Application of Robust Non-Parametric Methods in Modeling of a Geothermal Well Data

James Madegwa

Geothermal Development Company, P.O. Box 100746 – 00101, Nairobi

jmadegwa@gdc.co.ke

Keywords: Megawatt, Robust Regression.

ABSTRACT

To measure the output of a geothermal well, discharge tests are done for two to four months after completion of drilling and heating recovery. If Lip-pressure method is used for the discharge tests, the relevant data types includes; wellhead pressure, lip pressure and the weir height. This data exhibits skewness and excess kurtosis also known as heavy-tailed. This paper applies robust non-parametric estimation to fit these data. The model is known to be robust to outliers that characterize the data. Robustness signifies insensitivity to deviations from the strict model assumptions. The statistical characteristic of geothermal well data has been reviewed and a robust non-parametric model has been proposed that has been able to fit the data. A comparison between the robust method used and Ordinary Least Squares (OLS) method which follows the strict model assumptions has also been made with graphical illustrations. The robust method was able to fit the data and performed better than OLS method.

1. INTRODUCTION

In Kenya, Geothermal exploration dates back more than 40 years ago where it was recognized that the central Rift Valley could contain a geothermal energy resource, Heya (2002). In 1956 two wells were drilled with the second one X2 reaching a depth of 1035 meters was found to have a temperature of up to 2350C. It was not until 1970 when detailed exploration, with Government of Kenya (GoK) and United Nations Development Programme (UNDP) funding commenced investigations in Olkaria, Eburu and Bogoria. Olkaria was subsequently given a top priority after detailed review in 1972, Heya (2002).

During drilling there are several tests that take place to determine reservoir properties. When well drilling is completed, a wellhead is fitted at the top of the well to control the steam as well as enable other relevant tests to be carried. More important and critical test to this project is the discharge test which takes place between 2 – 4 months after completion of the drilling process. One of the basic tasks of a geothermal reservoir engineer is to measure the fluid flow from a discharging well and its energy content as well as to analyze the flow characteristics of the well. After 2 – 4 months heat up, the well is opened up and allowed to flow to the atmosphere. Geothermal high – temperature wells are usually discharged into a silencer which also acts as a steam – water separator at atmospheric pressure, Houssein (2008).

There are several methods applied to determine a geothermal well output using the discharge data, of interest is the lip pressure method which was used not only in calculating the output of wells. The lip pressure method is based on an empirical formula developed by Russel James (1962). The method is preferred to separator method (separates the steam-water flow into a flow of water and a flow of steam at the pressure of the separator) because it requires minimum hardware and instrumentation to obtain good results, Houssein (2008).

To use the lip pressure method, the steam – water mixture is discharged through an approximately sized pipe into a silencer or some other simple device to separate the steam and water phases at atmospheric pressure. Assuming that we have a fairly large amount of steam/ water mixture flowing at sonic velocity through an open – ended pipe to the atmosphere, the absolute pressure at the external end of the pipe is then proportional to the mass flow rate and enthalpy. The flow in geothermal wells is assumed to be isenthalpic (adiabatic), Houssein (2008). Water flow from the silencer is commonly measured by the weir – box method, Grant et al. (1982). Ofwona, Kipyego and Suwai (2011) reiterated that well MW-01 was discharged in May, 2011. In determining the flow characteristics of the well, James lip pressure method was used. Discharge pipes of varying sizes; 8", 6", 5", 4" and 3" were used during discharge test. From the discharge tests, well-head pressure (WHP), lip pressure (PC) and the weir height data is collected which are used in estimation of the well output.

The characteristics of the WHP, PC, weir and megawatts (MWe) deviate from the expected characteristic of a normal distribution. In a normal distribution the values of skew and kurtosis is 0 (i.e. the tails of the distribution are as they should be). If a distribution has values of skew or kurtosis above or below 0 then this indicates a deviation from normality.

Statistics					
		WHP	PC	Weir	MWe
N	Valid	138	138	138	138
	Missing	0	0	0	0
Skewness		-.429	-.327	1.137	-.667
Std. Error of Skewness		.206	.206	.206	.206

Kurtosis	-1.683	-1.876	3.748	-.523
Std. Error of Kurtosis	.410	.410	.410	.410

Table 1: Shows skewness and kurtosis values for Wellhead Pressure (WHP), Lip Pressure (PC), Weir height and Megwatts (MWe). WHP, PC and MWe are negatively skewed while weir is positively skewed. In the same way, WHP, PC and MWe are platykurtic while weir is leptokurtic which implies that the data is skewed and heavy - tailed (excess kurtosis).

