Journal of Statistical Modelling: Theory and Applications
Vol. 3, No. 3, 2022, pp. 15-29
Yazd University Press 2022

JSMTA

Research Paper

Introduction to non-parametric generalized additive models

Ahad Malekzadeh, Fatemeh Rajabi Naraki*
Department of Computer Science and Statistics, K.N. Toosi University of Technology, Tehran, Iran

Received: January 08, 2023/ Revised: May 22, 2023/ Accepted: May 23, 2023

Abstract: In order to investigate a series of data scenarios and determine the model governing the changes of a random variable over time, according to the variables affecting it, efficient methods have been developed in recent decades. One of these methods is the generalized additive model. By this modeling for data, it is possible to check the behavior of the non-linear data and even predict the future. In this article, we intend to express this method non-parametrically, in cases such as when the variable is independent, time series, or has a lag and implement the estimation of model parameters. Moreover, we will demonstrate the power and effectiveness of this method by presenting some examples.

Keywords: Distributed lag models; Generalized additive model; Generalized linear model; Penalized likelihood; Smooth function; Splines.
Mathematics Subject Classification (2010): 62J02, 62J12, 41A15.

1 Introduction

Whereas examining the information in a dataset requires a statistical analysis process, modeling is a proposal for the issue that a strategically understanding and interpretation can be made regarding the dataset. Statistical modeling is like a formal depiction of a theory. It is typically described as the mathematical relationship between random and non-random variables. In data analysis, when the relationships between variables follow a non-linear process, different approaches can be tried to model them. One approach that can be used to accommodate some non-linear effects into the model is using polynomial trends or adding a transformation to the response variable such

*Corresponding author: `Ftm23rjb@gmail.com`

as a logarithms transformation. But the non-linear relationship may not be easily recognizable and is avoided as this makes the model difficult for interpretation and specification. Segmented linear regression (SLR) can be introduced as the first proposed method to model this data. SLR, which is recognized as a piecewise regression or break-point model also, is a statistical approach in regression analysis where the independent variable is divided into some disjoint intervals and fits a distinct line segment for each interval so that each part of the time series can have different levels and trends. However, these models suppose a linear trend that may not be suitable in many positions.

Another proposed way for such models with non-linear relationship is generalized additive model (GAM) that can fit a smooth relationships between the variables without knowing the non-linear relationship in advance. GAMs were originally invented by Hastie and Tibshirani (1986). GAMs may be utilized in a variety of situations and exists in two modes, parametric and non-parametric. In parametric mode, it allows a wide range of distributions to be adopted for the response variable. The most prevalent distributions in GAM modeling are Normal, Gamma and Poisson distributions. However, due to the lack of a complete reference for the non-parametric case, in this paper we are going to focus on the non-parametric case of GAM. Furthermore, GAM can be useful when the variables are independent or the data are time series or even time series with lags.

In a comparison study, Pinilla and Negrin (2021) shown that SLR has poorer performance than GAM. Also, GAMs are widely utilized in various fields such as economics, biology, climate variables, clinical research, etc. Many researchers have worked on this topic to expand it and have published the results of their research on GAMs in the form of various books and articles, including James et al. (2013), Yee (2015), Wood (2017), etc.

As a look at other uses of GAM, the following can be mentioned, fisheries applications have been examined by Zhao et al. (2014), applications for air pollution and climate variability have been discussed by people like Dehghan et al. (2018) and Ravindra et al. (2019). In the context of investigation with high-capacity data such as genetic data analysis, a high-dimensional GAM approach is required, as discussed by Yang and Maiti (2020). Applications for economics discussed by Pinilla and Negrin (2021). GAM is used in deep learning with changes made by Chang et al. (2022) and etc.

Moreover, GAMs were introduced in time series epidemiology studies by Schwartz (1993) and have since been used and developed by many as a standard approach to solving the control variable problem (Schwartz, 1994b; Hoek et al., 1997; Kelsall et al., 1997; Bremner et al., 1999; Dominici et al., 2000). Also, in some research, a response variable may be measured repeatedly over time, and the values of a predictor may be time-dependent and also lagged. To check such data, two methods of GAMs and distributed lag models (DLM)s should be used. The DLM was first proposed by Allmon (1965) and then Corradi (1977) developed this method for greater flexibility by subjecting coefficients to non-parametric smoothing using spline functions. To utilize other covariate variable information and enter them into a DLM in a flexible way, the model is combined with the generalized additive model and Zanobetti et al. (2000) called the result "generalized additive DLM".

The remaining of the paper is organized as follows. In Section 2, we explain the range of application of non-parametric GAMs in independent, time series and distributed lagged data. Section 3 deals with estimation methods of model parameters. In Section 4, we implement a simulation to compare the models and eventually in Section 5, we provide some examples to show the efficiency of GAM method.

