International Journal of Engineering & Technology, 8 (1.4) (2019) 597-605

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

A Comparative Study of Best-Fit Algorithms for the Risk Assessment of Weather Conditions in Electricity Based on Iot

Saraswathi Sivamani[1,], Vasanth Ragu[1], Myeongbae Lee[1], Kyongryong Cho[1], Sungeon Cho[1] Changsun Shin[1] , Jangwoo Park[1] , Yongyun Cho[1*]

[1]Department of Information and Communication Engineering, Sunchon National University, Suncheon, Korea, Of Republic
**Corresponding author E-mail:yycho@sunchon.ac.kr*

Abstract

Three statistical methods, Generalized Additive Model (GAM), Generalized Linear Model (GLM) and Linear Mixed Effects Model (LME) are used to analyze the relationship between the electric pole vibration and the weather conditions. All the models were fitted individually to the respective weather conditions such as temperature, humidity, wind speed and wind direction. All the information from the sensors are processed and analyzed, where the pitch and the roll of the electric pole reveals the influence of the temperature over the respective data. Therefore, the model is fitted with the respect to the weather conditions obtained from different source and platform. In order to fit the model accurately, all three models implemented to pitch and roll, along with the weather conditions. The results show that the best model among the three is Generalized Additive Model, which is identified using AIC value, BIC value and the deviance explained. For more deep understanding and clearance, the residual fit is performed and the model validation is tested for normality using the Kolmogorov-Smirnov normality test. With the best-fit model, the risk assessment becomes more reliable, with either, minor or major causalities.

Keywords: *Generalized Linear Model; Generalized Addictive Model; Linear Mixed Effects Model; Model-fitting.*

1. Introduction

Dependency over the electricity has been growing more intensively, which makes electricity as an essential part of the life. All the services are related with electricity such as lights, televisions, gps navigation, hospitals and other auto industries [1- 4]. These changes has opened the door for utilities to compete with each other and against independent suppliers regardless of their geographic location. Although this change will benefit the consumer, utilities are going to face a highly unpredictable market and will need to make tough decisions regarding power generation and delivery.

Commonly, the improper assessments will result in greater damages in the large corporation [5]. To reduce the uncertainty, the different risk factors need to be found that are related to the electric outages or fault in the electric utility pole. In order to have a good understanding of the risk factors, the utility pole is tested with the natural causes and technical problems, which will provide emergency strategies and decision process.

Electric Utility malfunction caused by weather such as strong winds and cyclone can vary from minor to major losses [6, 7]. It can also lead to the major economic loss to damage of high cost materials. Similarly, the unknown factors that cause an electric fault can also be quite serious with the place, severity and other factors. Therefore, the risk assessment strategy can be more appropriate and effective to increase the efficiency of risk management. In addition to the power consumption process, the safety of the electric pole is supposed to be the major issue in the maintenance, due to the weather conditions. The weather conditions such as temperature and wind can cause elasticity in the conducting wires, resulting in the major impact and influence in the acceleration data of the electric pole. In most cases, instead of the intentional forces, the weather conditions can cause changes in the pitch and roll of the utility pole, with respect to the wind direction. Therefore, it is important to identify and remove the influential factors related with the temperature, to avoid the ambiguity in the model fitting or data analysis.

All the sensor data from the power devices on the electric pole are received and processed. Information from different sources such as weather station and electric pole pass through the multiple platform such as data storage layer, information-processing layer and the middleware, which uses the language platform such as R. In this paper, we have used three models such as Generalized Additive Model (GAM), Generalized Linear Model (GLM) and Linear Mixed Effects Model (LME) to find the best-fit model for the acceleration of the electric utility pole, influenced by the meteorological conditions. GAM model proves to be the best model after analyzing the AIC, BIC, R Squared and Deviance Explained values. To extend the validation of the proposed model, the normality test is conducted. With this best fit model, the risk assessment by external forces such as weather conditions can be identified and increase the safety of the electric utility pole.

.

2. Related Works

The potential damage to the electric utility pole can be due to the strong winds and storms. As Korea is mostly covered with the mountains, the poles in the high altitude places are most vulnerable to the strong winds, which may result in greater damages and outages. There is some previous work on storm-related utility outage predictions. Many researches focus on the forecasting the power outages that related to the high storms such as Zhu et al. [8] and Liu et al. [9, 10]. They discuss the risk and forecasting strategies of the power outages with various wind speeds. Liu et al. [11] studies explain the post disaster infrastructure and the system restoration time for the power outages, with respect to the storms. Similarly, the other natural disasters are cyclones and hurricanes are used in the studies to predict the power outages, in which Guikema et al [12] explains the risk analysis and the prior estimation of the damage caused over the electric power outages. Similarly, Ouyang et al [13] also explains the hurricane analysis in a multi-dimensional view. With the provided analysis, the power outage and the electric system resiliency are determined [14]. The study that surrounds the prediction of power outages are mostly related with the regression analysis. Madanat et al [15] explains the power outages causes relating to the infrastructure probabilities along with Poisson regression. Similarly, Domijan et al. [16] also uses Poisson regression to outnumber the power outage with the significance of the weather.

Apart from the weather conditions, the unknown conditions are also considered to study the power outages. The uncertainty of the unknown faults is identified by using Monte Carlo along with the Poisson regression is calculated by Zhou et al [17]. There are also more studies related the different probabilities to study the unknown fault detection [18, 19]. However, the models do not predict the data considering the uncertain data that might result in the providing valuable results. Although, there are many studies that concerns the weather conditions and the respective forecast models. Identification of the influential external forces on the observed data is neglected. In this paper, we analyze the influence of the external forces over the electric utility pole, for which the best fit model is identified. GAM, GLM and LME models are compared with each other, to find the best fitting model for the acceleration data.

3. Data and Methods

One-month electric utility pole data of the southern district obtained from the Korean Electric Power Cooperation(KEPCO) has been used for the analysis process. The observed data from the electric utility pole contains the Pitch and Roll data received from the gyroscope sensor. The analysis is based on data collected from accelerometers installed in the utility pole from the southern region of South Korea, Daegu. The collected data ranges from September 1, 2016 to September 30, 2016, where the sensor node measures acceleration, temperature and humidity from different devices installed on the utility pole. The five devices are transformer, load switch, load balance, utility pole, and communication enclosure. As our main of the analysis lies on the risk management of the utility pole, the acceleration data is analyzed. The weather data for the same month was obtained from the Korea Meteorological Agency (KMA), which contains the temperature, humidity, air pressure, vapor pressure, rain precipitation, radiation, wind speed, wind direction, cloud capacity and daylight hours.

