# Enhancing Geometrically Designed Spline (GeDS) Regression through Generalized Additive Models and Functional Gradient Boosting

Dimitrina S. Dimitrova[1], Vladimir K. Kaishev[1] and
Emilio Sáenz Guillén (presenter)[1]

26th International Conference on Computational Statistics (COMPSTAT 2024)

[1] Faculty of Actuarial Science and Insurance, Bayes Business School.
Email: emilio.saenz-guillen@bayes.city.ac.uk

BAYES BUSINESS SCHOOL
CITY, UNIVERSITY OF LONDON

JUSTUS-LIEBIG-UNIVERSITAT GIESSEN
iasc
COMPSTAT 2024

# 1. Motivation

- **Geometrically Designed Splines (GeDS)** (Kaishev et al., 2016, Dimitrova et al., 2023), — accurate and efficient tool for regression problems involving one or two covariates.

# 1. Motivation

- **Geometrically Designed Splines (GeDS)** (Kaishev et al., 2016, Dimitrova et al., 2023), — accurate and efficient tool for regression problems involving one or two covariates.
- GeD spline methodology is extended further by:
  1. **GAM-GeDS**: encompassing **Generalized Additive Models (GAM)**, thereby making GeDS highly multivariate.
  2. **FGB-GeDS**: incorporating **Functional Gradient Boosting (FGB)**, improving the construction of the underlying spline regression model.

# 1. Motivation

- **Geometrically Designed Splines (GeDS)** (Kaishev et al., 2016, Dimitrova et al., 2023), — accurate and efficient tool for regression problems involving one or two covariates.
- GeD spline methodology is extended further by:
  1. **GAM-GeDS**: encompassing **Generalized Additive Models (GAM)**, thereby making GeDS highly multivariate.
  2. **FGB-GeDS**: incorporating **Functional Gradient Boosting (FGB)**, improving the construction of the underlying spline regression model.

- Applications in highly multivariate contexts: AI (e.g., image recognition/processing); robotics (e.g. motion planning for humanoid robots).
- Implemented in the R package **GeDS**, available from CRAN: https://cran.r-project.org/package=GeDS

# 2. GeDS estimation method

Free-knot spline regression technique based on a ***residual-driven (locally-adaptive) knot insertion scheme*** that produces a piecewise linear spline fit, over which ***smoother higher order spline fits*** are subsequently built.

# 2. GeDS estimation method

Free-knot spline regression technique based on a ***residual-driven (locally-adaptive) knot insertion scheme*** that produces a piecewise linear spline fit, over which ***smoother higher order spline fits*** are subsequently built.

GeDS method unfolds into two phases:

- **STAGE A** constructs a least squares linear spline fit to the data.
  - Starting with a straight-line, LS fit, which is then sequentially "broken" by iteratively introducing knots at those points 'where the fit deviates most from the underlying functional shape determined by the data', based on a measure defined by residuals.

# 2. GeDS estimation method

Free-knot spline regression technique based on a ***residual-driven (locally-adaptive) knot insertion scheme*** that produces a piecewise linear spline fit, over which ***smoother higher order spline fits*** are subsequently built.

GeDS method unfolds into two phases:

- **STAGE A** constructs a least squares linear spline fit to the data.
  - Starting with a straight-line, LS fit, which is then sequentially "broken" by iteratively introducing knots at those points 'where the fit deviates most from the underlying functional shape determined by the data', based on a measure defined by residuals.
- **STAGE B**
  - Builds smoother higher order spline fits using Schoenberg's variation diminishing spline (VDS) approximation, based on the linear fit from Stage A.
  - For each higher spline order (quadratic, cubic...), compute the *averaging knot location* and re-estimate the spline coefficients by LS.

**Properties of GeDS estimated knots and regression coefficients:**

- *Schoenberg variation diminishing optimality* of the estimated knots (Kaishev et al., 2006b).
- *Asymptotic normality* of estimators in the case of normal noise, which allows for the construction of pointwise asymptotic confidence intervals that effectively converge (Kaishev et al., 2006a).
- Asymptotic conditions on the rate of growth of the knots for *negligible bias/variance ratio* (Kaishev et al., 2006a).

