

Research Article

A New Approach to Investigation of the Relationship of VLF Signals by Using Longitudinal Analysis Models

E. Guzel,¹ M. Yaşar,¹ M. B. Kılıç,² and M. Canyılmaz¹

¹ Department of Physics, Faculty of Science, Firat University, 23119 Elazığ, Turkey

² Department of Statistics, Faculty of Arts and Science, Middle East Technical University, 06800 Ankara, Turkey

Correspondence should be addressed to M. Yaşar; m.yasar3329@gmail.com

Received 27 April 2013; Accepted 24 October 2013

Academic Editor: Stefano Selleri

Copyright © 2013 E. Guzel et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Longitudinal analysis models are applied to analyze very low frequency (VLF) electromagnetic waves signals. They are useful in realization of nuance of relationship between parameters of concern which change over time. VLF data are so important for the communication and determining the disturbances in the lower ionosphere. In this study, we used a four-day VLF signal data which come from transmitter stations located at two different countries to Elazığ receiver system. A very important feature of this data is that there is a strong relationship between days and signals. At first, the exploratory data analysis was applied for identification of data. Then, marginal, random-effects, and transitional models were used to analyze the data, respectively. The marginal model is found more appropriate than the other models used for VLF data analyses. As a result of this process, a new method has been added to the analysis of VLF signal data.

1. Introduction

VLF radio waves (3–30 kHz) are guided by the spherical waveguide formed between the Earth's surface and the lower ionosphere and can propagate long distances [1]. VLF stations located in various places of the world have been used for different purposes, such as communication with submerged military submarines and determining remote sensing of transient D Ionospheric disturbance region [2]. The observed properties of VLF signals can be used to determine the spatial and temporal structure of localized disturbances in the lower ionosphere [3]. These disturbances are often manifested as sudden changes in the amplitude or phase of a VLF signal. Using VLF remote sensing techniques, D region disturbances have been observed in association with solar flares, meteor showers, auroral enhancements, gamma ray flares of extraterrestrial origin, and the direct and indirect effects of lightning [4, 5].

Statistical studies provide a very important “baseline” reference for case studies and a global view that can be used to identify different processes for a variety of different

geophysical conditions [6]. Statisticians, as well as others who apply statistical methods to data, have often made preliminary examinations of data in order to explore their behavior. In this sense, exploratory data analysis has long been a part of statistical practice [7]. In statistics, exploratory data analysis (EDA) is an approach to analyzing datasets to summarize their main characteristics in easy-to-understand form, often with visual graphs, without using a statistical model or having formulated a hypothesis. Exploratory data analysis was promoted by Hoaglin et al. to encourage statisticians visually to examine their datasets, to formulate hypotheses that could be tested on new datasets [8].

The possibility of correlations between responses needs to be taken into account in the data analysis. Various models can be used to handle such correlations. Longitudinal analysis studies are progressively widespread in many scientific research fields. The aim of making longitudinal study is to look at the variation of process across time periods [9]. From this point of view, marginal modeling allows for inferences about parameters averaged over the whole population. Random-effects modeling deliberately provides

inferences about the variability between respondents. Transitional model investigates the reasons for the change of the responses [10].

This paper presents a comparative study of longitudinal analysis techniques and models applied to experimental VLF signals and investigations of the relationship between two VLF transmitters.

The organization of this paper is as follows. Experimental systems including VLF transmitters and receiver are introduced in Section 2. Initial EDA analysis is given in Section 3. Section 4 discusses various possible models (marginal, random-effects, and transitional models) in some detail. Some concluding remarks are given in Section 5.

2. Experimental Systems

The experimental VLF signals used in this study have been obtained from VLF receiver system located at Firat University, Elazig, Turkey (38.40°N, 39.12°E). The wideband signal is detected by two orthogonal magnetic loop antennas, which are right-isosceles triangles (2.6 m base, 12 turns). The data are band pass filtered to a range of 9–45 kHz and sampled at 100 kHz with 16 bit resolution and GPS timing. The demodulated amplitude and phase of the narrowband signals are recorded typically during 13:00–06:00 [1]. The amplitude data is only included in this study. The GCPs of the Elazig VLF receiver system are shown in Figure 1 and summarized in Table 1.

