# The Influence of Heat Loss on Wind Generators to Implement Condition-Monitoring System Based on the Application of the Polynomial Regression Model

Khaled B. Abdusamad*‡, David Wenzhong Gao **, Eduard Muljadi***

*Department of Mechanical and Materials Engineering, University of Denver, Denver, CO 80208, USA

**Department of Electrical and Computer Engineering, University of Denver, Denver, CO 80210, USA

***National Renewable Energy Laboratory, Golden, CO 80401-3393, USA

(khaledabdusamad@yahoo.co.uk,wenzhong.gao@du.edu, eduard.muljadi@nrel.gov)

‡Corresponding Author:David Wenzhong Gao, University of Denver, Department of Electrical and Computer Engineering,2390 S. YorkStreet, Knudson Hall 302, Denver, CO 80210, USA, Tel: +1 303-871-6618, Fax: +1 303-871-2194, wenzhong.gao@du.edu

*Received: 24.03.2014 Accepted: 11.05.2014*

**Abstract-**This paper presents an application of a condition-monitoring system (CMS) based on a polynomial regression model (PRM) to study the influence of heat loss on a wind generator's temperatures. Monitoring the wind generator temperatures is a significant for efficient operation, and plays a key role in an effective CMS. Many techniques, including prediction models can be utilized to reliably forecast a wind generator's temperature during operation and avoid the occurrence of a failure. PRMs arewidely used in situations when the relationship between the response and the independent variables are curve-linear. These techniques can be used to construct a normal behavior model of an electrical generator's temperatures based on recorded data. Many independent variables affect a generator's temperature; however, the degree of influence ofeach independent variableonthe response is dissimilar. In many situations, adding a new independent variable to the model may cause unsatisfactory results;therefore, the selection of the variables should be very accurate. A generator's heat loss can be considered a significant independent variable that greatly influences the wind generator with respect to the other variables. A generator's heat loss can be estimated in intervals by analyzing the exchange in the heat between the hot and cold fluid through the heat exchangers of wind generators. A case study built on data collected from actual measurements demonstrates the adequacy of the proposed model.

**Keywords-** condition-monitoring system, polynomial regression model, heat loss, predicted generator temperature, independent variables.

## 1. Introduction

Wind generator failure caused by high temperatures, which can occur as a result of various reasons, has increasedremarkably. External vectors such as harsh climatic situations and variable electrical loads cause abnormal operation conditions.External vectors such as harsh climatic situations and variable electrical loads cause abnormal operation conditions. However, wind energy reliability is growing as a result of recently implemented advanced monitoring systems. Applying modern condition monitoring to wind generators increases the generated wind power and helps reduce operation and maintenance expenses, especially in wind turbines deployed offshore.A condition-monitoring system (CMS) can provide detailed information about a wind generator's condition by analyzing measured signals to predict and avoid an impending failure in the wind turbine's components.The temperature of a wind generatoris one of its most important operating characteristics, and it needs to be monitored. In wind generators, the electromagnetic losses produce a significant amount of heat, which in turn warms up the generator and causes a temperature increase in the stator bars that reduces the life span of the insulation materials. This problem challenges engineers to develop proper CMSs for wind turbine generator parts [1]–[3].

Researchers have improved several condition-monitoring techniques that can increasethe reliability of the wind energy industry. Analyzing temperature trends by using the nonlinear state estimation technique (NSET) is one of the proposed methodologies to perform condition monitoring on wind generators [4]. To detect a fault that can potentially occur as a result of high temperatures of wind generators, the differences between the predictions and the actual values are used as important indicators. M. Popa, B. Jensen, E. Ritchie, and I. Boldea discussed using the time and frequency domain analysis to apply condition monitoring to a wind generator [5].The authors emphasized that generator faults may be detected by monitoring the stator and rotor current line trends when both the stator and generator rotors are under an unbalanced force. They applied the machine current signature analysis (MCSA) method, which is a noninvasive online or offline monitoring technique to diagnose faults in generators, such as turn-to turn faults, broken rotor bars, and static or dynamic eccentricity. In [6], heat transfer analysis and fluid mechanics relations were used to develop a proper CMS on wind generators based on the increase in the generators' temperatures. This work presents a new CMS that is applied to wind generators that work with water-air heat exchangers. The results obtained from the proposed model show that the increase in the heat loss is not desirable with respect to the logarithmic average of the temperature differences of the generator heat exchanger. In [7], the authors presented a new condition-monitoring method based on applying a multiple linear regression model (MLRM) to a wind turbine generator. The method can be used to construct a standard conduct model of electrical generator temperatures based on a generator's recorded temperature data. The main idea of the proposed technique is to measure the correlation between the observed and predicted values of the criterion variables based on reordered generator temperatures.

To improve upon the results that were obtained by utilizing the approach in [7], variables needed to be considered and/or added. The heat loss of wind generators is a very significant independent variable that affects the generator temperature; therefore, it is worthwhile to focus on the wastedheat of wind generators. The heat loss of wind generators can be estimated from the thermal aspect by computing the heat exchange between the hot air that comes from the high temperatures of a generator's parts and the cold fluid inside a generator's heat exchangers [6], [8], [9]. This paper studies and evaluates the effect of heat loss on a generator's temperatures by employing regression models. Regression models are formulatedinseveral ways, such as the standard MLRM or a polynomial regression model (PRM). A regression model is an effective approach that can be utilized to predict and monitor temperatures inside wind turbine generators by evaluating the correlation between the observed values and the predicted values of the criterion variables based on recorded generator temperatures. The rest of the paper is organized as follows: Section 2 introduces background information about the MLRM and describes how model validation can be detected. Section 3 presents the statistical tests that can be applied to measure the standard MLRM adequacy. The transformation process to the PRM is explained in Section 4. To demonstrate the estimation of wind generator heat losses from the thermal side, Section 5 provides heat transfer analysis for the heat exchangers of wind generators. A case study is provided in Section 6 to demonstrate the methodology of the present work. Section 7 presents the results of the proposed method. Finally, the conclusion of the proposed work is given in Section 8.

## 2. Theoretical Background on the Standard MLRM

The MLRM is one of the most popular statistical techniques used to predict the behavior of a response (dependent variable) by modeling a group of explanatory variables (independent variables) [9]–[11]. The independent variables that affect the generator temperature can be defined as follows [4], [7]:

- *Generator power(GP):* A generator's power has a direct effect on a generator's temperature. The stator current in a generator rises when the electrical load is high, which leads to an increase of the generator's output power and the temperature of the generator.
- *Ambient or outside temperature (OT):* The considerable and frequent rise in the outside temperature leads to an increase ina generator's temperature.
- *Nacelle temperature (NT):* A nacelle's temperature is closely related to a generator's temperature, because the generator itself is located inside the nacelle component.
- *Cooling temperature (CT):* There is a strong relationship among the temperatures of the cold fluids inside the heat exchangers, which are used to cool a generator and the temperature of the generator. Heat exchangers that are used in wind generators play a remarkable role in providing proper operating conditions for the generator components.
- *Heat loss (HL):* The heat loss of a wind generator is a very significant variable that can be added to the proposed model as an independent variable because the heat loss warms up the active parts of a generator and causes an increase in the generator's temperature [6], [8], [9].

The generator stator winding temperature (GT) represents the dependent variable (response) in the proposed model and strongly correlates to the previous independent variables [3,6]. The multiple regression model might describe the relationship between the dependent variable (GT) and the independent variables as follows [10]–[12]:

$$GT = \beta_0 + \beta_1 CT + \beta_2 GP + \beta_3 HL + \beta_4 NT + \beta_5 OT + \varepsilon \qquad (1)$$

Where:

- $\beta_i$ $(i = 1,2, \dots, k)$ is the $i^{th}$ regression coefficient $\beta$ corresponding to the independent variables, and $k$ is the number of the independent variables.
- $\varepsilon$ is the residual, defined as the difference between the experimental and predicted value.

