Structured Additive Regression Models for Functional Data

Kumulative Habilitationsschrift
nach §12(1), Nr. 2 der Habilitationsordnung
der Ludwig-Maximilians-Universität München
für die Fakultät für Mathematik, Informatik und Statistik

vorgelegt von
Dr. Fabian Scheipl

im Juli 2017

This cumulative *Habilitationsschrift* is based on the scientific publications listed below in accordance with §12(1), Nr. 1 of the *Habilitationsordnung* as well as the additional publications submitted in lieu of a *Habilitationsschrift* in accordance with §12(1), Nr. 2 of the *Habilitationsordnung*. In an introductory report in Chapter 1, the publications submitted in lieu of a *Habilitationsschrift* are summarized and situated in a broader scientific context in accordance with §12(3) of the *Habilitationsordnung*.
In accordance with §12(2), Sentence 1 of the *Habilitationsordnung*, I describe my own contribution to the submitted works for which I'm not the first author below each reference (in parentheses).

Published works submitted for the *Habilitation*:

- F. Scheipl, T. Kneib, L. Fahrmeir (2013): Penalized Likelihood and Bayesian Function Selection in Regression Models. AStA Advances in Statistical Analysis, 97(4):349-385.
  Citation count: 8[1]
- S. N. Wood, F. Scheipl, J. J. Faraway (2013): Straightforward Intermediate Rank Tensor Product Smoothing in Mixed Models. *Statistics and Computing*, **23**(3): 341–360.
  Citation count: 37
  (Development & prototypical implementation of the basic idea described here in the R-package `amer`, later subsumed by `gamm4` and implemented as `t2` in `mgcv`; substantial contributions to the text.)
- J. Goldsmith, F. Scheipl (2014): Estimator Selection and Combination in Scalar-on-Function Regression. *Computational Statistics and Data Analysis*, **70**: 362–372.
  Citation count: 9
  (Implementation of the simulation study, of the common interface for the various methods and of the ensembling algorithm; substantial contributions to the text.)
- M. W. McLean, G. Hooker, A.-M. Staicu, F. Scheipl, D. Ruppert (2014): Functional Generalized Additive Models. *Journal of Computational and Graphical Statistics*, **23**(1): 249–269.
  Citation count: 67
  (Adviser on implementation adapted from my `pffr()`-function in the `refund`-package; contributions to the implementation, the simulation study and the application example.)
- K. Fuchs, F. Scheipl, S. Greven (2015): Penalized Scalar-on-functions Regression with Interaction Term. *Computational Statistics & Data Analysis*, **81**: 38–51.
  Citation count: 11
  (Advisor on implementation adapted from my `pffr()`-function in the `refund`-package; contributions to the implementation, the simulation study and the application example; substantial contributions to the text.)
- S. Brockhaus, F. Scheipl, T. Hothorn, S. Greven (2015): The Functional Linear Array Model. *Statistical Modelling*, **15**(3): 279–300.

[1] according to Google Scholar, accessed June 12, 2017

Citation count: 14
(Original development of basic idea & co-development of implementation in package `FDboost`; contributions to the text.)

- V. Obermeier, F. Scheipl, C. Heumann, J. Wassermann, H. Küchenhoff (2015): Flexible Distributed Lags for Modeling Earthquake Data, *Journal of the Royal Statistical Society: Series C*, **64**(2): 395–412.
  Citation count: 6
  (Basic idea and implementation of `mgcv`-interface; performed analysis and wrote the application section; substantial contribution to the text.)

- A. Gasparrini, F. Scheipl, B. Armstrong, M.G. Kenward (2017): A penalized framework for distributed lag non-linear models. *Biometrics*, *in press*, DOI: `10.1111/biom.12645`
  Citation count: 1
  (Basic idea and prototypical implementation of `mgcv`-interface; co-author of restructured `dlnm`-package; contributions to the text.)

