TESTING SIMPLE REGRESSION MODEL FOR COORDINATE TRANFORMATION BY COMPARING ITS PREDICTIVE RESULT FOR TWO REGIONS

Odutola Christian Amans[1], Wu Beiping[2], Yao Yevenyo Ziggah[3]

Department of Geodesy and Surveying Engineering, China University of Geosciences, Wuhan, CHINA.

[1] obydukwe@yahoo.co.uk, [2] 964048773@qq.com

ABSTRACT

Coordinate Transformation between different datum usually involve complicated equation and computational analysis that is time consuming. Moreover, systematic errors are inevitable in the course of the transformation process thereby leading to accuracy problem. The theoretical mathematical models are methods most surveyors are versatile with. As more non professional surveying discipline is being introduced to GPS/GIS technology, the need for simple coordinate transformation from global WGS84 to local planimetric coordinates is absolutely important to guarantee local positioning accuracy. The research has been able to compare the validity of regression models developed from random distribution of GPS points in two regions. The points distribution in the first region are relatively close with an average distance less than 350m while for the second region are relatively far apart with an average distance of at least 850m. The model predictive values from the region, whose random point distributions are close, are far better with minimal errors compared with theoretical model. While the result from region whose GPS points are far apart was unsatisfactory, pointing out factor to be considered when developing a regression model for coordinate transformation.

Keywords: Coordinate transformation, global WGS84, local planimetric coordinate

INTRODUCTION

Coordinate transformation has been a great challenge to non professional surveyors that regard Global Positioning System (GPS) device as inevitable tool in their daily routine activities. Relating a position in the GPS system to the same position in a local map system requires a coordinate transformation model. Several transformation procedures and relationships have been put forth by researchers in transforming global coordinates to cartesian coordinates and vice versa (Jijie,1994; Ralph,1995; Gerdan *et al*.,1990; Burtch,2006; Chanfang *et al*.,2010; Gullu *et al*.,2011; Pinar *et al*.,2011; Soler *et al*.,2012; Featherstone,1997).

Transformations between various coordinate systems involve not only complex, usually non-linear, algebraic formulas, but also some very specific numerical parameters which have been established officially as national and continental geodetic datum. Unfortunately, some transformation formulas for specific areas are based on estimation and inherently of errors; and therefore, positioning accuracy is degraded as such errors are accumulated during the transformation process. The more mathematical steps involve in coordinate transformations, the more error prone may be the desired result (Chih-Hung *et al.,* 2007).

Traditionally GPS global coordinates are transformed to their concerned local datum using different theoretical methods such as; Abridged Molodensky Transformation Model, Bursa-Wolf Model, Veis Transformation model, Molodensky-Badekas Transformation Model, Position dependent transformations, Projective Transformation, Two-Dimensional (2D) Projective Transformation Model, Three-Dimensional (3D) Polynomial Projective Transformation Model, Twelve-Parameter Linear Affine Transformation Model and many others details of this can be found in Yao Yenenyo Ziggah,2013. Instead of using the complex algorithms involve in this transformation models by non professionals, it is proposed that the coordinate transformation model derived from simple regression model can serve as an alternative in predicting Cartesian planimetric (2-D) coordinates for the involved regions in the study area.

Linear regression model is a technique for modeling and analyzing several variables. It models data using linear predictor function and estimates unknown model parameters from data. The statistical technique involve are found on different assumption such as; identification of the properties of parameters of the derived model, application of standard testing procedure and predicting variables of interest for given values of the explanatory variables . It simplifies transformation process and still holds the positioning accuracy. This method of coordinate transformation was carried out by Gomaa *et al.,* 2011, where they affirmed its application simplicity. The accuracy of their outcome was within 0.3m for their study area, which is fair for geomatics activities, for example GIS data collection. This no doubt will ease datum related research work, hence the need to test its performance in the study area.

