International Journal on Advanced Science Engineering Information Technology

Vol.10 (2020) No. 5
ISSN: 2088-5334

Using the Generalized Partial Linear Regression Model to Determine Climatic Factor Effect on Dust Storms in Baghdad Governorate

Laith Fadhil S. H[a,1], Auday Taha R[a], Waleed Mohammed Elaibi[a]

[a] *College of Administration and Economics, University Mustansiriyah, Baghdad, Iraq*
E-mail: [1]Laith@uomustansiriyah.edu.iq

***Abstract*— The phenomenon of dust that occurs in Iraq is one of the phenomena that cannot be completely controlled or partially processed in a short time. It also works to know how some climatic factors such as the average wind speed, relative humidity, atmospheric pressure above sea level, and the maximum temperature which represent explanatory (independent) variables (X_1, X_2, X_3, X_4), respectively. The effects on the number of occurrences of dust storms represent the adopted variable (Y) in Baghdad Governorate for the period from 2008 to mid-2013. The researchers used the Generalized Partial Linear Regression Model (GPLRM) consist of fourteen models, after determining the best link function for each model. Then we compared these models to determine the best model that represents this data with the best representation using the Akaike information criterion (AIC), the Schwartz information criterion (BIC), and the determination Coefficient criterion $(\mathbf{R^2})$. We also used the program (i- xplore) in the calculation, and we have concluded that the best model is the model in which the variable $(\mathbf{X_2})$ relative humidity and $(\mathbf{X_4})$ the maximum temperature in the parametric part, i.e., linear and stable. On the other hand, the variable $(\mathbf{X_1})$ average wind speed and the variable$(\mathbf{X_3})$ atmospheric pressure above sea level are non-parametric and their behavior is non-linear and unstable. The researchers consider that Baghdad Governorate suffers from this negative phenomenon as well as in general in Iraq. Besides, the effect of the variable $(\mathbf{X_2})$ relative humidity is a decreasingly negative effect, while the effect of the variable $(\mathbf{X_4})$ maximum temperature, it is an increasingly positive effect.**

***Keywords*— generalized partial linear regression model; parametric model; non-parametric model; link function; dust storms.**

I. INTRODUCTION

The Generalized partial linear regression model is a semi-parametric model representing the extended state of the semi-parametric models. It integrates the parametric models and the non-parametric models. It also represents an extension of the concept of expanded linear regression (GLMz) and that the methods for estimating this model depend mainly on the methods of estimating the extended linear regression. This model is applied to the data of the number of dust storms to study the effect of the explanatory variables (average wind speed, relative humidity, atmospheric pressure above sea level, and maximum temperature) on the dependent variable (number of dust storms) [1], [2]. Several models are specified of (GPLRM) models, and several link functions are used for each model. The best link function is determined. Then the best model representing these data is best represented using the criteria BIC, AIC,R^2 as well as knowing which explanatory variables behave linearly and parametrically and any of the variables that behave non-linearly and non-parametrically [3], [4] .

II. MATERIAL AND METHOD

A. General Partial Linear Regression Model (GPLRM)

The generalized partial linear regression model (GPLRM) is the extension of the concept of extended linear regression (GLMz) and its relationship to the semi-parameter model and that its estimation methods depend mainly on the estimate of (GLMz). The mathematical formula for the (GPLRM) [5]–[7] is as follows:

$$E(Y/X, Z) = G\{X^T\beta + m(Z)\} \quad (1)$$

Where:

- Y: Response variable vector.
- $X^T\beta$: It represents a linear combination of the function of the part parametric.
- $m(Z)$: It represents the function of the part non-parametric.
- G(.): Link function.

The principle of estimating this model is based on two common methods used by researchers:

- Estimate of $\hat{\beta}$ with $m(.)$ known.
- Estimate $m(.)$ with $\hat{\beta}$ known.

Some studies have shown the expectation and variance of the dependent variable Y [9], [10] by the following formula:

$$E(Y/X, Z) = \mu = G(\eta) = G(X^T\beta + m(Z)) \quad (2)$$

$$Var(Y/X, Z) = \sigma^2 \, Var(\mu) \quad (3)$$

Where:

σ^2 : dispersion parameter.

Accordingly, the researchers here conduct some of the following estimation methods, due to the importance of the (GPLRM) among the semi-parametric models.

B. Estimation Method of General Partial Linear Regression Model

For the purpose of estimating the model in reference to equation (1), the researchers used the Profile Likelihood Method. It is based on the separation of the estimation problem into a parametric part and a non-parametric part. The Simpler Profile Likelihood Method is one of the methods for estimating the model according to its own conditions, which we will explain in more detail so that we can estimate the model.