2 Some application of GAM

After the introduction of GAM method by Hastie and Tibshirani (1986), numerous research has been conducted on the application of this model in various data scenarios in successive years. In the following, we intend to examine the application of this model when the variables are independent, for time series data and time series data with lag.

2.1 GAM with independent variables

One of the broadest and most widely used methods for modeling independent observations is generalized linear models (GLM)s. At First, Nelder and Wedderburn (1972) introduced the GLM, which can be expressed as,

$$GLM : g(\mu_i) = \beta_1 x_{i1} + ... + \beta_p x_{ip},$$

where $\boldsymbol{x_i} = (x_{i1}, ..., x_{ip})'$ and $\mu_i = E(Y|\boldsymbol{x_i})$ is the mean function and g is a well-known link function that has the property of strict uniformity and second-order derivability in the range of μ. GLMs can be used as a guideline for modeling, whose specific cases include linear model (LM), Poisson regression, logistic regression, and etc.

According to Hastie and Tibshirani (1986), in general, the GAM is a non-parametric extension of the GLM that replaces the linear predictor $\sum \beta_j x_{ij}$ with the additive predictor term $\sum f_j(x_{ij})$. The following model specify the structure of the non-parametric GAM,

$$GAM : g(\mu_i) = \beta_0 + \sum_{j}^{p} f_j(x_{ij}),$$

where the f_j's are smooth functions of the independent variables.

Among the link functions that can be considered are, Log link function $g(\mu_i) = \log(\mu_i)$, Identity link function $g(\mu_i) = \mu_i$, Inverse link function $g(\mu_i) = \frac{1}{\mu_i}$, Probit link function $g(\mu_i) = \Phi^{-1}(\mu_i)$, where $\Phi(.)$ is the cumulative distribution function of the standard normal distribution, Logit link function $g(\mu_i) = \log\{\mu_i/(1-\mu_i)\}$, that corresponds to the inverse CDF of the standard logistic distribution, complementary log–log (cloglog) link function $g(\mu_i) = \log\{-\log(1-\mu_i)\}$, that is formed from the inverse CDF of the Gumbel (or log-Weibull) distribution.

2.2 Time series by using GAM

The method of GAM is an effective and flexible technique for performing non-linear regression analysis in time series studies. Time series analysis methods such as autoregressive integrated moving average (ARIMA) models are a common proposition, but are impossible to use in data samples with irregular intervals because the software usually requires the interval of observations in time. One suggestion for estimating trends in such time series data, is to use locally estimated scatterplot smoothing (LOESS) (Cleveland, 1979; Birks, 1998). However, GAMs have also shown good results as they allow non-parametric adjustments for non-linear seasonal effects, trends and variables. LOESS and GAMs estimate non-linear and smooth trends in time series data and can implement irregular spacing of samples in time, but GAM does not have the problem of LOESS as being considered a formal statistical inference method.

The GAM for a trend in a series of T observations y_t at observation times x_t $(t = 1, 2, ..., T)$ is defined as follows,

$$g(\mu_t) = \beta_0 + f(x_t),$$

where μ_t is the expected value of the response variable Y_t $(\mu \equiv E(Y_t))$, g is the link function, a monotonic and invertible function and $f(x_t)$ is a smooth function of time.

2.3 Generalized additive distributed lag model

In this part, we first examine a smooth DLM and then its extension to the GAM is discussed. Let $(x_t, y_t), t = 1, , T$ represent a data set sorted by time. Moreover, additional covariates may be measured, but are ignored here. The distributed lag model of order q is defined as follow,

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + ... + \beta_q x_{t-q} + \epsilon_t, t = 1, ..., T,$$

where the ϵ_t are error terms. As respect to the outcome variable y_t may depend on x_t at time t (Both today and in the previous days), it is necessary to put a restriction on β_l and turn it into a simple function of the lag number to solve the collinearity problem. Also, this restriction can be extended to polynomial DLM of order (q, p) ($\text{PDLM}_{(q,p)}$) where β_l is restricted to be the piecewise pth polynomial function of l, $\beta_l = \sum_{j=0}^{p} \tau_j l^j, l = 0, ..., q,$. After bringing the $\text{PDLM}_{(q,p)}$ formula, since polynomial models are not suitable for more localized structures, its extended form is written as follow.

$$\beta_l = \sum_{j=0}^{p} \tau_j l^j + \sum_{k=1}^{K} \nu_k (l - \kappa_k)_+^p,$$

where $\kappa_1, ..., \kappa_K$ is a set of K discrete numbers between 0 and q, which are called knots. The above equation is usually referred to as the spline regression function of l, which can be estimated using P-spline smoothing. According to what was said, a generalized additive DLM is implemented as follows

$$GADLM : g(\mu_i) = \alpha + \boldsymbol{\gamma^T z_t} + \sum_{j=1}^{d} f_j(s_{tj}) + \sum_{l=0}^{q} \beta_l x_{t-l},$$

where $\boldsymbol{z_t}$ is a vector of linearly modeled variables (often dummy variables), s_{tj} is the jth variable modeled as a smooth function, and g is a link function.

