GEOADDITIVE BAYESIAN MODEL FOR DATA WITH LIMITED SPATIAL INFORMATION

BY

ADENIKE OLUWAFUNMILOLA OLUBIYI

B.Sc(Unilorin); M.Sc. Statistics (Unibadan)

A Thesis in the Department of STATISTICS
submitted to the Faculty of Science in partial fulfilment of
the requirements for the Degree of

DOCTOR OF PHILOSOPHY

of the

UNIVERSITY OF IBADAN

April 21, 2015

GEOADDITIVE BAYESIAN MODEL FOR DATA WITH LIMITED SPATIAL INFORMATION

ADENIKE OLUWAFUNMILOLA OLUBIYI

Abstract

Large area estimation has been mostly accomplished using Geoadditive Models (GM) which combines the ideas of Geostatistics and additive models. The GM relaxes the classical assumptions of traditional parametric model by simultaneously incorporating linear and nonlinear, nonparametric effects of covariates, nonlinear interactions and spatial effects into a Geoadditive predictor. In the past, estimation of GM has been based on large area as a result of insufficient information in small areas. However, Bayesian approach allows out-of-sample information which can be used to augment the limited information in small areas. Hence, this study adopted the Geoadditive Bayesian model to estimate small areas with insufficient spatial information focusing on small district areas.

The GM by Kamman and Wand was specified by using Effect Coding (EC) to capture the spatial effect. The posterior was obtained by combining the likelihood (data) with the prior (out-of-sample) information. The likelihood and the prior information were assumed to be Gaussian and inverse gamma distribution respectively. The numerical solutions were obtained for the posterior distribution, which were not having a closed form solution, using Markov Chain Monte Carlo (MCMC) simulation technique. Finite difference and partial derivative methods were used to estimate other components of the Geoadditive Bayesian model. Kane analyser was used to collect vehicular emission (carbondioxide, carbonmonoxide and hydrocarbon). Information were also collected on age of vehicles, vehicle types (car and buses), vehicle uses (private and commercial) from 9211 vehicles for 3 years (2008-2011) covering 4 locations: Abeokuta, Sagamu, Ijebu-Ode and Sango-Ota. Data were also collected on respiratory health records of 9211 individuals (18 years and below) in six different hospitals on number of visits (nv) and diagnosis within the locality of the collection point of pollutants. Exploratory Data

Analysis (EDA) was carried out on emitted pollutants and age of vehicles. Autocorrelation plot was used to determine model performance.

The Geoadditive Bayesian model was :

$$exp[g_0(t) + \frac{1}{\sqrt{2\Pi\tau^2}} e^{\frac{-1}{2}\beta_j^2} \Sigma_{j=1}^{p} z_{ij} + \frac{1}{\sqrt{2\Pi\tau^2}} e^{\frac{-1}{2\tau^2}(\beta_j)^2}$$

$$+ \frac{1}{\sqrt{2\Pi\tau_j^2}} e^{\frac{-1}{2^2}(\beta^{spat})^2} + \frac{1}{\sqrt{2\Pi\tau_j^2}} e^{\frac{-1}{2^2}(\beta_{gi})^2}].exp \int_0^\infty exp(g_0(u) + \Sigma_{i=1}^{p} g_j(u) z_{ij}) du$$

where z_{ij},g_j, β_{spat} and β_j were non-linear time varying effect, linear time varying effect, spatial effect, and random component, respectively. The MCMC simulation technique gave the posterior means and the standard errors. This revealed that nv, diagnosis, vehicle uses, vehicle types jointly determine the health effect of pollutants on the individuals considered. Compared with Abeokuta individuals who lived in Sagamu (posterior mean = 0.036) were more likely to be affected by emitted pollutants while those in Sango-Ota (posterior = - 0.002) and Ijebu-Ode (posterior = - 0.015) were less likely to be affected. The EDA indicated non-linearity in the pollutants and age of vehicles. There were convergences of parameters at 250 Lag. A significant increase in the nonlinear effects was observed for age of vehicle (5years - 12years), Carbondioxide (10100 - 14400 ppm), Carbonmonoxide (0 - 25000 ppm) and hydrocarbon (4953 - 19812 ppm).

The derived Geoadditive Bayesian Model was found suitable and therefore recommended for estimating location effect of small areas with limited spatial information.

Keywords: Geoadditive Bayesian Model, Autocorrelation plot, Spatial Information.

Word Count : 467

Dedication

To the Almighty God and to my lovely children.

Acknowledgements

All praises are due to God Almighty, The Beginning and The End, for giving me strength and sound mind to carry out this research. The completion of this thesis would have been impossible without the assistance of many individuals.

First, I am grateful to Dr.A.A.Adelakun, my co-research-supervisor, who introduced me to so many people who helped me in the process of data collection and in obtaining the equipment used for collecting the emission parameters. I am indeed grateful for the fatherly and constructive role he played; may God almighty reward you abundantly.

Many thanks to Dr O.E. Olubusoye, the research supervisor, for his constructive criticisms, advice, fatherly role he played and encouragement at every stage of the thesis. I am indeed grateful to him for his assistance and patience in reading through the work. May God almighty bless you immensely sir.

To Mr. Kehinde, Mr. Kayode Lamina of Pescasen, and all the people who helped in the data collection, I say a big thank you. Also, thanks to the management of the various hospitals used for this study. I am particularly grateful for the opportunity they gave me to conduct the research and for the ethical approval given. I am also grateful to the following people who helped in one way or the other in the course of this work:, Mr. Ayoola Femi, Mr. Obisesan, Dr.(Mrs.) Otekunrin, Mr. Fadugba, Dr. Olujide Olubiyi, Pastor and Deaconess Dotun Olaitan, Dr. Ogbeide, Mr. & Mrs. Familusi and my colleagues at work. I am grateful for their useful suggestions, encouragement and support.

To my parents, Engr.& Mrs Ajibade, all my siblings and my in-laws for their support. I say a big thank you.

Finally, I say a big thank you to my lovely husband and my children

for their love, patience, understanding, support and encouragement even when the going got tough.

Certification

I certify that this work was carried out by Adenike Oluwafunmilola OLUBIYI in the Department of Statistics, University of Ibadan

. . . .

Supervisor

O. E. Olubusoye

B.Sc. Statistics (Ilorin), M.Sc., Ph.D. (Statistics) (Ibadan)

Department of Statistics,

University of Ibadan, Nigeria

Contents

List of Tables

List of Figures

List of Abbreviations

CO - Carbonmonoxide

CO_2 - Carbondioxide

COPD - Chronic Obstructive Pulmonary Disease

DEPs - Diesel Exhaust Particles

EA - Enumeration Area

EPA - Environmental Protection Agency

EMFAC - Emission Factor Model

FTP - Federal Test Procedure

HC - Hydrocarbon

HTBR - Hierarchical Tree Based Regression

LDV - Light Duty Vehicles

LGA - Local Government Area

NM-VOC - Non-methane Volatile Organic Components

NOx - Nitrogen Oxide

NO_2- Nitrogen dioxide

O_2 - Oxygen

O_3 - Ozone

PAH - Polyaromatic Hydrocarbon

PM_{10} - Particulate Matter

RES - Remote Emission Sensor

RSD - Remote Sensing Device

SO_2 - Sulphur dioxide

USEPA - United States Environmental Protection Agency

WCAS - Waterloo Centre of Atmospheric Sciences

WHO - World Health Organization

Chapter 1

INTRODUCTION

Vehicular emission remains a threat to environmental health, and it is expected to increase reasonably as vehicle ownership increases in the world. Over 600 million people globally are exposed to hazardous level of traffic - generated pollutants (UN, 1998). Human exposure to these air pollutants due to traffic is believed to constitute severe health problems, especially in urban areas where pollution levels are on the increase.
Vehicular transport represents a key factor in both urban and rural development as well as city revitalization worldwide.

In spite of the numerous benefits of motor vehicle transportation air pollution generated by motor vehicles remains a critical challenge in rural and urban areas. The major pollutants from traffic exhaust are particulates, nitrogen oxides, non-methane volatile organic components ($NM - VOC$) carbon monoxide, sulphur dioxides, polyaromatic hydrocarbon (PAHs) and lead. These pollutants are regularly released into the air in many parts of the world and, as said before, they constitute an important health hazard. For example, in the United States in 1993 alone, transportation sources were reported to be responsible for 7% of CO emissions, 45% of NOx, 36% of volatile organic compounds, and 22% of particulates. In the European Union, pollution control measures have been initiated over the past 20 years to reduce NO_2 levels, but these measures have been offset by increase in the number of vehicles on the road. The consequences of this was that in the UK, for example, average concentrations of NO_2 increased emissions by motor vehicle traffic. In the developing world, automotive air pollution is a problem mostly in large cities with regular heavy traffic congestion. Such cities include Mexico City, Bangkok and Lagos, among others. However, even in certain small urban centres such as Peshawar, Pakistan and Katmandu, Nepal, air pollution from motor vehicles remains a serious threat to human health. In addition to increased emissions by motor vehicle traffic, there are other stationary sources of air pollution which constitute significant threat to air quality.

All over the world, air pollution due to traffic congestion constitutes up to $90-95\%$ of the ambient CO levels, 80-90% of NOx, hydrocarbon and particulate matter, thus posing a serious threat to human health (Savile, 1993). Research conducted a decade and half ago has shown that transportation sources in the USA were responsible for 77% of CO levels, 80 - 90% of NOx, 36% of volatile organic compounds and 22% of particulate matter (USEPA, 1993). Similarly, in the UK, the average concentration of NO_2 was found to increase by 35% from 1986 to 1991 due to increase in vehicular emission (CEC, 1992). On the global level, various scholars (e.g. Seneca and Tausig, 1994 and Faucet and Sevingny, 1998) have all arrived at the same conclusion that transportation is the major culprit in air pollution, accounting for over 80% of total air pollutants.

Incomplete combustion in vehicular engines leads to the emission of carbon monoxide and a wide range of hydrocarbons, including aromatics and oxygenated species such as aldehydes. In addition, nitrogen oxides, which arise from the reaction of nitrogen and oxygen at high temperature produced in the combustion chamber, are also emitted in the exhaust gases. The use of lead and other metal- based compounds as octane improvers is no longer permitted in most countries. As a result, emissions of lead and other compounds have decreased in the past few years. Emission of nitrogen oxides is a function of fuel composition, engine type, and the power/ load conditions on the engine type.

The strong emissions of the vehicles are caused by leaking valve shafts and loose piston which lead to leaking of oil into the combustion chambers of the engines. When the oil does not burn away completely it produces blue plumes, bad odour and unburnt hydrocarbons. This situation occurs most often while a vehicle is descending a hill, which is when the engine operates like a pump and sucks oil into the burning chamber. The impact of motor vehicle emission extends far beyond the local area. The transportation sector is the most rapidly growing source of greenhouse gas emissions that is, emissions of chemicals that have the potential to contribute to global warming. These include CO_2, chlorofluorocarbons, NO, and CO. In 1990, about 22% of CO_2 emissions from fossil fuel use came from the transportation sector, especially in developed countries. However, the share of emissions from developing countries is expected to rise in the future because of the growing number of motor vehicles which use less efficient fuel-burning technologies in these countries.

Cities embody the diversity and energy of human pursuits. Urbanization brings about increase in population, which lead to corresponding increases in motor vehicles, either for private or for public transportation. The environmental costs of motor vehicle are hard to measure and vary according to local conditions. Also, most of the health hazards are as a result of increased mortality due to the presence of volatile organic compounds, NOx and SOx in the inhaled air. The rest of the hazards are due to minor illnesses from ozone (O_3), formed in the atmosphere from volatile organic compounds and NOx.

This is a clear indication that vehicle emissions are a major source of ambient air pollution and must be controlled if acceptable air quality is to be assured. In addition, there are numerous health problems associated with high concentration of these pollutants. For example, NO_2 is responsible for immune system impairment, exacerbation of asthma and chronic respiratory diseases: as well as reduced lung function and cardiovascular disease (Schwela, 2000). Particulates are dangerous and have been identified as facilitators in the development of lung cancer and increased rate of mortality (Schwela, 2000).

The two types of vehicle emissions are exhaust emission and evaporative emissions. The three major pollutants (HC, NOx, Carbonmonoxide (CO) and carbondioxide (CO_2) are exhaust pollutants. These major pollutants have direct impact on human health, but the CO_2 emissions do not directly impact human health. The impact of HC emission on human health includes respiratory problems, eye irritation and potential to cause cancer. CO emission reduces the flow of oxygen to the blood (Carboxyhemoglobin) which is responsible for heart diseases. NOx is one of the pollutants in the formation of Ozone and it contributes to the formation of acid rain (O_2 does not directly impact human health but is a green house gas which traps heat on earth and causes global warming.

Policy makers all over the world have been partially successful in imposing policies aim at guaranteeing air quality. In the US, the ambient levels of most pollutants have been reduced steadily since the 1960s while Europe has lagged behind, the US in emission control on motor vehicles. Africa is worse off in the attempt to eradicate vehicular pollutant. In Nigeria, the government has banned the importation of vehicles that have been in use for more than eight years. Good as this policy may be, what remains to be done is decide how to control the emissions from the existing old vehicles plying the streets and highways of Nigeria. Policies aim at reducing overall vehicle use, so as to minimize congestion or pollution, must be enforced. However, these policies really do little to reduce the twin effect of congestion and pollution. The problem of congestion is specific to location and time, whereas emissions are specific to vehicle characteristics and driving behaviour.

1.1 Background to the Study

Vehicle emission model has attracted considerable research attention and witnessed some improvement over the years. Several models have been developed by various researchers but majority of them only looked into the estimation of emissions from highway/ on-road vehicles while leaving out an assessment of the effect of the emission on the people living in the area. Over the years, research has suggested that existing mobile-source-emissions model accuracy in estimating emissions from highway vehicles is limited; some researchers have called into question the structure of existing models. A variety of problems in the existing model has been identified by researchers during the past five years. The most significant findings indicate that two major contributions to on-road emissions have been largely overlooked by existing data collection and analytical methods employed.

The model development process: superemitters, or vehicles that produce the very highest emissions rates under all operating conditions, contribute disproportionately to fleet emissions and are under-represented in the modeled vehicle fleet. High-power demand operations (hard accelerations, moderate accelerations at high speed, high-speed activity for vehicles with a low power-to-weight ratio, and so forth) have been identified as significant emissions-producing activities not adequately captured by existing modeling methods (Le Blanc et al 1994, Carlock 1992, Benson 1989, Groblicks 1990).

A single burst of or sharp acceleration may cause as much pollution as does the entire remaining trip (Carlock, 1993). These superemission episodes are known as enrichment events because excess fuel is delivered to the engine cylinders (thus enriching the fuel mixture). Superemitters and superemissions must both be addressed in modelling efforts. Current mobile emissions provide poor spatial estimates of emissions because they use regional default data for various factors that affect vehicle emissions rates (e.g fleet composition and vehicle operating modes). The aggregate default data does not allow variation over time and space.

Some of the traffic simulation and optimization models, such as TRANSTT-7F (Penic and Upchurch,1992), INTEGRATION (Van Aerde,1994), FREQ (Imada and May), NETSIM (Rathi and Santiago,1989), and INTRAS (Wicks and Liebermann,1980), have incorporated their own emission estimation methods, but none of these methods was tested or validated for the on-road driving vehicles and conditions. There are ongoing research efforts with respect to the development of new generation of modal emission models in the University of California at Riverside (An et al.,1997, Barth et al.,1997) and George Institute of Technology (Bachman et al.,1997).

The development of advanced infrared remote emission sensing technology brings a cost-effective and convenient instrument for collecting on-road vehicle exhaust emissions. Although, initially, the Remote Emission Sensor (RES)was proven to be useful in screening for High Emitter Vehicles (HEV) on the road (Bishop et al.,1994, Sorbe,1995, Jack et al.,1995), there are many advantages to the use of RES in emission model evaluation and emission model development. This is because the emission data collected by

RES will naturally reflect the on-road vehicle fleet combinations and current vehicular technologies.

Current research support the development of a modal model, which is based on the premise that modelling emissions from specific modes of vehicle operation will more accurately reflect on road emissions. Mobile source emissions are associated with specific engine and vehicle operating modes, each of which has a different emissions characteristic. Vehicle activity data and emissions rate factors are required for each emission-related operating mode, because modal activity data include a wide variety of measures and attributes, not simply the commonly used vehicle-miles travelled; a modal modelling regime will be much more complex than the models currently used.

The potential number of operating modes that would be necessary to incorporate into modal modelling was perceived to be large. Also, the availabilty of second-by-second data were inadequate for identifying causal relationships or developing reliable modal emission-rate algorithms. Preliminary research indicates that the relationships between operating modes and emissions for any given vehicle are complex. Emissions rates for certain models are affected by a variety of vehicles, environmental and perhaps even driver attributes. Numerical variables must be explicitly handled in a modal model, such as fleet composition, the incidence and types of superemitters, network characteristics, driver behaviour, and even fuel consumption. As complex as these factors might seem, however, new monitoring and modelling capabilities have made possible the development of advanced modal models.

The current mobile model, mobile 5a, also known as the EPA mobile source emission factor model, is a computer program that estimates the emissions of carbonmonoxide, hydrocarbon, and nitrogen oxides for eight different types of gasoline-fueled and diesel highway motor vehicles. The model consists of an integrated collection of mathematical equations and assumptions about the emissions from vehicles manufactured from 1960 to 2020. These are generally the cars produced in the 25 most recent model years which are assumed to be in operation in any given calendar year. The first mobile model was made available for use in 1978 (Abhishek, 2007); since then major updates and improvements to the model have been made and quite a bit more is now known about the complexity of the factors affecting vehicle emissions, as measurement devices have improved, and as more data have been collected. According to the agency, the improvements have resulted in the refinement of emissions estimates for evaporative emissions (such as occur when the fuel tank and system heat up) for the uncorrected in-use deterioration (wear and tear) that results from poor vehicle maintenance or tampering and for other factors.

In its simplest form, EPA's Mobile model allows the model user to produce a number-an estimated quantity of emissions for the three pollutants of concern-by multiplying the estimated emissions per mile for an average urban trip by the estimated number of trip miles travelled in an area. Over the years, researchers have learnt that vehicle emissions are highly complex. In order to compensate for the complexity and other emission-producing activities, EPA has periodically adjusted its basic formula-through the use of revised "correction factors" to approximate vehicle exhaust emissions in a range of situations. In essence, the correction factor is a multiplier added to the basic formula (miles travelled times emissions rate per mile), to adjust the model's output to closely reflect actual emissions. For the states, the Mobile model is a tool for constructing emissions inventories, creating control strategies,producing state implementation (SIP) and-subsequently-demonstrating control strategy effectiveness to EPA and others. For EPA, the Mobile model is a tool for evaluating the adequacy of a state's emissions inventory estimate, motor vehicle control strategies, and implementation plans.

In essence, the model's estimates provide EPA regulators with critical information that is used to evaluate the adequacy of a state's programme and the relative benefits of various policies to control motor vehicle emissions. Additionally, the model's estimate can affect state policy decision on issues such as the content and volatility of fuels, and some decisions on highway projects.

The Mobile model has been updated about 10 times since its introduction in 1978.

The following are limitations of Mobile 5a model:

1. Emission estimates for higher speeds, especially speeds in excess of 65mph
2. Representation of emissions from rapid acceleration and deceleration including aggressive driving behaviours
3. Representation of emissions immediately after engine start-up, known as cold start emissions.
4. Representation of emissions from air conditioner usage.
5. Representation of emissions from road grade such as when a car climbs a hill.
6. Representation of high emitting vehicles in the MOBILE model's supporting database.
7. Representation of emissions from lower polluting fuels, especially fuels with lower volatility
8. Representation of emissions system deterioration for vehicles with 50,000 or more odometer miles
9. Emissions estimates and assumptions for vehicle inspection and maintenance (I&M) programmes.
10. Estimates and assumptions for non-tail pipe evaporative emissions when the vehicle is not operating.
11. Emissions estimates assumptions for the inspection and maintenance (I&M) of heavy duty vehicles-those with a gross vehicle weight of 8,501 pounds or more.
12. Data characterizing vehicle fleet.
13. Greater distinctions in roadway classifications.
14. Quantifying the uncertainty of the model's estimates.

MOBILE model and Emfac model (used only in California) require the average speed as the sole descriptor of a vehicle's modal events and driving conditions. They are not sensitive to vehicle's modal events, such as acceleration/deceleration, cruise speed and idling; they cannot be used to effectively evaluate the traffic control and management strategies that are aimed at reducing vehicle emissions.

Also, the emission factors in MOBILE and EMFAC are derived from the FTP driving cycles of in-laboratory emission testing. Their capabilities in representing the vehicle emissions for the on-road driving conditions were not extensively investigated.

Aside these limitations, a number of independent evaluation field studies on the MOBILE models have indicated the unreliability of their results.

The need to cater for the shortfall of the Mobile model and get more accurate vehicle emission factor model for air quality modelling as well as the assessment of health effects, prompted the Waterloo Centre of Atmospheric Sciences (WCAS)'s decision to develop a detailed micro-scale model, MicroFac. This micro-scale model can provide accurate emissions from a vehicle fleet under specified conditions of meterology, fleet composition, vehicle age, distribution and speed. The Microfac emission model is now being coupled with high resolution dispersion models to produce a tool that will give the local concentrations of these emissions with high temporal and spatial resolution. These high resolution local concentrations will be incorporated into regional models in the future via a dual-kernel local-regional modelling procedure. This system will treat the chemistry and dispersion of the primary and secondary vehicle-related pollutant with the resolution of the local dispersion model and transfer them to the regional model when their spatial extent is comparable to that model's resolution. This will greatly increase the accuracy with which their effects on regional air quality can be determined.

Another model developed is the GIS-based modeling approaches which address the limitation of the current model. The GIS approach described requires significant time and effort to produce the data required. Cost associated with developing GIS-based emissions models, primarily those of model development, standardization, and integration of new data sources, are likely to be high. The GIS system, with integrated modeling and improved data-handling capabilities, will yield significantly improved spatial and temporal allocation of vehicle activity and emissions.

The limitations of GIS based model revolve mostly around the intensity of data required.

1.2 Statement Of the Problem

Some studies have shown that some pollutants are closely associated with human health. One of these pollutants is carbonmonoxide, which is one of the gases produced in the course of gas flaring. Carbonmonoxide poisoning can be difficult to diagnose because the symptoms mimic other illnesses. Both humans and animals exposed to gas flaring are at risk of carbonmonoxide poisoning and this could result in chronic health disease, anaemia or respiratory problems.

Diesel Exhaust Particles (DEPs) is another form of pollutant which is generated by heavy duty diesel engines in various industries. It can absorb 450 different organic compounds, including mutagenic and car urogenic Polycyclic Aromatic Hydrocarbon (PAH). These particles can remain unborn for a long period of time and will be deposited in the lungs. Because DEPs are a major component of particulate air pollution in most industrialized urban areas their effects on pulmonary infections are of great environmental and occupational concern. In Nigeria such effects can only be imagined. Studies have shown that chemical exposures can affect immunity in two major ways: by causing hypersensitivity reactions including allergy, which can be harmful to organs and tissues; and autoimmunity, in which immune cells attack self or by causing immunosuppression, a reduction in the responses and activities of the immune system. With increasing concern about environmental risk, the demand for reliable information has grown with alarming rapidity. Also, the need to get a suitable model to look at the vehicular emission and its effect on the populace, especially children (who are more susceptible to the health effects of these emission due to their weak immune systems and developing organs) has become very urgent indeed. This study will use Geoadditive Bayesian model in looking at the effect of essential vehicular pollutants [CO, CO_2,and HC] on the health of children.

1.2.1 Geoadditive Model and its problems

A geoadditive model uses mostly recorded observations in which there is provision for the assessment of geographical information of the location. Some of the limitations of geoadditive model include the inability to make use of the spatial effect when the number of districts/Local Government Areas considered is limited/small or when the units/districts considered are not neighbours. Also, it gives no significant difference when the district are not large.

For this research work, an effect coding method will be used for the spatial effect because of the small number of the local government areas and the fact that they are not neighbours. This makes the geoadditive model inappropriate for estimating the spatial effect using the longitude and latitude of the location.

1.2.2 Gap in Literature

Past studies have concentrated mostly on the use of the usual linear regression models which, however, are not appropriate in situations where effects of vehicular emission on the populace are to be determined.

Majority of the vehicle emission models developed by various researchers have only looked into estimating emissions from vehicles, leaving out the effect. The current mobile emissions provide poor spatial estimates of emissions because they use regional default data for various factors that affect vehicle emissions rates.

Another study carried out in Lagos by Ojolo et al. (2007) was based on the survey of the effects of vehicular emissions on human health in Nigeria. Data was elicited through the administration of copies of a questionnaire. Their analysis was also based on descriptive study through which the effect cannot be determined since their conclusion was based on observational information about the locations. There are problems with their methodology, which only considered the emission part and based the effect on observational information. As a result, the geoadditive Bayesian model that encompasses most of the known regression models and improved on their shortcomings using effect coding for the spatial effect was developed.

In the past, the use of Geoadditive model had been based on large areas because of sufficient information which is lacking in small areas. However, the Bayesian approach allows out-of-sample information which can be used to augment the limited information in small areas. Therefore, this study adopted the Bayesian technique to estimate small areas with insufficient spatial information, focusing on small district areas that are not neighbours.

1.2.3 Possible Research Questions

- Can Geoadditive model be used to determine the effect of vehicular emission on children?
- Can the model be used in the absence of sufficient Spatial information?
- Can the model be applied to data on location using effect coding to capture spatial effect?
- Can the model account for location effect?
- Will the model be able to combine the environmental effect and the health effect?

1.3 Aim and Objectives

The main objective of this research is to establish the effect of exhaust emission on children, using the effect coding for the spatial effect. However, the specific objectives are to:

- construct a geoadditive Bayesian model from the parent geoadditive model.,
- estimate and determine the components of the Geoadditive Bayesian Model.,
- apply the Geoadditive Bayesian Model to real life data on vehicular emission in Ogun state and investigate the relationship between the health of children and emitted pollutants

1.4 Justification of the Study

Environmental data are frequently asymmetrical and skewed to the right, having a long tail towards high concentrations. So the validity of classical assumptions of fixed covariate effects in traditional linear model are too rigid and restrictive. Consequently, there is the need for a more flexible approach that relaxes this assumption and a possible solution. A Geoadditive Model relaxes the classical assumptions of traditional parametric model by simultaneously incorporating linear and non-linear nonparametric effects of covariates, nonlinear interactions and spatial effects into a Geoadditive predictor. Also,with vehicular emission accounting for about 60% of the total emitted pollutants (Kpako, 2003), a study such as this is imperative indeed.

This research then considers a situation where the usual rigid assumption of additivity/linear model is extended to incorporate the non-linear covariate effect and at the same time considers the spatial effect. Geoadditive model(Kamman and Wand 2003) combine the idea of kriging and additivity while accounting for nonlinear covariates effects. Therefore, the Geoadditive model by Kamman and Wand was specified using effect coding to capture the spatial effect.

1.5 Definition of Relevant Terms

MARKOV RANDOM FIELD: This is a model in which a set of random variables have a Markov property described by an undirected graph. A Markov random field is similar to a Bayesian network in its representation of independence. Markov random field is popular when space is split into discrete contiguous geographic units (districts of a town, for example). In this case, a simple smoothing penalty is constructed based on the neighbourhood structure of the geographical units.

PRIORS FOR PARAMETERS: In classical inference, the sample data y is taken as random while population parameters θ, of dimension p, are taken as fixed. In Bayesian analysis, parameters themselves follow a probability distribution, knowledge about which is summarized in a prior distribution (θ). Often, a prior amounts to a form of modelling assumption or hypothesis about the nature of parameters, e.g in random effect models. A prior specifying the errors as spatially correlated is likely to be a working model assumption, rather than a true cummulation of knowledge. In many situations, existing knowledge may be difficult to summarize or elicit in the form of an informative prior and to reflect such essential prior ignorance, resort is made to non-informative priors. Examples are flat priors, that is, a parameter is uniformly distributed between $-\infty$ and ∞. It is possible that a prior is improper (does not integrate to 1 over its range); such priors may add to identifiability problems (Gelfand and Sahu, 1999) and so, many studies prefer to adopt weakly informative priors which are just proper.

POSTERIOR DENSITIES: In classical approaches such as maximum likelihood, inference is based on the likelihood of the data alone. In Bayesian models, the likelihood of the observed data y given parameters θ, denoted by $f(y/\theta)$ equivalently $L(\theta/y)$, is used to modify the prior beliefs (θ), with the updated knowledge summarized in a posterior density, (θ/y). Thus, updated beliefs are a function of prior knowledge and the sample evidence.

ASTHMA: Asthma is a chronic inflammatory disease of the airways which is associated with reversible airway obstructive, hyper responsiveness to triggers, clinical symptoms of wheezing, chest tightness, or cough and increased mucous production. It is a major respiratory illness among children and disproportionately affects minorities. Most children diagnosed with asthma have mild to moderate symptoms; however, there are those whose symptoms result in numerous visits to the hospital emergency room and multiple hospitalizations.

