Technical Report

Beyond unimodal regression: modelling multimodality with piecewise unimodal, mixture or additive regression

Claudia Köllmann, Katja Ickstadt, Roland Fried

8/2014

technische universität dortmund

SFB 876 Verfügbarkeit von Information durch Analyse unter Ressourcenbeschränkung

Part of the work on this technical report has been supported by Deutsche Forschungsgemeinschaft (DFG) within the Collaborative Research Center SFB 876 "Providing Information by Resource-Constrained Analysis", projects C3 and C4.

Speaker: Prof. Dr. Katharina Morik
Address: TU Dortmund University
Joseph-von-Fraunhofer-Str. 23
D-44227 Dortmund
Web: http://sfb876.tu-dortmund.de

Abstract

Research in the field of nonparametric shape constrained regression has been extensive and there is need for such methods in various application areas, since shape constraints can reflect prior knowledge about the underlying relationship. It is, for example, often natural that some intensity first increases and then decreases over time, which can be described by a unimodal shape constraint. But the prior knowledge in different applications is also of increasing complexity and data shapes may vary from few to plenty of modes and from piecewise unimodal to superpositions of unimodal function courses. Thus, we go beyond unimodal regression in this report and capture multimodality by employing piecewise unimodal regression, mixture regression or additive regression models.
We give an overview of the statistical methods, namely the unimodal spline regression approach by [1] and its aforementioned extensions for use with multimodal data. The usefulness of the methods is demonstrated by applying them to data sets from three different application areas: breath gas analysis, marine biology and astroparticle physics. Though the three application areas are quite different, the proposed extensions of unimodal regression yield very helpful results in each of it. This encourages using the methodologies proposed here in many other areas of application as well.

1 Introduction

Unimodal regression as a form of nonparametric shape constrained regression is a suitable choice in regression problems where the prior information about the underlying relationship between predictor and response is vague, but where it is (almost) certain that the response first increases and then decreases with higher values of the predictor. The first approach to unimodal regression was a pointwise nonparametric estimation procedure by [2]. The idea was to estimate two monotonic regressions and find the best location for the transition via least squares criterion with the drawback of possible discontinuity at the mode. A semi-parametric spline regression approach to unimodal regression was derived in [1] and its usefulness in dose-response analysis was presented. Simulations showed that the approach is advantageous in comparison to parametric as well as non-

parametric competitors. The method is based on the fact that using the B-spline basis, a spline can be restricted to be unimodal by choosing a unimodal sequence of B-spline coefficients with a fixed mode. Picking up the original idea by [2] the mode is selected from all possibilities based on the least squares criterion. The usage of spline functions guarantees the continuity of the fit and smoothness can be achieved by using penalized spline regression as also shown in [1].

In this article we demonstrate with three real data examples from different application areas that unimodal regression is also useful in situations where the relationship between two variables is not unimodal, but multimodal. In addition, the applications have increasing complexity and vary from few to plenty of modes and from piecewise unimodal to superpositions of unimodal function courses.

The first application area is breath gas analysis where ion mobility spectrometry (IMS) coupled with multicapillary columns (MCCs) is used to measure the amount of certain molecules (also called volatile organic compounds) in the air or in exhaled breath. Knowledge about the presence of such molecules and their concentrations can be used for medical purposes, for example to diagnose lung cancer (cp. [3]). An IMS-MCC data set is a matrix of measured intensities. Looking only at one row or one column at a time, the intensities are time series where the first time axis (rows) is drift time and the second (columns) is retention time. In this article we focus only on the observed intensities along the drift time, which is called a spectrum. Typically, the intensities in a spectrum fluctuate around zero and exhibit few peaks, see also Figures 1 and 2. At least one peak is always present at about 0.5 and does not carry information about the analyte: the so-called reaction ion peak (RIP). The other peak locations and their amplitudes provide information about the presence of different molecule types. A first aim of IMS analysis is data reduction in a way that every peak is described by a number of parameters that reproduce the characteristics of a spectrum (e.g., location and amplitude of the peaks) as appropriate as possible (cp. e.g. [4]). Subsequent analyses try to identify different molecules in the analyte with the help of those parameters or to classify samples into subgroups like healthy or not (see also [5]). In this article we only focus on the first analysis step.

For a whole IMS-MCC data set (rows and columns) a 2-dimensional mixture model was

used in [4] with a background component and several peak components, where each of the latter ones is based on the product of two three-parameter inverse Gaussian distributions, one over drift time and one over retention time. This means that in drift time direction the one-dimensional model is also a mixture of three-parameter inverse Gaussian distributions. As the name already says, there are only three parameters that describe each peak. Other approaches for modelling IMS or similar data (see, e.g., [6], [7], [8]) use different distributions, but also have in common that there are very few parameters. And although one of the aims is data reduction and the distributions allow for example for skewness, this representation of the data might be too restrictive regarding the shape of a peak. Since it is known that each peak is a unimodal function of the drift time, we propose instead to analyse a spectrum with multiple unimodal regressions. The number of modes is small in this data situation and we will apply models with piecewise and superposed unimodality.

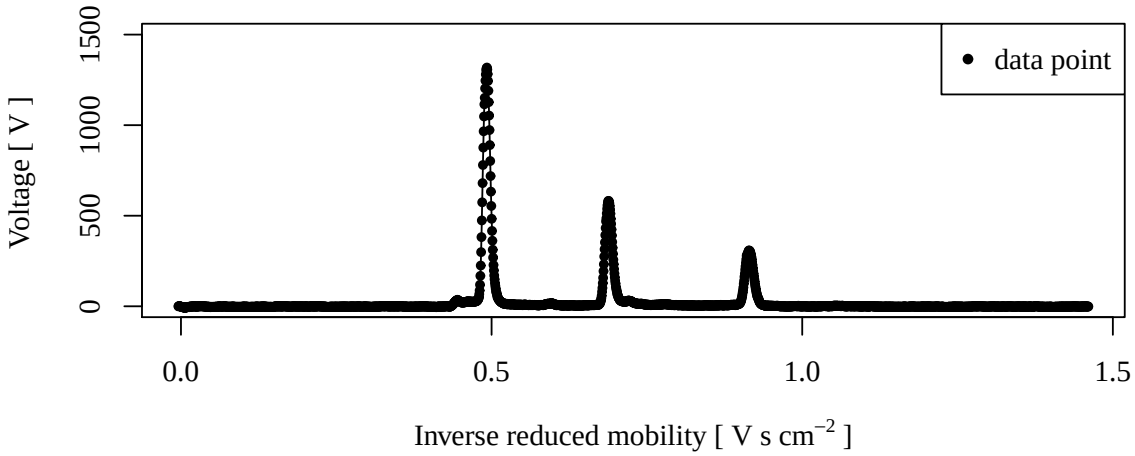

Figure 1. Spectrum A of the IMS example data set. The x-axis is inverse reduced mobility in [V s cm^{-2}], a transformation of drift time. The y-axis gives the voltage, measured in Volt [V].