3 Parameter estimation

Various methods have been proposed to estimate the parameters of the GAM. One of the parameter estimation methods is to use the backfitting algorithm, which was first introduced by Friedman and Stotzel (1981). Another approach, the local scoring method introduced by Hastie and Tibshirani (1986), is to update the Fisher score using a local estimate of the score. Moreover, Hastie and Tibshirani (1990) and Wood (2004) showed that the parameters can be estimated using the splines. In the following, we will introduce the estimation using the backfitting algorithm, the local scoring algorithm and the splines.

3.1 Backfitting algorithm

As mentioned, one of the parameter estimation methods is the use of the backfitting algorithm that was introduced by Friedman and Stotzel (1981). This approach is an iterative method where, at each step, one component is estimated while keeping the other components fixed, and then the algorithm moves component by component to reach convergence. This algorithm was proposed by Breiman and Fried (1985) for GAMs and its convergence was reviewed by Boja et al. (1989). In this part, the algorithm has been avoided, for more information about the backfitting algorithm, you can read the above articles.

3.2 Local scoring procedure algorithm

The proposed method Hastie and Tibshirani (1986) is a local scoring algorithm, which is an update of Fisher's score using local score estimation. Since smoothing generalizes the linear model, in the smoothed model $\eta = f(x)$, $f(.)$ can be estimated by iteratively smoothing the dependent variable set to X.

The local scoring algorithm includes the following 3 steps,
Step 1. Initialization: Set all smoothing functions to 0, for example,

$$\hat{f}_0 = g(E(Y)), \hat{f}_j^1 \equiv 0, m = 1.$$

Step 2. Construct the adjusted dependent variable as follows, $(m = m + 1)$

$$Z_i = \eta_i (y_i - \mu_i)(\frac{\partial \eta_i}{\partial \mu_i}),$$

so that,

$$\begin{aligned} \eta^{m-1} &= f_0 + \sum_{j=1}^{p} f_j^{m-1}(x_{ij}), \\ \eta^{m-1} &= g(\mu^{m-1}) \Rightarrow \mu^{m-1} = g^{-1}(\eta_i). \end{aligned}$$

Also make w_i weights as below,

$$w_i = (\frac{\partial \eta_i^{m-1}}{\partial \mu_i^{m-1}})^2 v_i^{-1}, v_i = Var(Y_i),$$

and then fit a weighted additive model to Z_i to obtain the estimate of the function f_j^m, based on the additive covariate η^m and the fitted value $\mu_i^m = p_i$.
Step 3. Repeat step 2 until the change in deviation is small enough,

$$D = 2\{\log(L(Y, \mu)) - \log(L(Y, \hat{\mu}_{LM}))\}.$$

For this, we can also consider the following expression,

$$\Delta(\eta^m, \eta^{m-1}) = \frac{\sum_{j=1}^{p} \|f_j^m - f_j^{m-1}\|}{\sum_{j=1}^{p} \|f_j^{m-1}\|}.$$

However, after considering this algorithm, you will realize that its estimation has problems such as convergence and lack of validity, and the use of this algorithm for estimation is ignored. Therefore, many authors have proposed more direct approaches to solve these problems, which can be referred to as the use of penalized likelihood.

3.3 Splines

The smoothing issue is one of the most powerful tools for data analysis in mathematics and statistics, also is used as the basis for many modern techniques. Due to the complexity of many statistical modeling, it is not possible to use a regular regression, also, in modeling, sometimes it is necessary to estimate a smoothing function so that it can cover and fit the data well. A commonly used suggestion for estimating the smoothing functions f_i in the GAM is splines, which should consider the space where f_i functions are located or are almost close to it, in order to convert the model into a linear model. Splines are used in order to mathematically make over flexible shapes, therefore the functions f_i which is a spline function with constant knote sequence $(\tau_1, ..., \tau_K)$ and degree d are written as follows using basis functions,

$$\boldsymbol{f(x)} = \sum_{i=1}^{p} \sum_{r=1}^{K+d+1} \beta_{ir} B_{ir}(x),$$

where $\boldsymbol{f(x)} = (f_1(x), ..., f_p(x))'$, B_{ir}'s are the collection of basis functions defined on a vector space V and β_{ir}'s are the associated spline coefficients. By placing the $f_i(x)$'s function in the model, it becomes a linear model that can be estimated using least squares.