# 3. Generalized Additive Models with GeDS

The **Generalized Additive Model (GAM)** assumes the response variable, $Y \sim E.F.$, and relates its conditional expectation, $\mu = \mathbb{E}[Y|X]$, to the predictor variables, $X_1, ..., X_P$, via a link function $g(\cdot)$:

$$g(\mu) = \alpha + \sum_{j=1}^{P} f_j(X_j), \text{ with } \mathbb{E}[f_j(X_j)] = 0, \quad j = 1, ..., P \qquad (1)$$

Hastie and Tibshirani, 1990 — *local-scoring* and *backfitting* algorithms in conjunction with scatterplot smoothers, to fit GAMs.

**GAM with GeD Splines**: Local-scoring algorithm using GeD splines as the function smoothers, $f_j$, within the backfitting algorithm.

# 3.1. Simulated data application

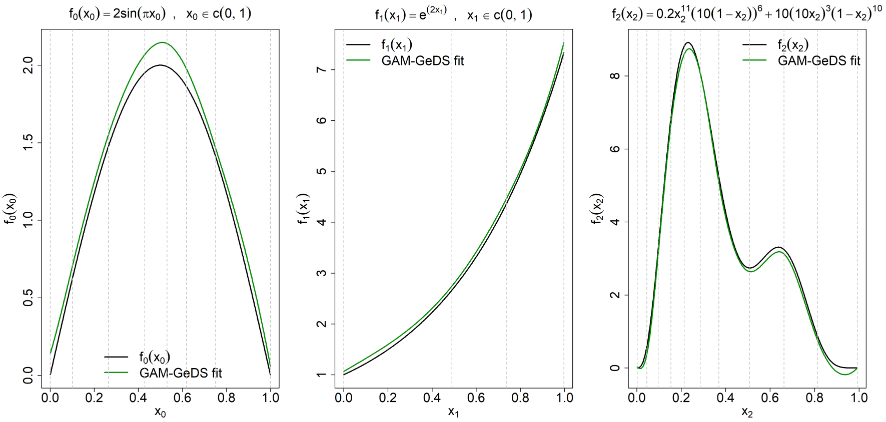
Consider the function (Gu and Wahba, 1991):

$$f(\mathbf{x}) = \underbrace{2 \times \sin(\pi \times x_0)}_{f_0(x_0)} + \underbrace{\exp(2x_1)}_{f_1(x_1)} + \underbrace{0.2x_2^{11}(10(1-x_2))^6 + 10(10x_2)^3(1-x_2)^{10}}_{f_2(x_2)}$$

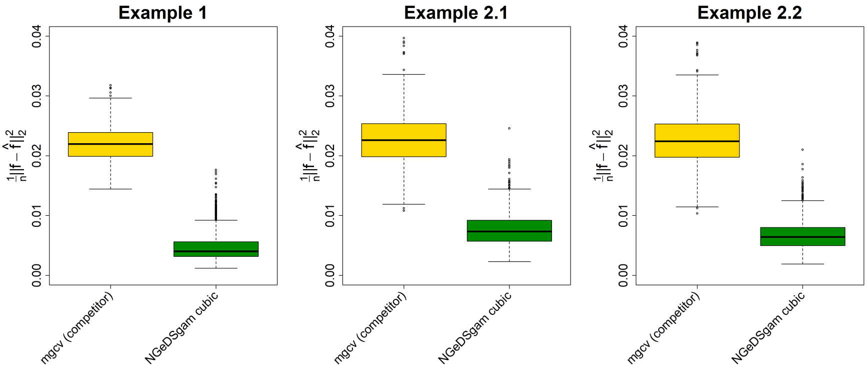
- **Example 1:** Fit $y = f(\mathbf{x}) + \epsilon$, $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$, and also include a noise predictor $x_3$.
- **Example 2:** replace $f_0(x_0)$ by a factor variable $x_0$ with 4 levels.
  - **2.1**: Include the noise predictor $x_3$.
  - **2.2**: Delete the noise predictor $x_3$.

