EMPIRICAL MODELS FOR FORECASTING GLOBAL SOLAR RADIATION ON HORIZONTAL SURFACE USING SUNSHINE HOURS AND TEMPERATURE DATA OVER IKEJA, LAGOS STATE, NIGERIA

[1]Ilori, B.Y., [2]Alabi N. O., [3]Ogun, C. A., [4]Awofodu J.O
[1,3 & 4]Department of Physical Science, Yaba College of Technology
[2]Department of Statistics, Yaba College of Technology
e-mail address: iloribabatunde7@gmail.com

Abstract

The measured data of global solar radiation on horizontal surfaces sunshine hours and temperature for Ikeja with latitude 6.39^0N, longitude 3.23^0E and altitude of 39.35 meters, during the period of (1990 – 2010) are analyzed. Generalized additive models (GAMs) were employed with scatterplot smoothers such as cubic smoothing splines and nearest neighbours to establish non-linear relationships that exist between solar radiation and the two random covariates. Generalized cross-validation re-sampling technique was employed in determining the effective degrees of freedom and the window size or span for smooth spline and local regression additive models respectively. Models accuracy was compared using the residual deviance for the GAMs. Composite GAM was selected through a nested analysis of deviance on four postulated models. **R** programming language packages such as " splines" , " gam" , " ggplot2" , " boot" and " lattice" amongst others were employed throughout the analysis. The selected model is given as $\overline{H}_{p(i)} = \overline{H}_{c(i)}\left[-1.38 + 0.28ks_i + 0.28\Delta T_i\right]$ Where, $\overline{H}_p$ is the predicted solar radiation, $\overline{H}_c$ is the calculated extraterrestrial solar radiation, **Ks** is the sunshine ratio and Δ**T** is the temperature difference (maximum minus minimum temperature). The developed model could be employed in predicting global solar radiation of a location that has the same geographical parameters as Ikeja, Lagos State, Nigeria.

Keywords: Solar Radiation, Generalized Additive Models, Generalized Cross-Validation, Smoothing Splines and Local Regression.

1.1 Introduction

Global solar radiation is of economic importance as renewable energy alternatives. More recently global solar radiation has been studied due to its importance in providing energy for earth climatic system. The successful design and effective utilization of solar energy system and devices for application in various facets of human needs such as power and water supply for industrial, agricultural, domestic uses and photo voltaic cell largely depends on the availability of information on solar radiation characteristics of the location which the system and device are to situated (Falayi and Rabiu, 2005). The best solar radiation information can be obtained from experimental measurements of the direct and diffuse component of solar insulation at the particular location (Ohunakin et al, 2013). Although solar radiation data are available in most meteorological stations, many stations in developing countries (including Nigeria) suffer from a shortage concern of these data. Thus alternative methods for estimating these data are required (Al-Salihi et al., 2010; Okundamiya

and Nzeoko, 2010). The alternative approach is to correlate the global solar radiation with the meteorological parameters at the place where the data is collected. The resultant correlation may then be used location of similar meteorological and geographical characteristics at which solar data are not available (El-Sebaii and Trabea, 2005). Several empirical models have been developed to calculate the solar radiation using various parameters such models include that of Augustina and Nnabuchi (2009), Mfon et al., (2013). Isikwe et al., (2012), Mfon et al., (2014), Ekwe et al., (2014), Ododo (1994), Okonkwo and Nwokoye (2014), Falaye et al., (2008) and Ohunakin et al., (2012).
This present work intends to develop a model for forecasting the global solar radiation on horizontal surfaces in Ikeja, Lagos State, particularly in areas with the same meteorological parameters as Ikeja.

1.2 Methodology

Solar regression model provides an alternative to actual measured weather station data by building and correlating the empirical relationship between solar radiation and the existing meteorological parameters. One significant advantage of this approach is that some of the meteorological parameters for example ambient temperature and sunshine hours can easily be measured in most places (Ayodele et al., 2016; Agbo, 2006). In this study, the solar radiation data which contains ambient temperature and sunshine hours were obtained from the Nigerian Meteorological Agency (NIMET). The data obtained cover a period of twenty-one years for Ikeja with latitude 6.39^{0}N, longitude 3.23^{0}E and an altitude of 39.35 meters, situated in the coastal region of Nigeria as shown on the map in **Figure** 1.