Additional published works submitted in lieu of a *Habilitationsschrift*:

The works below, successively introducing the core elements of the research program summarized in Section 1, are submitted in lieu of a *Habilitationsschrift*:

- A.E. Ivanescu, A.-M. Staicu, F. Scheipl, S. Greven (2015): Penalized Function-on-function Regression. *Computational Statistics*, **30**(2): 539–568.
  Citation count: 22
  (Methodological development & implementation of the model class in `R`; substantial contributions to planning, execution and evaluation of the simulation study as well as the text.)

- F. Scheipl, A.-M. Staicu, S. Greven (2015): Functional Additive Mixed Models. *Journal of Computational and Graphical Statistics*, **24**(2): 477–501.
  Citation count: 53

- F. Scheipl, S. Greven (2016): Identifiability in Penalized Function-on-Function Regression Models. *Electronic Journal of Statistics*, **10**(1), 495–526.
  Citation count: 16

- F. Scheipl, J. Gertheiss, S. Greven (2016): Generalized Functional Additive Mixed Models, *Electronic Journal of Statistics*, **10**(1): 1455–1492.
  Citation count: 12

- S. Greven, F. Scheipl (2017): A General Framework for Functional Regression Modelling (with discussions & rejoinder). *Statistical Modelling*, **17**(1): 1–35, 100–115.
  Citation count: 3
  (Substantial contribution to the text; prepared, executed, evaluated and documented application examples (and simulations for the rejoinder).)

Contents

1 Résumé

1.1 Introduction

Functional data, i.e., measurements of quantitative observables over a compact domain, presents many challenges and opportunities for statistical method development. Examples for such data that have come up in our applied work include long-time average meteorological measurements over the course of a year at a set of given locations [Ivanescu et al., 2015], light reflectance intensities of samples of certain substances over a specific wavelength domain (i.e., spectrographic data) [Brockhaus et al., 2015], feeding rates of pigs over the course of a day [Scheipl et al., 2016], or measures of neuronal health (fractional anisotropy and others, derived from DTI scans) along certain white matter tracts of the human brain [Scheipl et al., 2015].
More broadly speaking, functional data analysis (FDA) [Ramsay and Silverman, 2005, Ferraty and Vieu, 2006] has been conceived as the sub-field of statistics dealing with curve-, image- or array-valued data. It is generalized further by object-oriented data analysis [Marron and Alonso, 2014, and discussion therein], which encompasses the statistical analysis of shapes [Dryden and Mardia, 2016], manifolds, and other complex data objects such as trees that are elements of non-euclidean spaces. In the following, we will use the term FDA in the more traditional and narrower sense to refer to the analysis of data containing observations of smooth one-dimensional curves.

As an area of interest within applied and theoretical statistics, FDA is concerned with mathematical, methodological and computational aspects of data analysis dealing with such curves. Theoretical advances in FDA often require a solid footing in functional analysis and the theory of stochastic processes and stochastic differential equations to handle the infinite dimensionality of random functions and the operators associated with them [Ferraty et al., 2007, Manteiga and Vieu, 2007, Cuevas, 2014, e.g.]. On the other hand, actual measurements of such functions are, firstly, available only in discretized form of measurements of the functions over some grid of evaluation points and, secondly, the vectors constituted from these measurements can frequently be projected into significantly lower-dimensional spaces without any loss of practically relevant information. These two facts make it possible to transfer large parts of multivariate statistics and adapt them to the functional data setting without the need for re-developing theory and algorithms from scratch. In our case, such a pragmatic adaptation of established approaches originally developed for scalar data has allowed fairly rapid progress in methods and implementations for regression models with functional data.

In this résumé, I provide a brief synopsis of regression methodology summarizing the basic ideas behind the approaches we adapted and refined in the papers accompanying this résumé and describe the progress towards a general framework for multivariate additive regression models with functional data our work has been able to achieve so far and compare its capabilities to those of other approaches for functional regression that have been presented in the literature.