➥ Generate $1,000$ random samples, $\{X_i, Y_i\}_{i=1}^N$, with $N = 400$ for example 1 and $N = 200$ for example 2; $x_0, x_1, x_2, x_3 \sim$ Uniform$(0, 1)$.

# Example 1: GAM-GeDS (partial) fits



# Example 1, 2.1 & 2.2: MSE boxplots



# 4. Functional Gradient Boosting with GeDS

- **Functional Gradient Boosting** (FGB; Friedman, 2001).

**❊ FGB-GeDS deals with major limitations of mainstream boosting algorithms:**

# 4. Functional Gradient Boosting with GeDS

- **Functional Gradient Boosting** (FGB; Friedman, 2001).

**❊ FGB-GeDS deals with major limitations of mainstream boosting algorithms:**

- **"Prone to overfitting"**

➠ Optimal number of boosting iterations determined by a **stopping rule** based on a ratio of consecutive deviances.

# 4. Functional Gradient Boosting with GeDS

- **Functional Gradient Boosting** (FGB; Friedman, 2001).

**❊ FGB-GeDS deals with major limitations of mainstream boosting algorithms:**

- **"Prone to overfitting"**

➠ Optimal number of boosting iterations determined by a **stopping rule** based on a ratio of consecutive deviances.

- **"Large number of parameters and unstable performance"**

➠ Strength of the base learners is **automatically regulated by the GeDS** methodology and flexibly controlled through the GeDS parameters.

# 4. Functional Gradient Boosting with GeDS

- **Functional Gradient Boosting** (FGB; Friedman, 2001).

**FGB-GeDS deals with major limitations of mainstream boosting algorithms:**

- **"Prone to overfitting"**

➠ Optimal number of boosting iterations determined by a **stopping rule** based on a ratio of consecutive deviances.

- **"Large number of parameters and unstable performance"**

➠ Strength of the base learners is **automatically regulated by the GeDS** methodology and flexibly controlled through the GeDS parameters.

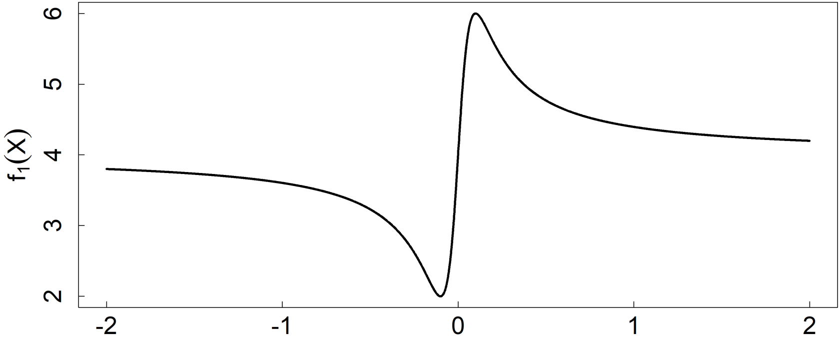
- **"Black-box models"**

➠ Final FGB-GeDS boosted model expressed as a **single spline model**, which simplifies its evaluation and enhances interpretability.

# 4.1. Simulated data application

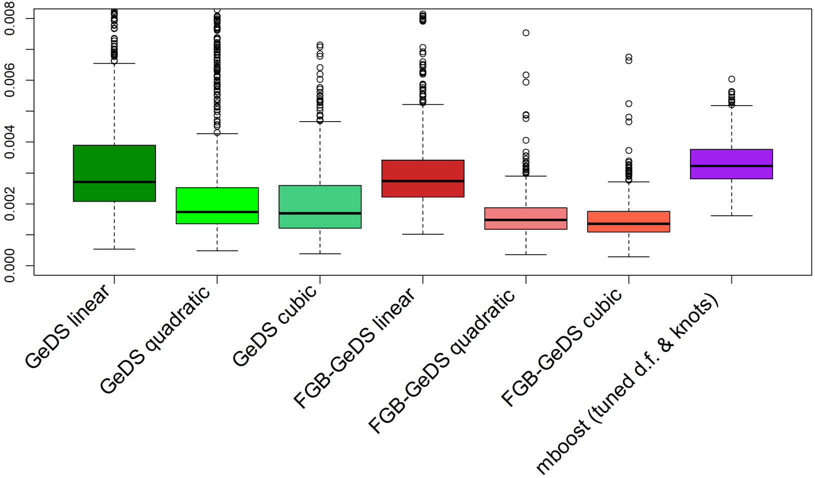
Consider the following function:

$$f_1(x) = 40\frac{x}{1+100x^2} + 4 \quad , \quad x \in c(-2, 2)$$

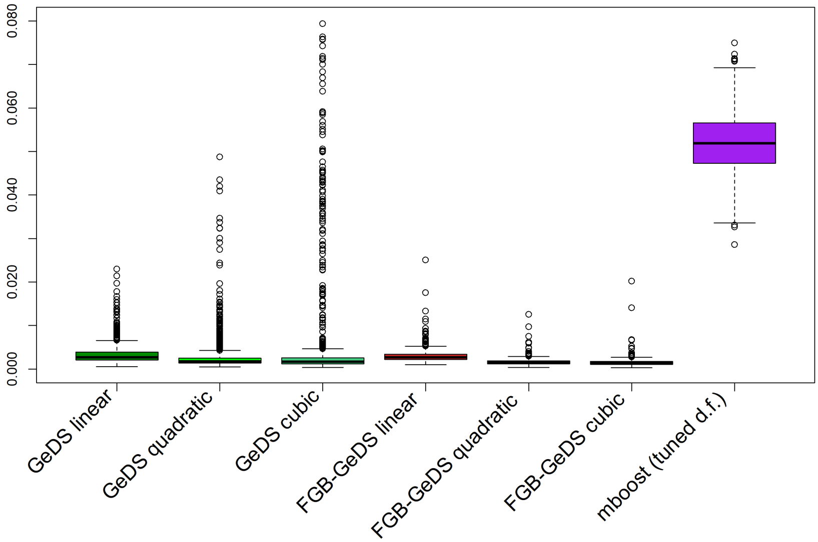



- For each, generate $1,000$ random samples, $\{X_i, Y_i\}_{i=1}^N$ with $Y_i \sim \mathcal{N}(\mu_i, \sigma)$, $\sigma = 0.2$, $\mu_i = \eta_i = f_1(X_i)$.

| | |
|---|---|
| **GeDS** | 10 int. knots |
| **FGB-GeDS** | Init. learner with 2 int. knots + 1 boosting iter. with 8 int. knots |
| **mboost** (competitor) | 10,000 boosting iter. with 36 int. knots per iter. |



➠ And setting **mboost** to have $10$ int. knots p/boosting iter. (i.e., $\simeq$ **FGB-GeDS**):



# 4.2 Task: Fourier Transform Computation of Materials Science Data

Given a sample, $\mathcal{L} = \{F(Q_i), Q_i\}_{i=1}^{N}$, $0 < Q_1 < \ldots < Q_N < \widetilde{Q}_{\text{max}}$, we are interested in estimating the **Fourier transform** (imaginary part):

$$G(r) = \frac{2}{\pi} \int_0^{Q_{\text{max}}} F(Q) \sin(Qr) dQ.$$

**Assuming $Q_{\text{max}}$ is known, this involves two steps**:

**Step 1.** Estimate $F(Q)$ through a GeDS fit $\equiv S(Q)$ to the sample $\mathcal{L}$.

**Step 2.** Compute $G(r)$ using the fitted GeDS model, $S(Q)$.

# 4.2 Task: Fourier Transform Computation of Materials Science Data

Given a sample, $\mathcal{L} = \{F(Q_i), Q_i\}_{i=1}^{N}$, $0 < Q_1 < \ldots < Q_N < \widetilde{Q}_{\text{max}}$, we are interested in estimating the **Fourier transform** (imaginary part):

$$G(r) = \frac{2}{\pi} \int_0^{Q_{\text{max}}} F(Q) \sin(Qr) dQ.$$

**Assuming $Q_{\text{max}}$ is known, this involves two steps**:

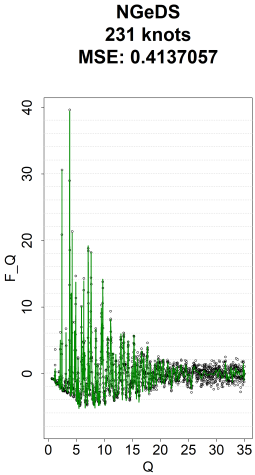
**Step 1.** Estimate $F(Q)$ through a GeDS fit $\equiv S(Q)$ to the sample $\mathcal{L}$.