Figure 1.Map of Lagos state, **Source**: Google Maps

Also the sunshine hour based regression model was amalgamated to produce a single model containing both change in temperature and sunshine ratio. The aim of this study is to develop empirical formulae based on sunshine hour ratio and change in temperature for predicting global solar radiation on the horizontal surface for location having the same latitude and topography with Ikeja. Though research works have been conducted at different instances and locations but with the use of software such as excel. This study involves the use if more sophisticated packages found on **R** Studio programming language, which gives wider statistical regression techniques options. So far the relationship between sunshine hours and temperature have more significance than other parameters hence the need to produce a model comprising sunshine hour and temperature which can easily be obtained for easy prediction of global solar radiation.

The global solar radiation data supplied by NIMET in Gum-Bellani distillate were converted to useful form $MJm^{-2}day^{-1}$ using conversion factor of 1.216 proposed by Ododo (1994). The most commonly used model which relates the global solar radiation to sunshine duration was first developed by Angstrom- Prescott type regression equation Angstrom, (1924) and modified by Prescott, (1924) and (Falayi et al., 2008). Solar radiation clearness index is computed from equation 1.

$$\frac{\overline{H}_m}{\overline{H}_c} = a + b(\frac{S_m}{S_c}) \tag{1}$$

where H_m is the monthly average daily global solar radiation on horizontal surface ($MJm^{-2}day^{-1}$), H_c is the calculated monthly average daily extra terrestrial radiation on a horizontal surface ($MJm^{-2}day^{-1}$). S_m is the monthly average daily number of hours of bright sunshine, S_c is the monthly average daily maximum number of hours of possible sunshine, where a and b are regression constant to be developed. S_c was computed for each month using equation 2.

$$S_c = \frac{2}{15}\omega_s \tag{2}$$

where, ω_s is the sunset hour angle given as

$$\omega_s = \cos^{-1}\left[-\tan\left(\frac{\pi\phi}{180}\right)\tan\delta\right] \tag{3}$$

The forecasting ability of this model is based on the pre-knowledge of extraterrestrial radiation (H_c), sunshine ratio (S_m/S_c), temperature. Extraterrestrial radiation in ($MJM^{-2}day^{-1}$) can be calculated for each day of the year at different latitude from equation 4.

$$\overline{H}_c = 37.59 * E_0 * \left[\cos\phi * \cos\delta * \sin\omega_s + \left(\frac{\pi\omega_s}{180}\right)\sin\phi * \sin\delta\right] \tag{4}$$

where E_0 is the eccentricity correction factor, 37.59 $MJM^{-2}day^{-1}$is a constant, which comprise the solar constant and other factors. While Φ is the latitude of the location under consideration.

$$E_0 = 1 + 0.033\cos\left(\frac{2\pi d_n}{365}\right) \tag{5}$$

where δ is the declination angle given in radian as

$$\delta = \frac{23.45\pi}{180}\sin\left[2\pi(\frac{d_n + 284}{365})\right] \tag{6}$$

1.3 Models setup, methods of estimation of parameters and results

The empirical analysis of the data begins with the analysis of the pairwise correlations between solar radiation, sunshine hours and changes in temperature. Scatterplots (**Figure** 2) show the association between solar radiation and its predictors at Ikeja during the period under review. It is not surprising that the nature of the interaction differs for each predictor, which mostly indicates the existence of nonlinear relationships with the solar radiation. However, the main objective of this current work is to establish an efficient model that can provide the best estimates for the solar radiation data at Ikeja from 1990 to 2010. Therefore, emphasis is on accuracy of the selected models through analysis of deviance.

Table 1: Correlation Analysis amongst the variables

	kt	**Ks**	**Te**
kt	1		
ks	0.51* t=11.373	1	
te	0.44* t=9.381	0.53* t=11.927	1

* Statistically significant at 1 per cent level. Personal Computations using **R** Studio

An initial test of significance conducted on the correlation coefficients between the variables indicated highly significant association signifying these predictors are covariates (**Table** 1).