Chemicals in vehicle exhaust are harmful to asthmatics. Exhaust can adversely affect lung function and may promote allergic reactions as well as airway constriction. All vehicles, especially diesel engines, emit very fine particles that deeply penetrate lungs and inflame the circulatory system, damaging cells and causing respiratory problems. Even short-term exposure to vehicle exhaust may harm asthmatics because asthmatic children are particularly sensitive to air pollution. New England status has some of the highest asthma rates in United States with about 9 percent of Connecticut's youth having the disease. Inhalation of vehicle emissions, even for short periods, may be harmful to asthmatics. A study revealed that children are 40

percent likely to have an attack on high outdoor pollution days.
CHRONIC OBSTRUCTIVE PULMONARY DISEASE (COPD): This is also known as chronic obstructive lung disease and encompasses two major disorders; emphysema and chronic bronchitis. Emphysema is a chronic disorder in which the walls and elasticity of the alveoli are damaged. Chronic bronchitis is characterized by the inflammation of the cells lining the inside of the bronchi, which increases the risk of infection and obstructs the airflow in and out of the lung. Smoking is responsible for approximately 80% of COPD cases while other forms of air pollution may also influence the development of these diseases. Symptoms include cough, production of mucous and shortness of breath. It is important to note that no cure exists for people suffering from COPD although healthy lifestyle and appropriate medication can help. Vehicle emissions are particularly harmful to people afflicted with COPD, such as chronic bronchitis. Significant and replicated associations have been found between increased Ozone levels and a range of adverse effects on the lungs. Several studies have shown an increased risk of hospital admission from COPD associated with high ozone level. There is also a relationship between the levels of PM_{10} and morbidity in patients with COPD. These associations were noted in Philadelphia, in the United States, where the major source of these particles is motor vehicles. Fine particle matter is especially harmful to the people with COPD and has been found to increase their hospital admission rates. High levels of PM_{10} are also associated with increased morbidity among those with the illness.
CARDIOVASCULAR DISEASE: Mortality and hospital admissions for myocardial infarction, congestive cardiac failure and cardiac arrhythmia increase with a rise in the concentrations of particulate and gaseous pollutants. As concentrations of airborne particles increase, those with cardiovascular disease may experience increasing severity of symptoms, rates of hospitalization and mortality. The risk of having a heart attack is greater for people exposed to pollution from heavy traffic, as well as for those living near air-polluted roadways.
KRIGING: The term kriging refers to a widely used method for interpolating or smoothing spatial data.Given a set of data y_i, i=1,....,n, at spatial location X_i,$X\epsilon\Re^2$, the simple kriging model for interpolating the underlying spatial

surface is

$$y_i = \mu + S(x_i) + \varepsilon_i \tag{1.1}$$

where S(x) is a zero-mean stationary stochastic process in $\Re^2$ and the ε_i are assumed to be independent zero-mean random variables with common variance σ_ε^2 and distributed independently of S (Cressie,1993). Interpolation at an arbitrary location $X_0 \epsilon \Re^2$ is done through

$$\hat{y}_0 = \bar{y} + \hat{S}(x_0) \tag{1.2}$$

where $\hat{S}(x_0)$ is the best linear predictor of $S(x_0)$ based on the data in y. For a known covariance structure of S, the resulting predictor is

$$\hat{S}(x_0) = C_0^T(C + \sigma_\varepsilon^2 I_n)(y - \mu I) \tag{1.3}$$

where:

$$C \equiv \begin{bmatrix} S(x_1) \\ \vdots \\ S(x_n) \end{bmatrix} \; and \;\; C_0 = \begin{bmatrix} cov\{S(x_0), & S(x_1)\} \\ \vdots & \\ cov\{S(x_0), & S(x_n)\} \end{bmatrix}$$

The practical implementation of (1.2) requires the definition of the covariance structure of S(x). The usual approach is to define a parsimonious model for Cov{S(x),S(x+h)}, estimate the required parameters to derive the estimates of C and $\hat{c}_0$ and then substitute in (1.1) to obtain:

$$\hat{y}_0 = \bar{y} + \hat{c}_0^T(\hat{C} + \hat{\sigma}_\varepsilon^2 I_n)(y - \bar{y}I) \qquad 1.3$$

usually a common assumption to simplify the covariance structure of S is the assumption of isotropy, that is

$$Cov\{S(x), S(x+h)\}$$

depends only on

$$\|h\| \tag{1.4}$$

this is a stronger assumption than stationary, because it says that the covariance is independent both of location and direction, and sometimes it could not be a valid condition (1.4) implies that

$$C = [c(\|x_i - x_j\|)]_{1 \le i,j \le n}$$

$$C(r) \equiv \sigma_s^2 C_0(r), \sigma_s^2 \equiv var[s(x)],$$

with C_0=1. The functions C and C_0 are respectively the covariance function and the correlation function of the isotropic process S(x) and they should be chosen to ensure that C is a valid covariance matrix.

1.6 Organization of the Thesis

This chapter, which is the first, has been devoted to a discussion of the current existing model for vehicular emissions, the statement of the problem, objectives of the study, justification for the study and definition of terms. Chapter two deals with the literature review for environmental data, models for vehicular emission and review of geoadditive model. Chapter three discusses the theoretical framework of the model and chapter four entails teasing out the methodology for the research. Chapter five shows the result of the various analyses while chapter six brings up the rear by summarizing the geoadditive Bayesian model and the model application. It also indicates, the contribution which the current study has made to knowledge, pointing out areas for further research and drawing a general conclusion for the study.

Chapter 2

REVIEW OF LITERATURE

2.1 Introduction

This chapter is divided into three sections: the first section discusses different models for environmental data, the second section focuses on different models for vehicular emissions and the last section is concerned with a review of different areas of application of geoadditive model.

2.2 Review of Models for Environmental Data

Over the past ten years, hundreds of studies have been published in peer-reviewed literature to demonstrate the special vulnerability to air pollution that exists among susceptible population with serious illnesses. Tens of millions of Americans suffer from these illnesses, which include asthma, chronic obstructive pulmonary disease (COPD), cardiovascular diseases and lung cancer. Also, at special risk are children, the elderly, those with compromised immune systems and those with specific generic traits. During the past decade, scientists have confirmed a relationship between the two forms of air pollution, ozone and particulate matter, and increased rates of mortality, especially among those with cardiovascular disease.

Bobak & Leon (1992) studied infant mortality based on a consideration of the ecological situation in the Czech Republic. They found an association between sulfur dioxide and Total Suspended Particles (TSP) on the one hand and infant mortality on the other, after controlling a number of potentially confounding variables (at the ecological level). The effects were specific to respiratory mortality in the post-neonatal period. These results were later confirmed in a nationwide case-control study based on the Czech national death and birth registers; this design allowed one to control for social and biological covariates at the individual level. The study found a strong effect of sulfur dioxide and TSP on post-neonatal mortality from respiratory causes: the relative risks, per $50\mu g/m^3$ increase in pollutant concentration, were 1.95 (95% CI 1.09 - 3.50) for sulfur dioxide and 1.74 (95% CI 1.01 - 2.98) for TSP.

In a similar way, Wjst et al. (1993) investigated traffic flow around a child's school. The traffic count method has the advantage of being likely to be a more valid measure than distance to roads. It is worth the effort, however, considering the daily movements of an individual. Throughout the day, an individual travels between home and work or school, experiencing a number of different exposure levels on the way. Recreational activities may also subject a person to different levels of exposure. Indeed, even within the residential area, exposure may vary depending on the time one spends indoors or outdoors. Indoor exposure to NO_2 may be high, with levels possibly higher than outdoors if a gas stove is used in the home. Another point to consider is the type of traffic exposure. Emissions vary greatly between cars and trucks. Some have approached this by analyzing data from different vehicles separately, suggesting truck pollution to be more detrimental to health than that from cars. It could be suggested that as car and truck pollution varies, for example, trucks produce a lot more particulate matter consisting of diesel particles than cars, that perhaps one should consider the effects of particulate separately from those of NO_2. This, however, is not without difficulties; as if one is exposed to traffic there will be a combined effect from a cocktail of pollutants produced by both trucks and cars.

A number of studies have used modeling to estimate pollution exposure. A model is capable of taking into account a whole range of factors that may affect exposure. As illustrated by Pershagen et al. (1995), exposures both at home and at day-care centres or for others at school or work can be considered, with these being adjusted for the time spent in each location. Factors considered in the models used in the studies have included vehicle type and density, presence and type of buildings on a street, meteorological conditions, street width and distance from house to the middle of the street, amongst other factors.

Even within a model, however, accounting for personal day-to-day exposures is still problematic. In order to take previous exposures into account, a cohort study would be necessary. Certainly, if one is trying to account for the prevalence of a disease like asthma, knowing previous exposure levels prior to the onset of the disease is important. To do this, one would need to look at the previous resident and ascertain the day-to-day exposures of that person during his/her life. An alternative approach would be to use a personal monitoring system. Both methods of assessing long-term exposure, however, would be very expensive. One could consider the use of monitoring stations already in place throughout the cities. The problem with using such stations is that they are generally widely dispersed while pollution levels may vary substantially within short distances, e.g. exponential decline in the concentration of certain pollutants with increasing distance from busy roads. Installing sufficient monitoring stations to adequately capture spatial variation in levels of pollution encountered over short distances, would be both impractical and expensive.

Vliet et al. (1997) examined whether motor vehicle exhaust from freeways has an effect on the respiratory health of children. A cross-sectional study was conducted using children attending schools situated less than 1000 m from major freeways in the Province of South Holland. The selected freeways carried between 80,000 and 150,000 vehicles per day. Separate counts for truck traffic indicated a range from 8000 to 17,500 trucks per day. A total of 13 schools, from which 1498 children were drawn, were asked to participate. From these children, 1068 usable questionnaires were obtained. Chronic respiratory symptoms reported in the questionnaire were analyzed with logistic regression. Distances from the freeway and (truck) traffic intensity were used as exposure variables. Cough, wheeze, runny nose, and doctor-diagnosed asthma were significantly more often reported for children living within 100 m from the freeway. Truck traffic intensity and the concentration of black smoke measured in schools were found to be significantly associated with chronic respiratory symptoms. These relationships were more pronounced in girls than in boys.

Loomis et al. (1999) conducted a time-series study of infant mortality in the south-western part of Mexico City from 1993 to 1995. Exposure included nitrogen dioxide, sulfur dioxide, ozone and particulate matter with particle size $< 2.5\mu m (PM_{2.5})$. A $10\mu g/m^3$ increase in the mean level of fine particles during the previous three days was associated with a 6.9% (95%, CI $2.5 - 11.3\%$) excess increase in infant deaths.

Buckeridge et al. (2002) worked on the effect of motor vehicle emissions on the respiratory health in an urban area. They developed an exposure model and implemented it using a geographic information system to estimate the average daily census enumeration area (EA) exposure to PM (2.5). Hospital admission diagnostic codes from 1990 to 1992 were used to measure respiratory and genitourinary conditions. Effect of EA exposure on hospital admissions was assessed using a Poisson mixed-effects model and the spatial distributions of the variables. It was shown that exposure to PM (2.5) has a significant effect on the admission rates for a subset of respiratory diagnoses (asthma, bronchitis, chronic obstructive pulmonary disease, pneumonia, upper respiratory tract infection) with a relative risk of 1.24 at 95% C.I. also a weaker effect of exposure on hospitalization for all respiratory conditions, and no effect on hospitalization for non-respiratory conditions.

Gauderman et al. (2004), examined the effect of air pollution on the lung development from 10-18 years, adopting a two-stage regression approach to relate the longitudinal pulmonary-function. The first stage model was a regression of each pulmonary-function measure on age to obtain separate, community-specific average growth curves for girls and boys. They accounted for the growth pattern using a linear spline model.

The second-stage model was a linear regression of the 24 sex and community-specific estimates of the growth in lung function over an eight-year period on the corresponding average levels of each air pollutant in each community. Inverses of the first-stage variances were incorporated and the model was refitted to estimate the sex-average effect of the pollutant. They were able to show that the ozone contributes to the acute health effects and that exposure to ambient air pollution is correlated with significant deficits in respiratory growth over an eight-year period.

Ferguson et al. (2004) considered the increasing prevalence of asthma and the effects of air pollution on asthma using modelled exposure approach.

The routine surveillance system recording spatial variation in pollutants level allows improved understanding of the link between road-traffic pollution and asthma. This could be used to help predict future health impact, particularly in cities and towns.

Another study carried out on Traffic-related air pollution near busy roads using East Bay children by Kim et al (2004) revealed that there were differences in concentrations between schools nearby and those more distant from major roads. Using a two-stage multiple-logistic regression model, there was an association between respiratory symptoms and traffic-related pollutants. Also, there was spatial variability in traffic pollutants and associated differences in respiratory symptoms in a region with good air quality. The findings support the view that traffic-related pollution is associated with respiratory symptoms in children.

Oyana and Rivers (2005) worked on geographic variations of childhood asthma hospitalization and outpatients visits in relation to the proximity to ambient pollution sources at a US-Canada border crossing. The effects of ambient pollution sources on individuals with asthma was demonstrated and they suggested that these sources are the contributing factors both in the west and east of the study area. Identification of asthma clusters associated with different sources may provide insights into how mixtures of pollutant interact and lead to the development of asthma in susceptible individuals.

Jennifer and Roger (2005) developed a computer model, using Geographical Information System (GIS), to quantify potential health effects of air pollution from a new energy waste facility on the surrounding urban population. The method was a development based on an existing computer spreadsheet model. The model was based on changes in ambient pollution monitoring sites resulting from policies to improve air quality in a local authority, assuming that levels changed in parallel across the whole area. The method links the spreadsheet to a GIS, so as to be able to calculate the health impact for modelled additional exposure experienced by resident population of each enumeration district.

The quantification relies on the simple equation:

$$\Delta E = \beta * \Delta C * P * E \tag{2.1}$$

where Δ E = (change in) background rate of events

β = exposure-response coefficient

Δ C = change in concentration of pollutant

P = population exposed.

The coefficients are derived from epidemiological studies, which show a clear association between increased exposure and increased effects. The anticipated emissions from the proposed incinerator were entered into the ADMS air pollution dispersion model to obtain contours of additional concentrations for the new source.

Using the GIS it was possible to see, at a glance, the population density variation across the study area in conjunction with the distribution of pollutant concentrations. The results from the GIS were then exported into MS Excel so that further calculations could be made.

The model changes in annual mean concentrations of PM_{10} were small, being $0.08\mu g/m^3$ at the most affected location. The average additional exposure over the area within 20km would be $0.002\mu g/m^3$. Even in the most affected location, concentration increases of this magnitude would be impossible to detect through the use of monitoring instruments, given current background concentrations of about $25\mu g/m^3$ and the precision of current instruments.

Combining a GIS system, dispersion model and a spreadsheet the feasibility of potential health impacts of sources of pollution was shown, and the underlying assumptions were examined using sensitivity analyses.

Maynard et al. (2007) considered the Mortality Risk Associated with short-term exposure to Traffic Particles and sulphates using Geographic Information System (GIS) based exposure model. Deterministic covariates such as traffic density and meteorology factors and a smooth function of latitude and longitude were incorporated. They found out that both traffic particles and particles from coal burning power plants were associated with increased mortality in Boston metropolitan area. The traffic particle association was more significant and larger; because Boston currently is in compliance with the current and proposed US Environmental Protection Agency (2006) $PM_{2.5}$ standard. This suggests that the current standards are not protective of public health.

Suglia et al. (2007), in their investigation of the association of Black carbon with cognition among children in a prospective Birth Cohort study, found a consistent relation between exposure to black carbon and reduced neurocognitive functioning across a number of domains in urban community-dwelling school-aged children.

Kandala et al. in their work entitled "Spatial Analysis of Risk Factors for Childhood Morbidity in Nigeria" (2007), investigated the impact of geographical factors and other important risk factors on diarrhea, cough and fever using Geoadditive Bayesian Semiparametric models. A higher prevalence of childhood diarrhea, cough, and fever was observed in the northern and eastern states, while a lower disease prevalence was observed in the western and southern states. Also, children from mothers with higher levels of education and those from poor households had significantly lower associations with diarrhea. Statistical method:

$$\eta_i = x^{'}\beta + w^{'}\gamma \tag{2.2}$$

with a geoadditive predictor, leading to the geoadditive regression model

$$\eta_i = f_1(x_{i1}) + ... + f_p(x_{ip}) + f_{spat}(s_i) + w_i^{'}\gamma \tag{2.3}$$

where $f_1, ..., f_p$ are nonlinear smooth effects of the metrical covariates.

Linden et al. (2007) considered carbonmonoxide in Ouagadougou, Burkina Faso. It is a comparison among urban background, roadside and In-Traffic measurements where they examined the spatial variations of carbonmonoxide (CO) in the urban environment of Ouagadougou, Burkina Faso. The results show significant differences between the three methods where average in-traffic values were 2-3times higher than average background values. During traffic congestions, these differences extended up to 6 and 20 times respectively. Results are discussed in relation to human exposure assessments and WHO guidelines.

Osuntogun and Koku (2007) also worked on the "Environmental Impacts of Urban Road Transportation" in south-western states of Nigeria. Their study was carried out in some locations associated with heavy traffic -eight in Lagos metropolis and four locations each in Ibadan and Ado-Ekiti.

Also, two locations in Lagos, one in Ibadan and one in Ado-Ekiti were selected. Air quality indicators, namely carbonmonoxide (CO), sulphur dioxide (SO_2), nitrogen dioxide (NO_2) and total suspended particulates, were estimated using automatic air monitors. The noise level at these locations was also determined with a noise meter. The blood samples of people at these high trafficked locations (such as commercial drivers, conductors, street traders and road traffic wardens) were also analyzed for lead content with a resultant high Pb concentration. Response to interviews indicate that these people suffer more from air pollution related diseases such as headaches, loss of vision, anaemia, forgetfulness and fatigue than those from the control locations.

Ojolo et al. (2007) investigated the effects of vehicular emissions on human health, vegetations and environments using three locations in Lagos (Oshodi, Mushin and Apapa) and the fourth location (Fola Agoro) as a control.

The investigation was carried out with the use of questionnaires and laboratory experiments. They observed that the people living around are affected by sleeplessness, runny nose, heavy eye, asthmatic attack and headache. The location was used to determine the impacts of vehicular emissions on the ecosystem.

Another study carried out by Oguntoke and Yusuff (2008) on vehicular emissions and the associated human health problems in Abeokuta metropolis revealed that there was significant variation in the volume of traffic and the concentration of the sampled gases between the periods of the day at the selected motorways. Also, there was a significant ($p > 0.05$) correlation between traffic volume/density and CO. Hence, traffic volume accounted for pollutant concentrations in air sampled along the selected motorways and the health problems suffered and reported which include cough and breathing impairment among others.

Sugha, Gryparis, Schwartz and Wright (2008) association between traffic-related black carbon exposure and lung function among urban women and concluded that exposure to traffic-related Black carbon, a component of particulate matter, independently predicted lung function in urban women, when adjusting for tobacco smoke, asthma diagnosis, and socio - economic status.

Abam and Unachukwu (2009), studied the impact of vehicular emission on ambient air quality in selected areas in Calabar. This was done by monitoring the selected areas in Calabar for CO_2, NO_2, CO, SO_2, particulate matter (PM_{10}) and noise level. The ambient temperature, wind directions, wind velocity and traffic count were also monitored.CO, SO_2, NO_2, and PM_{10} were determined using standard method; CO was monitored using a portable analyzer model 2002, SO_2 were collected using APM 410 and 415 sampler. The traffic count was done manually and the ambient temperature, wind direction and wind velocity were monitored with a portable weather station (Davis Ventage Pro-2, USA). The study revealed that transport-related pollution in Calabar is significant with possible severe health consequences. The study further revealed that pollution at traffic intersection was threatening and that motor vehicles remain the dominant sources of urban air pollution.

Olajire et al (2011) evaluated exposure to hazardous air pollutants along Oba Akran Road, Lagos, using multivariate analysis in order to determine the contribution of different sources. It was found that the main principal components extracted from the air pollution data, were related to gasoline combustion, oil combustion and ozone interaction. Also, there were relatively high exposure levels of CO and PM_{10} along Oba Akran Road. PM_{10} and CO along Oba Akran Road were highly traffic-related with possible severe health consequences.

In the paper, "Impact of Traffic Emission on Air Quality Standard in a Developing City in Nigeria", Jimoh and Ndoke (2011) show that the city of Minna is under the threat of traffic pollution. These findings could serve as base-line information for urban development vis-a-vis traffic management policy in Nigeria.

Despite the limitations, a model would appear to be the most practical way of assessing traffic-related exposure where routine surveillance is concerned. Information such as vehicle density, type of vehicle, risk of traffic congestion, presence of bus stops and street crossings, distance of residences to roads, street width, type of street, building presence and type of meteorological conditions (e.g. wind speed and direction, absolute temperature and temperature differences, global and gamma radiation) could be collected routinely for use in a variety of models for predicting exposure to NO_2 and PM_{10}. The model could be used to estimate exposures on all the streets within a certain radius of the home or place of work as dispersion of pollutants from these streets may also be affecting the individual. In a sophisticated model, it may be possible to make adjustment for the height of an individual's residency or place of work in high rise buildings to account for the vertical dispersion of pollutants. Such a system could also be used to estimate exposures at previous residences, work places or schools of an individual so that an assessment of lifelong exposure could be made as accurately and practically as possible. However, the latter might be too complicated for a routine monitoring system.

2.3 Review of Vehicular Emission Model

Various models are developed to estimate emissions. This is an area where substantial amount of research work is being conducted. Vehicular emission factors are critical aspects that are considered in the transportation planning process of freeway facilities.

2.3.1 Statistical Models

The various statistical models developed to estimate vehicular emissions are discussed as follows.

Fuel Consumption and Emission Modeling Considering Power Demand as a Predictor Variable

Abhishek in his (Msc) Thesis explained that his emission model was based on the instantaneous power demand experienced by the vehicle. The data were obtained from dynamometer testing. About 177 in-use Australian vehicles were used to collect the data. Motor vehicles are driven on dynamometer simulating on-road conditions covering a wide range of speed and loads.On-road instantaneous power is derived from vehicles mass, drag, velocity, acceleration and road-gradient. This model can be applied for any traffic situation if the on-road power demand is known. Validation of the model was carried out using an on-road power method i.e., by driving over 2281 links and 956 km recording the on-road velocity,acceleration, and gradient data. The models developed for estimating fuel consumption and emissions considering power demand as a predictor variable performed well for long trips.

$$HC(g/mile) = \alpha + \beta Z_{tot} \; ; \qquad Z_{tot} > 0 \tag{2.4}$$

$$= \alpha \qquad ; \qquad \leq 0 \tag{2.5}$$

HC = hydrocarbon

Z_{tot}= overall instantaneous total power demand in kW

α and β = vehicle parameters (Note: vehicle parameter can vary for each vehicle)

Microscopic Models Developed to Estimate the Fuel Consumption and Emission Rates

The models developed in a Masters Thesis by K. Ahn have two predictor variables: speed and acceleration. Eight light duty vehicles were used to collect the data. The data collected by the Oak Ridge National Laboratory were used to develop these models. The models were developed considering speed and accelerations as predictor variables on a 19 second-by-second basis for individual vehicles. Two types of mathematical models, nonlinear regression models and neural network models were studied as part of this research. To validate the models developed for fuel consumption and emission rates, three methods were adopted: FTP cycle test, US06 cycle test, and Generalization test.

Non-linear regression model

$$log(MOE_e) = \sum_{i=0}^{3}\sum_{j=0}^{3}(K^e_{i,j} \times s^i \times a^j) \tag{2.6}$$

where MOE_e=Fuel Consumption or emission rates (lt/hr or mg/s)

k = model regression coefficients

s = speed (m/s)

a = acceleration (m/s^2)

Neural network model

$$MOE_e = F^3(W^3F^2(W^2F^1(W^1p + b^1) + b^2) + b^3) \tag{2.7}$$

where MOE_e = fuel consumption or emission rates (lt/hr or mg/s)

$W^1; W^2;$ and W^3 = model coefficients

$b^1; b^2;$ and b^3 = bias matrices

p = an input vector containing pairs of (speed; acceleration) used as predictor variables.

F^1= nonlinear transfer function (hyperbolic tangent sigmoid, $F = \frac{1}{1+e^{-n}}$)

F^2 and F^3 = nonlinear transfer function (logarithmic sigmoid; F = $\frac{e^n - e^{-n}}{e^n + e^{-n}}$)

There are some limitations to these models. Start up emissions and ambient temperatures were not considered, which will significantly affect the fuel consumption and emission rates.

Statistical Model Developed for Estimating Nitrogen Oxide Emissions from Light Duty Gasoline Vehicles

This model considered engine load as the major factor, which affects the NOx emission rates. The predictor variables are modal activity variables, which are used to estimate the emission rate. The in-use vehicle emission testing database compiled by United States Environmental Protection Agency (USEPA) was employed in developing the model, which contains 17,417 test results on hot stabilized testing cycles. Furthermore, the data was constrained by limiting the types of vehicles to light duty vehicles (LDV). Therefore, a total of 13,012 vehicle test results, representing 7,151 unique vehicles were tested. This data set contains 114 variables in which 50 variables, were taken for the purpose of analysis. Two types of regression techniques, the Hierarchical tree based regression (HTBR) and the Ordinary least-squares (OLS), were used to develop the model.

This model is a complicated one as the inputs are derived from the combustion mechanism and simulation of the fuel flow characteristics from the intake through the combustion chambers to the exhaust system.

2.3.2 On-Road Emission Measurements

There are three different kinds of on-road emission measurements:

- Federal test procedure
- Remote sensing
- On-Board measurement

2.3.3 Federal Test Procedure (FTP): (FTP Review Project, EPA, May 1993)

The FTP is used to test vehicles for compliance with emission standards. The current test procedure used in the U.S. is referred to as FTP75. The FTP is conducted on a dynamometer for different driving cycles. The FTP is used to measure the concentrations of different pollutants, such as HC, CO, NOx and CO_2. Both the evaporative and exhaust emissions are measured by dynamometer testing under several simulated situations. Evaporative emissions are measured after heating the fuel tank to simulate heating by the sun, i.e. diurnal test, and then the car is driven for some time and parked with the hot engine, i.e. hot soak test. Exhaust emissions were measured by driving the vehicles on a dynamometer for different simulated driving cycles. The vehicle is run on the dynamometer under two conditions. The first condition is cold start, i.e. after a period of non-use, and the second condition is hot start, i.e. while the engine is still hot. The FTP considers factors like ambient temperature, humidity, vehicle speed, fuel consumption, aerodynamic loss and vehicle inertia. Although the dynamometer is a reliable method for emission estimation, the drawback is that the dynamometer testing method may not simulate real world driving conditions and it may not consider short term events that will cause high emissions.

2.3.4 Remote Sensing

The Remote Sensing Device (RSD) was developed in the late 1980s at the University of Denver (US Remote Sensing Experience, Niranjan Vescio CITA conference, 2002).The RSD collects data, like speed and acceleration, captures license plate, and emission measurement of pollutants, like CO, NOx, and HC. The RSD operates by continuously projecting two beams across the roadway . One is non-dispersive Infrared Spectroscopy, which is used to measure the concentrations of HC and CO and the other beam is Dispersive Ultraviolet Spectroscopy, which is used to measure the NOx emissions. As the vehicle passes through the beam, the emissions are calculated. The main advantage of a remote sensing device is that it identifies the high emitting vehicles and can measure a large number of on-road vehicles. The major disadvantage of using the remote sensing device is that it will not measure evaporative emissions. It gives the instantaneous estimate of emissions at a specific location, and it is not suitable for bad weather conditions.

2.3.5 On-Board Measurement

The on-board measurement system is used to measure the exhaust emissions from vehicles under real-world travel conditions. This methodology has advantages over both dynamometer and remote sensing methods. The dynamometer testing method does not measure the emissions for real-world conditions and the remote sensing device method measures the emissions at a particular location, whereas the on-board measurement measures the emission rates under real world conditions and for all driving conditions. In on-board emission measurements, there are many factors considered while measuring the emissions. Such factors include, like speed, different driving modes (idle, acceleration, deceleration and cruising), ambient temperature, humidity, and different traffic conditions. In Ahn's masters thesis, on-board data is used in developing a model for carbon dioxide emission rate. An analysis of the difference in emission rates while considering the accelerations and deceleration versus constant speed was also performed. For the analysis, the data from on-board and Mobile 6.2 were used. Mobile 6.2 does not consider variable accelerations and deceleration while estimating the emission rates.

2.4 Modelling Software

There are many software tools for estimating the vehicular emission rates of different pollutants. The most popular tools used in the U.S. to estimate emission rates are Mobile 6.2 and Emission Factor Model (EMFAC).

2.4.1 EMFAC

EMFAC is the emission factor model used to calculate the emission inventories of on-road vehicles in California. EMFAC is a model in which the emission rate data and activity data are combined to calculate the emission inventory. The emissions for the following pollutants are calculated: CO, NOx, HC, CO_2, lead, PM, and oxides of sulfur. Both exhaust and evaporative emissions are calculated for 13 different classes of vehicles. The model can estimate the emission rates for any calendar year between 1970 and 2040.

2.4.2 Mobile

A brief history of the Mobile source emission factor model is as follows :
Mobile 1: The first Mobile model was developed in 1978 to estimate the highway vehicle emission factors.
Mobile 2: In 1981, the model was updated with the new in-use data. The new data of emission controlled vehicles for higher ages and mileages was added to the model.
Mobile 3: In 1984, the model was updated with the new in-use data. In this updated version of Mobile, anti-tampering programme benefits were added to the model and eliminated the California vehicle emission rates.
Mobile 4: In 1989, the model was updated with the new in-use data. In Mobile 4, evaporative running losses were added for gasoline powered vehicles and modeled fuel volatility (RVP) effects on exhaust emission rates.
Mobile 4.1: In 1991, the model was updated with the new in-use data. In this updated version of Mobile, the impact of oxygenated fuels on CO was included, added many features, which allow the user to control more parameters that affect the emission levels, and included more inspection and maintenance (I/M) programme designs.
Mobile 5 and Mobile 5a: In 1993, the Mobile 5 model was updated with the new in-use data. In this updated version, the effects of reformulated gasoline and the impact of oxygenated fuels on HC emissions were added. Later, after four months, Mobile 5a was issued. Many errors, which were detected under

specific conditions, were corrected in this updated version.

Mobile 5b: In 1996, the model was updated with the inclusion of the impacts of onboard refueling vapor recovery system, reformulated gasoline requirements, and expanded calendar year range from 2020 to 2050 for which the emission rates can be estimated.

Mobile 6: In 2002, the model was updated by including the effects of air conditioning and high acceleration driving and expanding the classes of vehicles from eight to twenty eight.

Mobile 6.2: In 2004, the model was updated by adding the ability to estimate the emission factors for particulate matter and six air toxins.Each version of the Mobile model becomes more sophisticated in estimating the emission factor for different pollutants and different classes of vehicles. The new version of Mobile (Moblie 6.2) provides users more advanced options to modify the emission factor estimates according to specific times and geographic locations.

Mobile estimates the emission factor for different pollutants, like HC, CO and NOx for different classes of vehicles. The Mobile model was written in FORTRAN. FORTRAN is a computer programming language that is suitable for numeric computations and scientific computations. The Mobile model estimates emission factors for both exhaust emissions (tailpipe) and evaporative emissions. In estimating the emissions factors, the model considers various factors, including vehicle population, vehicle activity, and meteorological factors (temperature, humidity, and type of fuel). The interface of this modelling software is DOS.

Mobile 6.2 estimates the CO_2 emissions in a very simplistic way. The CO_2 calculations are based on the fuel economy performance estimates built into the model or supplied by the user. For other pollutants Mobile 6.2 considers various factors such as vehicle activity, speed, and meteorological data to estimate emission rates. But for the CO_2 pollutant Mobile 6.2 does not adjust to the speed, temperature, and other factors.

Some of the other modeling software tools which calculate the emission inventories at micro level include MEASURE, FRESIM, and TRANSIMS etc. These software tools are discussed below:

MEASURE: MEASURE, is built in a Geographic Information System (GIS) framework and is able to estimate emissions for specific vehicle and engine operating modes(acceleration, deceleration and idling etc.). In developing and validating the MEASURE modeling software, the EPA used vehicle activity and emission data collected from different techniques which include remote sensing devices, automobiles and trucks equipped with on-board instrumentation. The MEASURE model estimates both spatially and temporally vehicle activities that result in emissions.

FRESIM: FRESIM is a traffic simulation model used for freeway analysis. At the micro level of detail, traffic-simulation models can be combined with modal or instantaneous emissions models to predict emission inventories. Second by second vehicle trajectory data is generated and used as input to modal emission model. The resulting emission data from all the vehicles are then integrated to provide a total emission inventory.

The advantage of the micro level models is that they are best in estimating changes in emissions resulting from strategies that affect traffic flow and can account for the effects of the variance of driver behavior on emissions. The limitation of the microscale level models is: vehicle trajectory data which includes velocity-acceleration lookup tables may not be available, or may have old data, due to which emissions may not be calculated accurately.

2.5 Review of geoadditive models for other areas of application

Geoadditive model, which combines the idea of geostatistics and additive models, have been shown, over the years and by various researchers, to be very useful in some other areas. Kamman and Wand (2003) have shown that linear mixed models could be used for Geoadditive model fitting and inference. However, several other scholars, such as Wood (2006), have treated the same structure in other ways.