In fact, the issue here is the selection of the degree of smoothing and basis functions, which is an approach to control the smoother. According to Yee (2015), four general categories of smoothers have been proposed, which are regression smoothers (polynomial regression, P-splines, regression splines, etc), smoothing splines (cubic smoothing splines, P-splines, O-splines), local regression (Lowess, Loess, it generalizes to local likelihood, etc) and nearest neighbor smoothers (running means, running medians, etc).

Nevertheless, due to the flexibility and easier calculation, some of these approaches are used more. In smoothing splines, knots are considered at all points $x_1, ..., x_n$ and a regular regression is performed based on the natural spline. Natural splines are cubic splines that require the spline function f to satisfy the condition $f'' = f''' = 0$, which leads to additional constraints. Since smoothing splines use only the inputs as knots, they overrule the knot selection problem and simultaneously control overfitting by minimizing the coefficients of the estimated function. It should be noted that a special case of the more general class of thin plate splines defined by Wood (2003), are smoothing splines. The approach for these types of smoothers is to start with many basis functions (for example, n of them) and penalized for some properties of these basis functions in order to control the flexibility of the fit.

Another common approach to facilitate knot location selection in spline modeling is to use penalized splines. Instead of using smoothing splines, it is more convenient to smooth using Eilers and Marx (1996) ”Penalized B-splines”, also known as ”P-splines”. Their solution can be easily computed because it involves direct linear algebra calculations, so estimation can be done in a similar way to generalized linear models (GLMs). There is no need for backward fitting and all functions are estimated simultaneously. Consider the sample data $(x_1, y_1), , (x_n, y_n)$, a penalized spline is defined as follows,

$$\hat{\beta} = argmax_{\beta}\{l_{\beta}(x_1, y_1, , x_n, y_n) - \lambda.J_{\beta}\}, \tag{1}$$

where l_β represents the logarithm of the likelihood and J_β is a roughness penalty that shrinks if the spline function is ”smooth”. Nevertheless, maximizing this function means a trade-off between model fit and smoothness, which is controlled by the tuning parameter $\lambda \geq 0$. Therefore, if λ is too large, the data will be over-smoothed, and if λ is too small, the data will be under-smoothed. A suitable criterion for choosing λ is the use of cross-validation.

4 Simulation

After presenting the issues related to the fitting of GAMs, now we intend to measure and evaluate the models in order to compare the fitted models by criteria such as mean squared error (MSE), Akaike information criterion (AIC), etc. As you have seen, in order to fit f_i functions, different methods can be used, such as the method presented by Hastie and Tibshirani (1986) under the title of local scoring algorithm, or the method presented by Wood (2004) using the penalized likelihood approach, etc. In this simulation we use splines to fit the model. Since GLM is not utilized in time series data, we only use independent variables for simulation. In this regard, we will benefit from various simulations by creating and designing different scenarios of data with linear and non-linear trends.

First, we explain how to generate data and the equation designed for each of them, and after fitting the model to the data, we compare the models. Note that all results presented below are based on 10,000 iterations. To begin with, we have created two datasets manually with different numbers of samples 50, 100, 250 and 1000 with linear equations and quadratic form. Consider the following equation,

$$y = f(x) + \epsilon,$$

by generating the random variable x from a uniform distribution (arbitrarily any distribution) and by defining error sentences that are randomly made from the family of

Table 1: The results related to fitting the GAM to the data with the linear equation.

Methods	n	50	100	250	1000
GLM	MSE	3.82	3.91	3.97	4.00
	AIC	213.86	425.19	1059.09	4227.97
	BIC	219.60	433.00	1069.66	4242.69
	Deviance	191.00	391.33	992.27	3995.21
GAMS	MSE	3.66	3.84	3.94	3.99
	AIC	213.02	424.50	1058.46	4227.36
	BIC	220.20	434.13	1071.47	4245.41
	Deviance	182.93	383.55	984.40	3987.42

exponential distributions and defining a function for each f_i, we create the response variable y.

After generating the data, we fit two models, the generalized linear model (GLM) and the generalized additive model using smoothing splines (GAMS), and in order to compare them, we evaluated the MSE, AIC, Bayesian Information Criterion (BIC) and deviation value (Deviance). The first equation which is linear, is defined as follows,

$$f(x) = 0.5x.$$

The lower the value of the defined criteria, it indicates that the model fitted on the data is better and covers them well. According to the information in the Table 1, it seems that the value of the criteria defined in the GAMS model is lower than the other model and it can be said that this model performed better. As you can see, with the increase in the number of samples, the MSE value for both models increases very slightly, but the other three criteria (i.e. AIC, BIC, Deviance) have increased a lot, so it can be said that the results are almost dependent on the number of samples. Also, while n increases, the change in the criteria becomes less meaningful and there is no considerable difference between the GLM and GAM. The second equation which has a quadratic form, is designed as follows,

$$f(x) = (1 + x)^2 - 0.2x.$$

The information in Table 2 shows that the GAMS has clearly performed better than the GLM.