C. Profile Likelihood Method

A previous study separated the estimation problem into a parametric part and a non-parametric part. Since the general greatest possible function is inaccurate in semi-parametric models, the basic method of estimation is determined by $\underline{\beta}$ and then estimated by $m_\beta(.)$ depending on $\underline{\beta}$ which is used in building the greatest possible special structure [10]–[12]

Assuming the parameter Likelihood function is:

$$L(\beta) = \sum_{i=1}^{n} L\left(m_{i,\beta} + Y\right)$$

$$m_{i,\beta} = G\left\{X^T\beta + m_{\underline{\beta}}(Z_i)\right\}$$

and the function in the equation is the one that is maximized to estimate β ,and that the paved Likelihood function

$$L^H{}_{m_{\underline{\beta}}(Z)} = \sum k_H(Z - Z_i)L\left(m_{i,\beta}, Y\right)$$

where

$$m_{i,\beta} = G\left\{X^T\beta + m_{\underline{\beta}}(Z_i)\right\}$$

The function in the equation is the one that is maximized to be estimated as $m_{\underline{\beta}}(Z)$,and $k_H(Z - Z_i)$ represents the weight in the multivariate kernel function, and H represents a bandwidth matrix b.m. [13], [14].

If the maximized Likelihood partial function (Quasi − likelihood) or function Likelihood logarithmic variable Y and the valuated η is $\ell_i(\eta) = L\{G(\eta), y_i\}$. Thus, we expressed the first and second derivative of the function $\ell_i(\eta)$ for η with both $(\acute{\acute{\ell}}_i, \acute{\ell}_i)$ respectively by compensating $m_j = m_\beta(Z_i)$. Maximization of the Likelihood paved function in the equation is necessary for a solution $0 = \sum_{i=1}^{n} \acute{\ell}_i\,(X_i{}^T\beta + m_j)K_H\left(Z_i - Z_j\right)$, concerning m_j, whereas β maximizing the function likelihood parametric equation is necessary to resolve $0 = \sum_{i=1}^{n} \acute{\ell}_i\,(X_i{}^T\beta + m_j)(X_i + \acute{m}_i)$.

These two solutions lead to an estimate algorithm (GPLRM)

$$\beta^{new} = \beta^{old} - B^{-1}\sum_{i=1}^{n} \acute{\ell}_i\,(X_i{}^T\beta + m_j)\breve{X}$$

where B is the (Hessian) matrix and is described $B = \sum_{i=1}^{n} \acute{\acute{\ell}}_i\,(X_i{}^T\beta + m_j)\breve{X}_i X_i{}^T$, and $\breve{X}_j$ be described

$$\breve{X}_i = X_j + m_j{}'$$

$$= X_j - \frac{\sum_{i=1}^{n} \acute{\acute{\ell}}_i\,(X_i{}^T\beta + m_j)K_H\left(Z_i - Z_j\right)X_i}{\sum_{i=1}^{n} \acute{\acute{\ell}}_i\,(X_i{}^T\beta + m_j)K_H\left(Z_i - Z_j\right)}$$

$$m_j{}^{new} = m_j - \frac{\sum_{i=1}^{n} \acute{\ell}_i\,(X_i{}^T\beta + m_j)K_H\left(Z_i - Z_j\right)}{\sum_{i=1}^{n} \acute{\acute{\ell}}_i\,(X_i{}^T\beta + m_j)K_H\left(Z_i - Z_j\right)}$$

These formulas are complex, but when considering x as (Design Matrix), it can be rewritten β^{new} , in matrix form [15].

$$\beta^{new} = (\widetilde{X}^T W \breve{X})^{-1}\,\widetilde{X}^T W \breve{Z} \quad (4)$$

D. Simpler Profile Likelihood Method

When the linking function $G(.)$ in the form of (Identity), and the distribution of Y is normal, it has been shown by (Hardle, Muller, Magda) that a simplified method of estimation can be determined using the simpler profile likelihood method , where:

$$\ell'_i = -\frac{\left(Y_i - X_i^T - m_j\right)}{\sigma^2}$$

$$\ell''_i = -\frac{1}{\sigma^2}$$

$S = \dfrac{K_H(Z_i - Z_j)}{\sum K_H(Z_i - Z_j)}$, and for matrix paved non − parametric

$m_j^{new} = \dfrac{\sum\left(Y_i - X_i^T\right)K_H(Z_i - Z_j)}{\sum K_H(Z_i - Z_j)}$, then estimate for part non − parametric

$\underline{m}^{new} = S\left(\underline{Y} - X\underline{\beta}\right)$, and matrices forms

$\underline{m}^{new} = [m_1^{new} * m_2^{new} * \ldots * m_n^{new}]$, where

$\beta^{new} = \left(\widetilde{X}^T\widetilde{X}\right)^{-1}\widetilde{X}^T\widetilde{Y}$, and estimate for part parametric

where

$\widetilde{X} = (I - S)X$

$\widetilde{Y} = (I - S)Y$

and these represent estimates (Speckman) for model (PLM) going back to the style of estimate (GLM), in that, all steps of the iterative method can be found in a method (WLS) , then it can estimate (GPLRM) by weighted linear partial estimation method [16], [17]

The link function is the function that connects the first two parts, the first part of the random complex, which is represented by the random variable y, which has a distribution that belongs to the distributions of the exponential family. The second part is the regular complex which is represented by the formula of the arithmetic mean that is in the form of a linear structure known as a linear predictor in terms of explanatory variables [20], [21]. The link function is a monotonous function (ascending or descending), linear or nonlinear, and its form is as in equation (5).