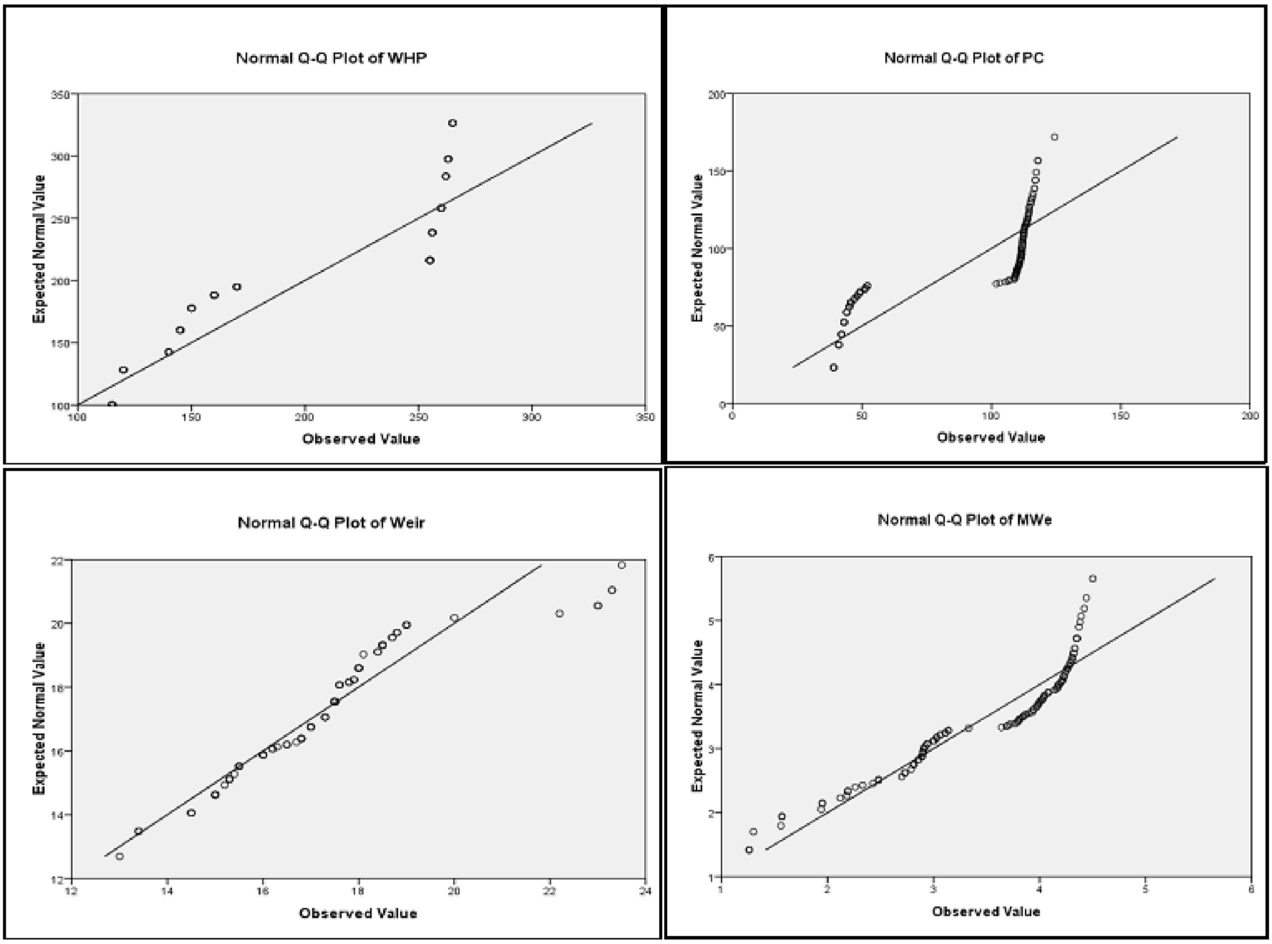

Figure 1: The Q-Q plot of WHP, PC, Weir and MWe data sets which shows high deviations from normality.