Study Area

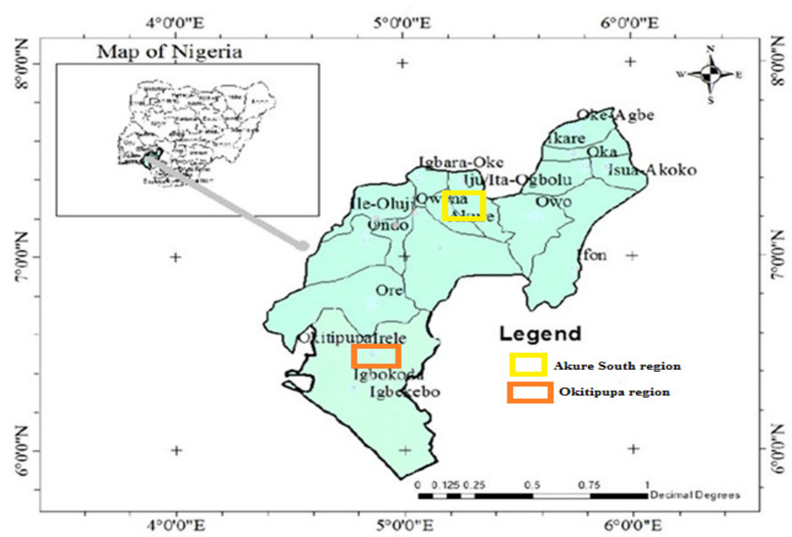


Figure 1

The study areas are small portion of about 19km^2 in Okitipupa and 56km^2 in Akure area of Ondo State, Nigeria as shown in figure 1 above. Okitipupa Local Government Area lies between Longitudes 4^0 31' and 4° 55' East of Greenwich Meridian and latitudes 6^0 48' and 6^0 28' North of the Equator , while Akure South Local Government Area lies between Longitudes 5^0 .00' and 5° 26' East of Greenwich Meridian and latitudes 7^0 10' and 7^0 16' North of the Equator.

Materials

The secondary source of data is the co-ordinates from the Co-ordinate Register of the Survey Department, Ministry of Lands, Housing and Environment, Akure, Ondo State in Nigeria. The two study area datum depends on Minna Datum of Nigeria that is based on Clarke 1880 ellipsoid utilizing the Universal Transverse Merecator (UTM) as a map projection. A dataset of forty two GPS stations each, with known precise WGS84 and local planimetric coordinates has been compiled and utilized in this research. The GPS points distribution in portion of Okitipupa area are relatively close to each other with an average distance less than 350m while those of Akure region are sparsely distributed with an average distance of at least 850m as shown in figure 2 and 3. They were obtained using differential GPS technique, as this is important to accommodate more number of parameters in the study area to enable extrapolation.

Plotted control points, data structures, descriptive and summary statistics for the entire study were produced using Matlab (2012b), Microsoft Excel 2010, and SPSS (Version 20). Then Arc GIS 10.0 was used to verify the degree of closeness of the data distribution.

