it is required to estimate only $\frac{n(n-1)}{2} + 1$ parameters.

2.3. **Equicorrelation and Compound Symmetry (CS).** In equicorrelation model we assume that all repeated measurements are equally correlated

$$\mathbf{R}_i = \sigma^2 \begin{bmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho & \rho & \rho & \cdots & 1 \end{bmatrix}$$

This structure is also called spherical or exchangeable and might be applicable to cluster data. In this structure ρ is called the intra-class correlation coefficient between two members of the same cluster and is a relative measure of the within cluster similarity [2]. A spacial case of autocorrelation is called compound symmetry where the correlation coefficient ρ is given by $\rho = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}$ for some σ_a^2 and σ_e^2. For compound symmetry case we have

$$\mathbf{R}_i = \begin{bmatrix} \sigma_a^2 + \sigma_e^2 & \sigma_a^2 & \sigma_a^2 & \cdots & \sigma_a^2 \\ \sigma_a^2 & \sigma_a^2 + \sigma_e^2 & \sigma_a^2 & \cdots & \sigma_a^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sigma_a^2 & \sigma_a^2 & \sigma_a^2 & \cdots & \sigma_a^2 + \sigma_e^2 \end{bmatrix}$$

2.4. **First Order Autoregressive (AR(1)).** The first-order autoregressive model assumes that measurements which are close to one another in time will show high correlations. Its structure is homogeneous, the variances are equal and the covariances between observations of the same object decrease exponentially as the lag increases. Hence, it comprises two parameters σ^2 and ρ which are the variance of the observations and that of the correlation between adjacent observations respectively. Hence, the covariance structure is given by

$$\mathbf{AR}(1) = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho^{n-1} & \rho^{n-2} & & \cdots & 1 \end{bmatrix}.$$

3. DATA ANALYSIS

In this section we will analyze the data concerning the read rate of the radio frequency identification. The data were collected in a laboratory with minimal interference from external factors. The data has been collected conducting an experiment by placing (attaching)the tag in the surface of one of the 7 materials (none [air], cardboard, plywood, waferboard, drywall, brass and water) and measured at 4 different time points (5 sec., 10 sec., 15 sec. and 20 sec.). The choice of the time is just to study the variation in the read rates within the industrial, scientific and medical (ISM) recommended time of 20 seconds [13]. The experiment was replicated five times. The data was collected in the following identical (fixed) conditions in order to avoid any environmental effects on the response.

Tag	Alien Squiggle-9540
Attenuation	0-dB
Tag Mounting	Vertical
Tag height from the ground	$3'8''$
Antenna height from the ground	$5'8''$
Antenna angle	15 degree
Reader	Motorola XR450
Antenna	Alien ALR-9611-CR

The aim of the present study is to examine whether there were systematic inter-individual differences in the intra-individual change in the read rate of RFID system over time as a consequence of the material that the tag is being placed. Throughout this note distance means distance between the tag and the antenna, time means time (in seconds) taken to get the readings, material means one of the 7 conditions (None [air], cardboard, plywood, waferboard, drywall, brass and water) in which the tag is being placed.

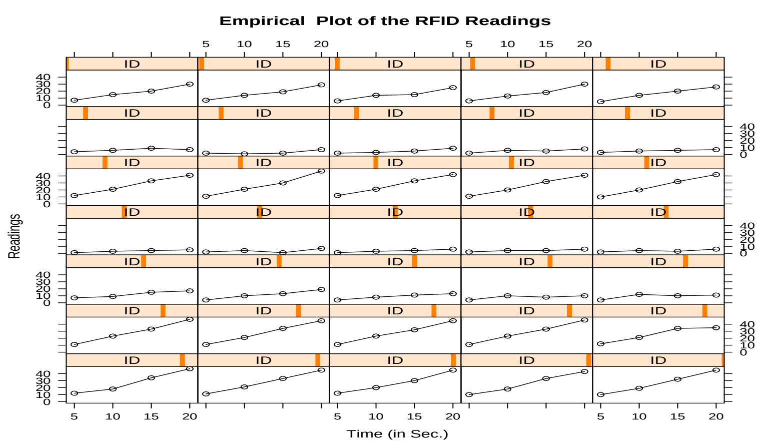

FIGURE 2. Empirical growth plot

Figure 2 displays the empirical growth of the readings as time elapsed. In this figure ID represents one of the seven conditions of the material. Each row represents a single material and the column represents simply the replications. This display is created by using R[9]. Observe that not all conditions are performing equally. This data can also be displayed in a spaghetti plot as illustrated in figure 3 below. This plot represents all the data measured at each time points.