1.1 Relationship between Wellhead Pressure, Lip Pressure, Weir height and Megawatts.

1.1.1 Scatterplot

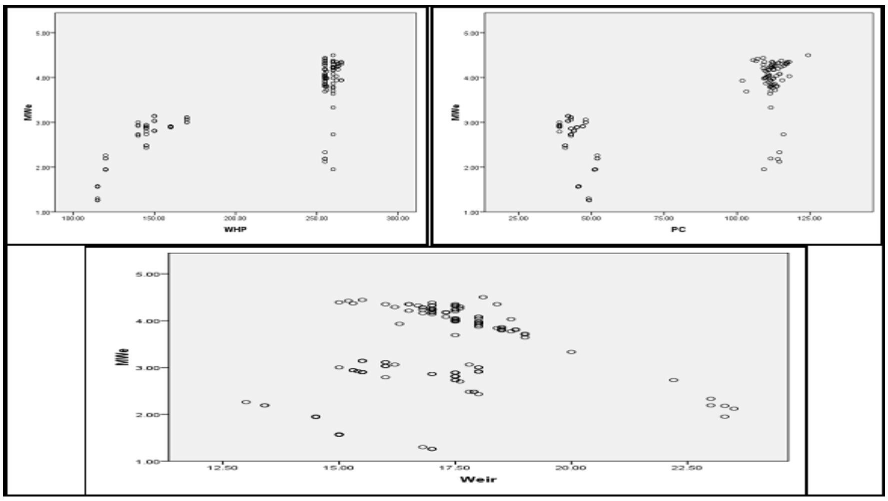

Figure 2: A scatterplot showing relationship between independent variables (WHP, PC, Weir) and the dependent variable (MWe).

As described by Andie (2009), a scatterplot is a graph that plots the score on one variable against the score on another. It provides information on the relationship between the variables, the kind of relationship and guides on whether there are any cases that are different from the others.

It can be seen from **Figure 2** above that it is difficult to precisely determine the relationship between the variables but it can generally be concluded that the relationship between all the three independent variables and the dependent variable is non-linear. The relationship between wellhead and lip pressure and megawatts are almost the same. Weir height presents a different scenario with the points distributed all over the graph.

1.1.2 Correlations

The table below provides the Pearson correlation coefficient for relationship between WHP, PC, Weir and MWe. There is significant relationship between WHP and PC with Megawatts while the relationship between Weir height and MWe is insignificant at the 0.01 level.

Correlations					
		WHP	PC	Weir	MWe
WHP	Pearson Correlation	1	.972**	.478**	.834**
	Sig. (2-tailed)		.000	.000	.000
	N	138	138	138	138
PC	Pearson Correlation	.972**	1	.446**	.756**
	Sig. (2-tailed)	.000		.000	.000
	N	138	138	138	138
Weir	Pearson Correlation	.478**	.446**	1	.056
	Sig. (2-tailed)	.000	.000		.517
	N	138	138	138	138
MWe	Pearson Correlation	.834**	.756**	.056	1
	Sig. (2-tailed)	.000	.000	.517	
	N	138	138	138	138
**. Correlation is significant at the 0.01 level (2-tailed).					

Table 2: Relationship between wellhead pressure, lip pressure, weir height and megawatts

2. LOCAL POLYNOMIAL REGRESSION

Local polynomial fitting is one of the local regression methodologies. Local regression is an approach to fitting curves and surfaces to data by smoothing i.e. it models a relation between a predictor variable and response variable. The general regression model is of the form;

$$y_i = m(x_{i,j}) + \in_i \tag{1}$$

where $m(x_i)$ is unknown function and $\in_i$ is an error term. Local polynomial smoothing approximates the unknown regression function $m(x_i)$ locally by a polynomial of order p , Fan and Gijbels (1992). If the $(p+1)th$ derivative of $m(x_i)$ at the point x_0 exists, then a Taylor expansion gives for x in a neighborhood of x_0 ,

$$m(x) \approx m(x_0) + m'(x_0)(x - x_0) + \frac{m''(x_0)}{2!}(x - x_0)^2 + ... + \frac{m^p(x_0)}{p!}(x - x_0)^p. \tag{2}$$

