Mathematics and Statistics 8(4): 472-479, 2020
DOI: 10.13189/ms.2020.080415 http://www.hrpub.org

Penalized Maximum Likelihood Estimation of Semiparametric Generalized Linear Models with Application to Climate Temperature Data

Azumah Karim[1,*], Ananda Omutokoh Kube[2], Bashiru Imoro Ibn Saeed[3]

[1]Pan African University Institute for Basic Sciences, Technology and Innovation, Kenya
[2]Department of Mathematics,Kenyatta University, Kenya
[3]Department of Mathematics and Statistics,Kumasi Technical University, Ghana

Received April 10, 2020; Revised June 4, 2020; Accepted June 16, 2020

Abstract Global temperature change is an important indicator of climate change. Climate time series data are characterized by trend, seasonal/cyclical as well as irregular components. Adequately modeling these components cannot be overemphasized. In this paper, we have proposed an approach of modeling temperature data using semiparametric additive generalized linear model. We have derived a penalized maximum likelihood estimation of the additive component of the semiparametric generalized linear models, that is, of regression coefficients and smooth functions. A statistical modeling with real time series data set was conducted on temperature data. The study has provided indications on the gain of using semiparametric modeling in situations where a signal component can be additively decomposed in to trend, cyclical and irregular components. Thus, we recommend semiparametric additive penalized models as an option to fit time series data sets in modelling the different component with different functions to adequately explain the relation inherent in data.

Keywords Cubic Spline, Fourier Function, Trend, Cyclical/Periodic, Maximum Likelihood, Semiparametric, Generalized Linear Models

1 Introduction

A systems or a signal is defined by a set of entities, named as components that are jointly connected. These connections are accountable for defining the various relationships and dependencies among all components. The relationships and the dependencies are often unknown. Therefore, the knowledge of the components and understanding their connections according to [1], is an important way to modeling the system in order to improve decision making, forecasting, controlling system among other things. Climate change effects on individual regions varies over time according to [2], and therefore, societies need conscious effort to mitigate or adapt to changes.

A key to effective managerial decision-making will be forecasts of climate change that are accurate and cost effective. Global temperature change is an important indicator of climate change, although by no means the only one, [3]. Also, global temperature is a popular metric for describing the state of global climate. However, its effects are felt locally, but the global distribution of climate response to many global climate changes is reasonably congruent in climate models, suggesting that the global metric measure is useful [4]. Climate change is far from intangible, it is presently defining the course of people's lives. Africa experienced extreme weather events and more irregularity in weather patterns, leading to serious outcomes for the people, who depend on land and some water bodies to survive, [5]

Most natural time series signals are composed of trend as well as seasonality (or periodicity) components which described their observed sequence, and extracting these components are a crucial subject or problem in many scientific arenas. Additive models are generally employed in offering identically flexible and real explanations of relationship in modeling, such as regression modeling. In the literature, several studies assumed a linear relationship in the modeling are too limiting. Yet, the regression models normally adopt independent errors. However, there are numerous applications in which the data are correlated such as time series data in natural sciences and as well as the environmental science settings. To overcome the drawback of a parametric estimation, we employ the semiparametric generalized linear modeling approach, where we

model the trend non-parametrically by cubic spline smoothing and the seasonal/cyclical by Fourier function. The study, proposed a one-dimensional curve setting to consider formulating a functional relationship of modeling the components using semiparametric generalized linear models framework. These components are to be estimated using different function to establish the functional relationships inherent in the time series signal. Generally, spline smoothing is one of the nonparametric methods regularly used in practice for function estimation where requirement of assumption about the shape of the unknown function is not needed. The trend component will be adequately estimated by using the cubic spline. The Fourier function comes in handy and well suited to model signals that exhibits some seasonal/cyclical components. In this regard, the knowledge of climatic variables is indispensable. The time series analysis of climatic variables is a way to learn evolution of climate change and its effects on the socio-economic development of the society.

2 Materials and Methods

Researchers such as [6, 7], originated the estimation of the trend and seasonality components from a one dimensional time series signal in Economics. This was followed by further prescribed advances in statistics, see [8, 9, 10, 11, 12]. The structural time-series model was formilized and described as the classical way of decomposition by [12, 13], According to [14], and [15] owing to the increasing global warming in the world, analyzing greenhouse gas emissions is an essential concern. To estimate future emissions, the following time series analysis models: moving average method, exponential smoothing method, and exponential smoothing with trend method were used to estimate greenhouse gas emissions. For suitability of their models, they performed a statistical analysis on their results based on mean error, mean absolute error and root mean square values to assess the performance of the formulated models. An indication of modeling climatic factors using time series analysis concepts, even though the complex evolution of greenhouse gas emissions such as periodic or otherwise oscillatory nature were not considered in this approach.