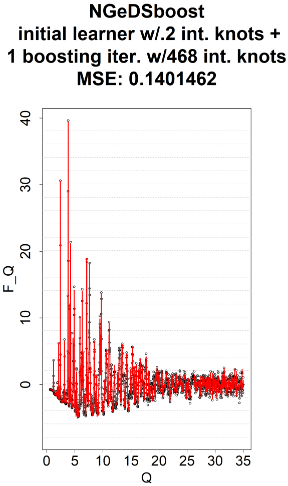
**Step 2.** Compute $G(r)$ using the fitted GeDS model, $S(Q)$.

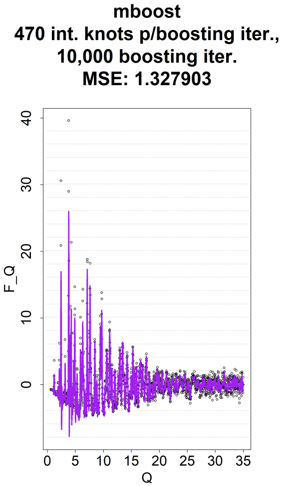
For the time being, let us assume $Q_{\text{max}} \equiv \widetilde{Q}_{\text{max}}$, though in general $Q_{\text{max}} < \widetilde{Q}_{\text{max}}$:

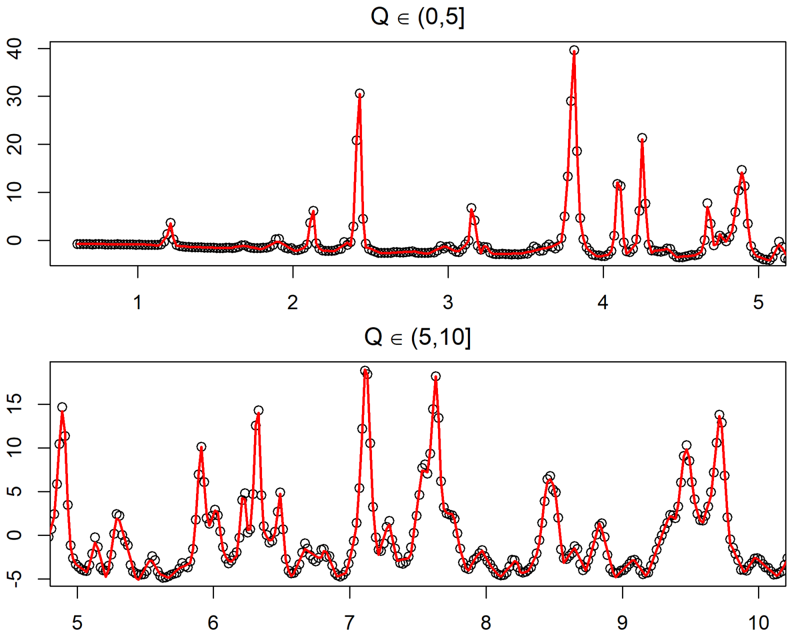
- Signal in the data prevails up to a certain point; beyond this, only noise remains.
- Sequential (and costly) data collection: cut off at the appropriate $Q_{\text{max}}$ for an optimal experimental design.

