Prediction of Salt Contamination on High Voltage Insulators in Rainy Season Using Regression Technique

Ahmad S. Ahmad, Hussein Ahmad, Md. Abdus Salam, T. Tamsir, Z. Buntat, M. W. Mustafa

Faculty of Electrical Engineering
Universiti Teknologi Malaysia
81310 Skudai – Johor
Malaysia
E-mail: ahmads@ieee.org

Abstract: The severity of contamination on the high voltage insulator surfaces is the significant factor in determining the level of outdoor insulation and in choosing the types of insulators. In the equatorial region, the most dangerous kind of contamination is salt contamination. In this paper, regression technique has been used to develop a modified Equivalent Salt Deposit Density (ESDD) mathematical model with respect to meteorological conditions. This model provides a useful way for predicting contamination level and for determining frequency of washing the insulators in a given contaminated area.

Keywords
ESDD, insulator, contamination, regression analysis

I. INTRODUCTION

The wet salts form a conducting layer on the surfaces. This layer provides an ideal path for the leakage current to flow from the high tension side to the ground side of the insulators. If the current flow reaches a critical stage flashover occurs. Since the meteorological parameters play a big role in contamination accumulation, it should be considered in the mathematical model. So far, the researchers have studied the effect of the meteorological parameters on ESDD or flashover in piecemeal manner. It has been found that some factors such as temperature and pH can play a significant role on salt solubility [1]. Reference [2] described that the contamination layer depends upon the wind speed and the distance between the seashore and the insulator; a mathematical relationship has been developed between these factors. Sometimes non-uniform contamination due to the unidirectional wind can be found on the surfaces of high voltage insulators as described in [3]. High temperature is responsible for a lower relative humidity which leads to lower flashover voltages; the relationship is described in [4, 5]. As regards the air pressure, it has been found that the flashover voltages of polluted insulators decrease as the pressure decreases [6, 7].

This paper proposes a new relationship of ESDD with six meteorological variables considered simultaneously. The variables are temperature, humidity, pressure, rainfall, wind speed and wind direction. Multiple linear regression technique has been used to predict ESDD from these six variables.

II. TEST LOCATION AND METHODOLOGY

Paka Thermal Power Station one of the power plants in the eastern region of Peninsular Malaysia which is suffering from the rapid build-up of salt contamination, is chosen for the study. Daily washed samples of glass insulators were installed on a special rig about 50 meters from the seacoast. The contamination collection process is carried out daily for two months period with 60 tests for determining ESDD. The first part of the recorded data has been used to develop the model. The second part is used to test the model. The data is collected during the rainy season. The relevant meteorological parameters such as ambient temperature, relative humidity, quantity of rainfall, pressure, wind speed and wind direction were measured at the power station corresponding to the period of ESDD measurements.

III. CLIMATE OF PENINSULAR MALAYSIA

The characteristic features of the climate of Malaysia are uniform temperature throughout the year, high relative humidity within 60% to 90% varying from place to place and from month to month. Excessive rainfall especially in the eastern region of the peninsula and during the rainy season. Winds are generally light [8]. It should be mentioned that in the test area, three obvious wind directions are recorded during the month from August to February: southeastern, eastern and northeastern. The normal direction of the wind during the most of the year months is southeast (SE), but this direction will change gradually from the month of October with the starting of rainy season. In January, the prevalent direction will be the northeast (NE). Fig. 1 shows the change of wind direction. Observations reveal that the wind speed increases in the rainy season causing turbulence in sea waves which can affect the high voltage insulators due to increased salt deposition.

0-7803-6355-8/00/$10.00©2000 IEEE

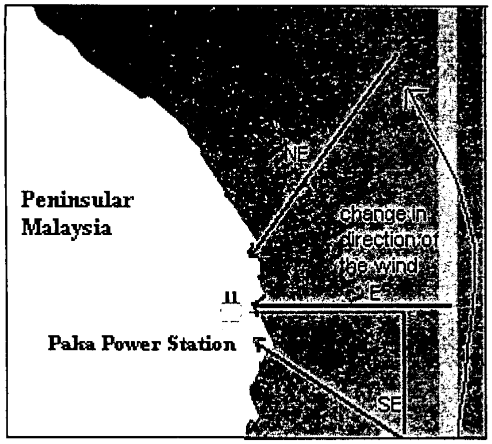

Fig. 1: The change in wind direction at Paka Power Station

IV. SIX PREDICTORS FOR THE MODEL

In the regression analysis, the ESDD is the dependent variable while the other parameters are the driving variables (independent). In our model there are six independent variables have been considered, temperature (X_1), humidity (X_2), pressure (X_3), rainfall (X_4), wind speed (X_5) and wind direction (X_6). The multiple regression equation for this case is written:

$$y = B_o + B_1 X_1 + B_2 X_2 + B_3 X_3 + B_4 X_4 + B_5 X_5 + B_6 X_6$$

and compute values for the B's in a way that satisfies the least squares criterion. The value obtain for B_o represents the predicted value of y when X_1, X_2, ...X_6 remain constant. Similarly, the value of B_2 gives us the predicted rate of response in y to the change in X_2 if X_1, X_3, ...X_6 remain constant, etc.. Figures 2-6 describe the correlation between ESDD and the meteorological parameters in piecewise method.

V. MODEL COEFFICIENTS

A. Coefficient of Determination *(R^2)*

The coefficient R^2 can be defined and interpreted as:

$$R^2 = 1 - \frac{\sum (y_i - \tilde{y})^2}{\sum (y_i - \bar{y})^2} \quad \ldots\ldots(1)$$

The coefficient R^2 has a range between 0 and 1. When the model fits the data well, it is clear that the value of R^2 is close to unity [9]. On the other hand, if there is no relationship between the independent variables and the dependent variable and the linear model gives a poor fit, the best predicted value for an observation y_i would be $\bar{y}$. In the absence of any relationship, the best estimate is the sample mean, for in that case the sample mean minimizes the sum of squared deviations. So in the absence of any linear relationship, R^2 will be near zero. The value of R^2 is therefore used as a summary measure to judge the fit of the linearity for our model. $R^2 = 0.807$ for the developed model, that is, 80% of the variability in the data is accounted for the model. The statistic R^2 should be used with caution since it is always possible to make R^2 unity by simply adding enough terms to the model. Also, R^2 will increase if we add a variable to the model, but it does not necessarily mean the new model is superior to the old one.

B. Multiple Correlation Coefficient *(R)*

The correlation coefficient cannot take values outside the units −1 and +1. A high absolute value of R indicates a close relationship, and a small value, a less relationship. When the absolute value of R is unity, the points fall exactly on a straight line and the relationship is perfect. When $R = 0$, the points scatter in all directions, and the varieties are linearity independent. The sign of R is the same as the sign of the covariance [10]. In our model, $R = 0.899$. From this, we can conclude that although the six variables tend to increase and decrease together, the positive correlation between the X's and y was especially strong and significant, and we can conclude that the sample of observations was drawn from a population in which a positive correlation existed.

C. Adjustment of R^2

The coefficient R^2 measures the proportion of the variation in y, which is explained by the predictor variables. If the y values were entirely in the space of n dimensions the K predictors variables would still explain some variation in y, on average $K/(n\text{-}1)$. This is then scaled to give a value of 1 (perfect explanation of y) when $R^2 = 1$.

$$Adj.\ R^2 = \frac{\left[R^2 - k/_{n-1}\right](n-1)}{(n-k-1)} \quad \ldots\ldots(2)$$

This adjusted value could even be negative if R^2 is small, which highlights the only problem with it. On the other hand the unadjusted value has a clear interpretation as the proportion of the variance of y explained by the predictor variable. In our model the adjusted value for $R^2 = 0.757$, which means that 76% of variance in y is explained by our variables.

VI. UNSTANDARDIZED AND STANDARDIZED MODEL

The multiple linear regression equation can be written base on the dependent and independent variables considering that all independent variables have positive response with the dependent variable Y as:

$$Y = B_o + B_1X_1 + B_2X_2 + B_3X_3 \ldots\ldots B_6X_6 \quad \ldots\ldots(3)$$

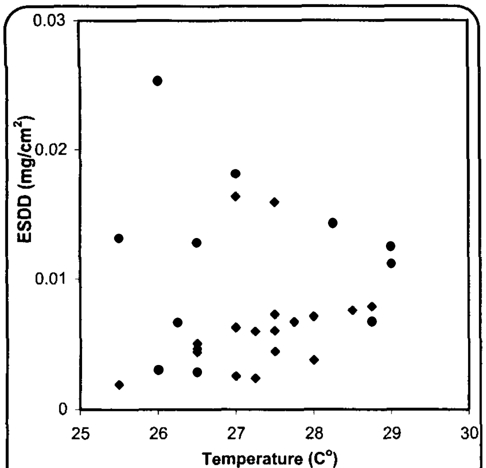

Fig. 2: ESDD vs. Temperature

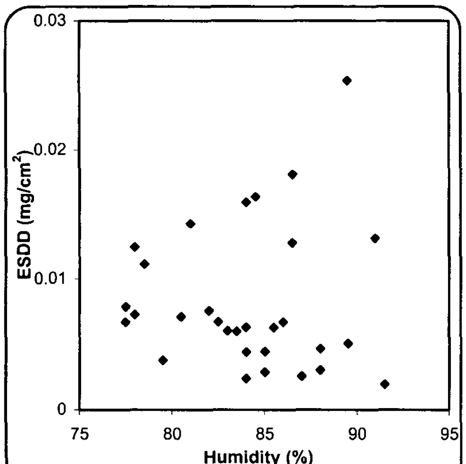

Fig. 3: ESDD vs. humidity

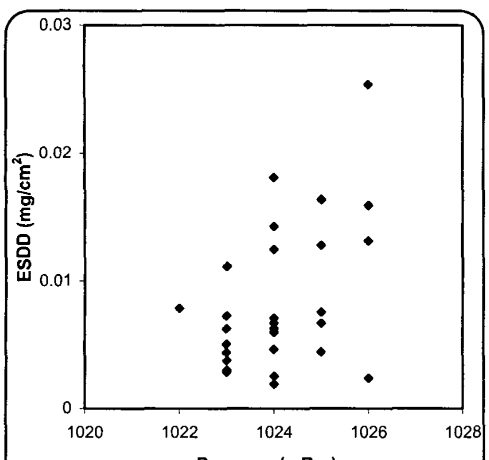

Fig. 4: ESDD vs. pressure

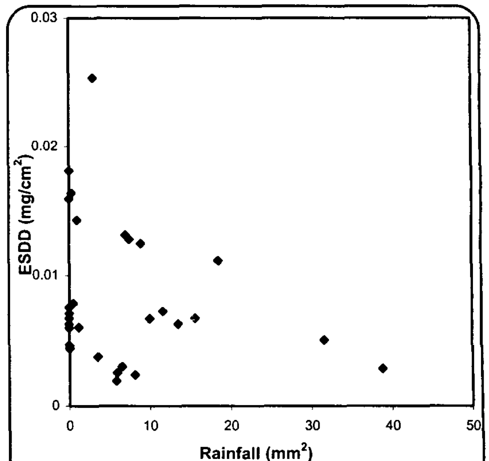

Fig. 5: ESDD vs. rainfall

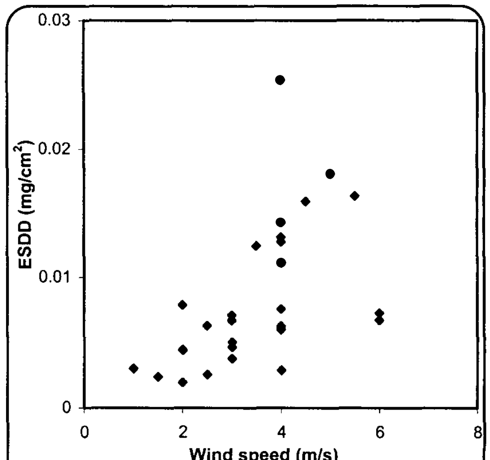

Fig. 6: ESDD vs. wind speed

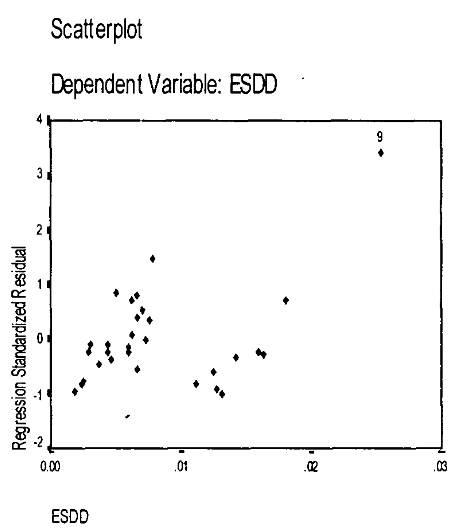

Figure 7: Analysis of residuals for the model

Using the least squares criterion to select the values for the B's, we obtain the result:

$$ESDD= -0.396 - 3.931E\text{-}04X_1 - 8.022E\text{-}05X_2 + 4.212E\text{-}04X_3 - 7.411E\text{-}05X_4 + 8.431E\text{-}04X_5 - 1.604E\text{-}04X_6 \quad (4)$$

From this equation we can predict the rates of response of ESDD to the change in the independent variables. A typical sample of temperature: for each 1° variation in temperature with no change in other variables, a predicted reduction of 3.931E-04 in ESDD since B_1 = temperature coefficient. The same way follows to predict the response in ESDD for every independent variable, taking into consideration the coefficient sign. It is noted that since B_o = -0.396, the predicted ESDD at a point of all independent coefficient variables = 0, is –0.396.