Table 2: The results related to fitting the GAM to the data with the quadratic equation.

Methods	n	50	100	250	1000
GLM	MSE	59.51	59.54	59.53	59.51
	AIC	34625.65	34627.62	34627.38	34625.33
	BIC	34645.20	34647.17	34646.93	34644.88
	Deviance	297560.21	297677.59	297662.42	297541.10
GAMS	MSE	3.99	3.99	3.99	3.99
	AIC	21133.64	21131.43	21133.38	21132.69
	BIC	21199.34	21197.13	21199.07	21198.38
	Deviance	19974.50	19965.73	19973.57	19970.78

Increasing the sample from 50 to 1000 has not had much effect on all four criteria in the models and we have encountered both an increase and a little decrease. The next generated dataset is from the ”mgcv” statistical package in R software, which uses the gamsim() function to generate a suitable dataset for GAMs. The approach used in the gamsim function is to use examples related to Gu and Wahba (1991), where we have used the first example to generate data. Table 3 shows the fact that according

Table 3: The results related to fitting the GAM to the data with gamsim() function.

Methods	n	50	100	250	1000
GLM	MSE	3.82	3.91	3.97	4.00
	AIC	213.86	425.19	1059.09	4227.97
	BIC	219.60	433.00	1069.66	4242.69
	Deviance	191.00	391.33	992.27	3995.21
GAMS	MSE	3.66	3.84	3.94	3.99
	AIC	213.02	424.50	1058.46	4227.36
	BIC	220.20	434.13	1071.47	4245.41
	Deviance	182.93	383.55	984.40	3987.42

to the measured values of MSE, AIC and deviation, GAMS performed better than the other model, but the point here is that the BIC value is slightly lower in the GLM model. Paying close attention to the available values for both models, we can see that the sample size had an effect on the models and with the increase in the number of samples, the measured values also increased.

Based on the results obtained in each part of the simulation, it can be concluded that the GAM that is fitted with smoothing splines has a relatively better performance in different data situations, which can be used without the need to detect the non-linear relationship in advance. In the end, it can be mentioned again that GAMs perform very well for non-linear data scenarios and can be easily used.

5 Examples

Here we intend to point out the effectiveness of the GAM method in data modeling by providing some examples. As mentioned, GAM is used in different data situations, and since it was shown in the simulation section which the splines method works better for parameter estimation, we use this method for all the examples defined below.

5.1 GAM with independent variables

The example related to this part uses ”Auto” data from the ”ISLR” package in R software. These data include 392 independent observations on 9 variables, available in James et al. (2013). Here the intention is to investigate the effect of the variables ”weight”, ”mpg”, ”horsepower” and ”displacement” on ”acceleration”.Therefore, a GAM was fitted to the data using the ”mgcv” package in R. To estimate the unknown smoothing functions f in R, the s function was used, which by default is the thin plate regression splines. We see the result of the model in Table 4.

By checking Table 4 for the significance of the variables, we find that all the explanatory variables are significant and the value of R^2 indicates the good performance of the model. It is also possible to show a visual representation in Figure 1 of the GAM on the data, based on which the performance of the model can be understood.

Table 4: A summary of the GAM fit to the Auto data is provided.

Parametric coefficients:				
	Estimate	Std. Error	t value	p-value Bootstrap
(intercept)	15.54133	0.07205	215.7	0.0000
Approximate significant of smooth terms:				
	edf	Ref.df	F	p-value Bootstrap
s(mpg)	6.382	7.515	3.479	0.0010
s(displacement)	1.000	1.000	46.055	0.0000
s(horsepower)	4.883	6.006	70.187	0.0000
s(weight)	3.785	4.800	41.135	0.0000
R-sq.(adj)=0.733	Deviance explained=74.4%		n=391	