The response GT, which is related to the predictor variables (independent variables X) and theregression coefficients$\beta=[\beta_0, \beta_1, \beta_2, \dots\dots, \beta_k]'$ can be estimated by using the method of least squares, as follows:

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \vdots \\ \beta_k \end{pmatrix}, \quad X = \begin{pmatrix} 1 & CT_1 & CT_2 \dots\dots CT_k \\ 1 & GP_1 GP_2 & \dots\dots GP_k \\ 1 & HL_1 HL_2 & \dots\dots HL_k \\ 1 & NT_1 NT_2 & \dots\dots NT_k \\ 1 & OT_1 OT_2 & \dots\dots OT_k \end{pmatrix} \qquad (2)$$

$$y = \begin{pmatrix} GT_1 \\ GT_2 \\ \vdots \\ \vdots \\ GT_n \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \vdots \\ \varepsilon_n \end{pmatrix}, \ \beta = (X'X)^{-1}X'y \qquad (3)$$

Then the predicted dependent variable ($\widehat{GT}$) can be formulated as follows:

$$\widehat{GT} = \beta_0 + \beta_1 CT + \beta_2 GP + \beta_3 HL + \beta_4 NT + \beta_5 OT \qquad (4)$$

The residuals ε (the difference between the observed values of GT and the corresponding predicted values$\widehat{GT}$) play an important role in evaluating the adequacy of the fitted regression model and the shape of the model. Themodeldeficienciesshow up clearly by analyzingtherelationship between the residuals ε and thecorrespondingfittedvalues$\hat{y}$, [10]–[12]. The error variance of term yis $\sigma^2$, which can be determined by using the following equation:

$$\sigma^2 = \frac{SS_{Res}}{n-p} \qquad (5)$$

Where n is the number of observations (experimental data points) and$SS_{Res}$ is the residual sum of squares, which can be given as follows [9]–[11]:

$$SS_{Res} = y'y - \beta'X'y \qquad (6)$$

The term $n - p$ is the residual degree of freedom, and $p = k - 1$, where k is the regression degree of freedom. The residual mean square can be determined as follows [10]–[12]:

$$MS_{Res} = \frac{SS_{Res}}{n-p} \qquad (7)$$

The total sum of squares $SS_T$is a statistical term that is partitioned into the sum of squares as a result of regression$SS_R$ and the residual sum of squares $SS_{Res}$. Thus,

$$SS_T = SS_R + SS_{Res} \qquad (8)$$

$$SS_R = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2, \ \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad (9)$$

The regression mean square can be defined as follows:

$$MS_R = \frac{SS_R}{k} \qquad (10)$$

The regression and residual mean squares are utilized to determine the existence of the linear relationship between the response and any of the independent variables based on the test for the regression significance. The model's coefficients must be within a range determined by utilizing the next formula [10]–[12]:

$$\beta_i - \left[t_{\frac{1-\alpha}{2},n-p}\right] * \sqrt{\delta^2 * C_{jj}} \leq \beta_{jj} \leq \beta_i + \left[t_{\frac{1-\alpha}{2},n-p}\right] * \sqrt{\delta^2 * C_{jj}} \qquad (11)$$

Where $\delta^2$ is the residual mean square, α is the confidence interval percent ($\alpha = 95\%$ in the model assumption), and $C_{jj}$ the jth diagonal element of the $(X'X)^{-1}$ matrix.To determine the importance level for each independent variable in the model, the length-scaling method can be used. The model coefficients are the criteria that are used to detect the importance level for each independent variable. The differences in the dimensions of the dependent variables and model coefficients affect the response values. To address this problem, it is beneficial to create dimensionless regression coefficients by utilizing the length-scaling method. The corrected sum of squares for an independent variable $X_i$:

$$S_{jj} = \sum_{i=1}^{k} (X_{ij} - \bar{X}_i)^2, \qquad (12)$$

Where i = 1, 2,..., k,j= 1, 2,..,n, and $X_{ij}$ is the independent variable i at a specific observation j. The average value of the dependent variable $\bar{X}_i = \frac{1}{n}\sum_{j=1}^{n} X_{ij}$ for the duration of the window of observation.The corrected sum of squares for the dependent variable $GT_i$:

$$S_{ii} = \sum_{i=1}^{k} (GT_{ij} - \overline{GT}_i)^2, \qquad (13)$$

The simple correlation between $X_i$ and $X_j$is $r_{ij}$:

$$r_{ij} = \frac{S_{ij}}{(S_{ii}S_{jj})^{1/2}} \qquad (14)$$

In this scaling, each new regressor is $w_i$. Based on the correlation matrix $w'w$ and the standardized regression coefficients matrix,$\hat{b}$can be estimated as follows:

$$w'w = \begin{pmatrix} 1 & r_{12} & r_{13} \dots\dots r_{1K} \\ r_{12} & 1 & X_{23} \ \dots\dots X_{2K} \\ r_{13} r_{23} & 1 & \dots\dots X_{3K} \\ \vdots & \vdots & \vdots \quad\quad \vdots \\ r_{1k} r_{2k} & r_{3k} & \dots\dots 1 \end{pmatrix} \qquad (15)$$

$$\hat{b} = (w'w)^{-1} r_{iy} = \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \vdots \\ \vdots \\ \hat{b}_k \end{bmatrix} \qquad (16)$$

Estimating the standardized regression coefficients for each independent variable is essential for measuring the reliability of each independent variable. Such estimations indicate the significance of each variable and term in the model [10]–[12]. The following section presents the required test that can be applied to measure the model adequacy.

## 3. Measure of Model Adequacy

Several techniques can be utilized to determine the validity of the proposed model. They can be summarized as follows:

### *3.1. Test of Individual Regression Coefficients*

This test is helpful in selecting the best independent variables for the model. The variance of the response probably increases with more independent variables; therefore, the model might be more efficient with the deletion of some of the independent variables in the model. Determining the correlation coefficients and the degree of significance for each variable areboth beneficial to defining the proper independent variables in the model. The hypothesis for testing any regression coefficient is as follows: If the significance of the regression coefficient's statistical value for a particular independent variable ($t_0$) is more than the proposed t value ($t_{TABLE}$), then the hypothesis of $H_0: \beta_i = 0$ is rejected and the independent variable strongly contributes to the model. This hypothesis can be interpreted as follows [10]–[13]:

$t_0$ should be $> t_{(1-\alpha)/2,k,n-k-1}$,

$$t_0 = \beta_i / \sqrt{SS_{Res}.C_{jj}} \tag{17}$$

### *3.2. Test of Significance of Regression*

A test of significance of regression is required to investigate whether a linear relationship between the response and any of the regressor variables is present. The statistical concept of this test emphasizes that at least one of the independent variables is clearly related to the model. The hypothesis of this test is as follows: If the significance of the regression statistical value ($F_0$) is more than the proposed F value ($F_{TABLE}$), then the hypothesis of $H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$ is rejected, which means [10-13]:

$F_0$ should be $> F_{1-\alpha,k,n-k-1}$,

$$F_0 = MS_R / MS_{Res} \tag{18}$$

### *3.3. The Test for Lack of Fit*

This test can be considered as an overall test of the model adequacy. The lack-of-fit test is helpful to determine whether the linear relationship fits the obtained data. The requirements of this test refer to the normality in the distribution of the model's residuals, the independence of the regressor, and the constant-variance in the model. Fulfilling these requirements confirms whether the tentative model adequately describes the data. To perform this test, the lack-of-fit sum of squares $SS_{LOF}$ and the pure-error sum of squares $SS_{PE}$ can be estimated as follows [10]–[12]:

$$SS_{LOF} = SS_{RES} - SS_{PE} \tag{19}$$

$$SS_{PE} = \sum_{i=1}^{m}\sum_{j=1}^{ni}(GT_{ij} - \overline{GT_i})^2, \tag{20}$$

Where $\overline{GT_i}$ is the generator temperature average of the i row of the generator temperature matrix, and $GT_{ij}$ is the generator temperature of the observation j at row i of the generator temperature matrix. The hypothesis of this test says that if the model adequately describes the data, then the assumption of (Lack-of-fit F test $F_{LOF} > F_{\alpha,df_{LOF},df_{PE}}$) will be rejected, where $F_{LOF}$ can be calculated from the next relation:

$$F_{LOF} = \frac{\text{mean square of lack of fit}}{\text{mean square of pure erroe}} = \frac{SS_{LOF}/df_{LOF}}{SS_{PE}/df_{PE}} \tag{21}$$

Where $df_{LOF}$ is the lack of fit degree of freedom, and $df_{PE}$ is the pure-error degree of freedom.

