Geometrically Designed Variable Knot Splines in Generalized (Non-)Linear Models

Dimitrina S. Dimitrova[a*], Vladimir K. Kaishev[a], Andrea Lattuada[b] and Richard J. Verrall[a]

[a]Faculty of Actuarial Science and Insurance, Bayes Business School (formerly Cass), City, University of London
[b]Department of Mathematical Sciences, Mathematical Finance and Econometrics, Catholic University of the Sacred Heart

August 6, 2022

Abstract

In this paper we extend the GeDS methodology, recently developed by Kaishev et al. (2016) for the Normal univariate spline regression case, to the more general GNM/GLM context. Our approach is to view the (non-)linear predictor as a spline with free knots which are estimated, along with the regression coefficients and the degree of the spline, using a two stage algorithm. In stage A, a linear (degree one) free-knot spline is fitted to the data applying iteratively re-weighted least squares. In stage B, a Schoenberg variation diminishing spline approximation to the fit from stage A is constructed, thus simultaneously producing spline fits of second, third and higher degrees. We demonstrate, based on a thorough numerical investigation that the nice properties of the Normal GeDS methodology carry over to its GNM extension and GeDS favourably compares with other existing spline methods.

The proposed GeDS GNM/GLM methodology is extended to the multivariate case of more than one independent variable by utilizing tensor product splines and their related shape preserving variation diminishing property.

Keywords: variable-knot spline regression; tensor product B-splines; Greville abscissae; control polygon; generalized non-linear models

*Corresponding author's address: Faculty of Actuarial Science and Insurance, Bayes Business School (formerly Cass), City, University of London, 106 Bunhill Row, London EC1Y 8TZ, UK. E-mail address: d.dimitrova@city.ac.uk

1 Introduction

Various spline based constructions to express and estimate the linear predictor in Generalized Linear Models (GLM) have recently been proposed among which the Semi-parametric Models (SPM) of Rupert et al. (2003), the Generalized Smoothing Spline (GSS) ANOVA models introduced in Wahba et al. (1995) (see also Gu 2013 and related implementation in the **R** package **gss** by Gu 2014), the Generalized Additive Models (GAM) of Hastie and Tibshirani (1990) and the Adaptive Bayesian Regression Splines proposed by Biller (2000). The scope and methodological background of each of these approaches can be summarised as follows. Common features of the SPM, GAM and the GSS models are their truly multivariate nature and the fact that the underlying estimator for the linear predictor part is some form of a smoothing spline or a related penalized version of it. In SPM the linear predictor is a $p-$spline introduced by Eilers and Marx (1996). SPM are coded in the **R** package **SemiPar** developed by Wand (2018) which is currently limited to handling only Normal, Poisson and Binomial data. In **SemiPar**, the p-spline linear predictor is in the form of a mixed model, in terms of radial basis functions. This has allowed for some algorithmic advantages, including automatic fitting of SPM using the generalized linear mixed model representation and (automatic) smoothing parameter estimation using restricted maximum likelihood.

Under the GAM methodology of Hastie and Tibshirani (1990) it is assumed that the linear predictor is in the form of an Additive Model (AM) in which the underlying additive functions can be selected from a pool of spline models. GAM is implemented e.g., in the **R** package **mgcv** developed by Wood (2006) where this pool includes p-splines, Thin Plate splines, cubic smoothing splines and shrinkage versions of them. If cubic smoothing splines are selected, they have a prescribed number of knots, whose locations are preselected following certain quantiles of the input covariate data and a smoothing penalty parameter is estimated. As a result of this, the method is quite fast, but as demonstrated in Section 4 and in the online supplement may experience significant under-fitting. For further advantages and disadvantages of alternative choices of spline models in **mgcv** we refer to Wood (2006) (see page 222 therein). An overall advantage of GAM is its simple additive structure which in the multivariate case allows for more straightforward interpretation of the model

estimation results, since one can inspect separately the effect of each covariate on the response variable.

In summary, in most of the aforementioned literature, the number and positions of the knots are fixed a priori and the (global) degree of smoothness is controlled by a smoothness penalty parameter estimated based on the data. However, as demonstrated in Section 4 there are many applications where it is desirable to have more flexibility in the spline predictor component in order to control the smoothness locally. A natural way to achieve this is by assuming that the knots are free parameters. It is directly verified however, that knots enter the spline predictor non-linearly and their estimation would require considering the broader class of generalized non-linear models (GNM) which include GLM as a special case. The GNM have been considered e.g. by Lane (1996) and Turner and Firth (2020). In such models, the predictor is nonlinear with respect to only a few of the parameters and is linear in the remaining ones. So the GNM structure seems not to have been set to naturally embed free-knot spline models in which the number of nonlinear (knot) parameters can be quite high. This could be among the reasons why GNM has not been used (to the best of our knowledge), to implement generalized non-linear spline models with free knots.

Such non-linear models have been considered in the Normal case by e.g. Lindstrom (1999) but the direct ML estimation of the knot locations requires non-linear least squares optimization, which is often computationally prohibitive with corresponding complications and limitations (due to e.g. multiple extrema, *knot confounding* or *lethargy* problems; cf. Jupp 1978, Zhou and Shen 2001, Beliakov 2004 and Kovács and Fekete 2019). Alternative data driven procedures to estimate the knot positions have been proposed by many authors in the Normal case (see e.g. Spiriti et al. 2018, Yeh et al. 2020). A brief review of the related literature on Normal free-knot spline regression is provided by Kaishev et al. (2016). These authors have recently proposed a geometrically motivated approach to variable knot spline regression estimation. More precisely, Kaishev et al. (2016) propose to view estimation as a geometrically motivated procedure consisting of two stages. At the first stage, A, a least squares linear spline is fitted to the data, starting from a straight line, by sequentially adding knots (one at a time), which can be interpreted as sequentially breaking a stick into smaller polygonal pieces specified by adjacent left and right end knots, added in the

model. At the second stage B, smoother higher order spline fits are built by viewing them as Schoenberg's variation diminishing spline (VDS) approximations to the piecewise linear spline fit from stage A. Geometrically, this means that these fits closely follow the shape of the piecewise linear spline fit from stage A. The latter is viewed as a control polygon capturing the shape of the curve underlying the noisy data, similarly as designers construct such control polygons in order to capture shapes of curves and surfaces in Computer Aided Geometric Design applications. They then compute smoother higher order Schoenberg's variation diminishing spline curves that closely follow the control polygon and hence the desired shape. Such geometrically designed (GeD) variable knot spline regression models have been referred to by Kaishev et al. (2016) as geometrically designed splines (GeDS).

As has been demonstrated, GeDS estimation avoids many of the limitations of the existing knot optimization methods, while at the same time producing competitive fits with a low number of knots for a wide range of signal-to-noise ratios and for both sparse and dense data points at a very low computational cost, utilizing a stopping rule controlled by two tuning parameters (see Kaishev et al. 2016). In contrast to non-linear estimation of the knot locations, in GeDs we utilize a two stage procedure that possesses some very attractive features, i.e. very fast geometrically motivated (data driven) estimation of knots and coefficients, whose number is adequately estimated to match the shape underlying the data, simultaneous estimation of the degree of the spline fit, based on the linear, quadratic, cubic, and possibly higher order spline fits produced simultaneously in stage B. These conclusions are made based on through comparison with some popular estimators such as SPM, GSS and GAM (see Section 4 and the online supplement to this paper). GeDS has also been successfully applied in estimating multivariate Archimedean copulas (c.f. Dimitrova et al. 2008) and in smoothing mortality data (c.f. Kaishev et al. 2009).

In summary, our contribution in this paper is in extending the GeDS methodology, originally developed for the univariate Normal case, to the more general GNM (GLM) context in which the response variable may have any distribution from the exponential family (EF) and also in extending it to two (and higher) dimensions so as the response variable may depend on more than one independent variable. Moreover, we have implemented this generalized GeDS methodology in an **R** package, named **GeDS**, available from the Com-

prehensive **R** Archive Network (CRAN) at `http://CRAN.R-project.org/package=GeDS`. The latter is used in the study of the numerical properties of GeDS and in thoroughly comparing GeDS with major existing spline fitting methodologies (see Section 4 and the online supplement). We demonstrate that the nice properties of the Normal GeDS methodology carry over to its GNM extension and GeDS favourably compares with the existing methods such as GAM, SPM and GSS implemented in **R**. In the Normal case, we have also compared with the adaptive P-splines of Yang and Hong (2017) and the **trendfilter** fitting method in the package **genlasso** of Arnold and Tibshirani (2020)).

The paper is organized as follows. In Section 2 we formulate the spline GNM (GLM) fitting problem and its solution utilizing the variation diminishing (shape preserving) properties of splines. The latter formulation allows to view estimation of the spline regression coefficients and knots as a geometric operation. In Section 3 we describe in detail the two stages of the GeDS GNM/GLM methodology. In Section 4 the numerical properties of GeDS are thoroughly investigated and compared to existing alternative estimators among which SPM, GSS and GAM. In Section 5 we describe a multivariate extension of the GeDS algorithm to the case of more than one explanatory variables utilizing tensor product splines and their related (shape preserving) variation diminishing property. Finally, in Section 6 we provide some conclusions and directions of future work.

2 The Spline GNM (GLM) fitting problem: a geometric perspective

In order to introduce GeDS in the GNM (GLM) framework and follow its characterizing geometric properties, we first introduce some notation and background on GLM and VDS approximation.

2.1 The GNM (GLM) fitting problem and the B-spline predictor

We denote the univariate response variable by y, assume that its mean, $\mu = \mu(x) = \mathbb{E}(y|x)$ depends on a single covariate x, and that its distribution belongs to the exponential family,

i.e. has density of the form

$$f(y;\vartheta,\phi) = \exp\left\{\frac{y\vartheta - b(\vartheta)}{a(\phi)} + c(y,\phi)\right\}, \tag{1}$$

where $\vartheta = \vartheta(\mu)$ is the natural parameter, ϕ is the scale parameter and a, b and c are functions that characterize the distribution. We also assume that the transformed mean $g(\mu) = \eta(\boldsymbol{\theta}; x)$, where $g(\cdot)$ is the link function, $\eta(\boldsymbol{\theta}; x)$ is the linear predictor and $\boldsymbol{\theta}$ is a vector of unknown parameters. In what follows we will also express μ as $\mu(x) = g^{-1}(\eta(\boldsymbol{\theta}; x))$, where $g^{-1}(\cdot)$, is the inverse function of $g(\cdot)$. Given a sample of observations, $\{x_i, y_i\}_{i=1}^N$, and preliminary selected functions, g and η, standard GLM procedures find the maximum likelihood estimate (MLE) of the vector of parameters $\boldsymbol{\theta}$, i.e., estimate $\mu(x) = g^{-1}(\eta(\boldsymbol{\theta}; x))$. For the purpose, the iteratively re-weighted least squares (IRLS) algorithm is applied to estimate $\boldsymbol{\theta}$ by regressing the (approximated) transformed responses, $g(y_i)$, on x_i, $i = 1, \ldots, N$ (see e.g. Charnes et al. 1976). The IRLS estimation is terminated when the deviance,

$$D(\mu) = 2\sum_{i=1}^{N} \{y_i(\vartheta(y_i) - \vartheta(\mu(x_i))) - b(\vartheta(y_i)) + b(\vartheta(\mu(x_i)))\},$$

usually considered as the goodness of fit measure, is minimized. In Section 5 we will extend the GLM setting to the multivariate case.

We will consider here a more general (non-linear) formulation of the GLM fitting problem, assuming that the predictor, $\eta(\boldsymbol{\theta}; x)$ is chosen to be a polynomial spline function, whose degree, number and location of knots are viewed as unknown parameters that need to be estimated. We denote by $S_{\mathbf{t}_{k,n}}$ the linear space of all n-th order spline functions defined on a set of non-decreasing knots $\mathbf{t}_{k,n} = \{t_i\}_{i=1}^{2n+k}$, where $t_n = a$, $t_{n+k+1} = b$. We will use splines with simple knots, except for the n left and right most knots which will be assumed coalescent, i.e.

$$\mathbf{t}_{k,n} = \{t_1 = \ldots = t_n < t_{n+1} < \ldots < t_{n+k} < t_{n+k+1} = \ldots = t_{2n+k}\}. \tag{2}$$

We further assume that the predictor component of the GLM is in the form of a spline

function $f \in S_{\mathbf{t}_{k,n}}$, expressed as

$$\eta(\boldsymbol{\theta}; x) = f(\mathbf{t}_{k,n}; x) = \boldsymbol{\theta}^T \mathbf{N}_n(x) = \sum_{i=1}^{p} \theta_i N_{i,n}(x), \tag{3}$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^T$ is a vector of real valued coefficients and $\mathbf{N}_n(x) = (N_{1,n}(x), \ldots, N_{p,n}(x))^T$, $p = n+k$, are B-splines of order n, defined on $\mathbf{t}_{k,n}$. It is well known that $\sum_{i=j-n+1}^{j} N_{i,n}(t) = 1$ for any $t \in [t_j, t_{j+1})$, $j = n, \ldots, n+k$, and $N_{i,n}(t) = 0$ for $t \notin [t_i, t_{i+n}]$. In the sequel, where necessary, we will emphasize the dependence of the spline $f(\mathbf{t}_{k,n}; x)$ on $\boldsymbol{\theta}$ by using the alternative notation $f(\mathbf{t}_{k,n}, \boldsymbol{\theta}; x)$.

The GLM fitting problem can now be more precisely stated as follows. Given a sample of observations $\{x_i, y_i\}_{i=1}^N$, estimate the order n of the spline predictor, $\eta \equiv f(\mathbf{t}_{k,n}, \boldsymbol{\theta}; x)$, its number of knots k, their locations $\mathbf{t}_{k,n}$ and the coefficients, $\boldsymbol{\theta}$, so that an appropriate goodness-of-fit measure based on the deviance

$$\begin{aligned} D(\mu) &= D(g^{-1}(f(\mathbf{t}_{k,n}, \boldsymbol{\theta}; x))) \\ &= 2\sum_{i=1}^{N} \left\{ y_i(\vartheta(y_i) - \vartheta(g^{-1}(f(\mathbf{t}_{k,n}, \boldsymbol{\theta}; x_i)))) - b(\vartheta(y_i)) + b(\vartheta(g^{-1}(f(\mathbf{t}_{k,n}, \boldsymbol{\theta}; x_i)))) \right\}, \end{aligned} \tag{4}$$

is minimized. In what follows, we will use the notation $D(\mu) := D(\boldsymbol{\theta}_p; k, n)$ in order to emphasize the dependence of the deviance on the number of spline coefficients p, knots k, and order n.

As already mentioned, under this formulation of the GLM fitting problem, the predictor component is, in general, non-linear in the unknown parameters, since $f(\mathbf{t}_{k,n}, \boldsymbol{\theta}; x)$ is non-linear with respect to the knots $\mathbf{t}_{k,n}$. Note that this is different from the GLM formulations involving smoothing splines where the number of knots and their locations are pre-determined (usually equidistant), i.e. are not considered unknown parameters. Therefore it will be more precise to say that the model considered here falls within the class of GNM (see e.g. Lane 1996 and Turner and Firth 2020).