$$\eta_i = G(\mu) = G(X^T\beta) \quad (5)$$

TABLE I
THE DISTRIBUTIONS OF THE EXPONENTIAL FAMILY - THE LINK FUNCTION TYPE

$\eta_i = G(\mu) = \mu$	**Identity**
$\eta_i = G(\mu) = \mathrm{Log}(\mu)$	**Log**
$\eta_i = G(\mu) = \log\frac{1}{1-\mu}$	**Logit**
$\eta_i = G(\mu) = \frac{1}{\mu}$	**Reciprocal**
$\eta_i = G(\mu) = \log(-\log(1-\mu))$	**Comp log − log**
$\eta_i = G(\mu) = \frac{1}{\mu^2}$	**Reciprocal^2**

E. Dust Storms

According to the introductory definition, low visibility is less than (1000 m) and the wind speed is more than (7 m / s), and dust storms can be defined from a geographical point of view as: a cloud of mobile soil with air in which the density of the dust increases so that the visibility range is less than (1 km) with wind speed (7 m / s) or more . Dust storms vary in intensity, size, intensity, and height. They range from (1-5500 m) and the distances covered by them ranging from tens of kilometers to thousands of kilometers, thus traveling across the continents and having the ability to carry large amounts of dust up to (4000 tons / mile) [18]. Dust also varies in terms of its composition and density, depending on the origin and the speed of it carrying wind. Dust pollutes the atmosphere and is harmful to public health, especially if the clay atoms are of a needle type. Likewise, it badly affects air, land, sea, and neighborhood transportation of all kinds and has many bad effects on other vital facilities [19]. Microns with several atoms are characterized by their relatively large size, where some diameters reach 100 microns and are present in the lower levels of the storm because of their relative gravity where the wind cannot carry them to higher levels when brown or gray depending on the color of dust carried by the wind during the storm. After a calm wind storm, dust turns to dust stuck to fading and then deposited on the Earth's surface by gravity [3].

III. RESULTS AND DISCUSSION

The data obtained is the number of times of the dust storms, the dependent variable, Y, and the explanatory variables, represent some climatic factors, which represent the average wind speed (X_1), relative humidity (X_2), atmospheric pressure above sea level (X_3) and the maximum temperature (X_4) These data were recorded and taken from the Weather Service Seismic monitoring - Climate section located in the Ministry of Transport and Communications building, which represent monthly data for the period from 2008 to mid-2013 at (66) views.

A. Building (GPLRM) in Case One Variable Non-Parametric

The non-parametric part consists of only one explanatory variable, while the rest of the explanatory variables exhibit a linear behavior, and the parameter part. And according to the equation (1) we will have four models of generalized partial linear regression models, and the beam width for each of these explanatory variables is as follows:

TABLE II
THE BANDWIDTH OF EACH EXPLANATORY VARIABLE

explanatory variable (x's)	**Bandwidth**
(X_1) average wind speed	0.68249422
(X_2) relative humidity	17.908332
(X_3) atmospheric pressure above sea level	7.8191514
(X_4) maximum temperature	11.445098

After finding the Bandwidth (B.W.) for each explanatory variable (X_1, X_2, X_3, X_4) we will estimate each model using the link functions according to the following distributions:
Then we determine the link function that gives us the best estimate of the model, using the coefficient of determination (R^2) Akaik's information criterion (AIC) Schwarz's Bayesian information criterion (BIC) according to the following Table 3.

TABLE III
LINK FUNCTIONS FOR THE FOLLOWING DISTRIBUTIONS USED TO ESTIMATE THE GPLR MODELS WHEN WE HAVE ONE VARIABLE, NOT PARAMETRIC

parametric variables	**Link functions**	**Gaussian**	**Poisson**	**Gamma**	**Inverse Gaussian**	**Negative Binomial**
$\mathbf{X_1}$	(R^2)	0.7031	---	---	---	---
	(AIC)	377.3921	---	---	---	---
	(BIC)	390.291	---	---	---	---
$\mathbf{X_2}$	(R^2)	0.7109	0.266	0.0008	---	**0.0137**
	(AIC)	375.8568	380.1146	754.4513	---	**525.9445**
	(BIC)	388.9691	393.1231	767.2379	---	**538.769**
$\mathbf{X_3}$	(R^2)	0.695	---	---	---	---
	(AIC)	379.7863	---	---	---	---
	(BIC)	393.3289	---	---	---	---
$\mathbf{X_4}$	(R^2)	0.7112	---	---	---	---
	(AIC)	377.9291	9916.80	---	---	---
	(BIC)	393.2644	9932.109	---	---	---

The non-parametric part consists of two explanatory variables, non-linear behavior, whereas the other two explanatory variables will be linear and constitute the parameter part. According to equation (1) we will have six models of generalized partial linear regression models and the bandwidth for each of these explanatory variables are presented in Table 4. After finding the Bandwidth (B.W.) for each explanatory variable (X_1, X_2, X_3, X_4) we estimated each model using the link functions according to the distributions in Table 5. Then we determine the link function that gives us the best estimate of the model, using the coefficient of determination (R^2) Akaik's information criterion (AIC) Schwarz's Bayesian information criterion (BIC) according to the following table 5

TABLE IV
THE BANDWIDTH OF EACH EXPLANATORY VARIABLE

explanatory variable (x's)	**Bandwidth**
(X_1) average wind speed	0.78020278
(X_2) relative humidity	20.472159
(X_3) atmospheric pressure above sea level	8.9385719
(X_4) maximum temperature	13.083624