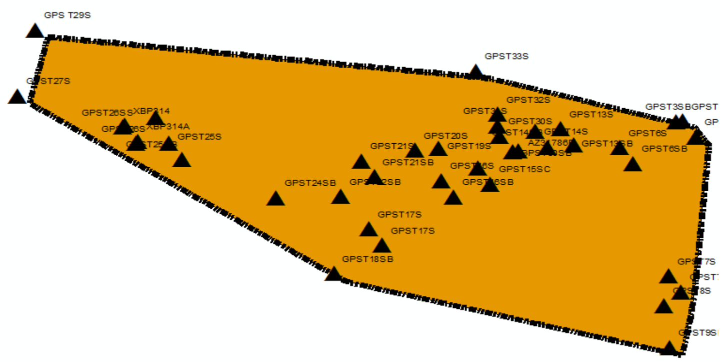


Figure 2. Portion of Okitipupa area

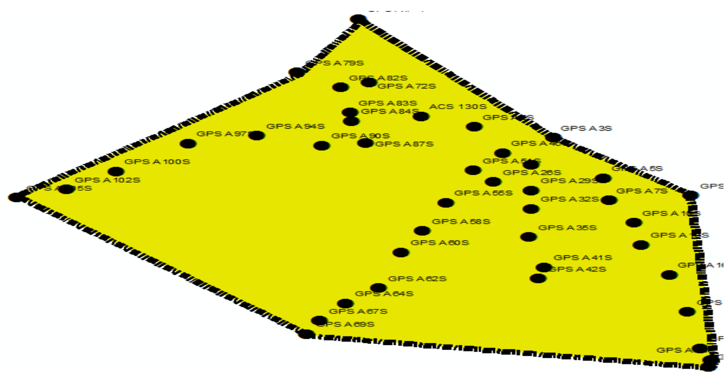


Figure 3. Portion of Akure area

DATA PROCESSING AND ANALYSIS

In estimating the simple transformation model using regression analysis, the primary goal is to determine the value of the regression function that is associated with a specific combination of predictor variable values of the local coordinates. The estimated values are computed by plugging the value(s) of WGS 84 $(\phi,\lambda,)$ direct coordinates into the regression model equations. The WGS 84 global geodetic coordinates $(\phi,\lambda,)$ must first be converted to decimal figures using matlab before plotting against the local cartesian coordinate (Northing and Easting). Finding the estimate of a parameter in a regression model is regression analysis collection of statistical tools. Regression models derivation can be found in many statistical literatures; example is Montgomery et at.,2011. The model is in the form $y_i = \beta_0 + \beta_1 x_i$; where y_i is the estimated coordinates (Northings / Eastings) in meters, x_i is the latitude /

ISSN-L: 2223-9553, ISSN: 2223-9944
Vol. 4 No. 6 November 2013

Longitude, β_0 is a constant, β_1 is the regression model coefficient (slope). German scientist Karl Gauss (1777–1855) proposed estimating the parameters β_0 and β_1 to minimize the sum of the squares of the vertical deviations. This approach of estimating the regression coefficients is the method of least squares.

Regression Model Sample Data

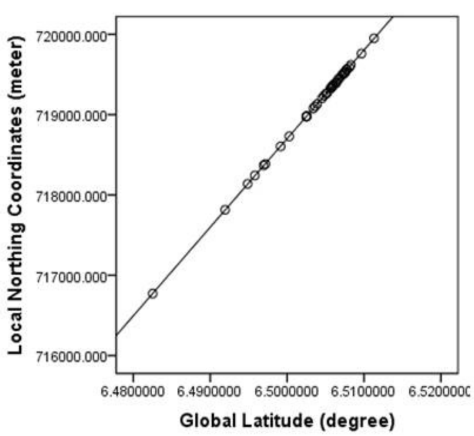
Table 1. Sample Data for Okitipupa region

	latitude WGS84			*Local Plan Coord*	*longitude WGS84*			*Local Plan Coord*
Point ID	*D*	*M*	*S*	*N*	*D*	*M*	*S*	*E*
GPST29S	6	30	51.364	720268.696	4	44	55.57172	693464.509
GPST33S	6	30	40.657	719950.597	4	46	36.49791	696566.712
GPST2S	6	30	27.929	719564.707	4	47	23.96435	698026.569
GPST2SB	6	30	23.854	719439.825	4	47	26.88738	698116.829
GPST7SB	6	29	44.900	718242.717	4	47	23.41094	698014.242

Table 2. Sample Data for Akure region

	Latitude WGS84			*Local Plan Coord*	*Longitude WGS84*			*Local Plan Coord*
Point ID	*D*	*M*	*S*	*N*	*D*	*M*	*S*	*E*
GPS A.F 1	7	17	33.379	806549.877	5	9	54.632	739138.232
GPS A3S	7	15	5.709	802032.516	5	12	12.771	743398.883
GPSXSO14B	7	13	53.607	799950.705	5	13	50.568	746321.799
GPS A22S	7	10	30.356	793587.453	5	14	4.019	746854.172
GPS A23S	7	10	22.032	793331.317	5	14	1.742	746785.533