2.2 The B-spline predictor and its control polygon

In order to solve this GNM estimation problem, without confronting the usual complications stemming from the non-linearity of the predictor $\eta \equiv f(\mathbf{t}_{k,n}, \boldsymbol{\theta}; x)$, we alternatively view $f(\mathbf{t}_{k,n}, \boldsymbol{\theta}; x)$ as a special case of a parametric spline curve $\mathbf{Q}(t)$, $t \in [a, b]$. A parametric spline curve $\mathbf{Q}(t)$ is given coordinate-wise as

$$\mathbf{Q}(t) = \{x(t), y(t)\} = \left\{ \sum_{i=1}^{p} \xi_i N_{i,n}(t), \sum_{i=1}^{p} \theta_i N_{i,n}(t) \right\}, \tag{5}$$

where t is a parameter, and $x(t)$ and $y(t)$ are spline functions, defined on one and the same set of knots $\mathbf{t}_{k,n}$, with coefficients ξ_i and θ_i, $i = 1, \ldots, p$, respectively. If the coefficients ξ_i in (5) are chosen to be the knot averages

$$\xi_i^* = \left(t_{i+1} + \cdots + t_{i+n-1}\right)/(n-1), \quad i = 1, \ldots, p, \tag{6}$$

then it is possible to show that the identity

$$x(t) = \sum_{i=1}^{p} \xi_i^* N_{i,n}(t) = t, \tag{7}$$

referred to as the linear precision property of B-splines, holds (see e.g. De Boor 2001). The values ξ_i^* given by (6) are known as the Greville abscissae. We will alternatively use the notation $\boldsymbol{\xi}^*(\mathbf{t}_{k,n})$, to indicate the dependence of the set $\boldsymbol{\xi}^*$ on the knots $\mathbf{t}_{k,n}$. In view of (5) and (7), the (predictor) spline function $f(\mathbf{t}_{k,n}, \boldsymbol{\theta}; x)$ can be expressed as a parametric spline curve

$$\mathbf{Q}^*(t) = \{t, f(\mathbf{t}_{k,n}, \boldsymbol{\theta}; t)\} = \left\{ \sum_{i=1}^{p} \xi_i^* N_{i,n}(t), \sum_{i=1}^{p} \theta_i N_{i,n}(t) \right\}, \tag{8}$$

where $t \in [a, b]$. In what follows, it will be convenient to use $\mathbf{Q}^*(t)$ and $f(\mathbf{t}_{k,n}, \boldsymbol{\theta}; t)$ interchangeably to emphasize that the GNM predictor, η, is in a functional spline curve form.

Interpretation (8), of the predictor $\eta \equiv f(\mathbf{t}_{k,n}, \boldsymbol{\theta}; x)$ as a parametric spline curve $\mathbf{Q}^*(t)$,

allows us to characterize the predictor spline curve $\mathbf{Q}^*(t)$ by a polygon, with vertices $\mathbf{c}_i = (\xi_i^*, \theta_i)$, $i = 1, \ldots, p$. The latter is closely related to $\mathbf{Q}^*(t)$, and is called the control polygon of $\mathbf{Q}^*(t)$, denoted by $\mathbf{C}_{\mathbf{Q}^*}(t)$, and constructed by connecting the points $\mathbf{c}_i = (\xi_i^*, \theta_i)$ with straight lines, (see Figure 1).

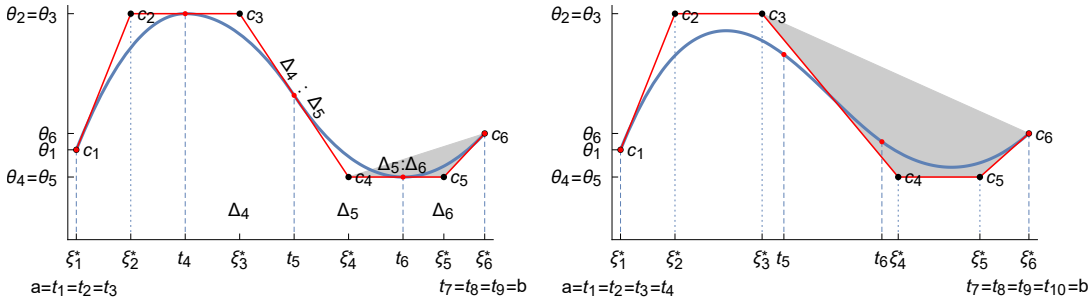


Figure 1: Quadratic, $n = 3$, (left panel) and cubic, $n = 4$, (right panel) functional spline curves $\mathbf{Q}^*(t)$ and their control polygons $\mathbf{C}_{\mathbf{Q}^*}(t)$. The spline coefficients $\boldsymbol{\theta}$ and the sets of knots $\boldsymbol{t}_{3,3}$ and $\boldsymbol{t}_{2,4}$ are arbitrarily chosen, and the corresponding Greville sites $\boldsymbol{\xi}^*(\mathbf{t}_{3,3})$ and $\boldsymbol{\xi}^*(\mathbf{t}_{2,4})$ are evaluated following (6).

This relationship is due to the fact that both the x and y coordinates of the vertices of the control polygon, $\mathbf{c}_i$, $i = 1, \ldots, p$, called control points, are related to the linear predictor spline curve $\mathbf{Q}^*(t)$. More precisely the x-coordinates ξ_i^*, are the Greville sites, (6), obtained from the knots $\mathbf{t}_{k,n}$, and the y-coordinates, θ_i, are simply the spline coefficients, $i = 1, \ldots, p$. As illustrated in Fig. 1 in the quadratic, $n = 3$, and cubic, $n = 4$, cases (left and right panels respectively), the curve closely follows its control polygon. In particular, it directly coincides with its polygon when $n = 2$, is in its close vicinity when $n = 3$, and deviates a bit further from it when $n = 4$. There are two reasons for this close relationship, one is that each segment of the curve lies within the convex hull of n consecutive vertexes of its polygon, as illustrated by the shaded areas on Fig. 1. The second is that, $\mathbf{Q}^*(t)$ is a Schoenberg's variation diminishing approximation of its polygon. This means that it preserves the shape of the polygon, i.e., if the latter is positive the curve is also positive, if it is increasing the curve is also increasing, if it is convex the curve is also convex. For further details and examples of the close geometric relationship between a parametric spline curve and its control polygon, we refer to Kaishev et al. (2016).

This relationship suggests that, given n and k, locating the knots $\mathbf{t}_{k,n}$ and finding the coefficients $\boldsymbol{\theta}$ of $f(\mathbf{t}_{k,n}, \boldsymbol{\theta}; x)$, based on the set of observations $\{y_i, x_i\}_{i=1}^N$, is equivalent to

finding the location of the x- and y-coordinates of the vertices of $\mathbf{C}_{f(\mathbf{t}_{k,n},\boldsymbol{\theta};x)}$. This establishes the important fact that estimation of $\mathbf{t}_{k,n}$ and $\boldsymbol{\theta}$ affects the geometrical position of the control polygon $\mathbf{C}_{f(\mathbf{t}_{k,n},\boldsymbol{\theta};x)}$, which, due to the shape preserving and convex hull properties, defines the location of the spline curve $f(\mathbf{t}_{k,n},\boldsymbol{\theta};x)$. Inversely, locating the vertices $\mathbf{c}_i$ of $\mathbf{C}_{f(\mathbf{t}_{k,n},\boldsymbol{\theta};x)}$ affects the knots $\mathbf{t}_{k,n}$, through (6), and the values of $\boldsymbol{\theta}$, and hence affects the position of the predictor curve $f(\mathbf{t}_{k,n},\boldsymbol{\theta};x)$. As in the Normal case, the latter conclusion motivates the construction, in stage A of GeDS, of a control polygon as an IRLS fit to the data, whose knots determine the knots $\mathbf{t}_{k,n}$, and whose B-spline coefficients are viewed as an initial estimate of $\boldsymbol{\theta}$, which is improved further in stage B (see Section 3). This is the basis of our approach to constructing the GeD variable knot spline predictor as part of the GNM estimation which is summarized in the next section.

3 The GeDS estimation method.

As mentioned above, the GeDS methodology in the Normal regression case incorporates two major stages, A and B. The main idea behind generalizing GeDS to fit GNM (GLM), is as follows. At Stage A, the spline predictor is expressed as a linear combination of second order (degree one) B-splines. Thus, starting with zero knots and adding knots, one by one at appropriate locations, the best piecewise linear (predictor) fit to the shape underlying the data is achieved in the transformed predictor scale. At each step of this sequential knot addition, the coefficients of the spline predictor are estimated by applying the IRLS procedure, used in estimating GLM. A stopping rule (model selector) based on a ratio of consecutive deviances is applied to exit Stage A.

At stage B, smoother, higher order spline representations of the GLM predictor component are found by viewing them as Schoenberg's VDS approximations to the linear spline fit from stage A. The latter VDS approximations are then slightly adjusted to become maximum likelihood fits to the data. Therefore, the nearly Schoenberg variation diminishing maximum likelihood fits of stage B, have very distinct geometric properties in that they closely follow the shape of the linear spline fit from stage A, viewed as a control polygon (see Section 2, Fig. 1).

It is worth noting that because of the sequential location of the knots, the remaining

unknown parameters at each step of stage A are the B-spline coefficients and therefore the spline predictor remains in the GLM class. This holds true also with respect to the final GeDS fit from stage B, as it preserves the (optimal) knot locations of stage A. As a result of that GeDS is essentially a geometrically motivated procedure that at stage A builds nested GLM spline predictor models, the final one of which is approximated by higher order GLM fits in stage B, in order to solve the GNM estimation problem formulated in Section 2.

3.1 Building the piecewise linear spline predictor in Stage A

The steps of stage A of the generalized GeDS are similar to those in the Normal case (cf. Kaishev et al. 2016)), employing appropriate IRLS fitting and a deviance-based stopping criterion, as opposed to LS fitting and a stopping rule based on residual sum of squares. Stage A of the generalized GeDS can be summarized as follows.

Stage A. Starting from a straight line fit and adding one knot at a time (see Section 3.2 for details), follow the IRLS procedure to find the linear spline fit $\hat{f}\left(\boldsymbol{\delta}_{\kappa,2}, \hat{\boldsymbol{\alpha}}_p; x\right) = \sum_{i=1}^{p} \hat{\alpha}_i N_{i,2}(x)$ with number of internal knots κ, number of B-splines (and parameters) $p = \kappa + 2$ and with a set of knots $\boldsymbol{\delta}_{\kappa,2} = \{\delta_1 = \delta_2 < \delta_3 < \ldots < \delta_{\kappa+2} < \delta_{\kappa+3} = \delta_{\kappa+4}\}$, such that

$$\hat{\phi}_\kappa = 1 - \exp\{\hat{\gamma}_0 + \hat{\gamma}_1 \kappa\} \geq \phi_{exit} \tag{9}$$

where $\hat{\phi}_\kappa$ is the least squares fit (with coefficients $\hat{\gamma}_0$ and $\hat{\gamma}_1$) to the sample $\{\phi_h, h\}_{h=q}^{\kappa}$, evaluated at κ, and $\phi_{exit} \in (0, 1)$ is a pre-specified threshold level (close to one). The sample $\{\phi_h, h\}_{h=q}^{\kappa}$ is obtained by evaluating, at each step in the knot addition process, the ratios of deviances

$$\phi_h := \frac{D(\hat{\boldsymbol{\alpha}}_{h+2}; h, 2)}{D(\hat{\boldsymbol{\alpha}}_{h+2-q}; h-q, 2)}, h = q, q+1, q+2, \ldots$$

where $D(\hat{\boldsymbol{\alpha}}_{h+2}; h, 2) := D(g^{-1}(f(\boldsymbol{\delta}_{h,2}, \hat{\boldsymbol{\alpha}}_{h+2}; x))$ is computed using (4) and $q \geq 1$, $\kappa \geq q+1$*. Testing the inequality in (9) serves as the stage A model selector. If the number of knots, κ, is such that the inequality in (9) is fulfilled for the first time in the knot addition

*This inequality stems from the fact that at least two values of ϕ_h are required in order to estimate the two coefficients, $\hat{\gamma}_0$ and $\hat{\gamma}_1$, in the exponent. However, one needs at least three values of ϕ_h, $h = q, q+1, q+2$ in order to have a smoothing instead of interpolation of ϕ_h, $h = q, q+1$. See also Remark 1.

process, then $\hat{f}\left(\boldsymbol{\delta}_{\kappa,2}, \hat{\boldsymbol{\alpha}}_p; x\right)$ has not significantly improved adding the last q more knots and therefore, $\hat{f}\left(\boldsymbol{\delta}_{\kappa-q,2}, \hat{\boldsymbol{\alpha}}_p; x\right)$ is the selected model which adequately reproduces the "shape" of the underlying data in the predictor scale.

Remark 1. Let us note that the stopping rule (model selector) in (9) represents an exponentially smoothed version of the rule

$$\phi_h := \frac{D(\hat{\boldsymbol{\alpha}}_{h+2}; h, 2)}{D(\hat{\boldsymbol{\alpha}}_{h+2-q}; h-q, 2)} \geq \phi_{exit} \quad h = q, q+1, q+2, \ldots \tag{10}$$

which is a direct generalization of the rule for the Normal GeDS regression proposed by Kaishev et al. (2016) (c.f. Equation (10) therein). Stopping rule (10) is applied at the first 2 steps in the knot inclusion process of stage A, but can also be used as an alternative to (9). The motivation behind the exponential smoothing of $\{\phi_h, h\}_{h=q}^{\kappa}$ is based on the empirical evidence that ϕ_h in (10), grows like $1-\exp\{-h\}$ as knots are added at appropriate locations (c.f. Steps 3-8 below), i.e. when $h = 1, 2, 3, \ldots$. Applying (9) as an alternative to (10) leads to more stability with respect to the number of inserted knots. This is demonstrated in the online supplement Dimitrova et al. (2022) (see Tables 1 and 2 therein), where stopping rules (9) and (10) are compared and contrasted to a third, alternative stopping rule. The latter is based on the fact that

$$[D(\hat{\boldsymbol{\alpha}}_{h+2-q}; h-q, 2) - D(\hat{\boldsymbol{\alpha}}_{h+2}; h, 2)] \sim \chi^2_q \quad h = q, q+1, q+2, \ldots$$

which follows from the Wilks's theorem, noting that the knot-addition scheme in stage A gives rise to nested linear GeDS models, due to the local properties of B-splines. Therefore, after each iteration in stage A, one may test whether the lastly added q coefficients are significant or not and decide to exit stage A. The latter means testing the null hypotheses, H_0 : *the reduced (simpler) model* $\hat{f}\left(\delta_{h-q,2}, \hat{\boldsymbol{\alpha}}_{h+2-q}; x\right)$ *is true* at significance level $1-\phi_{\text{exit}}$. If the realized difference $\Delta D = D(\hat{\boldsymbol{\alpha}}_{h+2-q}; h-q, 2) - D(\hat{\boldsymbol{\alpha}}_{h+2}; h, 2)$ is such that

$$\Pr\left[\chi^2_q \geq \Delta D\right] \geq 1 - \phi_{\text{exit}}, \tag{11}$$

then we cannot reject the null hypothesis and hence, exit stage A.

Note that on average, the number of knots selected with (11) decreases if ϕ_{exit} increases, in contrast to rules (9) and (10) where the higher ϕ_{exit} the more knots are added before exit.

We finally note also that other (number and location of) knots selection rules can be used as an alternative to the model selector in (9). The latter is well suited for the proposed GeDS GNM/GLM methodology mainly because of the nature of the GeDS knot insertion scheme which ensures consecutive fits are nested. In our previous work Kaishev et al. (2016), we have implemented and compared with e.g. the Stein's unbiased risk estimate (SURE) and the generalized cross-validation (GCV) as alternatives to the rule given by the analog of equation (10) in the Normal case. Furthermore, let us note that an information criterion in the spirit of Akaike (AIC) which combines (9) and (11) can be used as yet another alternative model selector.