Another study by [16], considered a generalized structural time-series modeling framework to analyzed the monthly records of mean temperature, one of the best central environmental factors, using classical stochastic processes, using the $SARIMA(0,1,2)(0,1,1)_{12}$ model and obtain a medium-term (10 years) forecast of the mean temperature in Erbil, community in Iraq, they predicted that the average temperature in Erbil, Iraq, will be stable for the next 10 years. Temperature could be seen as a localized climatic factor, concerted effort are required locally or globally to minimized the impact of climate , hence the need to scale this study globally. Succinctly, global mean surface temperature is widely used in the climate literature as a measure of the impact of human activity on the climate system , as a consequent, [17], employed the generalized least squares estimator, which they indicated has the capabilities to moderates most of the inaccuracy associated with the use of a naive area weighted average.

A study on the detection and future projection of climate change in the city of Rio de Janeiro by, [18], posited that, the average change in annual maximum (minimum) air temperatures may range between $2\,^{\circ}\mathrm{C}$ and $5\,^{\circ}\mathrm{C}$ ($2\,^{\circ}\mathrm{C}$ and $4\,^{\circ}\mathrm{C}$) above the current weather values in the late 21st Century. They also indicated that, the warm (cold) days and nights are becoming more (less) frequent each year, and for the future climate (2100) it has been projected that about 40% to 70% of the days and 55% to 85% of the nights will be hot, leading to no longer cold days and nights.

Also, a study by [19], on the trends and periodicities in the annual and seasonal temperature time series at fifteen weather stations within Ontario Great Lakes Basins were analyzed, for the period 1941-2005, where the researchers employed Fourier series analysis, t-test, and Mann-Kendall test, the study reveals that the extreme minimum temperature is increasing annually and seasonally, with statistically significant at many stations. Several researchers such as[20, 21] used a Deterministic-Stochastic Combine approaches to modeled the surface temperature.

Given the Climate time series observations, $y = (y_1, \cdots, y_n)$ collected on an ordered increasing time $t = (t_1, \cdots, t_n)$, we formulate a functional relationship that describe the characteristics of the data-set. For a given sequence of the pairs observations $\{(y_1, t_1), \cdots, (y_n, t_n)\}_{i=1}^{n}$, and, $t = (t_1, \cdots, t_n)$ in an increasing ordered compact real interval $T = [t_1, t_m] \subset \mathbb{R}$ define a function $\{f(t) : t \in T \subset \mathbb{R}\}$ that describes the relation between t and $y = (y_1, \cdots, y_n)$. The modeling of the surface temperature T_t is given by equation

$$T_t = L_t + C_t + E_t \tag{1}$$

where L_t represents the trend component, C_t is the cyclical or periodic component, and E_t, the stochastic term.

2.1 Modeling the trend component using splines models

A spline model is a piece-wise well-defined function with the separable pieces joined together employing continuity and smoothness constraints. They are worthwhile in explaining the relationship between a response variable and one or more independent variables when the relationship involves a curve or flexible model. The segments of a spline function are usually low order polynomials of up to third degree and the polynomial segments connect at a set of finite points known as knots.

Definition 1. Let $a < t_1 < \cdots < t_k < b$ be fixed points called knots. Let $t_0 = a$ and $t_{k+1} = b$. Generally, splines functions are piecewise polynomials joined together smoothly at the knots. Formally, a *spline function* of order m, is a real-valued function on the closed interval $[a, b]$.

Let L_t in equation (1) be a natural cubic spline function. On each interval, $[t_i, t_{i+1}],\ i = 1, \cdots, m-1$, where, L_t is given by a different cubic polynomial, l_i,defined as

$$L_t = \sum_{t=1}^{m-1} 1_{[t_i, t_{i+1})}(t) l_i(t) \tag{2}$$

Such that,

(i) l is piecewise polynomial of order m on the $[t_i, t_{i+1})$, $i = 0, 1, \cdots, k$

(ii) l has $m-2$ continuous derivatives and the $(m-1)$st derivative is a step function with jumps at the knots. For orders represented by $m = 2r$,the function l is a *natural spline function* of order $2r$ if, in addition to (i) and (ii), it satisfies the natural boundary conditions

(iii) $l^{(j)}(a) = l^{(j)}(b) = 0$,$j = r, \cdots, 2r-1$.

From equation (2) since L_t is natural cubic spline, the following must hold,

$$l_i(t) = a_i(t-t_i)^3 + b_i(t-t_i)^2 + c_i(t-t_i) + d_i,\ 1 \le i \le m-1 \tag{3}$$

and

$$l_{i-1}(t_i) = l_i(t),\ l'_{i-1}(t) = l'_i(t_t),\ l''_{i-1}(t) = l''_i(t_t),\ i = 2, \cdots, m-1$$

Now we introduce the following notations

$$l = (l_1, \cdots, l_n)', \quad where\ \ l_i = l(t_i), i = 1, \cdots, m$$

$$\gamma = (\gamma_2, \cdots, \gamma_{n-1})', \quad where\ \ \gamma_i = l''(t_i), i = 1, \cdots, m$$

From the definition of natural cubic spline, we have, $\gamma_1 = \gamma_m = 0$.