Plotted Diagram of the Research Data for Both Regions



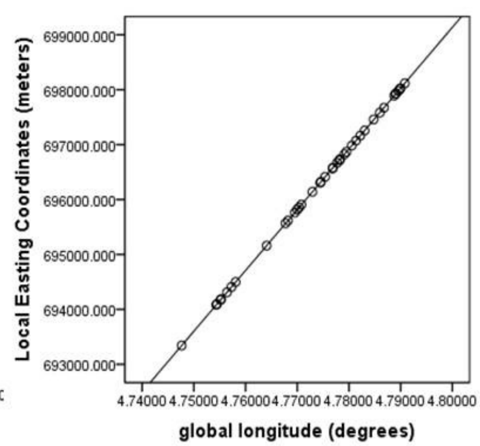


Figure 4. Plotted Data for portion of Okitipupa area

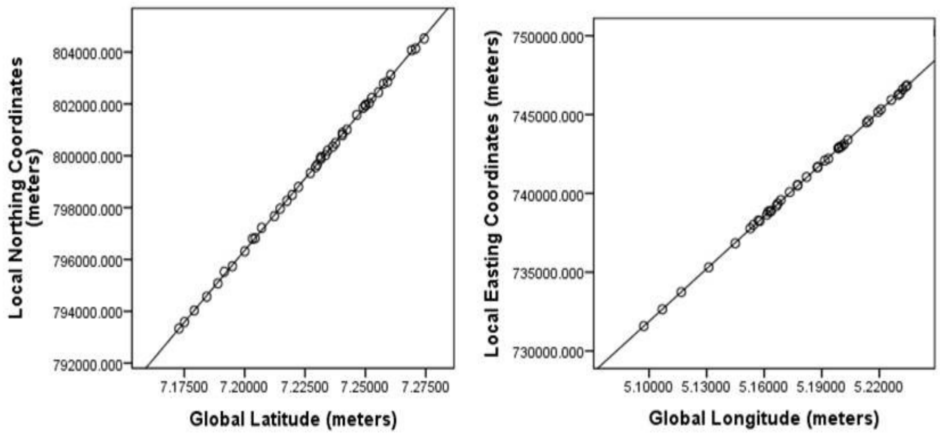


Figure 5. Plotted Data for portion of Akure area

Graph Analysis

The above graph illustrations show good positive correlation between local cartentian coordinate and WGS84 global coordinates in both regions, but the correlation is perfect in portion of okitipupa region as shown in the proportion of variation in the dependent variable (goodness of fit) (**R^2**) and correlation coefficient **R** between the observed and the predicted values of the dependent variables in tables 1, which is an indication of the models compliant. R^2 according to ANOVA partition forces is $0 \leq R^2 \leq 1$.

Table 3. Coefficient of determinant R^2 for both regions observation

Okitipupa		*Akure South*	
Local Northing coordinate against Global Latitude	Local Easting coordinate against Global Longitude	Local Northing coordinate against Global Latitude	Local Easting coordinate against Global longitude
1.000	1.000	0.9997	0.9999

Our large value of **R^2** in table's 3 suggests that the model has been successful in explaining the variability in the response. In finding the equation of the regression line, we assume that each observation y, can be described by the model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon \qquad i = 1, \quad 2, \ldots\ldots, n$$

Where $e_i = y_i - \hat{y}_i$ is called the *residual*. We will use the residuals to provide information about the adequacy of the two regions fitted models. However the estimated regression line is $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ with the derived regression fitted models for both regions displayed in tables 4 and 5. Where:

$\hat{y}$ = Estimated local Northing and Easting planimetric coordinate.

$\hat{\beta}_0$ = Value of the dependent variable when the independent variable are all zero (constant value).

$\hat{\beta}_1$ = Coefficient of independent variable (Slope).

x = WGS84 global coordinates.

The Developed Models

Table 4. Regression Model from 42 points that are at most 350m apart in Okitipupa region

Graph	*Regression Model (Expression equation)*
Y- Local Northing coordinate against Global Latitude	NORTHINGS COORD = 110315.589(Latitude $_{WGS84}$) + 1650.418
X- Local Easting coordinate against Global Longitude	EASTINGS COORD = 110682(Longitude $_{WGS84}$) + 167861

Table 5. Regression Model from 42 points that are at least 850m apart in Akure region

Graph	*Regression Model*
Y- Local Northing coordinate against Global Latitude	NORTHINGS COORD = 110956.228(Latitude $_{WGS84}$) – 2535.622
X- Local Easting coordinate against Global Longitude	EASTINGS COORD = 111164.349(Longitude $_{WGS84}$) + 164944.314

Checking the Adequacy of the Developed Regression Models

Model adequacy explains the basis on which the developed model between global and planimetric coordinate system can be accepted for prediction in the study areas.