<u>Scatterplots of likely association between Solar Radiation and its predictors at Ikeja between 1990 and 2010</u>

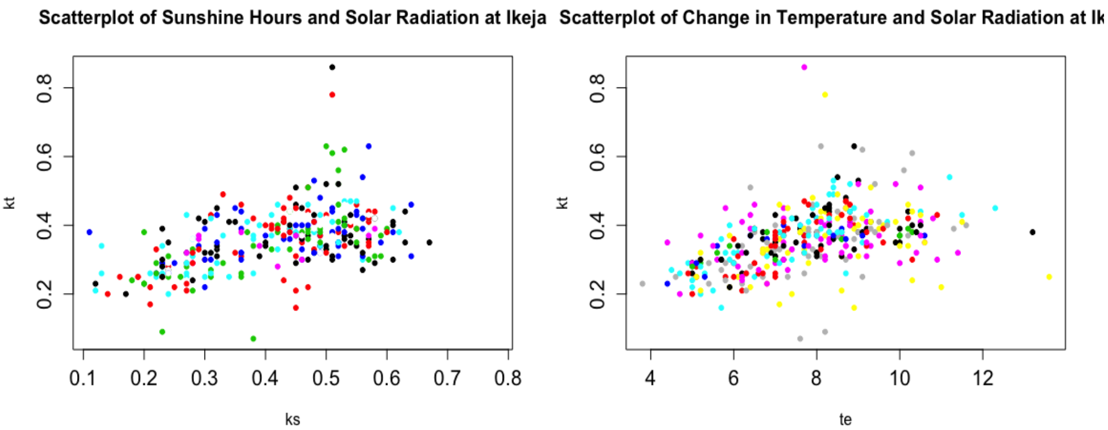

Figure 2: Scatter plots on solar radiation and covariates. These plots exhibit patterns of non-linearity. **Source**: Personal computation using **R** Studio ggplot2 package.

<u>1.4 Generalized Additive Model (GAM) on solar radiation using smoothing splines and local regression</u>

A powerful flexible regression-modeling tool was employed in generalized additive model (GAM) was employed. A GAM such as, specified in equation 7 will form the basis of all the models postulated in this paper. Although each unspecified univariate smooth function f_j in equation 7 is estimated using very fast non-parametric local scoring algorithm techniques such backfitting (Friedman and Stuetzle, 1981 & 1982) and memory-based procedure. The response and its predictors were transfomed into their natural log equivalents to rescale all the models parameters. The fit of a flexible regression analysis method such as GAM is necessary to ensure interpretability lost with the use of other non-linear regression statistical learning tools such as bagging, random forests, boosting and polynomials of higher degrees. GAM which was proposed by Hastie and Tibshirani (1986) is an extension of the generalized linear model[1] which replaces each linear predictor in a linear model with a smooth non-linear function $f_j(X_{ij})$. GAMs allow for implicit non-linear relationships between the response and the covariates without suffering from " curse of dimensionality" (Cho et al., 2013).
GAM model is postulated on solar radiation and the two covariates, sunshine hours and change in temperature as follows:

$$E(kt_i|ks, \Delta T) = f_0 + f_1(ks)_i + f_2(\Delta T)_i \qquad 7$$

The above equation is a specification on individual f_j which are standardized so that **E** [$f_j(.)$]=0, for each predictor or covariate[2] and then add together all of their contributions to the response variable. Non-linear f_j was fitted into each predictor allowing non-linear relationships missed by a multiple linear model.
Smoothing spline method was introduced, which results are summarized in **Table** 2 and **Figure** 3. This method involves fitting an additive model by using additive regression backfitting algorithm with weights (ARBAW) in which each f_1 and f_2 are smoothing splines[3] with 5.47 and 4.36 effective degrees of freedom df π respectively. Each function f_j contains smoother or smoothing parameter π determined by using re-sampling method known as generalized cross-validation (GCV) which is based on an orthogonal rotation of the matrix of residuals such that the diagonal elements of the hat or influence matrix[4] are as even as possible. Equation 8 is a function of residual sum of squares plus a smoothness penalty associated with each f_j .

$$E(kt_i|ks, \Delta T) = \sum_{i=1}^{n} (Y_i - g(X_i))^2 + \pi \int g^{||}(z)^2 dz \qquad 8$$