### *3.4. PredictedResidual Sum of Squares Statistic Test*

The Predicted Residual Sum of Squares (PRESS) is a measure of the regression model validity and the prediction of potential performance. The PRESS statistical value can be defined as the sum of the squared residuals. An observation that falls outside of the general trend of the data should be considered as adversely affecting the model. The presence of this variation can be exposed by computing the PRESS statistic value and comparing that value to the residual sum of squares. Therefore, the PRESS statistic is considered as a measure of how well the regression model will be able to predict new data. To compare PRESS to the residual sum of squares, a small value of PRESS is desirable in the proposed model. The PRESS value can be computed as follows [10]–[13]:

$$\text{PRESS statistic value} = \sum_{i=1}^{n}\left(\frac{GT_i - \widehat{GT_i}}{1 - h_{ij}}\right)^2 \tag{22}$$

Where $h_{ij}$ is the leverage for the ijth element of the hat matrix (H), $H = X(X'X)^{-1}X'$.

### *3.5. The Coefficient of Multiple Determination*

The coefficient of determination $R^2$ also measures the goodness of fit of the proposed model. The high value of $R^2$, however, does not necessarily denote that the regression model is suitable. In many cases, adding a new independent variable to the model may causes worse results. This situation occurs when the mean squared error for the new model is larger than the mean squared error of the old model—even though the new model will show an increased value of $R^2$. The coefficient of determination is defined as follows [10]–[13]:

$$R^2 = \frac{SS_R}{SS_T}, \tag{23}$$

### *3.6. Multicollinearity Test*

Multicollinearity is an undesirable case that might occur in the regression models when one or more of the independent variables are robustly correlated to each other. This condition negatively affects the obtained results. The high multicollinearity among the independent variables results in a large variance and covariance for the least-squares estimators of the regression coefficients. High multicollinearity causes the coefficients to be insignificant because the model coefficients will be unstable, and their standard errors will become large. A very simple way to measure the multicollinearity is to calculate the inflation factors (VIFs),which can be obtained from the following formula [10], [11], [14]:

$$\mathrm{VIF_j} = (1 - \mathrm{R_i^2})^{-1} \tag{23}$$

Where $\mathrm{R_i^2}$is the coefficient of determination obtained when a particular independent variable regressed with a degree of freedom equals the total number of variables (GT and Xs) $-1$. Many experts confirm that when the VIFs values exceed 5 (high multicollinearity), the regression coefficients of the model will be unreliable [10], [11], [14]. SPSS and Minitab statistical software [15], [16] automatically perform a tolerance analysis and will not adopt the model results with tolerance < 0.2 for each variable inserted into the regression model.

$$\mathrm{Tolerance} = \frac{1}{\mathrm{VIF}} \tag{24}$$

A transformation process is required when the proposed MLRM fails to exceed the previous statistical tests.Thisis presented in the following section.

## 4. The Transformation Process to the Polynomial Regression Models.

When the proposed model fails to exceed the statistical hypothesis tests that were explained previously, a transformation process on the independent variables is required. Even though the condition of the normal distribution is satisfied—i.e., the residual is almost normally distributed—this transformation process is needed because the relationship between the dependent variable GT and some of the independent variables is nonlinear. Therefore, a nonlinear function is required to fit the collected data [10]–[13]. Many nonlinear functions can be used for this purpose. PRMs are widely used in situations when the relationships between the response and the independent variables are curve-linear. For example, the second-order polynomial model with two variables would be:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1^2 + \beta_4 X_2^2 + \beta_5 X_1 X_2 \tag{25}$$

One of the most common transformation methods that works well with PRMs is the mean-centering technique. The methodology of this technique involves first calculating the mean of each independent variable. Then, by subtracting the old independent variable value at each observation from its mean, new data can be created for each variable as follows [10]–[14]:

$$X_{i_{NEW}} = X_{i_{old}} - \overline{X}_{i_{old}} \tag{26}$$

In addition to the applicability of this technique, the multicollinearity problem disappears because the mean-centering technique reduces the correlation between the independent variables and the variance for the least-squares estimators of the regression coefficients [9]–[11], [13]. Section 5 presents the mathematical analysis that is used to estimate the heat loss of a wind generator.

## 5. The Estimation of the Heat Loss of the Wind Generators.

A generator's heat losses, which directly increase a generator's temperature, can be estimated based on the air friction losses, generator bearing losses, iron losses, stator winding losses, and harmonic losses in the rotor of the generator [6], [8]. However, a generator's heat losses can be evaluated from the thermal aspect because the heat losses are equal to the heat gains of the fluid that flows through the heat exchanger of the wind generator. The heat gain of the cold fluid through the heat exchanger is supposedly equal to the generator's heat loss. Heat exchangers are used to transmit the heat from the hot fluid into cold fluid. The losses of the heat between the heat exchanger and surroundings could be neglected because such losses are very low compared to the heat exchange between the cold and hot fluids;therefore, heat exchangers are considered adiabatic devices—i.e., no heat is exchanged with the surrounding through the heat exchangers. Heat exchangers are classified according to their construction and the flow configuration. Many types of heat exchangers are used to provide a suitable cooling system in wind generators. Counterflow tube heat exchangers are the most widespread type that are used in wind generator cooling systems [17]–[19]. Figure 1 displays the mechanism of work in a cross-flow case of a tube heat exchanger model. The figure shows that one flow moves inside a set of tubes and another flow moves through the outer shell in the opposite direction.

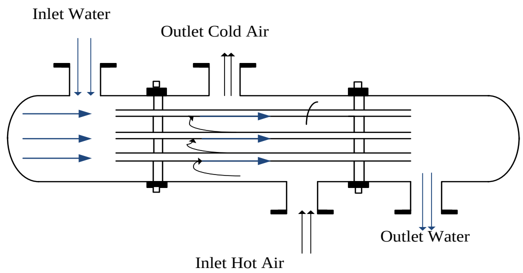


**Fig. 1.** Tube counterflow heat exchanger [17]–[19]

The heat losses from the generator $\mathrm{Q_h}$ are equal to the heat gains for the cold fluid of the heat exchanger $\mathrm{Q_c}$, based on the energy balance analysis between the cold and hot fluids, as shown in the following equations:

$$\mathrm{Q_h} = \mathrm{Q_c} \tag{27}$$

$$\dot{m}_h C_{p,h}(T_{h,i} - T_{h,o}) = \dot{m}_c C_{p,c}(T_{c,o} - T_{c,i}) \tag{28}$$

The aboveequations assume that the potential and kinetic energy are negligible, and the specific heat does not change over time for each fluid. $T_{h,i}$ and $T_{c,i}$are defined as the hot and cold fluid inlet temperatures of a wind generator's heat exchanger, respectively. $T_{h,o}$ and $T_{c,o}$are defined as the hot and cold fluid outlet temperatures of a wind generator's heat exchanger, respectively.$C_{p,h}$ and $C_{p,c}$are defined as the hot and cold fluid specific heat, respectively. To simplify the calculations, the average temperature of the inlet and outlet points of the hot or cold fluid are assumed to represent the hot and cold fluid inlet temperatures [17]–[19].Advanced heat exchanger design engineers commonly refer to the LMTD term, which is defined as the average logarithmic of the temperature differences between the hot and cold streams at each end of an exchanger with a constant flow. In other words, LMTD is the maximum mean temperature difference that can be achieved for any given set of inlet and outlet temperatures through the heat exchangers, which can be expressed as follows [16]–[18]:

$$\mathrm{LMTD} = \frac{\left[\left(\mathrm{T_{h,i}} - \mathrm{T_{c,o}}\right) - \left(\mathrm{T_{h,o}} - \mathrm{T_{c,i}}\right)\right]}{\ln\left[\frac{(\mathrm{T_{h,i}} - \mathrm{T_{c,o}})}{(\mathrm{T_{h,o}} - \mathrm{T_{c,i}})}\right]} \quad (29)$$