The second field of application is the analysis of dive phases of marine animals. Time-depth-recorders (TDRs) measure the diving depth of marine animals such as fur seals or whales. The resulting data sets typically contain measurements of several days. Especially during night time, the animals dive repeatedly in short time intervals to forage for

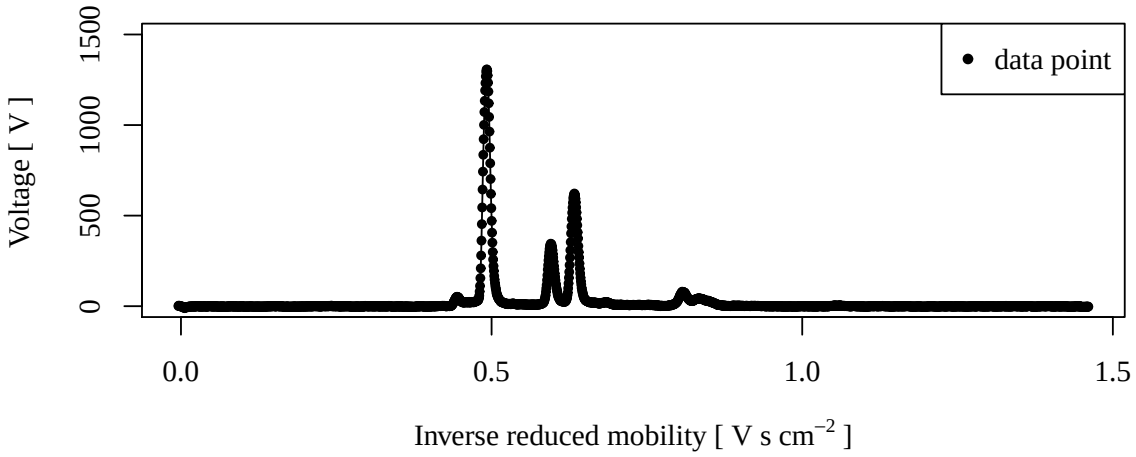

Figure 2. Spectrum B of the IMS example data set. The x-axis is inverse reduced mobility in [V s cm^{-2}], a transformation of drift time. The y-axis gives the voltage, measured in Volt [V].

food. An excerpt from such a TDR data set, taken from `R` package `diveMove` (cp. [9], [10]), is shown in Figure 3. Marine biologists are interested among other things in the detection of different dive phases like descent and ascent. For this purpose the dives have to be modelled, which is realized in version 1.3.9 of the R package by fitting multiple smoothing splines to the data. In this article we show that using piecewise unimodal regression splines is advantageous. The number of modes is much higher here than in the first application. Since the animal has to come back to the surface to draw breath, the time series can be modelled by piecewise unimodality.

The third application comes from astroparticle physics. The First G-APD Cherenkov Telescope (FACT) is used by astroparticle physicists to detect cosmic rays. These cosmic rays induce light flashes in the earth's atmosphere, which can be used to calculate the primary particle's properties. The camera of the telescope has several pixels and each pixel collects a signal, that is, a time series of measured voltages, see e.g. Figure 4. Each photon that hits a camera pixel causes a change in the signal that can be described by a unimodal loading curve with an amplitude of approximately 10 mV (see also [11]). The aim of the physicists is to detect the arrival times and numbers of photons to draw conclusions about the type of the triggering particle (gamma or hadron). A good overall fit

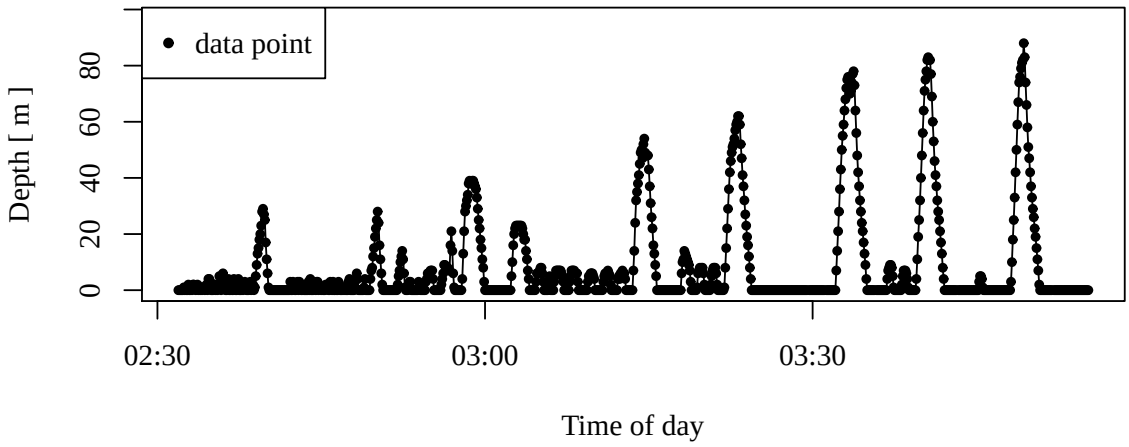

Figure 3. Excerpt from data set `divesTDR` (`R` package `diveMove`, [9], [10]). It displays the diving depth [in m] of a marine animal, which was recorded every five seconds between 02:31:55 a.m. and 03:55:15 a.m. on January 6th 2002.

is of interest, too, since the integral over the signal is used in subsequent analyses. The shape of the signal is similar to that of a loading and unloading condensator and thus physicists have derived a complicated parametric wave form for the change in the voltage when one or more photons arrive at a certain time (cp. [12]). This wave form is given by

$$U(t) = b + n_p \cdot U_0 \cdot I_{\{t \geq t0\}} \cdot \left(1 - e^{-\frac{t-t_0}{\tau_1}}\right) e^{-\frac{t-t_0}{\tau_2}} \tag{1}$$

when n_p photons arrive at time t_0. Since the telescope has been constructed quite recently, standard methods for the evaluation of measured signals are mostly heuristic and only applied on segments of a signal. [12] derived parameter estimates for waves of the form (1) from well-distinguished signals of *single* photons. As photons can arrive anytime, the measured voltage is a superposition of several loading curves (each corresponding to one or more photons). This suggests using a model with superposed parametric waves. One such parametric wave, that is the parameters b, n_p, U_0, t_0, τ_1 and τ_2, can be estimated, for example, using the least squares criterion. To ease the computational burden, we will only estimate the baseline b, the number of photons n_p and the arrival time t_0 of each wave and fix the other parameters at the following values derived in [12]: $\hat{U}_0 = 17.41$,

$\hat{\tau}_1 = 4.745$, $\hat{\tau}_2 = 31.81$. We will show that this approach is not flexible enough and we propose to use a superposition of unimodal regressions instead. The number of modes is moderate in this application.

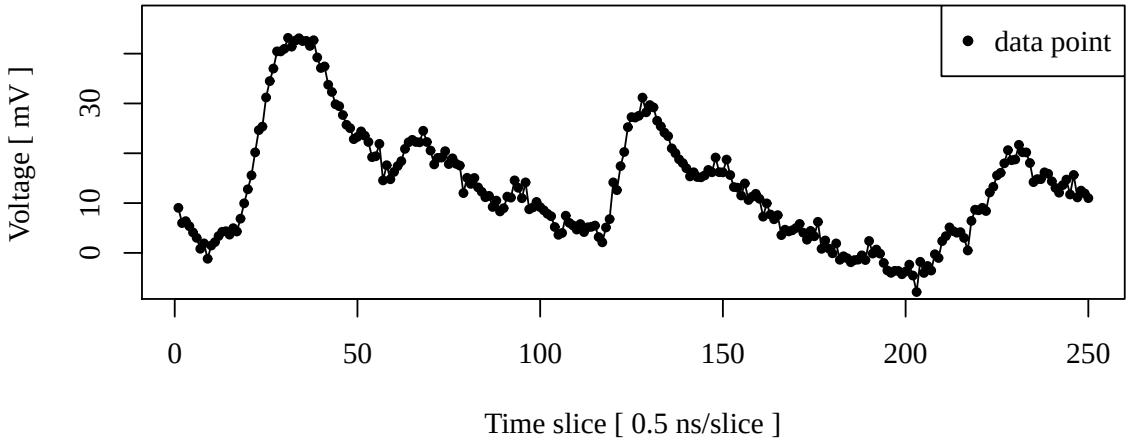

Figure 4. Example of a FACT time series. The x-axis is the number of the time slice/sample, where the slices are about 0.5 ns. The y-axis gives the measured voltages in millivolt [mV].