Extension of geoadditive models in the direction of generalized responses are contained in Fahrmeir and Echavarria (2006) and Zhao et al. (2006). Zhao et al. (2006) deal with exponential family models; whilst Fahrmeir and Echavarvia (2006) treat over-dispersed and zero-inflated count data. Each used a Bayesian mixed model framework, with fitting via MCMC, and provide applications.

The extension of Geoadditive models to survival data has seen considerably researched since 2003. Hennerfeind et al. (2006) developed geoadditive survival models for both geographical point data and count data. They take a Bayesian P-spline approach and use Gaussian and Markov random fields for the spatial components. Kneib & Fahrmeir (2004) lay out the mathematics under- pinning geoadditive hazard regression models. Kneib (2006) extends these models to handle interval censored data. Adebayo & Fahrmeir (2005) developed a geoaddditive discrete-time survival model and used it to analyze child mortality data. Ganguli & Wand (2006) also deal with georeferenced survival data, and use the low rank radial smothers of Kamman & wand (2003).Geoadditive models have also been adapted to model space-time data. Fahrmeir et al. (2004) and Kneib & Fahrmeir (2005) used low dimensional smooths, involving time and age, to model forest data, in conjunction with Gaussian and Markov random fields for the spatial effects. Gryparis et al. (2007) also involved space-time data, but their Geoadditive model is an elaborate one that includes latent variable structure for multiple exposures from mobile particulate matter.

Geoadditive models with missing data covariate are studied by French & Wand (2004). Chen & Ibrahim (2006) extended that work to Geoadditive models that allow for specification of the covariate distribution and the missing data mechanism.

Kamman & Wand (2003) used Geoadditive models to study the geographical variability of reproductive health outcomes (e.g birth weight) in upper Cape Cod, Massachusetts, USA. The study showed that geoadditive model is an effective vehicle for the analysis of spatial epidemiologic data and other applications where geographic point data are accompanied by covariate measurements. The low rank mixed model formulation allows a straight forward implementation and fast processing of large databases, thus facilitating the use of the model in the surveillance of disease clusters.

The Geoadditive model has been shown to be useful for an analysis of the upper Cape Cod Reproductive data. It properly accounts for all covariate information before producing disease maps. In the case of gestational age, it has been seen that no residual geographical effect is present. The birth weight analysis is slightly suggestive, but geographical variation cannot yet be concluded.

Nkurunziza et al. (2011) used Geoadditive model for modeling malaria in Burundi. The data analysis was carried out using monthly data; semiparametric model was used to model the effects of both climatic covariates and spatial effects on malaria distribution in Burundi.

Sarah and Wand (2011) developed a new method for generalized extreme value geoadditive model analysis via Variational Bayes which was fast for approximate inference in Bayesian Generalized Extreme value additive model analysis. Such models are useful for flexibly assessing the impact of continuous predictor variables on sample extreme. The new methodology allows large Bayesian models to be fitted and assessed without the significant computing cost of Monte Carlo methods.

Wand et al. (2011), in their work entitled Geoadditive Models to Assess Spatial Variation of HIV Infections among Women in Local Communities of Durban, South Africa" used geoadditive model to assess nonlinear geographical variation in HIV prevalence while simultaneously controlling for important demographic and sexual risk factors. A total of 3469 women who were screened for a phase III randomized trial were included in the current analysis. The study revealed significant geographic variability in HIV infection in the Ethekwini metropolitan municipal in KwaZulu-Natal, South Africa.

Adebayo et al. (2013) applied Geoadditive model to modelling geographical variations and determinants of use of modern family planning methods among women of reproductive age in Nigeria using reference coding. The study revealed considerable geographical variation in the use of modern family planning. Variation was evident with an increase between 2003 and 2005 followed by a decline between 2005 and 2007. The effect of respondent's age was non-linear, and use of modern family planning was found to differ significantly between never-married and currently/formerly married respondents.

2.6 Summary of literature review

Literature on various environmental models used in estimating environmental data,statistical models, measurement method and modelling software used in estimating the emission rates as well as geoadditive models for estimating other areas aside the combination of health effects and environmental has been reviewed. The next chapter describes the theoretical framework for the research specifying the prior and the posterior.

Chapter 3

MATERIALS AND METHODS

3.1 Introduction

This chapter discusses the framework for the model by looking at the priors and the posteriors which are utilized in developing the Geoadditive model. In particular, we apply a Geoadditive Bayesian model in order to estimate the effect of vehicular emission on individuals (18 years and below) in four local government areas in Ogun state. Six different hospitals in the state were observed for 3 years.

This chapter is divided into sections; the first section highlights the various structural additive models, while the second section discusses the general additive model and the third section discusses the geoadditive Bayesian model. The fourth section describes the prior and the properties of the posterior. Sixth section discusses the programme developed for the analysis. The last section describes the data sources and description of the data.

3.2 THEORETICAL FRAMEWORK

3.2.1 Structured Additive Regression Models

The Structured Additive Regression Models(STAR) is based on the framework of Bayesian Generalized Linear models (GLMs McCullagh and Nelder 1989 and Fahrmeir and Tutz, 2001). Generalized Linear Models assume that, given covariates X and unknown parameters γ,the distribution of the response variable y belongs to an exponential family with mean.$\mu = E(y/x, \gamma)$ linked to a linear predictor η by

$$\mu = h^{-1}(\eta), \eta = x^T\gamma \tag{3.1}$$

where h is a known link function and γ are unknown regression coefficients. In STAR models (Fahrmeir et al. 2004; Brezger and Lang 2006), the linear predictor is replaced by a more general and flexible, structured additive predictor.

$$\eta = f_1(1) + - - - - - + f_p(z) + x^T\gamma \tag{3.2}$$

with

$$\mu = E(y/x, z, \gamma, \theta) \tag{3.3}$$

and z represents a generic vector of all nonlinear modelled covariates. The vector θ comprises all parameters of the functions $f_1, - - -, f_p$.The functions f_j are possibly smooth functions encompassing various types of effects.e.g

- Nonlinear effects of continuous covariates:$f_j(z) = f(z_1)$
- Two-dimensional surfaces: $f_j(z) = f(z_1, z_2)$
- Spatially correlated effects:$f_j(z) = f_{spat}(z_s)$
- Varying coefficients: $f_j = z_1 f(z_2)$
- Spatially varying effects: $f_j(z) = z_1 f_{spat}(z_s) or f_j(z) = z_1 f(z_2, z_3)$
- Random intercepts with cluster index c: $f_j(z) = \beta_c$
- Random slopes with cluster index c:$f_j(z) = z_1\beta_c$

Structured additive regression models which cover a number of well known model classes as special cases include:

- Generalized Additive Model (Hastie & Tibshrani, 1990)
- Generalized Additive Mixed Model (Lin & Zhang, 1999)
- Varying Coefficient Models (Hastie & Tibshrani, 1993)
- Geographical Weighted Regression (Fothering, Brunsdon & Charlton, 2002)

- Geoadditive Model (Kamman & Wand, 2003)

The unified representation of a STAR predictor arises from the fact that all functions f_j may be specified by a basis function approach, where the vector of function evaluations $f_j = (f_j(z_1), ---, f_j(z_n))^T$ of the i =1,- - -,n observations can be written in matrix notation.

$$f_j = z_j \beta_j \tag{3.5}$$

where the design matrix z_j depends on the specific term structure chosen for f_j and β_j are unknown regression coefficients to be estimated. Hence, the predictor may be rewritten as

$$\eta = Z_1\beta_1 + ----- + Z_p\beta_p + X\gamma \tag{3.6}$$

where X corresponds to the usual design matrix for the linear effect.

3.2.2 General Additive Models For Nonlinear Regression Effects

A generalization of the smoothing prior is to generalize additive model in regression; such models provide an approach to modelling possible nonlinearity but avoiding the need to specify complex algebraic forms. Thus, for a metric outcome $y_1, ..., y_n$, assume there are corresponding values of a regressor variate $x_1, ..., x_n$ ordered such that

$$x_1 < x_2 < ... < x_n \tag{3.7}$$

The model for the observations may then be

$$Y_t = \beta_0 + f(x_t) + \varepsilon_t \tag{3.8}$$

where $\varepsilon_t \sim N(0, \sigma^2)$
Let $g_t = f(x_t)$ be the smooth function representing changing, possibly nonlinear, impact of x on y as it varies over its range.
It is common to assume Normal or random walks in the first, second or higher differences of the g_t.

A variant on this is when the smooth in the variable x modifies the effect of a predictor Z, with

$$Y_t = \beta_0 + Z_t\beta_{1t} + \varepsilon_t \tag{3.9}$$

It will commonly be the case that the x_t are unequally spaced, and it is then necessary in specifying the prior for $g_t(or\beta_{1t})$ to weight each preceding point differently. This means adjusting the precision such that wider spaced points are less tied to their predecessor than closer spaced points. Thus, suppose the x_t were irregularly spaced and that the spaces between points $\delta_1 = x_2 - x_1, \delta_2 = x_3 - x_2, ..., \delta_{n-1} = x_n - x_{n-1}$.

Fahrmeir and Lang (2001) shows that a first order random walk smoothness prior with Normal errors, would then be specified as

$$g_t \sim N(g_{t-1}, \delta_t\tau^2) \tag{3.10}$$

and a second order would be

$$g_t \sim .N(V_t, \delta_t\tau^2) \tag{3.11}$$

where

$$V_t = g_{t-1}(1 + \frac{\delta_t}{\delta_{t-1}}) - g_{t-2}(\frac{\delta t}{\delta_{t-1}}) \tag{3.12}$$

If there is equal spacing then the first and second order random walk priors are

$$g_t \sim N(g_{t-1}, \tau^2) \tag{3.12}$$

$$g_t \sim N(2g_{t-1} - g_{t-2}, \tau^2) \tag{3.13}$$

3.3 Geoadditive Bayesian Model

There has been much recent work on Bayesian Regression methodology that is sufficiently flexible to accommodate nonlinear and nonadditive relationship between the response variable and the predictors. Constructs used for such modelling include neural networks (Neal, 1996), Gaussian processes (Neal, 1999) and regression trees (Chipman et al 1998; Denison et al. 1998a). Spline-based methods have been investigated by (Denison et al. 1998b,1998c), (Shively et al ,1999), and (Smith and Kohn 1996,1997). These methods tend to involve a compromise; some of the computational and interpretive simplicity of a linear model is sacrificed to obtain a flexible regression function.

There is the need for a method that will retain such flexibility and not sacrifice the computational and interpretive simplicity of a linear model. A Geoadditive model in its real sense allows for spatial and temporal correlations as well as nonlinear effects of covariates and unobserved heterogeneity.

In this research a Geoadditive Bayesian model will be considered which caters for all the shortcomings of most of the Regression analysis as well as retain its flexibility to accommodate nonlinear and nonadditive relationship between the response variable and the predictors.

3.4 Specification of Prior

In the Bayesian framework, the unknown smooth function f_j, parameter β, and the variance parameter σ^2 are all considered as random, and therefore, have to be assigned suitable priors. The usual approach is to assign diffuse priors (uninformative prior) to the parameters of the fixed effects, that is β_j $\propto$ constant j = 1. . .p.

The unknown smooth functions f_j, j = 1- - - p, are estimated using the Bayesian p-spline basis approach (Lang & Brezger 2004). In this approach, it is assumed that the unknown functions can be approximated by a polynomial spline of degree l defined by a set of equally spaced knots.

$$\xi_0 = x_{min} < \xi_1 < ...\xi_{k-1} < \xi_k = x_{max} \tag{3.14}$$

(Omitting the subscript j for convenience) over the domain of X. The spline can be expressed as a linear combination of T =K+1, B-spline basis functions, that is

$$f(x) = \sum_{t=1}^{p} \beta_t B_t(x) \tag{3.15}$$

where B_t is the t-th basis function.

Now, let X be the $n \times p$ design matrix with (i,t)th element given by X (i,t) = $B_m(X_i)$

Then the Geoadditive model can be expressed in matrix notations as

$$\eta = X_1\beta_1 + X_2\beta_2 + ... + X_p\beta_p + V\gamma \tag{3.16}$$

Here β_j (j = 1. . . p) are the unknown regression coefficients, whereas the matrix V correspond to the design matrix for the linear effects. Eiler and Marx (1996), in their frequentist setting, first and second order difference in order to overcome difficulties involved with regression splines such as non-flexibility (for smaller number of knots) or over-fitting (for large number of knots). Lang and Brezger (2004) in their Bayesian setting proposed replacing differences with stochastic analogues of first and second order random walks given by:

$$\beta_t = \beta_{t-1} + U_t \tag{3.17}$$

(first order random walk) or

$$\beta_t = 2\beta_{t-1} - \beta_{t-2} + U_t \tag{3.18}$$

(second order random walk) with Gaussian errors Ut $\sim N(0, \tau^2)$ and diffuse (uninformative) priors $\beta_1 \propto$ constant or $\beta_1\beta_2 \propto$ constant (for initial values). The variance parameter τ^2 controls the amount of smoothness and is also referred to as the inverse smoothing parameter. The amount of smoothness is estimated by defining a hyperprior for the variance parameter τ^2. A usual approach is to assign a conjugate prior for τ^2 which is the inverse Gamma prior with hyperparameters a and b, $\tau^2 \sim IG(a, b)$. Common choices for a and b are a =1 and b = 0.005 (or b = 0.0005). Alternatively, one may take a = b = 0.001.

Brezger and Lang (2006) also suggest a general structure of the prior as

$$\beta_j/\tau_j^2 \propto \frac{1}{(\tau_j^2)^{(rk)(K_j)/2}} \exp\left(\frac{-1}{2\tau_j^2}\beta_j^{'}k_j\beta_j\right) \tag{3.19}$$

where K_j is a penalty matrix which depends on the prior assumptions regarding smoothness of f_j and the type of covariate.

3.5 Proprieties Of The Posterior In Mixed Model

3.5.1 Lemma 1: (Hennerfeind et al, 2006)

Consider the Gaussian mixed model defined as :

$$Y = V\gamma + Z_1\beta_1 + ... + Z_p\beta_p + \varepsilon \tag{3.20}$$

For observations $y = (y_1, ..., y_n)\prime$, with an additive predictor, and a Gaussian error vector

$$\varepsilon = (\varepsilon_1, ..., \varepsilon_n) \sim N(0, \tau_0^2 I)$$

The prior assumptions for the parameters γ and β_j, j = 1,. . .,p are flat priors.

$$P(\gamma) \equiv 1 \tag{3.21}$$

For the vector γ of fixed effects

$$P(\beta_j) \propto \tau_j{}^{(-rj)} exp\left(\frac{-1}{2\tau_j^2}\beta_j' K_j \beta_j\right) \tag{3.22}$$

(1) rank (V) = P, rank $(Z'RZ + K) = d$,

where P = dim (γ) ,

$$d = d_1 + ... + d_m = dim(\beta),$$

$$K = \text{ diag } (K_1, ..., K_m), \; R = IV(V'V) - 1V'$$

(2) the priors $P(\tau^2), j = 1, ..., p$, are proper,

and

$$\int P(\tau_j^2)\tau_0^{(-(n-p-(d-r)))} exp\left(\frac{-SSE}{2\tau_0^2}\right) \, d\tau_0^2 < \infty, \text{ where } \; r = r_1 + ... + r_m \tag{3.23}$$

Then the posterior distribution $P(\gamma, \beta, \tau^2/y)$ is proper.

With $d_j = dim(\beta_j)$ and $r_j = rank(k_j)$. For $r_j < d_j$ the prior for β_j is partially improper and assume the following conditions hold

3.5.2 Corollary (1): (Hennerfeind et al, 2006)

For a linear mixed model (3.20) with prior (3.22) and

$$P(\tau_j^2) \propto \frac{1}{(\tau_j^2)^{a_j+1}} exp\left(\frac{-b_j}{\tau_j^2}\right) \tag{3.24}$$

which are proper for $a_j > 0$, $b_j > 0$.

The posterior $P(\gamma, \beta, \tau^2/y)$ is proper if condition (i) of Lemma (1) and $a_j > 0, b_j > 0$, j = 1,. . ., p, $n - p - (d - r) + 2a_0 > 0$, $SSE + 2b_0 > 0$ hold.

3.6 Properties of the Posterior for Generalized Geo-additive Models

3.6.1 LEMMA(2): (Hennerfeind et al, 2006)

Consider a generalized linear mixed model with observation densities $f_i(y_i/\eta_i)$ predictor

$$\eta = V\gamma + Z_1\beta_1 + ... + Z_m\beta_1 + Z_0\beta_0 \tag{3.25}$$

And priors assumptions for the parameters γ and β_j, j = 1,..., p are the same .i.e a flat prior $P(\gamma) \equiv 1$ for the vector γ of fixed effects and

$$P(\beta_j) \propto \tau_j^{-r_j} exp\left(\frac{-1}{2\tau_j^2}\beta_j' K_j \beta_j\right) \tag{3.26}$$

With $d_j = dim(\beta_j)$ and r_j = rank(K_j). For $r_j < d_j$ the prior for β_j is partially improper. Priors for hyperparameters $\tau^2 = (\tau_0^2, ..., \tau_m^2)'$ are $P(\tau^2) = \prod_{(j=0)}^m P(\tau_j^2)$.

An important special case is inverse Gamma priors.

$$P(\tau_j^2) \propto \frac{1}{(\tau_j^2)^{(a_j+1)}} exp\left(\frac{-b_j}{\tau_j^2}\right) \tag{3.27}$$

Which are proper for $a_j > 0, b_j > 0$ and a possibly partially improper prior

$$P(\beta_0) \propto \tau_0^{-r_0} exp\left(\frac{-1}{2\tau_0^2}\beta_0' K_0 \beta_0\right) \tag{3.28}$$

with $r_0 = rank(K_0)$, such that

$$d_0 \geq d_j, r_0 \geq r_j, j = 1, ..., m$$

setting $Z_0 = I, \beta_0 = \varepsilon \sim N(0, \tau_0^2 I)$.

In geoadditive models $Z_0\beta_0$ will usually represent a spatial effect with a MRF or kriging prior, or an unstructured spatial effect.

Suppose that:

$$(i) \qquad \int f_i(y_i/\eta_i) d\eta_i < \infty$$

holds for observations $i = 1, ..., n^*$ and

$$(ii) \quad f_i(y_i/\eta_i) \leq M, i = n^* + 1, n$$

holds for the remaining observations.

Denote the corresponding sub matrices of V, Z and Z_0 by V^*, $Z^* = (Z_1^*, ..., Z_p^*), Z_0^*$, and assume:

$$(iii) \quad rank(Z_0^*) = d_0$$

the rank conditions (i) in Lemma (1) hold forV^*, Z^*

condition (ii) in Lemma (1) holds with r_0 replacing n and SSE replaced by SSE^* then the posterior $P(\gamma, \beta_0, \beta_1, ..., \beta_m, \tau_0^2, ..., \tau_m^2/y)$ is proper. The following corollary is easier to check.

3.6.2 Corollary (2): (Hennerfeind et al, 2006)

Assume that conditions (i), (ii) and the rank conditions for V^*, Z^*, Z_0^* in Lemma 2 hold, and that $r_0 - p - (d - r) > 0$

with $d = d_0 + ... + d_p$, $r = r_0 + ... + r_p$, and $a_j > 0, b_j > 0$, j = 0, . . . ,p hold for the inverse Gamma priors (3.28) then the posterior $P(\gamma, \beta_0, \beta_1, ..., \beta_p, \tau_0^2, ..., \tau_p^2/y)$ is proper

Proof

We consider first the simpler case of individual-specific random effects $\beta_0 \equiv \varepsilon \sim N(0, \tau_0^2 I)$ using the one -to - one relation.

$\eta = V\gamma + Z\beta + \varepsilon$ between η and ε, we consider propriety of $P(\eta, \gamma, \beta, \tau_0^2, \tau^2/y)$ instead of $P(\varepsilon, \gamma, \beta, \tau_0^2, \tau^2/y)$. Proceeding as in sun et al (1998), one starts from

$$P("\eta", \gamma, \beta, \tau_0^2, \tau^2/y) \propto P(y/\eta)P(\eta/\gamma, \beta)P(\beta)P(\tau_0^2)P(\tau^2) \tag{3.29}$$

Using (ii) and $\eta^{**} = (\eta_{n^*+1}, ..., \eta_n)$ one arrives at

$$P(\eta^*, \gamma, \beta, \tau_0^2, \tau^2/y) \propto \prod_{i=1}^{n^*} f_i(y_i/\eta_i)P(\eta^*/\gamma, \beta)P(\beta)P(\tau_0^2)P(\tau^2) \tag{3.30}$$

$$\propto \prod_{i=1}^{n^*} f_i(y_i/\eta_i)P(\gamma, \beta, \tau_0^2, \tau^2/\eta^*) \tag{3.31}$$

Applying Lemma (1) or Corollary (1) to

$$\eta^* = V^*\gamma + Z^*\beta + \varepsilon^*\varepsilon^* \sim N(0, \tau_0^2 I) \tag{3.32}$$

gives

$$P(\eta^*/y)\prod_{i=1}^{n^*} f_i(y_i/\eta_i) \tag{3.33}$$

and propriety follows from (i) for the general case

$$\eta = V\gamma + Z\beta + Z_0\beta_0 \tag{3.34}$$

with prior (ii) for β_0, we first decompose β_0 into a $(d_0 r_0)$ dimensional subvector. β_0^{fl} with flat prior $P(\beta_0^{fl}) \equiv 1$ and a r_0 - dimension subvector β_0^{pr} with a proper prior

$$\beta_0^{pr} \sim N(0, \tau_0^2 I)$$

$$\beta_0 = Z_0^{fl}\beta_0^{fl} + Z_0^{pr}\beta_0^{pr} \tag{3.35}$$

Where the $d_0 \times (d_0 - r_0)$ matrix Z_0^{fl} contains a basis of the nullspace of K_0. The matrix Z_0^{fl} is the identity vector 1 for P spline with first order random walk prior. Markov-random fields and 2d-P-splines with MRF prior for the coefficients. For P-splines with second-order random walk prior. It is a two column matrix whose first column is the identity vector and the second column is composed of the (equidistant) knots of the spline. Where the $d_0 \times (d_0 - r_0)$ matrix Z_0^{fl} contains a basis of the nullspace of K_0. The matrix Z_0^{fl} is the identity vector 1 for P-spline with first-order random walk prior. Markov-random fields and 2d-P-splines with MRF prior for the coefficients. For P-splines with second-order random walk prior. It is a two column matrix, the first column of which is the identity vector and the second column is composed of the (equidistant knots of the spline. The $d_0 \times r_0$ matrix Z_0^{pr} is given by

$$Z_0^{pr} = L(L'L)^{-1}$$

Where $L = S'\Lambda^{1/2}$ is obtained from the spectral decomposition $K_0 = S\Lambda S' of K_0$. It follows that

$$\beta_0^{pr} \sim N(0, \tau_0^2 I)$$

Defining $\bar{V} = (V, Z_0, Z_0^{fl})$, $\bar{\gamma}' = (\gamma, \beta_0^{fl})'$, $\bar{Z}_0 = Z_0 Z_0^{pr}$, we can rewrite the predictor as

$$\eta = \bar{V}\bar{\gamma} + Z\beta + Z_0\beta_0^{pr} \tag{3.36}$$

For identifiability reasons, the columns of $Z_0 Z_0^{fl}$ are not contained in the $(d_0 - r_0)$ column space of V, so that rank $\bar{(}V) = P + (d_0 - r_0)$.
Defining $\varepsilon_0 = Z_0 \beta_0^{pr}$, we have an additive mixed model.

$$\eta = \bar{V}\bar{\gamma} + Z\beta + \varepsilon_0 \tag{3.37}$$

3.7 BayesX Software

BayesX is a software tool for estimating structured additive regression models. Structured additive regression embraces several well-known regression models such as Generalized Additive Models (GAM), Generalized Adddditive Mixed Models (GAMM), Generalized Geoadditive Mixed Models (GGAMM), Dynamic Models, Varying Coefficient Models (VCM) and Geographical Weighted Regression within a unifying framework. Besides exponential family regression, BayesX also supports non-standard regression situations such as regression for categorical responses, hazard regression for continuous survival times, and continuous time multi-state models. Estimation of regression models can be achieved based on two different inferential concepts: Markov Chain Monte Carlo Simulation techniques corresponding to full Bayesian inference and mixed model methodology corresponding to penalised likelihood or empirical Bayes inference. There are currently three regression tools implemented in BayesX :

(1) MCMC Simulation techniques (Bayesreg Objects) : A fully Bayesian interpretation of structural additive regression models is obtained by specifying prior distributions for all unknown parameters. Estimation can be facilitated using Markov Chain Monte Carlo Simulation Techniques, a general and versatile concept for Bayesian inference. Bayesreg objects provide numerically efficient implementations of MCMC schemes for structured additive regression models.

(2) Mixed Model Based Estimation (Remlreg Objects) : An increasingly popular way to estimate semiparametric regression models is the representation of penalization approaches as mixed models. Within BayesX, this concept has been extended to structured additive regression models and several types of non-standard regression situations. The general idea is to take advantage of the close connection between penalty concepts and correspond-

ing random effects distributions. The smoothing parameters of the penalties then transform to variance components in the random effects (mixed) model. While the selection of smoothing parameters has been a difficult task for a long time, several estimation procedures for variance components in mixed models with marginal likelihood as the non-Gaussian counterpart have been particularly successful.

Remlreg objects employ mixed model methodology for the estimation of structured additive regression models. While regression coefficients are estimated based on penalized likelihood, restricted maximum likelihood or marginal likelihood estimation forms the basis for the determination of smoothing parameters. From a Bayesian perspective, this yields empirical Bayes/ posterior mode estimates for the structured additive regression models. However, estimates can also merely be interpreted as penalized likelihood estimates from a frequentist perspective.

(3) Penalized least squares including model selection (Stepwisereg Objects): BayesX provides a penalized least squares (respectively penalized likelihood) approach for estimating structured additive regression tools. In addition to the previously described estimation alternatives a powerful variable and model selection tool is included. Model choice and estimation of the parameters is done simultaneously. The algorithms are able to decide whether a particular covariate enters the model, decide whether a continuous covariate enters the model linearly or nonlinearly, decides whether a spatial effect enters the mode, decides whether a unit - or cluster specific heterogeneity effect enters the model, selects complex interaction effects (two dimensional surfaces, varying coefficient terms), and selects the degree of smoothness of nonlinear covariate, spatial or cluster specific heterogeneity effects.

Inference is based on penalized likelihood in combination with fast algorithms for selecting relevant covariates and model terms.

For this research the MCMC simulation techniques implemented is based on full Bayesian interpretation. The prior distribution of the parameter is specified for all unknown parameters and effect coding is used to capture the spatial effect bringing about the Geoadditive Bayesian Model.

3.8 Data Sources and Description

The study was carried out in Ogun state, one of the 36 states of the Federal Republic of Nigeria. It is located in the south-western zone of the country. It occupies a total land area of 16409.26 sq.km. The estimated population of the state by 1991 census was put at 3,214,161. The state is composed of 20 Local Government Areas. It is bounded in the North by Oyo and Osun, in the south by Lagos and in the west by Republic of Benin. The state capital, Abeokuta, is 100 km North of Lagos, the commercial nerve centre of the country. The state is located in latitude 6^o and $7^o40'N$ of the equator and longitude $2^o20'E$ and $4^o35'$East of the Greenwich meridian. The samples were taken from four local government areas of the state, namely Abeokuta, Sagamu, Ijebu-Ode and Sango-Ota. They are located in latitude $7^005'N$ and $7^o28'N$ of the Equator and longitude $3^o10'E$ and $3^o28'E$ of the meridian, latitude $6^o40'N$ and $6^o53'N$ of the equator and longitude $3^o29'E$ and $3^o50'E$ of the Greenwich meridian, $6^o40'N$ and $6^o50'N$ of the equator and longitude $3^o59'E$ and $41^o10'E$ of the Greenwich meridian, $6^o30'N$ and $6^o42'E$ of the equator and longitude $2^o57'E$ and $3^o20'E$ of the Greenwich meridian respectively. Six hospitals (one from each of the six LGAs) were selected for the study. The hospitals include Federal Medical Centre Abeokuta, Sacred Heart Hospital Lantoro, General Hospital Abeokuta, General Hospital Sango Ota, Olabisi Onabanjo University Teaching Hospital and General Hospital, Ijebu-Ode.

The study population include individuals of 18 years and below. A total number of 9211 respondents (individuals) were sampled; they were investigated for their health status. 1800 respondents were selected from Federal Medical Centre, 1500 from Sacred Heart Hospital, 1500 from State Hospital Abeokuta, 1500 from General Hospital Sango- Ota, 1800 from Olabisi Onabanjo University Teaching Hospital and 1111 from General Hospital, Ijebu-Ode. The selection was conducted over a period of 3 years (2008-2011).

These hospitals were selected with consideration for their proximity to the location of vehicular emission collection points. Ethical approvals were sought and obtained from the management of each hospital.

Three personnel were engaged from the health record of each hospital for the purpose of collecting information from patients' records. The personnel were trained on how to extract information about patients diagnosis, number of times the patients visited the hospital in a week and the patients outcome (dead or alive). The proximity of patients' residential address to areas where vehicular emission were collected was considered in the recruitment of the respondents.

The effects of vehicular emmision on health of individuals was monitored in selected areas of Ogun state. The areas were Abeokuta, Ijebu-Ode, Sagamu and Sango-Ota. Each of these areas has different sample points of which CO, CO_2, HC were monitored. The four areas have sample points selected for collection in the priority of high population, traffic congestion and proximity to the selected hospitals.

Five people (research assistants) were engaged and trained for the purpose of collecting data from moving vehicles .i.e both commercial and private vehicles in the second half of the observation collection. The data was collected for 5 days in a week and this went on for the study period of 3 years. An average of twelve vehicles were seen per day.

The equipment used for the collection of the emitted pollutants from the vehicle was called Kane Gas Analyzer. Kane Gas Analyzer is used to measure both the efficiency of combustion and the levels of pollutant gases. It is suitable for all appliances and burning natural gas i.e hydrocarbons. It accurately checks CO levels, measure O_2, CO/CO_2 ratio and efficiency. The result can be printed using the optional infrared printer. It works with multi-fuel, gases and oils. It has a detachable fuel probe with a low cost of ownership.

The analyser is attached to the smoke centre of the vehicle to collect the pollutants. In this study a total of 9211 vehicles were tested for their emission status.

3.9 Predictor Variables in the Analysis

The Predictor variables are:

Age of vehicle (Agev): This is the difference between the year of manufacturing of the vehicle and the year of study.

Vehicular Pollutants : A pollutant is a waste material that pollutes air, water or soil. Vehicular pollution is caused by the emission of exhaust into the surroundings. Three different vehicular pollutants were considered for the study, namely, Carbonmonoxide (CO), Carbondioxide ($CO_{2)}$) and Hydrocarbon (HC).

Fuel type (ft): Two fuel types were considered for the study namely petrol and diesel.

Type of vehicle (tv): for the study, vehicles were classified into three categories, namely, five passenger vehicle (cars), more than five passengers vehicles (buses) and others.

Vehicle-use (use): This implies what the vehicle is used for (i.e. either for private or commercial purpose).

Diagnosis (diag): Patients diagnosis was based on children with respiratory health problems such as asthma, pneumonia, bronchitis and other cardiovascular diseases.