[1] *The generalized linear model is a class of log likelihood-based regression models, which assume some parametric form for the predictors.*
[2] *The result of the correlation analysis shows that these random predictors are covariates in the solar radiation model.*
[3] *Smoothing splines result from minimizing a residual sum of squares criterion subject to a smoothness penalty. They are natural cubic splines with knots at every unique observation of X_i. The details are not covered in this research work. Additive regression backfitting algorithm with weights (ARBAW) fits a multiple predictors model by repeatedly updating the fit for each predictor in turn holding the others fixed. This approach also called local scoring algorithm uses the scatterplot smoothers to generalize the Fisher scoring procedure for calculating MLE (Hastie and Tibshirani, 1986).*
[4] *A hat (influence) matrix provides the estimates of vector of **E**(Y) when post-multiplied by the data vector y.*

where Y is the vector of observations on solar radiation, X is the vector on predictor variables, sunshine hours and change in temperature and π is the tuning or smoothing parameter which determines the effectiveness of the penalty term in equation 8. The integral part measures the overall change in the slope $g^I(z)$ over its entire range. An attempt was made to find the smoothing spline, g that minimizes the function in equation 8. As the smoothing parameter π approaches ∞, the EDF decreases[5] from n to 2, where n is the sample size (here, n=252). ARBAW allows for updating of a function using the partial residual. **R** programming language packages were used such as “splines” to compute the smoothing parameter, the effective degrees of freedom and diagnostic checks for the GAM. The model is estimated as

$$E(kt_i|ks, \Delta T) = -1.3629 + 0.2798(ks\ )_i + 0.2750(\Delta T)_i \qquad 9$$

The estimated average effects of changes in sunshine hours and change in temperature on solar radiation indicate rather positive similarities. This model shows intuitively that increase in both covariates result in increase in solar radiation over Ikeja.

Table 2: Summary of analysis of GAMs using the entire 252 observations data set

Model Type	# Effects	#Significant effects	Null Deviance	Residual Deviance	AIC	Null – Residual Deviance
Smoothing Splines	4	3	21.49	14.66	22.01	6.83
Local Regression	4	2	21.49	14.50	33.16	7.19
Composite GAM 1	4	2	21.49	14.57	30.08	6.92
Composite GAM 2	4	2	21.49	14.41	25.38	7.08

Source: Personal computation using **R** Studio “ gam” package.

<u>Smoothing Splines plots showing relationships between solar radiation and its predictors</u>

5 *If EDF reduce to 2, then all observations are utilized and the resulting curve tallies with the least squares regression line. Of course this will result in a model with low residual variance but high bias. Hence the EDF are determined in some ways like the cross validation method of re-sampling.*

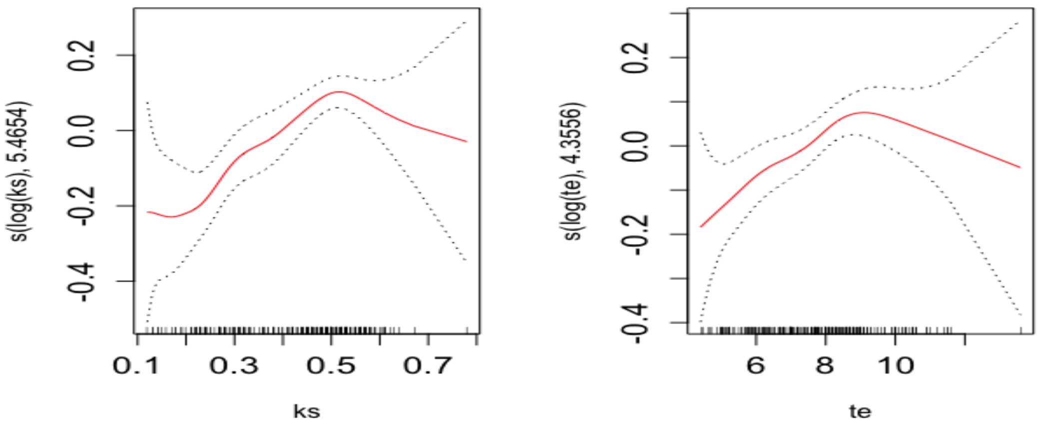

Figure 3: Plots of the relationships between each predictor and the solar radiation in the fitted model (equation 8). Each plot displays the fitted function and pointwise standard errors[6]. The two functions are smoothing splines with 5.47 and 4.36 EDF. **Source**: Personal computation using **R** Studio " splines" package.