3.1. Generalized Linear Model

To Generalized linear model (GLIM or GLM) was first popularized by McCullagh and Nelder [20, 21], which indicates a large class of models, in which the response variable follows an exponential family distribution along with GLM allows the linear model to relate to the response variables through the link function. Generalized linear models generalize the possible distributions of the residuals to the exponential family that includes all the distribution such as the binomial, Poisson, negative binomial, and gamma distributions as explained in the equation 1.

$$E(Y) = \quad \mu = \quad g^{-1}(X\beta) \tag{1}$$

Where, $E(Y)$ is the expected value of the *Y*, $X\beta$ is the linear predictor, and the g is the link function.

3.2. Generalized Additive Model

The Generalized Additive Model is similar to Generalized Linear Model, where the additive model is applied to the GLM properties. The spline function is used to fit the model with the linear function [22]. The response variables and the covariance are expressed by the following probability model as represented in the below equation 2.

$$\mu = \quad E(Y|x_1, \dots \quad x_\propto) \tag{2}$$

The link function indicates the link structure with the explanatory variable, and the link function structure for the predicted value is expressed as (3).

$$\eta = g(\mu) = s_0 + \sum_{i=1}^{p} s_j(x_j) \tag{3}$$

Here, the estimate is calculated by weighting the response variables calibrated by the link function. In the GAM, a non - parametric Smoothing Spline Function is used to describe the nonlinear relationship. In the GAM, it is necessary to estimate the smoothed function, which represents the legal effect.

$$s_k(x_k) = E[Y - s_0 - \sum_{k \neq i} s_j(x_j)|x_k] \tag{4}$$

The equation 4 is the conditional expected value of the additive model and is estimated by the back fitting algorithm

3.3. Linear Mixed Effects

Linear mixed-effects models [23] are extensions of linear regression models that are collected and grouped. These models explain the relationship between a response variable and independent variables that vary with one or more grouping variables. There are two parts in the mixed-effects model, fixed effects and random effects. Fixed-effects terms refer the conventional linear regression part, and the random effects are mostly drawn at random from a population. The common random effects are observations relates by associating group of data that have the same level of a grouping variable. The standard form of a linear mixed-effects model [24] is shown in the below equation 5.

$$y = \underbrace{X\beta}_{fixed} + \underbrace{Zb}_{random} + \underbrace{\varepsilon}_{error} \quad (5)$$

Where y is the n-by-1 response vector, and n is the number of observations, X is an n-by-p fixed-effects design matrix, β is a p-by-1 fixed-effects vector, Z is an n-by-q random-effects design matrix., b is a q-by-1 random-effects vector, ε is the n-by-1 observation error vector

4. Result and Discussion

The data collected from the 3-axis gyroscope, which is converted into a pitch angle and a roll angle. Although the data are perfect on its own, we need to filter and smooth data to increase the performance analysis. Based on the collected data, filtering was performed based on the parameters calculated through maximum likelihood estimation (MLE). Figure 1 shows the filtered data of pitch and roll and the figure 2 shows the correlation between the pitch and roll for different devices.