The remainder of the article is organized as follows: Section 2 gives an overview of the statistical models and methods. We first describe the unimodal spline regression approach by [1] before we extend it to be used in models for multimodal data: piecewise unimodal regression, mixtures of unimodal regressions and additive models. The three subsections of Section 3 then describe how the data sets from the different application areas were analysed and which results were gained. In Section 4 the results are discussed.

2 Methods

We first introduce some notational aspects. Let $\mathcal{T} = (\tau_j)_{-k}^{g+k+1}$ be the sequence of knots of a B-spline basis. That is, the basis consists of $d = g + k + 1$ (normalized) B-spline basis functions, $N_{j,k+1}(x)$, of degree $k \geq 1$ with knots $\tau_j, \ldots, \tau_{j+k+1}$. For each x the function values are given by the recursion formulae $N_{j,1}(x) = I_{[\tau_j, \tau_{j+1})}(x)$, $N_{j,k+1}(x) =$

$\frac{x-\tau_j}{\tau_{j+k}-\tau_j} N_{j,k}(x) + \frac{\tau_{j+k+1}-x}{\tau_{j+k+1}-\tau_{j+1}} N_{j+1,k}(x)$ for $j = -k, \ldots, g$. Since those functions form a basis of the space of spline functions of degree k with knots $\mathcal{T}$, every spline function s in this space can be written uniquely as $s(x) = \sum_{j=-k}^{g} \beta_j N_{j,k+1}(x)$ with so-called B-spline coefficients $\beta_{-k}, \ldots, \beta_g$ (see, e.g., [13]).

In addition, let (x_i, y_i), $i = 1, \ldots, n$, be pairs of observations of the model $Y_i = s(x_i) + \epsilon_i$ $= \sum_{j=-k}^{g} \beta_j N_{j,k+1}(x_i) + \epsilon_i$ with $\epsilon_i \underset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$, and let $\mathcal{N}_M(\mu, \mathbf{\Sigma})$ denote a multivariate normal distribution with mean μ and covariance matrix $\mathbf{\Sigma}$ truncated on the set $M \subset \mathbb{R}^d$. We suppose that an adequate estimate of $\sigma^2 \geq 0$ is available prior to model fitting.

The building block of each model used for analyzing the aforementioned data sets is the unimodal regression approach proposed in [1]. It is a univariate penalized spline regression under the shape constraint of unimodality. The shape constraint is imposed by finding the spline function $s(x) = \sum_{j=-k}^{g} \beta_j N_{j,k+1}(x)$ that minimizes the penalized residual sum of squares criterion,

$$\begin{aligned} pRSS &= \frac{1}{\sigma^2} \sum_{i=1}^{n} \left(y_i - \sum_{j=-k}^{g} \beta_j N_{j,k+1}(x_i) \right)^2 + \lambda \left\| \mathbf{V}^{-\frac{1}{2}} (\beta - \beta_0) \right\|_2^2 \\ &= \left\| \frac{1}{\sigma} (\mathbf{y} - \mathbf{B}\beta) \right\|_2^2 + \lambda \left\| \mathbf{V}^{-\frac{1}{2}} (\beta - \beta_0) \right\|_2^2, \end{aligned}$$

subject to the condition $\exists\, m :\ \beta_i \geq \beta_{i-1}\ \forall\ i \leq m$ and $\beta_i \leq \beta_{i-1}\ \forall\ i > m$ on the B-spline coefficients. Here we use cubic ($k = 3$) splines for all analyses.

Since the mode m of the coefficients is unknown, the minimization is performed for each possible choice of m and a decision is made subject to the residual sum of squares criterion. Choosing a large number of knot positions and placing a penalty on the B-spline coefficients enables an overfitting-underfitting compromise. The matrix $\mathbf{V} \in \mathbb{R}^{d \times d}$ in the *pRSS* criterion is a matrix of penalty coefficients and $\beta_0 \in \mathbb{R}^d$ is a vector of constants. The choice of $\mathbf{V}$ and β_0 determines the form of the penalty, e.g. differences of the B-spline coefficients as described in [14] or penalization against parametric functions as proposed in [1]. For the IMS data sets we use a zero order difference penalty, that is, we penalize against a constantly zero function since the intensities are very small over most of the x-axis. In the diving depth example, each unimodal piece is fitted with a second order difference penalty to introduce smoothness and the superposed regressions for the FACT

data set are penalized against a fitted parametric wave (see also equation (1)). That is, the penalty matrix $\mathbf{V}$ is the identity matrix and the vector β_0 takes the following values: $\beta_{0j} = \hat{U}(\tau_j^*) = \hat{b} + \hat{n} \cdot 17.41 \cdot I_{\{\tau_j^* \geq \hat{t_0}\}} \cdot \left(1 - e^{-\frac{\tau_j^* - \hat{t_0}}{4.745}}\right) e^{-\frac{\tau_j^* - \hat{t_0}}{31.81}}$, where τ_j^* are the knot averages, $j = -k, \ldots, g$.

The tuning parameter λ can be chosen via restricted maximum likelihood estimation (REML) or approximate REML, respectively, as described in [1]. This means that the coefficients are viewed as random with prior $\beta|\lambda \sim \mathcal{N}_{\mathcal{S}_m}(\beta_0, \lambda^{-1}\mathbf{V})$ or $\beta|\lambda \sim \mathcal{N}_{\mathbb{R}^d}(\beta_0, \lambda^{-1}\mathbf{V})$, respectively, and integrated out of the joint likelihood to get the restricted likelihood of λ. The tuning parameter that maximizes the restricted likelihood is selected.

The unimodal penalized spline regressions were performed using function `unireg` in R package `uniReg` [15]. For all three applications we use approximate REML to reduce the computational burden that arises from (repeatedly) estimating several unimodal regression functions.

For the IMS data sets we obtain an estimate of the variance from the first 800 measurements, since we know that these will definitely contain no peak, and fix σ^2 at this value prior to model fitting. For the diving depth data set the model coefficients and the variance are estimated iteratively (`unireg` with argument `abstol=0.01`), starting with an initial variance estimate of 2. For the fitting process of the FACT data we use a fixed standard deviation of 2 mV, which is the noise level's order of magnitude in the FACT measurements (see [11]).

Now one can take different approaches when modelling multimodal data using several unimodal regressions. The first and maybe simplest is a piecewise unimodal regression, that is, dividing the x-axis heuristically between each pair of modes and fitting separate unimodal splines. This implies that the underlying process that generates the observations is also divisible in some respects.