TABLE V
LINK FUNCTIONS FOR THE FOLLOWING DISTRIBUTIONS USED TO ESTIMATE THE GPLR MODELS WHEN THE TWO VARIABLES ARE NOT PARAMETRIC

parametric variables	**Link functions**	**Gaussian**	**Poisson**	**Gamma**	**Inverse Gaussian**	**Negative Binomial**
$\mathbf{X_1X_2}$	(R^2)	0.7139	0.2651	0.0008	---	**0.0136**
	(AIC)	375.352	380.702	753.2183	Nan	**526.1562**
	(BIC)	388.8623	393.8851	766.1919	Nan	**539.1597**
$\mathbf{X_1X_3}$	(R^2)	0.7071	---	---	---	---
	(AIC)	380.5229	---	---	---	---
	(BIC)	397.5759	---	---	---	---
$\mathbf{X_1X_4}$	(R^2)	0.7031	0.2595	0.0008	---	**0.0134**
	(AIC)	380.0667	385.5506	751.2303	---	**528.1408**
	(BIC)	395.7379	400.9524	766.2441	---	**543.2076**
$\mathbf{X_2X_3}$	(R^2)	0.7521	---	---	---	---
	(AIC)	376.5108	---	---	---	---
	(BIC)	400.6324	---	---	---	---
$\mathbf{X_2X_4}$	(R^2)	0.7623	0.2801	0.0008	---	**0.0144**
	(AIC)	374.9476	384.0618	764.867	---	**536.4391**
	(BIC)	400.2648	409.1621	789.4916	---	**561.1446**
$\mathbf{X_3X_4}$	(R^2)	0.7446	---	---	---	---
	(AIC)	379.0103	---	---	---	---
	(BIC)	403.6591	---	---	---	---

B. *Building* (GPLRM) in *Case Three Variables*

The non-parametric component consists of three explanatory variables that exhibit non-linear behavior, whereas the remaining explanatory variable exhibits a linear behavior of the parameter segment's component. According to equation (1) we will have four models of generalized partial linear regression models, and the bandwidth for each of these explanatory variables, as in Table 6. After finding the Bandwidth (B.W.) for each explanatory variable (X_1, X_2, X_3, X_4) we will estimate each model using the link functions according to the following distributions:

Then we determine the link function that gives us the best estimate of the model, using the coefficient of determination (R^2) Akaik's information criterion (AIC) Schwarz's Bayesian information criterion (BIC), as in Table 7.

TABLE VI
BANDWIDTH OF EACH EXPLANATORY VARIABLE

explanatory variable (x's)	**Bandwidth**
(X_1) average wind speed	0.85844717
(X_2) relative humidity	22.525256
(X_3) atmospheric pressure above sea level	9.8349967
(X_4) maximum temperature	14.395744

TABLE VII
LINK FUNCTIONS FOR DISTRIBUTIONS USED TO ESTIMATE THE GPLR MODELS WHEN THE THREE VARIABLES ARE NOT PARAMETRIC

parametric variables	**Link functions**	**Gaussian**	**Poisson**	**Gamma**	**Inverse Gaussian**	**Negative Binomial**
$\mathbf{X_1}$	(R^2)	0.7146	0.2633	0.0008	---	**0.0136**
	(AIC)	349.094	385.0658	753.4378	---	**529.4522**
	(BIC)	396.4331	402.0515	469.9756	---	**546.0514**
$\mathbf{X_2}$	(R^2)	0.7603	0.2808	0.0008	---	**0.0146**
	(AIC)	377.0251	385.0606	767.5061	---	**537.6034**
	(BIC)	403.8511	411.6314	793.5215	---	**563.698**
$\mathbf{X_3}$	(R^2)	0.7667	---	---	---	---
	(AIC)	378.4887	---	---	---	---
	(BIC)	408.5923	---	---	---	---
$\mathbf{X_4}$	(R^2)	0.7668	0.2814	0.0009	---	**0.0146**
	(AIC)	378.5821	387.7615	767.6433	---	**540.5878**
	(BIC)	408.6892	417.6254	796.9123	---	**569.9497**

From tables (2), (4) and (6), we find that the best model is when using the link function to distribute Gaussian, in other words the link function of type (Identity). The model has the lowest value of the Akaik's information criterion (AIC) and the lowest value for the Schwarz's Bayesian information criterion (BIC) and the highest proportion of the coefficient of determination (R^2) Compared to the rest of the functions.

C. Comparing Between Models

After determining the correlation function for the Gaussian distribution from tables (2), (4) and (6), we will determine the best model of the models that we obtained from these tables using the Akaik's information criterion (AIC) and the Schwarz's Bayesian information criterion (BIC) and the coefficient of determination (R^2) are presented in Table 8.