In addition to energy balance, heat transfer can be described based on the averagelogarithmic of the temperature difference, as follows [17]–[19]:

$$\mathrm{Q} = \mathrm{U.A.LMTD} \quad (30)$$

Where A is the area of the heat exchanger and U is the overall heat transfer coefficient, which can be calculated from the next formula [17]–[19]:

$$\mathrm{U} = \frac{1}{\frac{1}{\mathrm{h_h}} + \frac{1}{\mathrm{h_c}}} \quad (31)$$

Where $\mathrm{h_c}$ and $\mathrm{h_h}$ are the average internal and external heat coefficients of the cold and hot fluid, respectively [17]–[19].The heat losses of wind generators, therefore, can be estimated with the aid of Eq. (30). Figure 2 presents a flowchart that illustrates the methodology of the present work.

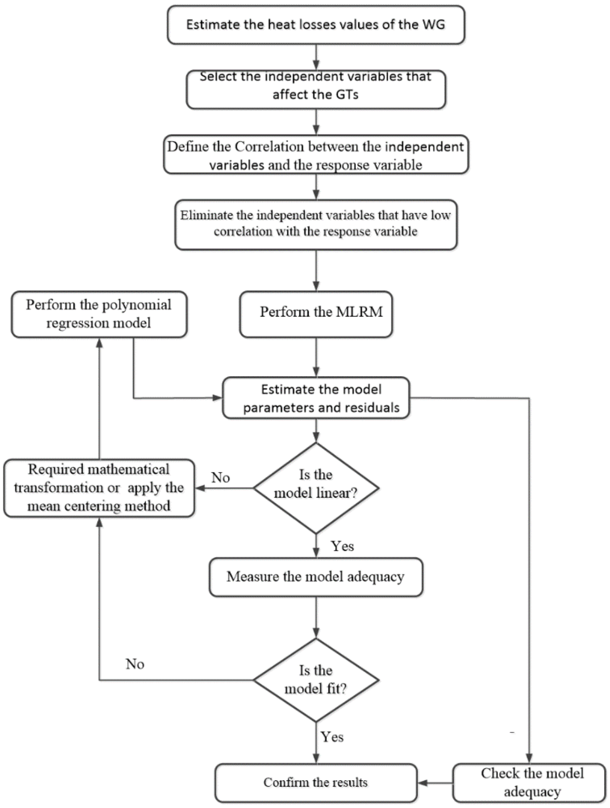


**Fig. 2.** The methodology of the proposed work

The previous flowchart indicates that a transformation process might be required if the model is not linear. A case study is presented in the following section to confirm the validity of the proposed technique.

## 6. Case Study.

This case study involves actual data collected from a variable-speed offshore wind turbine with rated power of 5 MW, 60Hz, three blades, 126m rotor diameter, and rated rotor speed of12.1 rpm. The wind turbine has a synchronous permanent magnet generator with a rated speed of 1,500 rpm; generator efficiency is 94.4% [20]. The coolingsystem of the generator includes a water-air counterflow heat exchanger with six cold-fluid pipes. Several temperature sensors are installed within the generator to measure the stator winding temperatures, or stator-core temperatures. The manufacturer handbook emphasizes that the generator temperature should not exceed 110C° to protect the electric generator, and the wind turbine will shut down when the generator temperature reaches 135 C°.

More additional temperature- and pressure-measuring devices are available to gauge the water inlet and outlet temperatures and pressure drop through the heat exchanger [20]. These devices can be installed at the inlet and outletslots of the heat exchanger.The required SCADA system provides enough details about the collected variables that are needed to be inserted into the proposed model. Further, the temperature of the air that enters the rotor and stator winding region can be measured based on the outside temperature. There is a valve to control the inlet mass flow rate of water to the heat exchanger, which is roughly 2.6 kg/s according to themanufacturer handbook. In addition, the air mass flow rates into the rotor and stator end windings are designed to equal almost 4.7 kg/s [20]. This information is necessary to estimate the heat loss values. The recorded data employed to test the validity of the proposed model represents 740 working days and the turbine's life during from 25,000 to 35,000 operating hours. The change in the data for each variable is very small every single day; therefore, the average of each day's collected data (gauged each second) is representative of the selected variables.

As mentioned previously, the variance of the predicted generator temperature increases as the number of independent variables increases. A variable selection process, therefore, is required to reduce and eliminate the variables that have the least correlation coefficient to the generator temperature. In addition, the significance of the regression coefficients for each independent variable in the model must be estimated to finally select the independent variables. By using Minitab or SPSS statistical software andinserting the related data into the model, the correlation coefficients and the degree of significance for each variable can be determined, as shown in Table 1. As shown, the outside temperature and nacelle temperature variables have the least correlation coefficients with respect to the generator temperature; moreover, they are not highly significant because the p-value for both of them $> 0.05$, where p $= (1 - \alpha)$ is the degree of significance and $\alpha = 95\%$ according to the assumption of this present work. In this context, when the degree of significance for any independent variable is less than 5%, the contribution of that variable in the model is very strong[10]–[13].

The significance of the regression coefficient statistical values ($t_0$) for the independent variable ($t_0$) are also very helpful in determining the best independent variables for the proposed model in terms of the accuracy of the results. Table 2 clarifies low values of ($t_0$) for the outside temperature variable and the nacelle temperature variable. The ($t_0$) values for both variables are less than the standard t value ($t_{TABLE}$), which is equal to 1.960 [10]–[13];therefore, the outside temperature and nacelle temperature variables must be eliminated and removed from the proposed model because they do not influence the response strongly. The initial analysis of variance for this model is shown in Table 3.

**Table 1.** The correlation coefficients and the degree of significance for the model's variables.

| The Variables | | GT | CT | HL | OT | NT |
|---|---|---|---|---|---|---|
| CT | Corr. Coeff. | 0.95 | | | | |
| | P-Value | 0.00 | | | | |
| HL | Corr. Coeff. | 0.93 | 0.87 | | | |
| | P-Value | 0.00 | 0.00 | | | |
| OT | Corr. Coeff. | 0.56 | 0.72 | 0.42 | | |
| | P-Value | 0.14 | 0.06 | 0.14 | | |
| NT | Corr. Coeff. | 0.51 | 0.45 | 0.47 | 0.88 | |
| | P-Value | 0.15 | 0.13 | 0.19 | 0.00 | |
| GP | Corr. Coeff. | 0.96 | 0.83 | 0.95 | 0.87 | 0.95 |
| | P-Value | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

**Table 2.**The significance of regression coefficientsstatistical values.

| The Independent Variables | CT | HL | OT | NT | GP |
|---|---|---|---|---|---|
| Significance of the Regression Coefficients Statistical Value $t_0$ | 3.9 | 3.5 | 1.8 | 1.7 | 3.6 |

**Table 3.**The initial analysis of variance of the proposed model.

| Analysis of Variance | | | | | |
|---|---|---|---|---|---|
| Source | DF | SS | MS | F | P |
| Regression | 3 | 16,866.24 | 5,622.1 | 3026 | |
| Residual Error | 736 | 1,367.53 | 1.858 | | |
| Lack of Fit | 371 | 812.83 | 2.190 | 1.46 | 0.000 |
| Pure Error | 369 | 554.7 | 1.5 | | |
| Total | 739 | 18,233.77 | | | |

PRESS = 1368.78 R-Sq. = 92.5%

The coefficient of determination $R^2$ is high (92%), which indicates that adding a new term may make the regression model worse. This situation occurs when the mean squared error for the new model $MS_{Res_{New}}$(when adding new independent variables to the model) is larger than the mean squared error of the older model $MS_{Res_{old}}$ (without adding new independent variables to the model). To determine the model adequacy, the PRESS and the lack-of-fit statistic values are very useful. First, it must be determined that the PRESS statistic value is greater than the residual sum of square. Second, the significance of the regression statistical value $F_0 = 3026 \gg F_{1-\alpha,k,n-k-1} = 2.60$, which means that at least one of the independent variables is strongly related to the model. In addition, the lack-of-fit statistic value $F_{LOF}$ = 1.46> $F_{\alpha,df_{LOF},df_{PE}}$ = 1.The initial regression equation is as follows:

$$\widehat{GT} = 1.14733 + 0.438984\,CT + 0.0149205\,GP + 0.0345665\,HL$$

Figures 3 and 4 present the three-dimensional surface plots that correlate the nominated independent variables to the generator temperature. It seems that the relationships between the independent variables and the dependent variable are not linear, which means that the transformation process is required.