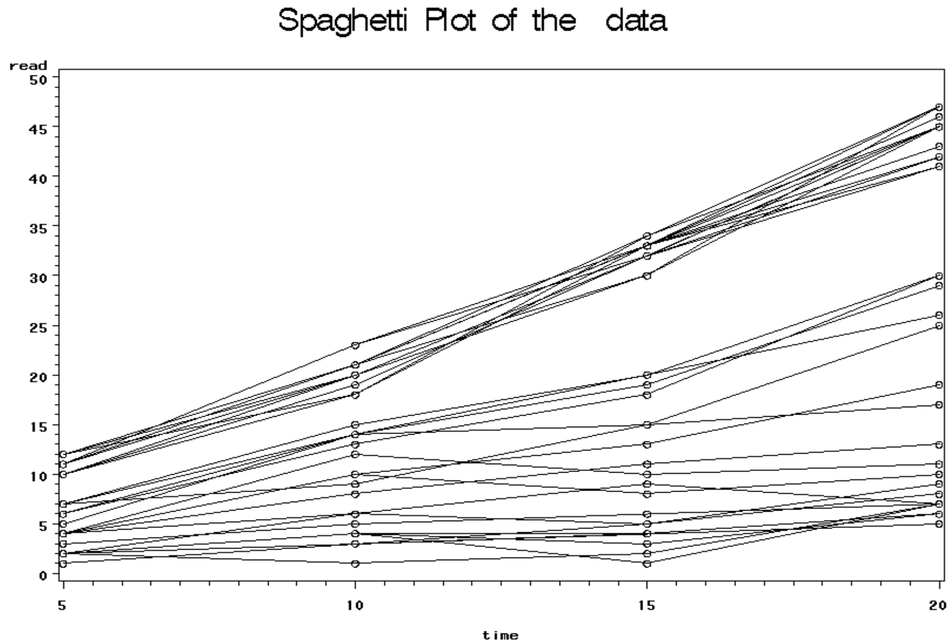

Figure 3. Spaghetti plot of the read rate data

In table 1 we provide a summary of the data when the tag is placed on different objects keeping distance between the tag and the antenna at 8 feet.

Material	5 Sec.	10 Sec.	15 Sec.	20 Sec.
Air	6.2	14	18.4	28.0
Drywall	2.6	4.2	5.4	7.6
Plywood	1.6	3.6	3.2	6.0
Waferboard	11.2	20.6	32.0	42.6
Cardboard	4.6	9.8	11.4	14.0
Brass	11.2	22.2	33.2	45.6
Water	11.0	19.2	32.4	45.0

Table 1. Mean readings when the tag is being placed on different materials

Figure 4 displays the mean reading values over time in response of the material in which the tag is being placed.

Observe that not all martials are responding equally to the time. When the tag is being attached on brass, water or waferboard we are consistently having high reading than the condition of having no material (Air) whereas plywood, drywall and cardboard have low read rate. In order to analyze the data we need to determine the covariance structure of $\mathbf{Z}_i\boldsymbol{\eta}_i$ and $\boldsymbol{\epsilon}_i$. We will be using the method of restricted maximum likelihood (REML) also known as residual maximum likelihood method to