TABLE VIII

VALUES OF EACH OF THE GPLR MODELS FOR EACH OF THE GENERALIZED PARTIAL LINEAR REGRESSION MODELS

NO. of Models	Parametric component	Non-Parametric component	(R^2)	(AIC)	(BIC)
1	$\mathbf{X_1}$	$\mathbf{m(X_2X_3X_4)}$	0.7146	379.094	**396.4331**
2	$\mathbf{X_2}$	$\mathbf{m(X_1X_3X_4)}$	0.7603	377.0251	**403.8511**
3	$\mathbf{X_3}$	$\mathbf{m(X_1X_2X_4)}$	0.7667	378.4887	**408.5023**
4	$\mathbf{X_4}$	$\mathbf{m(X_1X_2X_3)}$	0.7668	378.408	**408.6892**
5	$\mathbf{X_1X_2}$	$\mathbf{m(X_3X_4)}$	0.7139	375.352	**388.6623**
6	$\mathbf{X_1\ X_3}$	$\mathbf{m(X_2X_4)}$	0.7071	380.5229	**397.5759**
7	$\mathbf{X_1\ X_4}$	$\mathbf{m(X_2X_3)}$	0.7031	380.0667	**395.7379**
8	$\mathbf{X_2\ X_3}$	$\mathbf{m(X_1X_4)}$	0.7521	376.5108	**400.6324**
9	$\mathbf{X_2\ X_4}$	$\mathbf{m(X_1X_3)}$	0.7623	374.9476	**400.2648**
10	$\mathbf{X_3\ X_4}$	$\mathbf{m(X_1X_2)}$	0.7446	379.0103	**403.6591**
11	$\mathbf{X_1\ X_2\ X_3}$	$\mathbf{m(X_4)}$	0.7031	377.3921	**390.291**
12	$\mathbf{X_1\ X_2\ X_4}$	$\mathbf{m(X_3)}$	0.7109	375.8568	**388.9691**
13	$\mathbf{X_1\ X_3\ X_4}$	$\mathbf{m(X_2)}$	0.695	379.7863	**393.3289**
14	$\mathbf{X_2\ X_3\ X_4}$	$\mathbf{m(X_1)}$	0.7112	377.9291	**393.109**

By comparing the three criteria AIC , BIC , R^2, as in Table 8, the researchers determined the best generalized partial linear regression model (GPLRM) as follows:

1) First: the Akaik's information criterion: the researchers notice that the ninth model is the best because it had the lowest value for the Akaik's information criterion and its value was AIC = $\boxed{374.9476}$. And this represents parametric component (X_2) relative humidity and variable (X_4) maximum temperature. Either that variables (X_1) average wind speed and (X_3) atmospheric pressure above sea level. They represent the non-parametric component, as shown in Figure 1.

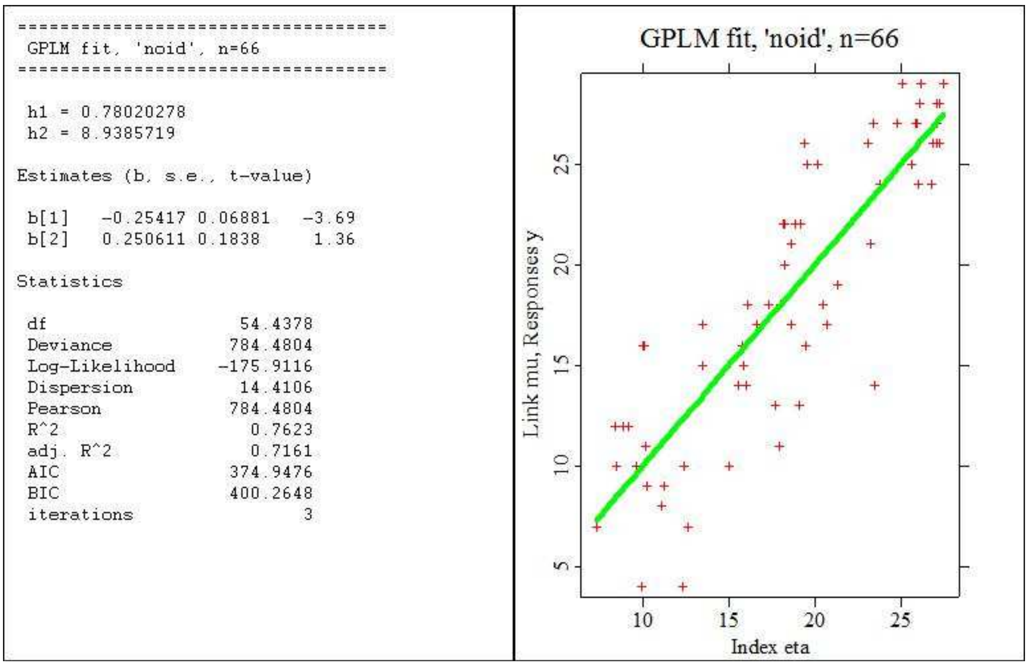

Fig. 1 GPLRM when the parameter component represents the second and fourth variables. As for the non-parametric component it consists of the first and third variables.

This model demonstrates from its parameter component that the variables achieve stability X_2 and X_4, so that the increase in one unit of the variable X_2 (relative humidity) will reduce the amount of dust concentrations by (0.25417) which is a negative and significant effect, and that the increase in one unit of the variable X_4 (Maximum temperature) will lead to an increase in the amount of polluted dust concentrations by (0.250611). Its unscientific component shows the instability of the variables X_1 (average wind speed) and X_3 (atmospheric pressure above sea level) and that their behavior is not linear.