The above polynomial is then fitted locally by a weighted least squares regression problem by minimizing;

$$\sum_{i=1}^{n} \{Y_i - \sum_{j=1}^{p} \beta_j (X_i - x_0)^j\}^2 K_h(X_i - x_0), \tag{3}$$

where h is a bandwidth and K is a kernel function, $K_h(\cdot) = K(\cdot / h)/h.$

If we let X be the design matrix of problem (3) defined as: $X = \begin{bmatrix} 1(X_1 - x_0)\cdots(X_1 - x_0)^p \\ \vdots \quad \vdots \quad \vdots \quad \vdots \\ 1(X_n - x_0)\cdots(X_n - x_0)^p \end{bmatrix}$, If we let also $y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$, $\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_n \end{bmatrix}$ and W be the $n \times n$ diagonal matrix of weights; $W = diag\{K_h(X_i - x_0)\}.$

thus weighted least squares problem (3) can be written as

$$\min_{\beta} (y - X\beta)^T W (y - X\beta), \tag{4}$$

with $\beta = (\beta_0, \cdots, \beta_p)^T$. By weighted least squares theory the solution can be given as;

$$\hat{\beta} = (X^T W X)^{-1} X^T W y. \tag{5}$$

In local polynomial fitting, the selection of the bandwidth h, order of the local polynomial p and kernel function K is of critical importance, Fan & Gijbels (1992). They further explain that since the modeling bias is controlled by bandwidth, selection of the order of local polynomial is less crucial. In the same vein, they observed that local polynomial smoothing is highly influenced by extreme observations (outliers) in the response variables which results in quite different fitted curve and therefore need for a robust method.

3. ROBUST LOCALLY WEIGHTED SCATTER PLOT SMOOTHING (LOWESS)

Locally weighted scatter plot smoothing is a straight forward application of the local polynomial regression discussed above (section 2). The method robustifies the local polynomial procedure, see Cleveland (1979). Robustifying locally weighted regression involves four main steps, Cleveland (1979).

Let $0 < d \le 1$ and consider nd a quantity referring to a portion of the data. If nd is rounded to the nearest integer denoted by r then, for given observations $x_1, \cdots, x_n$, the neighborhood of a given observation X_k is determined by its associated bandwidth h_k which is the distance from X_k to its r^{th} nearest neighbor i.e. h_k is the r^{th} smallest number among $|X_k - X_j|$, for $j = 1, \cdots, n$, Fan and Gijbels (1992).

Let also W be a weight function bearing the following properties;

i. $W(x) > 0$ for $|x| < 1$;

ii. $W(-x) = W(x)$;

iii. $W(x)$ is a non-increasing function for $x \ge 0$;

iv. $W(x) = 0$ for $|x| \ge 1$

Then the first step involves doing a locally weighted regression by assigning the weight;

$$K_i(X_k) = K\{h_k^{-1}(X_i - X_k)\} \tag{6}$$

which has the properties as defined above to each observation X_i. These weights are then used in the initial locally weighted polynomial regression

$$\sum_{i=1}^{n} \{Y_i - \sum_{j=1}^{p} \beta_j (X_i - x_0)^j\}^2 h_k^{-1}(X_k), \tag{7}$$

resulting to the estimates $\hat{\beta}_j$; $j = 0, \cdots, p.$ The fitted value $\hat{Y}_k$ is then estimated. We have $\hat{Y}_k = \hat{\beta}_0 = \hat{\beta}_0(X_k)$ which can be written as

$$\hat{Y}_k = \sum_{i=1}^{n} w_i(X_k)Y_i,$$

by the fact that local polynomial regression estimators are linear smoothers. The coefficients $w_i(X_k)$ depends on X_k as well as design points $\{X_i\}$. This step is done for each observation X_k which results to the initial fitted values $\hat{Y}_1, \cdots, \hat{Y}_n$. From this initial fit, the residuals r_k are then estimated;

$$r_k = Y_k - \hat{Y}_k, k = 1, \cdots, n. \quad (8)$$

In the second step, the robustness weights $\delta_i, i = 1, \cdots, n$ are calculated. If we let B be the bisquare weight function defined by

$$B(x) = (1 - x^2)^2, \text{ for } |x < 1|$$

$$= 0 \text{, for } |x| \geq 1.$$

Again let S be the median of $|r_k|$, then the robust weights are given by

$$\delta_i = B(r_i/(6S))$$

The third step starts by computing, for each k, a new fitted value $\hat{Y}_k$ by using the weight $\delta_i K_i(X_k)$ for $(X_i, Y_i), i = 1, \cdots, n.$

In the last step, the second and third steps are repeated N times to get the final fitted values $\hat{Y}_k, k = 1, \cdots, n$ which yields the final robust curve. The iterative fitting carried out in steps 2 - 4 is to achieve robust smoothed points in which outliers does not affect the results, Cleveland (1979).