3.1.1 Detailed description of stage A

Next we give a more detailed description of the steps of stage A of the generalized GeDS procedure.

Step 0. Let $n = 2$, $k = 0$, $p = n + k = 2$ with initial knot vector $\boldsymbol{\delta}_{0,2} = \{\delta_i\}_{i=1}^4$, such that, $a = \delta_1 = \delta_2 < \delta_3 = \delta_4 = b$. At the initial step of the IRLS procedure, set $l := 0$ and $\hat{\boldsymbol{\alpha}}_p^{(l)} = \left(\hat{\alpha}_1^{(l)}, \ldots, \hat{\alpha}_p^{(l)}\right)^T = (\alpha_1, \ldots, \alpha_p)^T$, where $\alpha_1, \ldots, \alpha_p$ are appropriate initial values and go to Step 1.

Step 1.

1. Evaluate $\hat{\mu}^{(l)}(x_i) = g^{-1}\left(\hat{f}\left(\boldsymbol{\delta}_{k,2}, \hat{\boldsymbol{\alpha}}_p^{(l)}; x_i\right)\right)$, where $\hat{f}\left(\boldsymbol{\delta}_{k,2}, \hat{\boldsymbol{\alpha}}_p^{(l)}; x_i\right) = \sum_{i=1}^p \hat{\alpha}_i^{(l)} N_{i,2}(x_i)$, is the linear spline predictor, fitted at the l-th IRLS iteration, and then calculate the transformed responses

$$g^{(l)}(y_i) \approx z_i^{(l)}(x_i) \quad = \quad g(\hat{\mu}^{(l)}(x_i)) + (y_i - \hat{\mu}^{(l)}(x_i)) g^{'}(\hat{\mu}^{(l)}(x_i)), \tag{12}$$

where, $i = 1, \ldots, N$ and $g^{'}(\mu) = \frac{\partial g(\mu)}{\partial \mu}$.

2. Calculate the weights

$$w_i^{(l)} = 1/Var[z_i^{(l)}|x_i] \;\; = \;\; 1/\left[\left(g^{'}(\hat{\mu}^{(l)}(x_i))\right)^2 Var[Y|x_i]\right]. \tag{13}$$

3. Perform a weighted linear regression of $z_i^{(l)}$ on x_i with weights $w_i^{(l)}$, $i = 1, \ldots, N$, i.e., find $\hat{\boldsymbol{\alpha}}_p^{(l+1)} = \left(\hat{\alpha}_1^{(l+1)}, \ldots, \hat{\alpha}_p^{(l+1)}\right)^T$.

4. Check if

$$\frac{\left|D(\hat{\boldsymbol{\alpha}}_p^{(l)}; k, 2) - D(\hat{\boldsymbol{\alpha}}_p^{(l+1)}; k, 2)\right|}{D(\hat{\boldsymbol{\alpha}}_p^{(l+1)}; k, 2)} \leq d_{irls}, \tag{14}$$

where the deviance $D(\hat{\boldsymbol{\alpha}}_p^{(\cdot)}; k, 2) := D(g^{-1}(f(\boldsymbol{\delta}_{k,2}, \hat{\boldsymbol{\alpha}}_p^{(\cdot)}; x))$ is computed using (4) and d_{irls} is an appropriate threshold level pre-selected in the IRLS estimation.

If (14) is not satisfied then set $l := l + 1$ and go to sub-step 1, otherwise calculate $w_i^{(l+1)}$, using (13), then calculate the weighted residuals

$$r_i = r(x_i) = w_i^{(l+1)} \left(y_i - \hat{\mu}^{(l+1)}(x_i)\right) g^{'}\left(\hat{\mu}^{(l+1)}(x_i)\right)$$

set $\hat{\boldsymbol{\alpha}}_p^{(l+1)} = \hat{\boldsymbol{\alpha}}_p$ and go to Step 2.

Step 2. If $k \geq q$ compute

$$\phi_k := \frac{D(\hat{\boldsymbol{\alpha}}_p^{(l+1)}; k, n)}{D(\hat{\boldsymbol{\alpha}}_{p-q}^{(l+1-q)}; k-q, n)} \tag{15}$$

and if $k \geq q+1$ find the least squares fit (with coefficients $\hat{\gamma}_0$ and $\hat{\gamma}_1$), $\hat{\phi}_h := 1 - \exp\{\hat{\gamma}_0 + \hat{\gamma}_1 \kappa\}$ to the sample $\{\phi_h, h\}_{h=q}^{\kappa}$.

Then, as discussed earlier for the model selector in (9), if

$$\hat{\phi}_k = 1 - \exp\{\hat{\gamma}_0 + \hat{\gamma}_1 k\} \geq \phi_{exit}$$

go to Stage B. In all other cases, move to Step 3.

Steps 3-8 follow correspondingly **Steps 2-7** of GeDS in the Normal case (c.f. Kaishev et al. 2016). In brief, a new knot, δ^*, is added according to a locally-adaptive knot insertion scheme involving a bias driven measure incorporating the weighted residuals, r_i, calculated at Step 1.4.

Step 9. Find i^*, $0 \leq i^* \leq k$ such that $\delta^* \in [\delta_{i^*+2}, \delta_{i^*+3}]$, set $l := 0$ and the initial values for the parameters, $\hat{\boldsymbol{\alpha}}_{p+1}^{(l)} = \left(\hat{\alpha}_1, \ldots, \hat{\alpha}_{i*+1}, \hat{f}\left(\boldsymbol{\delta}_{k,2}, \hat{\boldsymbol{\alpha}}_p; \delta^*\right), \hat{\alpha}_{i*+3}, \ldots, \hat{\alpha}_{p+1}\right)$ then set $p = p + 1$ and $k = k + 1$ and go to Step 1.

3.2 Detailed description of stage B

Let us now detail stage B of the generalized GeDS, which as in the Normal case consists of two parts B1 and B2.

Stage B1. Given the (final) fit $\hat{f}\left(\boldsymbol{\delta}_{\kappa,2}, \hat{\boldsymbol{\alpha}}_p; x\right)$, from stage A with κ internal knots, for each $n = 3, \ldots, n_{\max}$, calculate the knot placement $\overline{\mathbf{t}}_{\kappa-(n-2),n}$ with $\kappa - (n-2)$ internal knots defined as the averages of the knots $\boldsymbol{\delta}_{\kappa,2}$, i.e.

$$\bar{t}_{i+n} = (\delta_{i+2} + \cdots + \delta_{i+n})/(n-1), \qquad i = 1, \ldots, \kappa - (n-2). \tag{16}$$

As has been demonstrated in Kaishev et al. (2016), the choice of the knots $\overline{\mathbf{t}}_{\kappa-(n-2),n}$, according to (16), ensures that the n-th order spline predictor curve $f\left(\overline{\mathbf{t}}_{\kappa-(n-2),n}, \hat{\boldsymbol{\alpha}}_p; x\right)$ becomes nearly the VDS approximation to the fit, $\hat{f}\left(\boldsymbol{\delta}_{\kappa,2}, \hat{\boldsymbol{\alpha}}_p; x\right)$, from stage A, i.e. closely follows its shape, as explained in Section 2.2 (see Figure 1). Therefore the fit, $\hat{f}\left(\boldsymbol{\delta}_{\kappa,2}, \hat{\boldsymbol{\alpha}}_p; x\right)$ can be viewed as (nearly) the control polygon of the predictor curve $f\left(\overline{\mathbf{t}}_{\kappa-(n-2),n}, \hat{\boldsymbol{\alpha}}_p; x\right)$. Error bounds for this VDS approximation and the optimality properties of the knots are derived and discussed in Kaishev et al. (2016).

Let us note that, although the spline predictor curve $f\left(\overline{\mathbf{t}}_{\kappa-(n-2),n}, \hat{\boldsymbol{\alpha}}_p; x\right)$ obtained at stage B1, closely follow the shape of $\hat{f}\left(\boldsymbol{\delta}_{\kappa,2}, \hat{\boldsymbol{\alpha}}_p; x\right)$ and hence, the data, formally it is not a maximum likelihood estimate to the data $\{y_i, x_i\}_{i=1}^N$. In order to correct this and slightly adjust its position, in stage B2 its knots $\overline{\mathbf{t}}_{\kappa-(n-2),n}$ are preserved but its B-spline coefficients, $\hat{\boldsymbol{\alpha}}_p$, are treated as unknown parameters, denoted by $\boldsymbol{\theta}_p$, $p = \kappa+2$, which are then estimated in a final run of the IRLS procedure as follows.

Stage B2. For each fixed $n = 3, \ldots, n_{\max}$, find the ML estimates $\hat{\boldsymbol{\theta}}_p$ of the B-spline coefficients, $\boldsymbol{\theta}_p$, of the spline predictor curve $f\left(\overline{\mathbf{t}}_{\kappa-(n-2),n}, \boldsymbol{\theta}_p; x\right)$ from stage B1. For the purpose, set $l = 0$ and run the IRLS procedure similarly as in Step 1 of stage A, but with respect to the vector $\boldsymbol{\theta}_p$. More precisely, start by calculating $\hat{\mu}^{(l)}(x_i) = g^{-1}\left(\hat{f}\left(\boldsymbol{\delta}_{\kappa,2}, \hat{\boldsymbol{\alpha}}_p^{(l)}; x_i\right)\right)$, and

the transformed responses and weights, substituting $\hat{\mu}^{(l)}(x_i)$ in (12) and (13) respectively. Then perform a weighted linear regression of $z_i^{(l)}$ on x_i with weights $w_i^{(l)}$, $i = 1, \ldots, N$, to find $\hat{\boldsymbol{\theta}}_p^{(l+1)} = \left(\hat{\theta}_1^{(l+1)}, \ldots, \hat{\theta}_p^{(l+1)}\right)^T$. Similarly as in (14) check the inequality

$$\frac{\left|D(\hat{\boldsymbol{\theta}}_p^{(l)}; k, 2) - D(\hat{\boldsymbol{\theta}}_p^{(l+1)}; k, 2)\right|}{D(\hat{\boldsymbol{\theta}}_p^{(l+1)}; k, 2)} \leq d_{irls},$$

and if it is not satisfied then set $l := l + 1$ and perform the next IRLS iteration, otherwise exit with a final estimate $\hat{\boldsymbol{\theta}}_p$.

Among all fits $\hat{f}\left(\overline{\mathbf{t}}_{\kappa-(n-2),n}, \hat{\boldsymbol{\theta}}_p; x\right)$, of order $n = 2, \ldots, n_{\max}$, i.e. including the linear fit, $\hat{f}\left(\boldsymbol{\delta}_{\kappa,2}, \hat{\boldsymbol{\alpha}}_p; x\right)$ from stage A, choose the one of order $\hat{n}$, for which the deviance computed as in (4) is minimal.

In this way in stage B2, along with the number of knots and their locations, the order of the spline predictor is also estimated. This is an important feature of the proposed GeDS estimation method which is rarely offered by other spline GLM estimation procedures. Of course, any of the produced final fits of order $n \neq \hat{n}$ could be used, if other features were more desirable, for example if better smoothness were required.

4 Numerical study

In this section, we illustrate how the GeDS method, implemented in the **GeDS** package, can be applied to fit simulated and real data, assuming the response variable has a distribution from the EF. Without loss of generality, we simulate/use Normal, Gamma, Poisson and Binomial data in order to demonstrate the numerical properties of GeDS. Based on such data, we have also thoroughly compared the linear, quadratic and cubic GeDS fits to the GAM, GSS and SPM fits produced by the corresponding **R** packages **mgcv**, **gss** and **SemiPar**. As mentioned in the introduction, in the Normal case only, we also compare with the adaptive P-splines of Yang and Hong (2017) and the **trendfilter** fitting method in the package **genlasso** of Arnold and Tibshirani (2020)).

In the online supplement to this paper (see Dimitrova et al. 2022), we have also performed a thorough simulation study of the impact on the GeDS knot location and related

L^1 error, (i.e., distance to a "true" predictor function), of different assumptions and choices made, namely of: sample size ($N = 180, 500$); level of smoothness of the underlying function (smooth, medium-smooth and wiggly functions); value of the GeDS tuning parameter β; alternative model selection criteria (i.e. different stopping rules). We have tested GeDS on a series of simulated examples based on functions adopted in many other studies on variable knot spline methods (c.f. Schwetlick and Schütze 1995, Fan and Gijbels 1995, Donoho and Johnstone 1994, Luo and Wahba 1997 and more recently Kaishev et al. 2016) appropriately revised in order to be applicable in the GLM framework. These functions are used in order to compare GeDS to the alternative methods such as GAM, GSS and SPM and the corresponding results are presented in the online supplement Dimitrova et al. (2022).

In the next Section 4.1 we present the results for a test function, first considered by Schwetlick and Schütze (1995) (see Example 1 below), and we also compare GeDS with alternative (spline) estimators on the well known (simulated) Doppler example (see Example 2 below). In Section 4.2, we consider two real data examples, namely Example 3 based on data from materials science (c.f. Kimber et al. 2009) and Example 4 based on coal mining data (c.f. Eilers and Marx 1996).

4.1 Simulated examples

4.1.1 First simulated example

In this section, we illustrate the use of the GeDS regression and the properties of the GeDS fits compared to fits produced by alternative spline based GLM methodologies such as SPM, GAM and GSS. We start with the following simulated test example.

Example 1. We assume the "true" linear predictor is $\eta = f_1(x)$, where

$$f_1(x) = 40\frac{x}{1+100x^2} + 4, \qquad x \in [-2, 2]. \tag{17}$$

We have then generated random samples, $\{X_i, Y_i\}_{i=1}^N$, with correspondingly Poisson, Gamma, Normally and Binomially distributed response variable, y, and uniformly distributed explanatory variable, x, i.e.: $Y_i \sim \text{Poisson}(\mu_i)$, $Y_i \sim \text{Gamma}(\mu_i, \varphi)$ with $\varphi = 0.1$, $\mu_i = \exp\{\eta_i\}$

and $\eta_i = f_1(X_i)$; $Y_i \sim N(\mu_i, \sigma)$ with $\sigma = 0.2$, $\mu_i = \eta_i = f_1(X_i)$; $Y_i \sim \text{Binomial}(m, \mu_i)$ with $m = 50$, $\mu_i = \exp\{\eta_i\}/\left(1 + \exp\{\eta_i\}\right)$, $\eta_i = f_1(X_i) - 4$; and $X_i \sim U\left[-2, 2\right]$, $i = 1, \ldots N$, where N is the sample size ($N = 500$ and $N = 180$ are used in the comparison).