The second came the fifth model, and the value of the AICA criterion reached AIC = $\boxed{375.352}$, whose component represents the variable X_1 and the variable X_2, while the variables X_3 and X_4 represent the non-parametric component as in figure 2.

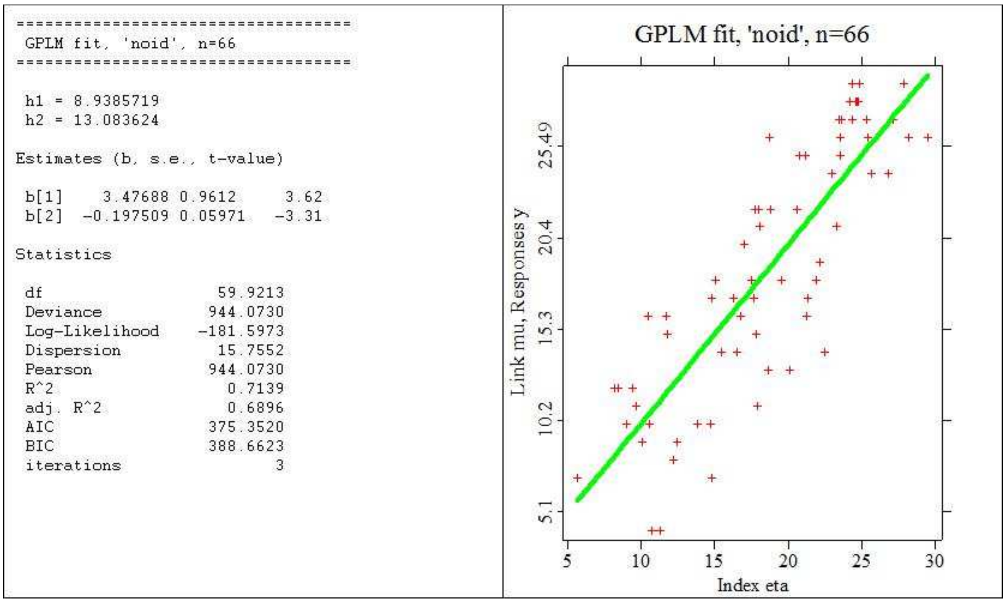

Fig.2 GPLRM when the parameter component represents the first and second variables. As for the non-parametric component, it consists of the third and fourth variables.

We note from its parameter component that stability verifies the variables X_1 and X_2, so that the increase by one unit of the variable X_1 will lead to an increase in the concentrations of polluted dust by (3.47688), which is a very big effect. While for the variable X_2, the increase by one unit will lead to a decrease in the amount of polluted dust concentrations to (0.197509), which negatively affects. As for its unscientific component, the variables in it X_3 and X_4 are non-linear and unstable.

2) *Second: the Schwarz's Bayesian information criterion:* We note that the fifth model is the best because it had the lowest value for the Schwartz criterion and its value was BIC = $\boxed{388.6623}$, then came in second place the twelfth model and the value of the Schwartz criterion BIC = $\boxed{388.9691}$ and its parameter component represents X_1 and X_2 and X_4 and the variable X_3 represents the non-parametric component as in figure No. (3):

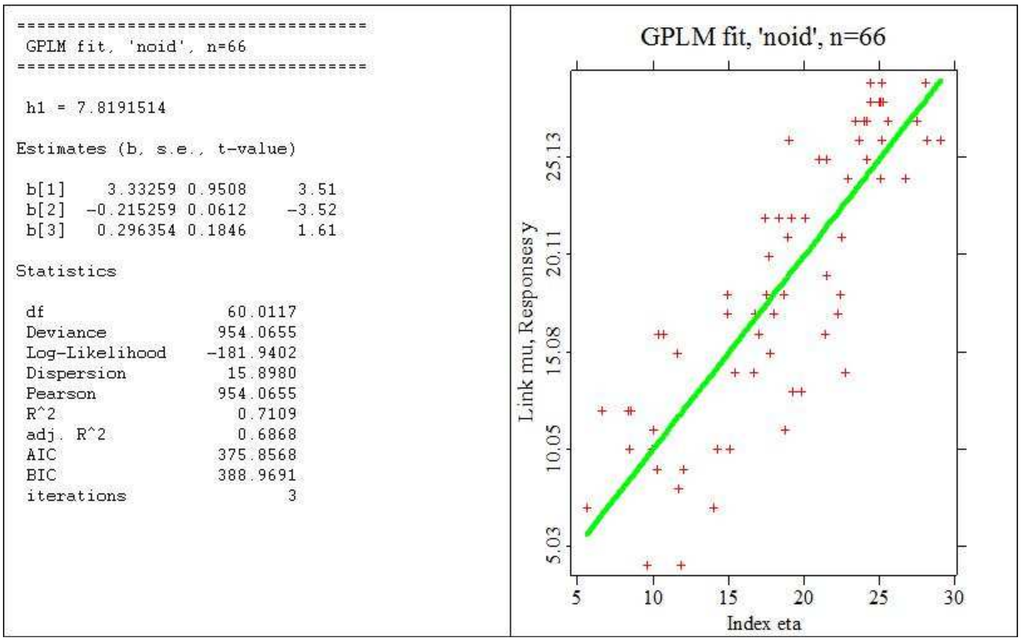

Fig. 3 GPLRM when it represents the first component parametric variables and the second and fourth either component Allamwalima consists of the third variable