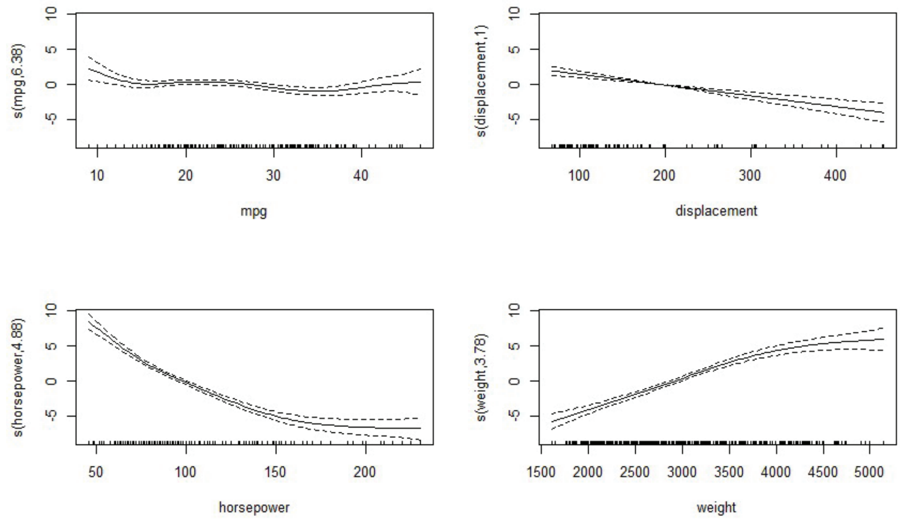

Figure 1: Plot related to data smoothing for Auto dataset.

5.2 GAM in time series

The example data is related to the total electricity consumption, which is collected as a time series with 672 observations. The data measurement procedure is every half hour once a day where we have 48 measurements per day and the average is considered for each day of the week. If we look at Figure 2 for visual observation, which shows electricity consumption in two weeks, we find that there are daily and seasonal effects that should be included in the model. The data and code for this example are available on the https://petolau.github.io/Analyzing-double-seasonal-time-series-with-GAM-in-R/github site.

After setting the daily and weekly variables, we fit a GAM to the data and use spline for the daily variable and P-spline for the weekly variable. You can see a summary of fitting model in the form of the following table (Table 5). According to the values in the table, it seems that the model is well fitted and the amount of R^2 indicates the convenient performance of the model.

By looking at Figure 3, you can see the effect of the variables on the amount of electricity consumption. As it is known, electricity consumption has clearly decreased

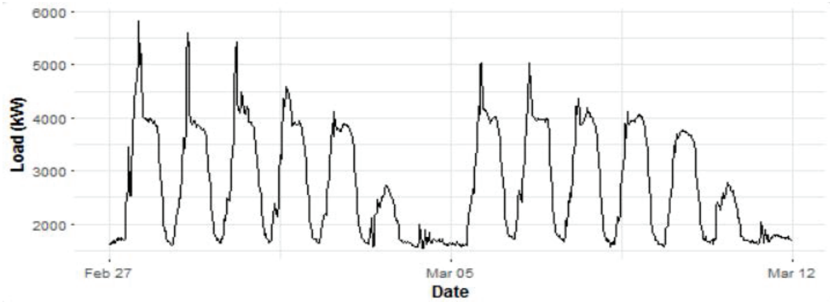

Figure 2: Time series data from two weeks of electricity consumption.

Table 5: The results of GAM fitting on time series data of electricity consumption.

Parametric coefficients:				
	Estimate	Std. Error	t value	p-value Bootstrap
(intercept)	2731.67	18.88	144.7	0.0000
Approximate significant of smooth terms:				
	edf	Ref.df	F	p-value Bootstrap
s(Daily)	10.159	12.688	119.8	0.0000
s(Weekly)	5.311	5.758	130.3	0.0000
R-sq.(adj)=0.772	Deviance explained=77.7%		n=672	

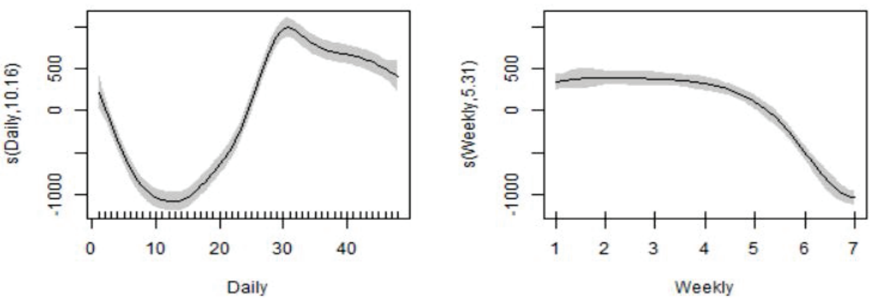

Figure 3: In the left and right panels of the figure, there are daily and weekly measurements, respectively.

on the weekend and the peak consumption is measured on the 30th.