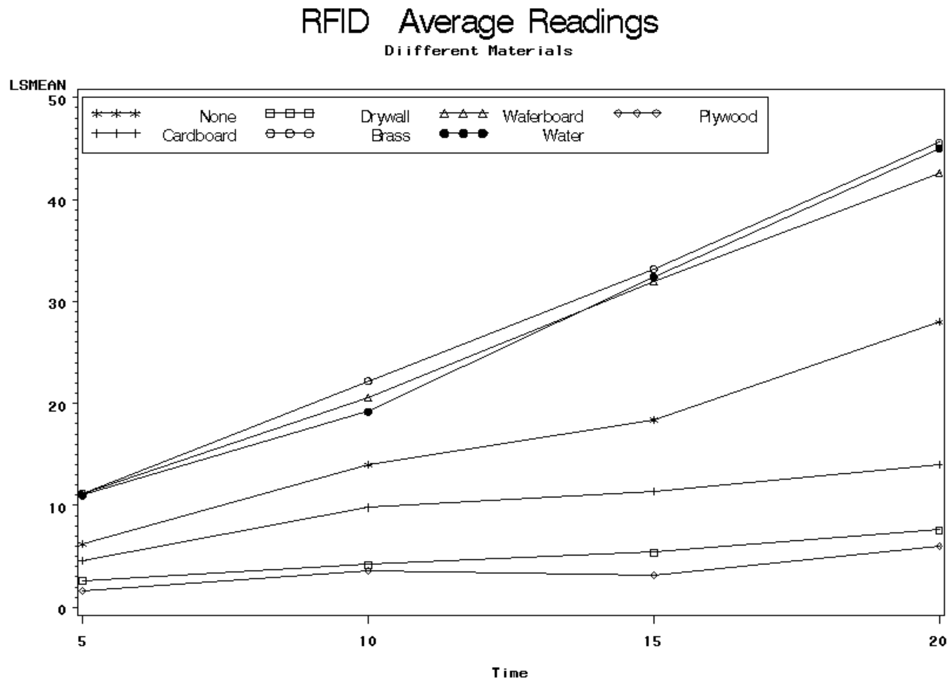

FIGURE 4. Mean readings as the tag is placed on different materials

estimate the covariance parameters. Proc mixed in SAS [10] estimates these parameter using REML. Note that within-subjects heterogeneity occurs when the variances across repeated measures are unequal while between-subjects heterogeneity occurs when covariance matrices differ across groups. As the data suggests in this study we will analyze the data assuming that there is a heterogeneity within subject and we will compare the model with the models that assume within-subject homogeneity. First we need to determine the best fit of the covariance structure of the subject data. The competing covariance structure are compound symmetry(CS), Unstructured (UN) and First order autoregressive(AR(1)) [4]. With the mixed model it is possible to select the covariance structure which best describes the subject data [3]. Widely used criteria to select the best covariance structures are: -2 times the residual log-likelihood (-2RLL), Akaike's Information Criterion (AIC) [1] or its corrected version for finite samples (AICC)[5] and the Bayesian Information Criterion (BIC) [11]. These criteria are indices of relative goodness-of-fit and may be used to compare models with different covariance structures and the same fixed effects. The -2RLL criterion, which measures the deviance between the data and the model, can be used to determine which covariance structure is the most adequate. The smaller this index is, the better is the fit. In order to find the parsimonious model -2RLL is not appropriate as not all covariance structure have the same number of parameters. Therefore, the AIC and BIC fit criteria have been proposed on the basis of the deviance but they penalize due to the number of parameters to be estimated,

$$AIC = -2RLL + 2d$$
$$BIC = -2RLL + d\ln(N)$$

where d is the number of parameters to be estimated and N is the sample size. Also note that in both of these measures smaller the value is the better is the fit.

The aim of study is to examine whether there is systematic interindividual differences in intraindividual change in the read rate over time as a consequence of the object in which the tag is being placed. Table below shows the variables and their levels included in this study.

Variable	Description
Material	The material in which the tag is being placed(7 different materials)
Distance	Distance between the tag and antenna $(2', 5', 8')$
Time	Time from 5 sec. to 20 sec.(4 levels)
Read	Number of readings

Table 2 compares the covariance structures for the subject data when we consider the effect of the material in which the tag is being placed and the distance between the tag and the antenna.