Based on Example 1, we compare the fits produced by GeDS with those obtained from the alternative estimators, GSS (package **gss**, function **gssanova**), GAM (package **mgcv**, function **gam**) and SPM (package **SemiPar**, function **spm**), and additionally, in the Normal case only, with the method of Yang and Hong (2017) (coded as Y&H; the software is provided by the authors as a supplementary material to their paper). For the purpose of comparison, for all four distributions, i.e., Poisson, Binomial, Gamma and Normal, we have run GeDS and the alternative fitting procedures 1000 times each, fitting them to 1000 data samples, generated as described in Example 1 with two sample sizes, $N = 180, 500$. The tuning parameter of the stopping criterion (rule) of stage A of GeDS, ϕ_{exit}, in all four cases was set to $\phi_{\text{exit}} = 0.995$. The values of β (the second tuning parameter in stage A, c.f. Kaishev et al. 2016) are $\beta = 0.2$ for the Poisson case, $\beta = 0.1$ for the Binomial and Gamma cases (these were chosen following the results from the sensitivity study given in Section 1 of the online supplement to this paper), and $\beta = 0.5$ for the Normal case (the default value suggested by Kaishev et al. 2016). The alternative methods were also first run with the default values of their corresponding tuning parameters and then all the methods, including GeDS, were run with appropriately tuned parameters. In **gss** the tuning parameter is α, and it tunes the stopping rule based on the GCV functional

$$GCV(\lambda) = \frac{n^{-1} Y^T (I - A(\lambda))^2 Y}{\left[n^{-1} tr(I - \alpha A(\lambda))\right]^2},$$

as defined in Gu (2014). By default $\alpha = 1$ for the binomial case, while $\alpha = 1.4$ otherwise. In **spm** the tuning parameter is directly the smoothing parameter λ. By default its value is automatically selected by the underlying REML procedure. For GAM the tuning parameters considered are the number of knots (the dimension of the basis) which, by default, is selected internally, and the adaptive smoothing parameter which by default is a global constant selected as part of the GCV criterion.

The results of these comparisons in the case of $N = 500$ are presented in Figs 2, 3, 4, 5 and 6 where, in each figure, panel (a) presents sample GeDS fits versus the true predictor

function $f_1(x)$; panel (b) presents sample curves of the alternative models versus $f_1(x)$; panel (c) - histogram of the number of internal knots, κ, of the linear GeDS fit; panel (d) - box plots of the L^1 distance, $\|f_1 - \hat{f}_1\|_1 = \int_{-2}^{2} |f_1(x) - \hat{f}\left(\overline{\mathbf{t}}_{\kappa-(n-2),n}, \hat{\boldsymbol{\theta}}_p; x\right)| dx$, to $f_1(x)$. Related numerical results are summarized in Tables 1 and 2. As can be seen looking at panels (a), in all four cases GeDS produces linear ($n = 2$), quadratic ($n = 3$) and cubic ($n = 4$) fits, $\hat{f}\left(\overline{\mathbf{t}}_{\kappa-(n-2),n}, \hat{\boldsymbol{\theta}}_p; x\right)$ of remarkable quality, as confirmed visually but also based on their L^1 distance summarized in panel (d). Furthermore, as can also be seen from panels (a), (b) and (d), and from the mean and median of the L^1 distance to the true predictor summarized in Table 1, for all four distributions the GeDS fits (of order 2, 3 and 4) outperform those produced by the alternative methods. The latter tends to wiggle around the true predictor function, with GAM more significantly away from f_1, especially around the origin. Overall for all four distributions, the quadratic GeDS performs better than the linear and cubic ones based on the corresponding mean L^1 values (c.f. Table 1), with the median L^1 for the cubic GeDS slightly better in the Normal and Poisson cases.

		GeDS(n = 2)	GeDS(n = 3)	GeDS(n = 4)	GAM	SPM	GSS
Normal	Mean	0.1588	**0.1342**	0.1398	0.8260	0.2131	0.2597
	Median	0.1554	0.1296	**0.1266**	0.8125	0.2100	0.2382
Poisson	Mean	0.1347	**0.1144**	0.1159	0.9750	0.1921	0.2335
	Median	0.1311	0.1058	**0.1026**	0.9377	0.1868	0.1944
Gamma	Mean	0.2396	**0.2174**	0.2699	0.8652	NA	0.3352
	Median	0.2239	**0.1928**	0.2277	0.8484	NA	0.3200
Binomial	Mean	0.2512	**0.2328**	0.3055	0.8137	0.2869	0.7294
	Median	0.2455	**0.2262**	0.2715	0.8002	0.2853	0.7292

Table 1: Example 1 - Mean and median L^1 distance to the true predictor function.

The distributions of the number of knots, κ (in general $\kappa - (n-2)$, with $n = 2$ for the linear GeDS fit), in panels (c) of Figs 2, 3, 4 and 5 are reasonably compact suggesting that the stopping rule, (9), is stable and consistently selects κ. This conclusion is also supported looking at the descriptive characteristics (mean, median, standard deviation and range) of the corresponding distributions of κ, summarized in Table 2. It should also be noted that in all four cases, i.e., Poisson, Binomial, Normal and Gamma samples, GeDS regression produces fits with relatively small number of estimated regression coefficients, $\kappa-(n-2)+n$, which on average is equal to $16+2$, $12+2$, $14+2$ and $11+2$ for the linear ($n = 2$) GeDS fit

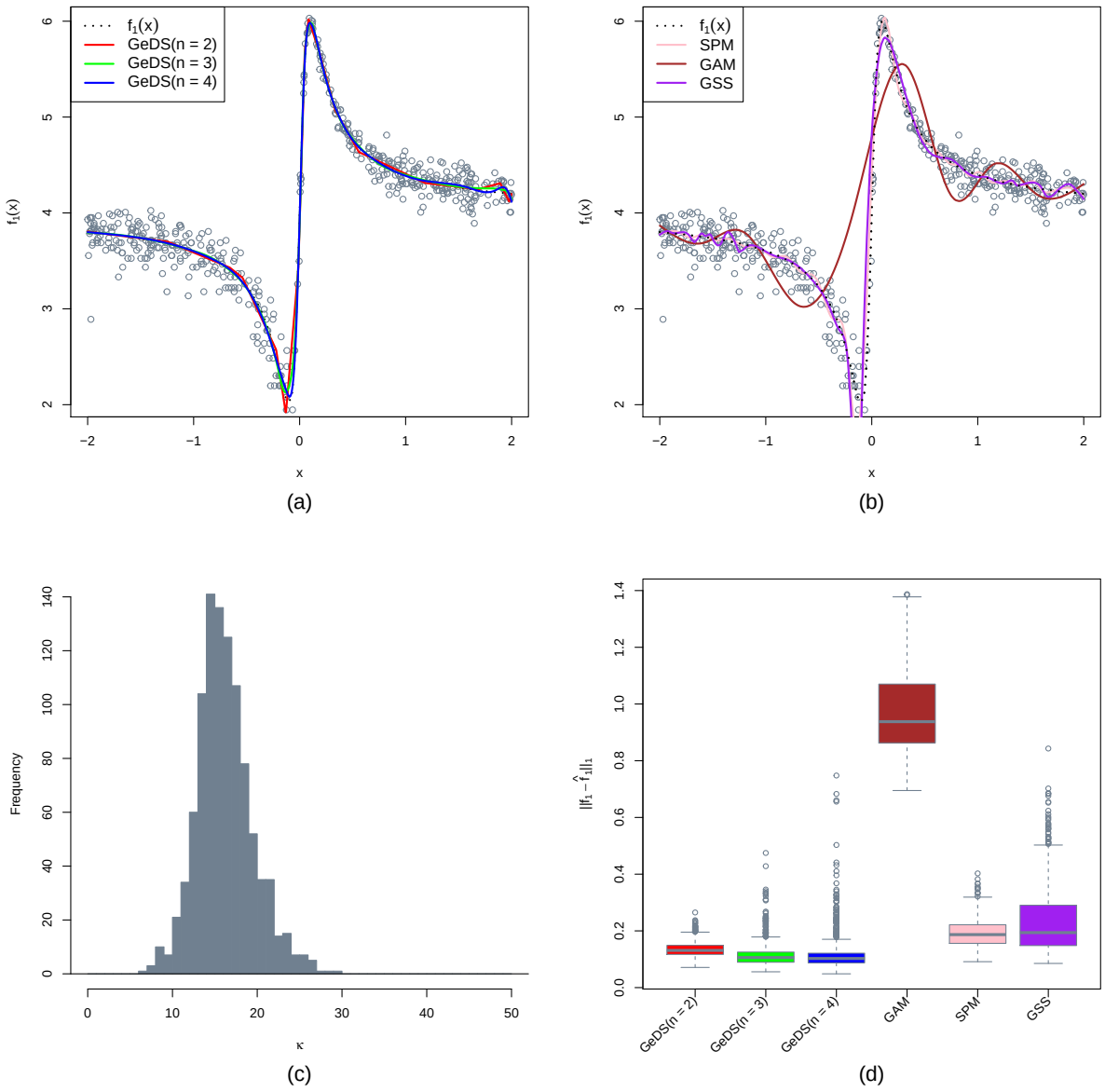


Figure 2: Comparison of the linear ($n = 2$), quadratic ($n = 3$) and cubic ($n = 4$) **GeDS** fits with the **mgcv**, **SemiPar** and **gss** models (on the predictor scale, with "true" predictor function $f_1(x)$), based on fitting 1000 Poisson samples (empty circles), generated according to Example 1.

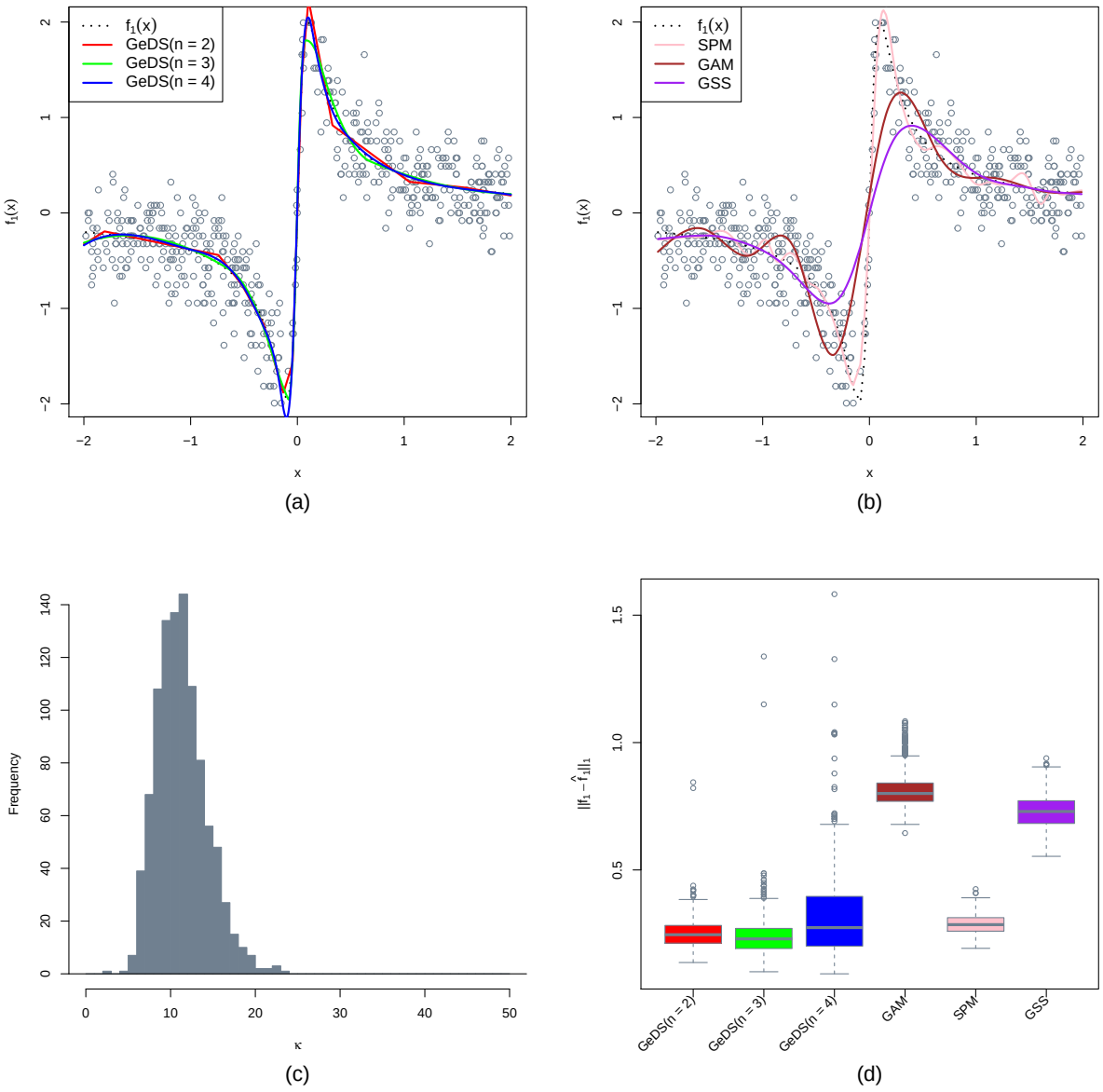


Figure 3: Comparison of the linear ($n = 2$), quadratic ($n = 3$) and cubic ($n = 4$) **GeDS** fits with the **mgcv**, **SemiPar** and **gss** models (on the predictor scale, with "true" predictor function $f_1(x)$), based on fitting 1000 Binomial samples (empty circles), generated according to Example 1.

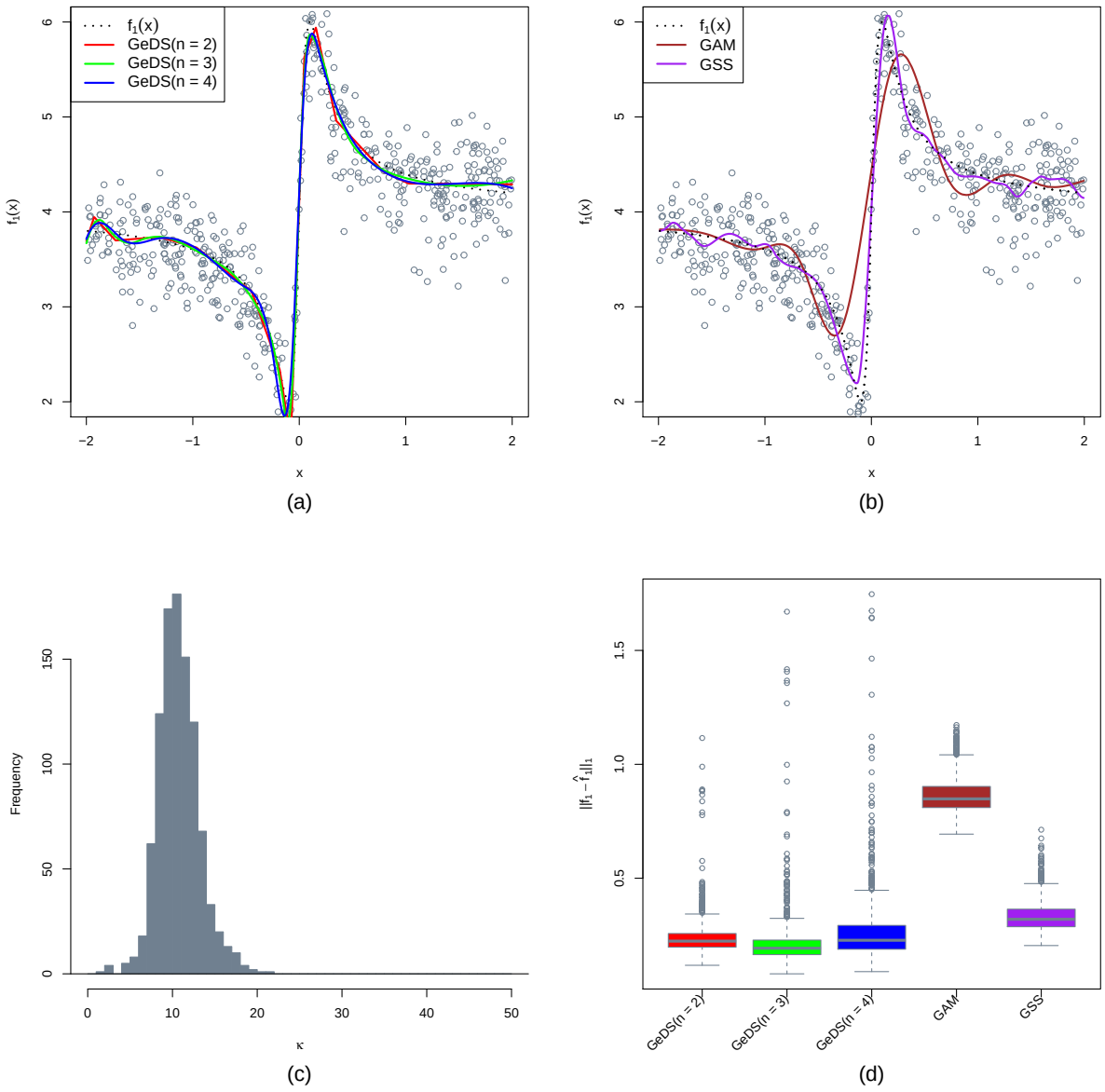


Figure 4: Comparison of the linear ($n = 2$), quadratic ($n = 3$) and cubic ($n = 4$) **GeDS** fits with the **mgcv** and **gss** models (on the predictor scale, with "true" predictor function $f_1(x)$), based on fitting 1000 Gamma samples (empty circles), generated according to Example 1. Note: **SemiPar** does not handle Gamma samples.