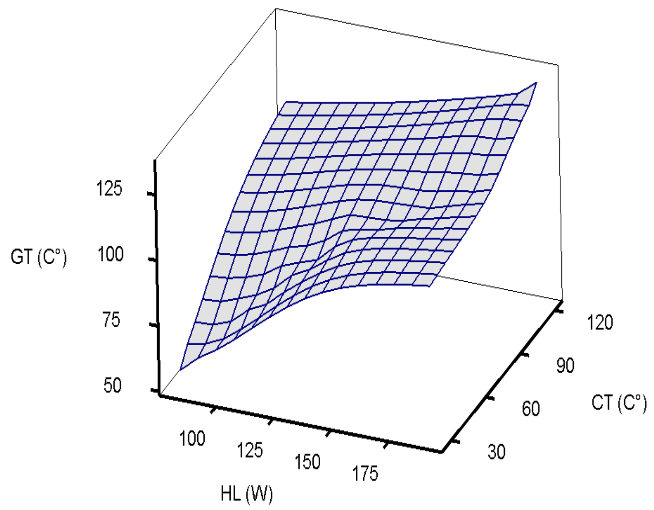


**Fig. 3.** The surface plot of CT, HL with GT [20]

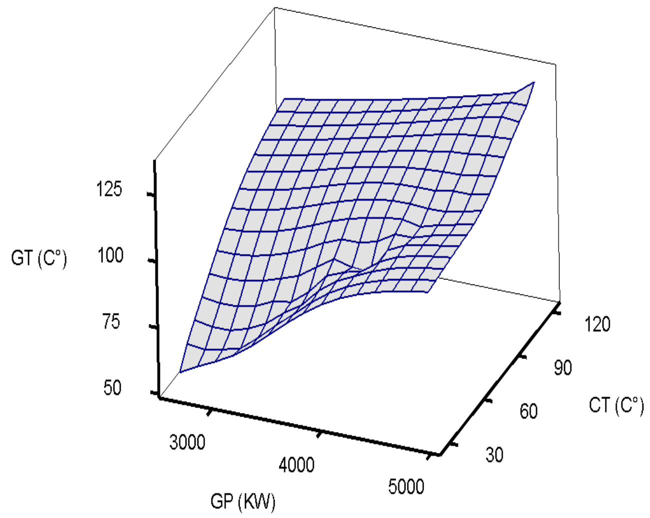


**Fig. 4**. The surface plot of CT, GP with GT [20]

The previous initial results indicated that the model does not fit the data, which means that the standard MLRM is not proper to the collected data because the relationship between the generator temperature and the nominated independent variables is not linear. The shape of the regression model can be clarified by defining the relationship between the generator temperature and each independent variable, as shown in as shown in Figs. 5, and 6.

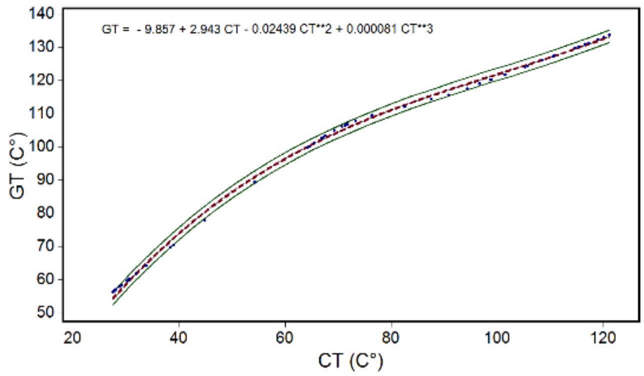


**Fig. 5.** The fitted curve between CT, GT variables

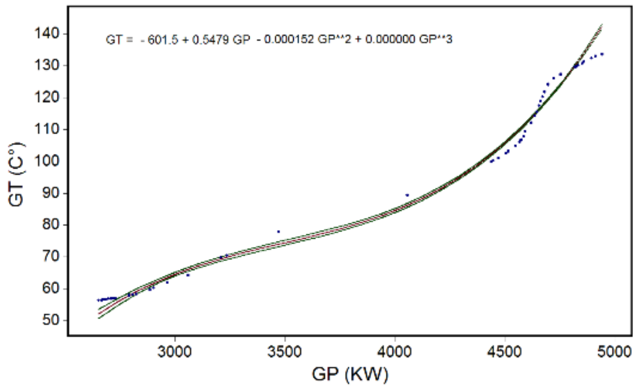


**Fig. 6.** The fitted curve between GP, GT variables

The antecedent graphical relations confirm that there are curve-linear relationships between the independent variables and the response. Further, it appears that the third-degree order is the best form for the proposed regression model that can be adopted to achieve logical results. Analysis of the residuals is an effective way to discover several types of model inadequacies.Figure 7 confirms that the pattern of the residuals versus the fitted values of the generator temperature is not contained in a horizontal band, which confirms that there are obvious defects in the model and a cubic term needs to be added to the model [10]–[13].

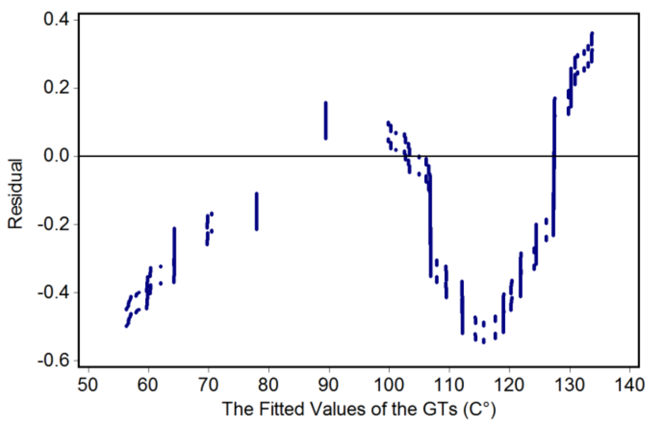


**Fig. 7.** The scatterplot of the residual versus the fitted values of the GTs.

Therefore, the appropriate transformation process is necessary to let the model exceed the proposed statistical tests.Polynomial models are very beneficial and widely used when the curve-linearity depicts the true response function. The fitting PRM of the third-order response surface in three independent variables is as follows:

$$\widehat{GT} = \beta_0 + \beta_1.CT + \beta_2.GP + \beta_3.HL + \beta_4.CT^2 + \beta_5.GP^2 + \beta_6.HL^2 + \beta_7.CT.GP + \beta_8.CT.HL + \beta_9.GP.HL + \beta_{10}.CT^3 + \beta_{11}.GP^3 + \beta_{12}.HL^3 + \beta_{13}.CT.GP.HL$$

The mean-centering method works perfectly with PRMs, where each independent variable is subtracted from its mean, and the new values of each variable are inserted into the PRM [10]–[13]. The final obtained results are presented in the following section.