Cleveland (1979) further acknowledges that in order to carry out robust locally weighted regression there are four items that must be selected. These are: h, the bandwidth which determines the amount of smoothing; p, the order of the polynomial; W, the function that determines the weights; and t the number of iterations.

3.1 Selection of order of polynomial, number of iterations, weight function and the bandwidth

3.1.1 Choosing the Order of the Polynomial p

This determines the degree of the polynomial to be used in fitting the data. When $p = 1$ then a linear equation is fitted but when $p = 2$ quadratic equations are fitted, $p = 0$ implies that linearity or rather local linearity does not exists. For this research, $p = 2$ was used in fitting the data because it produces a smooth curve that follows the data to acceptable degree.

3.2.2 Choosing the Number of Iterations t

To achieve robust smoothed points, the procedure described in Section 3 above was repeated as many times as possible to achieve the best fit.

3.3.3 Choosing the Weight Function W

W bears characteristics as described in Section 3 above. It decreases smoothly to 0 as x moves from 0 to 1. Gross (1976) carried out investigations and showed that Turkey's biweight function performs well for robust estimation of location parameter and for robust regression.

3.4.4 Selection of the Bandwidth h

Taking very small bandwidth results in a small approximation error (bias) while if it is large it creates a large modeling bias. In other words the bandwidth determines the amount of smoothing applied in estimation of $m(x)$; the fitted curve becomes smoother with larger values of h. The main aim in choosing of h is picking a value as large as possible to minimize the variability in the smoothed points without distorting the pattern in the data. Jacoby (2000) explains that the smoothing parameter is specified as a value between 0 and 1.

3.2 Results and findings

3.2.1 Optimal bandwidth

Bandwidth	Standard Error	Bias
0.01	12.20789	-3.385072
0.02	3.519923	-1.134348
0.03	0.5870768	-0.2062319
0.04	0.3433087	-0.122174
0.05	0.3433087	-0.1221707
0.06	4.015572e-07	-8.365687e-06
0.07	1.520978e-07	7.176888e-06
0.08	1.622617e-06	-1.519279e-06
0.09	8.001929e-06	0.0001170377
0.10	1.398312e-05	-0.0001202426
0.20	0.0003349328	0.0001960178
0.30	0.0005596501	0.000482664
0.75	0.001604839	0.0004933345

Table 3: **The table above provides the standard error and bias for various bandwidths**

The results in the table above shows that using the Gaussian kernel function the optimal bandwidth is 0.06 with the lowest standard error of 4.015572e-07. Increasing the bandwidth leads to over smoothing while decreasing the bandwidth leads to under smoothing which is manifested by the increasing standard errors below and above the optimal bandwidth. It can be deduced from the figures above that, large bandwidths reduce the variance by smoothing over a large number of points, leading to increased bias. In contrast, small bandwidths give higher variance but have less bias hence lower bias leads to closer match between estimated values and actual observed values which can be shown by closer matching values of estimated MWe values with bandwidth of 0.06 and the true MWe values.