This model demonstrates from its parameter component that stability is achieved in the variables X_1, X_2 and X_4, so that the increase in one unit of the variable X_1 will lead to an increase in dust concentrations by (3.33259), which is a very big effect. But for the variable X_2, increasing the intensity of one will lead to reducing the amount of concentrations of polluted dust by (0.215259) which is a negative effect. While increasing one unit of the X_4 variable will lead to an increase in the amount of polluted dust concentrations by (0.296354), which is a big effect, but its unscientific component shows the instability of the X_3 variable and its behavior is not linear.

3) Third: The determination coefficient criterion: We note that the fourth model is the best because it had the highest percentage of the determination coefficient and its value was $R^2 = \boxed{0.7668}$, whose parameter component consists of the variable X_4, while the rest of the variables X_1, X_2 and X_3 represent the non-parametric component as in figure. (4)

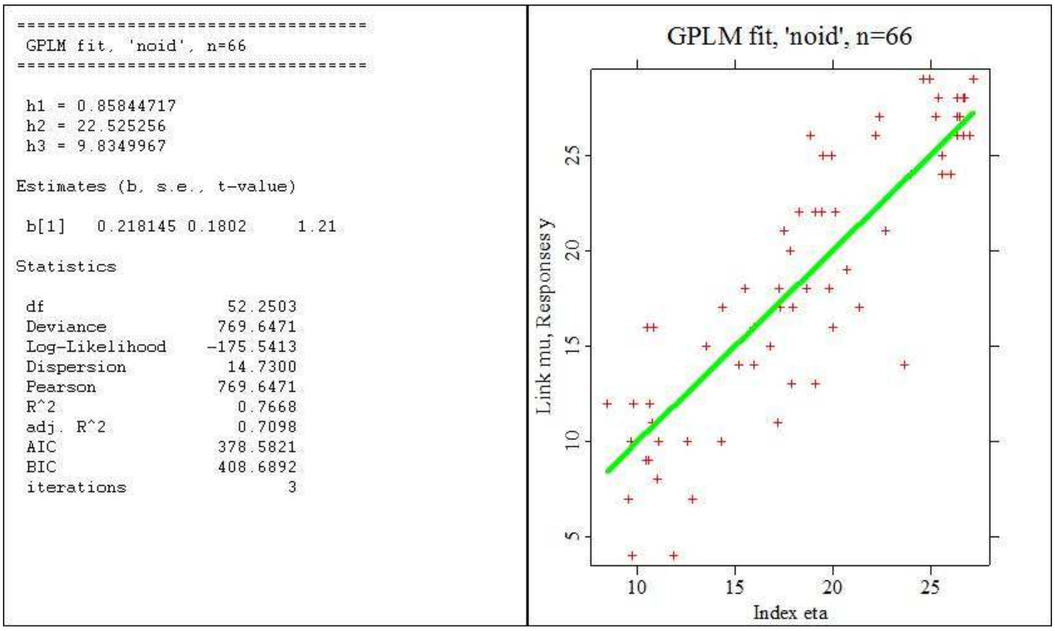

Fig. 4 (GPLRM) shows when the parameter component represents the fourth variable. As for the non-parameter component, it consists of the first, second and third variables

The model demonstrates from its parameter component that stability is achieved in variable X_4 and that increasing one unit of this variable will lead to an increase in the concentrations of polluted dust by (0.218145) which is a significant effect. Whereas its non-parametric component shows counting stability and non-linear behavior of the rest of the variables. The third model came in second place because it had the second highest proportion of the determination coefficient and reached $R^2 = \boxed{0.7667}$, whose parameter component represents the variable X_3. The variables remained X_1 and X_2 and X_4 component represents. This model shows from its parameter component that stability is achieved in the variable X_3 and that increasing one unit of this variable will lead to an increase in the concentrations of the amount of polluted dust by (0.0458546). To us, each model will be arranged according to its order of preference in relation to the standard and Table No. 9 clarifies this.

TABLE IX
ARRANGEMENT OF MODELS IN THE THREE STANDARDS R^2,AIC, BIC

Model	R^2	AIC	BIC
m3	2	9	13
m4	1	8	14
m5	**8**	**2**	1
m9	3	1	9
m12	10	3	2

From the comparison in table No. (9), we can determine that the (m5) model is the closest to the best model because it has the first (BIC), second (AIC) and eighth (R^2), where $R^2 = \boxed{0.7139}$.