The left panel in **Figure** 3 indicates that holding temperature fixed, solar radiation tends to rise with sunshine hours. Similarly, the right panel indicates that holding sunshine fixed, solar radiation tends to rise with changes in temperature. Furthermore, solar radiation over Ikeja was highest for intermediate values of sunshine and temperature and lowest for high and low sunshine hours and temperature values. The null model which estimates one parameter for the data set has a deviance of 21.49. The goodness-of-fit was tested by comparing the conducting the deviance test using the model deviance and residual degree of freedom. The p-value=1.00 was calculated which indicates no evidence of lack fit for the smooth spline model of equation 8. Therefore, the smoothing splines model in equation 3 is a " good fit" for the data. Since p = 2 (which is strictly less than the theoretical requirement that the number of predictors should be less or equal to 4), another model was postulated in equation 7 referred to as the local regression. Its methodology involves calculating the coefficients in equation 7 at a target point using only the nearby observations. For each function of the predictors f_j in equation 7, a span h = k/n was calculated of training observations closest to $\mathbf{x}_o$, k is the number of observations in the neighbourhood. Weight $W_{io} = W(x_i,x_o)$ was assigned to each point in the neighbourhood. A weighted least squares regression of Y_i on X_i using the W_{io} was fitted by minimizing

$$\sum_{i=1}^{n} W_{io}(Y_i - \beta_o - \beta_1 X_i)^2 \qquad 10$$

A linear form was assumed for the functions and calculated the span h using the efficient cross-validation approach. Cross-validation selected span of ~0.22 and ~0.23, for the functions sunshine hours and change in temperature respectively. Therefore, k the number of nearest observations is 55, 57 for f_1 and f_2 respectively. The lower the value of h, the " wigglier" the curve. Also, smaller values of h imply

[6] *The dotted curve is the estimated 95 per cent confidence interval for the smoothing splines.*

that the curve is more flexible and “ local” (**Figure** 4). The local regression fit on solar radiation results in equation 5 below.

$$E(kt_i|ks,\Delta T) = -1.3574 + 0.2844(ks)_i + 0.2743(\Delta T)_i \qquad 11$$

Local Regression plots showing relationships between solar radiation and its predictors

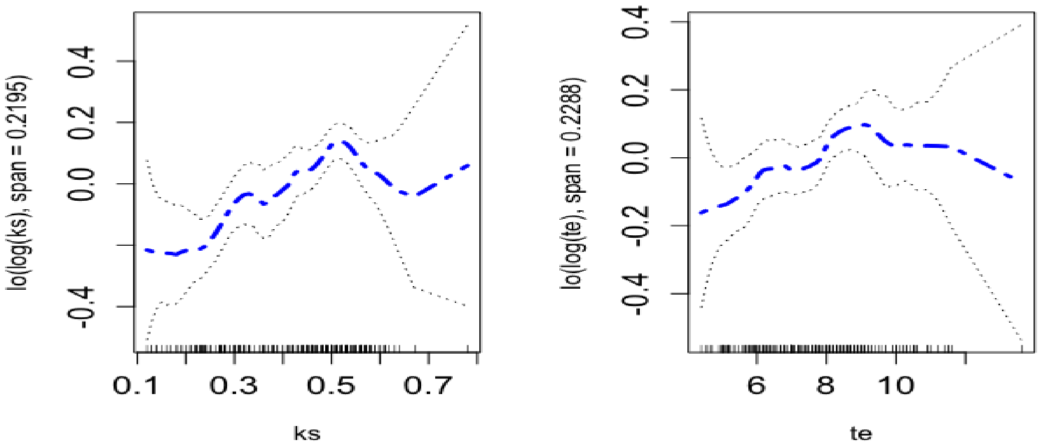

Figure 4: Plots of the relationships between each predictor and the solar radiation in the fitted model (equation 11). Each plot displays the fitted function and two times standard errors. The functions are local regression with h~0.22 and h~0.23 respectively. **Source**: Personal computation using **R** Studio “ gam” package.

The analysis of the plots in **Figure** 4 is obviously dissimilar to the smoothing splines in **Figure** 3. The curves on local regression depicted in **Figure** 4 are wigglier due largely to k relative to the sample size n. The left panel showing the plot on the additive model of the sunshine hours function indicate that solar radiation was lowest for high and low values of sunshine hours but rose sharply for values of sunshine hours higher than 0.7. The parametric effects of the predictors of solar radiation are all statistically significant at all levels (**Table** 2). Similar to the decision reached on the goodness of fit on equation 9, the local regression model in equation 11 is also a “ good fit” for the data on solar radiation over Ikeja.