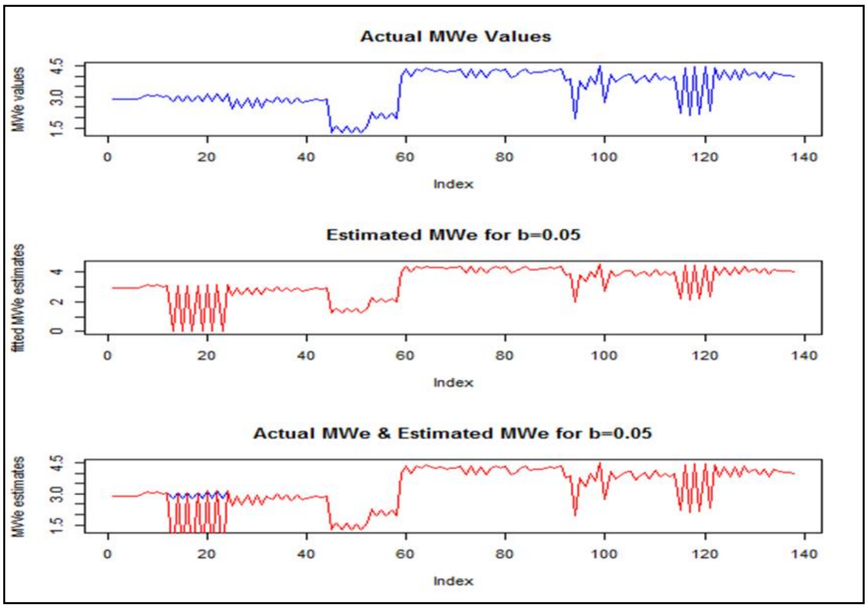

Figure 3: **Regression lines for the actual megawatts and the fitted megawatts for bandwidth = 0.05**

For bandwidth of 0.05 the estimated Megawatts values deviated from the 13th element to the 23rd element with the estimated values obtained as zero hence the high standard error.

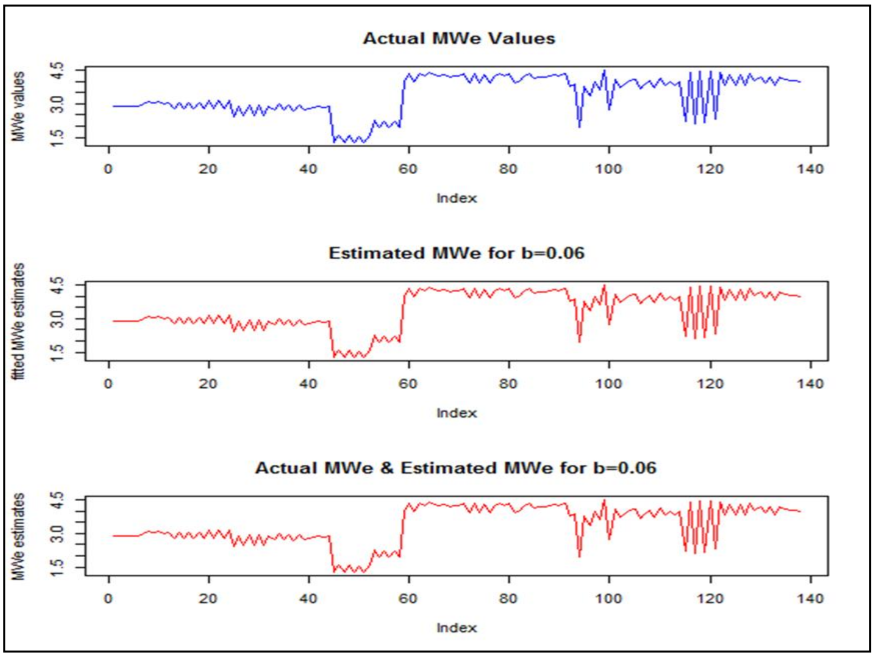

Figure 4: **Regression lines for the actual megawatts and the fitted megawatts for bandwidth = 0.06**

For bandwidth of 0.06 the estimated values were relatively very close to the actual values hence the similarity in the graphs between the actual and estimated values. For a bandwidth of 0.07, 0.08, 0.09, 0.10 the graphs were also nearly similarly with the actual graph except with an increase in the standard error hence indicating that increasing the bandwidth from 0.05 leads to over smoothing and an increase in standard error.