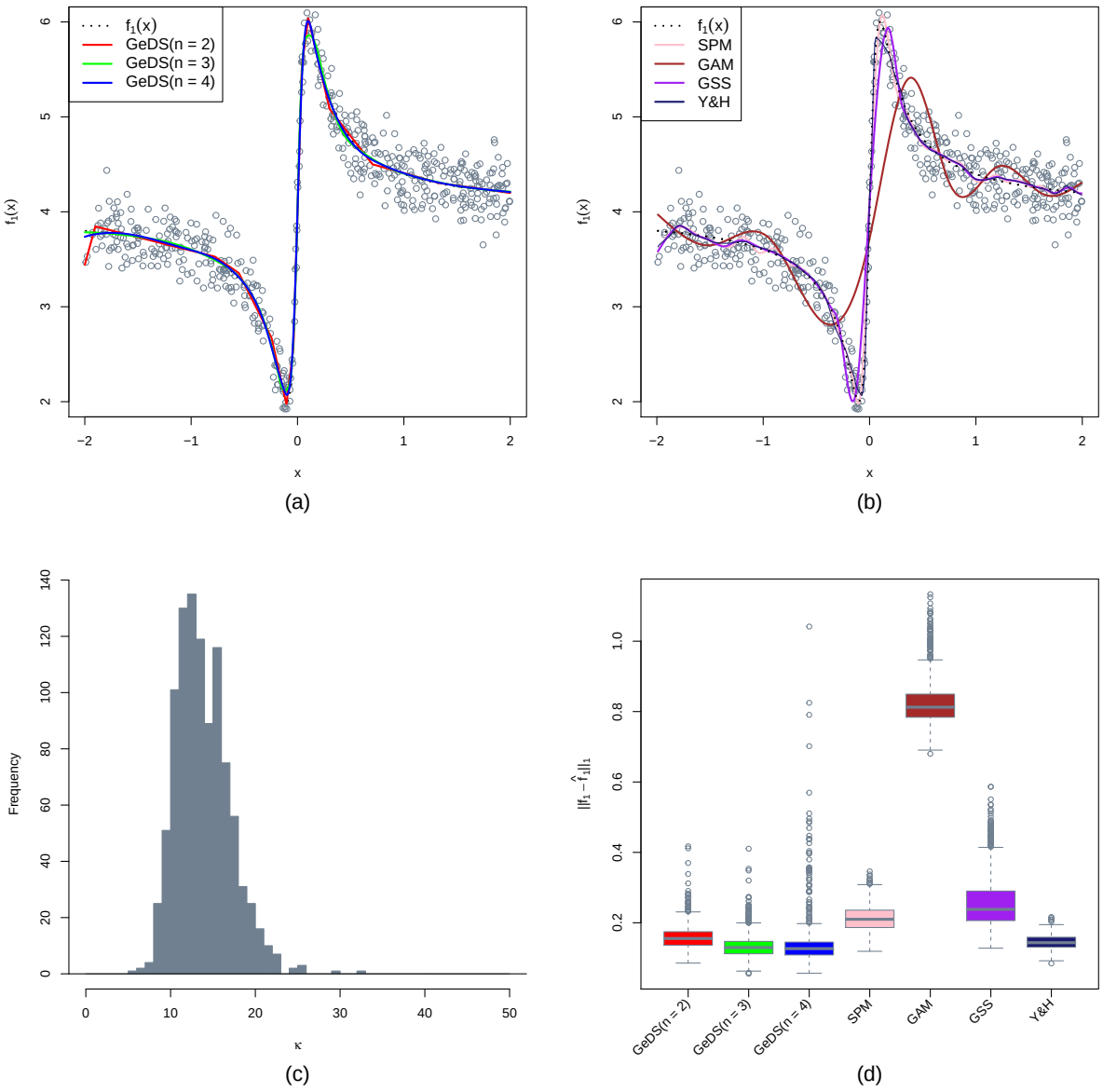


Figure 5: Comparison of the linear ($n = 2$), quadratic ($n = 3$) and cubic ($n = 4$) **GeDS** fits with the **mgcv**, **SemiPar**, **gss** and Y&H models (on the predictor scale, with "true" predictor function $f_1(x)$), based on fitting 1000 Normal samples (empty circles), generated according to Example 1.

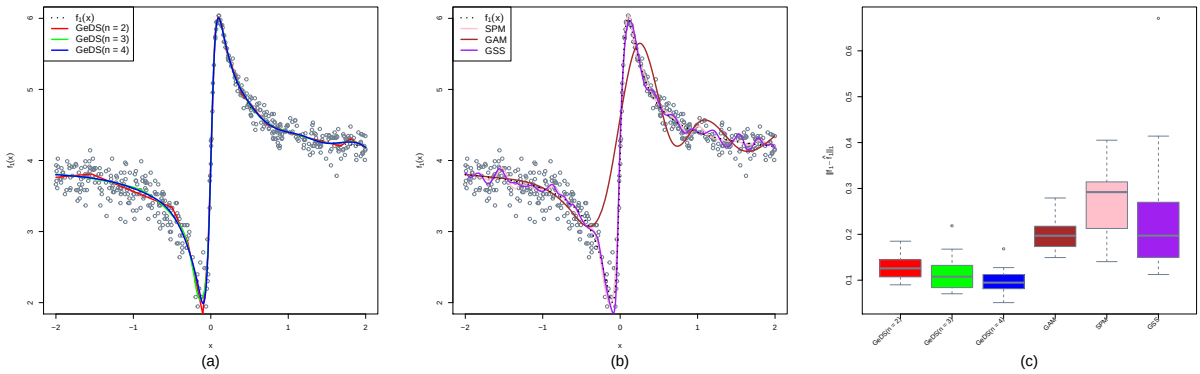


Figure 6: Comparison of the linear ($n = 2$), quadratic ($n = 3$) and cubic ($n = 4$) **GeDS** fits with the **mgcv**, **SemiPar** and **gss** models (on the predictor scale, with "true" predictor function $f_1(x)$), based on fitting 30 Poisson samples (empty circles), adjusting the corresponding tuning parameters.

(c.f. Table 2) and is the same (i.e. $\kappa + 2$) for the GeDS fits of higher order (see stage B in Section 3.2). Indicatively, the average reported degrees-of-freedom for the GAM models is about 10 in all four cases, and for the SPM about 33 in the Normal and Poisson case and about 16 in the Binomial case. On the next Example 2 presented in Section 4.1.2 below, we perform a thorough investigation of the degrees-of-freedom resulting from the GeDS fits and the alternative models where the advantage of GeDS is clearly demonstrated.

	Normal	Poisson	Gamma	Binomial
Median	14.00	16.00	11.00	12.00
Mean	14.35	16.70	11.26	11.93
SD	3.21	3.31	2.46	3.01
Min	6.00	7.00	2.00	3.00
Max	33.00	30.00	22.00	26.00

Table 2: Example 1 - Descriptive statistics for the distribution of number of internal knots κ.

In addition, here we have also compared the performance of GeDS and the alternative GAM, SPM, GSS models by fitting them to 30 Poisson samples, in each case appropriately adjusting the corresponding tuning parameters described earlier. As can be seen from Fig. 6, which presents the result of this comparison, GAM and GSS have significantly improved, although as before, GeDS has remained superior after adjusting the tuning parameters for all six models (GeDS, $n = 2, 3, 4$, GAM, SPM, GSS).

Lastly we note that the results obtained for the case of smaller sample size $N = 180$ are similar to those obtained for $N = 500$, namely GeDS produces good quality fits with low number of knots, compactly distributed around the mean. This suggests that GeDS performance is reasonably robust with respect to the sample size (as was also demonstrated by Kaishev et al. 2016 in the Normal case). For brevity, the graphical illustrations, detailed statistics and related conclusions for $N = 180$ are omitted.

Results of further tests including sensitivity of GeDS with respect to the choice of the stopping rule and the tuning parameters β, ϕ_{exit} and further comparisons with the alternatives GAM, SPM and GSS, based on four additional test functions commonly used in the spline literature are presented in the online supplement (see Dimitrova et al. 2022).

4.1.2 Second simulated example

Here, we present another simulated test example on the Doppler function which will be used to study the numerical performance and in particular, the (effective) degrees-of-freedom of GeDS and the alternatives as another way of measuring the efficiency of each method/model. The Doppler function is highly oscillating, especially near the origin, where most of the methods/models fail to recover it.

Example 2. We assume that the "true" linear predictor, $\eta = f_2(x)$, where

$$f_2(x) = 5\sqrt{x(1-x)}\sin\frac{2\pi(1+0.05)}{(x+0.05)}, \qquad x \in [0,1], \tag{18}$$

is the well-known Doppler function used by a number of authors, including Kaishev et al. (2016) (see the online supplement to the latter) and Yang and Hong (2017), to test their proposed methods in the Normal case. We have simulated random samples, $\{X_i, Y_i\}_{i=1}^N$, with Normally distributed response variable, y, and the explanatory variable, x, is on a uniform grid in $[0,1]$, i.e. $Y_i \sim N(\mu_i, \sigma)$ with $\sigma = 0.2$, $\mu_i = \eta_i = f_2(X_i)$; and $x_i = (i-1)/N$, $i = 1, \dots N$, where the sample size is $N = 400$.

Fig 7 presented next is analogues to Fig 4 of Yang and Hong (2017). It illustrates the Doppler output fits of GeDS and the alternatives based on one Normal data sample. It it clear that the GSS (**gss**) and SPM (**SemiPar**) fits are underperforming due to lacking

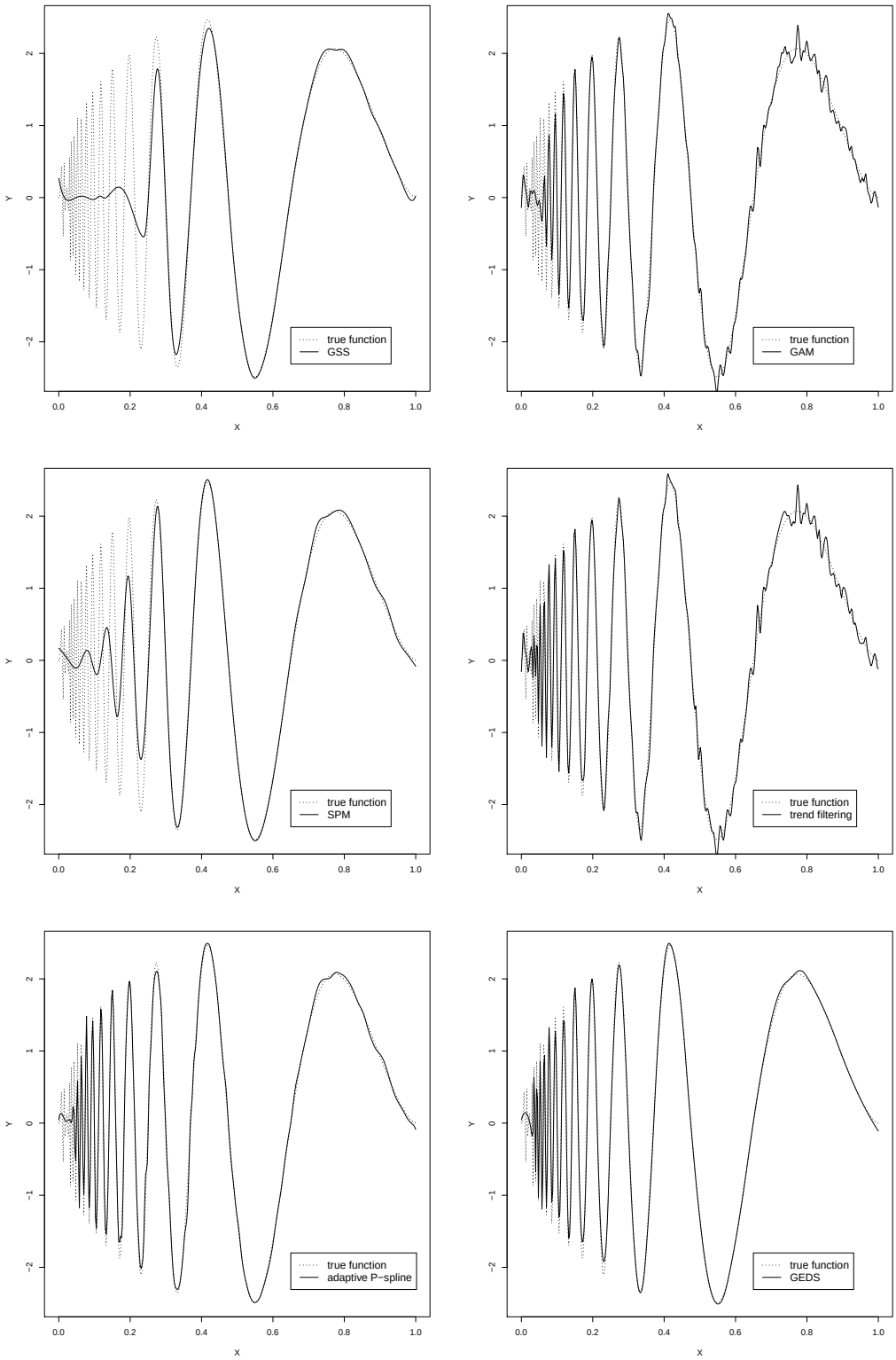


Figure 7: Comparison of the quadratic ($n = 3$) **GeDS** fit with the **gss**, **mgcv**, **SemiPar**, **trendfilter** method of Arnold and Tibshirani (2020)) and the adaptive P-splines of Yang and Hong (2017) (Y&H) (on the predictor scale, with "true" predictor function $f_2(x)$), based on one Normal sample, generated according to Example 2.