3. Exploratory Data Analysis (EDA)

Exploratory data analysis emphasizes looking at and analyzing residuals to help understand data and to investigate models. A fit describing the data and examining residuals often gives guidance for improving that description. Strong patterns usually indicate that further refinement of the fit will be a worthwhile aid to understand the structure in the raw data. A regression setting illustrates the use of residuals and some associated plots. EDA advocates exploring data for patterns and relationships without requiring prior hypotheses. EDA suggests that analyses are more scientifically useful and productive when data have been transformed to comply with basic assumptions. Exploratory analyses can incorporate methods of statistical inference but use them more as indicators of the strength of a relationship or the fit of a model than as confirmation of a hypothesis [7].

In this study, we analyzed serial 4 days of the VLF signals of June, 2008. The parameters used for the analysis are day, station name, signal amplitude, and start hour. We have used exploratory graphics to summarize our dataset. The means of the amplitude of the VLF signals for each station and everyday were calculated preliminarily and given by Table 2.

Spaghetti plots and Drawman displays which are the parts of the exploratory analysis are given in Figures 2 and 3.

As can be seen in Figure 2, there is no much difference between signals for DHO signal, but some deviation was seen at the HWU signal. In particular lines are not at the same direction between 2nd and 3rd. So, we can say that some

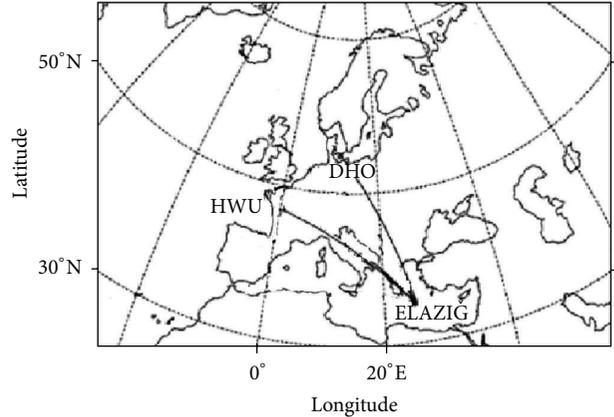


FIGURE 1: The GCPs connecting the Elazig receiver with the VLF stations.

physical events occurred at the Earth-ionosphere waveguide at these days.

Scatter plot is giving the relationship between two signals. For example, the relationship between 1st and 2nd day is shown by circular area in Figure 3. The pink and blue points are gathered at the corners separately. This means that there is a strong relationship between days and signals.

Correlation and covariance matrices are given in Tables 3 and 4, respectively. Signal correlation is extremely high, positive, and linear. Since all correlations are close to 0.99, the variances increase from 288.47 to 482.29. The compound symmetry (exchangeable) covariance structure is more suitable for this data.

4. Longitudinal Approach for the Analysis of VLF Signal Data

Longitudinal approach consists of coalition of regression and time-series analysis. Longitudinal models have an important role in the literature. Nevertheless, it is used mostly in social sciences, economy, sociology, and medicine fields. Longitudinal data analysis methods have been developed because a lot of databases have become usable to experimental investigations [11].

This section studies new approaches to the analysis of VLF data. We consider longitudinal analysis techniques: marginal, random-effects, and transitional models, respectively.

4.1. Marginal Model. In marginal model, the relation between the response vector and the explanatory variables at each time and how these relations change over time are given. In marginal modeling, the joint outcome is modeled by using a marginal probability for each response variable, and correlation structure, by using the odds ratio, parameterizes the relation between the two dependent variables [10]. The model is given in the following equation:

$$E(Y_{it}) = 48.55 + 1.74 * \text{Time} - 32 * \text{Signal} - 0.42 * (\text{Time} * \text{Signal}). \quad (1)$$

TABLE 1: The list of transmitters and receiver.