## 7. Results and Discussions.

To measure the adequacy and normality of the proposed PRM, the model's residual should be analyzed. Figure 8shows that the error term $\epsilon$ is almost normally distributed, and it is very close to the ideal normal probability plot because the majority of the residual points are approximately distributed along a straight line. Few points fall outside the fitting line (regression line);these can be neglected because they do not affect the general trend of the model.

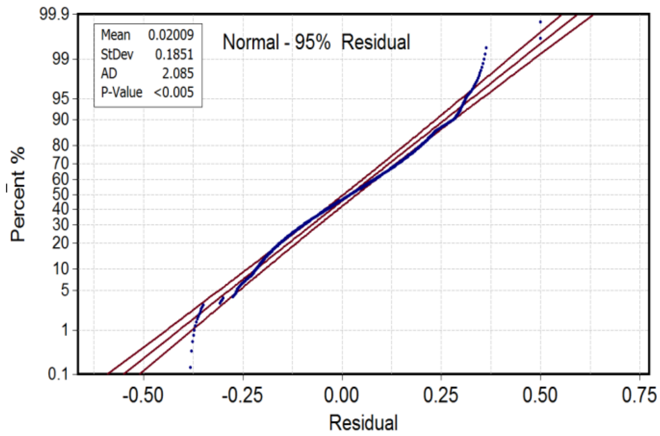


**Fig.8.** The residual normal probability plot

Figure 9 presents the residuals plot versus the predicted generator temperatures$\widehat{GT}_i$. The graph emphasizes that the residual distribution is perfectly normal because the majority of the points are contained in a horizontal band.The PRM that can be used to predict the generator temperature is as follows:

$$\widehat{GT} = 108.57 + 0.78CT - 0.0024GP + 47.45HL + 0.012CT^2$$
$$+4.27X10^{-5}GP^2 + 5943.31HL^2 - 5.34X10^{-5}CT^3$$
$$-2.85X10^{-7}GP^3 - 149805XHL^3 - 0.0014GP.CT$$
$$-0.16GP.HL - 8.84CT.HL + 0.0135GP.CT.HL$$

The final analysis of variance (ANOVA) for the proposed PRM is shown in Table 4.

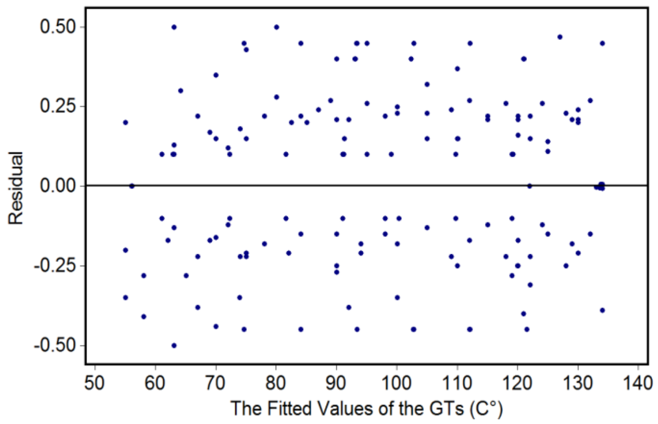


**Fig. 9.** The fitting GTs. values versus the residual plot

**Table 4.** The final analysis of variance (ANOVA) of theproposed model.

| ANOVA | | | | | |
|---|---|---|---|---|---|
| Source | DF | SS | MS | F | P |
| Regression | 13 | 10,473.71 | 805.67 | 633.4 | |
| Residual Error | 726 | 923.5 | 1.272 | | |
| Lack of Fit | 483 | 595.86 | 1.234 | 0.92 | 0.000 |
| Pure Error | 239 | 320.902 | 1.343 | | |
| Total | 739 | 17,835.832 | | | |

PRESS = 914.87 R-Sq. = 94.86%

The ANOVA table indicates that the model does not suffer from lack of fit with the data. It can consider that the proposed PRM describes the data because the statistic value $F_{LOF} = 0.92 < F_{\alpha,df_{LOF},df_{PE}} = 1$ ,and the PRESS statistic value = 914.87 $< SS_{Res}$= 916.762. Further, there is an improvement in the coefficient of determination $R^2$ value, which indicates that the PRM is more proper than the standard multiple regression model. The high value of the coefficient of determination confirms that there is no need to add independent variables to the mode and the selected variables are adequate. Finally, the results confirm that at least one of the independent variables relates strongly to the model because the significance of regression statistical value $F_0$ = 633.4 is still much larger than the$F_{1-\alpha,k,n-k-1}$ = 2.60.

The length-scaling method is required to determine the most important independent variable that strongly affects the proposed model [10]–[13]. Table 5 presents the most significant terms of the proposed PRM because they have the highest standardized coefficients values. The main purpose of determining the standardized coefficients of the model is specifying the most important terms in the model that strongly influence the generator temperature. For instance, increasing the standardized value of the heat loss variable by one unit increases the standardized value of the generator temperature by 9.86 units (dimensionless regression coefficient). In addition, the standardized value of the generator temperature increases by 4844.3 unitswhen increasing the standardized value of the third order of the heat loss variable by one unit. The high influence of the heat loss's third order term on the generator temperature can be seen.This indicates that the most significant variable that controls the generator temperature is the heat loss variable. The other independent variables differ in their impact on the model based on the weight of the standardized coefficient value of each term. Moreover, the table presents the coefficients' confidence intervals of the most significant terms in the model, which confirm that the coefficient of each term must be within the specified interval.

**Table 5.** The standardized coefficients of variance of the proposed model.

| Coefficients | | | |
|---|---|---|---|
| Model | 95% Confidence Lower Value | Interval Upper Value | Stand. Coeff. |
| HL | 28 | 67 | 9.86 |
| $HL^2$ | 5522 | 6364 | 214.46 |
| $HL^3$ | -159315 | -140294 | 4844.3 |
| CT | 0.565 | 0.924 | 6.87 |
| $CT^2$ | 0.009 | 0.024 | 2.25 |
| GP | -0.0035 | -0.00124 | 1.45 |
| CT.HL | -10.00 | -8.00 | 0.43 |
| GP.HL | -0.180 | -0.142 | 0.03 |
| GPCTHL | 0.0098 | 0.0162 | 0.0001 |

The last issue that should be mentioned in the ANOVA table analysis is the multicollinearity problem. The obtained results of the proposed model indicate that the model does not suffer from the multicollinearity problem because the variance inflation factors (VIFs) for the model's terms are very low ($< 5$), and the tolerance values are$> 0.2$. In addition, the degrees ofsignificance for the model terms are less than 0.05, which confirms that all terms are significant and contribute strongly in the model [10]–[14]. Table 6 presents the VIFs for the most important terms in the proposed model. The obtained VIFs emphasize that the independent variables are not correlated to each other, which increases the model reliability.

**Table 6.** VIF, tolerance, and the significant values for the most important terms in the model.

| Model | Tolerance | VIF | p |
|---|---|---|---|
| HL | 0.645 | 1.550 | 0.0014 |
| $HL^2$ | 0.540 | 1.85 | 0.0012 |
| $HL^3$ | 0.427 | 2.342 | 0.0018 |
| CT | 0.805 | 1.242 | 0.00185 |
| $CT^2$ | 0.873 | 1.145 | 0.0019 |
| GP | 0.922 | 1.085 | 0.00194 |
| CT.HL | 0.404 | 2.476 | 0.00196 |
| GP.HL | 0.443 | 2.256 | 0.00198 |
| GPCTHL | 0.459 | 2.178 | 0.00199 |

Contour plots are very beneficial and can be utilized to study the effect of the heat loss variable on the predicted generator temperatures. Figures 10, 11, and 12submit graphic representations of the relationships among three numeric variables. These figures present the influence of the heat loss on the predicted generator temperature with respect to the generator power variable. The predicted generator temperature is for contour levels, which are plotted as curves. As shown, the effect of the heat loss variable is regular in the first, second, and third order of the polynomial model.