1.2 Regression with functional data

1.2.1 From simple linear regression to structured additive models

The articles accompanying this résumé successively extend most of the flexibility and scope of current regression methodology that has been achieved for scalar data to functional data, i.e. for both functional responses and functional covariates. In order to summarize this development, this section lays the foundation by very briefly summarizing the basics of statistical regression methods.

1.2.1.1 Overview of regression models

In statistics, regression models formalize the stochastic relationship between a *response* or *dependent* variable y and *covariates* or *regressors* $x_1, \dots, x_R$. In the simplest case, the relationship is assumed to be linear and, given n realizations of tuples $\{y_i, x_{i1}, \dots, x_{iR}\}$, $i = 1, \dots, n$, the regression model can be written as $y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_{iR} x_{ir} + \epsilon_i$, where the linear predictor $\beta_0 + \sum_{r=1}^{R} \beta_r x_{ir}$ gives the average or expected value of the response variable y for a given combination of covariate values $x_{i1}, \dots, x_{iR}$, while $\epsilon_i, i = 1, \dots, n$ are the random deviations between these values and the observed data. Parameters $\beta_r, r = 1, \dots, R$ encode the strengths and the directions of the associations between the x_r and y. Another way to phrase this would be to say that the linear predictor $\beta_0 + \sum_{r=1}^{R} \beta_r x_{ir}$ is the *systematic component*, while ϵ represents *random error*, which is typically assumed to be independent and identically distributed. In other words, the relationship between the response and the covariates in a regression model is not deterministic but stochastic. Instead, the linear regression model assumes a certain linear structure of the conditional expectation of the response $E(y|x_i, \dots, x_R) = f(x_i, \dots, x_R) = \sum_{r=1}^{R} \beta_r x_{ir}$. The role of statistical inference is to estimate the shape of this conditional expectation function $f(x_i, \dots, x_R)$. In the special case of a linear regression model for scalar data, this reduces to estimating the regression coefficients $\beta_r, r = 1, \dots, R$ that yield the best fit to the observed data, according to some criterion of optimality [Fahrmeir et al., 2013, e.g.].

The simple model above, first introduced by Galton [1894], has undergone much refinement. The three subsections below describe aspects of the model equation that have been successively refined and (implicit) model assumptions that have been relaxed or done away with altogether: 1) more flexible and versatile structure of the systematic component, 2) more general distributional assumptions and loss functions, and 3) multiple inferential frameworks. The works submitted for this *Habilitationsschrift* represent a similar effort to both generalize the scope of and weaken or remove assumptions in regression models for functional data within a broad overarching framework.

1.2.1.2 Refinements of the systematic component of regression models for scalar data

First, modern regression methods for scalar data allow for a realistic and versatile structure of the systematic component:

- Whereas the simple linear model above only admits separate associations of each covariate with the response, modern regression also includes *interaction* effects (e.g., $\beta_{rr'}(x_r x_r')$ where the combination of values of multiple covariates are jointly associated with the response.
- Whereas the simple linear model above only admits covariates that are measured at least on an ordinal scale (or that are dichotomous), modern regression also includes effects of covariates measured on nominal scales. Such covariates are typically encoded via contrast vectors – in the simplest case, a binary vector encoding presence or absence is created for each level of the nominal covariate except one *reference category*, and the regression coefficient associated with such a vector then represents the expected average difference of observations in the respective category to those in the reference category.
- Whereas the simple model above only admits *linear* associations, modern regression also provides methods for estimating non-linear associations, either by kernel-based regression methods [e.g. Härdle, 1990, Hastie et al., 2011, Ch. 6] or (penalized) splines [e.g. Hastie and Tibshirani, 1990, Wood, 2006a]. This constitutes the field of study known as "(generalized) additive models" or "structured additive regression models". Note that this idea also extends to non-linear interaction effects of multiple predictors [Friedman, 1991, Gu and Wahba, 1993, Wood, 2006b, Wood et al., 2013] and non-linear varying coefficient models [Hastie and Tibshirani, 1993]. Non-linear effects of spatial or spatio-temporal locations in geo-additive regression models [Kammann and Wand, 2003, Fahrmeir et al., 2004] can be represented as such non-linear interaction effects of coordinates as well.