Frequency (kHz)	Station	Location	Distance between transmitters and receiver (km)
18.3	HWU	Le Blanc, France	3,170
23.4	DHO	Rhauderfehn, Germany	2,860
Receiver	FF	Elazığ, Turkey	

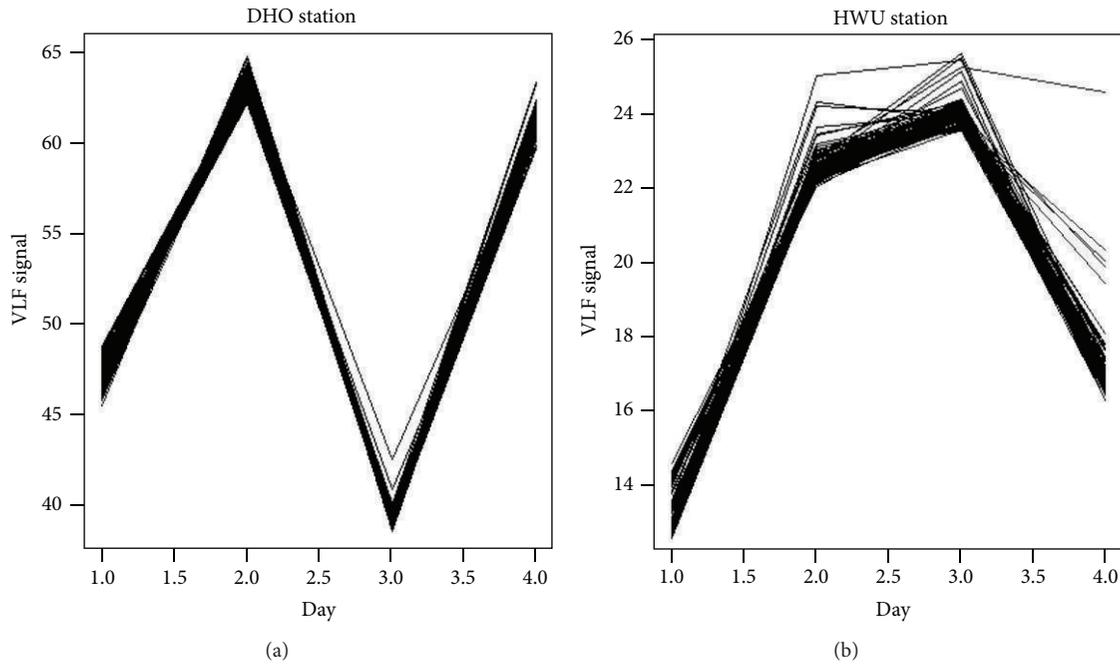
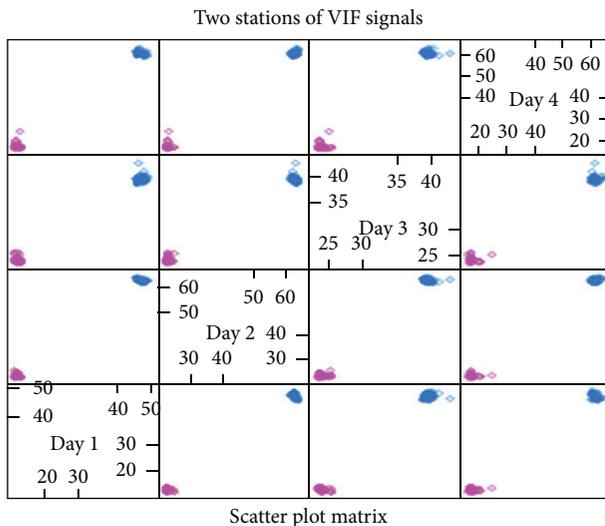


FIGURE 2: Spaghetti plots of VLF signals for the two stations.