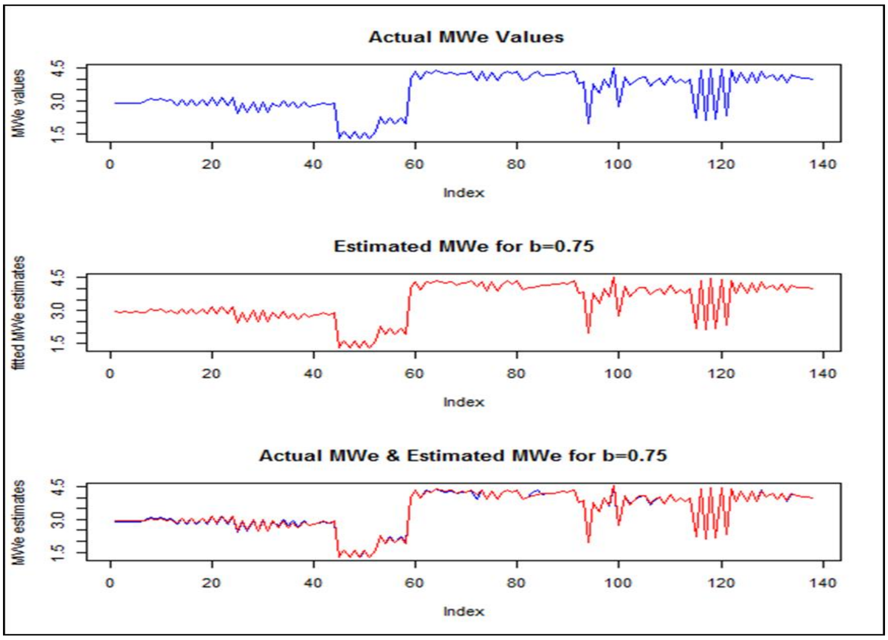

Figure 5: For bandwidth of 0.20, 0.30 and 0.75 the standard errors significantly increased and the disparities between the observed and fitted values was more pronounced as shown in the figure above.

3.2.2 Comparison with ordinary least squares (OLS)

Locally weighted scatter plot smoothing performs better than the ordinary least squares (OLS) at a bandwidth of 0.06. The former produces a lower error (4.015572e-07) compared to ordinary least squares (0.06029956). These results show that the robust method produce better estimates of predicted values.

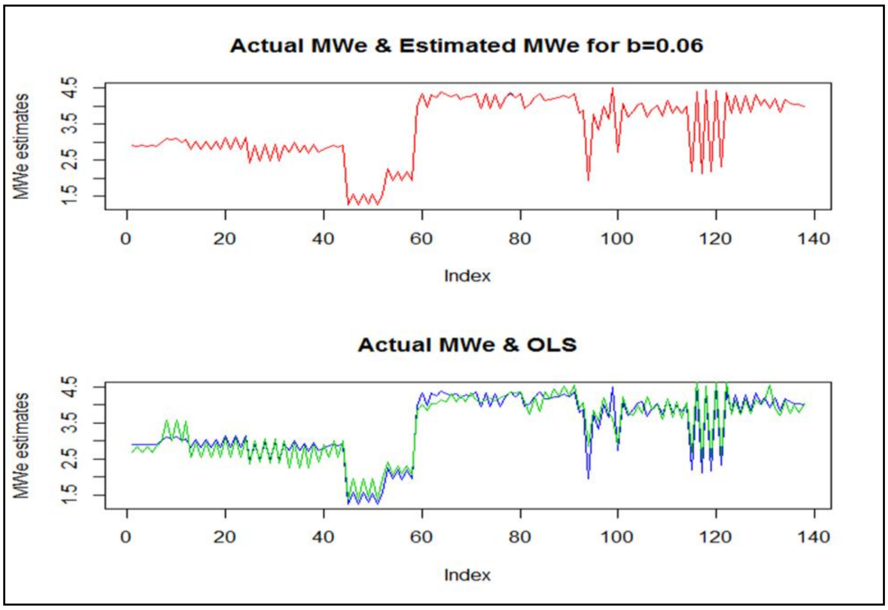

Figure 6: **Graph comparing robust regression and ordinary least squares regression.**

From the graph it is seen that while the actual values of Megawatts (MWe) fitted nearly perfectly with the estimated values for bandwidth of 0.06, on the other hand the estimated values for the OLS deviated so much from the actual values as shown by the green lines. While it's easy to see the difference between the actual MWe values and OLS estimates, it's not easy to distinguish the actual MWe values with those obtained by the smoothing function since the two graphs superimpose to near exact of another as shown in **Figure 6** below.