When the means and standard deviations of the predictors and criterion variables are significantly different, round-off error, which results from the inability to maintain a sufficient number of significant digits in the computations, may cause the partial regression coefficient to be erroneous. Thus most multiple regression analysis are computed using a standardized model:

$$Z_Y = \beta_1 Z_1 + \beta_2 Z_2 + \ldots\ldots \beta_6 Z_6 \quad (5)$$

In which β_j $(j=1, 2, \ldots 6)$ are called standardized partial regression coefficients, and Z_Y and Z_j $(j=1, 2\ldots..6)$ are the criterion variable and the predictor variables, respectively, expressed in standardized form especially, for $i= 1,2\ldots..n$, they are computed by

$$Z_y = \frac{y_i - \bar{y}}{S_y} \quad (6)$$

and

$$Z_j = \frac{X_{ij} - \bar{X}_j}{S_j} \quad (7)$$

in which S_y is the standard deviation of the criterion variable and S_j $(j=1, 2\ldots.6)$ are the standard deviations of the predictors variables. We can write our standardized model as:

$$ESDD= -0.07Z_1 -0.058Z_2 +0.081Z_3 -0.126Z_4 +0.19Z_5 - 0.744Z_6 \quad (8)$$

It can be shown that the standardized partial regression coefficient *(β's)* and the partial regression coefficient *(B's)* are related by

$$B_j = \frac{\beta_j S_y}{S_j} \quad (9)$$

A. Correlation Matrix

A convenient way to display a set of correlation coefficients is in the correlation matrix as shown in Table 1. From the table, it shows that some of the predictor variables are highly correlated with ESDD, particularly, wind direction being the largest and then, wind speed; the correlation coefficients are -0.868 and 0.529 respectively. This mean that wind direction has the greatest influence on ESDD then followed by wind speed and atmospheric pressure. Some predictor variables are highly correlated with each other, particularly between temperature and humidity, the inverse correlation coefficient between them is -0.912. Some relationships between the variables were not clear from the coefficients, may be due to the fact that these variables have a high correlation with a third variable such as pressure in our case, it has a significance correlation with these two variables, temperature and humidity. The correlation coefficients are –0.263 with temperature and 0.373 with humidity.

B. Analysis of Residuals

As in fitting any linear model, analysis of residuals from a regression model is necessary to determine the adequacy of the least squares fit. It is helpful to examine a normal probability plots, plot of residuals versus fitted values and plot of residuals versus each regression variable. These are mainly concerned with checking the residuals to be assured that the model with its assumption is reasonable for the data. In general, when the model is correct, the standardized residuals tend to fall between +2 and –2 and are randomly distributed about zero. A normal probability plot of residuals from our linear regression model is shown in Figure 7. From the examination of the graph, there is one point which has a large residual. The primary feature of these plots is the large residual associated with run number (9). The standard residual for this run is (3.4). Since the standardized values of these run is larger than ∓ 2, we suspect that this run is potential outlier. The outlier is a peculiarity and indicates a data point, which is not at all typical of the rest of the data. It is possible that the data recording error has been made. However, since there is no way to verify this and since there is no other non-statistical basis for deleting this run, we decided to leave this run in the model or for further determinations. After removing this run, we can conclude that the model specification is satisfactory, and we proceed with the analysis. Dropping this point from the data and re-estimating the regression equation can check the conjecture.