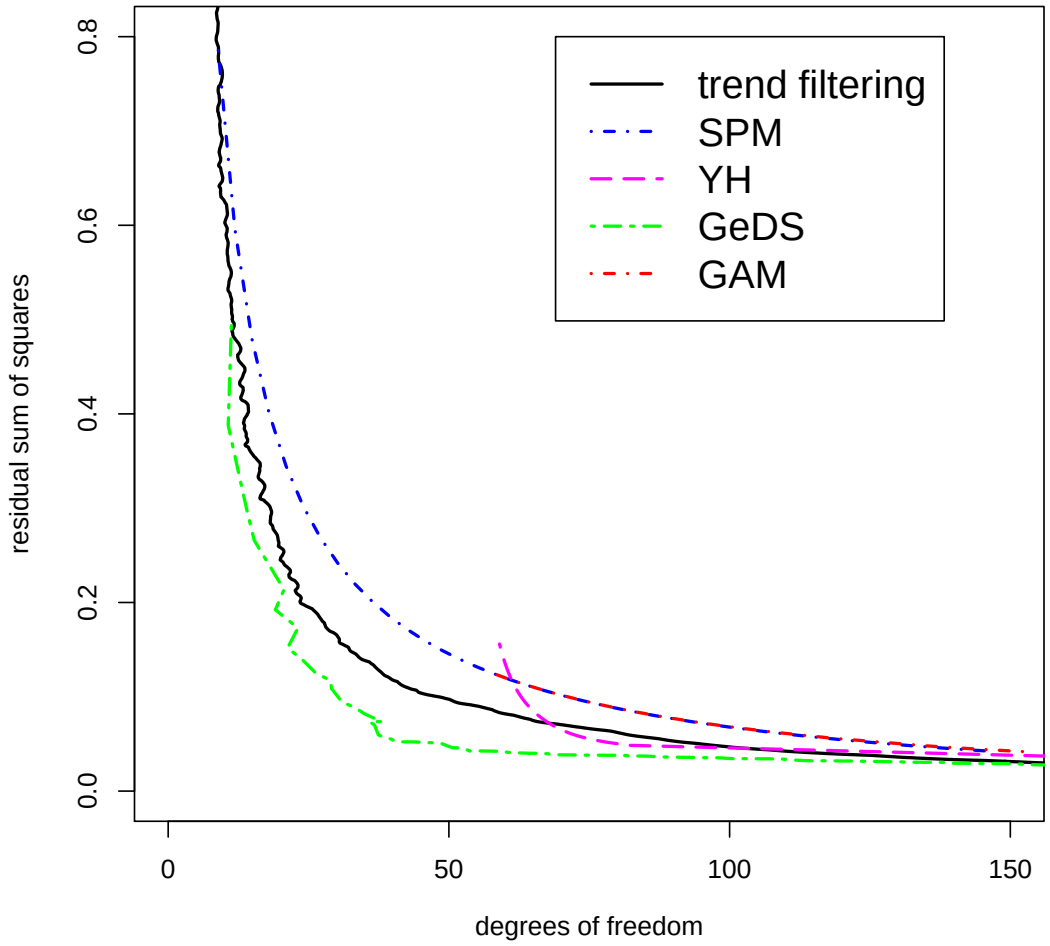


Figure 8: The residual sum of squares against the degrees-of-freedom for the quadratic ($n = 3$) **GeDS** fit, **mgcv**, **SemiPar**, **trendfilter** method of Arnold and Tibshirani (2020)) and the adaptive P-splines of Yang and Hong (2017) (Y&H), based on the results for 100 perturbed Normal samples (Example 2). The (generalized) degrees-of-freedom are calculated applying Ye (1998)'s approach.

local adaptivity, whereas in the GAM (**mgcv**) and **trendfilter** estimates some overfitting is observed in the less oscillating part of the function. The GeDS model together with the adaptive P-splines model of Yang and Hong (2017) are fitting the Doppler function well even close to the origin, where GeDS is slightly better (see the bottom panels of Fig 7).

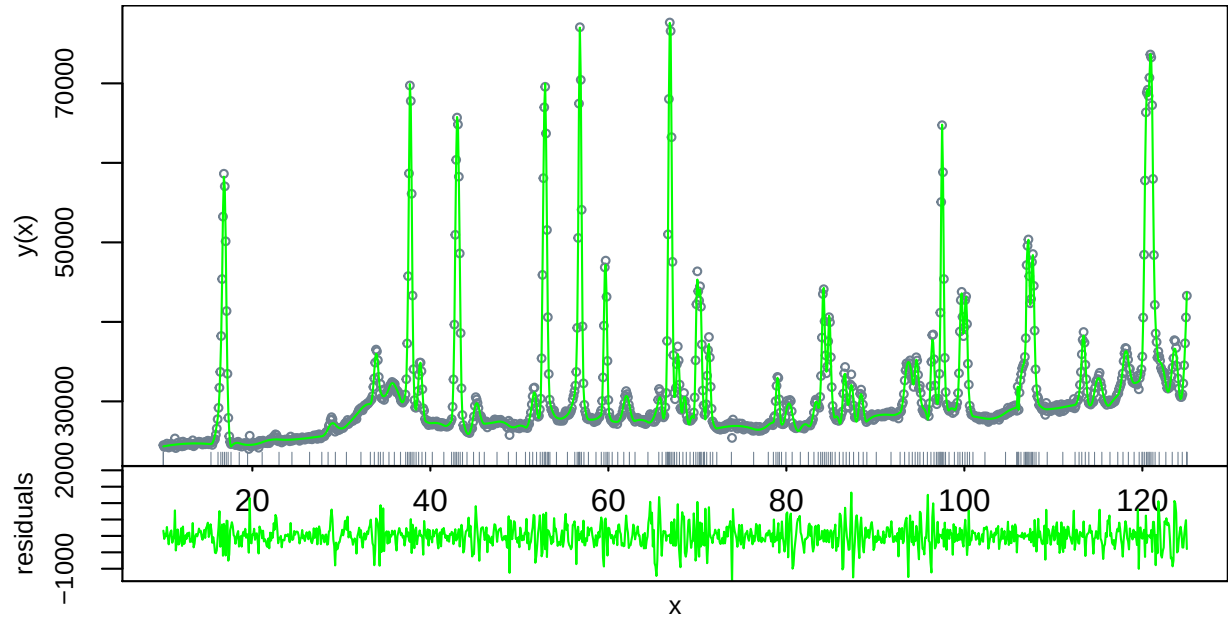
Next, we employ the concept of generalized degrees of freedom as proposed by Ye (1998) in order to make an objective comparison between GeDS and the alternative models in terms of complexity (see also Shen and Ye 2002, Hansen and Sokol 2014, as well as Tibshirani 2014 and Yang and Hong 2017). In Fig 8, we give the residual sum of squares plotted against the degrees-of-freedom for the quadratic ($n = 3$) **GeDS** fit, **mgcv**, **SemiPar**, **trendfilter** method of Arnold and Tibshirani (2020)) and the adaptive P-splines of Yang and Hong (2017) (Y&H), based on the results for 100 perturbed Normal samples following Ye (1998) (see Section 2 therein). As can be seen, the GeDS method delivers better (numerical) performance with fewer parameters when compared to the alternatives, including the Y&H adaptive P-splines. The latter is also confirmed on the real data examples presented in the next Section 4.2.

4.2 Real data examples

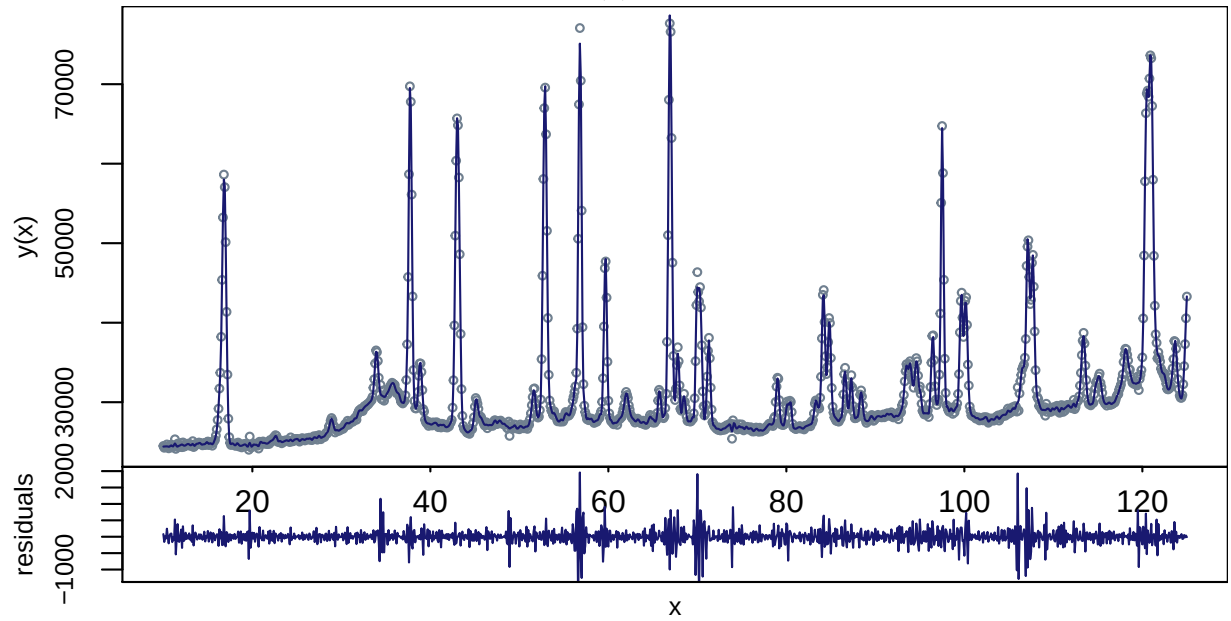
In what follows, we illustrate the performance of GeDS on two real data examples from materials science and coal mining.

Example 3. Our first data set originates from a superconductivity study of Barium-Ferrum-Arsenide ($BaFe_2As_2$) through a neutron diffraction experiment, carried out by Kimber et al. (2009). The data has been used by Kaishev et al. (2016) to illustrate GeDS in the Normal case and here, in addition, we will consider the Gamma case. It includes the sample $\{x_i, y_i\}_{i=1}^N$ of $N = 1151$ observations of the neutron diffraction intensity, y_i viewed as a function of the angle of dispersion of the neutrons, x_i. As can be seen from Fig. 9, this functional dependence is highly non-linear with numerous intensity peaks occurring at certain angle values. Smoothing out the noise while at the same time adequately capturing the peaks is of utmost importance since, as highlighted by Kimber et al. (2009), their location, height (and area) carries information about the structural (conductivity) properties of the $BaFe_2As_2$ superconductor. For a more detailed description of the dataset and the

experiment we refer to Kaishev et al. (2016) and Kimber et al. (2009).



(a)



(b)

Figure 9: Estimated fits to $BaFe_2As_2$ data and their corresponding residuals, assuming Normally distributed response variable: (a) Quadratic GeDS fit; (b) Y&H fit.

Assuming a Normally distributed response variable, the number of knots estimated by GeDS (with tuning parameters $\phi_{exit} = 0.99$, $\beta = 0.6$) is 227 ($n = 3$) and the square root

of the sum of squared residuals[†] is 11653.99. As can be seen from Fig. 9 (a), the GeDS fit, obtained in 9.6 sec, captures really well the peaks of the response variable and at the same time is nicely fitting the flat segments. We have also run all alternative models mentioned above, with default as well as tuned values of the parameters, and the majority significantly underfit the data with only the adaptive P-spline model of Yang and Hong (2017) (Y&H) producing an acceptable output (in about 600 sec), see Fig. 9 (b). The latter captures well the peaks of the data but a closer look reveals substantial overfitting in the flat segments. This is also confirmed by the fact that the minimum of the stopping criterion is achieved for a value of the smoothing parameter, λ, very close to zero and the reported degrees of freedom are 515.69. In fact, as λ is close to zero, the Y&H fit is a spline regression with 515 knots which is over twice as many knots as for the GeDS fit in Fig. 9 (a).

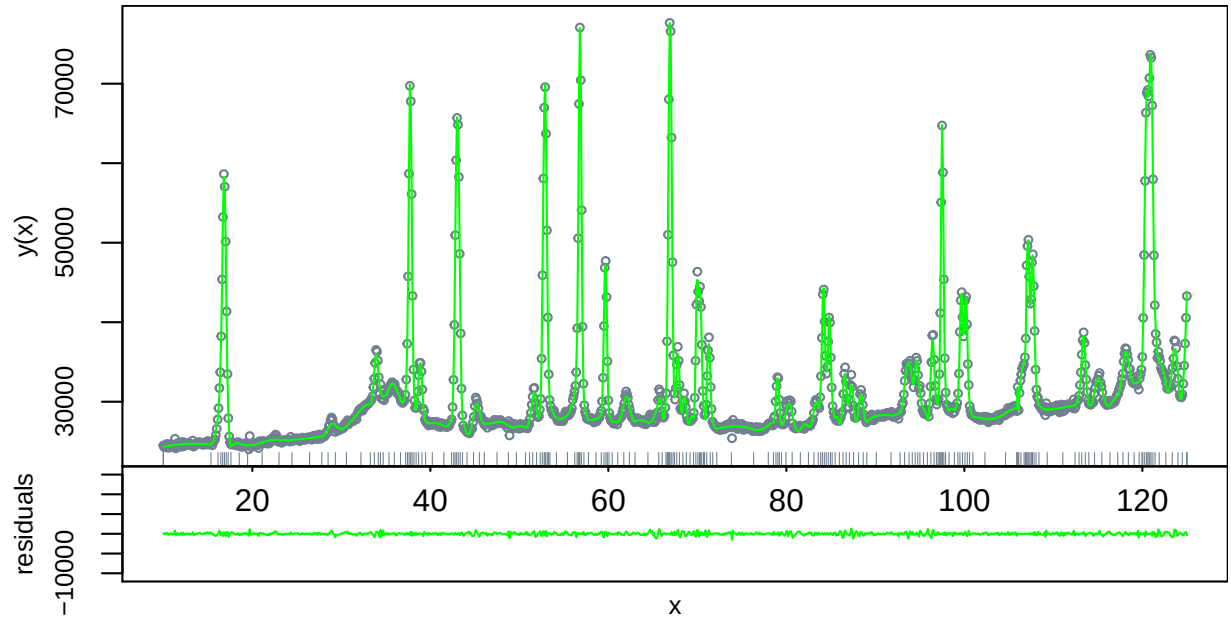
In addition, we fit the same dataset assuming that the response variable follows a Gamma distribution. This seems reasonable as it is continuous and positive with likely dependent mean and variance, as suggested by the data. We use the log link function as it provides more numerical stability than the canonical one. The resulting quadratic GeDS fit with 238 knots (assuming the ratio of deviances stopping rule in (10) and tuning parameters $\phi_{exit} = 0.995$, $\beta = 0.6$ and $q = 3$) is obtained in 15.9 sec and is illustrated in panel (a) of Fig. 10. As can be seen, comparing the plotted residuals, it is of similar high quality as the Normal fit. We have also tested the alternative models (GAM, GSS) and obtained a somewhat acceptable GAM model by specifying the dimension of the basis (number of coefficients) to be 241 (i.e. the same as in the quadratic GeDS spline fit $238 + 3$). However, comparing the corresponding residual plots, one can see that the GAM fit (obtained in 6.6 sec), illustrated in panel (b), is clearly inferior to that of the GeDS fit illustrated in panel (a) of Fig. 10 and the GAM fit also has the tendency to overfit the flat segments of the data.