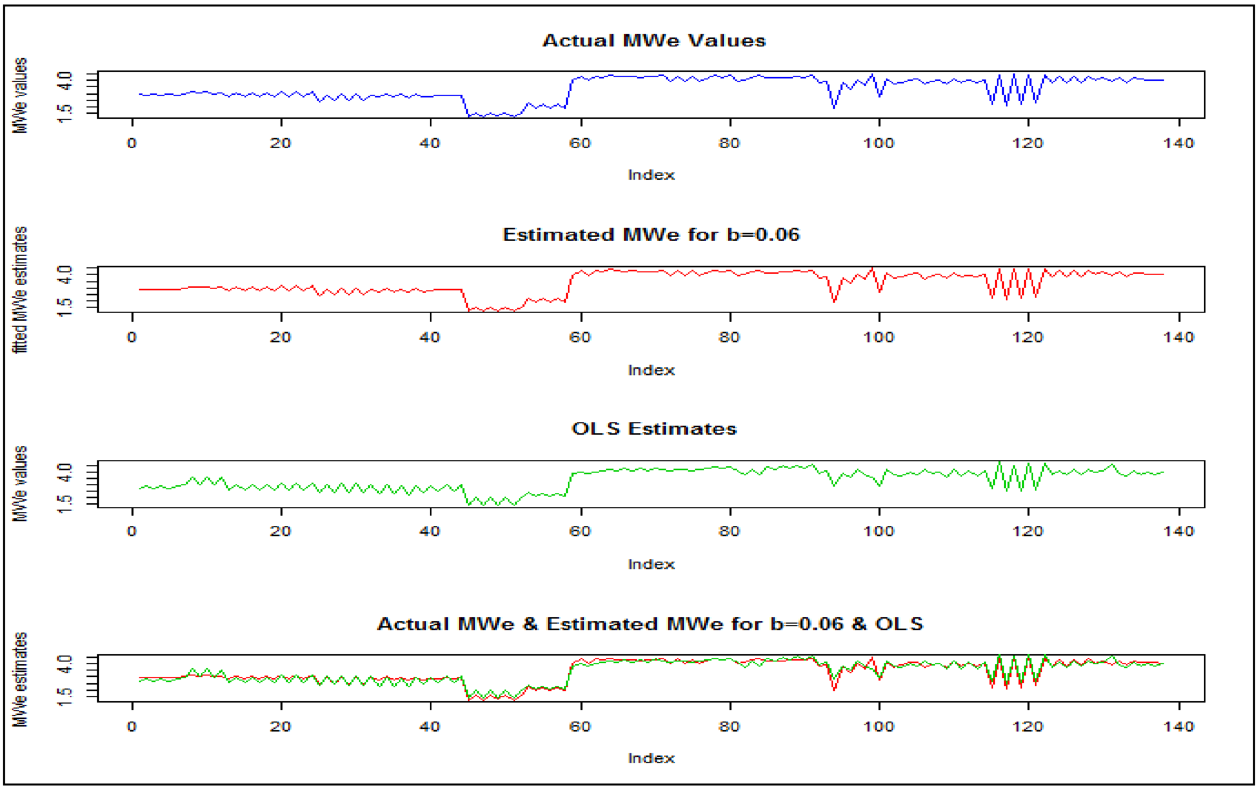

Figure 7: Graph showing comparison between fitted well data using the OLS method and robust LOWESS smoothing with bandwidth of 0.06.

5. CONCLUSSION

From the results it can be concluded that robust multivariate locally weighted scatter plot smoothing technique using the Gaussian kernel with a bandwidth of 0.06 is able to model the wells data. The smoothing technique performs better than the ordinary least squares method which follows strict normality assumptions.

REFERENCES

Fan, J., and Gijbels, I.: Local polynomial modeling and its applications; Monographs on statistics and applied probability. Chapman & Hall, London (1996).

Cleveland, S.: Robust Locally Weighted Regression and Smoothing Scatterplots. Journal of the American Statistical Association, **74**, (1979), 829-836.

Cleveland, S., and Devlin, J.: Locally weighted regression: An approach to regression analysis by local fitting. Journal of the American Statistical Association, **83**, (1988), 596-610.

Cleveland, S., and Grosse, E.: Computational Methods for Local Regression. Statistics and Computing, **1**, (1191), 47-62.

Hardle, W., and Marron, S.: Optimal bandwidth selection in non-parametric regression function estimation. The Annals of Statistics, **13**, (1985), 1466 - 1481.

Meseret, Z.: Overview of geothermal resource exploration and development in the East African Rift System, *Proceedings,* 4th African Rift Geothermal Conference, UN Headquarters, Nairobi, Kenya (2012)

Houssein, E.: Geothermal resource assessment through well testing and production response modeling, United Nations University, Iceland (2008).