Number of visit (nv): Number of visits is the number of times a child visits the hospital in a week on respiratory health related problems.

Location : four locations were selected for the study and Abeokuta, which was the largest of them, was used as the reference category.

Chapter 4

RESULTS AND DISCUSSIONS

4.1 Introduction

This chapter is devoted to the application of geoadditive bayesian model in the area of environmental and health effects. The concept of geoadditive Bayesian model is presented. In section 4.2, the model is presented, while section 4.3 presents the model extension, incorporating a new random components using the Bayesian approach. In section 4.4 the estimate of all the components of the model using the concept of finite difference and partial derivatives is presented, and in section 4.5 and 4.6 the prior and the posterior for the geoadditive Bayesian model are specified. In section 4.7 we discuss the properties of the posterior and, in section 4.8 the R code for implementing structured additive regression model is written.

Section 4.9 application presents the result of the estimation and the analysis. Finally, performance of the model is assessed and discussed.

4.2 The Model

Geoadditive Models, introduced by Kammann and Wand(2003), analyze the spatial distribution of the study variable while accounting for possible linear and non-linear covariate effects.Under the additivity assumption they can handle such covariate effects by combining the ideas of additive models and kriging, both represented as linear mixed model. According to Kamman and Wand (2003), incorporation of a geographical component can be achieved by expressing kriging as a linear mixed model and merging it with an additive model such as equation 4.1:

$$Y = X\beta + Zb + \varepsilon \tag{4.1}$$

$$cov\begin{pmatrix} u \\ \varepsilon \end{pmatrix} = 0, cov\begin{pmatrix} u \\ \varepsilon \end{pmatrix} = \begin{pmatrix} \sigma_s^2 I & 0 & 0 \\ 0 & \sigma_t^2 I & 0 \\ 0 & 0 & \sigma_\varepsilon^2 I \end{pmatrix}$$

to obtain a single mixed model, which we call the Geoadditive model. Suppose that the data are $(x_i, y_i), 1 \leq i \leq n$, where the $y_i s$ are scalar and $X_i \epsilon \Re^2$ represents geographical location, the simple universal kriging model for such data is

$$y_i = \beta_0 + \beta_1' x_i + S(x_i) + \varepsilon_i \tag{4.2}$$

where S(x) : $x\epsilon^2$ is a stationary zero-mean stochastic process and the ε_i are assumed to be independent zero-mean random variables with common variance σ_ε^2 and distributed independently of S (e.g Cressie, 1996). Prediction at an arbitrary location $x_0 \varepsilon \Re^2$ is typically done through an expression of the form.

$$\hat{y}(x_0) = \hat{\beta_0} + \hat{\beta_1}' x_0 + \hat{S}(x_i) + \varepsilon_i \tag{4.3}$$

where $\hat{\beta_0}$ and $\hat{\beta_1}$ are estimates of β_0 and β_1 respectively and $\hat{S}(x_0)$ is an empirical best linear unbiased prediction of $S(x_0)$. However, for fitting purposes, we should reparameterized to :

$$y = X\beta + \tilde{Z}\tilde{u} + \varepsilon \tag{4.4}$$

where $\tilde{Z} = Z\Omega^{-\frac{1}{2}}$ and $cov(\tilde{u}) = \sigma_x^2 I$, and utilize the variance component structure. In view of equation (4.1) and (4.4) the geoadditive model

$$Y_i = \beta_0 + f(s_i) + g(t_i) + \beta_1' x_i + S(x_i) + \varepsilon_i \tag{4.5}$$

$$\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$$

where f and g are unspecified smooth functions of s_i and t_i, and S is an unspecified bivariate smooth functions.Then the model has representation is now trivial to formulate as a single linear mixed model. Put

$X = [1 \;\; s_i \;\; t_i \;\; x_i^T]_{1 \leq i \leq n}$,

where Z_s and Z_t are defined and has representation

$$Y = X\beta + Zu + \varepsilon \tag{4.6}$$

where Z $= [Z_s|Z_t|Z_x]$, $Z_x = \widetilde{Z}$

$$u = \begin{pmatrix} u^s \\ u^t \\ \tilde{u} \end{pmatrix}, cov \begin{pmatrix} u \\ \varepsilon \end{pmatrix} = \begin{pmatrix} \sigma_s^2 I & 0 & 0 & 0 \\ 0 & \sigma_t^2 I & 0 & 0 \\ 0 & 0 & \sigma_x^2 I & 0 \\ 0 & 0 & 0 & \sigma_\varepsilon^2 I \end{pmatrix}$$

4.3 Model Extension

Kamman and Wand (2003) combine the idea of kriging and additive model . Heinnerfeind et al. (2006) used Cox model in their application and they applied it to effect of area of residence on Coronary Artery Disease bypass Graft(CABG).

A possible extension by Hennerfeind et al (2006) is :

$$\eta_{it} = \beta_0 + Z_{it}' f(t) + \sum f_j(x_{ij}) + f_{spat}(s_i) + V_{it}'\beta \tag{4.7}$$

where: β - baseline effect

$Z_i(t)$ - time varying effect

f(t) - non linear effect

$f_j(x_{ij})$- unknown smooth function of x

s_i- is the district where subject i resides

f_{spat} is the spatial smoothing function for geographical location s_i where subject i resides

$f_{str}(s_i)$ modelled by a Markov random field prior

$z_i(t)$ time varying effect of covariates z_j

In this work, we will consider the geoadditive model of this form using Bayesian approach;

$$\eta_i(t) = \beta_0 + Z_i^T(t) f_i(t) + \sum_{j=1}^{p} f_j(x_{ij}) + f_{sp}(s_i) + \beta x_i + V_i^T(t)\gamma \tag{4.8}$$

In the equation above, we define;

$\eta_i(t)$ as Linear predictor

β_0 as a vector matrix of the form $\beta = (\beta_1^{'}, ..., \beta_m^{'})^{'}$

$Z_i^T(t)$ as the transpose of the matrix $Z_i = (Z_1, ..., Z_m)$

$f_i(t)$ as the time varying effects of covariates Z_i

$f_i(x_{ij})$ as the nonlinear effect of a continuous covariate x_j

$f_{sp}(s_i)$ as a spatial effect

$\beta_{x_i} = g_i = f_x$ as a new random component which could be the uncorrelated spatial effect and $V_i^{'}(t)$ is a fixed effect.

$\beta_x \sim N(0, \tau^2)$.

Assuming that we consider (4.8) under generalized linear and additive mixed models, equation (4.8) is seen as predictor for the observations Y_i, such that:

$$Y_i = \beta_0 + f(s_i) + g(t_i) + \beta_i^{'} x_i + \varepsilon_i \tag{4.9}$$

It follows apparently that the models under consideration have the densities of the form (Sun et al., 1999)

$$f_i(y_i/x_i) \tag{4.10}$$

The partially improper priors in the predictor are: $\beta_0, f_i(t), g_i$ and γ, such that we have a flat prior

$$P(\beta_0) \equiv 1 \tag{4.11}$$

$$P(\gamma) \equiv 1 \tag{4.12}$$

Since β_0 and γ are vectors of fixed effects. Because $f_i(t), g_i = \beta_x = f_x$ are vectors of varying or random effects, their flat prior are given as

$$P(f_i(t))\alpha\tau_i^{-\tau_i} \exp\left(\frac{-1}{2\tau_i^2} \acute{f}_i(t) k_i f_i^{(t)}\right) \tag{4.13}$$

$$P(g_i) = P(\beta_x) = P(f_x) \propto \tau^{-i\tau} \exp\left(\frac{-1}{2\tau_i^2} g_i D_i g_i\right) \tag{4.14}$$

It is good to note that the priors for the hyper-parameters $\tau^2 = (\tau_0^2 + \tau_1^2 + \tau_2^2 + \tau_3^2 + \tau_4^2 + \tau_{5^2} + \tau_6^2)$ are given by:-

$$P(\tau^2) = \prod_{i=0}^{6} P(\tau_i^2) \tag{4.15}$$

This is because we are considering 6 different hospitals. For the term $Z_i^T(t)f_i(t)$ for i=0, we obtain $Z_0^{'}(t)f_0(t)$ which usually represents a spatial effect with MRF or Kriging prior, or an unstructured spatial effect and all these are a types of a random effect. Recall that in a geoadditive model with predictor:

$$\eta = V\beta + Z_1\beta_1 + Z_2\beta_2 + ... + Z_m\beta_m + Z_0\beta_0 \tag{4.16}$$

$Z_0\beta_0$ will usually represent a spatial effect with a Markov random field or Kriging prior, or an unstructured spatial effect. Bringing this understanding into our new model whose predictor is given as

$$\eta_i(t) = \beta_0 + Z_i^{'}(t)f_i(t) + \sum_{j=i}^{p} f_j(x_{ij}) + f_{sp}(s_i) + \beta x_i + V_i^{'}(t)\gamma \tag{4.17}$$

Where $\beta_x \sim N(0, \tau^2)$ $i : 1(1)m, j = 1(1)P$ The above implies that:
For i=1, we have

$$\eta_1(t) = \beta_0 + Z_1^{'}(t)f_1(t) + \sum_{j=1}^{p} f_j(x_{1j}) + f_{sp}(s_1) + \beta x_1 + V_1^{'}(t)\gamma \tag{4.18}$$

For i=2,we have:

$$\eta_2(t) = \beta_0 + Z_2^{'}(t)f_2(t) + \sum_{j=2}^{p} f_j(x_{2j}) + f_{sp}(s_2) + \beta x_2 + V_2^{'}(t)\gamma \tag{4.19}$$

$\vdots$

For $i = m$, we have

$$\eta_m(t) = \beta_0 + Z_m^{'}(t)f_m(t) + \sum_{j=m}^{p} f_j(x_{mj}) + f_{sp}(s_m) + \beta x_m + V_m^{'}(t)\gamma \tag{4.20}$$

In all these equations $\eta = (\eta_1, \eta_2, ..., \eta_m)$.
The vector of evaluations of the function $Z_i(t)$

$$\begin{aligned}
Z_1 &= Z_1(t_1), Z_1(t_2), Z_1(t_3), ...Z_1(t)^{'} \\
Z_2 &= Z_2(t_1), Z_2(t_2), Z_2(t_3), ...Z_2(t)^{'} \\
&= \vdots \quad \vdots \quad \vdots \\
Z_m &= Z_m(t_1), Z_m(t_2), Z_m(t_3), ...Z_m(t)^{'}
\end{aligned}$$

$$f_j = (f_j(x_{ij}), f_j(x_{2j}), ..., f_j(x_{mj}))^{'}, \quad j = 1, 2, ..., p$$

$$f_{spat} = (f_{spat}(s_1), f_{spat}(s_2), ..., f_{spat}(s_m))'$$

$\beta = (\beta_{x1}, \beta_{x2}, ..., \beta_{xm})'$ vector of random component

From (4.18) to (4.20) the matrix form are:

$$\eta = \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \\ \eta_5 \\ \vdots \\ \eta_m \end{bmatrix}' \quad \beta_{xi} = \begin{bmatrix} \beta_{x1} \\ \beta_{x2} \\ \beta_{x3} \\ \beta_{x4} \\ \beta_{x5} \\ \vdots \\ \beta_{xm} \end{bmatrix}' \quad Z = \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \\ Z_5 \\ \vdots \\ Z_m \end{bmatrix}' \quad f(t) = \begin{bmatrix} f_1(t) \\ f_2(t) \\ f_3(t) \\ f_4(t) \\ f_5(t) \\ \vdots \\ f_m(t) \end{bmatrix}'$$

$$fj(x) = \begin{bmatrix} \sum_{j=1}^{p} fj(x_{i1}) \\ \sum_{j=1}^{p} fj(x_{i2}) \\ \sum_{j=1}^{p} fj(x_{i3}) \\ \sum_{j=1}^{p} fj(x_{i4}) \\ \sum_{j=1}^{p} fj(x_{i5}) \\ \vdots \\ \sum_{j=1}^{p} fj(x_{ip}) \end{bmatrix}' \quad f_{spat}(s_1) = \begin{bmatrix} f_{spat}(s_1) \\ f_{spat}(s_2) \\ f_{spat}(s_3) \\ f_{spat}(s_4) \\ f_{spat}(s_5) \\ \vdots \\ f_m(t) \end{bmatrix}'$$

The equation (4.18) to (4.20) can be compactly written as :

$$\eta = \beta_0 + Zf(t) + f(x) + f_{spat} + f_x + V\gamma \tag{4.21}$$

$$\eta = \beta_0 + Z_1 f_1(t) + ... + Z_5 f_5(t) + Z_m f_m(t) + f(x) + f_{spat} + f_x + V\gamma \tag{4.22}$$

Let $V\gamma \cong Z_0 f_0(t)$ then (4.22) becomes

$$\eta = \beta_0 + Z_1 f_1(t) + ... + Z_m f_m(t) + f(x) + f_{spat} + f_x + Z_0 f_0(t) \tag{4.23}$$

Another possibility is to take:

$f(x) + f_{spat}(s) + f_x + V\gamma$ to be equal to $Z_0 f_0(t)$ such that

$$\eta = \beta_0 + Z_1 f_1(t) + ... + Z_m f_m(t) + Z_0 f_0(t) \tag{4.24}$$

It follows that if $Zf(t) = Z_1 f_1(t) + ... + Z_m f_m(t)$ then

$$\eta = \beta_0 + Zf(t) + Z_0 f_0(t) \tag{4.25}$$

Choosing f(t) as γ and $f_0(t)$ as β_0, we have

$$\eta = \beta_0 + Z\gamma + Z_0\beta_0 \tag{4.26}$$

It is also possible for us to write (4.22) as follows:

$$\eta = Z_0\beta_0 + Zf(t) + ... + \xi + V\gamma \tag{4.27}$$

$$\eta = Z_0\beta_0 + Z_1 f_1(t) + ... + Z_m f_m(t) + f(x) + f_{spat} + f_x + V\gamma \tag{4.28}$$

Where $Z_0 = I$ is an identity matrix. By this expression it follows that

$$\eta = Z_0\beta_0 + Zf(t) + \xi + V\gamma \tag{4.29}$$

Where

$$\xi = f(x) + f_{spat} + f_x \tag{4.30}$$

The expressions in (4.23), (4.25), (4.26) and (4.29) are all acceptable since the terms in them are in vector form,if we decompose β_0 into a $(d_0 - r_0)$ dimensional sub-vector β_0^{fl} with flat prior $P(\beta_0^n) \equiv 1$ and a r_0 dimensional sub-vector β_0^{pr} with a prior

$$\beta_0^{pr} \sim N(0.\tau_0^2)I$$

Then,

$$\beta_0 = Z_0^{fl}\beta_0^{fl} + Z_0^{pr}\beta_0^{pr} \tag{4.31}$$

where the $d_0 \times (d_0 - r_0)$ matrix Z_0^{fl} contains a basis if the null space of K_0, Z_0^{fl} is an identity vector 1 for P splines with first order random walk prior , Markov random field and $2d$- P - splines with Markov random field prior for the coefficients.

The $d_0 \times r_0$ matrix Z_0^{pr} is given by $Z_0^{pr} = L(L^T L)^{-1}$, where

$$L = S^T\lambda \tag{4.32}$$

Equation (4.32) is obtained from the spatial decomposition

$$L = \lambda S^T \ \text{ of } \ K_0 \tag{4.33}$$

It follows that

$$\beta_0 = Z_0^{fl}\beta_0^{fl} + Z_0^{pr}\beta_0^{pr} \tag{4.34}$$

Our aim is to obtain a model y_i whose predictor is η_i and to determine the corresponding terms in our predictor. Thus, our take off is i and $i+1$ be two mesh points in the associated interval such that (4.28) becomes:

$$\eta_i = Z_0\beta_0 + Z_1 f_1(t) + ... + Z_m f_m(t) + f_i(x_{ij}) + f_{spat}(s_i) + f_{x_i} + V_i\gamma = y_i \tag{4.35}$$

and

$$\eta_{i+1} = Z_0\beta_0 + Z_1 f_1(t) + ... + Z_m f_m Z_{m+1} f_{m+1}(t) +$$
$$f_{i+1}(x_{i+1}) + f_{spat}(s_i) + f_{x_i+1} + V_{i+1}\gamma = y_{i+1} \tag{4.36}$$

The assumption behind the expressions in (4.35) and (4.36) is that we allow η_i and η_{i+1} to coincide with y_i and y_{i+1} respectively so as to be able to determine the undetermined coefficients in the predictor, using differentiation method and finite difference approach. Using(4.5)under the above assumption and that the derivatives of η_i and η_{i+1} exist continuously, we obtain

$$\eta_i(t) = \beta_0 + Z_i^T(t) f_i(t) + \sum_{j=i}^{p} f_j(x_{ij}) + f_{sp}(s_i) + \beta x_i + V_i^T(t)\gamma = y_i \tag{4.37}$$

$$\eta_{i+1}(t) = \beta_0 + Z_{i+1}^T(t) f_{i+1}(t) + \sum_{j=i}^{p} f_j(x_{i+1_j}) + f_{sp}(s_{i+1}) + \beta x_{i+1} + V_{i+1}^T(t)\gamma = y_{i+1} \tag{4.38}$$

It follows that:

$$\eta_{i+1}(t) - \eta_i(t) \tag{4.39}$$

$$= y_{i+1} - y_i = \beta_0 - \beta_0 + Z_{i+1}^T(t) f_i(t) - Z_i^T(t) f_i(t) +$$
$$\sum_{j=i}^{p} f_j(x_{i+1j}) - \sum_{j=i}^{p} f_j(x_{ij}) + f_{sp}(s_{i+1}) - f_{sp}(s_i)$$
$$+\beta x_{i+1} - \beta x_i + V_{i+1}^T(t)\gamma - V_i^T(t)\gamma \tag{4.40}$$

From (4.39), we obtain the sum difference of

$$\eta_{i+1}(t) - \eta_i(t) = y_{i+1} - y_i = \beta_0 - \beta_0 + Z_{i+1}^T(t) f_i(t) - Z_i^T(t) f_i(t) +$$
$$\sum_{j=i}^{p} [f_j(x_{i+1j}) - f_j(x_{ij})] + f_{sp}(s_{i+1}) - f_{sp}(s_i)$$

$$+\beta x_{i+1} - \beta x_i + V_{i+1}^T(t)\gamma - V_i^T(t)\gamma \tag{4.41}$$

This leads to

$$y_{i+1} - y_i = \eta_{i+1} - \eta_i \tag{4.42}$$

$$y_{i+1} - y_i = Z_{i+1}^T(t)f_i(t) - Z_i^T(t)f_i(t) +$$

$$\sum_{j=i}^{p}[f_j(x_{i+1j}) + f_{sp}(s_i) - f_j(x_{ij})] + f_{sp}(s_{i+1}) - f_{sp}(s_i)$$

$$+\beta x_{i+1} - \beta x_i + V_{i+1}^T(t)\gamma - V_i^T(t)\gamma \tag{4.43}$$

4.4 Determination of the Undetermined Components

By using the concept of finite difference and partial derivatives, the undetermined components of (4.8) are obtained as follows

$$\frac{d\eta_i}{dt} = Z_i^T(t)f_i'(t) + [Z_i^T(t)]'f_i(t) + [V_i^T(t)]'\gamma \tag{4.44}$$

$$\frac{d\eta_i}{dx_{ij}} = \sum_{j=1}^{p} f_i'(x_{ij}) \tag{4.45}$$

$$\frac{d\eta_i}{ds_i} = f_{sp}'(s_i) \tag{4.46}$$

$$\frac{d\eta_i}{dx} = \beta_{x_i}' \tag{4.47}$$

From (4.44)

$$\frac{d\eta_i}{dt} = Z_i^T \frac{df_i}{dx_i} + \frac{d}{dt} Z_i^T f_i(t) + \frac{d}{dt}[V_i^T(t)]\gamma \tag{4.48}$$

From (4.45)

$$\frac{d\eta_i}{dx_{ij}} = \sum_{j=1}^{p} \frac{d}{d} f_i(x_{ij}) \tag{4.49}$$

From (4.46)

$$\frac{d\eta_i}{ds_i} = \frac{d}{ds} f_{sp}(s_i) \tag{4.50}$$

From (4.47)

$$\frac{d\eta_i}{dx_i} = \frac{d}{dx}\beta_{x_i} \tag{4.51}$$

According to Heinnerfield et *al.* (2006), the baseline hazard rate can be reparameterized by choosing $\beta_0 = \log \lambda_0(t)$ and observation model is given by

$$\lambda_i(t) = \lambda I(t : x_i, z_i, s_i, v_i) = \exp \eta_i(t)$$

with Geoadditive predictor as defined in equation(4.8). Correspondingly, for the unknown functions $f_j(x_{ij})$, βx_i, we assume Bayesian P-spline priors as in Lang and Brezger (2004). For the Random Walk priors, Fahrmier and Lang (2001) suggested that they may be used as smoothness priors for the baseline effect and time-varying covariates effects in a piecewise exponential model (Heinnerfeind et al. 2006)
In addition, the basic idea of P-spline regression by Eilers and Marx (1996) will be used to approximate the unknown function as a linear combination of B-spline basic function B_t:

$$f_j(x_{ij}) = \sum_{t=1}^{p} \beta_t B_t(x_j) \tag{4.52}$$

where B_t are B-splines of degree I defined over a grid of equally spaced knots

$$x_{\min} = \xi_0 < \xi_1 < \ldots < \xi_s = x_{\max}; d_j = I + s$$

According to Heinnerfeind et al. (2006) in a simulation study $f_{spat}(s_i)$ was model as MRF using trigonometric functions to simulate for the spatial effect and the nonlinear function.

$$f_{spat}(s_i) = \sin(x_{s_i}, y_{s_i}) \tag{4.53}$$

$$f_j(x_{ij}) = \sin(x_i) \tag{4.54}$$

The baseline hazard rate $\lambda_0(t)$ is set to $3t^2$ which is a Weibull Hazard rate, so that

$$\beta_0(t) = \log 3t^2 \tag{4.55}$$

To determine the undetermined coefficient of our predictor, we will be making use of cubic spline and the random nature in the distribution of prime numbers for a better performance setting $f_{spat}(s_i)$ as ;

$$f_{spat}(s_i) = \sin(x_{s_i}, y_{s_i}) + \cos(x_{s_i}, y_{s_i}) \tag{4.56}$$

According to Hennerfeind et al. (2006) as shown in (4.52) and (4.53) we choose $f_j(x_{ij})$ to be a replica of $f_{spat}(s_i)$ such that

$$f_j(x_{ij}) = \sin(x_i) + \cos(x_i) \tag{4.57}$$

Thus,

$$f_i(t) = P + \log(Pt^3) + 1 \tag{4.58}$$

where P is an arbitrary prime number.

Also, we set our baseline hazard rate $\lambda_0(t)$ to Pt^3, so that $\beta_0(t) = log(Pt^3)$ For β_x we make use of a function which exhibits a random nature aside the known trigonometric function that we have already used. Thus,

$$\beta_x \equiv e^{Px} \tag{4.59}$$

where P is also a prime number since we are using the random nature in the distribution of primes. γ_i is a fixed effect and we assume that is given by

$$\gamma_i = \pm\Gamma^{-1}P \tag{4.60}$$

For the non-linear time-varying effect $Z^T(t)$, we model this by using a matrix representation of the form

$$Z^T(t) = \begin{bmatrix} \cos(Pt) & \cos t & 0 \\ \cos t & \cos(Pt) & e^{2Pt} \\ 0 & e^{2Pt} & \cos(Pt)^2 \end{bmatrix}$$

P$\subseteq$prime number and the linear time-varying effect matrix as;

$$V^T(t) = \begin{bmatrix} 3P & P^2t & 0 \\ P^2 & 3Pt & 3P^2t \\ 0 & 3P^2t & 1 \end{bmatrix}$$

The above mentioned matrices can be extended to fit into our model but must always be square symmetric matrices. Thus, we are able to determine the stepwise difference (forward difference) of our predictor as presented in (4.37) and (4.40)

$$y_{i+1} = y_i + \begin{bmatrix} \cos(Pt_{i+1}) & \cos t_{i+1} & 0 \\ \cos t_{i+1} & \cos(Pt_{i+1}) & e^{2Pt_{i+1}} \\ 0 & e^{2Pt_{i+1}} & \cos(Pt_{i+1})^2 \end{bmatrix} \begin{bmatrix} 1 \\ P \\ \log(Pt_{i+1})^3 \end{bmatrix}$$

$$-\begin{bmatrix} \cos(Pt_i) & \cos t_i & 0 \\ \cos t_i & \cos(Pt_i) & e^{2Pt_i} \\ 0 & e^{2Pt_i} & \cos(Pt_i)^2 \end{bmatrix} \begin{bmatrix} 1 \\ P \\ \log(Pt_i)^3 \end{bmatrix}$$

$$+\sum_{j=1}^{P} \Big[[\sin(x_{i+1}) + \cos(x_{i+1})] - [\sin(x_i) + \cos(x_i)] \Big]$$

$+$

$$\Big[[\sin(x_{s_{i+1}}, y_{s_{i+1}}) + \cos(x_{s_{i+1}}), y_{s_{i+1}}] - [\sin(x_{s_i}, y_{s_i}) + \cos(x_{s_i}), y_{s_i}] \Big]$$

$+$

$$[e^{Px_{i+1}} - e^{Px_i}]$$

$+$

$$\Gamma^{-1}(P) \begin{bmatrix} 3P & P^2 t_{i+1} & 0 \\ P^2 & 3Pt_{i+1} & 3P^2 t_{i+1} \\ 0 & 3P^2 t_{i+1} & 1 \end{bmatrix} - \Gamma^{-1}(P) \begin{bmatrix} 3P & P^2 t_i & 0 \\ P^2 & 3Pt_i & 3P^2 t_i \\ 0 & 3P^2 t_i & 1 \end{bmatrix}$$

From (4.44);

$$\frac{d\eta_i}{dt} = Z_i^T(t) \left[\frac{d}{dt}[1 + P + \log(Pt^3)] \right]$$

$+$

$$\frac{d}{dt} \begin{bmatrix} \cos(Pt) & \cos t & 0 \\ \cos t & \cos(Pt) & e^{2Pt} \\ 0 & e^{2Pt} & \cos(Pt)^2 \end{bmatrix} [1 + P + \log(Pt^3)]$$

$+$

$$\frac{d}{dt} \begin{bmatrix} 3P & P^2 t & 0 \\ P^2 & 3Pt & 3P^2 t \\ 0 & 3P^2 t & 1 \end{bmatrix} \gamma$$

$$\frac{d\eta_i}{dt} = \begin{bmatrix} \cos(Pt) & \cos t & 0 \\ \cos t & \cos(Pt) & e^{2Pt} \\ 0 & e^{2Pt} & \cos(Pt)^2 \end{bmatrix} \left[\frac{3Pt^2}{Pt^3} \right] +$$

$$\begin{bmatrix} -P\sin(Pt) & -\sin t & 0 \\ -\sin t & -P\sin(Pt) & 2Pe^{2Pt} \\ 0 & 2Pe^{2Pt} & -2P\sin(Pt)^2 \end{bmatrix} [1 + P + \log(Pt^3)]$$

$+$

$$\gamma \begin{bmatrix} 0 & P^2 & 0 \\ P^2 & 3P & 3P^2 \\ 0 & 3P^2 & 0 \end{bmatrix}$$

$$= \frac{3}{t} \begin{bmatrix} \cos(Pt) & \cos t & 0 \\ \cos t & \cos(Pt) & e^{2Pt} \\ 0 & e^{2Pt} & \cos(Pt)^2 \end{bmatrix}$$

$$+[1 + P + \log(Pt^3)] \begin{bmatrix} -P\sin(Pt) & -\sin t & 0 \\ -\sin t & -P\sin(Pt) & 2Pe^{2Pt} \\ 0 & 2Pe^{2Pt} & -2Pt\sin(Pt)^2 \end{bmatrix}$$

$+$

$$\gamma \begin{bmatrix} 0 & P^2 & 0 \\ P^2 & 3P & 3P^2 \\ 0 & 3P^2 & 0 \end{bmatrix}$$

From (4.45);

$$\frac{d\eta_i}{dx_{ij}} = \sum_{j=1}^{P} [\cos(x_i) - \sin(x_i)]$$

From (4.46);

$$\frac{d\eta_i}{ds_i} = \sum_{j=1}^{P} [\cos(x_{s_i}, y_{s_i}) - \sin(x_{s_i}, y_{s_i})]$$

$$\frac{d\eta_i}{dx_{ij}} = Pe^{Px_i}$$

4.5 Specification of Priors for the Geoadditive Bayesian Model

Brezger and Lang (2006) suggested a general structure of the prior as

$$\beta_j \setminus \tau_j^2 \propto \tau_j^{-r_j} exp\left(\frac{-1}{2\tau_j^2}\beta_j' K_j \beta_j\right) \tag{4.61}$$

In our work we will assume diffuse priors or weakly informative prior for the fixed effect that is $P(\gamma) \propto const$. This is the appropriate choice when there is no prior knowledge. The uncorrelated random effects are assumed to be i.i.d Gaussian, $\beta_x \sim N(0, \tau_b^2)$. τ_b^2 we assign a conjugate prior for τ^2 which is the inverse Gamma prior with hyper parameters a and b i.e. $\tau^2 \sim IG(a, b)$. The effects of the continuous covariates are modelled by cubic p-splines with 20 equidistant knots and second order random walk penalty.Common choices for a and b are a=1 and b=0.005 (or b=0.0005).12,000 iterations of the MCMC were run using BayesX package with a burn-in phase of 2,000 iterations.Thinning was applied to the Markov Chain to reduce autocorrelations, by requiring the programme to store only every 10th sampled parameter. Alternatively, one may take a=b=0.001, we take a = b= 0.001 as a standard choice and to test for the sensitivity other values of a and b were considered. We also consider cases of when a=b=0.00001,a=1,b=0.005 and a=1,b=0.00005.

For the continuous(smooth) functions $f_1 ... f_6$,a second order random walk prior is considered for f defined as follows. Consider the case of a metrical covariate x with equally spaced observations $x_i, i = 1, ..., m, m \leq n$ (n is the number of observations). Suppose that $x_1 < ... < x_t < ... < x_m$ is an ordered sequence of distinct values for a covariate and define $f(t) = f(x_t)$. The second order random walk is then defined by

$$f(t) = 2\beta_{t-1} - \beta_{t-2} + U_t \tag{4.62}$$

with Gaussian errors $U_t \sim N(0, \tau^2)$.