It should also be noted that on this real data example we have tested tuning the parameter β increasing it from 0.1 to 0.6 which has significantly improved the quality of the fit (respectively increasing the number of knots). One can therefore conclude that for very wiggly functions, with sharp peaks, it is natural to put more weight on the residual height,

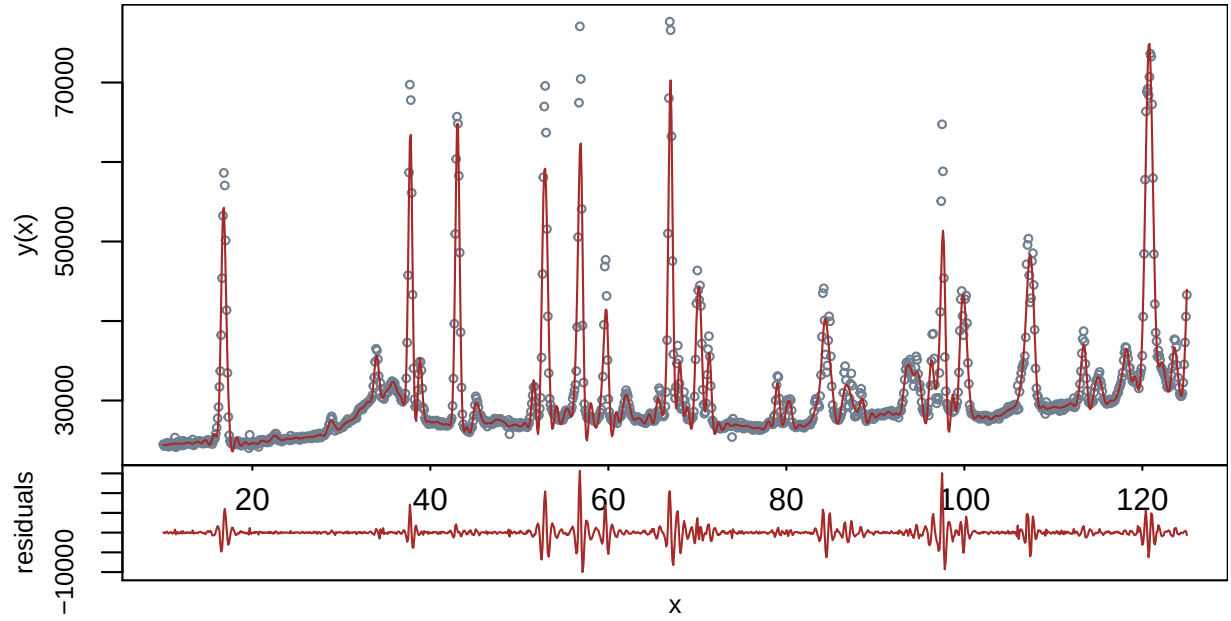
[†]Note that the value reported in Kaishev et al. (2016) (see Section 4.2 therein) is 10847 which is obtained by excluding a few end-point observations, as was done by Kimber et al. (2009) for the Rietveld fit.

i.e. select $\beta = 0.5, 0.6$, rather than on the residual range (c.f. Kaishev et al. 2016).

Overall, given the sharply peaking (unsmooth) nature of the underlying dependence, in both the Normal and Gamma case GeDS performs very well, capturing the required locations and heights of the peaks, without oversmoothing flat regions.



(a)



(b)

Figure 10: Estimated fits to $BaFe_2As_2$ data and their corresponding residuals, assuming Gamma distributed response variable: (a) Quadratic GeDS fit; (b) GAM fit.

Example 4. In our next example we fit GeDS to the coal mining data from Eilers and Marx (1996).

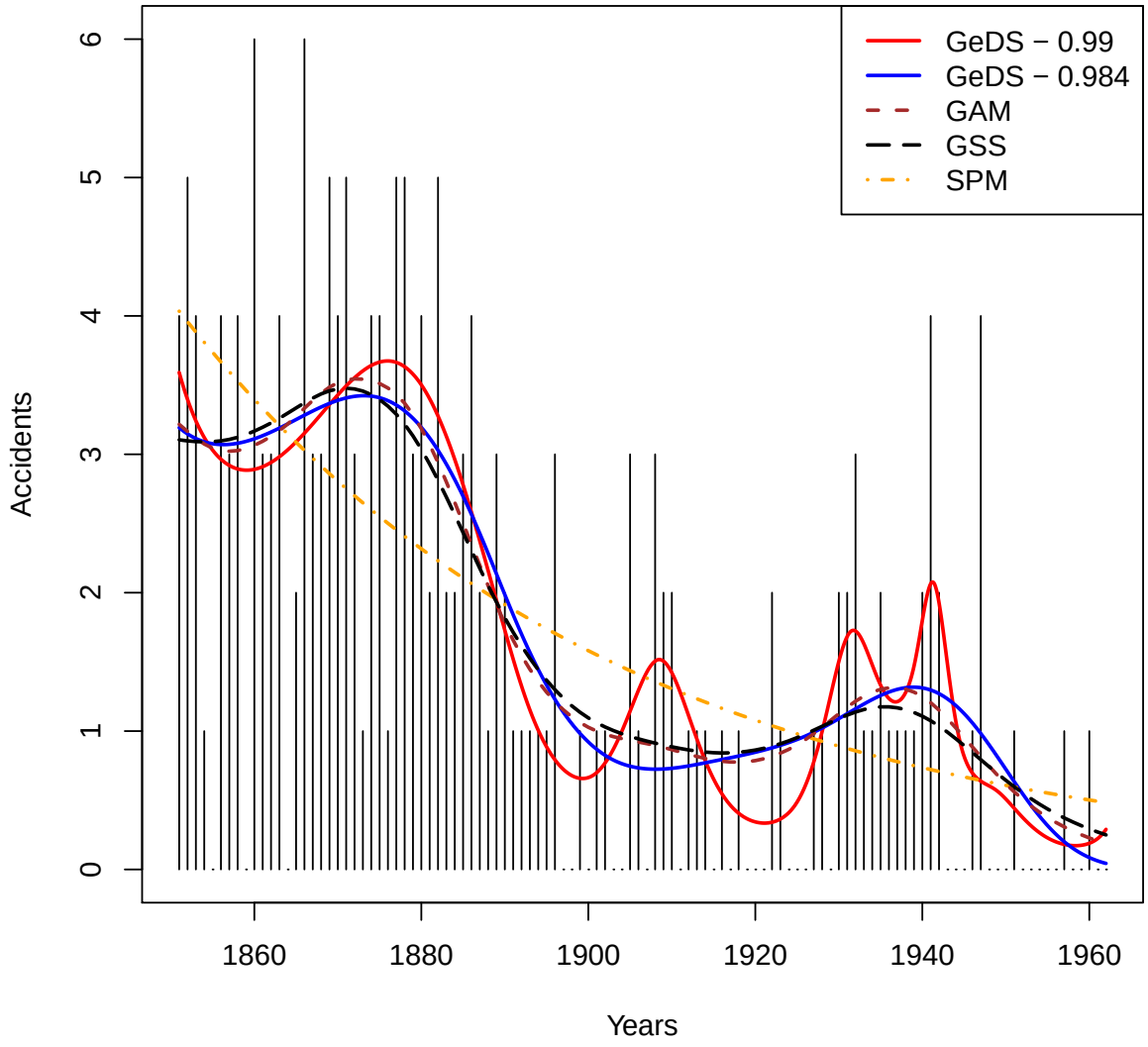


Figure 11: Comparison of the cubic GeDS fits, red ($\phi_{exit} = 0.99$) and blue ($\phi_{exit} = 0.984$) lines, with the alternative estimators GAM, (brown dashed line), GSS (black dashed line) and SPM (orange dotted and dashed line).

This dataset includes annual number of severe accidents in the United Kingdom coal mines for the period from 1851 to 1962. We apply the Poisson GeDS to estimate the expected number of accidents as a function of time under the exponentially smoothed ratio

of deviances stopping rule given in (9) and with (the default values) $\phi_{exit} = 0.99$ and $\beta = 0.2$. The resulting cubic spline fit with 12 knots is the red solid line in Fig. 11 which may be argued to overfit the data. By tuning ϕ_{exit} to 0.984, we have obtained the cubic GeDS fit with four knots illustrated by the blue line in Fig. 11.

We note that the resulting GeDS fit is similar to the fits produced by the alternative estimators GAM (representing the p-spline fit by Eilers and Marx 1996) and GSS. In addition we have also fitted an SPM model and as can be seen, it underfits the data and deviates significantly from the rest of the estimators, (c.f. Fig. 11).

5 Multivariate extension

We demonstrate here that the GeDS methodology can be successfully extended to the multivariate case of more than one explanatory variable. For the sake of simplicity and without loss of generality we consider the case of two covariates, x_1 and x_2 and generalization to higher dimensions is straightforward. We consider the two dimensional predictor $\eta(\boldsymbol{\theta}; \boldsymbol{x})$ where $\boldsymbol{x} = (x_1, x_2) \in \mathcal{D} = [a_1, b_1] \times [a_2, b_2]$. We assume that our sample of observations is now $\{x_{i1}, x_{i2}, y_i\}_{i=1}^N$ where $(x_{i1}, x_{i2}) \in \mathcal{D}$ and y_i has a distribution from the Exponential Family defined as in (1). We now assume that the predictor component of the GLM is in the form of a bivariate spline function

$$\eta(\boldsymbol{\theta}; \boldsymbol{x}) = f(\mathbf{T}_{k_1 \times k_2}; \boldsymbol{x}) = \boldsymbol{\theta}^T \left(\mathbf{N}_{n_1}(\mathbf{t}_{1;k_1,n_1}; x_1) \otimes \mathbf{N}_{n_2}(\mathbf{t}_{2;k_2,n_2}; x_2)\right) = \sum_{i=1}^{p_1} \sum_{j=1}^{p_2} \theta_{ij} N_{i,n_1}(\mathbf{t}_{1;k_1,n_1}; x_1) N_{j,n_2}(\mathbf{t}_{2;k_2,n_2}; x_2),$$

where $\mathbf{T}_{k_1 \times k_2} = \mathbf{t}_{1;k_1,n_1} \times \mathbf{t}_{2;k_2,n_2}$ and $\mathbf{t}_{1;k_1,n_1}, \mathbf{t}_{2;k_2,n_2}$ are sets of knots with respect to x_1 and x_2 with correspondingly k_1 and k_2 internal knots defined as in (2); $p_1 = n_1 + k_1$ and $p_2 = n_2 + k_2$, $\otimes$ and $\times$ are correspondingly the tensor product and the Cartesian product; $\mathbf{N}_{n_1}(\mathbf{t}_{1;k_1,n_1}; x_1) = (N_{1,n_1}(\mathbf{t}_{1;k_1,n_1}; x_1), \ldots, N_{p_1,n_1}(\mathbf{t}_{1;k_1,n_1}; x_1))^T$ and $\mathbf{N}_{n_2}(\mathbf{t}_{2;k_2,n_2}; x_2) = (N_{1,n_2}(\mathbf{t}_{2;k_2,n_2}; x_2), \ldots, N_{p_2,n_2}(\mathbf{t}_{2;k_2,n_2}; x_2))^T$ are vectors of B-spline basis functions of order correspondingly n_1 and n_2 defined on the sets of knots $\mathbf{t}_{1;k_1,n_1}$ and $\mathbf{t}_{2;k_2,n_2}$ and where $\boldsymbol{\theta} = (\theta_{11}, \ldots, \theta_{1p_2}, \theta_{21}, \ldots, \theta_{2p_2}, \ldots, \theta_{p_1p_2})^T$, is a vector of B-spline coefficients.

Similarly as in the univariate case the construction of bivariate and multivariate GeDS GNM/GLM models is based on the variation diminishing, shape preserving property, which directly carries over from the univariate to the multivariate case (c.f. Lyche and Mørken 2011). As in the univariate case the GeDS algorithm has two stages, **A** and **B**.

Stage A. In the first step of stage A, a bivariate linear spline $\hat{f}(\boldsymbol{\Delta}_{\kappa_1\times\kappa_2}, \hat{\boldsymbol{\alpha}}_p; \boldsymbol{x})$ with zero internal knots ($\kappa_1 = \kappa_2 = 0$) and $\boldsymbol{\Delta}_{\kappa_1\times\kappa_2} = \boldsymbol{\delta}_{1;k_1,2} \times \boldsymbol{\delta}_{2;k_2,2}$ is fitted to the data and the corresponding residuals are analysed as follows. The sample space is subdivided into m_1 and m_2 strips with respect to each of the coordinates x_1 and x_2 and groups of residuals of similar sign falling within each of the strips are identified, similarly as in the univariate case (see steps 2–7 of stage A of Kaishev et al. (2016)). Knots are then added, one at a time, in either x_1 or x_2 at a location defined by the group of residuals having maximum weight measure, ω_j, similarly as in steps 5-7 of stage A as described in Kaishev et al. (2016). The spline with the new knot set is then fitted and residuals are recomputed. This knot addition process is terminated following the stopping rule described in step 2 of stage A, or an alternative rule, (as described in Section 3).

Stage B. Given the final fit $\hat{f}(\boldsymbol{\Delta}_{k_1\times k_2}, \hat{\boldsymbol{\alpha}}_p; \boldsymbol{x})$ from stage A with κ_1 and κ_2 internal knots in each coordinate respectively, in stage B1 calculate the knot placements $\bar{\mathbf{T}}_{\kappa_2\times\kappa_2} = \bar{\mathbf{t}}_{1;\kappa_1-(n-2),n} \times \bar{\mathbf{t}}_{2;\kappa_2-(n-2),n}$ similarly as in (16). As in the univariate case, we note that although the spline predictor surface $f(\bar{\mathbf{T}}_{\kappa_2\times\kappa_2}, \hat{\boldsymbol{\alpha}}_p; \boldsymbol{x})$, $p = p_1p_2$, closely following the shape of the linear spline fit $\hat{f}(\boldsymbol{\Delta}_{\kappa_1\times\kappa_2}, \hat{\boldsymbol{\alpha}}_p; \boldsymbol{x})$ from stage A, is not a maximum likelihood estimate of the data. Therefore in stage **B2**, as in the univariate case, the spline regression coefficients $\hat{\boldsymbol{\alpha}}_p$ are assumed unknown parameters and the IRLS procedure is applied in order to compute the ML estimate of the vector of p_1p_2 unknown B-spline regression coefficients $\boldsymbol{\theta}$ (see stage B2 in Section 3).

5.1 Detailed description of Stage A

Following is a more detailed description of stage A.

Step 0. Similarly as in the univariate case, set $n_1 = n_2 = 2$ and $k_1 = k_2 = 0$ $p_1 = p_1 = 2$ with initial knot mesh $\boldsymbol{\Delta}_{0\times0} = \boldsymbol{\delta}_{1;0,2} \times \boldsymbol{\delta}_{2;0,2}$, where $\boldsymbol{\delta}_{1;0,2} = \{\delta_{1i}\}_{i=1}^4$ and $\boldsymbol{\delta}_{2;0,2} = \{\delta_{2i}\}_{i=1}^4$ such that $a_1 = \delta_{11} = \delta_{12} \leq \delta_{13} = \delta_{14} = b_1$ and $a_2 = \delta_{21} = \delta_{22} \leq \delta_{23} = \delta_{24} = b_2$;

Step 1. Divide the sample space $\mathcal{D}$ into M_1 rectangular strips in x_1 such that $\mathcal{D}_{1j} = [a_1 + (j-1)(b_1 - a_1)/M_1, a_1 + j(b_1 - a_1)/M_1] \times [a_2, b_2]$, $j = 1, \ldots, M_1$. Consider the sets of indexes $I_{1j} = \{i \in \{1, \ldots, N\} : (x_{i1}, x_{i2}) \in \mathcal{D}_{1j}\}$, $j = 1, \ldots, M_1$, and sort the indexes in each set I_{1j} so that $\forall i_s, i_{s+1} \in I_{1j}$, $x_{i_s 2} \leq x_{i_{s+1} 2}$. Similarly, divide the sample space $\mathcal{D}$ into M_2 rectangular strips in x_2 such that $\mathcal{D}_{2j} = [a_1, b_1] \times [a_2 + (j-1)(b_2 - a_2)/M_2, a_2 + j(b_2 - a_2)/M_2]$, $j = 1, \ldots, M_2$. Consider the sets of indexes $I_{2j} = \{i \in \{1, \ldots, N\} : (x_{i1}, x_{i2}) \in \mathcal{D}_{2j}\}$, $j = 1, \ldots, M_2$, and sort the indexes in each set I_{2j} so that $\forall i_s, i_{s+1} \in I_{2j}$, $x_{i_s 1} \leq x_{i_{s+1} 1}$.