Composite generalized additive models that comprise a blend of smoothing spline and local regression are now introduced. Firstly, a GAM (equation 12) was estimated whereby the function on sunshine hours was a smoothing spline with EDF=5.47 and the function on change in temperature was a local regression with span=0.23. This model is referred to as Composite GAM 1.

$$E(Kt_i|ks,\Delta T) = -1.3385 + 0.2874(ks)_i + 0.2664(\Delta T)_i \qquad 12$$

$$E(kt_i|ks,\Delta T) = -1.3829 + 0.2765(ks)_i + 0.2833(\Delta T)_i \qquad 13$$

Then, the second model (equation 13) was estimated with local regression function on sunshine hours with span=0.22 and smoothing spline function with EDF=4.36 on change in temperature referred to as Composite GAM 2. The estimated coefficients do not vary widely with the estimates of the models on smoothing splines and local regression in equations 9 and 11 respectively. Table 2 shows the summary of these models.

Composite GAM 1 plots showing relationships between solar radiation and its predictors

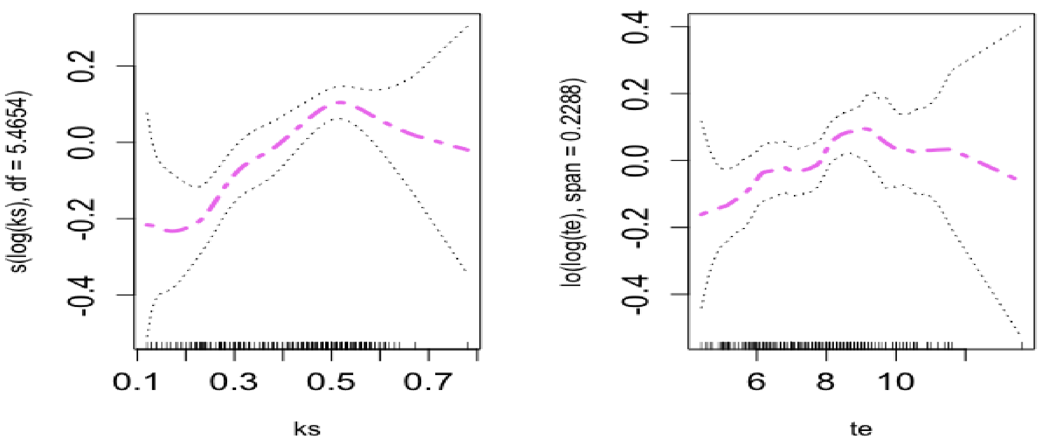

Figure 5: Composite GAM 1 with smoothing spline function on sunshine hours and local regression function on change in temperature. **Source**: Personal computation using **R** Studio “ gam” package.

Composite GAM 2 plots showing relationships between solar radiation and its predictors

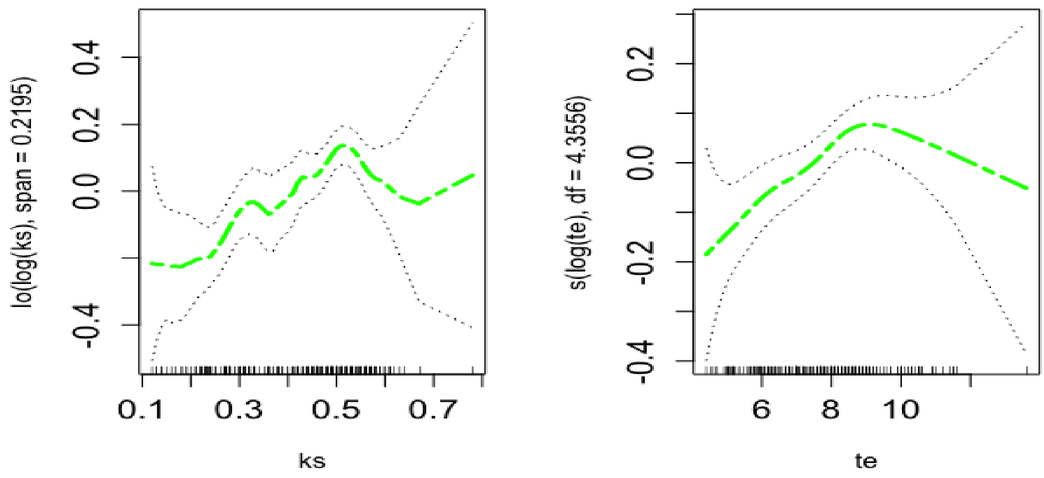

Figure 6: Composite GAM 2 with local regression function on sunshine hours and smoothing spline function on change in temperature. **Source**: Personal computation using **R** Studio “ gam” package.