We know l''_i to be a linear function of the form:

$$l''_i(t) = \frac{\gamma_i}{h_i}(t_{i+1} - t) + \frac{\gamma_{i+1}}{h_i}(t - t_i), \quad t \in [t_i, t_{i+1}] \tag{4}$$

where $h_i = t_{i+1} - t_i$, for $i = 1, \cdots, m-1$. Integrating,equation(4) twice,we obtain l_i as:

$$l_i(t) = \frac{\gamma_i}{6h_i}(t_{i+1} - t)^3 + \frac{\gamma_{i+1}}{6h_i}(t - t_i)^3 + C_1(t - t_i) + C_2(t_{i+i} - t),\ t \in [t_i, t_{i+1}] \tag{5}$$

We need the conditions $l_i(t) = z_i$ and $l_i(t_{i+1})$ to determine C_1 and C_2, giving rise to

$$l_i(t) = \frac{\gamma_i}{6h_i}(t_{i+1} - t)^3 + \frac{\gamma_{i+1}}{6h_i}(t - t_i)^3 + \left(\frac{z_{i+1}}{h_i} - \frac{\gamma_{i+1}h_i}{6}\right)(t - t_i) + \left(\frac{z_i}{h_i} - \frac{\gamma_i h_i}{6}\right)(t_{i+1} - t) \tag{6}$$

Now, using the condition $l'_{i-1}(t_i) = l'_i(t_i)$, we obtain

$$\frac{1}{6}h_{i-1}\gamma_{i-1} + \frac{1}{3}\left(h_i + \frac{1}{6}h_{i-1}\right)\gamma_i + h_i\gamma_{i+1} = \frac{1}{h_i}(z_{i+1} - z_i) - \frac{1}{h_{i-1}}(z_i - z_{i-1}), for\ \ i = 1, \cdots, m-1. \tag{7}$$

From equation (7), corresponds to the following system of equations

$$R\gamma = Qz \tag{8}$$

These matrices are simplified as $z = (z_1, \cdots, z_m) \in \mathbb{R}^m$,$\gamma = (\gamma_2, \cdots, \gamma_{m-1}) \in \mathbb{R}^{m-2}$

Where Q is a tridiagonal $(m-2) \times m$ matrix given by;

$$\begin{aligned} Q_{i,i} &= \frac{1}{h_i} \\ Q_{i,i+1} &= -(\frac{1}{h_i} + \frac{1}{h_{i+1}}), \\ Q_{i,i+2} &= \frac{1}{h_{i+1}} \end{aligned} \tag{9}$$

is

$$Q = \begin{bmatrix} h_1^{-1} & 0 & \cdots & 0 \\ -h_1^{-1} - h_2^{-1} & h_2^{-1} & \cdots & 0 \\ h_2^{-1} & -h_2^{-1} - h_3^{-1} & \cdots & 0 \\ 0 & h_3^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h_{m-1}^{-1} \end{bmatrix}_{m \times (m-2)} \tag{10}$$

for $i = 1, \cdots, m-2$, and R is the symmetric tridiagonal $(m-2) \times (m-2)$ matrix given by

$$R_{i,i-1} = R_{i,i+1} = \frac{h_i}{6}, i = 2, \cdots, m-3, \tag{11}$$

$$R_{i,i} = \frac{(h_{i+1} + h_i)}{3}, i = 1, \cdots, m-2 \tag{12}$$

as

$$R = \begin{bmatrix} \frac{1}{3}(h_1 + h_3) & \frac{1}{6}h_2 & \cdots & 0 \\ \frac{1}{6}h_2 & \frac{1}{3}(h_2 + h_3) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{3}(h_{m-2} + h_{m-1}) \end{bmatrix} \tag{13}$$

Here equation (8) is solved for γ , and $\gamma_2, \cdots, \gamma_{m-1}$ is then plugged into equation (7).

We can specify a natural cubic spline by giving its value and second derivative at each knot t_i.

Define

$$l = (l_1, \cdots, l_n)', \ \ where\ \ l_i = l(t_i)$$

$$\gamma = (\gamma_2, \cdots, \gamma_{n-1})', \quad where\ \ \gamma_i = l''(t_i)$$

Which defines the curve natural cubic spline, l fully.