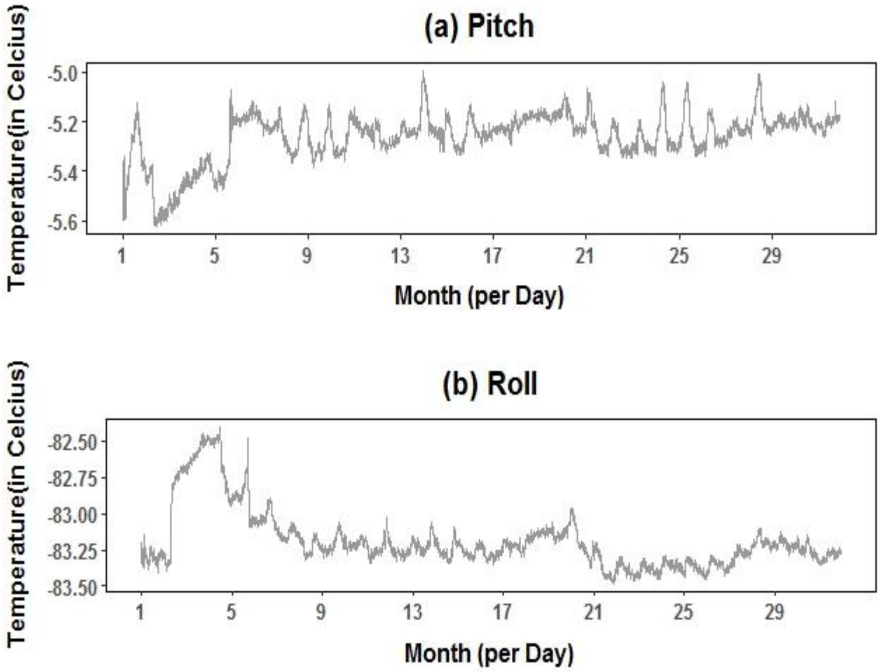

Fig 1: Filtered acceleration data of pitch and roll for one day

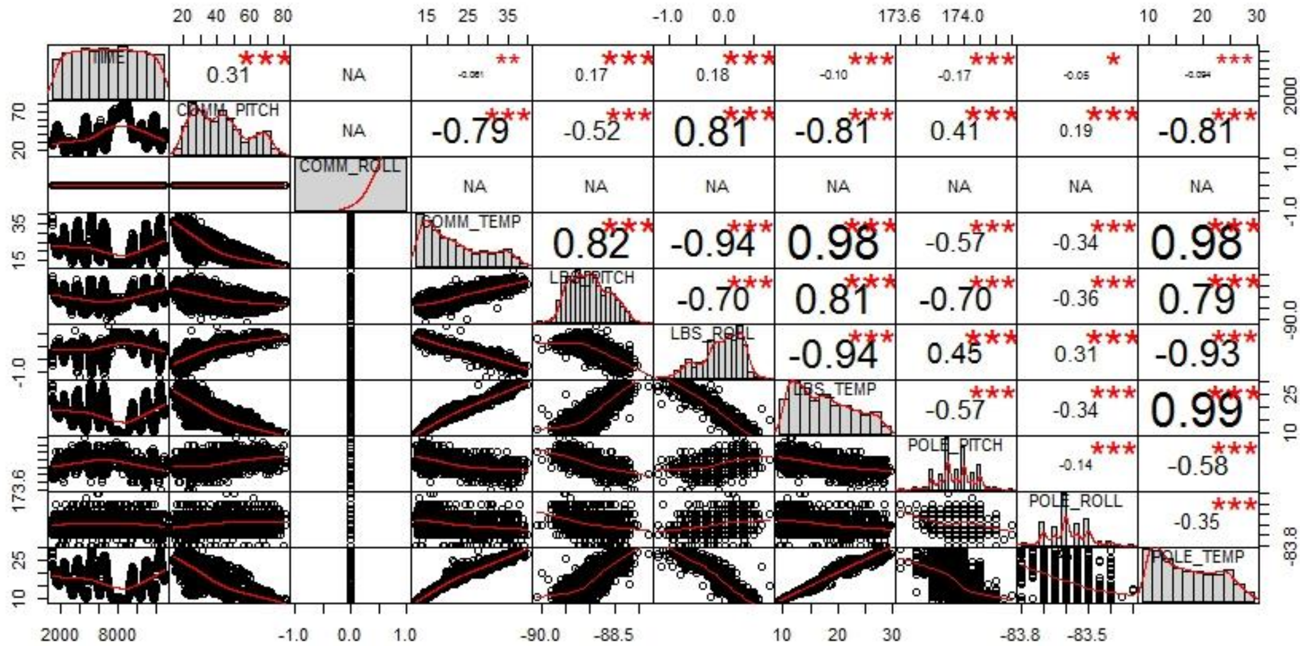

Fig 2: Correlation plot for the pitch and roll of different devices

Most of the electric utility pole risks are based on the weather conditions. Therefore, we have used the weather data from the Korea Meteorological Agency (KMA). The obtained weather data consists of atmospheric pressure, temperature, wind direction, wind speed, relative humidity, precipitation, precipitation, solar radiation, sunshine time, ground temperature, vertical temperature, ground temperature, soil moisture, groundwater level and visibility. As the meteorological conditions have most influence and correlation factors over the acceleration data, it is important to select the importance explanatory variables for the modelling. Multicollinearity means the reliability of the model coefficients, due to the high correlation between the explanatory variables [25]. So, the multicollinearity needs to be removed to find a best fit model. The linear model is fitted with the temperature and the other data in the meteorological observation.

Table 1 shows the result of the linear fit model that contains the t-value and p-value. It is noted that the high correlation occurs between the Humidity, precipitation, vapor pressure, atmospheric pressure, dew point temperature, daylight hours, and cloudiness. In order to solve the problem of overfitting, we have used Variation inflation factor (VIF) [26] to remove the mutlicollinearity for the temperature, which is highly correlated with the pitch and roll of the utility pole.

Table 1: Summary of the linear model for weather data.

Weather Data	t-value	p-value
Humidity	-128.762	$< 0.02*10^{-14}$
Rain Precipitation	3.174	0.0015
Wind Speed	-1.340	0.1860
Wind Direction	-0.032	0.7623
Vapor Pressure	3.630	0.0003
Air Pressure	-10.761	$< 0.02*10^{-14}$
Rain Dew Point	18.515	$< 0.02*10^{-14}$
Hours of Daylight	6.436	$< 0.02*10^{-8}$
Solar Radiation	0.178	0.8591
Cloud Cap	3.335	0.0008