Another approach is a mixture of regressions model, where each component of the mixture is a unimodal regression and each observation stems from the different components with a certain probability. Using this model the random variable Y given x and the

parameters θ is stated to have the density

$$g(y|x,\theta) = \sum_{\ell=1}^{L} \pi_\ell \; \varphi\left(y|\mu_\ell(x), \sigma^2\right),$$

where $\varphi\left(y|\mu_\ell(x), \sigma^2\right)$ is the density of a normal distribution with variance σ^2 and the mean $\mu_\ell(x)$ is given by a unimodal spline function $\mu_\ell(x) = \sum_{j=-k}^{g} \beta_{\ell,j} N_{j,k+1}(x)$ with coefficients $\beta_{\ell,-k} \leq \ldots \leq \beta_{\ell,m_\ell-1} \leq \beta_{\ell,m_\ell} \geq \beta_{\ell,m_\ell+1} \geq \ldots \geq \beta_{\ell,g}$. That is, the parameter vector θ is given by $\theta = (\pi_1, \ldots, \pi_L, \beta_{1,-k}, \ldots, \beta_{L,g})$. This model does not only yield regression functions for all components but can also classify each data point into the subgroups specified by the components. Mixture regression models can be fitted with the EM algorithm (see [16], [17]) and are implemented in the R package `flexmix` (see [17]). The number of components L can be determined with the help of a model selection criterium, for example AIC.

The third approach for modelling multimodal data is to describe the observations as a superposition of L unimodal functions. To achieve this we employ additive models of the form

$$y_i = \alpha + \sum_{\ell=1}^{L} f_\ell(x_i) + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2),$$

where each layer f_ℓ is a unimodal spline, that is, $f_\ell(x) = \sum_{j=-k}^{g} \beta_{\ell,j} N_{j,k+1}(x)$ with coefficients $\beta_{\ell,-k} \leq \ldots \leq \beta_{\ell,m_\ell-1} \leq \beta_{\ell,m_\ell} \geq \beta_{\ell,m_\ell+1} \geq \ldots \geq \beta_{\ell,g}$. Such models can be fitted using the so-called backfitting algorithm (cp. [18]), which is given by

1. Initialize $\hat{\alpha} = \frac{1}{n} \sum_{i=1}^{n} y_i$ and $\hat{f}_\ell(x) \equiv 0 \; \forall \ell$.
2. For $\ell = 1, \ldots, L$: calculate $\hat{f}_\ell$ from data $(x_i, \tilde{y}_i)$ with $\tilde{y}_i = y_i - \hat{\alpha} - \sum_{k \neq \ell} \hat{f}_k(x_i)$.
3. Center the function estimates around zero: $\hat{f}_\ell = \hat{f}_\ell - \frac{1}{n} \sum_{i=1}^{n} \hat{f}_\ell(x_i)$.
4. Repeat steps 2 and 3 until convergence.

Initializations for the function estimates in each layer (step 1) other than zero functions might lead to faster convergence or more appropiate results. Hence, some time should be spent regarding that issue.

The additive models we are looking at are not traditional ones. In contrast to commonly

applied additive models (see e.g. [18]) we have only one regressor and this is used in all layers. The number L of layers is thus not simply the number of regressors. Sometimes the specific application might enforce a fixed number of layers or it can be determined with the help of a model selection criterium. As it is also commonly used for selecting the number of components in mixture regression models, we use the AIC for our analyses.

3 Results

In the following, the methods presented in Section 2 are applied to data sets from three application areas: breath gas analysis, marine biology and astroparticle physics. All analyses were performed using R (version 3.1.1, [19]).

3.1 Breath gas analysis with IMS

Each IMS-MCC measurement consists of only few, mostly well-distinguished peaks (see for example spectrum A in Figure 1). Thus, a first approach to model this kind of data is piecewise unimodal regression. We determined three unimodal pieces using a threshold of 50 on the measured voltages and fitted separate unimodal splines with zero order difference penalty and $g = 100$ inner knots. In Figure 5 we see that each of the three peaks is reproduced nicely using this procedure. But problems can occur when the peaks from different molecules are closer to each other as for example in spectrum B (see Figure 6). The second and third peak are so close that their tails might overlap, which means that the intensity measured in between them might result from both types of molecules, i.e., it is a superposition of both concentrations. This cannot be modelled appropriately with piecewise unimodal regression. A mixture model as used in [4] and described in Section 2 is also unable to reflect this characteristic of the data, because therein each observation stems with certain probabilities from either of the peaks and both peaks do not reach the height of the superposed concentration. Thus we propose to fit an additive model that can describe those cumulated intensities much better. Each layer is a unimodal spline regression with zero order difference penalty and $g = 250$ inner knots. Figures 7 and 8 show the estimated layers of the additive model fitted to spectrum

A and the fitted global function, respectively. The number of layers was chosen with the help of the model selection criterium AIC. That means, we fitted additive models with two to five layers and the one with four layers had the lowest AIC value. Figures 9 and 10 show the analoguos plots for spectrum B of the IMS data set. Here, additive models with two to six layers were fitted and the model with five layers was preferable with regard to AIC. It seems that the RIP is appropriately described in both data sets using two layers. The remaining peaks are modelled with one layer each and it is obvious that the two close peaks in spectrum B are now much better represented than in the piecewise model.

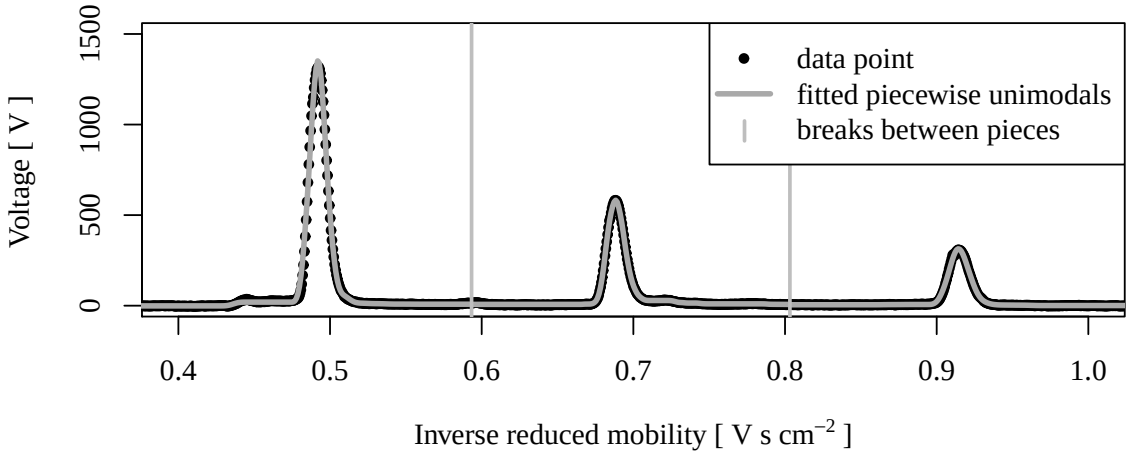

Figure 5. Close-up of IMS spectrum A with fitted piecewise unimodal regressions. The x-axis was devided into pieces according to a threshold of 50 on the intensities (y-axis) and a unimodal zero order difference penalized spline was fitted to each of the three pieces. The breaks between the pieces are indicated by vertikal lines.