C. Model Test

The model has been tested by using different data, which has been collected from the same rainy season. Fig. (8) shows the fitness of the prediction curves of ESDD values for this months with the measured ESDD values which determined on the site. Reasonable result has been found

Table 1: Matrix correlation

	ESDD	Temperature	Humidity	Pressure	Rainfall	Wind speed	Wind direction
ESDD	1.000	0.036	0.041	0.453	-0.231	0.529	-0.868
Temperature	0.036	1.000	-0.912	-0.263	-0.100	0.154	-0.043
Humidity	0.041	-0.912	1.000	0.373	0.044	-0.178	-0.059
Pressure	0.453	-0.263	0.373	1.000	-0.267	0.133	-0.425
Rainfall	-0.231	-0.100	0.044	-0.267	1.000	-0.050	0.106
Wind speed	0.529	0.154	-0.178	0.133	-0.050	1.000	-0.434
Wind direction	-0.868	-0.043	-0.059	-0.425	0.106	-0.434	1.000

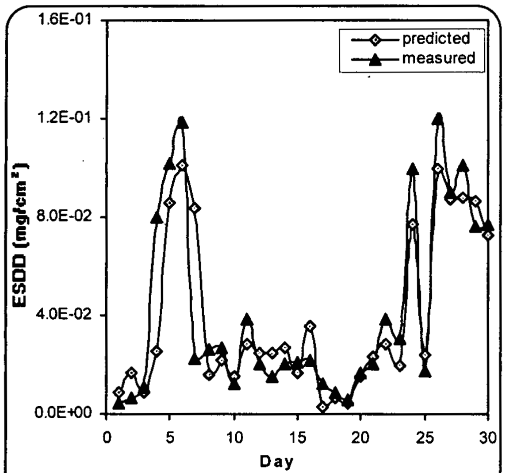

Fig. 7: ESDD predicted values with measured values

and the close curves reflects the fitness of the model with the real data which means that this model is valid for prediction of ESDD values for the eastern coastal region of the Peninsular Malaysia.

VII. CONCLUSIONS

- Regression technique incorporating multivariate statistics could be a useful way to develop ESDD mathematical model and predicting the contamination severity in given areas considering located meteorological parameters.
- The wind direction has the largest effect on ESDD followed by wind speed and air pressure.
- The predictor variables such as temperature, humidity, rainfall and wind direction have negative regression coefficients sign which means that ESDD increases when all these variables decrease.
- The interrelation coefficient is high between the temperature and humidity, -0.912. However, it is found that these parameters do not contribute any significant effect to ESDD value.

VIII. ACKNOWLEDGEMENTS

The authors wish to thank the general manager of Sultan Ismail Power Station and all the staff for their help and assistance and for providing all the facility to carry out this work. Also the authors wish to thank IRPA for providing the financial support to this project.

IX. REEERENCES

[1] G. N. Ramos, M. T. R. Campillo and Naito, "A Study on the Characteristics of Various Conductive Contaminants Accumulated on HV Insulators", IEEE Trans.PD, Vol. 8,pp.1842-1850,1993.

[2] O. E. Gouda, "Influence of Pollution on HV Insulators", Conference Record of the 1990 IEEE International Symposium on Electrical Insulation, Toronto, Canada, pp. 195-198, June 3-6, 1990.

[3] T. Fujimura, K. Naito and Y. Suzuki, "Dc Flashover Voltage Characteristics of Contaminated Insulators", IEEE Transactions on Electrical Insulation, Vol. EI-16, No. 3, pp. 189-198, June 1981.

[4] Kazuhiko Takasu, Takatoshi Shindo and Noboru Arai, "Natural Contamination Test of Insulators With DC Voltage Energization at Inland Areas", IEEE Transactions on Power Delivery, Vol. 3, No. 4.1847-1853, October 1988.

[5] Zheng J. C., Wang Z. and Liu Y. W, "Influence of Humidity on Flashover in Air in the Presence of Dielectric Surfaces", Proceedings of IEEE Conference Region 10 on Computer, Communication, Control and Power Engineering, TENCON' 93, pp. 443-449, 1993.

[6] Zhang Renyu and Zheng Jianchao, "Progress in Outdoor Insulation Research in China", IEEE Transactions on Electrical Insulation, Vol. 25, No. 6, pp. 1125-1137, December 1990.

[7] Liu Xianggheng and Bai Jianqun, "Selection of Insulation Level of HVAC Power Lines of Operating in High Altitude Polluted Area", Proceedings of the Second IEEE International Conference on Properties and Applications of Dielectric Materials, Vol. 1, pp. 268-271, 1988.

[8] Malaysian Meteorological Service, "Annual Summary of Meteorological Observations", 1996.

[9] Douglas C. Montgomery, Design and Analysis of Experiments: John Wiley & Sons. 1991.

[10] Own L. Davies and Peter L. Goldsmith, Statistical Methods in Research and Production: Oliver and Boyd Tweeddle Court. 1972.