A second order random walk penalizes deviations from the linear trend $2\beta_{t-1} - \beta_{t-2}$. Let $\beta = (\beta_0' .....\beta_m')'$ denote the vector of all regression coefficients and γ the vector of fixed effects, and $\tau^2 = \tau_0^2.....\tau_m^2$ the vector of all variance components. Fully Bayesian inference is based on the entire posterior distribution

$$p(\beta, \gamma, \tau^2 \backslash x) \propto L(\beta, \gamma, \tau^2) p(\beta, \gamma, \tau^2) \tag{4.63}$$

Due to conditional independence assumptions, the joint prior factorizes into

$$p(\beta, \gamma, \tau^2) = \{\prod_{j=0}^{m} p(\beta_j \backslash \tau_j^2) p(\tau_j^2)\} p(\gamma) \tag{4.64}$$

The last factor can be omitted for diffuse fixed effect priors. In which case these can be expressed as:

$$\prod(\theta) = \prod(\tau^2) \prod(\gamma) \prod(\beta/\tau_j^2, \gamma) \tag{4.65}$$

$$\prod(\theta) = \prod(\tau_j^2) \prod(\beta/\tau_j^2) \tag{4.66}$$

where $\beta \sim N(0, \tau_j^2)$ is chosen to be normally distributed and $\tau_j^2 \sim IG(a_j, b_j)$

$$\prod(\tau_j^2) = \frac{b_j^{a_j}}{\Gamma a} (\tau_j^2)^{-a_j-1} e^{\frac{b_j}{\tau^2}}$$

$$\prod(\beta/\tau_j^2) = \frac{1}{\sqrt{2\Pi\tau_j^2}} e^{\frac{-1}{2\tau_j^2}(\beta)^2}$$

$$\prod(\theta) = \prod(\tau_j^2) = \frac{b_j^{a_j}}{\Gamma a} (\tau_j^2)^{-a_j-1} e^{\frac{b_j}{\tau^2}} \times \frac{1}{\sqrt{2\Pi\tau_j^2}} e^{\frac{-1}{2\tau_j^2}(\beta)^2}$$

Where for τ_j^2 :

$$a_j^{'} = a_j + \frac{rank(k_j)}{2}$$

and

$$b_j^{'} = b_j + \frac{1}{2}\beta_j^{'} k_j \beta_j$$

$$\prod(\theta) = \frac{b_j^{a_j}}{\Gamma a_j} (\tau_j^2)^{-a_j-1} e^{\frac{-b_j}{\tau_j^2}} \times \frac{1}{\sqrt{2\Pi\tau_j^2}} e^{\frac{-1}{2\tau_j^2}\beta^2}$$

$$= \frac{b_j^{a_j}}{\Gamma a_j \sqrt{2} (\tau_j^2)^{n/2}} e^{\frac{-1}{2\tau_j^2}\beta^2} (\tau_j^2)^{-a_j-1}$$

$$= \frac{1}{(\tau_j^2)^{\frac{(rk(k_j)}{2}}} e^{\frac{-1}{2\tau_j^2}\beta_j^{'} k_j \beta_j} (\tau_j^2)^{-a_j-1} e^{\frac{-b_j}{\tau_j^2}}$$

4.6 Properties of the Posteriors

Consider a geoadditive model with predictor

$$\eta = V\gamma + Z_1\beta_1 + - - - + Z_m\beta_m + Z_0\beta_0 \tag{4.67}$$

in a generic form, where $Z_0\beta_0$ corresponds to an effect with prior.

$$\beta_0 \sim \tau_0^{-r_0} exp(\frac{1}{2\tau_0^2}\beta_0' K_0 \beta_0) \tag{4.68}$$

such that

$$dim(\beta_0) = d_0 \geq d_j, rank(K_0) = r_0 \geq r_j$$

, j =1- - - m.

This assumption is usually fulfilled for the spatial component or for a high dimensional vector of group or individual specific correlated random effects. Denote by $\eta_u, V_u, Z_u = (Z_{iu}, - - -, Z_{mu}), Z_{ou}$ the (sub-) predictor and sub-design matrices corresponding to uncensored observations. Assume that the following conditions hold

(C1) rank(Vu)= rank(V)= p = dim(γ)

$$rank(Z_{ju}) = rank_{Z_j} = d_j = dim(\beta_j), j = 0, - - -, m$$

$$rank(Z_u' R Z_u + K) = d \tag{4.69}$$

where

$$d = d_1 + - - - + d_m, K = diag(k_1, - - - k_m), R = I - V_u(V_u' V_u)^{-1} V_u'$$

(C2) The priors $p(\tau_j^2), j = 1, - - -, m$, are proper,

and $\int p(\tau_0^2)\tau_0^{-(r_0-p-(d-r))} d\tau_0^2 < \infty$,

where r $= r_1 + - - - + r_m$

Theorem: If conditions (C1),(C2) hold then the posterior $p(\gamma, \beta, \beta_0, \tau^2, \tau_0^2/y)$, where $\tau^2 = (\tau_1^2, - - -, \tau_m^2)'$ and $\beta = (\beta_1, - - -, \beta_m)$, is proper.

Corollary: Assume proper inverse Gamma priors for τ_j^2 with

$$a_j > 0, b_j > 0, j = 0, - - -, m,$$

and

$$r_0 - p - (d - r) - (d_0 - r_0) > 0$$

If condition (C1) holds, then the posterior

$$p\left(\gamma, \beta, \beta_0, \tau^2, \tau_0^2 \setminus y\right)$$

is proper.

$$p + d = \text{rank}\begin{bmatrix} V_u' V_u & V_u' Z_u \\ Z_u' V_u & Z_u' Z_u + K \end{bmatrix}$$

Proof: The density of observation i is given by

$$f_i(t_i/\eta_i(t_i) = \lambda_i(t_i)^{\delta_i} S_i(t_i),$$

where

$$\lambda_i(t_i) = exp(\eta_i(t_i), S_i(t_i) = exp(-\int_0^{t_i} \lambda_i(s)ds)$$

For censored observations ($\delta_i = 0$), we have $f_i(t_i \setminus \eta_i(t_i)) = S_i(t_i) \leq 1$, so that condition

$$f_i(y_i/\eta_i) \leq M, i = n^* + 1, - - -, n$$

holds.

For uncensored observations ($\delta_i = 1$)

$$f_i(t_i/\eta_i(t_i)) = \lambda_i(t_i) S_i(t_i)$$

setting $\eta_i := \eta_i(t_i), \lambda_i := \lambda_i(t)$, we obtain

$$\int_0^\infty f_i(t_i \setminus \eta_i) d\eta_i = \int_0^\infty \lambda_i S_i(t_i) \lambda_i^{-1} d\lambda_i = \int_0^\infty S_i(t_i) d\lambda_i$$

so that

$$\int f_i(y_i \setminus \eta_i) d\eta_i) < \infty = \int_0^\infty S_i(t_i) d\lambda_i < \infty$$

4.7 Specification of Posterior for the Geoadditive Bayesian Model

The posterior is obtained by combining the prior (out-of-sample) information with the likelihood (data).

The prior is obtain to be :

$$\frac{1}{(\tau_j^2)^{\frac{(rk(k_j)}{2}}} e^{\frac{-1}{2\tau_j^2}\beta_j' k_j \beta_j} (\tau_j^2)^{-a_j - 1} e^{\frac{-b_j}{\tau_j^2}}$$

Posterior:

$$\prod(\theta) = \prod(\tau^2)\prod(\beta/\tau^2)L(\beta,\gamma,\tau^2)$$
$$= \frac{1}{(\tau_j^2)^{\frac{rk(k_j)}{2}}} e^{\frac{-1}{2\tau_j^2}\beta_j' k_j \beta_j} (\tau_j^2)^{(-a_j-1)e^{-b_j/\tau_j^2}}$$
$$= exp[g_0(t) + \frac{1}{\sqrt{2\Pi\tau^2}} e^{\frac{-1}{2}(\beta_j)^2} \sum_{j=1}^{p} z_{ij} + \frac{1}{\sqrt{2\Pi\tau^2}} e^{\frac{-1}{2\tau^2}(\beta_j)^2} + \frac{1}{\sqrt{2\Pi\tau^2}}$$
$$e^{\frac{-1}{2\tau^2}(\beta^{spat})^2} + \frac{1}{\sqrt{2\Pi\tau^2}} e^{\frac{-1}{2\tau^2}(\beta_{gi})^2}].exp\int_0^{t_i} exp(g_0(u) + \sum_{j=1}^{p} g_j(u)z_{ij})du$$

4.8 R Code for Implementing Structured Additive Regression Model

♯♯ stepwise algorithm

♯♯ generate some data

$n \leftarrow 1000$

♯♯ regressors

$dat \leftarrow data.frame(x_1 = runi + (n, -3, 3), x_2 = runif(n)$ x_3= runif(n,3,6), x_4= runif(n,0,1))

♯♯ response

$dat\$y \leftarrow with(dat \leftarrow, 1.5 + sin(x_1) + 0.6 + x_2 + rnorm(n, sd = 0.6)$

♯♯ estimate the model with STEP

$b \leftarrow bayesX(y \sim sx(x_1) + sx(x_2) + sx(x_3) + sx(x_4),$

method=”STEP”, algorithm= C descent1”, CI=”MCMC select”, iter = 10000,step=10, data=dat)

Summary (b)

plot(b)

♯♯ a probit example

set.seed(111)

n← 1000

dat ← data.frame(x ← runi(n,-3,3))

$dat\$z$ ←with (dat,sin(x)+rnorm(n))

dat \$y ← rep(0,n)

dat\$y[dat\$z¿ 0] ← 1

```
b ← bayesX (y ∼ sx(x), family = "binomial probit", data = dat)
Summary(b)
plot(b)
♯♯ estimate varying coefficient models
set.seed(333)
n ← 1000
dat ←.frame(x=runif(n,-3,3),id= factor(rep(1:4,n/4))]
♯♯ response
dat$ y ← with(dat,1.5+sin(x)∗ c(-1,0.2,1,5)[id]+rnorm(n,sd=0.6))
♯♯ estimate model
b ← bayesX(y ∼ sx(x,by = id,center=TRUE),
method= "REML",data=dat)
summary(b)
plot(b,resid=TRUE,cex.resid=0.1)
♯♯ End(Not run)
```

4.9 Model Estimation and Analysis

4.9.1 EXPLORATIVE DATA ANALYSIS

The study included 9211 vehicles with their age ranging from 2 to 39 years. The average of these vehicles was 11.21 years and a standard deviation of 8.35 years. The standard deviation showed a wide variation in the ages of vehicles considered. The hydrocarbon collected ranged from 0 to 19812 ppm with an average of 1500.84 ppm and standard deviation 2072.201 ppm. Carbonmonoxide ranged from 0 to 14370 ppm with an average of 1487 ppm (SD = 1678.9 ppm). Also, carbondioxide ranged from 0 to 13240 ppm with an average of 7586.4 ppm (SD = 4652 ppm).

Over nine thousand respondents were seen, 30.6% of whom had asthma, 21.64% had pneumonia, 7.72% had bronchitis and 40.4% had cardiovascular diseases.

Furthermore, 94.88% of the vehicle sampled ran on petrol while 5.12% ran on diesel.

In addition, 52.40% vehicle seen were commercial vehicles while 47.60%

were privately owned vehicles.

Finally, 53.56% of the vehicles were cars while 42.63% were buses and 3.81% were categorized as other types of vehicles.

The data were subjected to explorative analysis and the behavior of the various variables considered was shown. Box plot and histogram of the various vehicular pollutants were shown and the plot of carbonmonoxide and hydrocarbon revealed that there are a lot of outliers and extreme values, which really attracted our attention because it shows that the people were keenly affected by vehicular emission. Also, their distribution is positively skewed, which implies that it could be modelled by many positively fitting distributions. Other emission variables considered also showed that nonlinear relationship exists among all the emission variables, since they follow a nonlinear pattern in their distribution.

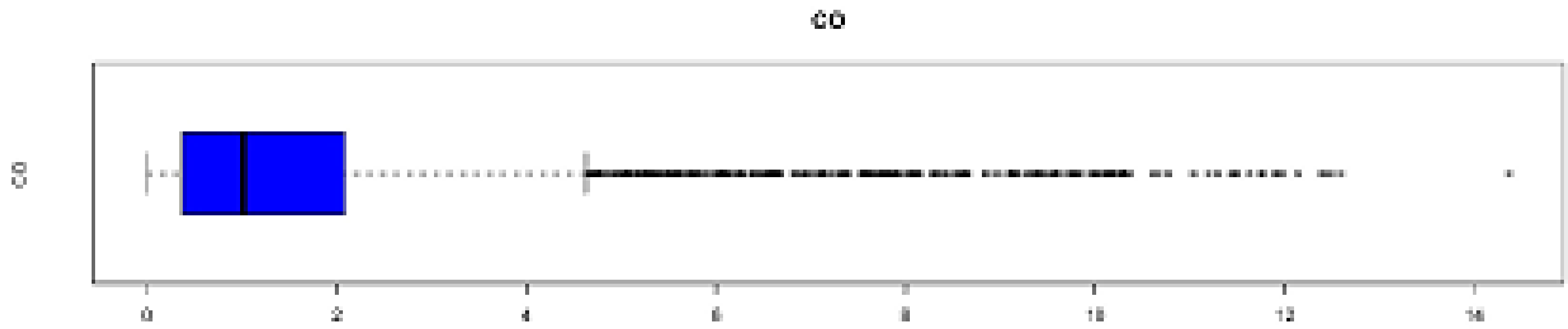


Figure 4.1: Picture Showing Distribution of Pollutants using Box-Plot

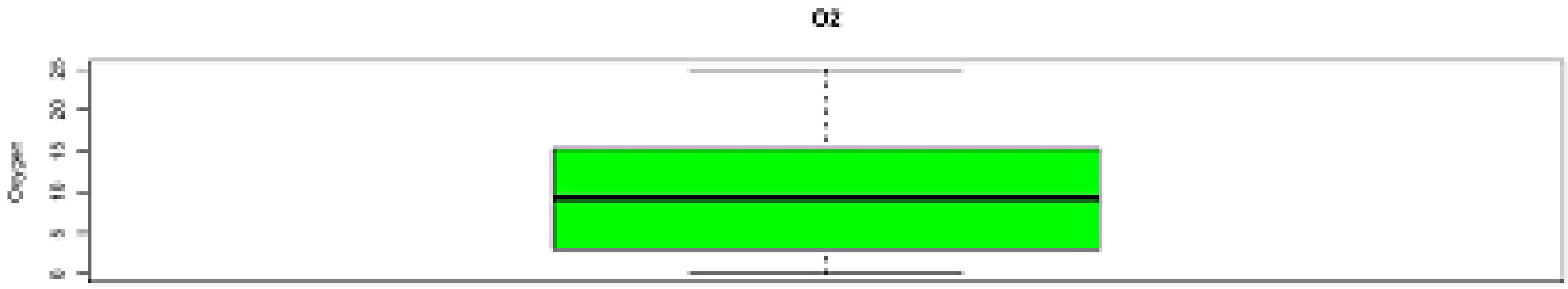


Figure 4.2: Picture Showing Distribution of Pollutants using Box-Plot

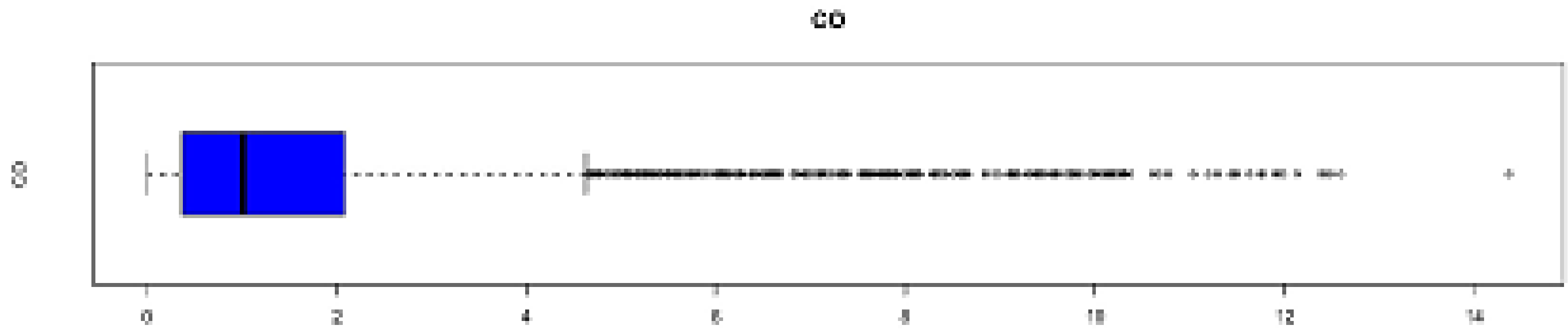


Figure 4.3: Picture Showing Distribution of Pollutants using Box-Plot

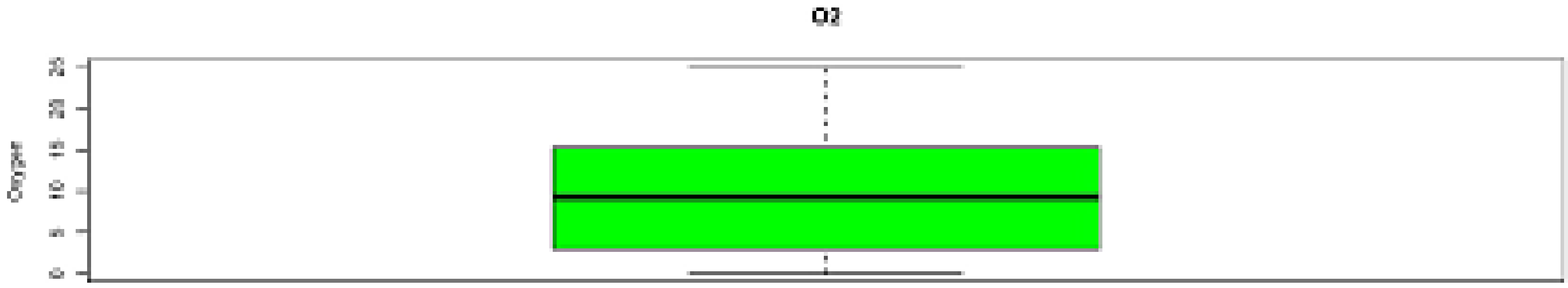


Figure 4.4: Picture Showing Distribution of Pollutants using Box-Plot

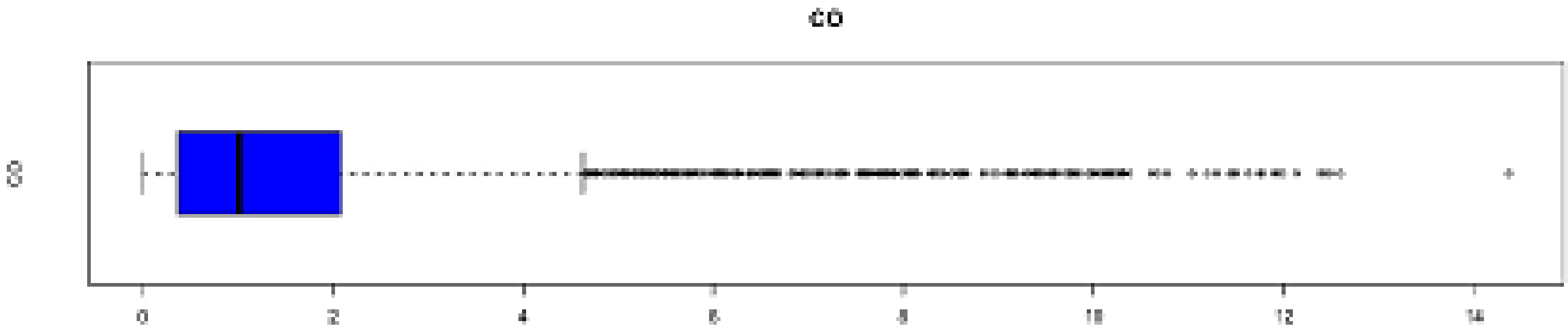


Figure 4.5: Picture Showing Distribution of Pollutants using histogram

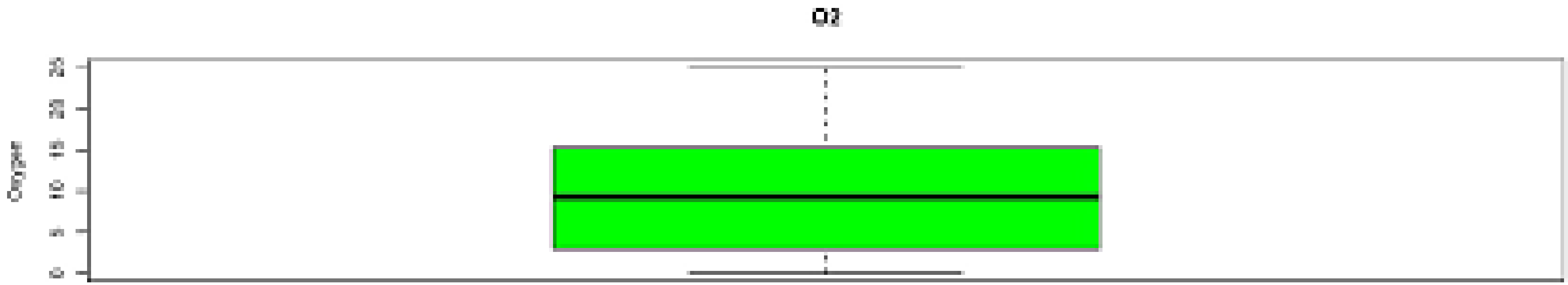


Figure 4.6: Picture Showing Distribution of Pollutants using histogram

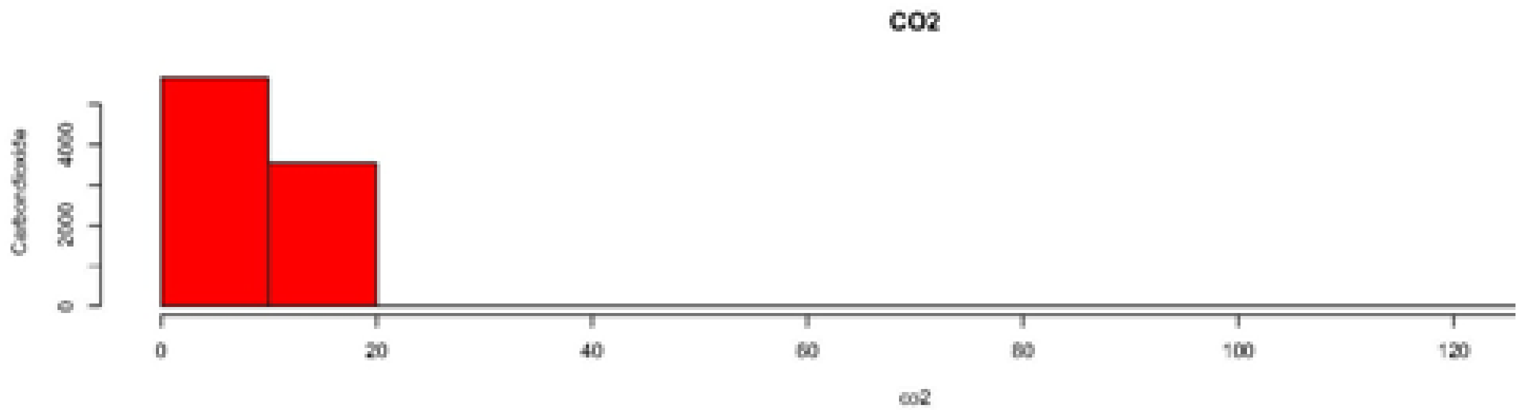


Figure 4.7: Picture Showing Distribution of Pollutants using histogram

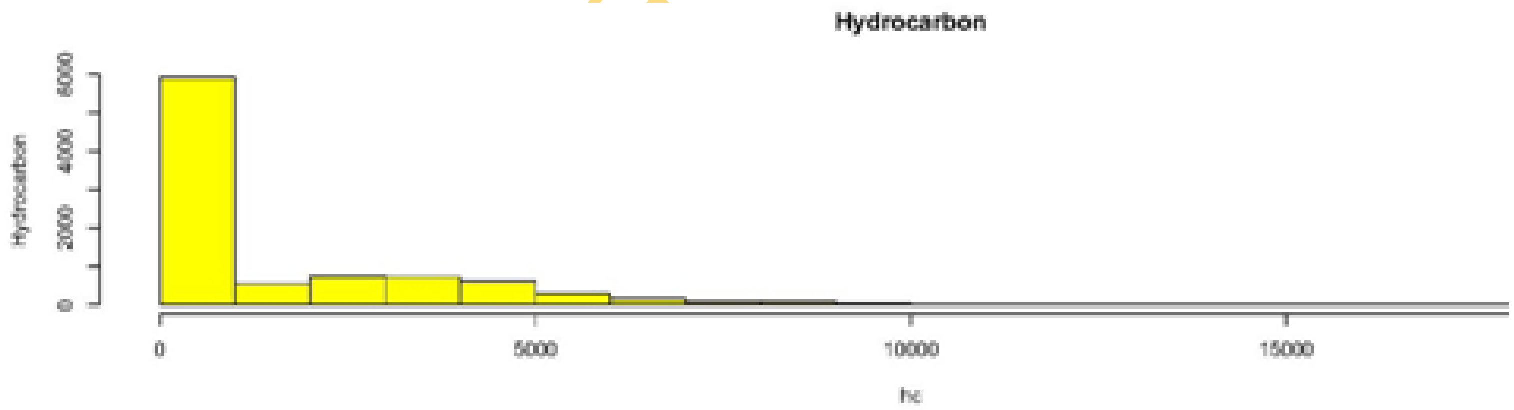


Figure 4.8: Picture Showing Distribution of Pollutants using histogram

4.9.2 The Model Estimation

The Geoadditive Bayesian Model is:

$$exp[g_0(t) + \frac{1}{\sqrt{2\Pi\tau^2}} e^{\frac{-1}{2}\beta_j^2} \Sigma_{j=1}^{p} z_{ij} + \frac{1}{\sqrt{2\Pi\tau^2}} e^{\frac{-1}{2\tau^2}(\beta_j)^2}$$

$$+ \frac{1}{\sqrt{2\Pi\tau_j^2}} e^{\frac{-1}{2^2}(\beta^{spat})^2} + \frac{1}{\sqrt{2\Pi\tau_j^2}} e^{\frac{-1}{2^2}(\beta_{gi})^2}].exp \int_0^{\infty} exp(g_0(u) + \Sigma_{i=1}^{p} g_j(u) z_{ij}) du$$

where z_{ij},g_j, β_{spat} and β_j were non-linear time varying effect, linear time varying effect, spatial effect, and random component, respectively.
The MCMC analysis is implemented with a burn-in period of 2000 iterations and iterations of 12000. Thinning factor of 10 was specified which forces BayesX to store only every 10th sampled parameter, leading to a random sample of length 1200 for every parameter in our work. The model diagnostic was based on the Deviance Information Criterion (DIC) given by DIC = $\overline{D}(\theta)$ + pD where $\overline{D}$ is the posterior mean of the deviance and pD is the effective number of parameter (which is similar, but not equal, to degrees of freedom) (see Spiegelhalter et al., 2002). The model with the lowest DIC is considered best. Sensitivity to the choice of priors was investigated by varying the priors for all the parameters.

The dependent variable is the Patient Outcome (rst). Results were obtained for all the effect without the spatial effect. The original model proposed by Hennerfeind et al. (2006) contains a spatial component which, in turn helps to capture the spatial structure present in the relationship. However, in our analysis, we have not used the spatial component as there were only four local government areas involved and they do not share a common boundary, so it was difficult to get the spatial information. Effect Coding was used to capture the spatial effect for the locations. In place of the spatial effect, we employed the effect coding in our model using three of them (namely Sagamu, Sango-Ota and Ijebu-Ode) with the fourth local government, Abeokuta, used as reference .

4.9.3 Fixed Linear Effect

A fixed effects model is a statistical model that represents the observed quantities in terms of explanatory variables that are treated as if the quantities were non-random. Fixed effects are parameters associated with an entire population or with certain repeated levels of experimental factors.

For the purpose of this work a diffuse prior was used for the fixed effect, while the effect of the continuous covariates were modelled by cubic p-splines with 20 equidistant knots and second order random walk penalty. Result for the fixed effect model are presented in Tables 4 to 7. The posterior mean, standard deviation and credible intervals are the comparison criteria.

The four models for the geoadditive Bayesian model are as follows:

(1) a = b =0.001, iterations = 12,000, burn-in = 2000, step = 10, n = 9211, prior = diffuse prior

(2) a = b = 0.00001, iterations = 12,000, burn-in = 2000, step =10, n = 9211, prior = diffuse prior

(3) a =1, b = 0.005, iterations = 12,000, burn-in = 2000, step = 10, n =9211, prior = diffuse prior

(4) a = 1, b = 0.00005, iterations = 12,000, burn-in = 2000, step = 10, n=9211, prior = diffuse prior

Table 1: Posterior means, median, SD and 95% credible interval using a=b=0.001

variable	mean	median	SD	SE	lower quantile	upper quantile
Const	0.3598	0.3594	0.0646	0.0007	0.2312	0.4951
nv	0.6751	0.6753	0.0048	0.00005	0.6657	0.6843
use	0.0058	0.0059	0.0041	0.00004	-0.0022	0.01389
ft	-0.0102	-0.0105	0.0078	0.00008	-0.0255	0.0049
diag	0.0005	0.0005	0.0009	0.000009	-0.0014	0.0023
car	0.0132	0.0132	0.0094	0.00010	-0.0055	0.0324
bus	-0.0118	-0.0118	0.0057	0.00006	-0.00233	-0.000084
Abeokuta	0 (reference)	-	-	-	-	-
Ijeb	-0.0148	-0.0145	0.0072	0.00008	-0.028	-0.00081
sang	-0.0020	-0.0019	0.0059	0.00006	-0.0135	0.0097
sagm	0.0036	0.0036	0.0053	0.00006	-0.0065	0.0135

Table 1 gives results for the fixed effect for all the variables. The output gives posterior means along with their standard deviation and 95% credible intervals. From the table, we can observe that the number of visit (nv),vehicle use (use),diagnosis (diag) and vehicle type(.i.e car) jointly determine the health status of the child. Out of these, the effect of number of visits (nv) is statistically significant which implies that the children tend to visit the hospital more frequently with exposure to emitted pollutant and depreciation in their health status. This suggests that frequency of visit to the hospital is an indicator of health effect of vehicular emission on an individual.

The table also shows that the effects of emission from cars are more likely to affect the individual more than buses, although the effect is not statistically significant. This is because cars are more frequently used for commercial purpose in the state than bus. Result for location of individual shows that, compared with individuals who lived in Abeokuta, individual who resided in Sagamu were more likely to be affected by emitted pollutants while those from Sango and Ijebu Ode were less likely. However, none of these is statistically significant. Keeping in view the fact that bigger towns tend to have more relatively new vehicles, it is not conducive for better health status of individuals. But then, we may interpret that there seems to be a simultaneous relationship with larger town,vehicular emission and effect on the children living in the locality.

Table 2: Posterior means, median, SD and 95% credible interval using a=b=0.00001

variable	**mean**	**median**	**SD**	**SE**	**lower quantile**	**upper quantile**
Const	0.3656	0.3650	0.0433	0.0005	0.2748	0.4436
nv	0.6750	0.6748	0.0049	0.00005	0.6653	0.6849
use	0.0057	0.0056	0.0040	0.00004	-0.0022	0.0138
ft	-0.0088	-0.0090	0.0074	0.00008	-0.0239	0.0059
diag	0.0005	0.0005	0.00097	0.00001	-0.0014	0.0023
car	0.0.0124	0.0.0128	0.0089	0.00009	-0.0055	0.0295
bus	-0.0112	-0.0111	0.0054	0.00006	-0.0216	-0.00021
ijeb	-0.0150	-0.0151	0.0069	0.00007	-0.0282	-0.0020
sang	-0.0017	-0.0016	0.0060	0.00006	-0.0131	0.0099
sagm	0.0038	0.0038	0.0054	0.00006	-0.0069	0.0145

Table 2 provides the results of the model for hyper prior with a=b=0.00001 reveals that the number of times the child visit the hospital, what the vehicle is used for, the diagnosis made, and vehicle-type contributes to patients outcome. The table shows similar results with table 2 except that the vehicle used (in which both the effect of using car and bus) are significant even though using a bus has negative relationship with health effect of vehicular emission on individual. Also, fuel type(ft), though not statistically significant, has a negative relationship with health effect of vehicular emission on the individual.