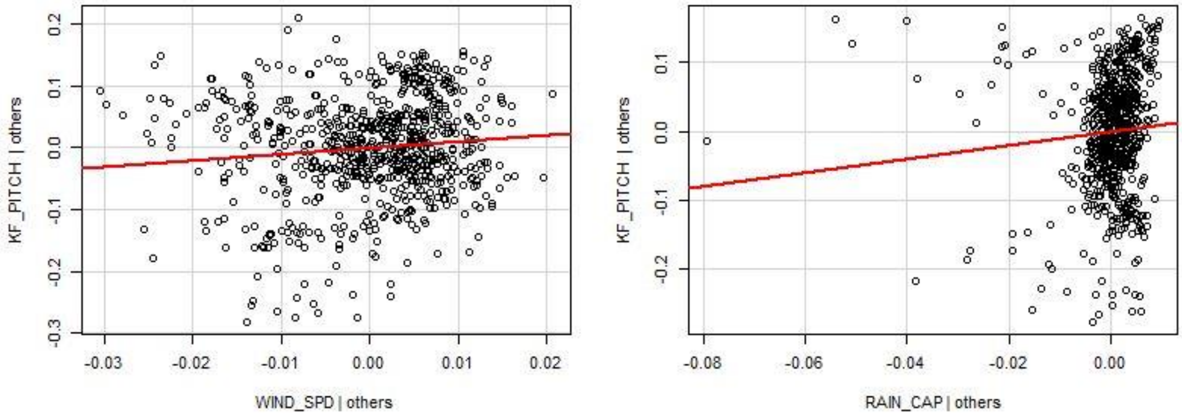

Fig 3: Leverage plot for Wind speed and Rain Precipitation

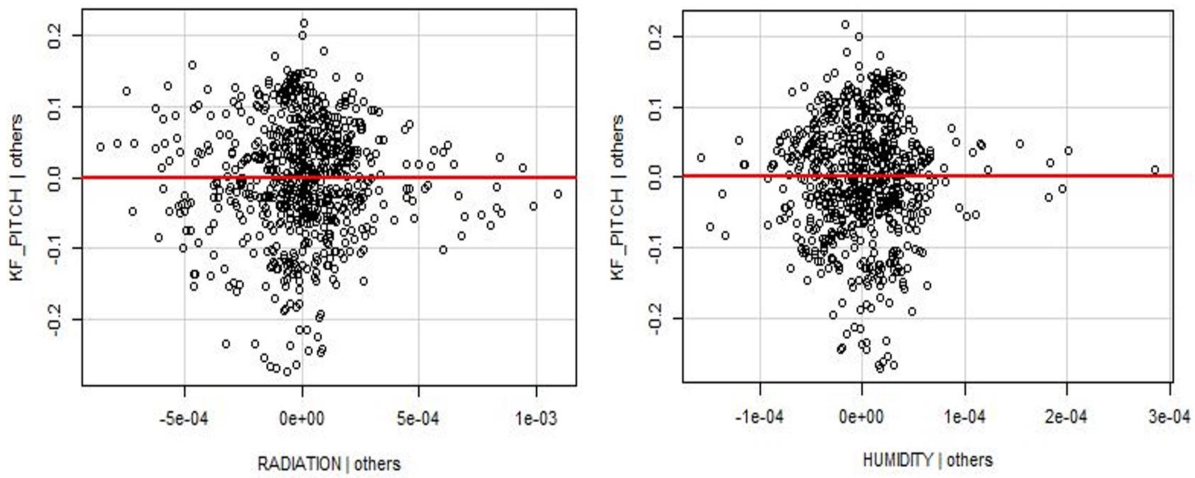

Fig 4: Leverage plot of Radiation and Humidity

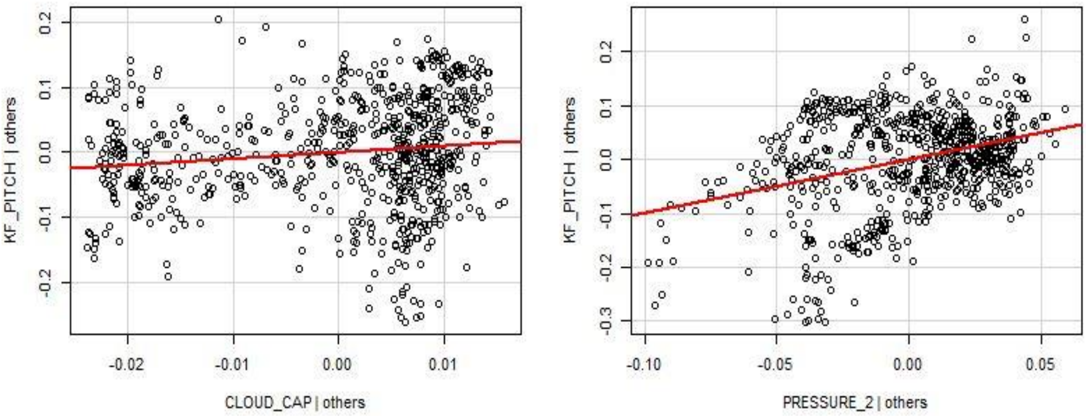

Fig 5: Leverage plot of Cloud and Pressure

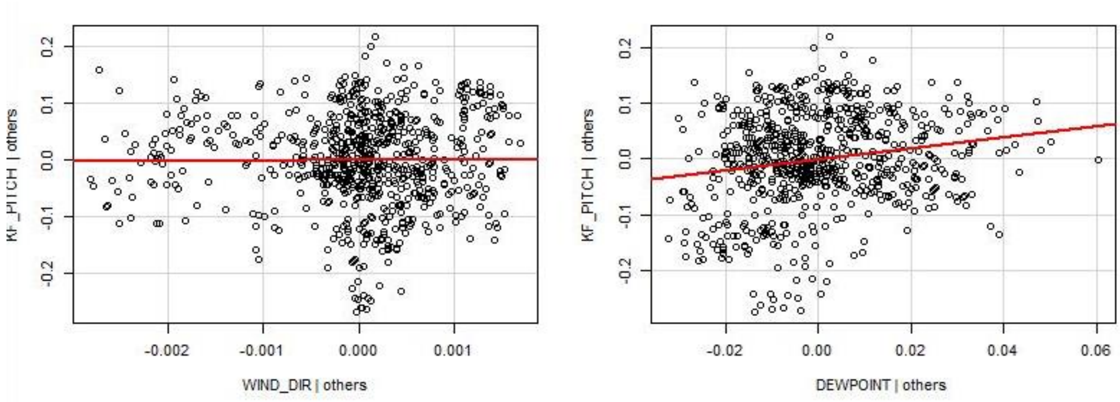

Fig 6: Leverage plot of Wind Direction and Dew point

Table 2: Summary of VIF data before and after Multicollinerity.

Weather Data	VIF (All data)	VIF (Multicollinerity data removed)
Temperature	105.8432	2.4846
Humidity	103.8234	-
Rain Precipitation	1.3128	1.2621
Wind Speed	1.3413	1.3199
Wind Direction	1.0884	1.0813
Vapor Pressure	98.5901	-
Air Pressure	1.6244	1.3999
Rain Dew Point	138.73806	1.9488
Hours of Daylight	4.4760	1.7927
Solar Radiation	16.8662	-
Cloud Cap	1.4226	1.1702

Temperature, humidity, dew point temperature, and irradiation. Factor verification was performed repeatedly except for variables with high correlation except for temperature. We selected the explanatory variables of temperature, precipitation, wind speed, wind direction, atmospheric pressure, dew point temperature, solar radiation, and cloud at 2.5 VIF upper limit value. Table 2 shows the result of the verification of the dispersion expansion factor (VIF). Linear model with response variables as pitch. As a result of adaptation, weakness factors have an explanatory power of 37% with respect to pitch. The linearity of the explanatory variables are explained in detailed using the leverage plots

Table 3: Comparison test data between three models

Model	AIC	BIC	R-Squared	Deviance Explained
Generalized Linear Model (GLM)	-1513.136	-1471.91	0.28637	28.6%
Generalized Additive Model (GAM)	-1807.917	-1629.721	0.54	56.4%
Linear Mixed Effects (LME)	-1.785.811	-1617.403	0.524	54.07%

The figure 2 shows the leverage plot of explanatory variables, wind speed and rain precipitation. It indicates the non-linearity similar to the plot of cloud capacity and air pressure as shown in the figure 4. The leverage plot of the dew point in the figure 5, where the plot of the wind direction shows linearity. Similarly, the leverage plot of the humidity and the radiation shows linearity with the straight line at zero. The linearity is exactly 0, indicating the high correlation between the variables.