- ◇ HWU
- ◇ DHO

FIGURE 3: Scatter plot of VLF signals for the two stations.

TABLE 2: The means of the VLF signals.

Station	Day 1	Day 2	Day 3	Day 4
HWU	13.45	22.70	24.14	17.35
DHO	47.30	63.63	39.58	61.13

TABLE 3: Correlation matrix of the two stations.

Day	1st	2nd	3rd	4th
1st	1	0.998	0.997	0.997
2nd	—	1	0.997	0.998
3rd	—	—	1	0.997
4th	—	—	—	1

The parameters are interpreted as the constant (intercept) and slopes of each fixed-effect predictor such as time (day), signal (signal amplitude), and interaction of time and signal. The intercept is interpreted as the mean of VLF signals when all the predictors have a value of zero. A one-unit increase in the predictor time to new experiences time corresponds

TABLE 4: Covariance matrix of the two stations.

Day	1st	2nd	3rd	4th
1st	288.47	347.93	131.34	372.24
2nd	—	421.20	158.76	450.25
3rd	—	—	60.088	169.83
4th	—	—	—	482.29

to a 1.74 increase in the outcome VLF signals for each station. Since, the coefficient for HWU signal is negative, HWU signal has lower VLF times compared to DHO signal when all other factors are kept constant. In other words, VLF signal time is expected to be lower when the HWU signal is used. This is a result expected for the station. The models of coefficients are statistically significant, because robust z values have a high number. The residuals seem high. The residuals and coefficients for marginal model are given in Table 5.

4.2. Random-Effects Model. In random-effects model, a natural heterogeneity across individuals in their regression coefficients is investigated and the heterogeneity can be handled by a probability distribution. Correlation among observations for the same individual arises from their unobservable variables [10]. The model is given in the following equation:

$$E(Y_{it}) = 48.55 + 1.74 * \text{Time} - 32.43 * \text{Station} - 0.42 * (\text{Time} * \text{Signal}) + \hat{b}_{oi}. \quad (2)$$

This relation can be rewritten for the first VLF signals:

$$E(Y_{1t}) = 48.55 + 1.74 * \text{Time} - 32.43 * \text{Station} - 0.42 * (\text{Time} * \text{Signal}) + 0. \quad (3)$$

In the random-effects model, we have the same model as marginal model and the predictor coefficients are interpreted as the coefficients in the marginal model. The fixed coefficients for random-effects model are given in Table 6. The coefficient sign is an expected result as the coefficient of signal. The interaction between signal and time is not statistically significant, because of $t = 0.90 < 1.96$.

The random effects of coefficients of the output provide estimations for the random effects in the form of variances and standard deviations (Table 7). The variance is equal to 0. The variance estimates are of interest here because we can add them together to find the total variance and then divide that total by each random effect to see what proportion of the random effect variance is attributable to each random effect (similar to R^2 in traditional regression). We sum the variance components, $0 + 55.28 = 55.28$, and divide this total variance by our variance to give us the proportion of variance accounted for, which indicates whether or not this is meaningful ($0/55.28 = 0.0$). If all the percentages for each random effect are equal to zero, then the random effects are not present and linear mixed modeling is not appropriate. You must remove the random effects from the model and use general linear or generalized linear modeling instead.

On the other hand, the output simply provides the correlations among the fixed effects variables given in Table 8.

This can be used to assess multicollinearity. The predictors are related to time and intercept (-0.913), also signal and interaction of signal and time. Therefore, multicollinearity is a concern. As a result, the point estimates for the fixed effects are exactly the same as the marginal model.

The two rows of fit statistics showed the Akaike Information Criterion and the Bayesian Information Criterion, the log-likelihood, the deviance for the maximum likelihood criterion, and the deviance for the REML criterion in Table 9.