Notice the similarity in the flexibilities exhibited by the curves (**Figure** 5 and **Figure** 6) and the earlier two models. Tests of goodness-of-fit reveal that the two composite GAMs are also adequate for the solar radiation data (**Table** 2).

1.5 Analysis of deviance on postulated models

These four models were compared using a nested analysis of deviance (**Table** 3). The Composite GAM 2 model clearly achieves lower residual deviance compared to all the other models. The probability of achieving a lower deviance of 0.16 for one degree of freedom in a nested analysis is about 0.28, which is statistically insignificant, at all levels. Regardless, we decide Composite GAM 2 model in equation 7 is sufficient for the data on Ikeja solar radiation because of its slightly lower residual deviance. Also, the predicted values of solar radiation were compared over Ikeja from 1980 through 1989 using the four postulated models (**Figure** 7).

Table 3: Analysis of deviance on Smoothing Spline, Local regression models, Composite GAM 1 and Composite GAM 2

Model	Residual df	Residual Deviance	Deviance	Pr(>Chi-square)
Smooth Spline	241.18			
Local Regression	232.51	8.67	0.36	0.73
Composite GAM 1	236.38	-3.87	-0.27	0.34
Composite GAM 2	237.31	-0.92	0.16	0.28

Source: Personal computation using R Studio “ gam” package.

<u>Forecasting Solar Radiation between 1980 and 1989 using the postulated models</u>

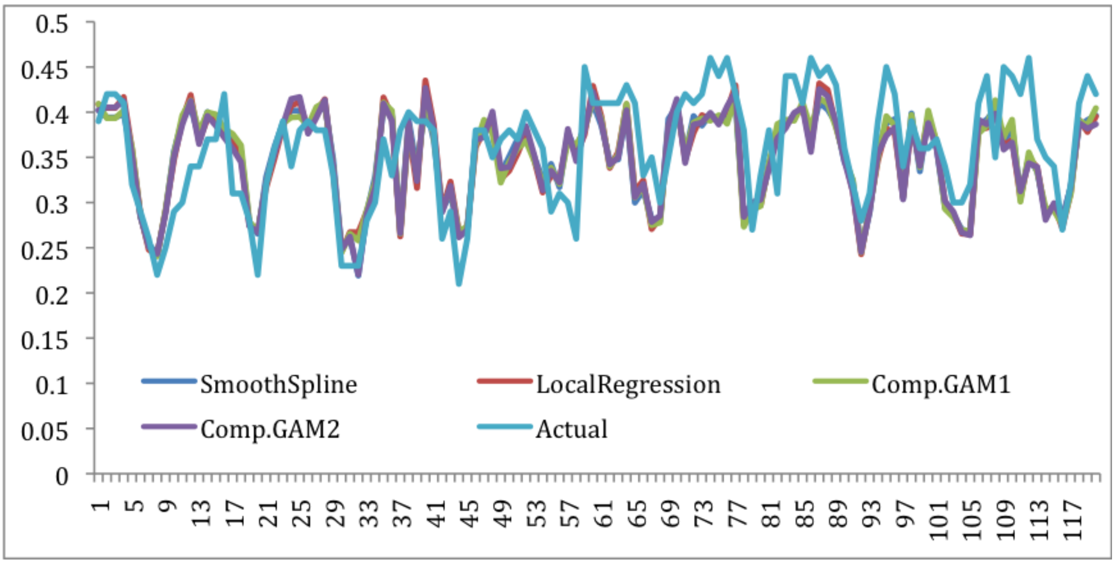

Figure 7: Plots of predicted values of the solar radiation over Ikeja from 1980 to 1989 using four postulated models. It is obvious from the plots that differences in the models are small throughout the test data set. **Source**: Personal computation using **R** Studio lattice package.

1.6 Conclusion

The effects of the two predictors of solar radiation such as sunshine hours and change in temperature over Ikeja between 1990 through 2010 were studied. Four powerful generalized additive models were estimated for prediction and interpretation purposes. The four models revealed positive relationship between solar radiation and the two predictors. The predicted values generated using these models indicate an insignificant difference between the models. However, due to slightly lower residual deviance and probability of achieving this deviance, the Composite GAM 2 model was decided to be sufficient for the solar radiation data at Ikeja.