5.3 GAM in time series with lag

Example data provided for GADLM are from daily time series of mortality, air pollution and meteorological variables in a study in Italy, measured for the ten-year period 1980-1989. In this example, two variables of daily mortality data (counting natural death certificates of Milan residents) and total suspended particulate matter (TSP) as pollution measures are used. Here, fitted a Poisson additive distributed model with mortality as the response variable. A traditional model that includes the TSP effect on the same day and a DLM that includes the TSP effect distributed on the same day and the previous 45 days are fitted to the data. In DLM, penalized splines are used for smooth expressions. In this model, variables of relative humidity, temperature and number of days are considered as covariates to measure their effect on mortality rate. Since other unmeasured covariates such as diet and smoking have seasonal patterns, they have no effect in the short term and can be ignored. Further details on this example and data are provided in Zanobetti et al. (2000).

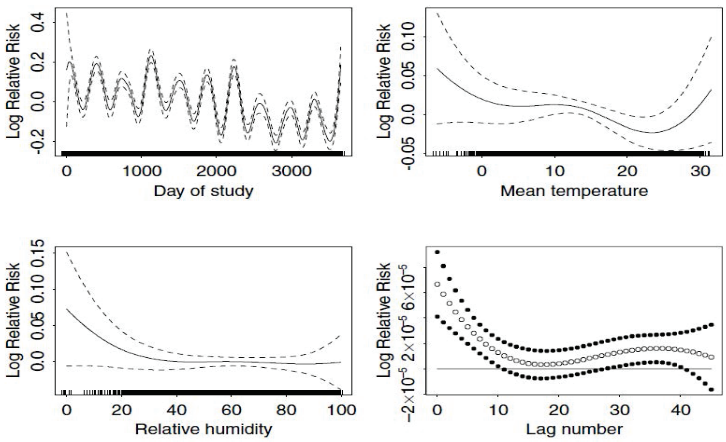

Figure 4: The smooth curve components for Milan mortality data after generalized additive DLM fitting.

Figure 4 illustrates a visual representation of the smooth functions that showed almost a U-shaped relationship between the mean temperature and relative humidity with mortality. In fact, people are more vulnerable to extreme amounts of these factors. Note that, the dashed lines or dots around the fitting region correspond to ± twice the estimated standard deviation of each estimate.

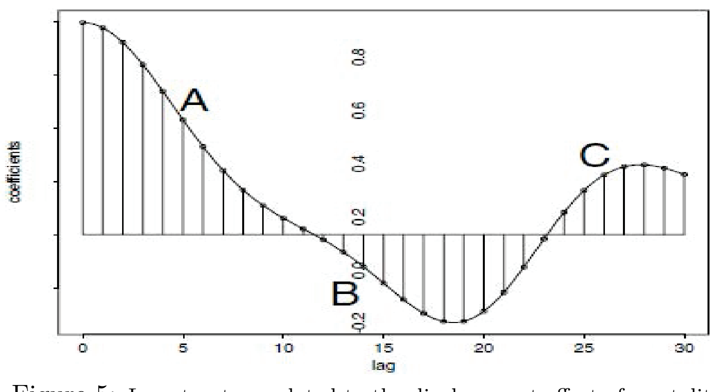

Figure 5: Lag structure related to the displacement effect of mortality.

Figure 5 shows how the effect of TSP is decomposed and divided into three parts A, B and C. As it is clear in the figure, the vertical axis shows the coefficients β (the effect of pollution a few days ago on today's mortality rate) and the horizontal axis shows the delay. A shows that for low lag, the sum of the coefficients is positive, in fact, we see the positive effect of pollution on mortality in a few days. As shown in Figure 5, some coefficients in B are negative, which indicates that the level of pollution has a negative effect in a longer period, and then there is a period C in which the regression coefficients are once again positive. The components defined above, as well as the cumulative effect of $A+B+C$ have been investigated and the results are shown in Table 6.

Table 6: Coefficient amount of β, the standard errors, the relative risks and a confidence interval for the decomposition of effect TSP are shown.

	$\hat{\beta}$	SE	RR	$CI5\%$	$CI95\%$
A	0.000424	0.000106	1.037	1.019	1.056
B	0.000018	0.000027	1.002	0.997	1.006
C	0.000307	0.000165	1.027	0.998	1.056
$A+B+C$	0.000749	0.000161	1.067	1.038	1.096

6 Discussion and conclusions

Many analysts need knowledge of powerful statistical tools to work with data. GAM provides a good method for modeling non-linear process data that can predetermine the shape of the non-linear relationship without the need for an analyst. GAM has many applications in different data situations such as when the variable is independent, time series, or has a lag. In addition, it is utilized in various academic and research fields like climate variables, economics, clinical research, biology, etc. In this article, we tried to review explanations about non-parametric GAM and its application and show the power and efficiency of this statistical method by providing examples and simulations.

References

Allmon, S. (1965). The distributed lag between capital appropriations and expenditures. *Econometrica*, **33**: 178–196.

Birks, H.J.B. (1998). Numerical tools in palaeolimnology-progress, potentialities, and problems. *Journal of Paleolimnol*, **20**: 307–332.