The smaller deviance indicates better fit; it may be useful to add a random slope or to find the better fit or can high variance components. So, we add random slope; the results of the model are given in the following equation:

$$E(Y_{it}) = 48.55 + 1.74 * \text{Time} - 32 * \text{Station} - 0.42 * (\text{Time} * \text{Station}) + \hat{b}_{oi} + \hat{b}_{ti} * \text{Time}. \quad (4)$$

For instance, for the first VLF signals,

$$E(Y_{1t}) = 48.55 + 1.74 * \text{Time} - 32 * \text{Station} - 0.42 * (\text{Time} * \text{Station}) + 1.035697e - 18. * \text{Time}. \quad (5)$$

The deviance of time is smaller than others (time, signal, signal * time). So, we select time covariate as random part. The coefficient of sign is an expected result for signal. The variances for both random intercept (0.0) and slope ($4.2152e-14$) are very low. Since the random coefficients are very low, it may not be beneficial to add the random coefficient for this data. For this situation, to test, we can use ANOVA. If we compare random-effects models with ANOVA, then the corresponding P value of 1 implies that estimation of coefficient may not be beneficial enough at 5% significance level. The histograms of the residuals for two models are given in Figure 4. For two random-effects model is exactly same histograms of residuals.

In this study, we have also studied subsequence of original data and made a simulation study using Bayesian procedure. The purpose of Bayesian approach is to provide the posterior variance of model factor from the likelihood function and the prior variance [12]. We can estimate random- and fixed-effect coefficients by using the highest posterior density (HPD) intervals for the parameters of an MCMC distribution. Random effects of coefficients are not different for each VLF signal. They change between 0.3 and 0.4. We can estimate both lower and higher coefficients. The lower coefficient of intercept is equal to 46.84 and the higher coefficient of intercept is equal to 50.36. Similarly, the lower coefficient of time is equal to 1.12 and the higher coefficient of time is equal to 2.41.

4.3. Transitional Model. In transitional models, the probability distribution of the outcome of individual i at time t , Y_{it} , is a function of the individual's covariates at time t , X_{it} , and the individual's outcome history Y_{i1}, \dots, Y_{it-1} , $t > 1$. Such

TABLE 5: The residuals and coefficients for marginal model.

(a)					
Min.	1Q	Median	3Q	Max.	
-15.03	-4.24	3.32	4.85	12.96	
(b)					
Coefficients	Estimate	Naive S.E.	Naive Z.	Robust S.E.	Robust Z.
Intercept	48.55	0.96	50.35	0.08	545.74
Time	1.74	0.38	4.52	0.03	47.13
Signal	-32.43	1.36	—	0.11	-286.12
Time * Signal	-0.42	0.54	-0.78	0.05	-8.49