Table 3: Posterior means, median, SD and 95% credible interval using a =1,b = 0.005

variable	**mean**	**median**	**SD**	**SE**	**lower quantile**	**upper quantile**
Const	0.3610	0.3624	0.05820	0.0006	0.2487	0.4783
nv	0.6749	0.6748	0.0050	0.00005	0.6655	0.6849
use	0.0057	0.0057	0.0042	0.00004	-0.0026	0.0137
ft	-0.0102	-0.0098	0.0076	0.00008	-0.0257	0.0042
diag	0.0006	0.0006	0.0010	0.00001	-0.0014	0.0024
car	0.0130	0.0130	0.0088	0.00009	-0.0039	0.0309
bus	-0.0117	-0.0116	0.0054	0.00006	-0.0223	-0.0007
Abeokuta	0(reference)	-	-	-	-	-
ijeb	-0.01462	-0.0143	0.0071	0.00007	-0.0283	-0.0010
sang	-0.0022	-0.0021	0.0057	0.00006	-0.0129	0.0092
sagm	0.0038	0.0037	0.0050	0.00005	-0.0056	0.0141

Table 3 also reveals that the number of time(s) the child visits the hospital, what the vehicle is used for, the diagnosis made, and vehicle type contributes to patients outcome.

Tables 3 and 4 show a similar result with number of visits (nv), vehicle use (use), fuel type (ft), diagnosis (diag), type of vehicle (i.e car and bus) jointly determine the health effect of vehicular emission on children. Out of these, the effect of number of visits is statistically significant with number of visits having the largest value,suggesting that frequency of visit to the hospital is an indicator of health effect of vehicular emission on children.

Table 4: Posterior means, median, SD and 95% credible interval using a=1,b=0.00005

variable	mean	median	SD	SE	lower quantile	upper quantile
Const	0.3673	0.3686	0.0348	0.0004	0.2966	0.4349
nv	0.6752	0.6751	0.0049	0.00005	0.6661	0.6850
use	0.0059	0.0060	0.0040	0.00004	-0.0019	-0.0136
ft	-0.0079	-0.0080	0.0074	0.00008	-0.0219	0.0070
diag	0.0006	0.0007	0.0009	0.000009	-0.0011	0.0024
car	0.0131	0.0130	0.0089	0.00009	-0.0039	0.0308
bus	-0.0116	-0.0116	0.0056	0.00006	-0.0220	-0.0003
Abeokuta	0	-	-	-	-	-
ijeb	-0.0132	-0.0135	0.0073	0.00008	-0.0272	0.0008
sang	-0.0026	-0.0026	0.0055	0.00006	-0.0126	0.0008
sagm	0.0033	0.0029	0.0052	0.00005	-0.0059	0.0136

The various tables show that the fixed effects of all variables chose different hyperprior value. The result gives posterior means and median along with their standard deviations,standard errors and credible intervals. From the table, we can observe that the number of visits, vehicle use and diagnosis jointly contribute to outcome, which is 'dead' or 'alive'; that is, they have significant effect on the health of the child. Also, the posterior standard deviation for all the factors considered are very small, indicating that the fixed effects are not heavily skewed/tailed, which implied that they are not largely spread out. This makes our knowledge about the parameter to be precise. The small standard error for all the factors considered is an indication that our posterior mean is a more accurate reflection of the actual population.

From the research, using different hyper prior for the nonlinear effect, the effect of age is significant only when the age of the vehicle is between 5 years and approximately 12 years. Also, the effects of carbonmonoxide, carbondioxide and hydrocarbon are on the increase. The effect of CO is increasing between 10100 and 14400ppm, while CO_2 is increasing between 0 and 15000ppm and HC is increasing between 4953 and 19812 ppm. Using a = 1, b= 0.00005, effect of age of vehicle is increasing between 7 years and 12 years, carbonmonoxide between 0 - 2500 ppm and 10800 and 14400, carbondioxide between 0 and 15 ppm and hydrocarbon between 5000 ppm and 19812 ppm.

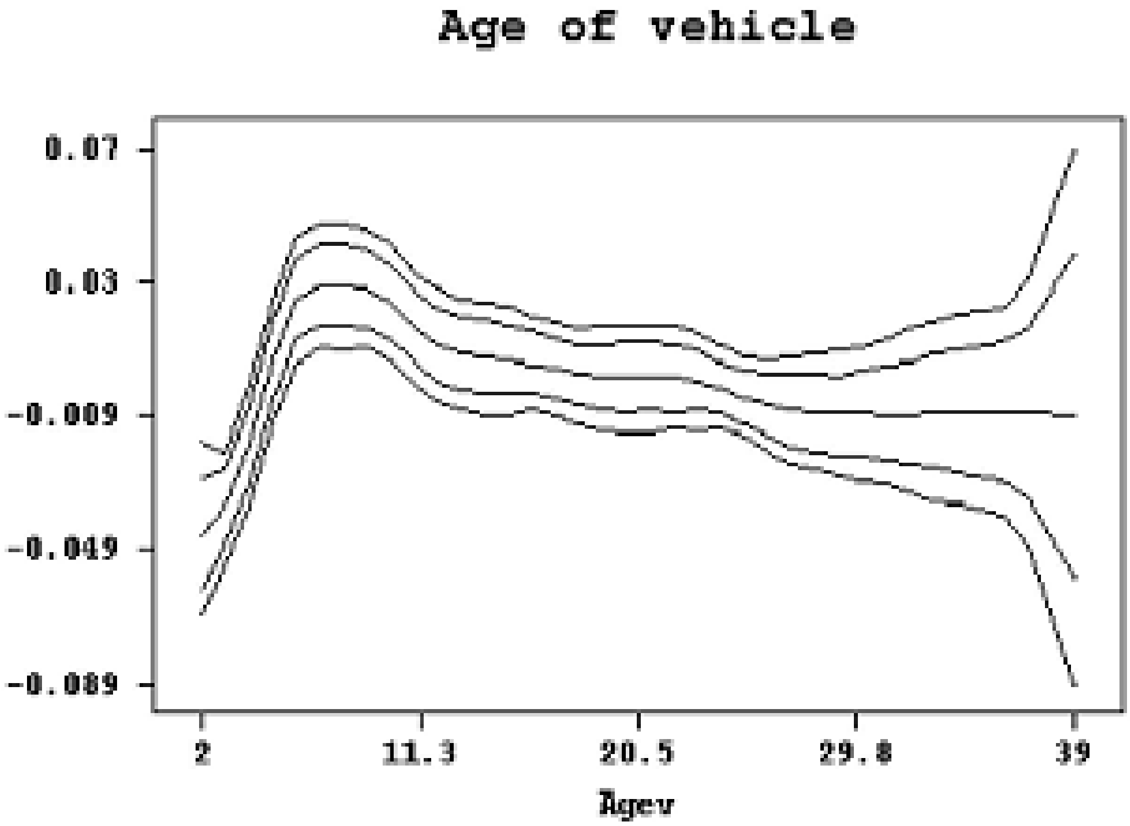


Figure 4.9: Posterior means of the nonlinear effect of age of vehicle and carbonmonoxide and its 95% C.I

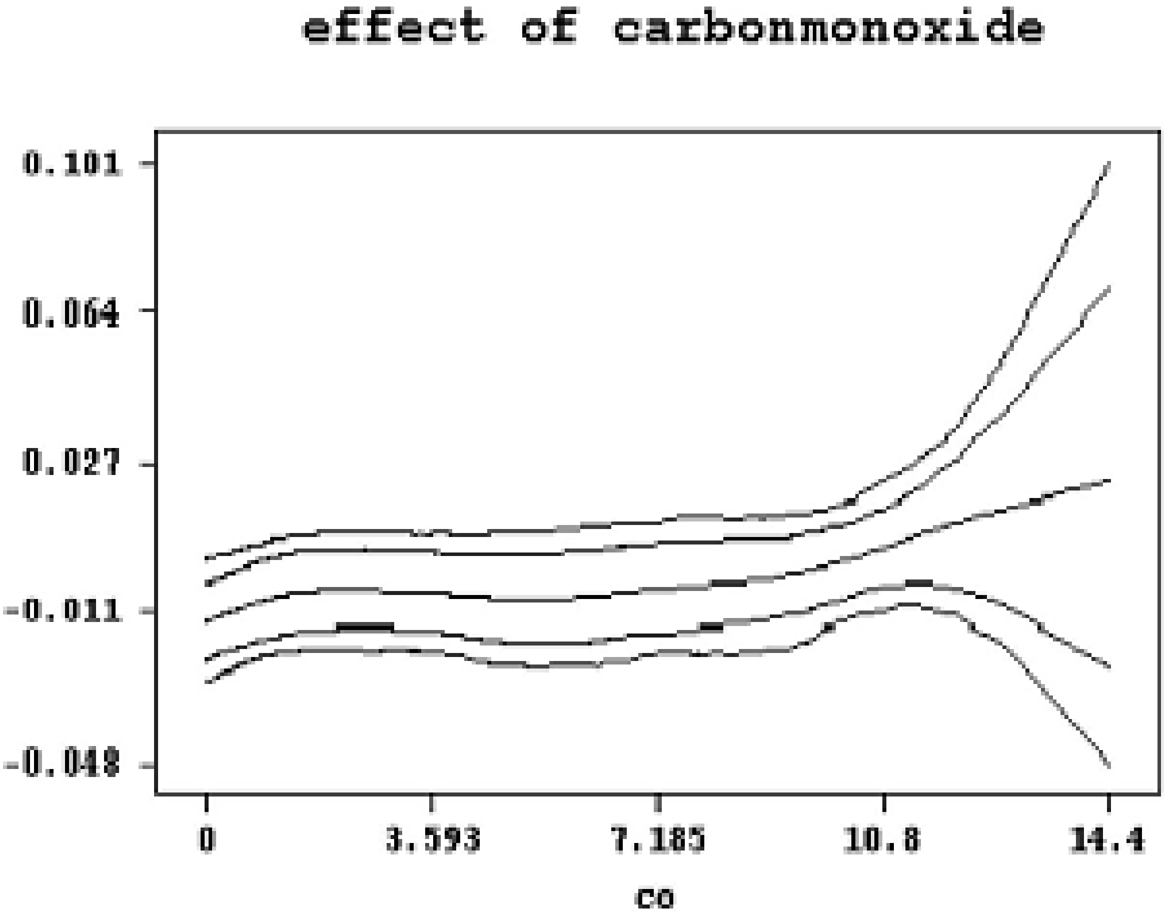


Figure 4.10: Posterior means of the nonlinear effect of age of vehicle and carbonmonoxide and its 95% C.I

The overall mean age of vehicle was 11.21 years (SD = 8.35 years) demonstrating a wide variation in age. For HC the overall mean was 1500.84 ppm and SD = 2072.28 ppm. For CO_2 the overall mean was 7590 and SD = 4650 ppm demonstrating a wide variation in distribution of CO_2, while the overall mean for CO was 1490ppm and SD was 1680ppm.
The ages of the vehicle reach their peak at 39 years for the three years and the flow of the ages of vehicles reached its minimum at 2 years. A high number of vehicles fall within the ages of 23 and 39 years, and the frequency reduces from 20 years to 16 years and picks a little at 17 years before dropping to 3 years. Rate of emission tends to increase with age of vehicle and lack of maintenance.

Fig 4.9 and 4.10 presents the nonlinear effect of age of vehicle and carbonmonoxide in model 4.5.1 with 95% credible interval. The nonlinear effect of age of vehicle and effect of carbonmonoxide was presented and the upper-left plot suggests that health status of the child in a given period is associated/significant with the age of the vehicle. The above results may be explained as follows: As the age of vehicle increases the tendency to discharge pollutant increases because of the depreciation in engine capability. This is because majority of the cars tested for are as a result of lack of maintenance of the vehicle and ageing of the vehicle which makes the engine to wear out. The right side also shows that the age of vehicle is significant to the health status of the child but it remains constant because majority of the old vehicles will be in the same state and as they get older they tend to be taken off the road. This explains why the effect of age of vehicle is increasing between 8 years and 12 years. The plot of the nonlinear effects shows that patients outcome increases with increasing age of vehicle where there was a sharp increase in the effect with the age of vehicle between 10 years and 12 years and that there was a steady and consistent increase in the health effect afterward. A significant increase in the nonlinear effect was observed for carbonmonoxide between 0 to 25000 ppm. Also, the plot suggests that carbonmonoxide is constant except between 10800 and 14400 ppm, which indicates incomplete combustion of fuel (i. e. there were vehicles that do not have enough oxygen to fuel ratio at the point of combustion, suggesting that most of the vehicles in these categories are old and are not properly

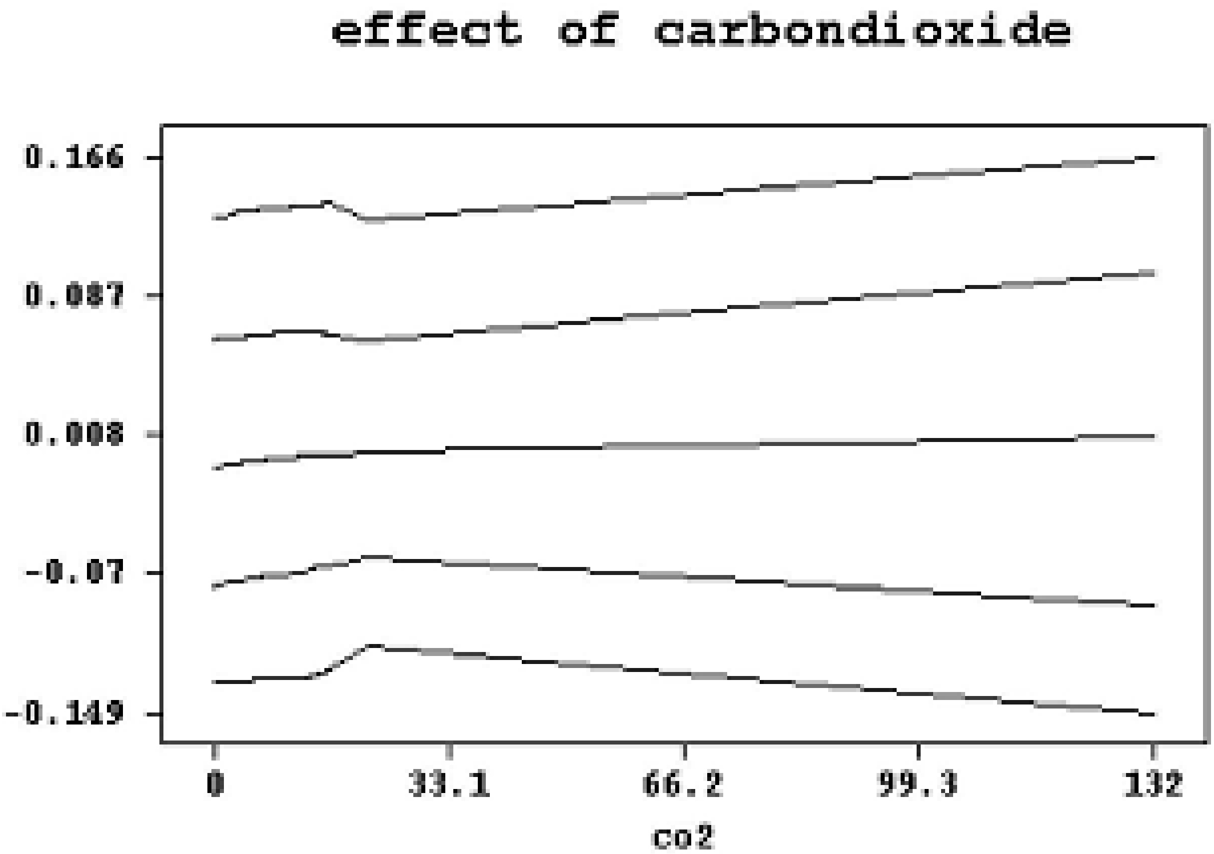


Figure 4.11: Nonlinear effect of carbondioxide using cubic p-spline

maintained).

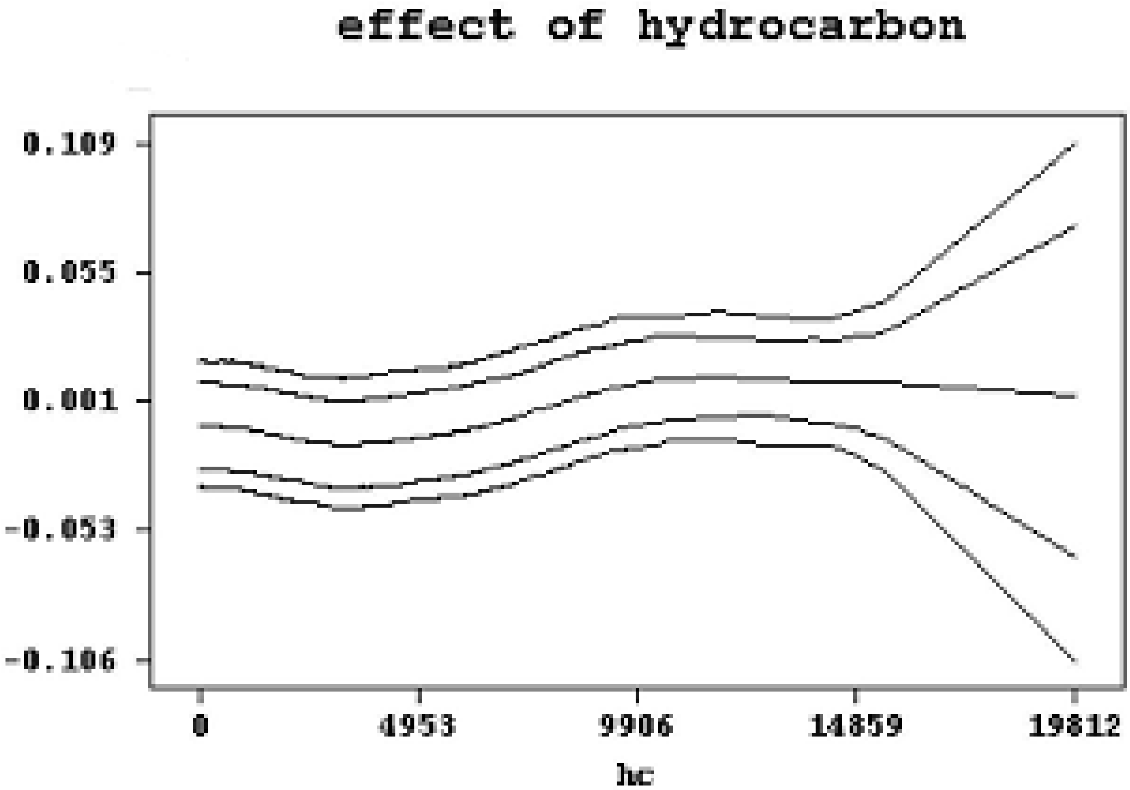


Figure 4.12: Nonlinear effect of Hydrocarbon using cubic p-spline

Fig 4.11 and 4.12 presents the nonlinear effect of carbondioxide and Hydrocarbon in model 4.5.1 with 95% credible interval. It was observed that patients outcome tends to be increasing with increase in Hydrocarbon where there was a significant increase between 4953 to 19812 ppm and there was a steady and consistent decline in the effect on patients outcome. This implies that patient outcome increased with every unit increase in emission rate of hydrocarbon up until 19812 ppm before stabilizing. In case of carbondioxide, the effect was almost the same at all levels. This indicated that either low or on the high side, it was having a negative effect on the health of children.

Table 5: Summary of estimated variance components

prior	IG,a=b=0.001	a=b=0.00001	a=1,b=0.005	a=1,b=0.00005
mean(Agev)	0.00119	0.00125	0.00067	0.0005
lower quantile(Agev)	0.00012	0.00011	0.00010	0.00007
upper quantile(Agev)	0.0053	0.0052	0.00312	0.0021
SD(AgeV)	0.00142	0.00149	0.00084	0.0006
mean(co)	0.00004	0.000005	0.00008	0.000003
lower quantile(co)	0.000008	0.0000002	0.00002	0.0000004
upper quantile(co)	0.00014	0.00003	0.0002	0.00001
SD(co)	0.00004	0.000010	0.00006	0.000003
mean(co2)	0.00026	0.00005	0.0002	0.00000
lower quantile(co2)	0.000011	0.00003	0.00042	0.0000004
upper quantile(co2)	0.00191	0.00042	0.0008	0.00002
SD(co2)	0.00053	0.00028	0.0004	0.00001
mean(hc)	0.00006	0.00001	0.0001	0.000005
lower quantile(hc)	0.0000001	0.0000002	0.00003	0.0000005
upper quantile(hc)	0.00023	0.00008	0.0003	0.00003
SD(hc)	0.00007	0.00003	0.00008	0.000009

The variance components have similar magnitude, which suggests that the unobservable heterogeneity between hospitals explains a portion of total variation similar to that explained by the observable covariates. These results suggest that the number of visits,type of vehicles used and diagnosis have positive relationship with health effect (i.e. health outcome of children in a given period of time). In contrast, the results suggest that the health effect of a child is negatively related/associated with fuel type and type of vehicle.nv (number of visit) has the largest value, suggesting that number of visits to the hospital has larger effect on the health outcome of the child.

Table 6 : Summary of Estimated Results for the Smoothing Parameter

prior	**a=b=0.001**	**a=b=0.00001**	**a=1,b=0.005**	**a=1,b=0.00005**
mean(Agev)	21.90(9.8011)	20.86(9.8810)	31.8317(9.1297)	44.8584(8.5392)
SD(Agev)	22.771	22.55	25.8089	39.2731
lower quantile(Agev)	1.856(14.5662)	2.5406(13.98)	3.1047(13.5965)	4.5503(12.8464)
upper quantile(Agev)	86.6382(7.4757)	135.512(6.81)	102.383(7.2206)	146.502(6.6932)
mean(co)	410.341(4.3259)	10928.8(1.9510)	170.941(5.2989)	8261.11(2.0860)
SD(co)	295.121	15589	88.5274	6774.73
lower quantile(co)	68.9344(6.4729)	316.45(4.5986)	41.2705(7.2090)	772.577(3.7165)
upper quantile(co)	1151.21(3.3713)	54554.2(1.3522)	375.134(4.1486)	25899.7(1.5898)
mean(co2)	192.521(2.1959)	6952.58(1.195)	100.419(2.4168)	6692.48(1.2011)
SD(co2)	220.556	12261.3	71.211	5995.29
lower quantile(co2)	5.0627(3.5006)	23.3937(2.9450)	12.5228(3.170225)	482.131(1.9085)
upper quantile(co2)	880.676(1.7233)	42381.9(1.0374)	286.152(2.0696)	22933.4(1.0673)
mean(hc)	313.717(4.0577)	5118.1(1.9904)	141.351(4.8913)	5068.07(1.9955)
SD(hc)	241.877	8747.25	82.6247	4905.44
min(hc)	41.7314(6.3852)	128.722(4.9968)	28.5389(6.8948)	437.345(3.7446)
max(hc)	952.691(3.0852)	32848.9(1.2793)	348.663(3.9560)	18606.61(1.4366)

The four nonlinear effects have significant effects on patient outcomes having smooths with degrees of freedom ranging from 1.04 to 14.6. Also, the smoothing parameters and variance components for the different choices of hyperprior shows that the results vary. This implies that the results are sensitive to the choice of hyperprior.

Table 7 : Deviance Information Criteria using different hyper prior

hyper prior	PD	DIC
a=b=0.001	36.0419	9248.5205
a=b=0.00001	31.6236	9238.6967
a=1,b=0.005	36.1438	9253.375
a=1,b=0.00005	28.4879	9250.292

Model diagnostics were based on the deviance information criterion (DIC) Spiegelhalter et al 2002 given by DIC $= \overline{D}(\theta)$ + pD,where $\overline{D}$ is the posterior mean of the deviance, measuring how well a model fits the data.

According to Spiegelhalter et al (2002) DIC can be used to compare models. We first note that DIC is subject to Monte Carlo sampling error, since it is a function of stochastic quantities generated under an MCMC sampling scheme. In any case, a check on the different hyperprior using DIC shows that the model is better improved with the choice of hyperprior a = b= 0.00001 with sample size of 9211 revealing the sensitivity of the model to the choice of hyperprior. The model with lowest value is the best, which implies that the choice of a = b =0.00001 has hyperprior performs better than when we choose hyperprior of a = b= 0.001 and a =1, b = 0.005 and a =1, b= 0.00005.

4.10 Model Performance

The performance of the model was checked using autocorrelation plot of 250 lag for the scale parameter, fixed effects and non linear effect and the result of the plot shows convergence.

Convergence imply good mixing of parameters. Mixing refers to the degree to which the Markov Chain explores the support of the posterior distribution. Good mixing implies that the estimated models are not with highly correlated variables.

MCMC often results in strong autocorrelation among samples that can result in imprecise posterior inference. To circumvent this, it is useful to thin the sample by only retaining every kth sample, where k is passed to the sampler via the thin argument.

The result of the autocorrelation plot for the scale parameters in Figure 4.13 using Bayesian p-spline with second order random walk penalty showed a good mixing properties with 12,000 iterations, 2000 burn-in, step =10 and convergence at lag 250. The results of the autocorrelation plot for the fixed effects, age of vehicle, carbonmonoxide and hydrocarbon were also shown in Figure 4.14 to 4.17. The autocorrelation plot for the fixed effects, age of vehicle, carbonmonoxide and hydrocarbon showed that the estimated models are not with highly correlated variables and that the Markov Chain was able to explore the support of the posterior distribution.

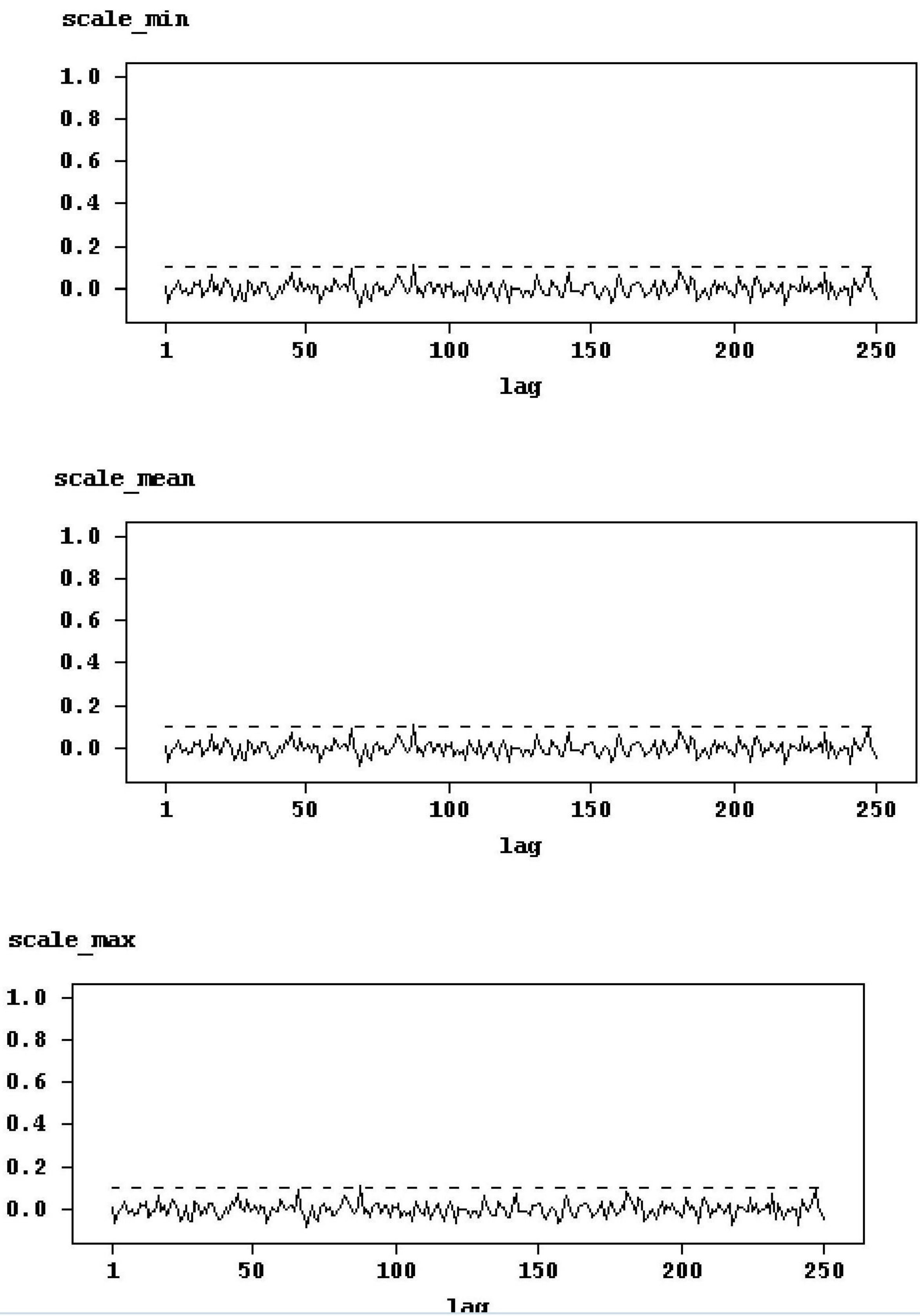


Figure 4.13: Plot of the Scale Parameter showing minimum, mean and maximum autocorrelation function for the scale parameter

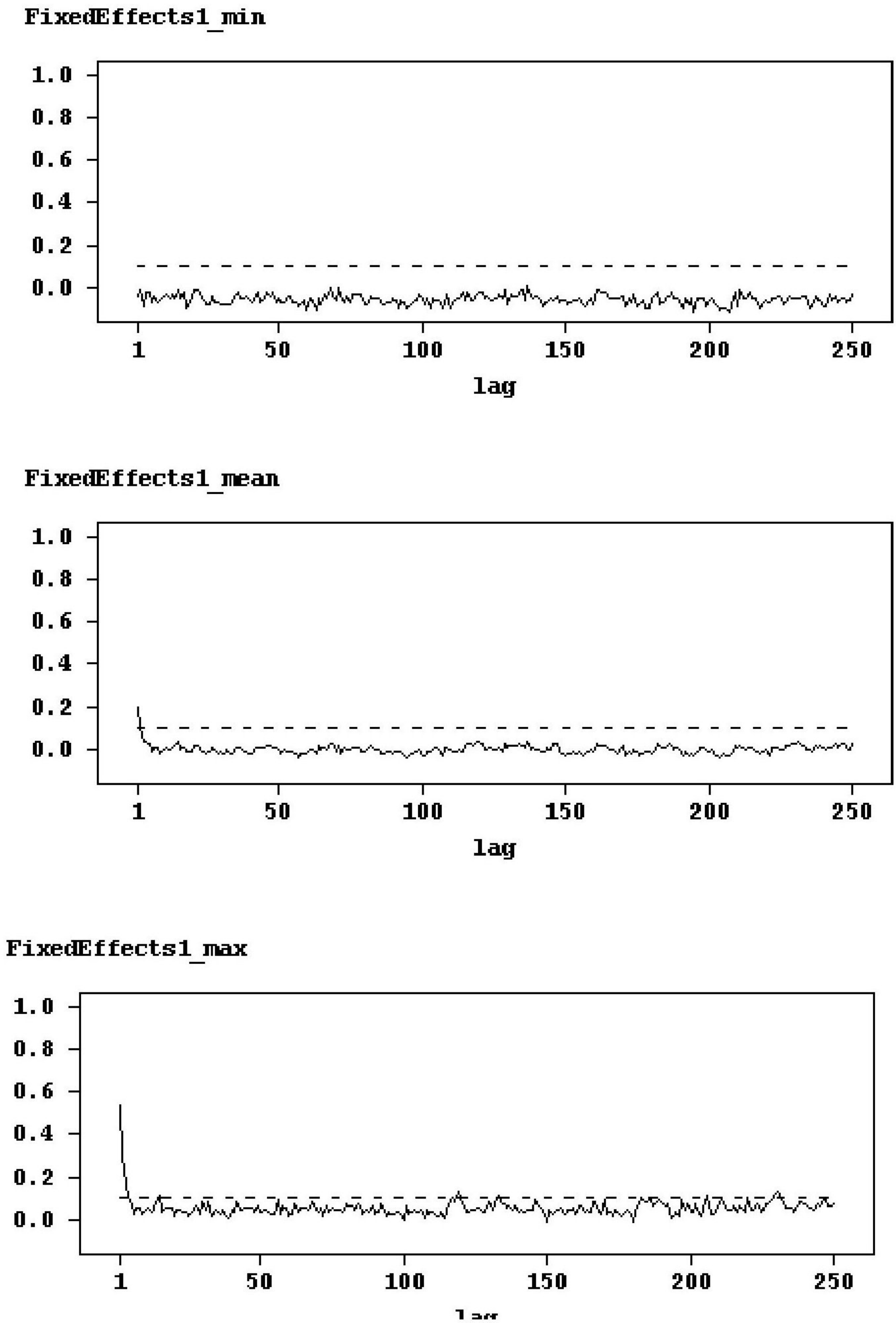


Figure 4.14: Plot of the Fixed Effect showing the minimum, mean and maximum autocorrelation function.