Letting l be a cubic interpolation spline interpolating the data $\{(t_i; y_i)\}_{t=1\cdots,m}$, it consists of piecewise cubic polynomials l_i such that,

$$l_i(t_i) = y_i, \quad l_i(t_{i+1}) = y_{i+1},\ \ i = 1 \cdots, m-1$$

together with the conditions

$$l'_i(t_{i+1}) = l'_{i+1}(t_{i+1}), \quad , and\ \ l''_i(t_{i+1}) = l''_{i+1}(t_{i+1}),\ \ i = 1 \cdots, m-1$$

We are left with two degrees of freedom. The natural cubic spline is the cubic interpolation spline that furthermore enforces,

$$l'(t_i) = l''(t_m) = 0$$

In general, we want the resulting function to exhibit some degree of smoothness. The general approach to a formal generalization of this is to introduce the roughness measure,

$$P(l) = \int_{t_1}^{t_m} \{l''(t)\}^2 dt \tag{14}$$

which clearly measure the total curvature of the smoothing function. A fundamental results in spline theory is that the natural cubic spline (the cubic spline imposing, $l'(t_i) = l''(t_m) = 0$ ensures the smoothest fit by minimizing equation (14) among all C^2 interpolation functions

Theorem 2.1. *The vectors* $\boldsymbol{l}$ *and* γ *defines a natural cubic spline function,* l*, if and only if the equation*(8) *is satisfy.*

Theorem 2.2. *With* l *denoting the natural cubic interpolation spline, we have for any interpolation function* $f \in C^2$ *such that* $P(f) \geq P(l)$*, where* $P(l)$ *is defined in equation* (14)*, with equality if and only if* $f = l$*.*

Proof. Since $l \in C^2$ is a cubic spline, we have that,

$$l'''(t) = k_i, \forall t \in (t_i, t_{i+1}),\ \ for\ some\ k_i \in \mathbb{R}, i = 1, \cdots, m-1 \tag{15}$$

$P(f) = 0$ if and only if $f''(t) = 0$, for all t, that is, if and only if f is a first order polynomial. For all, $x, y \in \mathbb{R}$, $x^2 - y^2 = (x-y)^2 + 2(x-y)y$, and let $\tau = f - l$. This implies that,

$$\begin{aligned} P(f) - P(l) &= \int_{t_1}^{t_m} [f''(t)]^2 dt - \int_{t_1}^{t_m} [l''(t)]^2 dt \\ &= \int_{t_1}^{t_m} [f''(t) - l''(t)]^2 + \\ &\quad 2\int_{t_1}^{t_m} [f''(t) - l''(t)]\, l''(t) dt \\ &= \int_{t_1}^{t_m} [\tau''(t)]^2 dt + 2\int_{t_1}^{t_m} \tau''(t) l''(t) dt \\ &= P(\tau) + 2\int_{t_1}^{t_m} \tau''(t) l''(t) dt \end{aligned}$$

Now,

$$P(\tau) = 0, \quad \Leftrightarrow \tau'' = 0 \Leftrightarrow f'' = l''$$

so we have $P(\tau) \geq 0$, with equality if and only if $f'' = l''$. Considering second term, we see that,

$$\begin{aligned} &2\int_{t_1}^{t_m} \tau''(t) l''(t) dt \\ &= 2\left([\tau'(t) l''(t)]_{t_1}^{t_m} - 2\sum_{i=1}^{m-1} \int_{t_i}^{t_{i+1}} \tau'(t) l'''(t) \right) \\ &= 2\left[\tau'(t) l''(t)\right]_{t_1}^{t_m} - 2\sum_{i=1}^{m-1} k_i \left(\tau(t_{i+1}) - \tau(t_i)\right) \end{aligned}$$

using partial integration, the continuity of the derivatives up to and including second order, and (15). Recall the definition of τ,

$$\begin{aligned} \tau(t_{i+1}) - \tau(t_i) = f(t_{i+1}) - l(t_{i+1}) - f(t_i) + l(t_i) = \\ y_{i+1} - y_{i+1} - y_i + y_i = 0 \end{aligned}$$

making the second term zero. Remaining with,

$$2\left[\tau'(t) l''(t)\right]_{t_1}^{t_m} = 2\left(\tau'(t_m) l''(t_m) - \tau'(t_1) l''(t_1)\right)$$

We note that by letting,

$$l''(t_1) = l''(t_m) = 0 \tag{16}$$

$(or\ \ with\ \ l' = f', in\ \ which\ \ case\ \ l = f)$, we have,

$$2\int_{t_1}^{t_m} \tau''(t) l''(t) dt = 0$$

This immediately allows us to conclude,

$$P(f) - P(l) = P(\tau) > 0$$

with equality if and only if $l = f$. □

Using standard notation proposed by[24, 22], the roughness measure,

$$\int_{t_1}^{t_m} \{l''(t)\}^2 dt = l^T K l \tag{17}$$

given that $K = Q^T R^{-1} Q$, Q and R are defined as in equation (10)and equations (13) respectively. Thus we have proven the theorem under the boundary conditions (16), which was exactly the defining property of the natural cubic spline.