After removing the multicollinearity, the model fitting is performed. It is clear that the model is a linear model, but to find the proper fit, we have compared the three models such as Generalized Linear Model, Generalized Additive Model and Linear Mixed Effects Model, due to the size of the dataset and linearity. Even though, the entire three models are suitable for the linear model, three models are compared to find the best-fit model. The explanatory variables used to fit the model include Temperature, Wind Speed, Wind direction, Rain Precipitation, Air Pressure, Rain Dew point and Cloud Capacity. As shown in the table 3, the Akaike Information Criterion (AIC), Bayesian Information Criterion

(BIC), R-Squared value and the Deviance Explained values are compared between the three values.

AIC measures the relative quality of the models for the better selection from the given set of data. GAM shows the best fit with less AIC value, while the GAM and LME model shows a slightly high value. Similarly, the Bayesian Information Criterion also helps to select the model by estimating the log likelihood function. As same as AIC values, the BIC value also shows GAM model with less BIC value. The R-squared values measures the closeness of the fitted regression line, which indicates higher the value, better the model fit. The Deviance explained is close to the R-squared value represented in the percentage value. Both the R-squared and the deviance explained give the high value towards the GAM model. Among the three models, Generalized Additive Model seems like perfect fit model for the pitch and the meteorological conditions.

Hence, in this paper, we use a generalization method model with weather factors as explanatory variables. The model is assumed normal distribution for the explanatory variables and connection functions, where the significant explanatory variables selected for consideration of multicollinearity are temperature, precipitation, wind speed, wind direction, air pressure, dew point, radiation, and cloud cover giving the below equation 6.

$$g(\mu) = s_0 + s_1(\text{Temperature}) + s_2(\text{Precipitation}) + s_3(\text{Wind Speed}) + s_4(\text{Wind direction}) + s_5(\text{Air Pressure}) + s_6(\text{DewPoint}) + s_7(\text{Daylight}) + s_8(\text{Cloud Capacity}) \quad (6)$$

Figure 7 shows the results of temperature. The temperature is zero at 23 ° C. The influence is negative, if less than 23 ° C, and positive if it is greater than 23 ° C. Figure 8 shows the results of the rain precipitation. If the precipitation is 0, there is no effect, and the negative influence is overall.

As shown in the figure 9, in the case of wind velocity, there is no effect at 0, and overall positive influence. In the case of wind direction, the degree of influence changes around 160 ° is shown in the figure 10.