Fig 4.13 and 4.14 presents the nonlinear effect of carbon dioxide with 95% credible interval. The upper-left plot of fig 4.4 suggests that health status of the child is associated/significant with the level of carbondioxide being dispersed into the atmosphere. The above results may be explained as follows: as the vehicle consumes more fuel it increases the amount of carbon dioxide that is released into the atmosphere.

The right side also shows that carbondioxide is significant to health status of the child but it remains constant because majority of the vehicles are not efficient in their fuel consumption. This explains why most of the vehicles tested exceeded the standard that was set for the state. The result for the nonlinear effect of hydrocarbon with its 95% credible interval also shows that the health status of the child is associated/ significant with effect of hydrocarbon. This may be explained as follows: the poor maintenance coupled with the age of the vehicle gives rise to oil leakage into the combustion chamber of the engines through the leaking end of the valve shafts and loose piston rings. This explains why majority of the vehicles tested exceeded the standard set for the state, thus establishing the relationship between the effect which emission has on the health of the children living in the neighbourhood.

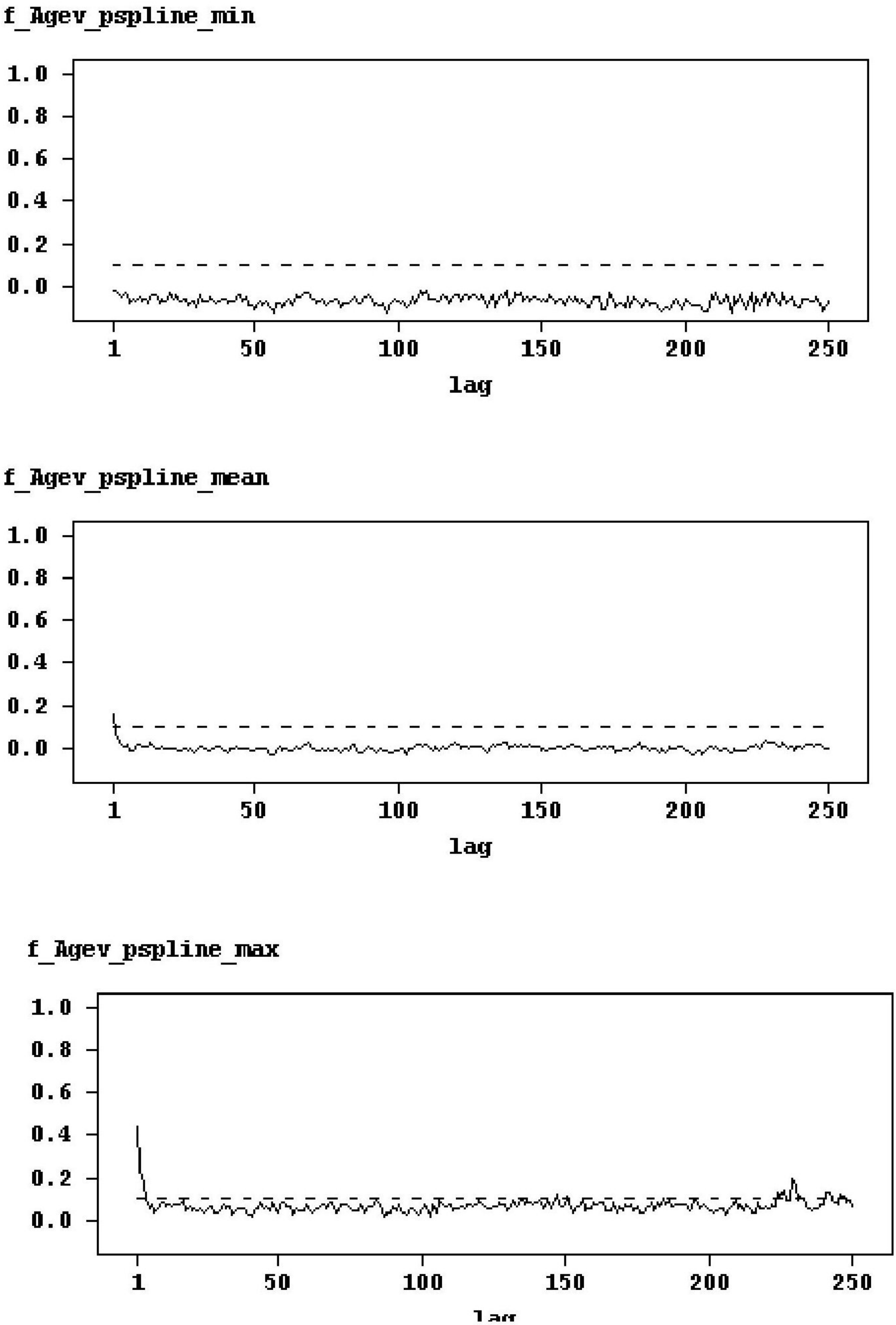


Figure 4.15: Plot of Age of Vehicle showing that a good behaviour of the chain was obtained

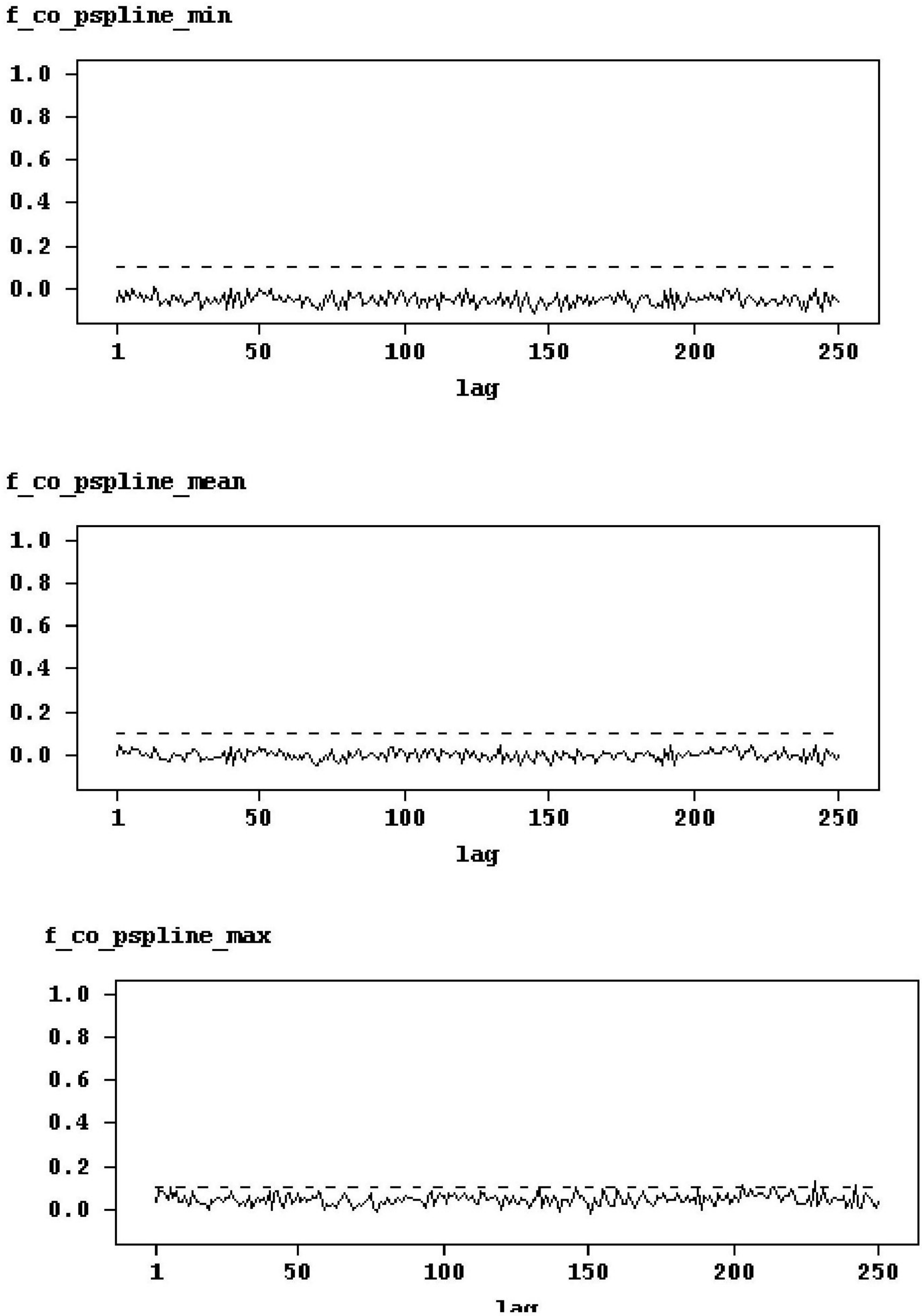


Figure 4.16: Plot of Carbonmonoxide showing convergence of parameters at 250 lag

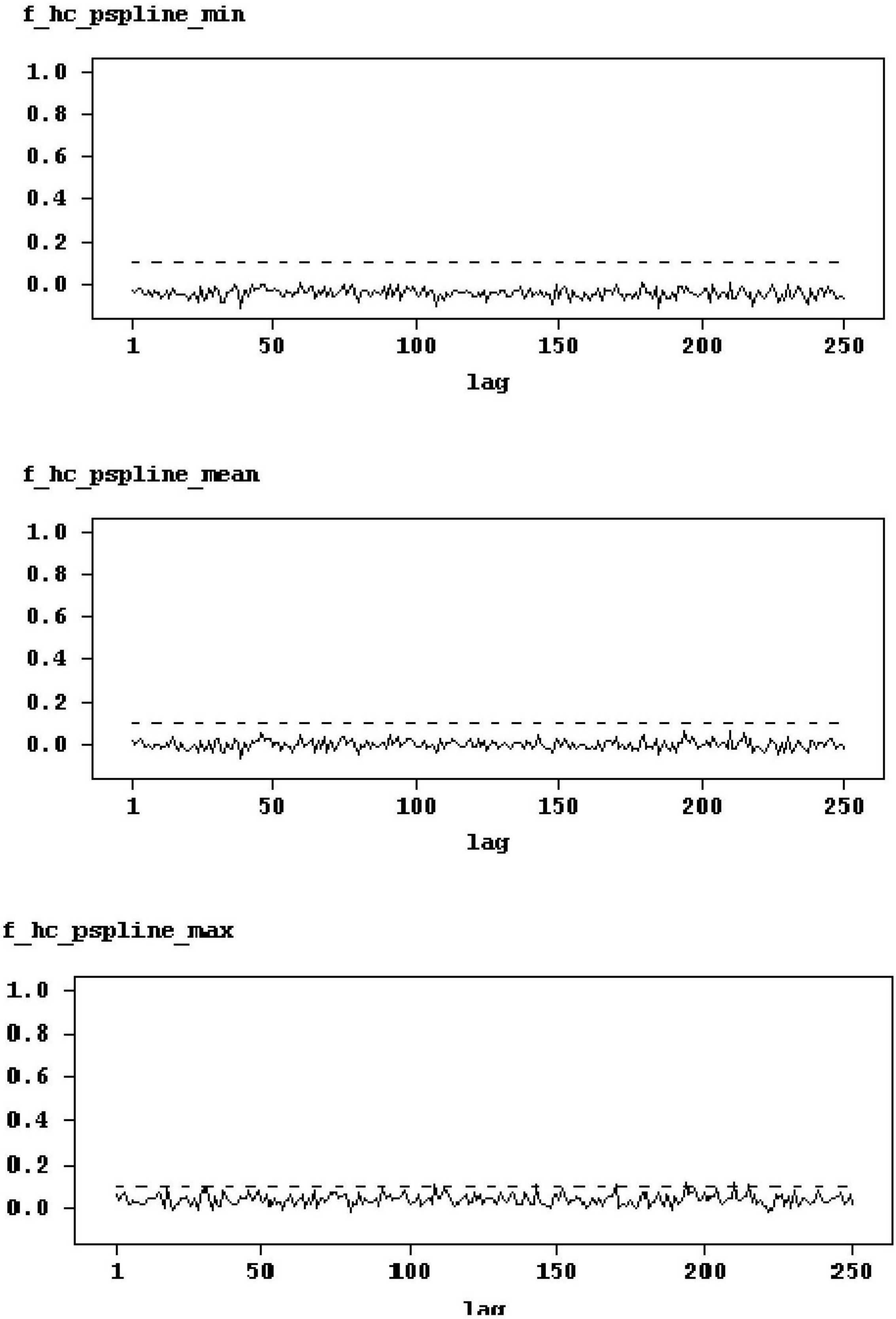


Figure 4.17: Plot of Hydrocarbon showing minimum, mean and maximum autocorrelation function.

Chapter 5

CONCLUSIONS AND RECOMMENDATIONS

5.1 Introduction

In this chapter, we draw appropriate conclusions from the results obtained, and indicate the contribution that has been made to knowledge. Appropriate recommendations are also offered. In section 5.2, we present the conclusions, in section 5.3 we present the contribution of the study. Section 5.4 present the recommendations.

5.2 Conclusion

The estimation of geoadditive model had been based on a large area as a result of insufficient information in small areas. In the geoadditive Bayesian model considered, a new random component was introduced and Bayesian approach was used in the application using effect coding for the spatial effect. The parameters considered were number of visits, age of vehicle, carbondioxide, carbonmonoxide, hydrocarbon, vehicle type, vehicle use, diagnosis, location and fuel type. The conclusions from the work are as follows.

The geoadditive Bayesian model constructed from the parent Geoadditive model was able to combine the linear and nonlinear components of the model and was able to assess the effects of vehicular emission on the children living in the environment. The component of the geoadditive Bayesian model was obtained using the concepts of *finite difference and partial derivatives.* The nonlinear time-varying effect was modeled using a matrix representation. Also, cubic spline and the random nature in the distribution of prime numbers were used. The prior was obtained, and the random component is distributed Gaussian $\beta_x \sim N(0, \tau_b^2)$. A conjugate prior was assigned on the variance component τ^2 which is an inverse gamma prior with hyper parameters a and b: $\tau^2 \sim IG(a, b)$. The posterior was obtained from the prior and the likelihood. An MCMC (Markov Chain Monte Carlo) simulation technique was used to draw samples from the posterior using BAYESX software. From these samples, quantities such as posterior mean, posterior standard deviation and quantiles (which, in turn, gives associated credible interval) was estimated.

Using the Effect Coding for the location with different hyperprior, we concluded that children living in Sagamu (posterior means = 0.0036, 0.0038, 0.0038 and 0.0033) were more affected by emitted pollutants than those from Sango-Ota (-0.0020, -0.0017, -0.0022 and -0.0026) and Ijebu-Ode (-0.0148, -0.0150, -0.0146 and -0.0132), especially when compared with individuals living in Abeokuta.

The geoadditive Bayesian model using different hyperprior revealed that vehicle use and vehicle type jointly determine the status of a child, implying that they were indicators of health effect of vehicular emission on individuals. These also determine the number of hospital visits and the diagnosis made on these children (e.g. asthma, pneumonia, bronchitis and cardiovascular diseases). The nonlinear effect of age of vehicle, carbonmonoxide, carbondioxide and hydrocarbon suggests that health status of the child in a given period is associated with the emitted pollutants.

The use of a geoadditive Bayesian model was found to be suitable for assessing the location effect on the health of the individual considered. Moreover, even with limited spatial information geoadditive model was still found to be applicable with good results using the Effect Coding with the Bayesian approach.

The posterior was not in a closed form. An MCMC (Markov Chain Monte Carlo) simulation technique was used to draw samples from the posterior using BAYESX software.From these samples,quantities such as posterior mean,posterior standard deviation and quantiles (which, in turn,gives associated credible interval) was estimated.

5.3 Contributions of the Study

Geoadditive model has been applied in other areas using the longitude and latitude for the spatial effect. In this study, we have been able to introduce another method which is the use of effect coding for the spatial effect instead of the previous use of longitude and latitude. The geoadditive Bayesian model adopted from the parent geoadditive model is able to estimate small areas with insufficient spatial information using Bayesian approach. It also serves as an improvement on most of the shortcomings of other known regression models. With this method, we are able to determine the location effect of exhaust emission from vehicle on the individual considered.

The effect coding used in this study are particularly useful when small district areas with no common boundaries are considered.

Our approach has been found suitable in combining the linear and the nonlinear effect. Thus, it is more preferable to the previous methods used by past researchers. Also, it has the advantage of reduced experimental costs since it is not compulsory that a large number of areas be involved.

5.4 Recommendations

Exploratory analysis of the nonlinear effect shows that there are extreme values in the data which indicate that the children in the locality are really suffering from pollution. This is an area that can be explored in the future. Also, the research only used some locality in one state; it (the work) can be extended to cover the entire 36 states of Nigeria so as to explore the efficiency of Geoadditive Bayesian Regression Model using the spatial effect and compare it to the effect coding used in this work.

REFERENCES

Abam, F.I. and Unachukwu, G.O. 2009. Vehicular Emissions and Air Quality Standards in Nigeria. *Euro Journals Publishing, Inc.* 34, 550-560.

Abhishek Yerramalla 2007. Vehicular Emissions Models using Mobile 6.2 and Field Data. Msc Dissertation. *University of Texas at Arlington.*

Adebayo, S.B. and Fahrmeir, L. 2005. Analysing Child Mortality in Nigeria with Geoadditive Discrete-Time Survival Models. *Statistics in Medicine,* 24,709-728

Adebayo, S.B, Gayawan E, Ujuju C and Ankomah, A. 2013. Modelling Geographical Variations and Determinants of use of Modern Family Planning Methods Among Women of Reproductive age in Nigeria. *Journal of Biosoc.sciences* 45, 57-77.

Ahn, K. 1998. Microscopic fuel consumption and Emission modelling. Msc thesis *Virginia Polytechnic Institute and state University, Blacksburg, Virginia.*

Alderman, B.W, Baron A.E, Saviz Da. 1987. Maternal Exposure to Neighborhood Carbon Monoxide and Risk of Low Infant Birth Weight. *Public Health Reports,* 102: 410414.

An, F., Barth, M., Norbeck, J. and Ross, M. 1997. Development of a Comprehensive Modal Emissions Model. Operating under Hot - Stabilized Conditions, *Transp.Res.Rec.,* 1587, 73-84.

Anselin, L. 2001. Spatial Econometrics. In: Baltagi, B. (ed), A Companion to Theoretical Econometrics. *Oxford: Basil Blackwell,* pp. 310-330.

Anselin, L. and Florax, R. (eds) 2002. *Advances in Spatial Econometrics.* Berlin: *Springer-Verlag.*

Atienza N., Garcia-Heras J., Munoz-Pichardo, J.M. and Villa, R. 2008. An Application of Mixture Distributions in Modelization of Length of Hospital Stay. *Statistics in Medicine* 27, 1403-1410.

Augustin, N.H., Lanng,S., Musio, M. and von Wilpert, K. 2007. A Spatial Model for the Needle Losses of Pine-trees in the forest of Baden-Wurttemberg: An Application of Bayesian Structured Additive Regression. *Applied Statistics*, 56, 29-50.

Austrup, S.E. 2005. Survey of air pollution in Cotonou, Benin-air Monitoring and biomarkers *Science of Total Environment.* 358(1-1), 85-96

Bachman, J.G., Wadsworth, K.N., O'Mailey P.M., Johnston, L.D. and Schulenberg, J.S. 1997. Smoking, Drinking and Drug use in Young Adulthood: The Impacts of New Freedoms and New Responsibilities, *Lawrence Erlbaum, Mahwah, N.J*

Banerjee S.,Wall M.and Carlin, B.P. 2003. Frailty Modelling for Spatially Correlated Survival Data with Application to Infant Mortality in Minnesota.*Biostatistics* 4, 123-142.

Barth M., Younglove T., Wenze T., Scora G.,An F., Ross M., and Norbeck, J. 1997. Analysis of Modal Emissions from a Diverse in use Vehicle fleet, 76th Annual Transportation Research Board Meeting. *Washington D.C.*

Benson, P. 1989. CALINE 4 - A dispersion Model for Predicting Pollutant Concentrations Near Roadways. Report FHWA/CA/TL- 84/15. *Division of New Technology and Ralifornia Department of Transportation, Sacramento.*

Bishop G., Beaton S., Peterson J., Guenther P., McVey I., Zhang, Y. and Stedman, D. 1994. "Results of 1991 CO and HC Remote sensing in California", Proceedings of the 4th CRC On-Road Vehicle Emission Workshop in San Diego,CA. *Coordinating Research Council,Inc:Atlanta,GA,.*

Bobak, M. and Leon, D.A. 1992. Air Pollution and Infant Mortality in the Czech Republic, 1986-88. *Lancet* 340 (8826): 1010-4

Brezger and Lang 2003. Generalized Structured Additive Regression Based on Bayesian P-Splines. *Paper* (321).

Brezger A. and Lang S. 2006. Generalized Additive Regression Based on Bayesian P-Splines. *Computational Statistics and Data Analysis*, 50, 967-991.

Brown, H. and Prescott, R. 2006. *Applied Mixed Models in Medicine*, Second Edition. *John Wiley & Sons, Ltd.* NewYork

Brown, D.R, Wargo, J., Wargo, L. and Alderman, N. 2006. The Harmful Effects of Vehicle Exhaust. A Case for Policy Change. *Creative Advertising and Publishing Services, West Hartford.* 62 : 351-352

Brunekreef, B. 2005. Out of Africa, *Occupational and Environmental Medicine*

Buckeridge, D.L, Glazer, R., Harvey, B.J., Escobar, M., Amrheen,C., and Frank, J. 2002. Effect of Motor Vehicle Emissions on Respiratory Health in an Urban Area. *Environ Health Perspect* 110: 293-300.

Cadle, S.H., Belian, T.C., Black, K.V. and Carlock, M.A. 2006. Real-World Vehicle Emissions: A Summary of the 15th Coordinating Research Council On-Road Vehicle Emissions Workshop. *Being a presentation presented at Coordinating Research Council Workshop.*

CEC 1992. The state of the environment in the European Community. Overview, vol.3. *Commission of the Communities, Brussels and Belgium*

Carlock, M. 1992. Overview of Exhaust Emission Factor Models. Proc., Transportation Modeling Tips and Trip - Ups, *Air and Waste Management Association, Pittsburgh, Pa.*

Carlock, M. 1993. "An Analysis of High Emitting Vehicles in the On-Road Vehicle Fleet", California Air Resources Board. Overview.Vol.3, *Commission of the Communities,Brussels and Belgium.*

Chandler, R.E. and Scott, M.E. 2011. *Statistical Methods for Trend Detection and Analysis in the Environmental Sciences. Published by John Wiley & Sons,Ltd,* The Atrium, Southern Gate, Chichester, United Kingdom.

Chen, Q.X. and Ibrahim, J.G. (2006). Semiparametric Models for Missing Covariate and Response in Data in Regression Models. *Biometrics,*62, 177-184.

Chipman, H.A.,George, E.I and McCulloch, R.E. 1998. Bayesian CART Model Search , *Journal of the American Statistical Association* 93, 935-948.

Ciccone G, Forastiere F, Agabiti N, Biggeri A, Bisanti L, Chellini E, Corbo G, Dell'Orco V, Dalmasso P, Veante Tf, Galassi C, Piffer S, Renzoni E, Rusconi F, Sestini P, Viegi G, Sidria collaborative group 1998. Road traffic and adverse respiratory effects in children. *Occupational and Environmental Medicine,* 55: 771-778.

Collins, J.J., Kasap, H.S., Holland, W.W. 1971. Environmental Factors in Child Mortality in England and Wales. *American Journal of Epidemiology,* 93:10 -22.

Congdon, P. 2007. Bayesian Modelling Strategies for Spatially Varying Regression Coefficients. A Multivariate Perspective for Multiple Outcomes.*Computative Statistics and Data Analysis.* 51(5), 2586-2601.

Cressie, N. 1993. *Statistics for Spatial Data, John Wiley and Sons, New York.*

David, R., Wand, M.P and Carroll, R.J. 2008. *Semiparametric Regression,* Published by Cambridge University Press New York,USA.

Denison, D.G.T., Mallick, B.K and Smith, A.F.M. 1998a. Automatic Bayesian Curve Fitting. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 60 : 333-350.

Denison, D.G.T., Mallick, B.K. and Smith, A.F.M. 1998b. A Bayesian CART Algorithm, *Biometrika* 85, 363-377.

Denison, D.G.T., Mallick, B.K. and Smith, A.F.M. 1998c. Bayesian MARS. *Statist. comput.* 8: 337-346.

Dominici, F., McDermott, A., Zeger, S.L and Samet, J.M. 2002b. On Generalized Additive Models in Time Series Studies of Air Pollution and Health. *American Journal of Epidemiology,*156, 3, 1-11

Eilers, P.H.C., and Marx, B.D. 1996. Flexible Smoothing Using B-splines and Penalised Likelihood (with comments and rejoinder). *Statistical Science.* 11(2), 89 - 121

Elliott, P., Wakefield, J., Best, N. and Briggs, D.(eds) 2000. *Spatial Epidemiology: Methods and Applications.* Oxford: *Oxford University Press.*

Elspeth, C., Ferguson, Ravi Maheswaran and Mark Daly 2004. Road-Traffic Pollution and Asthma using Modelled Exposure Approach. *International Journal of Health Geography* V.3, 2004.

Erica Moen 2008. Vehicle Emissions and Health Impacts in Abuja, Nigeria. Bsc Dissertation.

Faboya, O.O. 1997. Industrial Pollution and Waste Management. pp 26-35 in Akinjide Osuntokun (ed), Dimensions of Environmental Problems in Nigeria, Ibadan *Davidson Press.*

Fahrmeir, L. and Lang, S. 2001. Bayesian Inference for Generalized Additive Mixed Models Based on Markov Random Field Priors *Journal of the Statistical Society, ser.c,* 50, 201-220.

Fahrmeir, L. and Echavarria, L.O. 2006. Structured Additive Regression for Overdispersed and Zero-inflated Count Data. *Applied Stochastic Models in Business and Industry,* 22, 351-369.

Fahrmeir, L., Kneib, T. and Lang, S. 2004. Penalized Structured Additive Regression for Space-time Data; a Bayesian Perspective. *Statistica Sinica,* 14, 715-745.

Fahrmeir, L. and Tutz, G. 2001. Multivariate Statistical Modelling Based on Generalized Linear Models. *Springer, New York.*

Faize, A. and Sturm, P. 2000. New Direction: Air Pollution and Road Traffic in Developing Countries. *Atmospheric Environment.* 34(27): 4745-4746

Fotheringham, A.S., Brunsdon, C., Charlton, M.E. 2002. *Geographically Weighted Regression, The Analysis of Spatially Varying Relationship.* Chichester, Wiley.

French, J.L. and Wand, M.P. 2004. Generalized Additive Models for Cancer mapping with incomplete covariates. *Biostatistics,* 5, 177-191.

Fu, L. 2001. Assessment of Vehicle Pollution in China. *Journal of the Air and Waste Management*: 51(5): 658-68

Ganguli, B. and Wand, M.P. 2006. Additive models for Geo-referenced Failure Time Data. *Statistics in Medicine,* 25, 2469-2482.

Goyal, S. 2006. Understanding urban vehicular pollution Problem vis-a-vis Ambient Air quality Case Study of Megacity (Delhi,India). *Environmental Monitoring and Assessment,* 119: 557-569

Greenhouse Gas Emissions from the U.S. "Transportation sector, 1990-2003" 2006. U.S. Environmental Protection Agency and office of Transportation and Air quality.

Groblicki, P.J. 1990. Presentation at the California Air Resources Board Public Meeting on the Emission Inventory Process. General Motors Research Laboratories. *Warren Mich*

Gryparis, A., Coull, B.A., Schwartz, J. and Suh, H.H. 2007. Semiparametric Latent Variable Regression Models for Spatiotemporal Modeling of mobile source particles in the greater Boston area. *Applied Statistics,* 56, 183-209.

Hastie, T.J. and Tibshirani, R.J. 1990. Generalized Addditive Models. *Chapman and Hall, New York.*

Hastie T.J. and Tibshirani, R.J. 1993. Varying Coefficient Models. *Journal of the Royal Statistical Society B,* 55, pp.757-796.

Haskew H. 1997. "Diurnal Emissions from in-use vehicles", Seventh CRC On-Road Vehicle Emissions Workshop, April 9-11

Hennerfeind, A., Brezger, A. and Fahmeir, L. 2006. Geoadditive Survival Models. *Journal of the American Statistical Association,* 101,1065-1075.

Ibrahim, B.G. 2009. Strategic Approach to Reducing Vehicle Emissions in Nigeria: Role of Fleet Operators. Being a lecture delivered at Safety Managers Training Programme at FRSC Academy, Jos.