Breiman, L. and Fried, J. (1985). Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association*, **80**(391): 580–598.

Bremner, S.A., Anderson, H.R., Atkinson, R.W., McMichael, A.S., Strachan, D.P., Bland, J.M. and Bower, J.S. (1999). Short term association between outdoor air pollution and mortality in London 1992-4. *Occupational and Environmental Medicine*, **56**(4): 237–244.

Buja, A., Hastie, T. and Tibshirani, R. (1989). Linear smoothers and additive models. *The Annals of Statistics*, **17**(2): 453–510.

Chang, C-H., Caruana, R., Goldenberg, A. (2022). NODE-GAM: Neural generalized additive model for interpretable deep learning. *arXiv preprint arXiv:2106.01613.*

Cleveland, W.S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, **74**(368): 829–836.

Corradi, C. (1977). Smooth distributed lag estimators and smoothing spline functions in Hilbert spaces. *Journal of Econometrics*, **5**(2): 211–220.

Dehghan, A., Khanjani, N., Bahrampour, A., Goudarzi, G., Yunesian, M. (2018). The relation between air pollution and respiratory deaths in Tehran, Iran-using generalized additive models. *BMC pulmonary medicine*, **18**(1): 1-9.

Dominici, F., Samet, J. and Zeger, S. L. (2000). Combining evidence on air pollution and daily mortality from the largest 20 US cities: a hierarchical modeling strategy. *Journal of the Royal Statistical Society*, **163**(3): 263–284.

Eilers, P.H.C. and Marx, B.D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, **11**(2): 89–121.

Friedman, J.H. and Stuetzle, W. (1981). Projection pursuit regression. *Journal of the American Statistical Association*, **76**(376): 817–823.

Hastie, T.J. and Tibshirani, R.J. (1986). Generalized additive models. *Statistical Science*, **3**(1): 297-318.

Hastie, T.J. and Tibshirani, R.J. (1990). *Generalized Additive Models*. London: Chapman and Hall.

Hoek, G., Schwartz, J., Groot, B. and Eilers, P. (1997). Effects of ambient particulate matter and ozone on daily mortality in Rotterdam, The Netherlands. *Archives of Environmental Health: An International Journal*, **52**(6): 455–463.

Gu, CH. and Wahba, G. (1991). Minimizing GCV/GML scores with multiple smoothing parameters via the Newton method. *SIAM Journal on Scientific and Statistical Computing*, **12**(2): 383–398

James, G., Witten, D., Hastie, T., Tibshirani, R. (2013). *An Introduction to Statistical Learning.* Vol. 112, New York: Springer.

Kelsall, J.E., Samet, J.M., Zeger, S.L. and Xu, J. (1997). Air pollution and mortality in Philadelphia: 1974-1988. *Americal Journal of Epidemiology*, **146**(9): 750–762.

Nelder, J. and Wedderburn, R. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, **135**(3): 370–384.

Pinilla, J. and Negrín, M. (2021). Non-Parametric generalized additive models as a tool for evaluating policy interventions. *Mathematics*, **9**(4): 299.

Ravindra, K., Rattan, P., Mor, S., Aggarwal, A.N. (2019). Generalized additive models: Building evidence of air pollution, climate change and human health. *Environment International*, **132**: 104987.

Schwartz, J. (1993). Air pollution and mortality in Birmingham, AL. *American Journal of Epidemiology*, **137**(10): 1136–1147.

Schwartz, J. (1994). Nonparametric smoothing in the analysis of air pollution and respiratory illness. *The Canadian Journal of Statistic*, **22**(4): 471–487.

Wood, S.N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **65**(1): 95–114.

Wood, S.N. (2004). *Generalized Additive Models: an introduction with R.* first ed, New York: Chapman and Hall/CRC.

Wood, S.N. (2017). *Generalized Additive Models: an Introduction with R.* 2nd ed, CRC Press.

Yang, K. and Maiti, T. (2020). Ultra high dimensional generalised additive model: Unified Theory and Methods. *Scandinavian Journal of Statistics*, **49**(3): 917–942.

Yee, T.W. (2015). *Vector Generalized Linear and Additive Models with an Implementation in R.* New York: Springer.

Zanobetti, A., Wand, M.P., Schwartz, J., Ryan, L.M. (2000). Generalized additive distributed lag models: Quantifying mortality displacement. *Biostatistics*, **1**(3): 279–292.

Zhao, J., Cao, J., Tian, S., Chen, S., Zhang, S., Wang, Z. and Zhou, X. (2014). A comparison between two GAM models in quantifying relationships of environmental variables with fish richness and diversity indices. *Aquatic ecology*, **48**: 297–312