Confidence interval values on the model parameter; slope β_1

This range of value has a 95% chance of including the population value of the regression coefficient, under the assumption that the GPS observations are normally and independently distributed. *A level* $(1-\alpha)$ confidence interval for β_1 in both regions is given by;

$$\hat{\beta}_1 - t * SE(\hat{\beta}_1) \leq \beta_1 \leq \hat{\beta}_1 + t * SE(\hat{\beta}_1)$$

Where: $t* = t_{\alpha/2,\ n-2}$

$SE =$ Standard error

A similar test can be performed for β_0, but this is rarely of interest, so there is no need. However,

The interpretation of the confidence interval is that we are 95% confident that the true slope β_1 of the regression line is between the values displayed on table 6 and 7.

Table 6. Confidence interval for the slope of Okitipupa region

Okitipupa Region	
Y- Local Northing coordinate against Global Latitude	X- Local Easting coordinate against Global Longitude
$110051.689 \leq \beta_1 \leq 110579.489$	$110619.011 \leq \beta_1 \leq 110726.265$

Table 7. Confidence interval for the slope of Akure regio

Akure Region	
Y- Local Northing coordinate against Global Latitude	X- Local Easting coordinate against Global Longitude
$110335.910 \leq \beta_1 \leq 111576.547$	$110759.634 \leq \beta_1 \leq 111569.065$

Hypothesis Test for the Regression Parameters

This was performed at 5% significance to know from regression model coefficient and slope if there is a significant linear relationship between Latitude/Longitude (global) and X/Y(local planimetric) coordinates. If there is, then the slope will be significantly different from zero.

Testing hypothesis: $H_0 : \beta_1 = 0$ (Slope is equal to zero)

$H_1 : \beta_1 \neq 0$ (Slope is not equal to zero)

Significant level: $\alpha = 0.05$

Test statistics $T = \dfrac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)}$

Decision Rule: Reject H_0 if $|t| > t_{\alpha/2,\ n-2}$; from student t- distribution table, $t_{(0.025,\ 40)} = 2.021$

Conclusion: If the calculated $|t|$ is greater than $t_{\alpha/2,\ n-2}$, then the null hypothesis will be rejected.

The values in the table below is a clear evidence that the calculated **| t |** exceeds the appropriate critical values t (0.025, 13) = 2.021, which is a clear indication that the data provides convincing evidence that β_1 is different from zero. Conclusively, there is a non-zero (positive) association between our global and cartesian planimetric coordinate system, hence H0: $\beta_1 = 0$ is rejected.

Table 8. Calculated $|t|$ for non -zero slope

Okitipupa Region		*Akure South Region*	
Y- Local Northing coordinate against Global Latitude	X- Local Easting coordinate against Global Longitude	Y- Local Northing coordinate against Global Latitude	X- Local Easting coordinate against Global Longitude
845.526	4174.329	361.798	555.578

The Standard Error

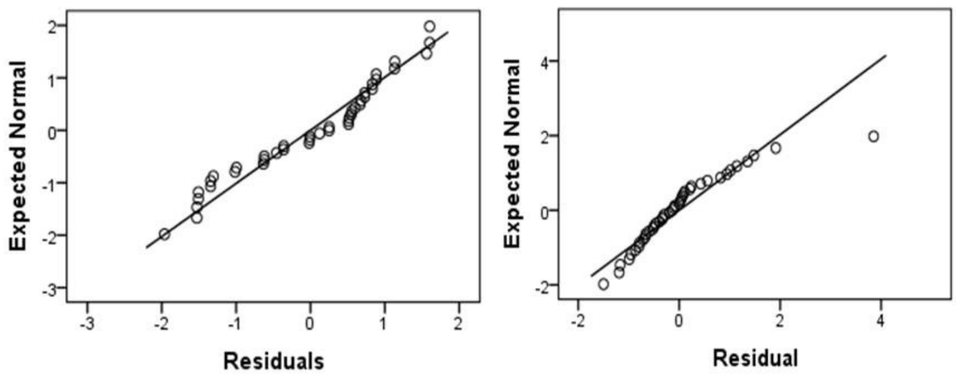
This is a measure of how much the value of the test statistics varies from the data used. It is the standard deviation of the sample distribution for the statistics. It measures the extent of precision the models estimates the values of X and Y (local planimetric) coordinates for both region as shown in table 9 below.