Imada, T. and May, A.D. 1985. A priority Lane Simulation Model, Technical Document UCB - ITS - TD - 85 - 1, *California Department of Transportation, Berkeley.*

Iyoha, M.A. 2009. The Environmental effects of oil industry activities on the Nigerian Economy. A theoretical Analysis: Paper Presented at National Conference on the management of Nigeria's Petroleum Resources organized by the Department of Economics, *Delta State University.*

Jack, M.D., Ahlgren, W., Alves, J.F and Palen, E.J. 1995: "Remote and On-Board Instrumentation for Automotive Emissions Monitoring, "75th Annual Transportation Research Board Meeting, *Washington D.C.*

Gauderman, James.W., Edward, Avol, M.S., Frank Gilliland., Hita, Vora. M.S., Duncan Thomas, Kiros Berhane, Rob Mcconnell, Nino Kuenzli, Fred Lurmann , Edward Rapport, M.S., Helene Margolis, David Bates, and John Peters. 2004. The Effect of Air Pollution on Lung Development from 10 to 18 years of age. *N.Engl. Journal of Medicine* 351: 1057-1067

Jenny Linden, Sofia Thorsson, Ingegard Eliasson 2007. Carbonmonoxide in Ouagadougou, Burkina Faso - A Comparison Between Urban Background,Roadside and in - traffic measurements. 0049-6979, *GUP* 64535

Jerome, A. 2000. "Use of Economic instruments for Environmental Management in Nigeria". Paper Presented at Workshop on Environmental Management in Nigeria and Administration (NCEMA)

Kammann, E.E. and Wand, M.P. 2003 Geoadditive Models *Journal of the Royal Statistical Society, ser.C,* 52, 1-18.

Kammann, E.E. and Wand, M.P. 2001. Low-rank Radial Smoothing/ kriging via Restricted Maximum Likelihood. *Harvard University. Boston.* (Unpublished.)

Kim, J.J., Smorodinsky, S., Lipsett, M., Singer, B.C., Hodgson, A.T., Ostro, B. 2004. Traffic - related air pollution near busy roads, the East Bay children's Respiratory Health study. *Am J.Respir Crit care Med.* 170(5): 520-526.

Kneib, T. and Fahrmeir, L. 2004. Structured Additive Regression for Multi categorical space-time data. A Mixed Model Approach, SFB 386 Discussion paper 377, *University of Munich.*

Kneib, T. and Fahrmeir, L. 2006. Structured additive regression for categorical space-time data: A Mixed Model Approach. *Biometrics,* 62, 109-118.

Kneib, T. 2006. Mixed Model-based Inference in Geoadditive Regression for Interval-censored Survival Times. *Computational Statistics and Data Analysis,* 51, 777-792.

Kneib, T. and Fahrmeir. L. 2007. A mixed Model Approach for Geoadditive Hazard Regression. *Scandinavian Journal of Statistics* 34:207-228

Kneib.T., Hothorn,T. and Tutz, G. 2007. Variable Selection and Model Choice in Geoadditive Regression Models. Technical Report, Department of Statistics, *Ludwig-Maximilians. University of Munich.*

Kneib,T. and Fahrmeir, L. 2004. A Mixed Model Approach for Geoadditive Hazard Regression. SFB 386, Department of Statistics, *Ludwig-Maximilians, University of Munich.*

Kneib,T. and Fahrmeir, L. 2004. A Mixed Model approach for Structured Hazard Regression: SFB 386. Discussion Paper 400, *University of Munich.*

Koku, C.A., Osuntogun, B.A. 2007. Environmental-Impacts of Road Transport in South-Western States of Nigeria. *Journal of Applied Sciences* 7(16) : 2536-2360

Lang, S. and Brezger, A. 2004. Bayesian P-splines *Journal of computational and graphical statistics,* 13, 183-212.

Lang, S., Adebayo, S.B., Fahrmeir, L. and Steiner, W. J. 2003. Bayesian Geoadditive Seemingly Unrelated Regression. *Computational Statistics,* 18, 163-192.

Lawson, A. 2001 Statistical Methods in Spatial Epidemiology. New York : *Wiley.*

Lawson, A., Biggeri, A. and Williams, F. 1999. A Review of Modeling Approaches in Health Risk Assessment Around Putative Sources. In: Lawson. A., Biggery, A., Bohning, D., Lesaffre, E. et al. (eds), Disease Mapping and Risk assessment for public Health. New York:Wiley.

Lee, M.P. 2004. *Bayesian Statistics An Introduction* (Third Edition). *Published by Hodder Arnold,* (a member of the Hodder Headline Group, London.)

Lesage, J. 1999. The Theory and Practice of Spatial Econometrics. Web Manuscript (http://www.spatial-econometrics.com/). Department of Economics. University of Toledo.

Lin, X. and Zhang, D. 1999. Inference in Generalized Additive Mixed Models by using Smoothing Splines. *Journal of the Royal Statistical Society, Series B* 61, 381-400.

Loomis, D., Castillejos, M., Gold, D.R., McDonnell, W., Burja-Aburto, V.H. 1999. Air Pollution and Infant Mortality in Mexico City. *Epidemiology* Vol 10, pp 118-123.

Magbagbelola, N.O. 2001. The use of Economic Instruments for Industrial Pollution Abatement in Nigeria: Application to the Lagos Lagoon. Selected papers, Annual Conferences of the Nigerian Economic Society held in Port-Harcourt

Maynard, D., Coull, B.A., Gryparis, A., Schwartz, J. 2007. Mortality Risk associated with short - term Exposure to Traffic Particles. *Environ Health Perspect* 115(5) : A262.

Marx, B.D. and Eilers, P. 1998. Direct Generalized Additive Modeling with Penalized Likelihood *Computational Statistics and Data Analysis,* 28, 193-209.

McCullagh, P. and Nedler, J.A. 1989. Generalised Linear Models, Second Edition. *Chapman and Hall*/CRC Monographs on Statistics and Applied Probability

Ndoke, P. N., Jimoh, D.O. 2011. Impact of Traffic Emission on Air Quality in a Developing City of Nigeria. *Department of Civil Engineering, Federal University of Technology.* vol 8(4).

Neal, R.M. 1994. Bayesian Learning for Neural Networks. PhD. Thesis, Department of Computer Science, University of Toronto.

Neal, R.M. 1996. Bayesian Learning for Neural Networks, Lecture Notes in Statistics No. 118, New York: *Springer-Verlag.*

Ngianga-Bakwin Kandala, Chen Ji., Nigel Stallard., Saverio Stranges, Francesco P. Cappuc 2007. Spatial Analysis of Risk Factors for childhood Morbidity in Nigeria. *Academic Publishers*

Nkurunziza, H., Gebhardt, A. and Pilz, J. 2011. Geoadditive Modelling of Malaria in Burundi. *Biomed Central Journal* 10:234

Olajire, A.A., Azeez, L. and Oluyemi, E.A. 2011. Exposure to Hazardous Air Pollutants along Oba-Akran Road, Lagos Nigeria. *Chemosphere* 84, 1044 - 1051.

Oguntoke, O. and Yusuff, A.S. 2008. Air Pollution Arising from Vehicular Emissions and the Associated Human Health problems in Abeokuta Metropolis, Nigeria. *An International Journal of Agricultural Sciences, Environment and Technology series A.* 8(2): 119-132.

Ojo, O.O.S. and Awokola, O.S. 2012. Investigation of Air Pollution from Automobiles at Intersections on Some Selected Major Roads in Ogbomosho, South Western, Nigeria. *Journal of Mechanical and Civil Engineering.* 1(4), 31-35.

Ojolo S.J., Oke S.A., Dinrifo R.R and Eboda F.Y. 2007. A Survey on the Effects of Vehicle Emissions on Human Health in Nigeria. *Journal of Rural and Tropical Public Health* 6: 16-23.

Omoleke I.I. 2004. Management of Environmental Pollution in Ibadan, An African City: The Challenges of Health Hazard Facing Government and the People. *Journal of Human Ecology* 15(4): 265-275.

Osuntogun, B.A. and Koku, C.A. 2007. Environmental Impact of Urban Road Transportation in South-western states of Nigeria. *Journal of Applied Sciences*, 7(16): 2356-2360.

Oyana, J. Tonny and Rivers, A. Patrick 2005. Geographic Variations of Childhood Asthma Hospitalization and Outpatient Visits and Proximity to Ambient Pollution Sources at a US - Canada Border Crossing. *International Journal of Health Geographics* 4:14

Pedersen, K., Przychodzka,m., Civis,M. and Hinson, A.V. 2003. Environmental Impact Assessment of Petrol Usage. *Environmental Studies, Aarhus University, Finlandsgade* 12-14 DK-8200 AARHUS N, Denmark

Pershagen, G., Rylander, E., Norberg, S., Eriksson, M., Nordvall, S.L. 1995. Air Pollution involving Nitrogen dioxide Exposure and Wheezing Bronchitis in Children. *International Journal of Epidermiology,* 24: 1147-1153.

Penic, M.A. and Upchurch, J. 1992. TRANSYT-7F: Enchancement for Fuel Consumption, "Pollution Emissions and User Costs, *Transportation Research Record* 1360: 104-111.

Raaschou-Nelson, O. 1995. Traffic-related Air pollution: Exposure and

Health effects in Copenhagen Street Cleaners. *Archives of Environmental Health* 50(3): 207-13.

Raaschou-Nelson, O., Hertel, O., Thomson, B.L. and Olsen, J.H. 2001. Air pollution from Traffic at the Residence of Children with Cancer. *American Journal of Epidemiology,* 153: 433-443.

Rao, S.T., Sistla, G and Henry, R. 1992. Statistical Analysis of Trends in Urban Air Quality. *Journal of the Air Waste Management Association.* 42, 1204-1211

Rathi, A.K. and Santiago, A.J. 1989. "The New Netsim TRAF-NETSIM", TRB 68th Annual Meeting. *Washington D.C.*

Ritz, B., Yu, F. 1999. The effect of Ambient Carbonmonoxide on Low Birth Weight Among Children Born in Southern California Between 1989 and 1993. *Environmental Health Perspectives,*107: 17-25.

Sarah E.N. and Wand M.P. 2011. Generalized Extreme value Geoadditive Model Analysis via Variational Bayes. *Procedia Environmental Sciences* 3, 8-13.

Saville, S.B. 1993. Automotive Options and Quality Management in Developing Countries Industrial Environment. 16(1-2); 20, 32

Schwela, D. 2000. Air Pollution and Health in Urban Areas. Reviews on Environmental Health. 15(12): 13-24.

Seneca, J.J. and Tausig, M.K. 1994. Environmental Economics, *Engle Wood Cliffs, Prentice Hall.*

Shivey, T.S., Kohn, R. and Wood, S. 1999. Variable Selection and Function Estimation in Additive Nonparametric Regression using a Data-Based prior. *J.Amer. Statist.Assoc.* 94, 777-806.

Smith, M. and Kohn, R. 1996. Nonparametric Regression using Bayesian Variable Selection. *J.Econometric* 75: 317-343.

Smith, M. and Kohn, R. 1997. A Bayesian Approach to Nonparametric Bivariate Regression. *J. Amer. Statist. Assoc.* 92 : 1522-1535

Sorbe, N. 1995. Hughes Employee Vehicle Exhaust Remote Sensing and Emissions Evaluation Project, Report Prepared for the Mobile Source Air Pollution Reduction Review committee (MSRC) under the AB2766 Program,*Hughes Environmental Systems,Inc.*

Speckman, P.L and Sun, D. 2003. Fully Bayesian Spline Smoothing and Intrinsic Autoregressive Priors. *Biometrica* Trust 90(2), 289-302.

Suglia, S.F., Gryparis, A., Schwartz, J. and Wright, R.J. 2008. Association Between Traffic-Related Black Carbon Exposure and Lung function Among Urban women. *Environmental Health perspectives* 2008; 116 (10) : 1333-1337.

Sun, D., Tsutakawa, R.K., Zhuoqiong, H. 1998. Propriety of Posteriors with Improper Priors with Improper Priors in Hierarchical Linear Mixed Model. *Statistica Sinica* 11(2001), 77-95.

Sun, D., Tsutakawa, R.K. and Speckman, P.L. 1999. Posterior Distribution of Hierarchical Models Using CAR (1) distributions *Biometrika,* 86, 341-350.

Taiwo, O. 2005. The Case of Lagos-Air Quality Improvement Project. *An Article Published by Lagos Metropolitan Area Transport Authority.* $http://www.cleanairnet.org/ssa/1414/articles - 69320_Taiwo.pdf$

Thomas, J. Le Blanc , Michael, L. Scott., Brian Marsh, Evangelos Markatos, Cezary Dubnicki, Mark Crovella and Tim Becker 1992. The Psyche Parallel Operating System. *In: IEEE TC on Operating Systems Newsletter*, 6(1): 11-13.

UNFPA."State of world population 2007: unleashing the potential of urban growth". http://www.unfpa.org/swp/2007/english/introduction

United Nation. Prospect of World Urbanization 1998 (Population Study No.112) New York.

USEPA 1993. Guide to Environmental Issues, Doc.No 520/B-94-01. *United States Environmental Protection Agency, Washington, DC, USA*

USEPA 1994. National Air Quality and Emissions Trends Report, pp.2,6,46,52. *United States Environmental Protection Agency, Washington, DC,WA*

Van Aerde M. 1994. "Integration User's Guide", Department of Civil Engineering,Queen's University Kingston, Ontario, Canada.

van Vliet P., Knape M., de Hartog J., Janssen N., Harssema H and Bunekreef B. 1997. Motor Vehicle Exhaust and Chronic Respiratory Symptoms in Children Living Near Freeways. *Environmental Research,* 74: 122-132

Wand H, Whitaker C and Ramjee G. 2011. Geoadditive Models to Assess Spatial Variation of HIV Infections Among Women in Local Communities of Durban,South Africa. *International Journal of Health Geographics*, 10 : 28.

Wicks, D.A. and Liebermann, E.B. 1980. "Development and Testing of INTRAS", A Microscopic Freeway Simulation Model, Final Report Vol. 1, FHWA/RD - 80/106, *Federal Highway Administration,Washington D.C.*

William, M.B. 2003. Introduction to Bayesian Statistics, *Published by John Wiley & Sons, Inc.New Zealand.*

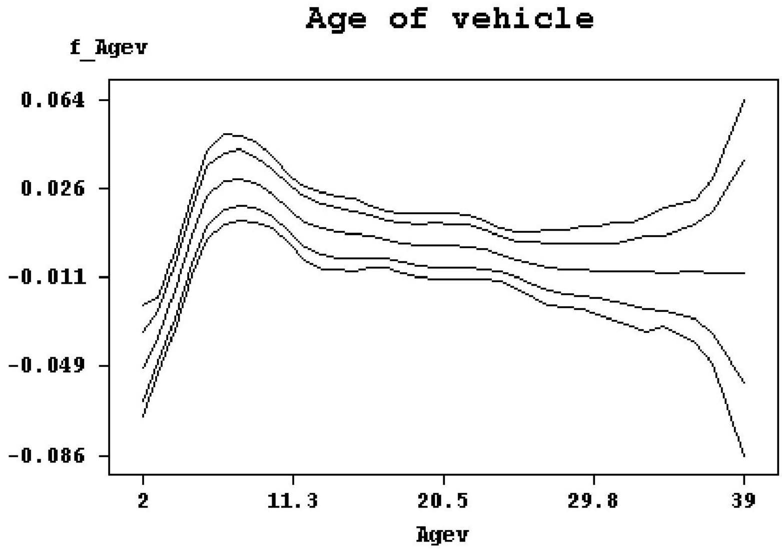
Wilpert, K. 2007. A Spatial Model for the Needle Losses of Pine-Trees in the Forest of Badan-Wurttemberg : An Application of Bayesian structured additive regression. *Applied Statistics*, 56, 29-50.

Wjst, M., Reitmeir, P., Wulff, A., Nicolai, T., von Leoffelholz-Colberg, E.F., von Mutius, E. 1993. Road Traffic and adverse effects on Respiratory Health in Children. *British Medical Journal* 307: 596-600.

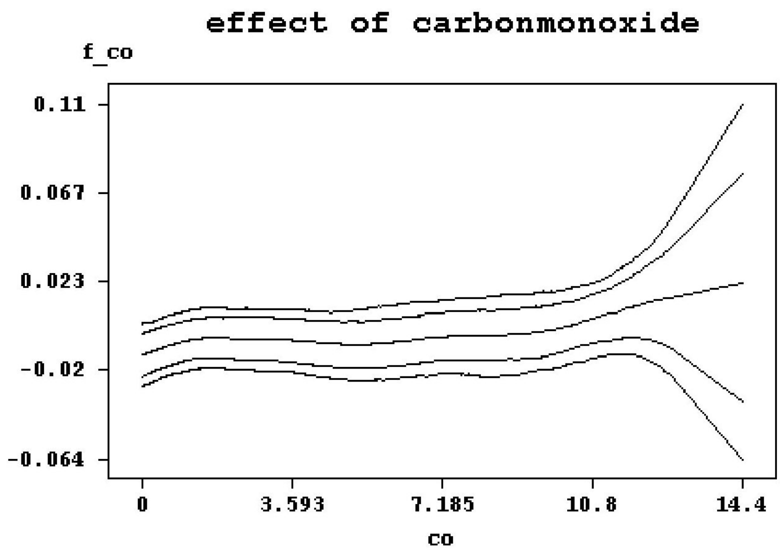
Wood, S.N. 2006. Generalized Additive Models, An Introduction with Chapman R. and Hall / CRC.

Zhao, Y., Staudenmayer, J., Coull, B.A. and Wand, M.P. 2006. General Design Bayesian Generalized Linear Mixed Models. *Statistical Science,* 21(1), 35-51.

APPENDIX

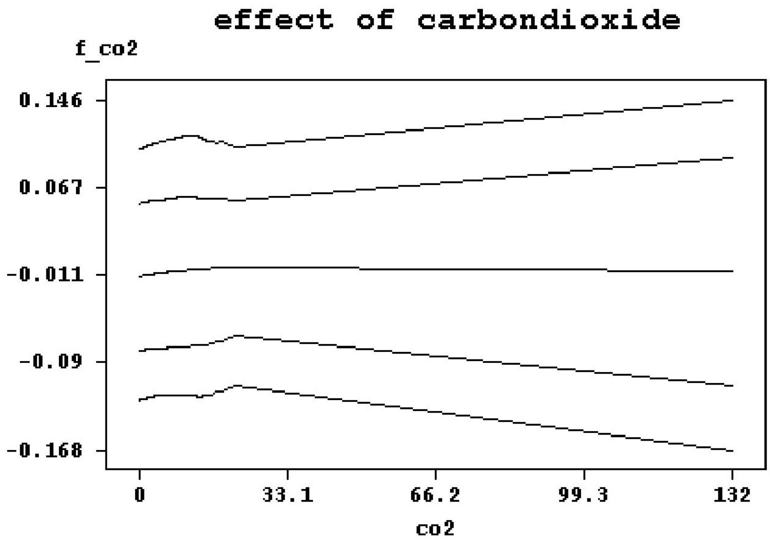


a=1,b=0.005

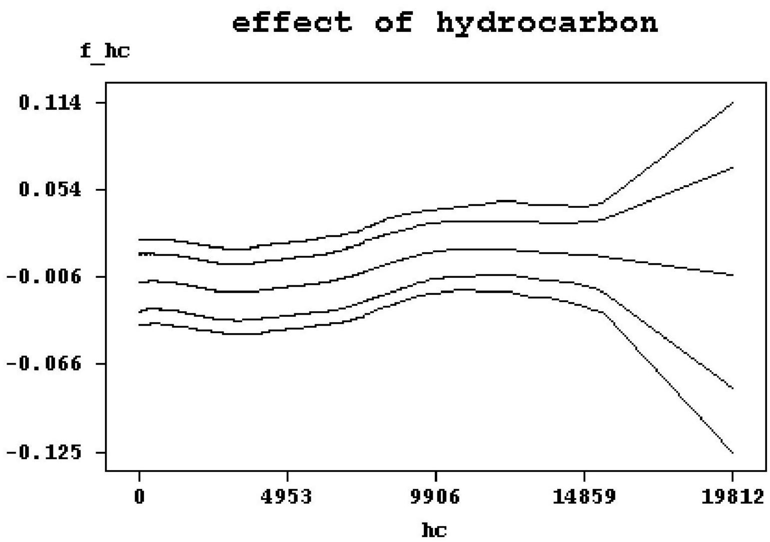


a=1,b=0.005

Figure 5.1: Effect of Age of vehicle and Effect of Carbon monoxide

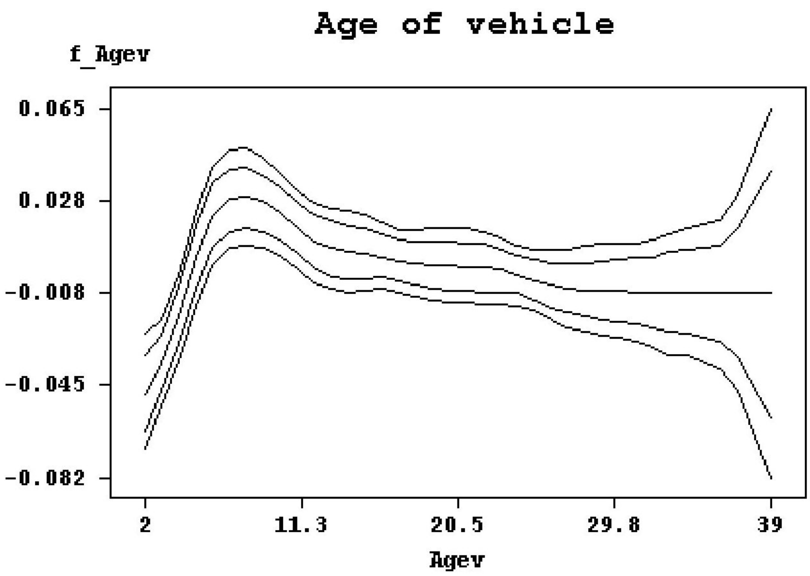


a=1,b=0.005

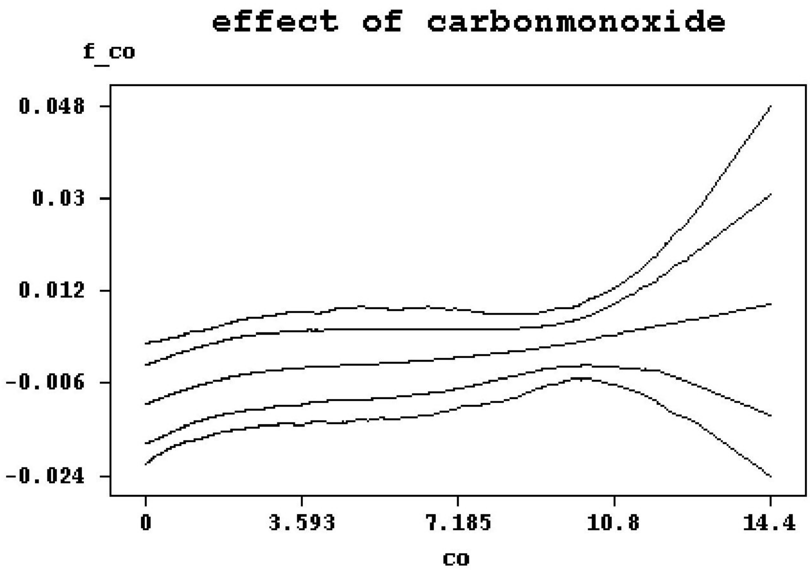


a=1,b=0.005

Figure 5.2: Effect of Carbondioxide and Effect of hydrocarbon

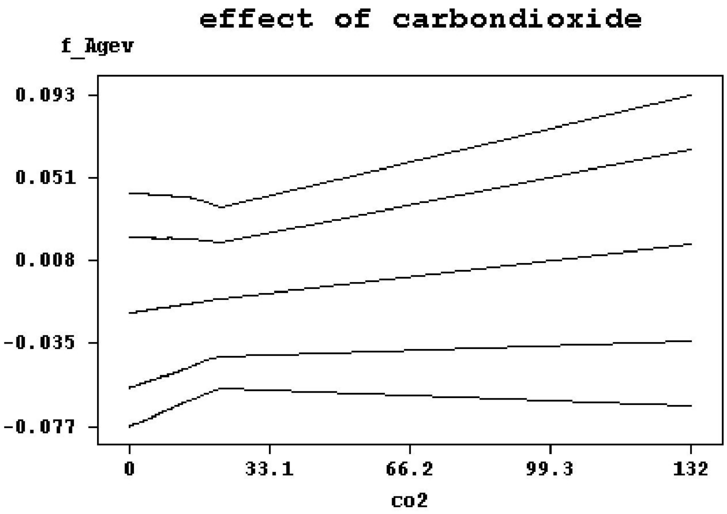


a=1,b=0.00005

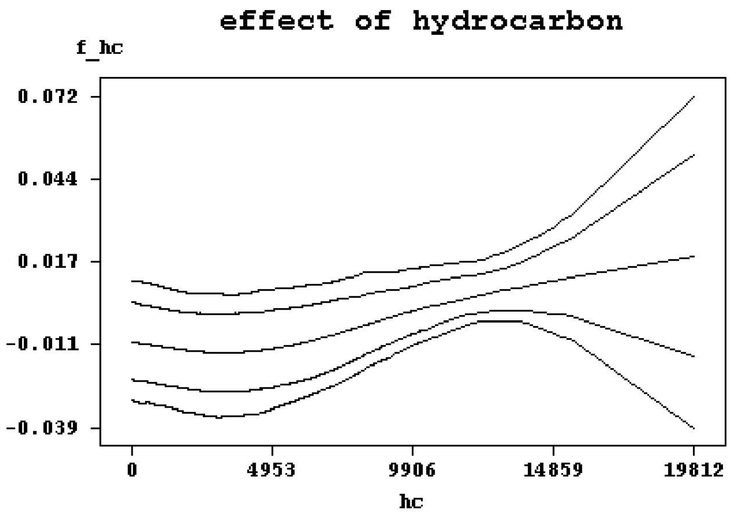


a=1,b=0.00005

Figure 5.3: Effect of Age of vehicle and Effect of Carbon monoxide

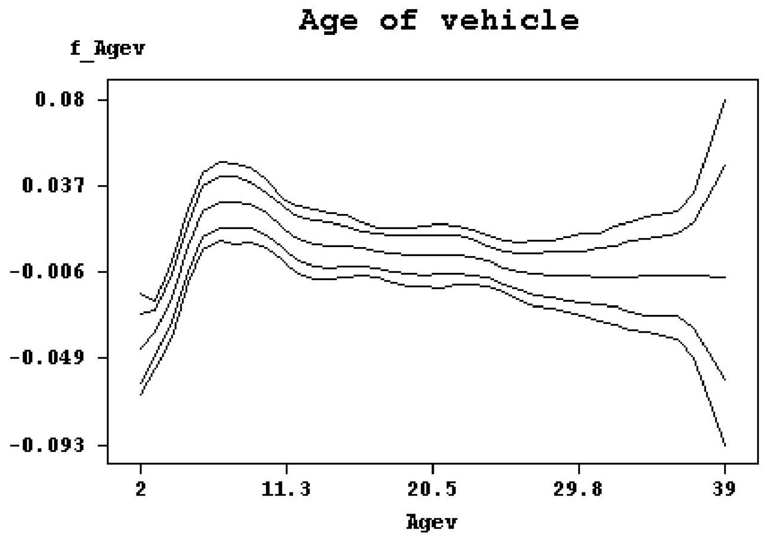


a=1,b=0.00005



a=1,b=0.00005

Figure 5.4: Effect of Carbondioxide and Effect of hydrocarbon



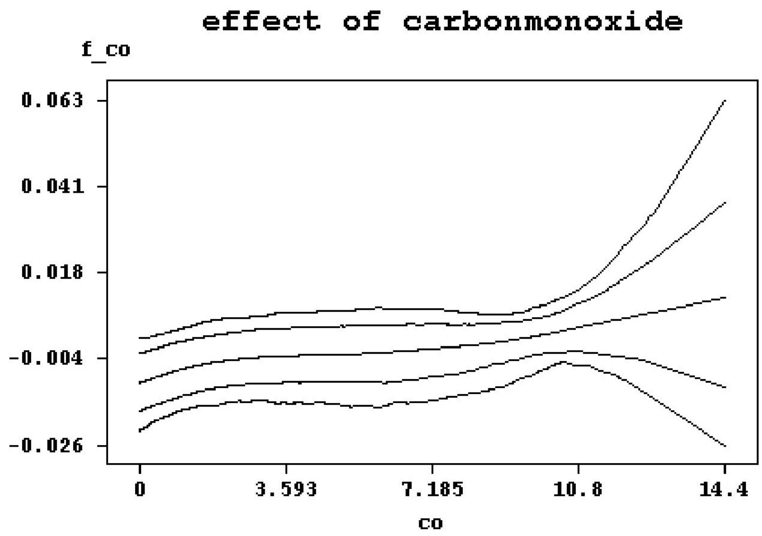
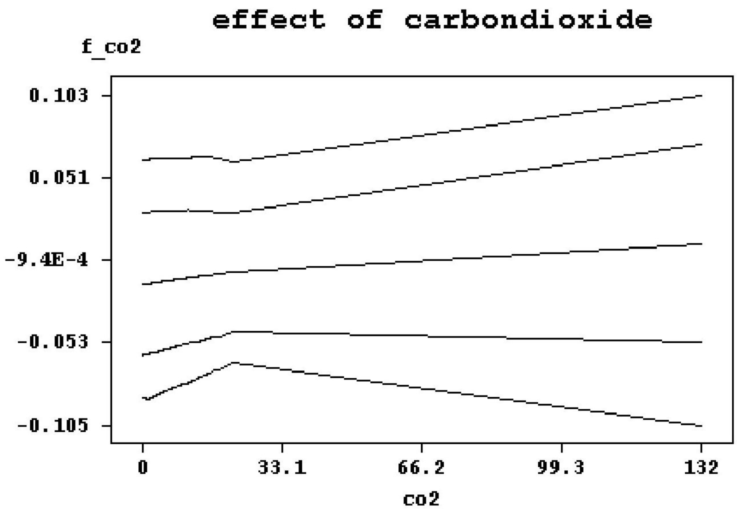
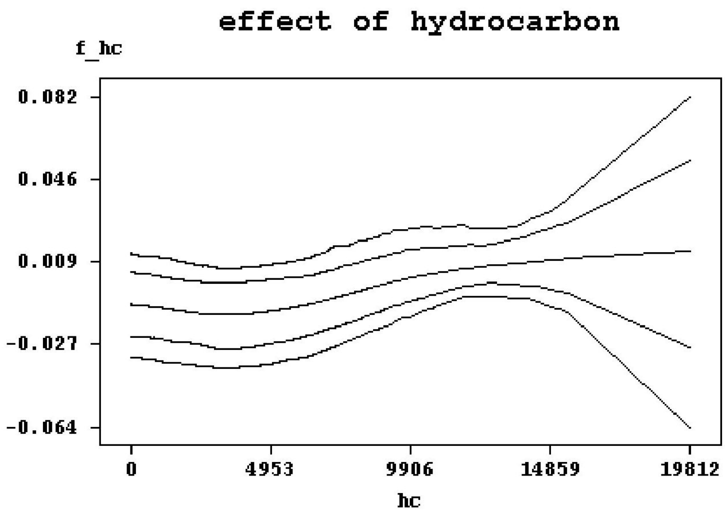


Figure 5.5: Effect of Age of vehicle and Effect of Carbon monoxide



a=b=0.00001



a=b=0.00001

Figure 5.6: Effect of Carbondioxide and Effect of hydrocarbon

Figure 5.7: Kane Gas Analyser

Figure 5.8: A Moving Vehicle