2.2 Modeling the Periodicity/Seasonality

The periodicity term, C_t, defined by equation (1) as

$$C_t(k) = \sum_j A_k \cos\left(2\pi f_k t + \varphi_k\right) \tag{18}$$

with A as the amplitude, φ the phase, f the frequency of the sinusoidal wave in cycle per unit time, and k as the order of the sinusoids. The right hand side of equation (18), is a nonlinear function of the phase variable, φ. Using the following trigonometric identity

$$\cos\left(A + B\right) = \cos\left(A\right)\cos\left(B\right) - \sin\left(A\right)\sin\left(B\right)$$

Then, equation (18), can written in a form by [23] as

$$\begin{aligned} C(k) &= \sum_j A_k \cos\left(2\pi f_k t + \varphi_k\right) \\ &= \sum_j \left(A_k \cos 2\pi f_k t \cos\varphi_k - A_k \sin 2\pi f_k t \sin\varphi_k\right) \\ &= \sum_j \left(M_k \cos 2\pi f_k t + N_k \sin 2\pi f_k t\right) \end{aligned} \tag{19}$$

where $M_k = A_k \cos\varphi_k$ and $N_k = -A_k \sin\varphi_k$. The values of M_k and N_k are obtained using the least squares estimate proposed by [23]

2.3 Method of estimation

Using the idea of semi-parametric modeling as espoused by several authors under various degree of generality, for example see[24, 25, 26, 27, 28, 29]. In this paper we consider the semi-parametric generalized linear models as

$$f(y_i,\theta_i,\phi)=exp\left(\frac{y_i\theta_i-b(\theta_i)}{\phi}+c(y_i,\phi)\right) \tag{20}$$

where

$$\theta_i=Z_i^T\beta+l(t_i), \text{for i}=1,\ldots,n \tag{21}$$

where Z_i and t_i are both possibly vector-valued, are two sets of functions for the i^{th} response and the $p-vector\beta$ and function l are to be estimated. We consider a one-dimensional t and the roughness of the curve $l(t_t)$ by its integrated squared second derivative. Gien that $\mu_i=E(Y_i;\theta_i,\phi)=b'(\theta_i)$ Following [24], the link function G is defined as

$$G(\mu_i)=Z_i^T\beta+l(t_i) \tag{22}$$

In the literature, several authors considered the used of penalized likelihood in generalized linear models, see, [25, 29, 24, 27, 26, 28, 30, 31, 32, 33] The penalized log-likelihood of the semi-parametric generalized linear models is defined as

$$\Pi=\ell\left(\theta,\phi\right)-\frac{1}{2}\lambda\int l''(t)^2dt \tag{23}$$

from equation (23), equations (17)

$$\ell\left(\theta,\phi\right)=\sum_{i=1}^{n}\left(\frac{Y_i\theta_i-b(\theta_i)}{\phi}+c(Y_i,\phi)\right)$$

$$\int_{t_1}^{t_m}\left\{l''(t)\right\}^2dt=l^TKl$$

and also from equation (23)

$$G(b'(\theta_i))=Z_i^T\beta+l(t_i) \tag{24}$$

to be maximized over l and β.Hence, our penalized log-likelihood of the semi-parametric generalized linear models is defined as

$$\Pi=\ell\left(\theta,\phi\right)-\frac{1}{2}\lambda l^TKl$$

Let $\hat{\beta}$ and $\hat{l}$ be estimates of β and l respectively. Let U be $n-vector$ score function, A be an $n\times n$ information matrix.

$$U=\left(\frac{\partial\Pi}{\partial\theta}\right)$$

$$A=E\left(-\frac{\partial^2\Pi}{\partial\theta\partial\theta^T}\right)$$

Let also, the matrices of partial derivatives of function in equation (21) with respect to β and l to be define as

$$D=\frac{\partial\theta}{\partial\beta}$$

$$E=\frac{\partial\theta}{\partial l}$$

The to solution to the maximum penalized likelihood estimates $(\hat{\beta},\hat{1})$ is from the modified likelihood, $D^TU=0$, $E^TU=\lambda Kl$, which are nonlinear, hence the iterative solution via the Newton-Raphson algorithms with the expected second derivatives which iteratively replacing the trial estimates (β,l), at which U, A,D and E are evaluated by (β^{new},l^{new}), where

$$\begin{pmatrix}\mathbf{D^TAD} & \mathbf{D^TAE}\\ \mathbf{E^TAD} & \mathbf{E^TAD}+\lambda\mathbf{K}\end{pmatrix}\begin{pmatrix}\beta^{\mathbf{new}}-\beta\\ \mathbf{l^{new}}-\mathbf{l}\end{pmatrix}=\begin{pmatrix}\mathbf{D^TU}\\ \mathbf{E^TU}-\lambda\mathbf{Kl}\end{pmatrix} \tag{25}$$