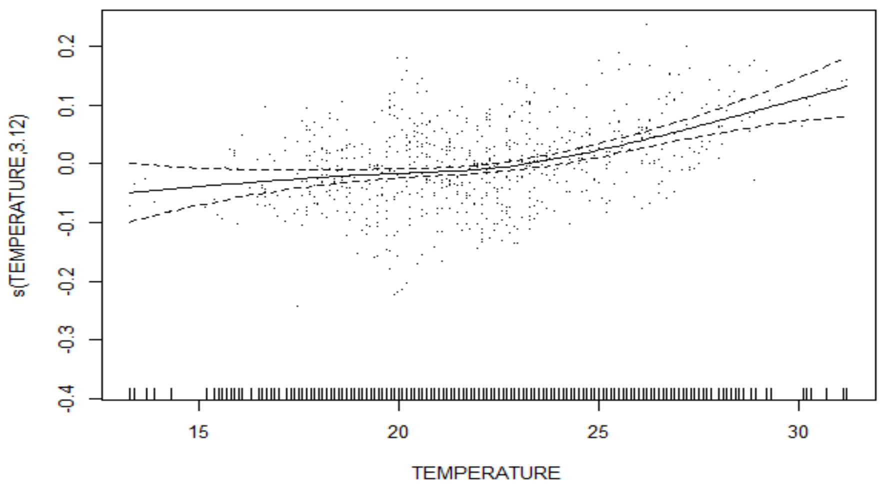

Fig 7: Plot for Temperature in GAM

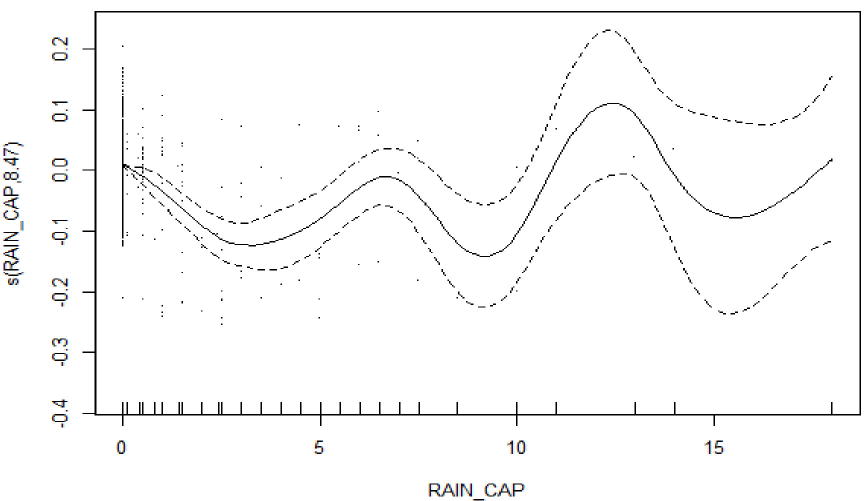

Fig 8: Plot for Rain Precipitation in GAM

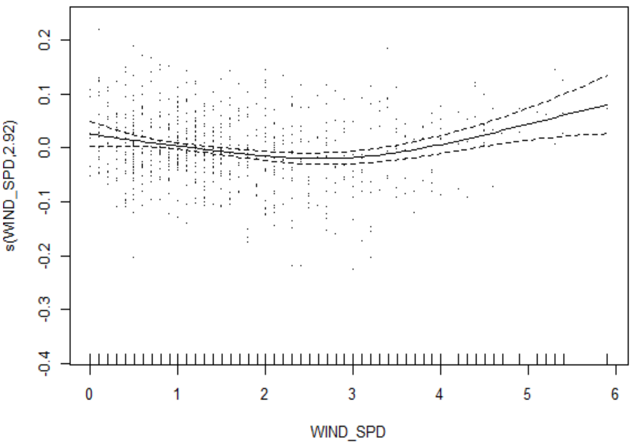

Fig 9: Plot for Wind speed in GAM

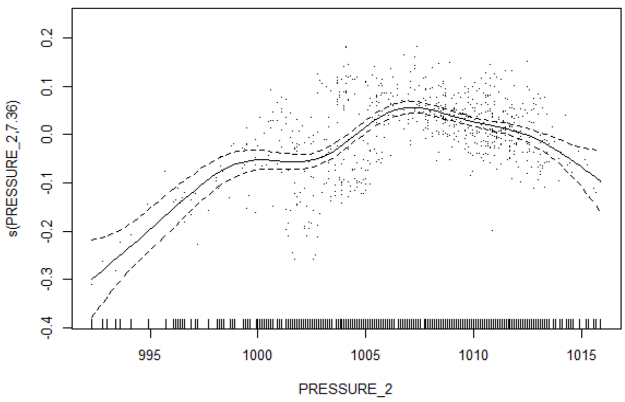

Fig 11: Plot for Air pressure in GAM

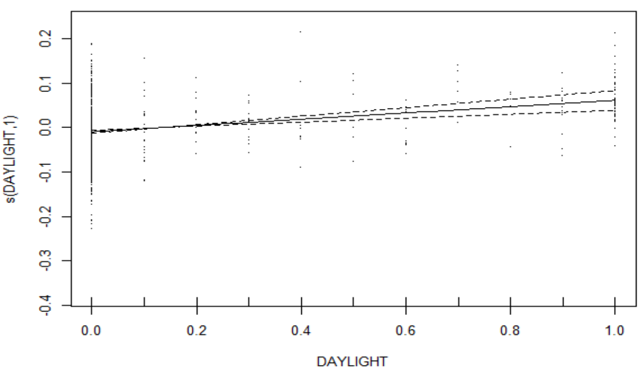

Fig 13 : Plot for Daylight in GAM

Figure 11 shows the results of air pressure and precipitation. There is no effect at 1,001hpa and 1,008hpa, and negative effects are observed when the result is less than 1,001hpa and 1,008hpa. If the dew point temperature is less than 17, it indicates the negative influence as shown in the figure 12.Figure 13 and 14 shows the plots of daylight hours and cloudiness respectively. In the case of daylight hours, no influence is observed at 0.1 hr. The parameter

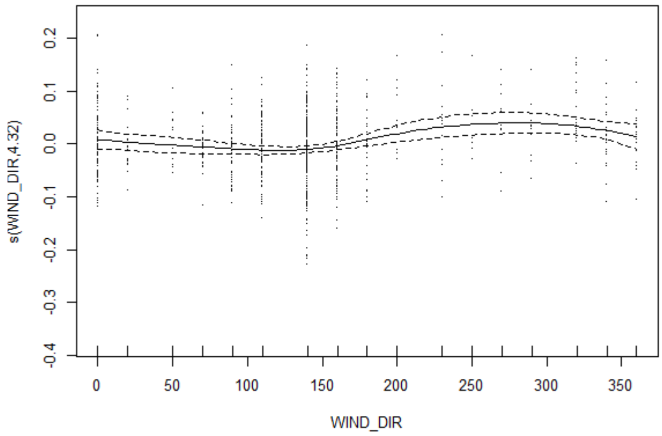

Fig 10: Plot for Wind Direction in GAM

of the estimated smoothing function is "1" and the linearity is high.

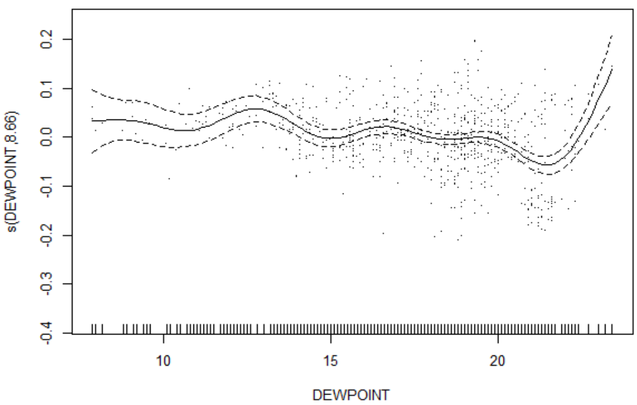

Fig 12: Plot for Dew point in GAM

In the case of cloudiness, the influence of sound and the amount of

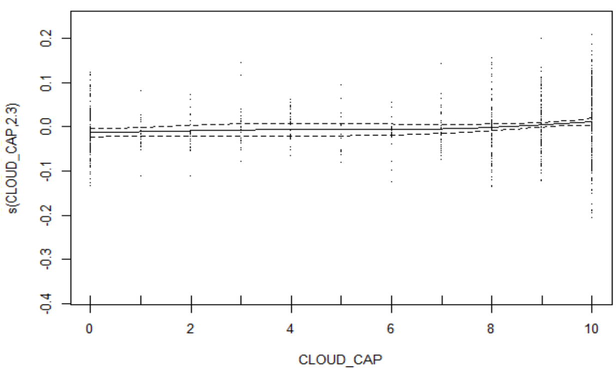

Fig 14: Plot of Cloud capacity in GAM

influence are shown based on the equation. With the definite values, the extracted residuals are obtained from the GAM model and the values are fitted for the smoother and refined model. The explanatory variables used at 5% significance level and significance level 0.05 show very significant results. The effect of atmospheric pressure, daylight time, and cloudiness was very high among the explanatory variables.