Reference

1. Agbo,S.N and Oparaku, O.U (2006), " Positive and future prospect of solar water heating in Nigeria" , The pecific J. Science and technology,7(2):191-195
2. Al-Selihi, Kadun, M.M. & Mohammed, A.J. (2010). " Estimation of global solar radiation on horizontal surface using meteorological measurement for different cities in Iraq" , Asian Journal of Scientific Research, 3(4):240-248.
3. Augustine & Nnabuchi (2003). " Empirical model for the correlation of global solar radiation with meteorological data for Enugu" , 10(1).
4. Augustine, C. and Nnabuchi, M.N. (2009). " Correlation between Sunshine hours and global solar radiation in Warri, Nigeria" , Researchgate.net/publication/288698034, volume 10, number 2.
5. Ayodele, R., Ayodeji, S.O. & Chukwuka, G.M. (2006). " On the global solar radiation prediction methods" , Journal of renewable and sustainable energy.
6. Ekwe M.C., Joshua J.K. & Igwe J.E. (2014). " Estimation of Daily Global Solar Radiation at Owerri, Imo State, Nigeria from hours of sunshine,

minimum temperature and relative humidity", International Journal of Applied Research and Studies.

7. El-Sebaii, A.A, & Trabea, A.A. (2005). " Estimation of global soalr radiation on horizontal surface over Egypt", Egypt Journal of Solids, 28(1).
8. Falayi, E.O. & Rabiu, A.B. (2005). " Modelling Global Solar radiation using sunshine duration data", Nigeria Journal of Physics, 17:181-186.
9. Friedman, J.H. & Stuetzle, W. (1981). " Projection Pursuit Regression", Journal of American Statistical Association, 76:817-823.
10. Friedman, J.H. & Stuetzle, W. (1982). " Smoothing of Scatterplots", Technical Report, Orion 3, Department of Statistics, Stanford University.
11. Haeran, C., Yannig, G., Xavier, B. & Qiwei, Y. (2013). " Modeling and Forecasting Daily Electricity Load Curves: A Hybrid Approach", Journal of the American Statistical Association, 108:501,7-21.
12. Hastie, T. and Tibshirani, R. (1986). " Generalized additive models (with discussion)". Statistical Science 1:297– 318.
13. Hastie, T. J. and Tibshirani, R. J. (1990). Generalized Additive Models, Chapman & Hall/CRC.
14. Isikwe, B.C., Amah, A.N. & Agada, P.O. (2012). " Empirical model for the estimation of global solar radiation in Makurdi, Nigeria", Global Journal of Science Frontier Research Physics and Space Science, 12(1).
15. Mfon, D.U., Sunday, O.U., & Ye-Obong, N.U. (2013). " Estimation of of global solar radiation on horizontal surface from sunshine hours and other meteorological paprameters for Calabar, Nigeria", Journal of Asian Scientific Research, 3(11):1083-1089.
16. Mfon D.U., Sunday, O.U. & Ye-Obong, N.U. (2014). " Estimation of global solar radiation on horizontal surface from sunshine hours over PortHarcourt, Nigeria", Journal of Electrical and Electronics Engineering Research, 6(1),pp. 1-5.
17. Muyiwa, S. & Adaramola (2012), " Estimating global solar radiation using common meteorological data in Akure", Journal of Renewable Energy.
18. Ododo (1994). " New model for prediction of solar radiation in Nigeria", paper presented at 2nd OAU/STRC conference on new, Renewable and solar energies, at Bamako Mali, May16-2000.
19. Ohunakin, O.S, Adaramola, M.S., Oyewola, O.M. & Fagbenle, R.O (2012). " Correlation for estimating solar radiation using sunshine hours and temperature measurement in Oshogbo, Osun State, Nigeria".
20. Okonkwo, G.N. & Nwokoye, A.O.C (2014). " Estimating Global Solar Radiation from temperature data in Minna location", European Scientific Journal, vol(10).
21. Okundaniya, M.S. & Nzeoko, A.N. (2010). " Empirical model for estimating global solar radiation on horizontal surfaces for selected cities in six geopolitical zones in Nigeria".