Structure	CS	UN	AR(1)
Covariance parameters	2	10	2
-2 RLL	2586.4	1862.2	2378.6
AIC	2590.4	1882.2	2382.6
AICC	2590.4	1882.8	2382.6
BIC	2595.7	1908.7	2387.6
Chi-Square	197.20	921.40	405.05
DF	1	9	1
$Pr > ChiSq$	< 0.001	< 0.001	< 0.001

Table 2. Fit statistics and null model likelihood ratio test

Observe that the unstructured(UN) covariance matrix best fit the subject data. Now after determining the covariance structure we study the significance of the fixed effect. We have the following fixed effect table:

Observe that all fixed effects are statistically significant at $\alpha = 0.05$ but the distance is not significant at $\alpha = 0.01$.

Also note that under the unstructured covariance criteria we have the least square means of the material as in the table 3 along with its 95% confidence interval.

Effect	Num. DF	Denom. DF	F-value	$Pr > F$
Material	6	97	3.26	0.0058
Distance	1	97	4.48	0.0369
Time	3	97	321.11	< 0.001
Material* Time	18	97	2.25	0.0062

Material	Estimate	SE	DF	t -Value	$Pr > \|t\|$	Lower	Upper
None(Air)	22.13	2.098	97	10.59	$< .0001$	17.98	26.28
Drywall	21.38	2.098	97	10.23	$< .0001$	17.23	25.53
Waferboard	27.53	2.098	97	13.17	$< .0001$	23.38	31.68
Plywood	21.62	2.098	97	10.34	$< .0001$	17.47	25.77
Cardboard	21.42	2.098	97	10.24	$< .0001$	17.27	25.57
Brass	28.83	2.098	97	13.79	$< .0001$	24.68	32.98
Water	29.45	2.098	97	14.09	$< .0001$	25.30	33.60

TABLE 3. Least Squares Means output for material

4. CONCLUDING REMARKS

In RFID system it is extremely important to identify the position of the tag and antenna and the time it takes to give the best readings. But it has been observed that the object in which the tag is being attached(placed) is also equally important. Because of the nature of the data, in this study, we investigate the best covariance structure of the subject data. It has been observed that the RFID system possesses unstructured (UN) covariance structure. We expect that this study will serve as a reference and help to advance future research in this area.

REFERENCES

[1] Akaike, H. (1974) A new look at the statistical model identification. *IEEE Transaction on Automatic Control*, AC-19, 716–723

[2] Annette J. Dobson (2002)*An itroduction to generalized Linear Models*. Chapman & Hall/ CRC

[3] Bono R., A. Jaume & B. Nekane (2007) Using linear mixed models in longitudinal studies: Application of SAS PROC MIXED. *Revista Electrónica de Metodología Aplicada* Vol. 12 pp. 15–31.

[4] Hedekar, D. & Gibbons, R.(2006)*Longitudinal Data Analysis*, Wiley Series in Probability and Statistics

[5] Hurvich, C. M., & Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika bf 76*, 297–307.

[6] Kator, Corinne (2008) RFID basics. *Modern Materials Handling* Vol.63 no. 2, 38–40 .

[7] Kharif, Olga (2006) What's lurking in that RFID tag? http://www.businessweek.com.

[8] Littell, R. C., Pendergast, J., and Natarajan, R. (2000). Modelling covariance structure in the analysis of repeated measures data. *Statistics in Medicine* **19**, 1793–1819.

[9] R: *A language for data analysis and graphics.*

[10] SAS Institute Inc. (2004).SAS Online Doc 9.1.3. Cary, NC: SAS Institute Inc.

[11] Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* **6**, 461–464.

[12] Wright, S. P., and Wolfinger, R. D. (1997). *Repeated measures analysis using MIXED models: Some simulation results.* In T. G. Gregoire, D. R. Brillinger, P. J. Diggle, E. Russek-Cohen, W. G. Warren and R. D. Wolfinger (Eds.), Modelling longitudinal and spatially correlated data (147–158). New York: Springer-Verlag.

[13] FCC Regulations for ISM Band Devices: http://www.semtech.com