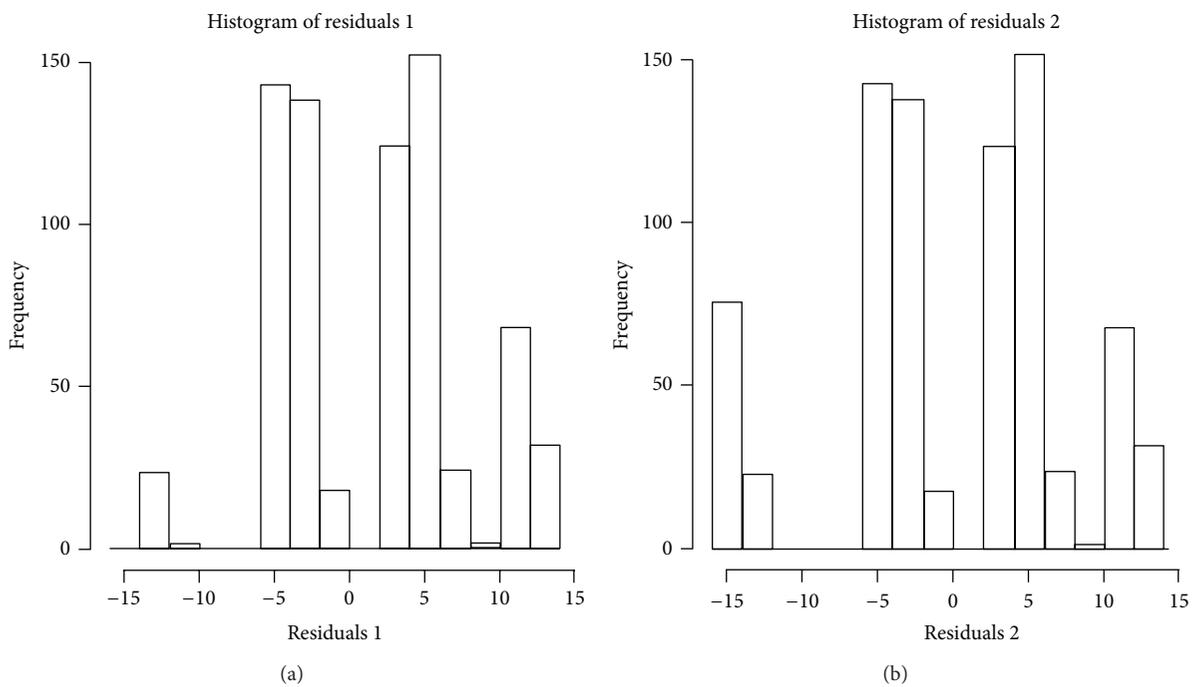


FIGURE 4: Comparing residuals of the two models.

TABLE 6: Fixed effects for random-effects model.

Coefficient	Estimate	Std. error	t value
Intercept	48.5557	0.9151	53.06
Time	1.7436	0.3341	5.22
Signal	-32.4309	1.2941	-25.06
Time * Signal	-0.4276	0.4726	-0.90

TABLE 7: The variances and standard deviations for random-effects model.

Variance	Standard deviations
0.000	0.0000
55.287	7.4717

TABLE 8: Correlation of fixed effects.

Variable	Intercept	Time	Signal
Time	-0.913	—	—
Signal	-0.707	0.645	—
Time * Signal	0.645	-0.707	-0.913

TABLE 9: Model selection of criteria.

AIC	BIC	LogLik	Deviance	REML Dev.
5497	5525	-2743	5484	5485

TABLE 10: The residuals and coefficients for transitional model.

(a)					
Min.	1Q	Median	3Q	Max.	
-3.93	-2.5	-0.29	0.96	9.31	
(b)					
Coefficients	Estimate	Naive S.E.	Naive Z.	Robust S.E.	Robust Z.
Intercept	117.76	1.09	107.37	0.67	174.3
Time	-5.28	0.21	-24.37	0.07	-69.06
Signal	-81.18	1.04	-77.35	0.58	-138.61
Ylag 1	-1.04	0.017	-60.35	0.01	-90.81
Time * Signal	8.19	0.33	24.65	0.14	58.25

models are appropriate when there is a natural sequencing of the responses. The model is given in the following equation:

$$E(Y_{it}) = 117.76 - 5.28 * \text{Time} - 81.19 * \text{Station} + 8.19 * (\text{Time} * \text{Station}) - 1.04 * Y_{t-1}. \quad (6)$$

The lag1 of response and the covariates have been included for transitional model and the Y_{t-1} of coefficient is -1.04 values. This value is not an expected result. The parameter of estimates highly changed in this model. This probably may be a result of high correlation among these variables. The residuals and coefficients are given in Table 10.

The time coefficient of signal is negative and this is an expected result. The models of coefficients are statistically significant, because robust z values are of a high number. This study may be modeled with transitional model, but there is a multicollinearity problem.

5. Result and Discussion

The main theme of presenting this paper is also to illustrate the use of various models in the analysis of VLF signals. In this paper, we used a four-day VLF signal data which come from transmitter stations located at different two countries to Elazığ receiver system. The aim of this study is to use new statistical method for analyses of VLF signal datasets and to make a comparison among the signals taken from two different stations. Therefore, longitudinal analysis techniques were applied over VLF signal data taken from two different stations. Longitudinal approach includes study where the value of the output and covariates are seen. Moreover, it provides the research and contrast to changes in the response of relationship point over time.