Making the estimation of non-linear effects feasible – at least in low dimensions – is one of the core achievements of modern statistics that has been used to achieve flexible regression models for functional data as well. More specifically, our framework utilizes low-rank penalized spline estimation [Ruppert et al., 2003, Wood, 2006a, Fahrmeir et al., 2013]. A special case of univariate functional regression models using penalized splines was first described in Marx and Eilers [1999].

1.2.1.3 Refinements of the stochastic component of regression models for scalar data

Second, modern regression methods for scalar data also allow for a more realistic and versatile structure of the model's stochastic component.

- Whereas the simple linear model is suitable only for scalar continuous responses and typically assumes Gaussianity of the errors, modern regression models have been extended to responses from any exponential family [Nelder and Wedderburn, 1972], ordinal [McCullagh, 1980] and multinomial [McCullagh and Nelder, 1984] responses as well as more general conditional response distributions for robust regression and multivariate or compositional data [Wood et al., 2016, e.g.].
- Whereas likelihood-based or Bayesian inference for the simple linear model above assumes conditional independence between observations $i = 1, \dots, n$ and a homogeneous variance of the random errors, modern regression models can also explicitly model

covariance and/or heteroskedasticity arising from a known spatio-temporal or grouping structure of the data units. *Marginal* models (also known as generalized estimating equations (GEE) models) [Liang and Zeger, 1986] focus on effects averaged over the data units' population and account for such structures by modifying the optimization criterion used for inference, whereas *conditional* models include batches of random effects in the additive predictor itself, which are used to explicitly model the stochastic effects of inter-temporal, spatial as well as inter- or intra-subject variability and dependence. This constitutes the field of study known as "mixed models" or "random effect models" [McCulloch and Neuhaus, 2001, Fahrmeir et al., 2013, e.g.].

- Whereas classical regression models estimate a linear function describing the conditional expectation of the response, regression models for location, scale and shape [Rigby and Stasinopoulos, 2005, Mayr et al., 2012] include multiple additive predictors for modeling additional conditional moments of the response distribution. Going even further, Hothorn et al. [2014] describe a framework that allows to model the entire conditional distribution of a scalar response, not just moments or parameters of pre-specified conditional response distributions. Related in spirit, quantile regression [Koenker, 2005, e.g.] defines models that explicitly target conditional quantiles of the response.

1.2.1.4 Inferential paradigms

Another aspect of the increasing flexibilization of modern regression methods is the availability of (implementations of) *multiple* inferential approaches for most of the models described in the previous paragraphs. A multitude of (penalized) maximum likelihood-based estimation methods with frequentist or empirical Bayesian approaches for testing hypotheses and providing interval estimates is now available in the form of well-documented and reliable software packages [Wood, 2016b, Bates et al., 2015, Huang et al., 2016, e.g.], but recent years have also seen increasing availability of and recognition for powerful implementations of fully Bayesian methods [Belitz et al., 2017, Umlauf et al., 2016, Stan Development Team, 2016, Wood, 2016a, Herrick, 2015, e.g.]. In addition, boosting-based approaches [Bühlmann and Hothorn, 2007, c.f.] have undergone rapid development and also provide high performing and flexible implementations [Hothorn et al., 2016, Hofner et al., 2016, Brockhaus and Rügamer, 2016].