3.2 Analysis of dive phases of marine animals

An approach to determine dive phases (descent and ascent) in TDR data is implemented in the `R` package `diveMove` (cp. [9], [10]). In the current version (1.3.9) the procedure starts similar to our first approach for the IMS data with heuristically dividing the diving depth time series into dives using a depth threshold of three meters and fitting a smooth-

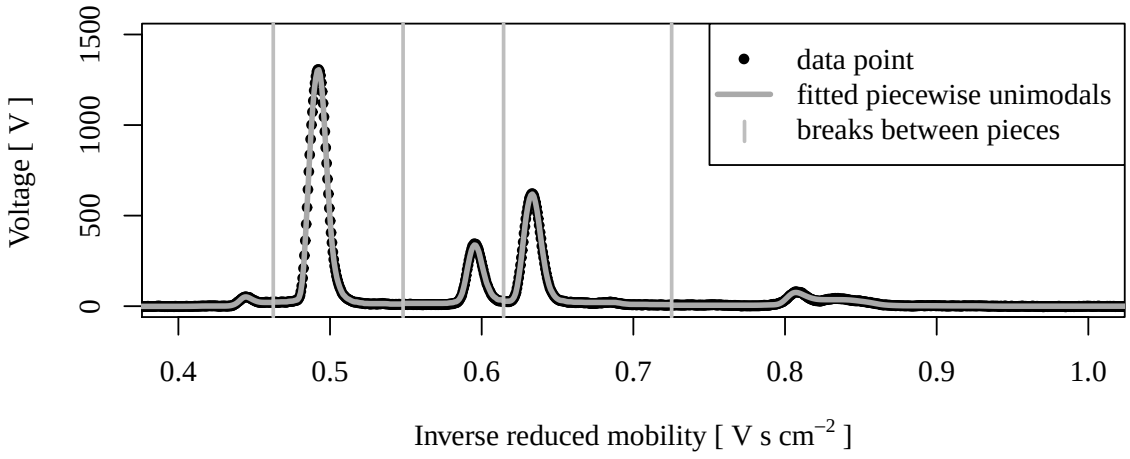

Figure 6. Close-up of IMS spectrum B with fitted piecewise unimodal regressions. The x-axis was devided into pieces according to a threshold of 50 on the intensities (y-axis) and a unimodal zero order difference penalized spline was fitted to each of the 5 pieces. The breaks between the pieces are indicated by vertikal lines.

ing spline to each dive (cp. Figure 11). Afterwards the derivative of the smoothing spline is used to identify the descent and ascent phase of each dive. This determination can be problematic since the uniqueness of the turning point depends on the choice of the smoothing parameter. This can be seen in Figure 12: for a smoothing parameter chosen via data-driven cross-validation the derivative of the smoothing spline is quite wiggly and crosses the interesting region around zero derivative several times. For a manually chosen (larger) smoothing parameter the derivative gets smoother and the zero line is only crossed once. However, such a manual choice might be a difficult task for users. If we replace the smoothing spline in the first analysis step by a unimodal second order difference penalized spline with $g = 25$ inner knots (i.e., using piecewise unimodal regression), the derivative has only one sign change and the turning point from descent to ascent is unique, irrespective of the tuning parameter value. This is also shown in Figure 12. Thus, a very welcome side effect of the unimodality constraint is that in constrast to the smoothing spline approach the choice of the tuning parameter has per construction no influence on the uniqueness of the turning point. It can be chosen via data-driven REML estimation and the user is not confronted with this task.

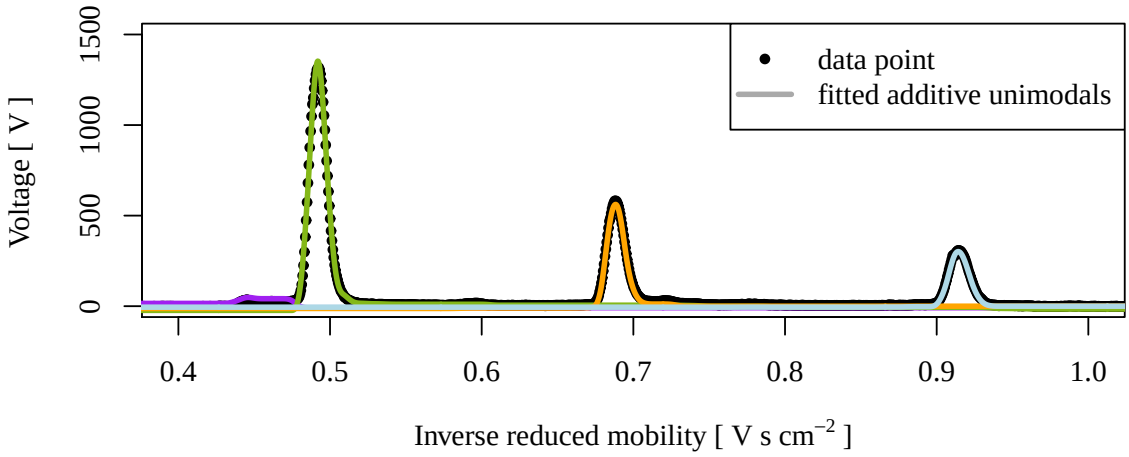

Figure 7. Close-up of IMS spectrum A with layers of the fitted additive model. Each of the four layers (marked with different colors) is a unimodal spline regression with zero order difference penalty.

3.3 Astrophysics data analysis

The loading curves caused by each photon hitting a FACT camera pixel are known to have a unimodal shape. As also already noted in the introduction, single and multiple photons can arrive at any time so that the measured voltage is a superposition of several loading curves, suggesting the use of an additive model. First, we look at an additive model where each layer f_ℓ is not a unimodal spline but a parametric wave of the form (1). The number of layers was chosen with regard to the AIC values of models with three to nine layers. Figure 13 shows the seven estimated individual waves (dashed lines) and their superposition (solid line), which was the model with lowest AIC. We can see that the superpositon of those parametric waves is a nice starting point for modelling such data, but that it is not able to describe all characteristics since it underestimates the local maxima and overestimates for example the falling edge between time points 180 and 200. Thus, the model might need some more flexibility and we also fitted an additive model where each layer is a unimodal spline ($g = 100$ inner knots) penalized against a fitted parametric wave (see also Section 2). The number of layers was chosen to be equal to that of the parametric additive model, namely 7. Figure 14 shows the fitted individual

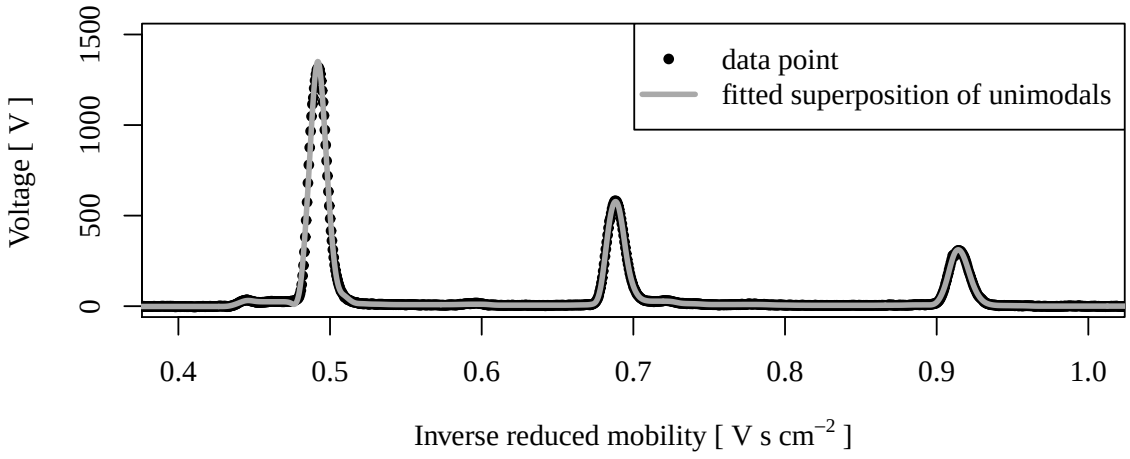

Figure 8. Close-up of IMS spectrum A together with the superpositon of the four layers displayed in Figure 7.