The outcome of study includes correlation and covariance matrix for the two stations, spaghetti (longitudinal) plots and scatter plot of VLF signals for the two stations, the residuals and coefficients for marginal model, fixed effects for random-effects model, the residuals and coefficients for transitional model, and comparing residuals of two models.

In consequence of correlation and covariance matrix made for two stations were determined a very high intimate between signal and data. Spaghetti (longitudinal) and scatter plots show that there are differences between stations. For

the DHO station, high correlation between day and signal is determined, but for the HWU station deviations at some days are detected.

For this study, as can be seen in Figure 4 and Table 6, random-effects model and coefficients of each VLF signal are not different from each other because random-effects model has not reduced deviations which occurred in marginal model. Moreover, these models show that simulation study and subsequence of data are not suitable for model VLF signals. For transitional model, there is multicollinearity problem. On the other hand, as may be seen from consequence in Table 5, the marginal model is more appropriate than the other models to the modeling of the VLF signals data. As a result of this process, a new method has been added to the analysis of VLF signal data.

Acknowledgment

The authors wish to thank Associate Professor Dr. Ozlem Ilk for her assistance and contributions.

References

- [1] E. Guzel, M. Canyılmaz, and M. Turk, "Application of wavelet-based denoising techniques to remote sensing very low frequency signals," *Radio Science*, vol. 46, no. 2, pp. 1-9, 2011.
- [2] M. P. Johnson, *VLF imaging of lightning-induced ionospheric disturbances [Ph.D. thesis]*, Stanford University, Stanford, Calif, USA, 2000.
- [3] G. Bainbridge and U. S. Inan, "Ionospheric D region electron density profiles derived from the measured interference pattern of VLF waveguide modes," *Radio Science*, vol. 38, no. 4, p. 1077, 2003.
- [4] U. S. Inan, J. V. Rodriguez, and V. P. Idone, "VLF signatures of lightning-induced heating and ionization the nighttime D region," *Geophysical Research Letters*, vol. 20, no. 21, pp. 2355-2358, 1993.
- [5] U. S. Inan, S. A. Cummer, and R. A. Marshall, "A survey of ELF and VLF research on lightning-ionosphere interactions and causative discharges," *Journal of Geophysical Research*, vol. 115, no. 6, Article ID A00E36, 2010.
- [6] A. J. Smith, R. B. Horne, and N. P. Meredith, "The statistics of natural ELF/VLF waves derived from a long continuous set of ground-based observations at high latitude," *Journal of*

Atmospheric and Solar-Terrestrial Physics, vol. 72, no. 5-6, pp. 463–475, 2010.

- [7] P. F. Velleman and D. C. Hoaglin, *Applications, Basics, and Computing of Exploratory Data Analysis*, Duxbury Press, Boston, Mass, USA, 1981.
- [8] D. C. Hoaglin, F. Mosteller, and J. W. Tukey, *Understanding Robust and Exploratory Data Analysis*, John Wiley & Sons, New York, NY, USA, 1983.
- [9] M. Nakai and W. Ke, “Statistical models for longitudinal data analysis,” *Applied Mathematical Sciences*, vol. 3, pp. 1979–1989, 2009.
- [10] Z. R. Ghahroodi, M. Ganjali, and I. Kazemi, “Models for longitudinal analysis of binary response data for identifying the effects of different treatments on insomnia,” *Applied Mathematical Sciences*, vol. 4, no. 61–64, pp. 3067–3082, 2010.
- [11] E. W. Frees, *Longitudinal and Panel Data: Analysis and Applications in the Social Sciences*, Cambridge University Press, Cambridge, UK, 2004.
- [12] Z. Zhang, K. Lai, Z. Lu, and X. Tong, “Bayesian inference and application of robust growth curve models using student’s t distribution,” *Structural Equation Modeling*, vol. 20, pp. 47–78, 2013.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