Table 9. Standard error for the slope estimate $SE(\hat{\beta}_1)$

Okitipupa region		*Akure South Region*	
Y- Local Northing coordinate against Global Latitude	X- Local Easting coordinate against Global Longitude	Y- Local Northing coordinate against Global Latitude	X- Local Easting coordinate against Global Longitude
4.448	2.01	53.244	43.949

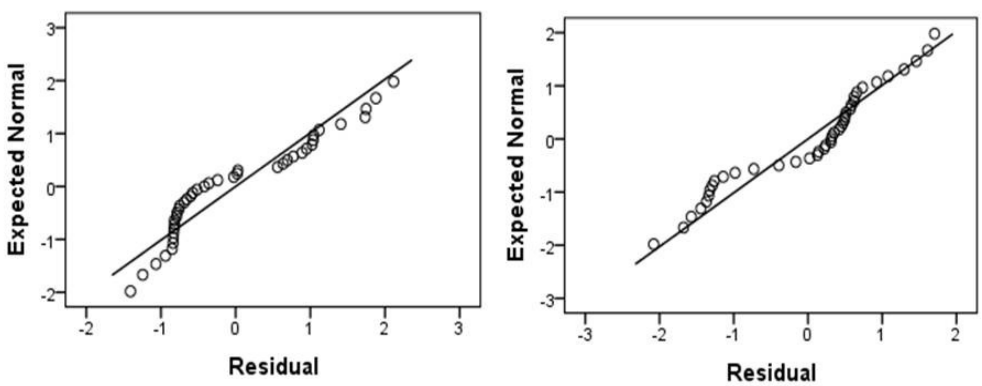
Test of Normality for Residuals

Another way of checking the adequacy of the developed model is to know if the observation is normally distributed by examining the residuals. The residuals provide information about the adequacy of the fitted models, as they describe the error in the fit of the model to the ith observation y_i by checking the assumption of normality and constant variance. In developing a regression model, the residuals ought to be normally distributed, and this normality was tested through graphical approach as shown below;