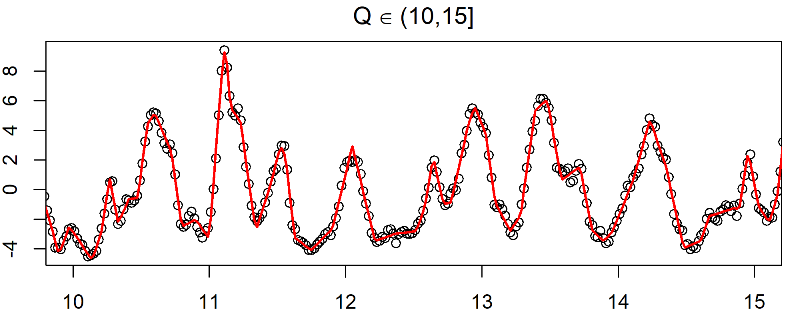
# Step 1. Fit $F(Q)$, e.g, with a GeDS model

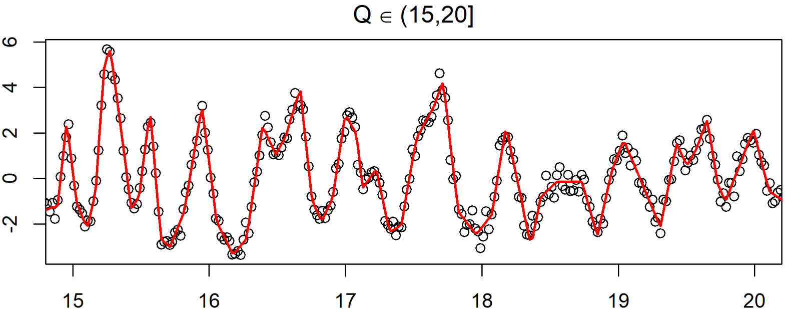


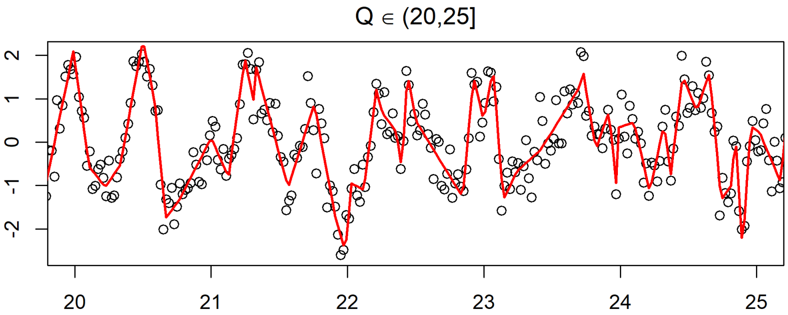



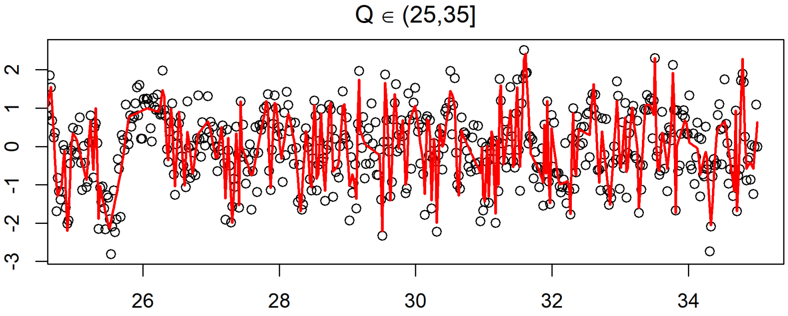




Q ∈ (0,5]
40
30
20
10
0
1
2
3
4
5
Q ∈ (5,10]
15
10
5
0
-5
5
6
7
8
9
10


Q ∈ (10,15]
-4
0
2
4
6
8
10
11
12
13
14
15


Q ∈ (15,20]
-2
0
2
4
6
15
16
17
18
19
20


Q ∈ (20,25]
2
1
0
-1
-2
20
21
22
23
24
25


Q ∈ (25,35]
2
1
0
-1
-2
-3
26
28
30
32
34

# Step 2. Compute the Fourier transform

## Proposition

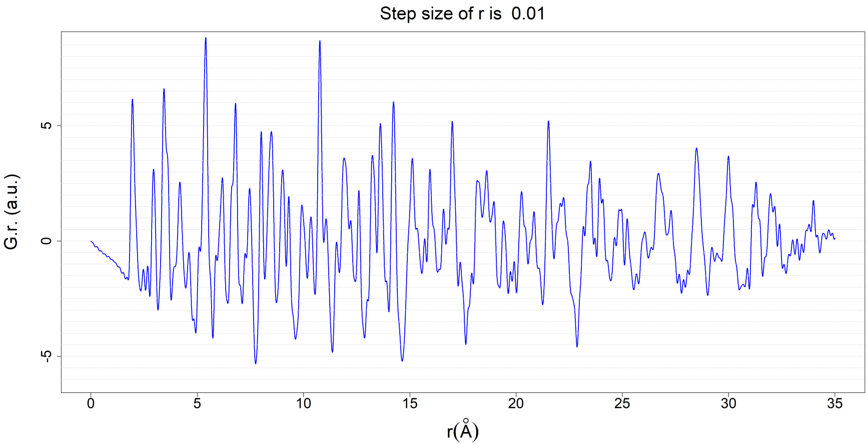
For the $\sin()$ transform,

$$G(r) = \frac{2}{\pi} \int_0^{Q_{\text{max}}} F(Q) \sin(Qr) dQ$$

of the function $F(Q)$, approximated by $S(Q)$ of order $n = 2s$, $s = 1, 2, 3, \ldots$ we have

$$G(r) \approx \frac{(-1)^s 2(n-1)!}{\pi r^n} \sum_{i=1}^{p} \hat{\theta}_i \left(t_{i+n} - t_i\right) \sum_{j=i}^{i+n} \frac{\sin(t_j r)}{\prod_{\substack{l=i \\ l \neq j}}^{i+n} (t_j - t_l)},$$