equivalently equation (25) is written as

$$\begin{pmatrix}\mathbf{D^TAD} & \mathbf{D^TAE}\\ \mathbf{E^TAD} & \mathbf{E^TAE}+\lambda\mathbf{K}\end{pmatrix}\begin{pmatrix}\beta^{\mathbf{new}}\\ \mathbf{l^{new}}\end{pmatrix}=\begin{pmatrix}\mathbf{D^T}\\ \mathbf{E^T}\end{pmatrix}\mathbb{AY} \tag{26}$$

where

$$\mathbf{Y}=\mathbf{A^{-1}U}+\mathbf{D}\beta+\mathbf{El}$$

[34], showed that

$$\beta^{\mathbf{new}}=\left(\mathbf{D^TAD}\right)^{-\mathbf{1}}\mathbf{D^TA}\left(\mathbf{Y}-\mathbf{l^{New}}\right)$$
$$\mathbf{l^{new}}=\mathbf{S}\left(\mathbf{Y}-\mathbf{D}\beta^{\mathbf{new}}\right)$$

where $\mathbf{S}=\left(\mathbf{A}+\lambda\mathbf{K}\right)^{-\mathbf{1}}\mathbf{A}$ will always converged.

2.4 Model Diagnostics and selection of smoothing parameter

The goodness-of-fit of the model is assessed by deviance and its appropriate degrees of freedom is by

$$\triangle=2\left\{sup_\theta L(\theta)-L(\theta(\hat{\beta},\hat{l}))\right\}$$

$$v=n-tr(S)-tr\left[\left(D^TA\left(I-S\right)D\right)^{-1}D^TA\left(I-S\right)^2D\right]$$

[35], and the smoothing parameter is choosing via automatic algorithms of the generalized cross-validation,[36]

2.5 Data

This study adopts a secondary data on monthly temperature from the World Bank Climate Change Knowledge Portal (CCKP), an online tool that provides access to comprehensive global and country data information related to climate change and development on temperature of Ghana for period 1901 to 2016. The open R source is employed in analyzing the data.

3 Results and Discussion

From the Figure (1), we study the patterns of the monthly temperature for the period 1901 to 2016. The plot shows that, the data is characterized with an increasing trend and a periodic oscillations. To further understand these characterization,the

sample autocorrelation and sample partial autocorrelation plots are presenting in Figure (2) below.

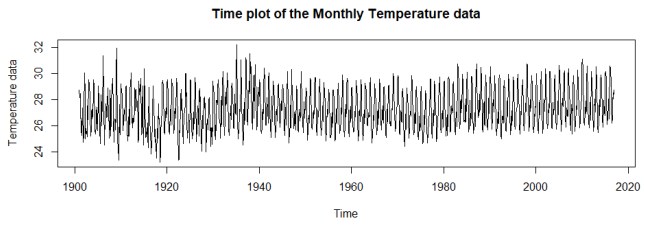

Figure 1. The time plot indicate that there is an increasing treng over period and clearly exhibiting periodic or cyclical patterns observed during period of consideration

The sample autocorrelation and sample partial autocorrelation plots,Figure (2), indicates the presents of trend and cyclical patterns, hence the oscillations observed therein.

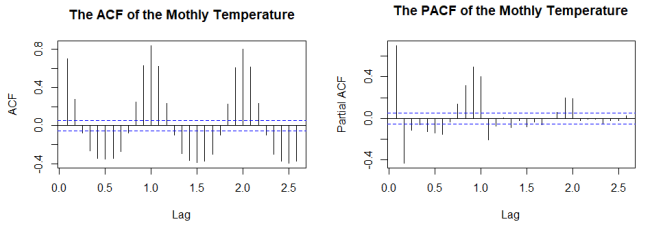

Figure 2. A critical observation of the ACF and PACF shows that the monthly temperature data is characterised with nonstationary tendencies as a results of the trend and the seasonal or cyclical variation inherent in the temperature data

To model the system, we first obtain the frequencies that adequately explains the cyclical patterns of the temperature data using the spectral analysis by ploting the periodogram to identify clearly the number of harmonic frequencies that describes periodic components in the Figure(3) as the frequency components to be used in equation (18)

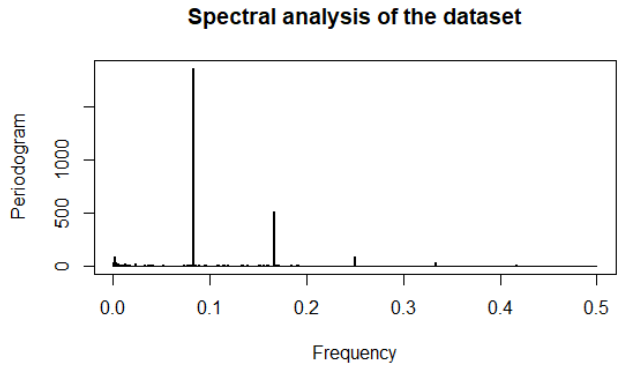

Figure 3. The periodogram shows that three significant harmonics(0.0842, 0.1667 and 0.0014) corresponding to the period of approximately 12, 6, and 714 respectively

Using the identified frequencies in the Figure3, we formulate the parametric component in the equation (21) and the time t for the duration between 1901 to 2016 as nonparametric smoothing component. Hence the semiparametric consideration in this study to adquately capture these components in the temperature data.