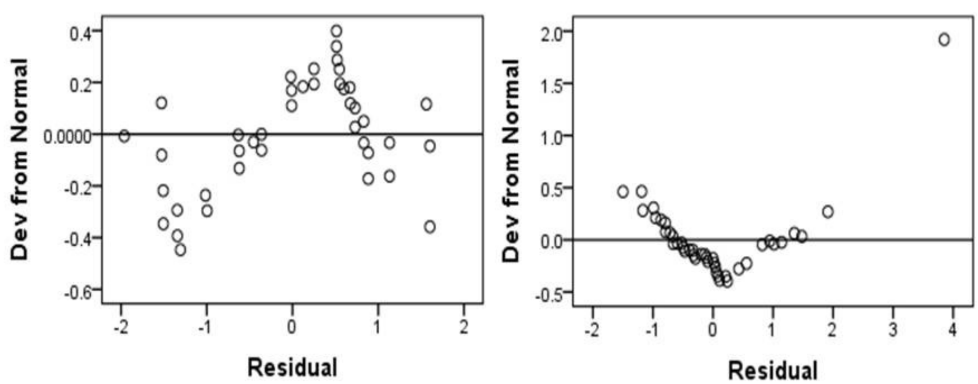
Latitude Longitude

Figure 6. Expected normal Residual Plot for portion of Okitipupa Region



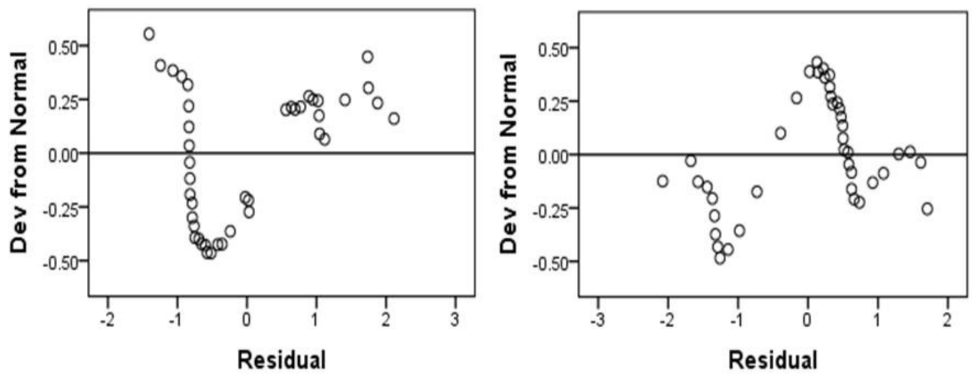
Latitude Longitude

Figure 7. Expected Normal Residual Plot for portion of Akure Region.



Latitude Longitude

Figure 8. Residual Plot Deviation from Normal for portion of Okitipupa Region



Latitude Longitude

Figure 9. Residual Plot Deviation from Normal for portion of Akure Region

Comparing the expected normal residuals plot in figure 6 and 7, it can be seen from the illustrations that more of the observation residuals from Okitipupa region are closer to the line compared to Akure region. This shows that the expected normal distribution of the residuals is better in Okitipupa region. While figure 8 and 9 shows the extent of deviation of the residual observation from normal. The residuals are closer to the zero line in Okitipupa region compared to Akure South region, indicating that residual deviation from normal is higher in Akure south region. Therefore residual distribution from both region shows that Okitipupa region are better normally distributed, meaning its model will predict a better accuracy.

Testing Models of both Regions for Prediction

Ten points where chosen from within both regions to test the models level of prediction. Table 8 and 9 show the results of predicted values and their corresponding residuals from the developed models.

Table 10. Okitipupa tested points and their predicted result

Point ID	*Existing Coordinate*		*Testing Ponits (Akure)*		*Predicted Coordinate*		*Residual*	
	E	*$LONG_{WGS84}$*	*LAT_{WGS84}*	*LAT_{WGS84}*	*E*	*N*	*ΔE*	*ΔN*
P1	697924.310	718367.622	4.7890273801	6.4969376578	697920.128	718367.246	4.1815	0.376
P2	697894.143	718138.488	4.7887473505	6.4948668674	697889.134	718138.921	5.0088	-0.433
P3	697165.629	719502.948	4.7822050740	6.5072271404	697165.022	719501.764	0.6070	1.184
P4	696982.910	719485.315	4.7805526942	6.5070735213	696982.133	719484.826	0.7767	0.489
P5	696721.678	719526.060	4.7781923990	6.5074502454	696720.891	719526.364	0.7869	-0.304
P6	696733.837	719449.288	4.7782998879	6.5067557130	696732.788	719449.785	1.0488	-0.497
P7	696824.047	719324.438	4.7791114215	6.5056239765	696822.610	719325.000	1.4366	-0.562
P8	696719.606	719618.255	4.7781766011	6.5082839061	696719.143	719618.283	0.4634	-0.028
P9	696869.620	719334.678	4.7795237405	6.5057151187	696868.247	719335.049	1.3734	-0.371
P10	697073.133	719360.392	4.7813643482	6.5059411319	697071.969	719359.969	1.1642	0.423