1.2.2 Structured additive models for functional data

The simple but rather effective idea underlying most of the work presented here is that the easiest way to transfer the scope and flexibility of modern regression for scalar data to the functional setting is to attempt to re-phrase models for functional data in terms of the available models for scalar data. If such a re-phrasing can be achieved, we can then adapt and re-use the body of methodological advances achieved in the scalar setting in the functional setting. Thus, the two central steps are 1) re-writing the model for functional responses as a model for each observed function value (i.e., scalar responses) and 2) incorporating the regularity and dependence structure induced by the functional nature of the response (more precisely: the dependency between *scalar* evaluations of a random function at different argument values) into the structure of the additive predictor itself. The first step boils down to simply concatenating all observed function values into one very long vector of responses and repeating the rows with associated covariate values that are constant over the responses' domain accordingly in order to set up the design matrix and response vector for the model. The second is achieved by enforcing a certain amount of smoothness over the responses' domain by regularizing all the effects that vary over it and encoding constraints like monotonicity or periodicity of functional responses directly into the structure of the additive predictor. This is done via suitably defined regularized basis expansions (c.f. Sections 1.2.2.2, 1.2.2.3).

We adapt the notation of Brockhaus, Scheipl, Hothorn, and Greven [2015] and Scheipl et al. [2016] to formalize the model class that was developed and validated in the papers submitted for this *Habilitationsschrift*. Note that the description below refers to the most general form of this class and omits some technical details required for dealing with certain non-standard effect types or with functional responses that are not observed over a common grid of evaluation points, i.e., sparse or irregular functional data. Section 1.3 lays out the contributions of the submitted works to the general framework in broad strokes.

1.2.2.1 Model

We consider data $(Y(t), X) \subset \mathcal{Y} \times \mathcal{X}$ consisting of functional responses $Y(t)$ and a covariate vector $X = (X_1, \ldots, X_R)$. $\mathcal{X}$ is the product space of the appropriate spaces for the various scalar or functional covariates $X_r, r = 1, \ldots, R$ in X. In the following, let $\mathcal{Y}$ be the space of square integrable functions $L^2(\mathcal{T}, \mu)$. The domain $\mathcal{T}$ of the functional responses is an interval over the real numbers, $\mathcal{T} = [t_1, t_2]$, with $t_1,\ t_2 \in \mathbb{R}$, and μ is the Lebesgue measure[2]. Under the assumption that $Y(t)$ given X follows a pointwise conditional distribution $F_{Y|X}$, a structured additive regression model can be written as

$$\boldsymbol{\xi}(Y(t)|X = x) = h(x,t) = \sum_{j=1}^{J} h_j(x,t). \tag{1}$$

The model's stochastic component is here represented via the function $\boldsymbol{\xi}(\cdot)$ operating on the conditional response distribution $F_{Y|X}$, most often its expectation but also the median or some quantile. Specifically, for the class of generalized additive models that are the focus of Ivanescu, Staicu, Scheipl, and Greven [2015], Scheipl et al. [2015] and Scheipl et al. [2016], $\boldsymbol{\xi} = g \circ \mathbb{E}$ is the expectation composed with the link function g mapping the response onto the linear predictor.

Note that each functional response $y_i(t), i = 1, \ldots, n$ is only observed on a grid of evaluation points, so that the function is actually represented by a vector $\boldsymbol{y}_i = (y_i(t_{i1}), \ldots, y_i(t_{iT_i}))^T$. To rephrase the problem of estimating (1) in terms of scalar data, we concatenate $\boldsymbol{y} = (\boldsymbol{y}_1^T, \ldots, \boldsymbol{y}_n^T)^T \in \mathbb{R}^{\sum_i T_i}$ and repeat all the entries in X that do not vary over $\mathcal{T}$ so that they

[2] Note that – with some abuse of notation – most of the subsequent development is valid for scalar responses as well, if $\mathcal{T}$ is a single point and μ is the Dirac measure.

match up with the entries in $\boldsymbol{y}$. We can then rewrite (1) in terms of the scalar responses $y_i(t_l)$ as

$$\boldsymbol{\xi}(y_i(t_l)|X = x_i) = h(x_i, t_l) = \sum_{j=1}^{J} h_j(x_i, t_l).$$