where $r \in \mathbb{R}^+$, $p = k + n$; $\hat{\theta}_i$, $i = 1, \ldots, p$ are the GeDS regression coefficients.


Step size of r is 0.01
G.r. (a.u.)
5
0
-5
0
5
10
15
20
25
30
35
r(Å)

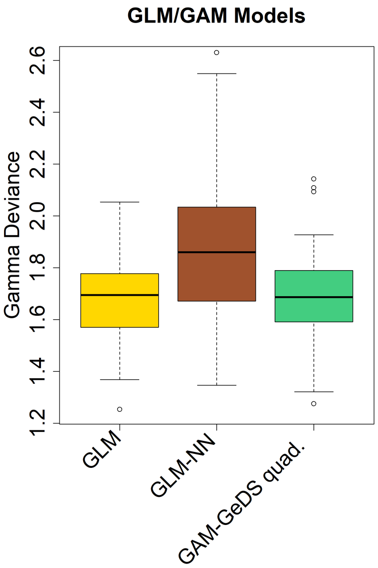
# 5. Insurance data application

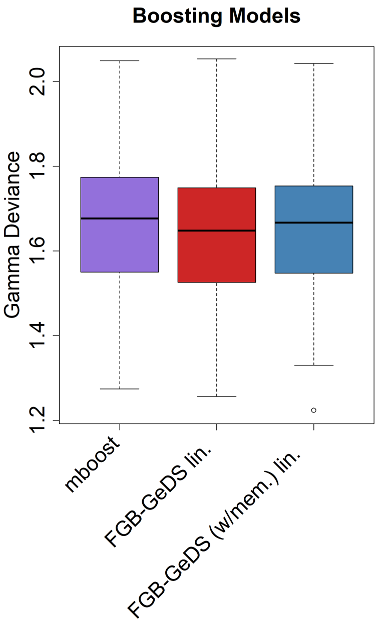
Motorcycle insurance data `swmotorcycle` available through the `R` package `CASdatasets` (Dutang and Charpentier, 2020).

$\longrightarrow$ We follow Delong et al., 2021 and model **gamma claim sizes**:

1. Gamma GLM regression + Gamma Neural Network regression.
2. `mboost`: FGB with P-splines.
3. GAM-GeDS.
4. FGB-GeDS.

- *Response*: `ClaimAmount/ClaimNb`, i.e., the average claim size.
- *Covariates*: `OwnerAge`; `Gender`; `Area`, `RiskClass`; `VehAge`.
- *Train/Test split*: **80%/20%**.
  - Simulate $\underline{100}$ different splits of data.


GLM/GAM Models
Gamma Deviance
1.2
1.4
1.6
1.8
2.0
2.2
2.4
2.6
GLM
GLM-NN
GAM-GeDS quad.