unimodal splines (dashed lines) and their superposition (solid line) when the tuning parameters are estimated via REML. The estimated superposed signal fits the data better than the parametric one, also in the local maxima and the falling edge. The individual unimodal splines on the other hand can be devided into two classes: most of them are nearly everywhere zero and only three of them exhibit higher amplitudes. It seems that the flexibility brought into the model by using splines is too much, since the individual splines are so wiggly that they explain the observations of more than one bunch of arriving photons. Figure 15 also shows an additive model with REML penalized unimodal splines, but here, the backfitting algorithm was initialized with the estimated waves of the additive parametric model. This results in several individual splines that differ from the zero line and that can be distinguished into different photon bunches, while the superposed signal fits the data much better than the superposed paramatric waves (cp. Figure 16) and thus the integral of the curve can be determined more precisely prior to its use in subsequent analyses.

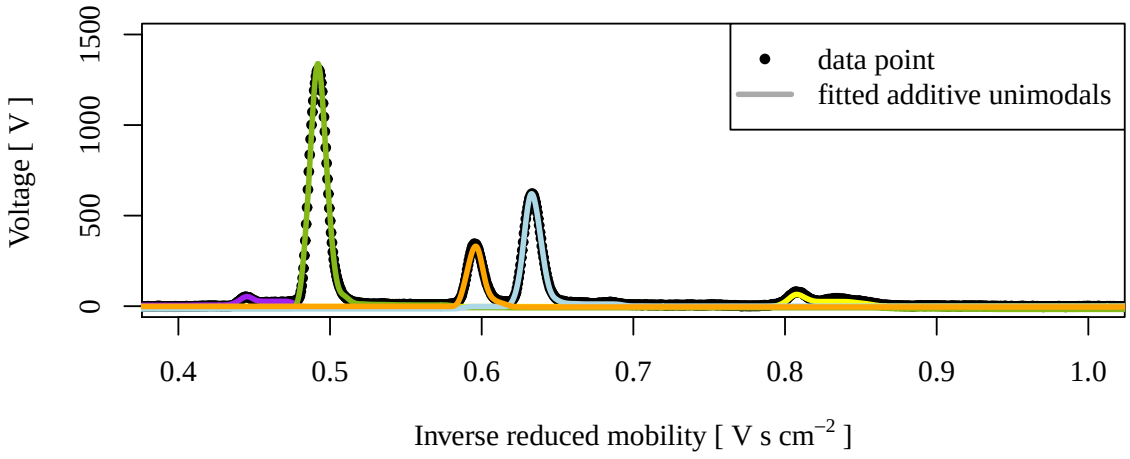

Figure 9. Close-up of IMS spectrum B with layers of the fitted additive model. Each of the five layers (marked with different colors) is a unimodal spline regression with zero order difference penalty.

4 Discussion

In this article we used the unimodal regression approach by [1] to analyse data from three different application areas: breath gas analysis, marine biology and astroparticle physics. We have seen that this approach is not only useful when a unimodal relationship between dependent and independent variable is likely (as for example in dose-response analysis as shown in [1]), but also in data situations where the relationship is multimodal and has increasing complexity: from few to plenty of modes and from piecewise unimodality to superpositions of unimodal functions. Table 1 summarizes the different data situations and gives recommendations for the model choice.

In comparison to parametric models as, e.g., in dose-response analysis or the wave form in Section 3.3, spline regression is a very flexible tool. Prior knowledge about the shape of the underlying relationship can be incorporated by using a shape constraint (here: unimodality) or by a "parametric" penalty (e.g., the wave form as a composition of two condensator loading curves).

Another nice characteristic of splines is the simplicity of calculating derivatives and that the derivatives of shape constrained splines also "inherit" shape properties. In the case of

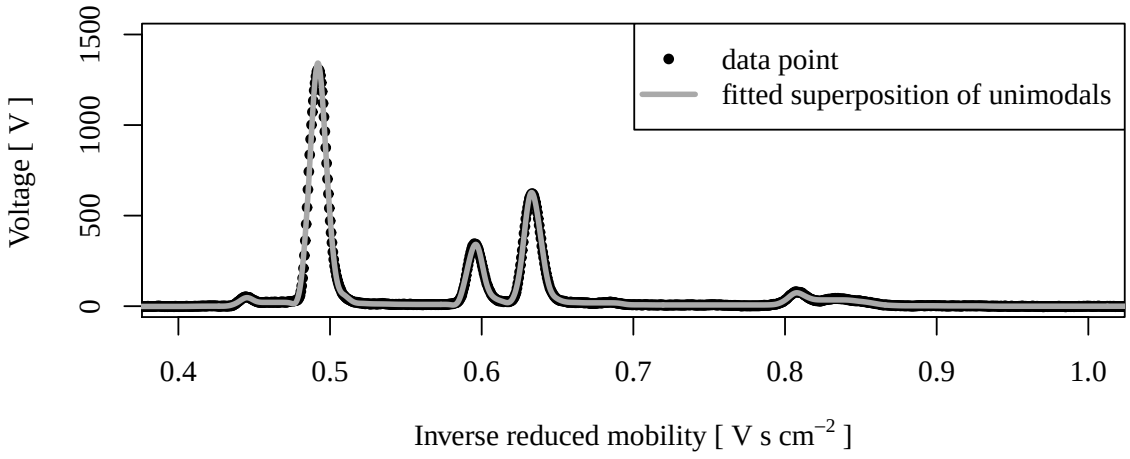

Figure 10. Close-up of IMS spectrum B together with the superpositon of the four layers displayed in Figure 9.

Table 1. Recommended approaches for different data situtations

# of modes	overlap	deriv.	recommended model	example
one	-	simple	unimodal regression as in [1]	see [1]
few to many	-	simple	piecewise unimodal regression	Sec. 3.2
few to several	weighted	hard	mixture model with unimodals	
few to several	superposed	hard	additive model with unimodals	Sec. 3.1, 3.3

This table gives an overview of the different data situations where unimodal regression or one of the proposed more complex models based on unimodal regression are applicable. Depending on the number of modes and their overlap the table states the recommended model and refers to corresponding examples. In addition, the column "deriv." informs if derivatives are easily obtained or not.

the marine biology data, the monotonicity of the first derivative simplified the subsequent analyses, namely the detection of decent and ascent phase of the dive by finding the zero of the derivative.