1.2.2.2 Basis representation of model terms

The model's structural component, the additive predictor $h(x, t)$, is the sum of effects $h_j(x, t)$. Since effects h_j can depend on – potentially overlapping – subsets X_j of covariates and not just single elements of X this formulation includes interactions, varying coefficients or moderator effects as well. This makes the imposition of an additive structure on the predictor less restrictive and increases the capacity of the model. Each h_j is a function over the product space $\mathcal{X}_j \times \mathcal{T}$, i.e., the joint domain of covariate subset X_j used in h_j and the functional response's domain.

As in the formulations of additive regression models for scalar responses and covariates that have proven to be the most fruitful [i.e., the regression methodology exemplified by Ruppert et al., 2003, Wood, 2006a, Schmid and Hothorn, 2008, Fahrmeir et al., 2013, e.g.], each h_j is represented in terms of a penalized basis expansion in the framework presented here[3]. Under the assumption that each effect can be represented faithfully by such a basis, we let

$$\begin{aligned} h_j(x, t) &= \sum_{\ell=1}^{K_{Xj}} \sum_{v=1}^{K_{Yj}} b_{Xj\ell}(x) b_{Yjv}(t) \theta_{j\ell v}, \qquad \text{i.e.,} \\ h_j(x, t) &= \left(\boldsymbol{b}_{Xj}(x)^\top \otimes \boldsymbol{b}_{Yj}(t)^\top \right) \boldsymbol{\theta}_j, \end{aligned} \tag{2}$$

where $\boldsymbol{b}_{Xj} : \mathcal{X}_j \to \mathbb{R}^{K_{Xj}}$ is a vector of K_{Xj} basis functions $b_{Xj\ell}(x_{ji}), \ell = 1, \ldots, K_{Xj}$ in the covariate(s) associated with h_j, $\boldsymbol{b}_{Yj} : \mathcal{T} \to \mathbb{R}^{K_{Yj}}$ is a vector of K_{Yj} basis functions $b_{Yjv}(t), v = 1, \ldots, K_{Yj}$ over the domain of the response, $\boldsymbol{\theta}_j \in \mathbb{R}^{K_{Xj}K_{Yj}}$ is the vector of coefficients associated with the tensor product basis functions created by cross-multiplying the marginal bases and $\otimes$ is the Kronecker product[4]. Note that this framework allows the construction of linear or smooth effects of scalar covariates or effects of functional covariates observed on a regular grid, both constant or varying over $\mathcal{T}$, as well as scalar or functional random effects. Table 1 and Section 2.3 in Scheipl et al. [2016] provide a detailed account of the types of effects available and how the respective bases are constructed for actual data, specifically for the general case where functional responses are not observed on identical grids.

1.2.2.3 Penalization of model terms

[3]For most types of effects, this basis will be a tensor product of marginal spline bases, but in some cases functional principal component (FPC) bases that are estimated from the observed data or tensor products between splines and FPCs are used [Scheipl et al., 2016]

[4]In the case of scalar-on-function regression, $\boldsymbol{b}_{Yj}(t) \equiv 1$ with $K_{Yj} = 1, \boldsymbol{P}_{Yj} \equiv 0 \; \forall j$.

Each term is associated with an anisotropic quadratic penalty for regularization. The penalty matrix $\boldsymbol{P}_j(\lambda_{Xj}, \lambda_{Yj})$ is constructed by taking the Kronecker sum of the marginal penalty matrices $\boldsymbol{P}_{Xj} \in \mathbb{R}^{K_{Xj} \times K_{Xj}}$ and $\boldsymbol{P}_{Yj} \in \mathbb{R}^{K_{Yj} \times K_{Yj}}$ associated with $\boldsymbol{b}_{Xj}$ and $\boldsymbol{b}_{Yj}$, respectively [Wood, 2006a, Sec. 4.1.8]:

$$\boldsymbol{P}_j(\lambda_{Xj}, \lambda_{Yj}) = \lambda_{Xj}(\boldsymbol{P}_{Xj} \otimes \boldsymbol{I}_{K_{Yj}}) + \lambda_{Yj}(\boldsymbol{I}_{K_{Xj}} \otimes \boldsymbol{P}_{Yj}), \tag{3}$$

where $\lambda_{Xj}, \lambda_{Yj} \geq 0$ are the corresponding smoothing parameters that control the relative complexity or roughness of the effect over $\mathcal{X}_j$ and $\mathcal{T}$, respectively.