Boosting Models
Gamma Deviance
1.2
1.4
1.6
1.8
2.0
mboost
FGB-GeDS lin.
FGB-GeDS (w/mem.) lin.

## Table 1: GLM/GAM Models

| | Gamma Deviance | | | Internal knots |
|---|---|---|---|---|
| | Train Data | Test Data | Time (sec.) | (OwnerAge+VehAge) |
| GLM | 1.585727 | 1.694797 | 0.008708 | - |
| GLM NN | 1.719903 | **1.859394** | 167.224576 | - |
| GAM-GeDS quadratic | 1.557612 | **1.686492** | 0.671260 | 5 |

## Table 2: Boosting Models

| | Gamma Deviance | | | Internal knots | Boosting |
|---|---|---|---|---|---|
| | Train Data | Test Data | Time (sec.) | p/boosting iter. (OwnerAge+VehAge) | iterations |
| mboost | 1.610290 | **1.676810** | 0.156095 | 4 | 100 |
| FGB-GeDS linear (2 starting knots) | 1.575972 | **1.648345** | 0.130963 | 2 | 1 |
| FGB-GeDS w/mem. linear (1 starting knot) | 1.575536 | 1.667158 | 0.129040 | 1 | 3 |

# Concluding remarks

- GeDS is able to perform well both with more intricate, wiggly data, as well as with more disperse data.
- Broad scope of applications (insurance data, materials science data

  Further extensions:
  - Quantile regression (Hendricks and Koenker, 1992).
  - Varying coefficients regression (Hastie and Tibshirani, 1993).
  - Density estimation.

- Delong, Ł., Lindholm, M., & Wüthrich, M. V. (2021).Making tweedie's compound poisson model more accessible. *European Actuarial Journal*, *11*(1), 185–226. https://doi.org/10.1007/s13385-021-00264-3
- Dimitrova, D. S., Kaishev, V. K., Lattuada, A., & Verrall, R. J. (2023).Geometrically designed variable knot splines in generalized (non-)linear models. *Applied Mathematics and Computation*, *436*, 127493. https://doi.org/https://doi.org/10.1016/j.amc.2022.127493
- Dutang, C., & Charpentier, A. (2020). *Casdatasets: Insurance datasets* [R package version 1.0-11].
- Friedman, J. H. (2001).Greedy function approximation: A gradient boosting machine.. *The Annals of Statistics*, *29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451
- Gu, C., & Wahba, G. (1991).Minimizing gcv/gml scores with multiple smoothing parameters via the newton method. *SIAM J. Sci. Comput.*, *12*, 383–398. https://api.semanticscholar.org/CorpusID:5789455
- Hastie, T., & Tibshirani, R. (1990).Generalized additive models. *Monographs on statistics and applied probability. Chapman & Hall*, *43*, 335.
- Hastie, T., & Tibshirani, R. (1993).Varying-coefficient models. *Journal of the Royal Statistical Society. Series B (Methodological)*, *55*(4), 757–796. Retrieved November 6, 2023, from http://www.jstor.org/stable/2345993
- Hendricks, W., & Koenker, R. (1992).Hierarchical spline models for conditional quantiles and the demand for electricity. *Journal of the American Statistical Association*, *87*(417), 58–68. Retrieved November 26, 2023, from http://www.jstor.org/stable/2290452
- Kaishev, V. K., Dimitrova, D. S., Haberman, S., & Verrall, R. J. (2006a). *Geometrically designed, variable knot regression splines: Asymptotics and inference* (Statistical Research Paper No. 28). Faculty of Actuarial Science & Insurance, City University London. London, UK.
- Kaishev, V. K., Dimitrova, D. S., Haberman, S., & Verrall, R. J. (2006b). *Geometrically designed, variable knot regression splines: Variation diminish optimality of knots* (Statistical Research Paper No. 29). Faculty of Actuarial Science & Insurance, City University London. London, UK.
- Kaishev, V. K., Dimitrova, D. S., Haberman, S., & Verrall, R. J. (2016).Geometrically designed, variable knot regression splines. *Computational Statistics*, *31*(3), 1079–1105. https://doi.org/10.1007/s00180-015-0621-7