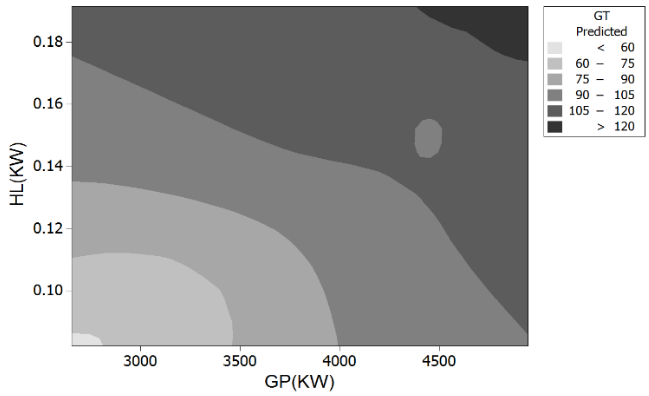


Fig. 10. The contour plot of the $\widehat{GT}$based on the first order of the hat loss in the PRM with respect to the GP.

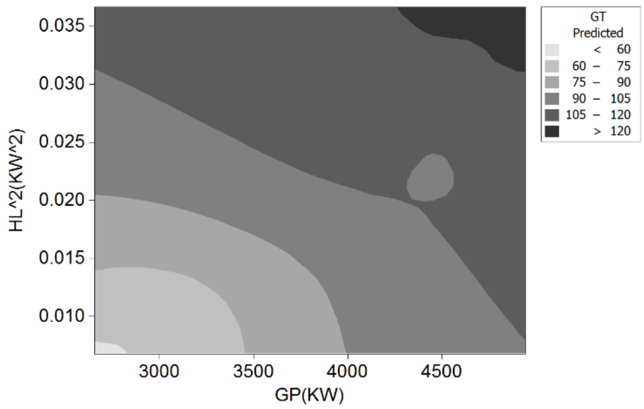


Fig. 11. The contour plot of the $\widehat{GT}$ based on the second Order of the heat loss in the PRM with respect to the GP.

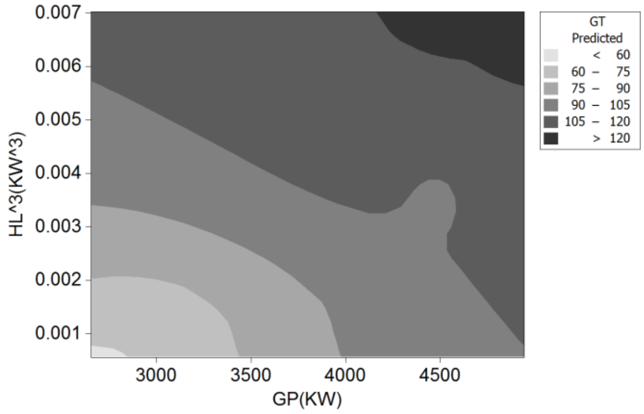


Fig. 12. The contour plot of the $\widehat{GT}$ based on the first order of the heat loss in the PRM with respect to the CT.

On the other hand, there is a very obvious decrease in the zone area of the predicted generator temperature when it exceeds 120 C° comparedto the effect of the cooling temperature variable on the predicted generator temperature as shown in Figs 13,14, and 15. It is very obvious that each term in the proposed PRM controls the predicted generator temperature based on the weight of its coefficients. Further, the influence of the heat loss variable on the predicted generator temperature is higher than the influence of the cooling temperature variable because the increase in the predicted generator temperature as a result of the influence of the heat loss variable is more than the influence of the cooling temperature variable.

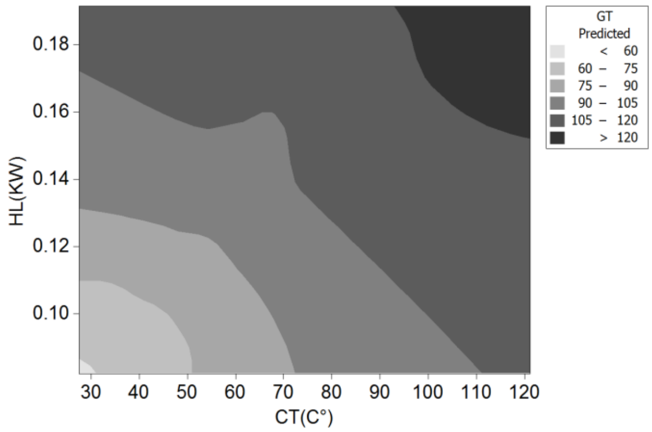


Fig. 13. The contour plot of the $\widehat{GT}$ based on the first order of the heat loss in the PRM with respect to the CT.

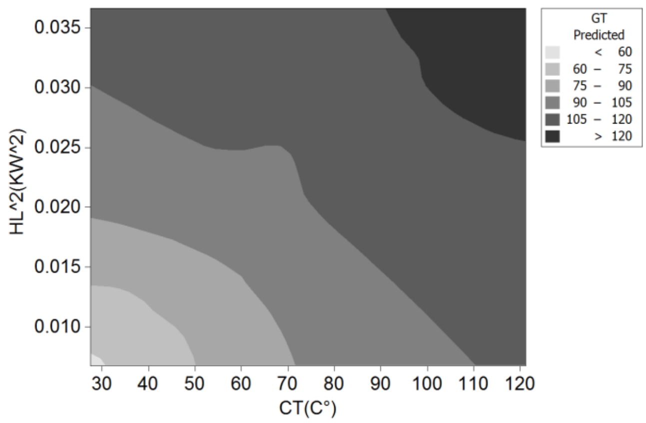


Fig. 14. The contour plot of the $\widehat{GT}$ based on the second order of the heat loss in the PRM with respect to the CT.

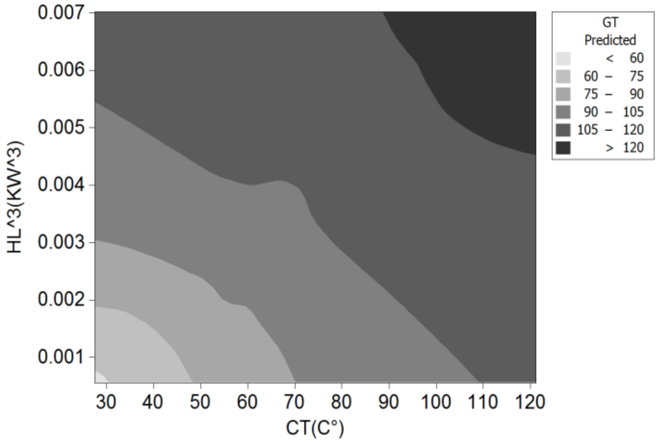


Fig. 15. The contour plot of the $\widehat{GT}$ based on the third order of the heat loss in the PRM with respect to the CT.

## 8. Conclusion.

To perform an effective condition-monitoring system, this paper proposes an application of the PRM based on the study of the heat loss's influence on a wind generator's temperatures. The proposed technique can be utilized to address the technical problems resulting from high temperatures on wind generators. This leads to reduced maintenance and operation costs and increased wind generators' reliability. The knowledge of CMS can be employed with the aid of regression techniques to create a prediction model for generator temperatures. The heat loss, cooling temperature, and generator power variables are the most significant variables that strongly affect generator temperature, according to the correlation coefficients and the degree of significance for each variable with the response. The transformation process to the PRM was required because the relationships between the response and the independent variables were curve-linear. The data behavior controls the regression model type and the need for the transformation process. The main concept of the proposed technique is measuring the correlation between the observed values and the predicted values of the criterion variables based on historical data. The obtained results confirm that the heat loss variable is the most influential variable on generator temperatures based on the standardized coefficients of the model. However, the other nominated independent variables obviously affect generator temperatures based on the weight of their coefficients. When comparing the regression technique with previous methods, such as the nonlinear state estimatemethod (NSET) to predicate the generator temperature, the proposed method has the advantages of determining the effect of the independent variables on the dependent variable accurately. Future work is required to apply this method to different data with other operational wind turbines to detect faulty conditions and advance the prediction of potential failures.