Note that Sections 1.2.2.2 and 1.2.2.3 together provide an extremely versatile way to pick and choose combinations of bases and penalties that can adequately represent structural assumptions or constraints for a broad variety of challenging settings. The submitted works contain successful applications to functional data with longitudinal [Scheipl et al., 2016, Greven and Scheipl [2017a]] as well as spatial [Scheipl et al., 2015, see Supplement] structure, accommodating, for example, periodicity constraints [Scheipl et al., 2015] or sparsely observed functional responses [Ivanescu et al., 2015].

1.2.2.4 Estimation & inference

If the transformation functions in (1) are restricted to $\boldsymbol{\xi} = g \circ \mathbb{E}$, i.e., the expectation operator over some probability distribution composed with a known link function g mapping the response onto the linear predictor, estimation for the proposed framework can be performed via the penalized likelihood algorithms described in Wood et al. [2016] and Wood et al. [2017] and implemented for scalar data in R-package `mgcv` [Wood, 2016b]. This is the inferential paradigm used in the submitted works Ivanescu et al. [2015], Scheipl et al. [2015], Scheipl et al. [2016], Scheipl and Greven [2016], and parts of Greven and Scheipl [2017a][5]. Since the functional data model is re-phrased as a model for scalar data that are independent conditional on the additive predictor $h(x, t)$, confidence intervals and tests available for conventional generalized additive mixed models for scalar responses as well as related model selection procedures can be used directly. Empirical evaluation corroborates the satisfactory performance of these methods in the functional setting for both Gaussian [Scheipl et al., 2015] and non-Gaussian [Scheipl et al., 2016] functional responses, while the discussions by Kokoszka and Reimherr [2017] and Morris [2017] and our rejoinder [Greven and Scheipl, 2017b] provide some additional perspective on this issue.

More recent developments of our framework like Brockhaus et al. [2015] and Brockhaus et al. [2016] implement the full generality of transformation functions specified in (1) via component-wise gradient boosting and even extend it to models with multiple additive predictors for distributional regression (i.e., generalized additive models for location, scale and shape (GAMLSS), c.f. Mayr et al. [2012]). While this approach currently only provides

[5] It is also the inference framework used for scalar responses in the submitted works Wood et al. [2013], McLean et al. [2014], Fuchs et al. [2015], Obermeier et al. [2015] and Gasparrini et al. [2017], as well as some of the methods compared in Goldsmith and Scheipl [2014].

resampling-based methods for uncertainty quantification, it is tailored to perform variable and model selection even in the *small n, large p*-setting, with strong asymptotic guarantees based on the notion of *stability selection* [Meinshausen and Bühlmann, 2010; Shah and Samworth, 2013; Hofner et al., 2015; see Scheipl et al., 2013, for a comparison with other function selection methods].

1.2.2.5 Implementation

The `pffr()` (functional responses) and `pfr()` (scalar responses) functions in R-package `refund` (Huang, Scheipl, Goldsmith, Gellar, Harezlak, McLean, Swihart, Xiao, Crainiceanu, and Reiss [2016]) implement convenient wrappers for the entirety of `mgcv`'s functionality that provide data transformations and customized identifiability constraints required for the functional data setting as well as functions for pre-processing functional data and constructing basis representations for different kinds of effects of functional covariates.