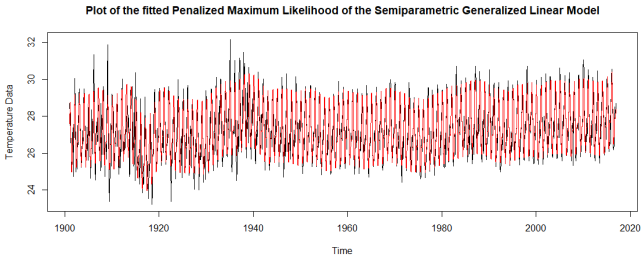

Figure 4. The fitted model approximately estimates the curve, this shows that the various components in the data set are adequately explained by cubic spline function, the Fourier functions for the trend and the cyclical/periodic components respectively.

From the observations in Figure 1 and Figure3, we fitted the temperature date with our proposed penalized maximum likelihood estimation of the semiparametric generalized linear models to obtain the fitted curve. The fitted curve and the temperature data are plotted in Figure 5. The fitted curve adequately fits the temperature data. Also, a plot of the distribution of residuals indicates that the residuals are approximately normally distributed or a white noise.

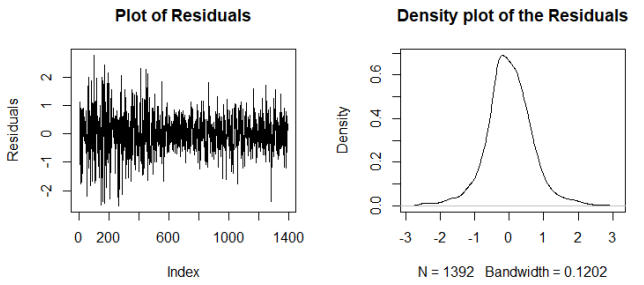

Figure 5. The residuals plot of the model indicates that, the penalized maximum likelihood estimates of the semiparametric generalized linear modeling of the Temperature data, adequately model the Temperature data, using the Cubic spline for the trend, Fourier function for the cyclical component. The density plot of the residuals also showed that , the residuals are approximately normally distributed

4 Conclusions

In this paper, we have proposed an approach of inference and diagnostics for the semiparametric additive generalized linear model. Specifically, we have derived a penalized maximum likelihood estimation of the additive component of the semiparametric generalized linear models, that is, of regression coefficients and smooth functions. Lastly, we have conducted a statistical modeling with real time series data set(Temperature data). The study has provided indications on the gain of using semiparametric modeling in situations where a Temperature data components can be additively decomposed in to various components, other component contributing differently to

the model. Thus, we recommend semiparametric additive penalized models as an option to fit time series data sets (Temperature data) in modelling the different component with different functions to adequately explain the relation inherent in time series data set.

Acknowledgments

The authors extend their appreciation to Pan African University Institute for Basic Sciences, Technology and Innovation and African Union for Supporting this research.

REFERENCES

[1] Sumway, R. H. and Stoffer, D. S. (2006). Time series analysis and its applications with r examples.

[2] IPCC (2001). Climate change 2001: The scientific basis. Contribution of working group i to the third assessment report of the intergovernmental panel on climate change [Houghton, J.T.,Y. Ding, D.J. Griggs, M. Noguer, P.J. van Der Linden, X. dai, K. Maskell, and C.A. Johnson (eds.)]. Cambridge University press, Cambridge, United Kingdom and New York, NY, USA, 881pp. Technical report.

[3] Romilly, P. (2005). Time series modelling of global mean temperature for managerial decision-making. *Journal of environmental management*, 76(1):61–70.

[4] Hansen, J., Sato, M., Ruedy, R., Lo, K., Lea, D. W., and Medina-Elizade, M. (2006). Global temperature change. *Proceedings of the National Academy of Sciences*, 103(39):14288–14293.

[5] Godfrey, A., Burton, M., and LeRoux-Rutledge, E. (2012). "africa talks climate". comparing audience understandings of climate change in ten african countries. *The handbook of global media research*, pages 504–520.

[6] Nerlove, M. (1964). Spectral analysis of seasonal adjustment procedures. *Econometrica: Journal of the Econometric Society*, pages 241–286.