Table 11. Akure South region tested point and their predicted result

Point ID	*Existing Coordinate*		*Testing Ponits (Akure)*		*Predicted Coordinate*		*Residual*	
	E	*$LONG_{WGS84}$*	*LAT_{WGS84}*	*LAT_{WGS84}*	*E*	*N*	*ΔE*	*ΔN*
P1	740506.023	802852.212	5.1773968828	7.2591208369	740486.268	802909.045	19.755	-56.833
P2	741678.593	802448.611	5.1879936126	7.2554223903	741664.247	802498.679	14.346	-50.068
P3	746244.937	795524.773	5.2298218511	7.1915549894	746314.055	795412.193	-69.118	112.580
P4	743115.275	797219.929	5.2015700934	7.2070148106	743173.467	797127.557	-58.192	92.372
P5	739356.656	804138.569	5.1670482010	7.2707992231	739335.864	804204.834	20.792	-66.265
P6	742197.318	801575.783	5.1934514934	7.2464282139	742270.968	801500.719	-73.650	75.064
P7	742986.089	796802.147	5.2003825888	7.2032440329	743041.459	796709.165	-55.370	92.982
P8	738243.192	801852.171	5.1576704454	7.2490978702	738293.391	801796.934	-50.199	55.237
P9	738649.150	804073.386	5.1614411351	7.2691588307	738712.558	804022.823	-63.408	50.563
P10	738859.850	803112.096	5.1633069540	7.2604602773	738919.970	803057.664	-60.120	54.432

A good appraisal of the predicted values from the developed models of both regions show that regression model for the portion of Okitipupa region has a far better predictive accuracy compared to the portion of Akure region. This reveals significance of relative distance between random distributions of GPS points when developing a regression model for coordinate prediction. High deviation of predicted values in Akure region as shown in table 11 is ascribed to the wide range distances between the points used in developing the model as compared to Okitipupa region.

CONCLUSION

This paper presents simple method of coordinate transformation for a portion of Okitipupa and Akure regions in Ondo state, using developed regression models. Standard reference points where used in both regions in developing the models. The distance between the random distributions of GPS points in Okitipupa region was less than 350m apart while that of Akure was at least 850m apart. The derived model performance for Akure region was not satisfactory compared to Okitipupa region, revealing the extent to which distances between random distributions of points can be a factor when developing regression model for local coordinate prediction. It is therefore advisable to limit the distances between the scattered points to less than 350m when developing a regression model for local coordinate prediction, to minimize errors in predictive values. However the developed regression model for a portion of Okitipupa can be conveniently utilized by anybody within Longitudes 4^0 45’50”.64 and 4° 47’26”.86 East of Greenwich Meridian and latitudes 6^0 30’9”.00 and 6^0 30’40.66 North of the Equator for transforming global coordinate to local planimtric coordinate.

ACKNOWLEDGEMENT

The author is highly indebted to China University of Geosciences (Wuhan) and Department, Ministry of Lands, Housing and Environment, Akure, Ondo State in Nigeria.

REFERENCES

[1] Blueman, A. (2011). *Elementary Statistics Step by Step* (ISBN-10: 0077460391)

[2] Greaves, M. (2001). The Os's New Coordinate Transformation for Great Britain – GPS to Osgb36 National Grid Transformation. *Geomatics World*, November / December 2001, pp. 34-36.

[3] Nwilo, P.C, Fajemirokun F.A, Ezeigbo C.U, Oyewusi A. M. & Ayodele E.G. (2007). A Study of Some Transformation Procedures for Nigerian Geodetic Network.

[4] Gomaa, M. D. (2000). Optimum Geodetic Datum Transformation Techniques for GPS Surveying in Egypt.

[5] Gomaa, M. D. (2011). Simple Precise Coordinates Transformations for Geomatics Applications in Makkah Metropolitan Area, Saudi Arabia.

[6] Ziggah, Y. Y. (2013). *Determination of GPS Coordinate Transformation Parameters*: A Case Study of Ghana Geodetic Reference Network. Lambert Academic Publishing.

[7] Montgomery, D. C., Runger, G. C. & Hubele, N. F. (2013). *Engineering Statistics*. New York: Wiley. (ASIN: B004K3BN6I)

[8] Chih-Hung, W. (2007). Direct Transformation of Coordinates For GPS Positioning Using The Techniques of Genetic Programming And Symbolic Regression on Partitioned Data.

[9] Hossain, M. Z. (2011). The Use of Box-Cox Transformation Technique in Economic and Statistical Analyses. *Journal of Emerging Trends in Economics and Management Sciences, 2*(1), 32-39.

[10] http://En.Wikipedia.Org/Wiki/Okitipupa.

[11] http://En.Wikipedia.Org/Wiki/Akure_South.

[12] http://en.wikipedia.org/wiki/Linear_regression.

[13] http://en.wikipedia.org/wiki/Regression_analysis.