Of course there are situations where unimodality is not as likely as for example in the case of the peaks in IMS data or the FACT loading curves. Actually, one could argue that the dive of a marine animal is not stricly unimodal since the animal might also descend to a certain depth, make some smaller upward and downward movements and then ascend to the surface again. Those wiggles at the bottom are flattened out by the unimodal spline approach and the turning point that fits the data best divides the dive into descent and

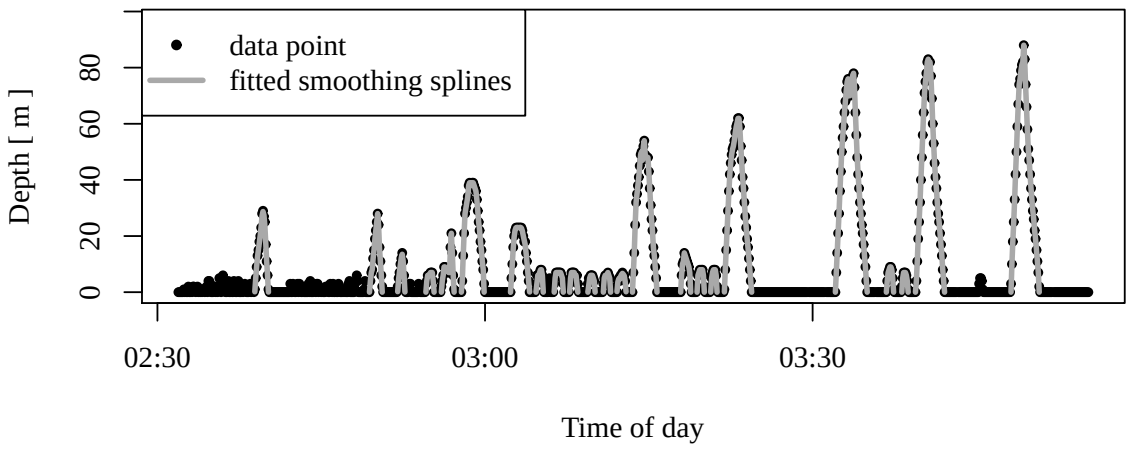

Figure 11. Excerpt from data set `divesTDR` and the smoothing splines fitted with `R` package `diveMove`.

ascent. This might be a simplification of a dive, but it makes subsequent analyses much easier and more accessibly to users.

When fitting more complex models such as the additive model with several unimodal layers in the FACT example, caution is advised when initializing the backfitting algorithm. The spline functions are so flexible that a good initialization (other than zero functions) is needed to arrive at a sensible solution.

The analyses in this article only provide an indication for the usefulness of unimodal regression in the presented applications. Especially since subsequent analysis steps like classification or integration are common in the described situations, systematic evaluations of the impact of the modelling step on the final outcome are needed. Such performance studies can be conducted, for example, along the lines of [5], where different peak detection methods for IMS data were evaluated.

Acknowledgments

We would like to thank SFB 876 projects TB1 and C3 for providing the IMS-MCC and FACT data examples and for their appreciated comments on the here presented

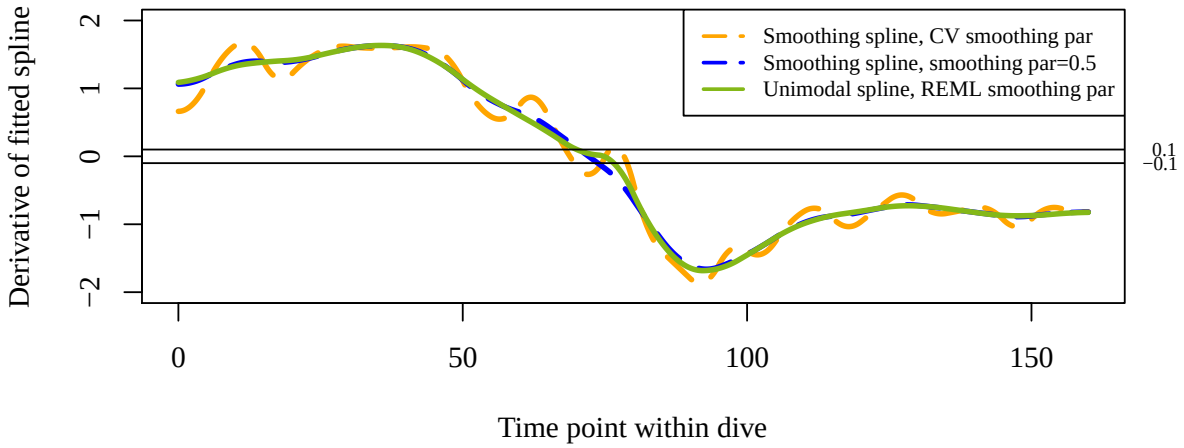

Figure 12. Derivatives of fitted smoothing splines (smoothing parameter chosen via cross-validation and manually) and unimodal spline (tuning parameter chosen via REML) for one of the dives in Figure 11. The x-axis is the number of the time point within the dive and the y-axis shows the value of the derivatives.

research. We are also grateful to Sebastián P. Luque for the valuable discussion about the application of our methods to the diving depth data.

References

[1] Köllmann C, Bornkamp B, Ickstadt K (2014). Unimodal regression using Bernstein-Schoenberg-splines and penalties. Biometrics. doi: 10.1111/biom.12193.

[2] Frisén M (1986) Unimodal regression. The Statistician 35: 479-485.

[3] Westhoff M, Litterst P, Freitag L, Urfer W, Bader S, et al. (2009) Ion mobility spectrometry for the detection of volatile organic compounds in exhaled breath of lung cancer patients. Thorax 64: 744-748.

[4] Kopczynski D, Baumbach JI, Rahmann S (2012) Peak modeling for Ion Mobility Spectrometry measurements. Proceedings of the 20th European Signal Processing Conference (EUSIPCO) : 1801-1805.

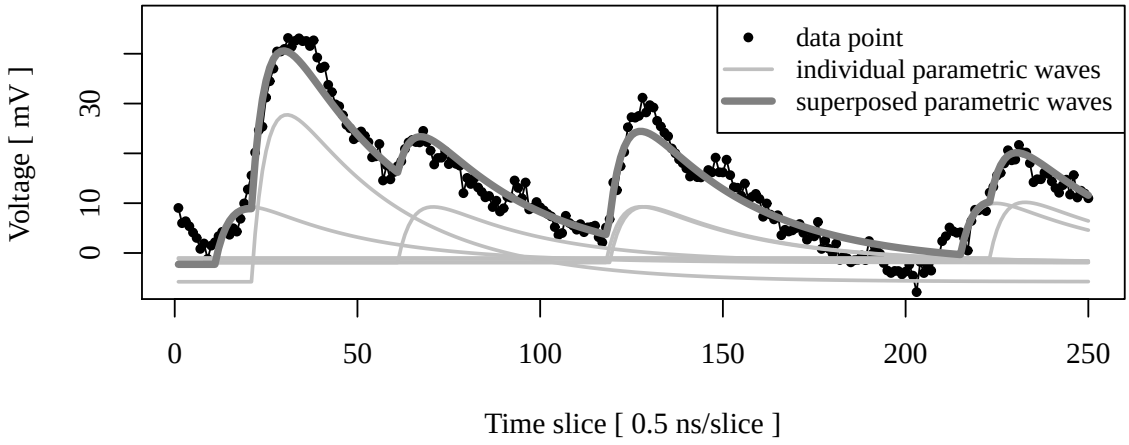

Figure 13. FACT time series with the individual layers (lightgrey) and the superposed signal (darkgrey) of the fitted additive parametric model. Each of the seven layers is a parametric wave function of form (1).

[5] Hauschild AC, Kopczynski D, D'Addario M, Baumbach J, Rahmann S, et al. (2013) Peak detection method evaluation for ion mobility spectrometry by using machine learning approaches. Metabolites 3: 277-293.

[6] Vogtland D, Baumbach J (209) Breit-wigner-function and ims-signals. International Journal for Ion Mobility Spectrometry 12: 109-114.