[7] Godfrey, M. D. and Karreman, H. F. (1964). *A spectrum analysis of seasonal adjustment*. Citeseer.

[8] Grether, D. M. and Nerlove, M. (1970). Some properties of" optimal" seasonal adjustment. *Econometrica: Journal of the Econometric Society*, pages 682–703.

[9] Cleveland, W. P. and Tiao, G. C. (1976). Decomposition of seasonal time series: A model for the census x-11 program. *Journal of the American statistical Association*, 71(355):581–587.

[10] Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control (Revised Edition)*. Holden-Day, revised edition.

[11] Hillmer, S. C. and Tiao, G. C. (1982). An arima-model-based approach to seasonal adjustment. *Journal of the American Statistical Association*, 77(377):63–70.

[12] Harvey, A. C. and Todd, P. (1983). Forecasting economic time series with structural and box-jenkins models: A case study. *Journal of Business & Economic Statistics*, 1(4):299–307.

[13] Peter J. Brockwell, R. A. D. (2009). *Time Series: Theory and Methods, Second Edition (Springer Series in Statistics)*. Springer, 2nd ed. 1991. 2nd printing edition.

[14] Akcan, S., Kuvvetli, Y., and Kocyigit, H. (2018). Time series analysis models for estimation of greenhouse gas emitted by different sectors in turkey. *Human and Ecological Risk Assessment: An International Journal*, 24(2):522–533.

[15] Murat, M., Malinowska, I., Gos, M., and Krzyszczak, J. (2018). Forecasting daily meteorological time series using arima and regression models. *International agrophysics*, 32(2):253–264.

[16] Chawsheen, T. A. and Broom, M. (2017). Seasonal time-series modeling and forecasting of monthly mean temperature for decision making in the kurdistan region of iraq. *Journal of Statistical Theory and Practice*, 11(4):604–633.

[17] Cowtan, K., Jacobs, P., Thorne, P., and Wilkinson, R. (2018). Statistical analysis of coverage error in simple global temperature estimators. *Dynamics and Statistics of the Climate System*, 3(1):dzy003.

[18] Acquaotta, F. and Fratianni, S. (2013). Analysis on long precipitation series in piedmont (north-west italy). *American Journal of Climate Change*, (2):14–24.

[19] Ahmed, S. I., Rudra, R., Dickinson, T., and Ahmed, M. (2014). Trend and periodicity of temperature time series in ontario. *American Journal of Climate Change*, 3(03):272.

[20] Ye, L., Yang, G., Van Ranst, E., and Tang, H. (2013). Time-series modeling and prediction of global monthly absolute temperature for environmental decision making. *Advances in Atmospheric Sciences*, 30(2):382–396.

[21] Mudelsee, M. (2018). Trend analysis of climate time series: A review of methods. *Earth-science reviews*.

[22] Reinsch, C. H. (1967). Smoothing by spline functions. *Numerische mathematik*, 10(3):177–183.

[23] Bloomfield, P. (2000). *Fourier Analysis of time series an introduction*. Wiley Series in Probability and Statistics. Wiley-Interscience, 2? edition.

[24] Green, P. J. and Yandell, B. S. (1985). Semi-parametric generalized linear models. In *Generalized linear models*, pages 44–55. Springer.

[25] Green, P. J. and Silverman, B. W. (1993). *Nonparametric regression and generalized linear models: a roughness penalty approach*. Chapman and Hall/CRC.

[26] Eubank, R. L. (1999). *Nonparametric regression and spline smoothing*. CRC press.

[27] Rice, J. and Rosenblatt, M. (1981). Integrated mean squared error of a smoothing spline. *Journal of approximation theory*, 33(4):353–369.

[28] Rice, J. and Rosenblatt, M. (1983). Smoothing splines: regression, derivatives and deconvolution. *The annals of Statistics*, pages 141–156.

[29] Wahba, G. (1990). *Spline models for observational data*, volume 59. Siam.

[30] Yu, Y. and Ruppert, D. (2002). Penalized spline estimation for partially linear single-index models. *Journal of the American Statistical Association*, 97(460):1042–1054.

[31] Carroll, R. J., Fan, J., Gijbels, I., and Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association*, 92(438):477–489.

[32] Chen, J. (2010). *Semiparametric Methods for the Generalized Linear Model*. PhD thesis, Virginia Tech.

[33] LeBlanc, M. and Crowley, J. (1995). Semiparametric regression functionals. *Journal of the American Statistical Association*, 90(429):95–105.

[34] Green, P. J. (1985). Linear models for field trials, smoothing and cross-validation. *Biometrika*, 72(3):527–537.

[35] Green, P. J. (1985 b). Penalized likelihood for general semi-parametric regression models. Technical Report 2819, Mathematics Research Center University of Wisconsin.

[36] Wahba, G. et al. (1977). A survey of some smoothing problems and the method of generalized cross-validation for solving them.