Step 2. Similarly as in step 1 of the univariate case, apply the IRLS procedure to find a bivariate ML spline fit $\hat{f}(\boldsymbol{\Delta}_{k_1 \times k_2}, \hat{\boldsymbol{\alpha}}_p; \boldsymbol{x})$ and compute the residuals r_i and the weights w_i, $i = 1, \ldots, N$.

Step 3. Similarly as in the univariate case, a stopping rule based on the deviances at consecutive iterations is checked and, if fulfilled, move to stage B, otherwise continue with step 4.

Stage 4. Find the position of a new knot as follows.

i) For each $j = 1, \ldots, M_1$, consider the pairs $\{(r_i, x_{i1})\}_{i \in I_{2j}}$ and find a candidate knot δ^*_{1j} and its cluster weight ω^*_{1j} according to Steps 3–8 of the univariate GeDS procedure described in Section 3.1.1. Let $j^*_1 := \arg\max_j \omega^*_{1j}$ and set $\delta^*_1 := \delta^*_{1j^*_1}$, $\omega^*_1 := \omega^*_{1j^*_1}$.

ii) Similarly, for each $j = 1, \ldots, M_2$, consider the pairs $\{(r_i, x_{i2})\}_{i \in I_{1j}}$ and find a candidate knot δ^*_{2j} and its cluster weight ω^*_{2j}. Let $j^*_2 := \arg\max_j \omega^*_{2j}$ and set $\delta^*_2 := \delta^*_{2j^*_2}$, $\omega^*_2 := \omega^*_{2j^*_2}$.

iii) If $\omega^*_1 \geq \omega^*_2$, then a new knot δ^*_1 is added so that $\boldsymbol{\delta}_{1;k_1+1,2} := \boldsymbol{\delta}_{1;k_1,2} \bigcup \{\delta^*_1\}$. Otherwise a new knot δ^*_2 is added so that $\boldsymbol{\delta}_{2;k_2+1,2} := \boldsymbol{\delta}_{2;k_2,2} \bigcup \{\delta^*_2\}$.

iv) Go back to step 2.

5.2 Numerical illustration

In this section we illustrate the multivariate GeDS methodology on a bivariate example assuming Normally distributed response.

Example 5. We assume that the "true" predictor is

$$\eta = f(x_1, x_2) = \sin(2x_1)\sin(2x_2), \qquad (x_1, x_2) \in [0, 3]^2. \tag{19}$$

We obtain a simulated dataset $\{x_{i1}, x_{i2}, y_i\}_{i=1}^{N}$ for which $y_i = f(x_{i1}, x_{i2}) + \epsilon_i$, where $\epsilon_i \sim Normal(0, 0.01)$, (x_{i1}, x_{i2}) are randomly scattered within $[0, 3]^2$, following a uniform distribution and $N = 400$.

This dataset is fitted with the two dimensional version of GeDS, implemented in the **R** package **GeDS**.

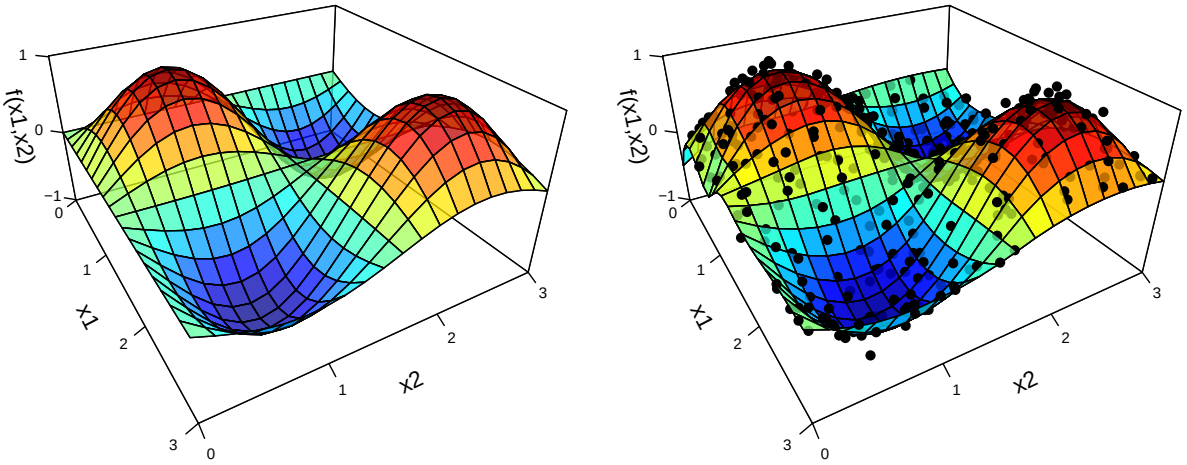


Figure 12: Panel (a) - the "true" function $f(x, y)$ from (19); panel (b) - the bi-quadratic GeDS spline fit to the 400 data points (green and red dots) obtained by adding Normal noise to $f(x, y)$.

The resulting bivariate GeDS fit is illustrated in panel (b) of Fig. 12. Comparing it with the true predictor function plotted in panel (a), one can see that the final bi-quadratic **GeDS** fit with $k_1 = 4$ and $k_2 = 2$ internal knots has reproduced it remarkably well, which is confirmed by the relatively low RSS= 3.702, by using just 400 observation.

6 Conclusions

We have demonstrated that the GeDS methodology developed for the Normal case by Kaishev et al. (2016) successfully extends to the more general Exponential Family of distributions in the context of GNM/GLM models. This has been confirmed by the thorough sensitivity tests, real data examples and comparisons with alternative spline methods such as GAM, GSS and SPM, illustrated on various underlying test functions in the paper and in the online supplement. We have also shown that the intuitive geometric interpretation, which allows one to follow the entire fitting process is preserved for GeDS generalized to the exponential family in the univariate as well as in the multivariate GNM/GLM framework with more than one covariate.

The GeDS estimation procedure can be flexibly tuned by the choice of the two parameters, ϕ_{exit} and β (see Section 3.1), using prior information (if available, otherwise estimated) about the smoothness of the underlying function.

The shape preserving variation diminishing knot positioning of GeDS avoids "knot confounding" and "lethargy" problems which have been reported for other methods (c.f. Zhou and Shen 2001 and Jupp 1978). It should also be noted that GeDS does not rely on costly non-linear optimization as is the case for other free-knot spline methods and model selection criteria, which is problematic for e.g. highly spatially inhomogeneous (multivariate) functions requiring many knots.

Another useful feature of the methodology is that it produces simultaneously linear, quadratic, cubic, and possibly higher order spline fits and offers the flexibility to choose the degree of the final fit providing best compromise between smoothness and accuracy.

GeDS has strong practical appeal especially in fast and accurate real-time smoothing when the response variable has any distribution from the exponential family and the predictor is a highly spatially inhomogeneous (possibly multivariate) function. Practitioners, academics and other researchers are encouraged to apply the GeDS methodology by making use of its **R** implementation available from the Comprehensive **R** Archive Network (CRAN) at `http://CRAN.R-project.org/package=GeDS`.

The numerical results (see Section 4 and the online supplement Dimitrova et al. 2022) have demonstrated that GeDS favourably compares with existing methods, producing fits

with a parsimonious number of knots capturing the shape of the underlying predictor functions both in the case when the latter are smooth or wiggly.

In section 5 we have illustrated the multivariate extension of GeDS. The latter generalization encompasses a broad spectrum of applications where fitting multivariate data is needed, e.g. image and signal processing, pattern recongnition, statistical data analysis, actuarial science (non-life insurance pricing), astronomy, physics and chemistry where a particular response variable is observed as a function of a set of covariates. In an ongoing research we are considering extending multivariate GeDS to cover the family of generalized additive spline models, which are useful in applications when the underlying predictor function has an additive (with respect to the explanatory variables) structure.

Application of GeDS to smoothing real mortality and claims reserving data is also part of an ongoing research.

SUPPLEMENTARY MATERIAL

As mentioned and referenced in the body of this paper, the following supplemental materials are available online.

Online supplement to: Geometrically Designed Variable Knot Splines in Generalized (Non-)Linear Models This is a supporting document containing further simulated examples and sensitivity tests (Dimitrova et al. 2022).

R-package GeDS: R-package GeDS containing code for the GNM (GLM) spline fitting method described in the article. The package also contains all datasets used as real data examples in the article. (R-package GeDS is available from CRAN at `http://CRAN.R-project.org/package=GeDS`.)

References

Arnold, T.B., Tibshirani, R.J. (2020). **genlasso**: Path algorithm for generalized lasso problems. *R package version 1.3*. URL `https://cran.r-project.org/web/packages/genlasso/index.html`.

Beliakov, G. (2004). Least squares splines with free knots: global optimization approach. *Applied Mathematics and Computation*, **149**(3), 783–798.

Biller, C. (2000). Adaptive Bayesian regression splines in semiparametric generalized linear models. *J. Comput. and Graph. Stat.*, **9**, 122–140.

Charnes, A., Frome, E.L. and Yu, P.L. (1976). The equivalence of generalized least squares and maximum likelihood estimates in the exponential family. *J. Am. Stat. Assoc.*, **71**, 169–171.

Currie, I.D. (2016). On fitting generalized linear and non-linear models of mortality. *Scandinavian Actuarial Journal*, **2016**(4), 356–83.

De Boor, C. (2001). *A practical guide to splines, revised Edition*. Springer, New York.

Dimitrova, D.S., Kaishev, V.K. and Penev, S.I. (2008). GeD spline estimation of multivariate Archimedean copulas. *Computational Statistics and Data Analysis*, **52**(7): 3570–3582.

Dimitrova, D.S., Kaishev, V.K., Lattuada, A., Verrall, R.J. (2022). Online supplement to: Geometrically Designed Variable Knot Splines in Generalized (Non-)Linear Models.

Donoho, D. and Johnstone, I. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, **81**, 425–455.

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

Fan, J. and Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *J. R. Stat. Soc. B*, **57**, 371–394.

Friedman, J.H. (1991). Multivariate adaptive regression splines (with discussion), *Annals of Statistics*, **19**(1), 1–141.

Gu, C. (2013). *Smoothing Spline ANOVA Models (2nd Ed)*. Springer-Verlag.

Gu, C. (2014). Smoothing Spline ANOVA Models: R Package **gss**. *Journal of Statistical Software*, **58**(5), 1–25.

Hansen, N. and Sokol, A. (2014). Degrees of freedom for nonlinear least squares estimation. *eprint arXiv:1402.2997*v3.

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

Jupp, D. (1978). Approximation to data by splines with free knots. *SIAM J. Num. Analysis.*, **15**, 328–343.

Kaishev, V.K., Dimitrova, D.S., Haberman, S., Verrall, R.J. (2016). Geometrically designed, variable knot regression splines. *Computational Statistics*. **31**: 1079–1105.

Kaishev, V.K., Haberman, S. and Dimitrova, D.S. (2009). Spline graduation of crude mortality rates for the English Life Table 16. Office for National Statistics (ONS) .

Kimber, S. A. J., Kreyssig, A., Zhang, Y. Z., Jeschke, H. O., Valenti, R., Yokaichiya, F., Colombier, E., Yan, J., Hansen, T. C., Chatterji, T., McQueeney, R. J., Canfield, P. C., Goldman, A. I. and Argyriou, D. N. (2009). Similarities between structural distortions under pressure and chemical doping in superconducting $BaFe_2As_2$. *Nature Materials*, **8**, 471–475.

Kovács, P. and Fekete, A.M. (2019). Nonlinear least-squares spline fitting with variable knots. *Applied Mathematics and Computation***354**, 490—501.

Lane, P.W. (1996). Generalized Nonlinear Models, In: *Proceedings in Computational Statistics, COMPSTAT* (ed. A. Prat), pp. 331–336, Wien: Physica-Verlag.

Lyche, T. and Mørken, K. (2011). *Spline Methods*. Available at `http://www.uio.no/studier/emner/matnat/ifi/INF-MAT5340/v11/undervisningsmateriale/book.pdf`.

Lindstrom, M. J. (1999). Penalized estimation of free-knot splines. *J. Comput. and Graph. Stat.*, **8**, 2, 333–352.

Luo, Z. and Wahba, G. (1997). Hybrid adaptive splines. *J. Am. Stat. Assoc.*, **92**, 107–115.

R Core Team (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL `http://www.R-project.org/`.

Rupert, D., Wand, M.P., and Carrol, R.J. (2003). Semiparametric Regression. *Cambridge University Press.*

Schwetlick, H. and Schütze, T. (1995). Least squares approximation by splines with free knots. *BIT Numer. Math.***35**, 854–866.

Shen, X. and Ye, J. (2002). Adaptive Model Selection. *Journal of the American Statistical Association***97**, 210—221.

Spiriti, S. Smith, P. and Lecuyer P. (2018). **freeknotsplines**: Free-Knot Splines. *R package version 1.0.* URL `https://cran.r-project.org/web/packages/freeknotsplines/index.html`.

Tibshirani, R.J. (2014). Adaptive piecewise polynomial estimation via trend filtering. *Annals of statistics*, **42**(1), 285–323.

Turner, H., Firth, D. (2020). **gnm**: Generalized Nonlinear Models. *R package version 1.0-8.* URL `https://cran.r-project.org/web/packages/gnm/index.html`.

Wand, M.P. (2018). **SemiPar**: Semiparametic Regression. *R package version 1.0-4.1.* URL `http://CRAN.R-project.org/package=SemiPar`.

Wahba, G., Wang, Y., Gu, C., Klein, R. and Klein, B. (1995). Smoothing spline ANOVA for exponential families, with application to the Wisconsin Epidemiological Study of Diabetic Retinopathy. *Annals of Statistics* **23**, 1865—95. Neyman Lecture.

Wood, S.N. (2000). Modelling and smoothing parameter estimation with multiple quadratic penalties. *Journal of the Royal Statistical Society (B)* **62**(2): 413–28.

Wood, S.N. (2006). *Generalized Additive Models: An Introduction with R.* Chapman and Hall.

Yang, L. and Hong, Y. (2017). Adaptive penalized splines for data smoothing. *Computational Statistics and Data Analysis* **108**: 70–83.

Ye, J. (1998). On Measuring and Correcting the Effects of Data Mining and Model Selection. *Journal of the American Statistical Association*, **93**(441): 120–131.

Yeh, R., Nashed, Y.S.G., Peterka, T., Tricoche, X. (2020). Fast Automatic Knot Placement Method for Accurate B-spline Curve Fitting. *Computer-Aided Design* **128**: 102905.

Zhou, S. and Shen, X. (2001). Spatially adaptive regression splines and accurate knot selection schemes. *J Am Stat Assoc* **96**: 247–259.