[7] Bödeker B, Baumbach J (2009) Analytical description of ims-signals. International Journal for Ion Mobility Spectrometry 12: 103-108.

[8] Rossoni E, Feng J (2006) A nonparametric approach to extract information from interspike interval data. Journal of Neuroscience Methods 150: 30-40.

[9] Luque SP (2007) Diving behaviour analysis in R. R News 7: 8–14.

[10] Luque SP, Fried R (2011) Recursive filtering for zero offset correction of diving depth time series with GNU R package diveMove. PLoS ONE 6: e15850.

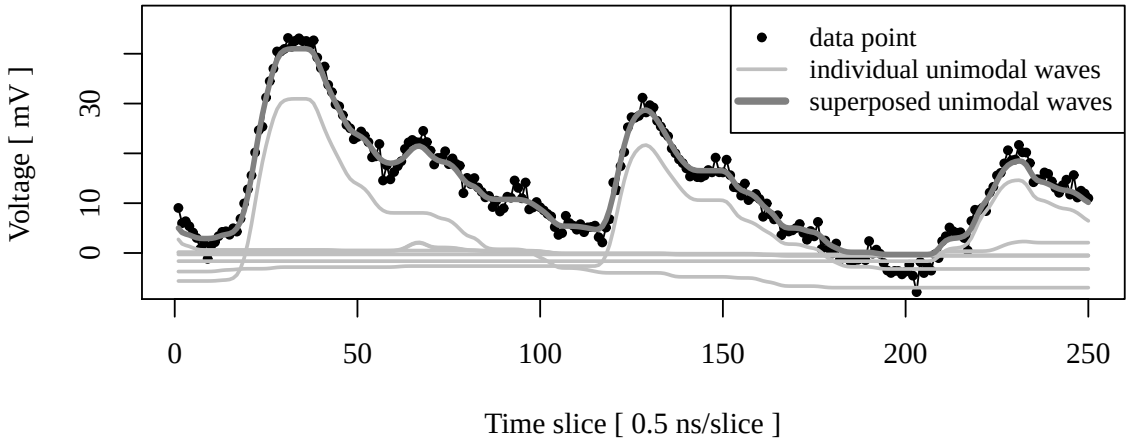

Figure 14. FACT time series with the individual layers (lightgrey) and the superposed signal (darkgrey) of the fitted additive spline model. Each of the seven layers is a unimodal spline penalized against a fitted parametric wave function of form (1).

[11] Anderhub H, Backes M, Biland A, Boccone V, Braun I, et al. (2013) Design and operation of fact - the first g-apd cherenkov telescope. Journal of Instrumentation 8.

[12] Buß JB (2013) FACT - Signal Calibration: Gain Calibration and Development of a Single Photon Pulse Template for the FACT Camera. Diploma thesis, TU Dortmund University.

[13] Dierckx P (1993) Curve and surface fitting with splines. Oxford Science Publications. Oxford: Clarendon Press.

[14] Eilers PHC, Marx BD (1996) Flexible smoothing with B-splines and penalties. Statistical Science 11: 89-121.

[15] Köllmann C (2014) uniReg: Unimodal penalized spline regression using B-splines. URL `http://cran.R-project.org/package=uniReg`. R package version 1.0.

[16] Dempster A, Laird N, Rubin D (1977) Maximum likelihood from incomplete data via the em-algorithm. Journal of the Royal Statistical Society B 39: 1-38.

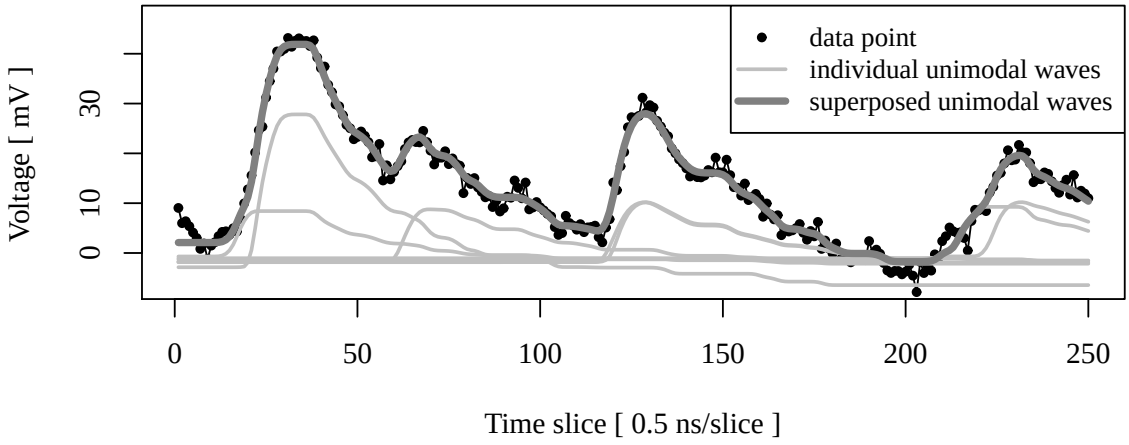

Figure 15. FACT time series with the individual layers (lightgrey) and the superposed signal (darkgrey) of the fitted additive spline model. Each of the seven layers is a unimodal spline penalized against a fitted parametric wave function of form (1). Here, the backfitting algorithm started at the parametric solution of Figure 13.

[17] Grün B, Leisch F (2008) Flexmix version 2: Finite mixtures with concomitant variables and varying and constant parameters. Journal of Statistical Software 28: 1-35.

[18] Hastie T, Tibshirani R, Friedman J (2009) The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Series in Statistics. Springer, second edition.

[19] R Core Team (2014) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL `http://www.R-project.org/`.

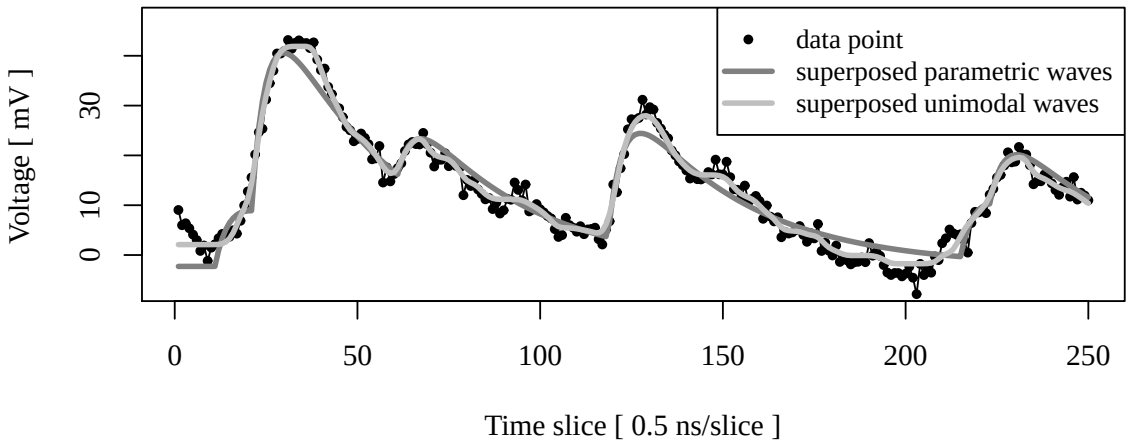

Figure 16. FACT time series with a comparison of the superposed signal estimates of Figure 13 (parametric waves, darkgrey) and Figure 15 (penalized splines, lightgrey).