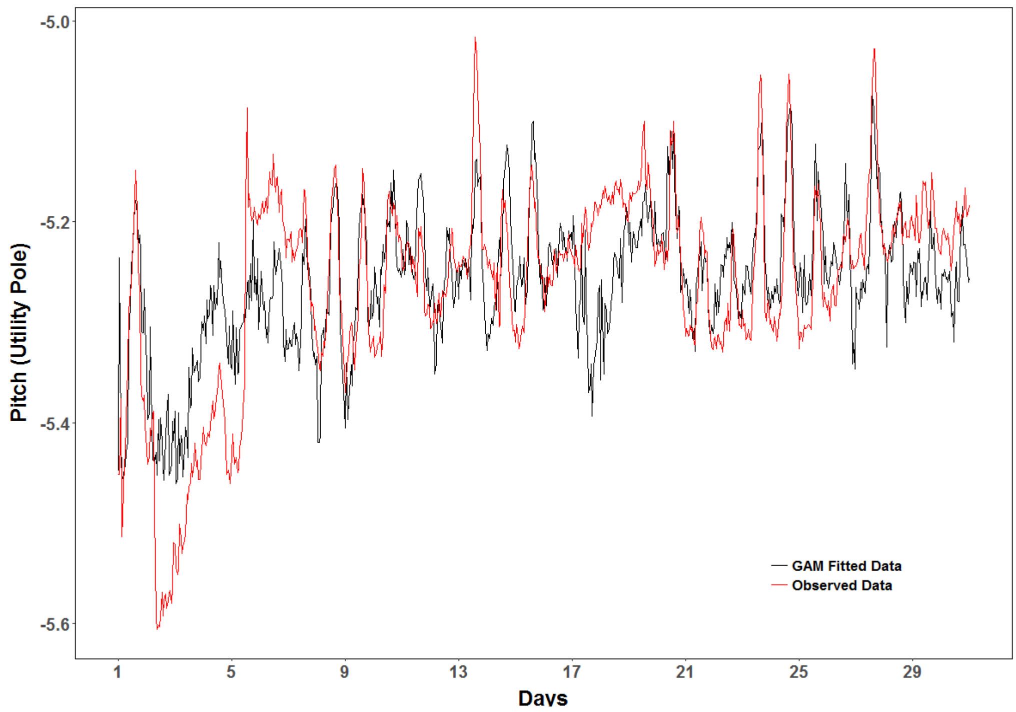

Fig 15: Plot of the Observed and fitted values for the acceleration of pitch

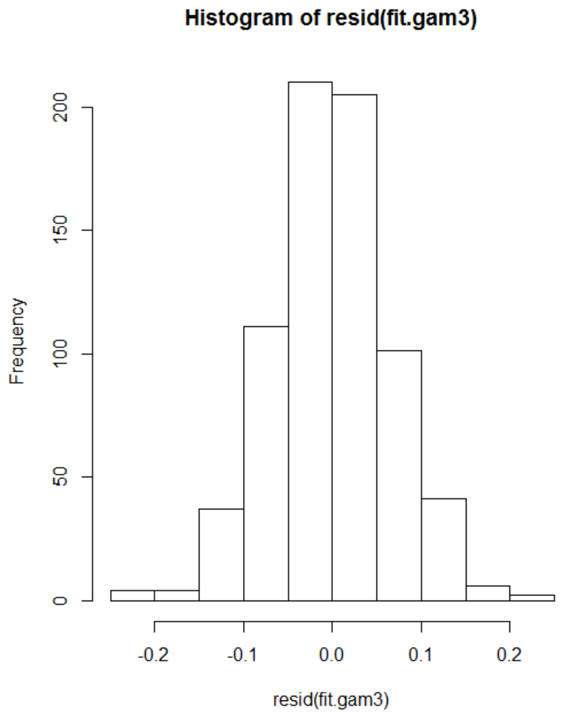

Fig 16: Histogram for the residual in GAM

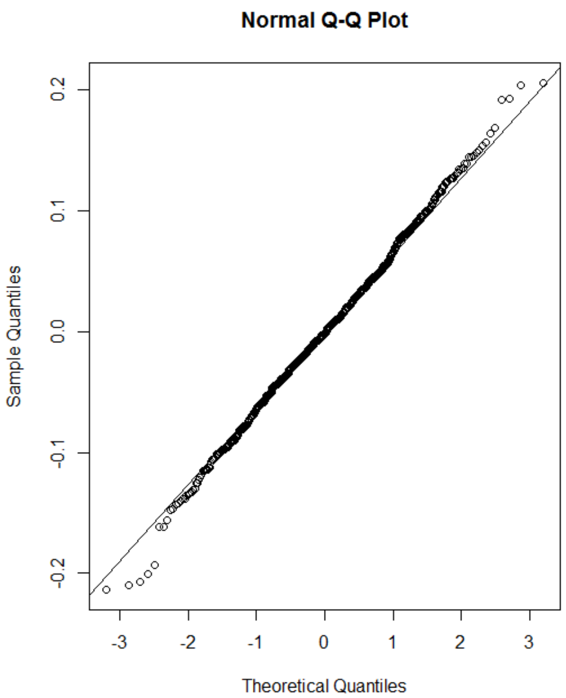

Fig 17: Histogram and Q-Qplot for the residual in GAM

Figure 15 shows the plot of observed and the fitted values for the acceleration pitch. The validation of the fitted models are analyzed through the residuals. For the better quality analysis of the model, firstly, the distribution check assures the normal distribution on the model and the Q-Q Plot verifies the normality of the distribution.

As shown in the figure 16 and 17, the histogram and Q-Q plot proves the normality of the model

Table 4: Result of Kolmogorav-Smirnov Model

Type	Value
D-value	0. 032333
p-value	0.4381

In addition to this, Kolmogorov-Smirnov normality test [27] is performed to find the goodness-of-fit and the results are tabulated in table 4. The p-value is 0.4381, which is less than 0.5, proving the normality of the model. Many studies prove the test to be effective in the analysis of the histogram, with the help of p-value [28]. Referring to the model and test, the risk assessment can be identified with any weather related conditions or natural disasters such as hurricane and wind storms.