IV. CONCLUSION

The best model is the model in which the behavior of the variable (X_2) relative humidity and the variable (X_4) the maximum temperature is a stable linear behavior in the parametric component and the variables (X_1) wind speed rate and (X_3) atmospheric pressure above sea level, their behavior is non-linear and independent in the non-parametric part. The mathematical formula of the model (13) is:

$$\hat{y} = g\big(-0.25417X_2 + 0.250611X_4 + m(X_1, X_3)\big)$$

From the model in the formula (13) we conclude that the variable (X_2) relative humidity has a decreasing negative effect, i.e., increasing one unit of it will lead to a decrease in the number of dust storms by (0.25417) units. The variable is the maximum temperature (X_4), then its effect is positive increasing and that increasing one unit from it will increase the number of dust storms by (0.250611) units. From the model in the formula (13) we conclude that the variable wind speed rate (X_1) is unstable, non-linear and non-parametric, as well as the variable (X_3) air pressure above sea level is unstable, non-linear and non-parametric, and it can be said that this case represents a negative problem that suffers from it Baghdad Governorate in particular, and Iraq in general.

By studying the number of dust storms as a variable dependent on the explanatory variables, the average wind speed (X_1), relative humidity (X_2), atmospheric pressure above sea level (X_3) and maximum temperature (X_4), we conclude that the lack of green belts and afforestation causes an increase in dust storms. The criteria that have been applied are considered very important criteria in the statistical analysis to compare the preference of the models, which are the Kaikai standard, the Schwartz criterion, and the determination factor.

REFERENCES

[1] Hardle, W., Muller. M, Sperlich, S & Werwatz, A, (2004)," Nonparametric and Semi parametric Models an Introduction ", Spring Edition.

[2] Hardle. W, Mammen E. and Muller, M., (1998),"Testing Parametric versus Semi parametric Modeling in Generalized Linear Models", Journal of the American Statistical Association, Vol.93, No.444, pp. (1461-1474).

[3] Liu, X., (2011) , " penalized variable selection for semiparametric regression models " , submitted in partial fulfillment of the requirements for the degree doctor of philosophy – University of Rochester , New York.

[4] Abdous .B, Kokonendj.C,Senga.T., (2012) ,"On Semi Parametric Regression for count explanatory variables", Journal of Statistical Planning and Inference, pp.(1537-1548).

[5] Akkus, O. , (2011) , " Xplore package for the popular parametric and semi-parametric single index models " , Journal of science ,vol.24 , No.4, pp.(753-762) .

[6] Al-Hussaini, Maryam Abdul-Hussein, (2014), "Building and applying linear mixed regression models in the environmental field", Master Thesis in Statistics, College of Administration and Economics, University of Baghdad.

[7] Al-sharot, Muhammad H. and Khanjar, Muhammad T., (2016) ,"Comparing some semi-parametric estimators using simulation" ,Republic of Iraq , College of Computer Science and Mathematics , Al-Qadisiyah University.

[8] Sharaf, H. K., Ishak, M. R., Sapuan, S. M., Yidris, N., & Fattahi, A. (2020). Experimental and numerical investigation of the mechanical behavior of full-scale wooden cross arm in the transmission towers in terms of load-deflection test. Journal of Materials Research and Technology, 9(4), 7937-7946.

[9] Sharaf, H. K., Ishak, M. R., Sapuan, S. M., & Yidris, N. (2020). Conceptual design of the cross-arm for the application in the transmission towers by using TRIZ–morphological chart–ANP methods. Journal of Materials Research and Technology, 9(4), 9182-9188.

[10] Donald W. K., Andrews, (1991), "Asymptotic normality of series estimators for Non parametric And Semi Parametric Regression Models" EconometricaVol.59, No.2, pp. (307-345).

[11] Engle, R. F., Granger, C.W. J., Rice, J. & Weiss, A., (1986) ,"Semi parametric estimators of the relation between weather and electricity "sales, Journal of the American Statistical Association, Vol. 81, No. 394, pp.(310-320).

[12] Eubank R.L., Whitney P., (1989),"Convergence Rates for Estimation in Certain partially linear models", journal of statistical planning and inference, Vol. 23, pp. (33-43), north Holland.

[13] Gao.Jiti, (1995), "Asymptotic Theory for Partly linear models Communications in Statistics- Theory and Methods", 24:8.

[14] German Aneiros & Alejandro Quintela, (2001)," Asymptotic properties in partial linear models under dependence", Socicdad de Estadlsticac Invcstigacin Operative Test, Vol. 10, No.2, pp. (333-355).

[15] Haggag. Magda, (2007), "On the Estimation of a Semi parametric Generalized Linear Model", Interstat.atat, Journals .net articles. 0707004, pdf.

[16] Hardle W and Huang L.S, (2013), "Analysis of Deviance in Generalized Partial Linear Models", SFB649, Humboldt-Universitat zu Berlin, Spandauer Strabe. 1, (D-10178), Berlin.

[17] Hardle W, and Liang, H. & Gao., J., (2000),"Partilly Linear Models", Heidelberg: Physica –Verlag.

[18] Muller, Marlene, (2000), "Semi parametric Extension GLMz", der Humboldt- Universitat zu Berlin.

[19] Qasem , Habeb ,S. ,(2018), " Comparison parametric & semi-parametric single index model With application " , Master Thesis, College of Administration and Economics ,Al-Mustansiriya University, Baghdad.

[20] Ruppert, D., Wand, M. and Carroll, R., (2003)," Semi parametric Regression", Cambridge University Press.

[21] Saleh, Tariq A., (2016) ."Some semi-parametric methods in estimating and choosing the variable for a single index model", PhD thesis, College of Administration and Economics, University of Baghdad.