### Acknowledgements

This work was supported in part by National Science Foundation Grant 0844707.The authors gratefully acknowledge the help of Dr. Kathryn Johnson, who collects the data measurements from Colorado School of Mines and National Renewable Energy Laboratory. This work was also supported by the U.S. Department of Energy under Contract No. DE-AC36-08-GO28308 with the National Renewable Energy Laboratory.

### References

[1] Lu, Bin, et al, A review of recent advances in wind turbine condition monitoring and fault diagnosis, Power Electronics and Machines in Wind Applications, 2009. PEMWA 2009, IEEE, 2009

[2] Avelino J. Gonzalez, M. Stanley Balowin, J. Stein, and N. E. Nilsson, Monitoring and Diagnosis of Turbine-Driven Generator, Electric Power Research Institute, Prentice Hall, Englewood Cliffs, New Jersey 07632,1995.

[3] J. F. Manwell, J.G. Mcgowan, and A.L. Rogers, Wind Energy Explained Theory, Design, and Application, WILLY, Second edition2009.

[4] Guo, Peng, D. Infield, and X. Yang, Wind turbine generator condition-monitoring using temperature trend analysis, Sustainable Energy. IEEE Transactions on 3.1 (2012): 124-133.

[5] Mihet Popa, Birgitte-Bak Jensen, Ewen Ritchie, and Ion Boldea, Condition Monitoring of Wind Generators, IEEE Industry Applications Society 38th Annual Meeting, IAS'03, Salt Lake City, Utah USA, IEEE Signal Processing Society, October 2003, Vol. 3, pp. 1839-1846.

[6] K. B. Abdusamad, and D. W. Gao, The Application of Heat Transfer Analysis in Condition Monitoring System of Wind Generator, PES Asia-Pacific Power and Energy Engineering Conference. IEEE 2013.

[7] K. B. Abdusamad, David W. Gao, and Eduard Muljadi, A condition monitoring system for wind turbine generator temperature by applying multiple linear regression model, North American Power Symposium (NAPS), 2013. IEEE, 2013.

[8] P. Preecha, and J. Dejvises, The power losses calculation technique of electrical machines using the heat transfer theory, Power Engineering Conference, 2007. IPEC 2007. International. IEEE, 2007.

[9] Nandi, Subhasis, Seungdeog Choi, and Homayoun Meshgin-kelk, Electric Machines: Modeling, Condition Monitoring, and Fault Diagnosis, CRC Press, 2012.

[10] C. D. Montgomery, A. E. Peck, and G. Vining, Introduction to Linear Regression Analysis,WILEY, Fifth edition 2012.

[11] N.H. Bingham, and John M. Fry: Regression linear models in statistic, Springer-Verlag London Limited 2010.

[12] Ali S. Hadi, and Bertram Price, Regression analysis by example, Third edition, Wiley Series in probability and statistics, 2000.

[13] Achinity Haldar, and Sankaran Mahadevan, Probability, Reliability, and Statistical Method in Engineering Design, WILLY 2000.

[14] Wold, Svante, et al, The collinearity problem in linear regression,and the partial least squares (PLS) approach to generalized inverses, SIAM Journal on Scientific and Statistical Computing 5.3 (1984): 735-743

[15] IBM SPSS software, http://www-01.ibm.com/software/ analytics/products/statistics/ index.html.

[16] Minitab 16 software, http://www.minitab.com.

[17] F. P. Incropera, A. S. Lavine, and D. P. DeWitt, Fundamentals of heat and mass transfer, John Wiley & Sons Incorporated, 2011.

[18] J. H. Lienhard. A heat transfer textbook, Courier Dover Publications, 2011.

[19] R. K. Shah, D. P. Sekulić. Fundamental of Heat Exchanger Design, Published online, 2007.

[20] Data of a variable speed wind turbine 5 MW rated power, three phase permanent magnetic type 440/660 V 60 Hz. Provided by Dr. Kathryn, Colorado School of Mines, 2013.

| **Nomenclature** | |
|---|---|
| *GT* | Generator temperature |
| *GP* | Generator power |
| *OT* | Ambient or outside temperature |
| *NT* | Nacelle temperature |
| *CT* | Generator cooling temperature |
| $y$ | Dependent variable (experimental value). |
| $\hat{y}$ | The predicted dependent variable in the model |
| $\bar{y}$ | The experimental value mean (mean value of y) |
| $k$ | Number of independent variables = The regression degree of freedom |
| $X_i, i = 1,2,..,k$ | The $i^{th}$ independent variable from total set of k variables |
| $\beta_i, i = 1,2,..,k$ | The $i^{th}$ coefficient corresponding to $X_i$ |
| $\beta_0$ | The intercept coefficient (or constant) |
| $i=1,2,3,....,k$ | Independent variables' index |
| $n$ | Number of observations ( experimental data points) |
| $\varepsilon$ | Residual (the difference between the experimental and predicted value) |
| $SS_{RES}$ | The residual sum square |
| $SS_R$ | The regression sum square |
| $SS_T$ | The total sum of squares |
| $MS_{RES}$ | The residual mean sum square |
| $MS_R$ | The regression mean sum square |
| $\alpha$ | The confidence interval percent |
| $\sigma^2$ | The error variance of term |
| $\delta^2$ | The residual mean square |
| $C_{jj}$ | The $j^{th}$diagonal element of the $(X'\,X\,)^{-1}$ matrix |
| $\beta_{jj}$ | The $j^{th}$diagonal element of the $(\beta'\beta\,)^{-1}$ matrix |
| $F_0$ | The significance of regression statistical value. |
| $t$ | Statistical value (the ratio of the coefficient to its standard error) |
| $R^2$ | The coefficient of determination. |
| $S_{ii}$, $i = 1,2,\dots,k$ | The corrected sum of squares for regressor $X_i$ |
| $S_{jj}$, $j = 1,2,,,.,n$ | The corrected sum of squares for regressor $X_j$ |
| $r_{ij}$ | The correlation between the regressor $X_i$ and $X_j$ |
| $\hat{b}_i$, $i=1,2,....,k$ | The standardized regression coefficients. |
| $w_i, i = 1,2,\dots,k$ | The new length standardized regression scaling (the independent variables importance in the model) |
| $SS_{LOF}$ | The lack-of-fit sum of squares. |
| $SS_{PE}$ | The pure-error sum of squares |
| $F_{LOF}$ | The Lack-of-fit statistical value |
| $df_{LOF}$ | The lack-of-fit degree of freedom. |
| $df_{PE}$ | The pure error degree of freedom. |
| $h_{ij}$ | The $ij^{th}$ element of the hat matrix H. |
| *VIF* | The variance inflation factor. |
| $\widehat{GT}$ | The predicted generator temperature. |
| $\bar{X}_{i_{old}}$, $i= 1,2,\dots,k$ | The mean value of the old data. |
| P | The degree of significant |
| $p$ | The regression degree of freedom$-1$ |
| $H$ | The hat matrix |
| $t_0$ | The significance of the regression coefficient's statistical value |
| $Q_h$ | The heat loss of the hot fluid |
| $Q_c$ | The heat gain of the cold fluid |
| $\dot{m}_h$ | The mas flow rate of the hot fluid |
| $\dot{m}_c$ | The mass flow rate of the cold fluid |
| $c_{p,h}$ | The specific heat of the hot fluid |
| $c_{p,h}$ | The specific heat of the cold fluid |
| $T_{h,i}$ | The inlet temperature of the hot fluid |
| $T_{c,i}$ | The inlet temperature of the cold fluid |
| $T_{h,o}$ | The outlet temperature of the hot fluid |
| $T_{c,o}$ | The outlet temperature of the cold fluid |
| *LMTD* | The logarithmic average of the temperature difference |
| $A$ | The area pf the wind generator's heat exchanger |
| $U$ | The overall heat transfer coefficient |
| $h_h$ | The average internal and external heat coefficients of the hot fluid |
| $h_c$ | The average internal and external heat coefficients of the cold fluid |