5. Conclusion

In this paper, we have compared the three models Generalized Additive Model (GAM), Generalized Linear Model (GLM) and Linear Mixed-Effects Model (LME) for fitting the acceleration data with the weather conditions. In order to compare the models, the observed data was filtered using Kalman filter. Model overfitting over the weather conditions such as Humidity, Vapor Pressure and Radiation was reduced using the Variation inflation factor (VIF). By comparing the three models, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), R-Squared and Deviance Explained values for GAM were obtained as -1807.917, -1629.721, 0.54 and 56.4% respectively, resulting in the best model. Although, the Histogram and Q-Qplot of the residual shows the fitness of the model, the goodness of fit has been verified using the Kolmogorov-Smirnov Tests with the D-value of 0. 032333 and p-value of 0.4381. The test validates the normal distribution towards the null hypothesis. With the fitted GAM model for acceleration data, the risk assessment for the electric utility pole related with the weather conditions can be identified with the residual values. For the future work, we are planning to implement a user interface for monitoring the devices, to identify the risk through the alert messages.

Acknowledgement

This work was supported by the Korea Institute of Energy Technology Evaluation and Planning(KETEP) and the Ministry of Trade, Industry & Energy(MOTIE) of the Republic of Korea (No. 20172010000730).

References

[1] H. Fraser, "The importance of an active demand side in the electricity industry," The Electricity Journal, vol.14, no.9, pp.52-73, 2001.
[2] J. Stern, "Electricity and telecommunications regulatory institutions in small and developing countries," Utilities Policy, vol.9, no.3, pp.131-157, 2000.
[3] H.D.Kutzbach, 2000. "Trends in power and machinery," Journal of Agricultural Engineering Research, vol.7, no.3, pp.237-247, 2000.
[4] A.S. Szklo, J.B.Soares, and M.T.Tolmasquim, "Energy consumption indicators and CHP technical potential in the Brazilian hospital sector," Energy Conversion and Management, vol.45, no.13, pp.2075-2091, 2004..
[5] K. Kim, and Y.Cho, "Estimation of power outage costs in the industrial sector of South Korea," Energy Policy, vol.101, pp.236-245, 2017.
[6] W. Li, Risk assessment of power systems: models, methods, and applications, 2nd ed. John Wiley & Sons, 2014.
[7] Y. Wang, C. Chen, J. Wang, and R. Baldick, "Research on resilience of power systems under natural disasters—A review," IEEE Transactions on Power Systems, vol.31, no.2, pp.1604-1613, 2016.
[8] D. Zhu, D. Cheng, R.P. Broadwater, and C. Scirbona, "Storm modeling for prediction of power distribution system outages," Electric power systems research, vol.77, no.8, pp.973-979, 2007.
[9] H. Liu, R.A. Davidson, D.V. Rosowsky, and J.R. Stedinger, "Negative binomial regression of electric power outages in hurricanes" Journal of infrastructure systems, vol.11, no.4, pp.258-267, 2005.
[10] H. Liu, R.A. Davidson, and T.V. Apanasovich, "Spatial generalized linear mixed models of electric power outages due to hurricanes and ice storms," Reliability Engineering & System Safety, vol.93, no.6, pp.897-912, 2008
[11] H. Liu, R.A. Davidson and T.V. Apanasovich, "Statistical forecasting of electric power restoration times in hurricanes and ice storms," IEEE Transactions on Power Systems, vol.22, no.4, pp.2270-2279, 2007.
[12] S.D. Guikema, S.M Quiring, and S.R. Han, "Prestorm estimation of hurricane damage to electric power distribution systems," Risk analysis, vol.30, no.12, pp.1744-1752, 2010.
[13] M. Ouyang, and L. Dueñas-Osorio, "Multi-dimensional hurricane resilience assessment of electric power systems," Structural Safety, vol. 48, pp.15-24, 2014.
[14] R.J. Campbell, "Weather-related power outages and electric system resiliency,". Washington, DC: Congressional Research Service, Library of Congress, Aug. 2012.
[15] S. Madanat, and W.H.W. Ibrahim, "Poisson regression models of infrastructure transition probabilities," Journal of Transportation Engineering, vol.121, no.3, pp.267-272, 1995.
[16] A. Domijan Jr, R.K. Matavalam, A. Montenegro, W.S. Wilcox, Y.S. Joo, L. Delforn, J.R. Diaz, L. Davis and J.D. Agostini, "Effects of norman weather conditions on interuptions in distribution systems," International journal of power & energy systems, vol.25, no.1, pp.54-61, 2005.
[17] Y. Zhou, A. Pahwa, and S.S. Yang, "Modeling weather-related failures of overhead distribution lines," IEEE Transactions on Power Systems, vol.21, no.4, pp.1683-1690, 2006.
[18] D. K. Mohanta, P.K. Sadhu and R. Chakrabarti, "Fuzzy reliability evaluation of captive power plant maintenance scheduling incorporating uncertain forced outage rate and load representation," Electric Power Systems Research, vol.72, no.1, pp.73-84, 2004.
[19] [19]J. S. Simonoff, C.E. Restrepo and R. Zimmerman, "Risk-Management and Risk-Analysis-Based Decision Tools for Attacks on Electric Power," Risk Analysis, vol. 27, no.3, pp.547-570, 2007.
[20] P. McCullagh and J. A. Nelder, Generalized Linear Models, Monograph on Statistics and Applied Probability, 1989.
[21] P. McCullagh, "Generalized linear models," European Journal of Operational Research, vol.16, no.3, pp.285-292, 1984.
[22] T.J. Hastie, and R.J. Tibshirani, Generalized additive models, John Wiley & Sons, 1990.
[23] A.L. Oberg and D. W. Mahoney, "Linear mixed effects models," Methods in Molecular Biology™, Humana Press, vol. 404, pp.213-234, 2007.
[24] "Linear Mixed-Effects Models," https://in.mathworks.com/help/stats/linear-mixed-effects-models.html?s_tid=gn_loc_drop
[25] H.M. Blalock, "Correlated independent variables: The problem of multicollinearity," Social Forces, vol.42, no.2, pp.233-237, 1963.
[26] C. Robinson and P.E. Schumacker, "Interaction effects: centering, variance inflation factor, and interpretation issues," Multiple Linear Regression Viewpoints, vol.35, no.1, pp.6-11, 2009.
[27] R. Wilcox, "Kolmogorov–smirnov test," Encyclopedia of biostatistics, 2005.
[28] I.T. Young, "Proof without prejudice: use of the Kolmogorov-Smirnov test for the analysis of histograms from flow systems and other sources," Journal of Histochemistry & Cytochemistry, vol.25, no.7, pp.935-941, 1977.