Component-wise gradient boosting for functional data is implemented in R-package `FDboost` [Brockhaus and Rügamer, 2016], which uses R-package `mboost` [Hothorn et al., 2016] as its computational back-end. For a subset of effects with a certain structure and functional responses on a regular grid, the linear array model trick [Currie et al., 2006] implemented in `FDboost` allows for large reductions in both memory consumption and computational effort.

Nonlinear effects of functional covariates for scalar responses [McLean et al., 2014] are implemented in the `af()` function in the `refund` package.

1.2.2.6 Comparison to other frameworks

Functional regression methods have been developed in a number of paradigms beside the one we have pursued in the publications submitted here. Comprehensive recent reviews that present a fairly broad cross-section of the methodological diversity in this field are Morris [2015], Wang et al. [2016], and Reiss et al. [2016]. The following is a short synopsis of Greven and Scheipl [2017a, Section 1.4] and Greven and Scheipl [2017b, Table 3] that distinguishes four complementary frameworks for representing and modelling functional data.

- The approach laid out in the seminal FDA books of Ramsay and Silverman [2005] and Ramsay et al. [2009] and implemented in the `fda` [Ramsay et al., 2014] packages for MATLAB and R treats smoothed basis representations of the observed vectors of function evaluations as the primary unit of analysis, as opposed to the vectors of evaluations themselves. These objects can then be treated as true functions, which makes some mathematical considerations easier and sidesteps the computational difficulties of our approach when dealing with large numbers of densely observed functions. However, any measurement error associated with functions treated in this way is not accounted for after the pre-smoothing step, densely observed functions are required and available implementations are mostly limited to simple special cases of the model class we have described.

- Ferraty and Vieu [2006] describe fully non-parametric approaches for functional data mostly based on kernel methods. They are able to describe highly non-linear and non-additive association structures, but available implementations like the `fda.usc` [Febrero-Bande and Oviedo de la Fuente, 2012] package for R are fairly rudimentary and to the best of my knowledge, no extensions to multiple regression models or to generalized functional data have been published.
- A class of flexible, computationally efficient and powerful methods projects functional data into low-dimensional coefficient spaces for given basis representations and performs subsequent analysis on the projected coefficient vectors, often reducing the problem of functional data regression into multivariate regression problems in similar fashion to our proposed reduction of functional regression to regularized pointwise models of scalar functional evaluations. This approach was introduced in Morris and Carroll [2006] and subsequent papers from this group and implemented for fully Bayesian MCMC based inference with Wavelet bases in the `WFMM` software [Herrick, 2015]. This approach provides great flexibility for the specification of the additive predictor and excellent computational scalability for large data sets, but extensions to e.g. sparse or irregular functional data or generalized functional data have not been provided and the publicly available implementation does not provide the full flexibility of the proposed model class.
- Finally, a computationally challenging approach based on Gaussian Process models for functional data with linear effects of scalar covariates has been described in Shi and Choi [2011] and related papers from this group and implemented partly in the R package `GPFDA` [Shi and Cheng, 2014].

1.3 Research program

In more detail, the contributions of the submitted publications to the research program outlined above are as follows:

Wood, Scheipl, and Faraway [2013]

This paper describes a convenient computational simplification for penalized tensor product spline bases that are central to our implementation of additive models for functional data described in later publications.

Ivanescu, Staicu, Scheipl, and Greven [2015]

Estimation and inference for penalized spline-based linear effects of functional covariates $x(s)$ (i.e., $h_j(x,t) = \int x(s)\beta(s,t)ds$) on independent functional responses with Gaussian errors are described and empirically validated in Ivanescu et al. [2015].

Scheipl and Greven [2016]

Scheipl and Greven [2016] discusses the problem of finite sample identifiability of such effects under both spline-based and functional principal component based modeling approaches from an applied perspective and provides empirically validated criteria for identifying non-identifiable data settings and model specifications.

Scheipl, Staicu, and Greven [2015]

This central paper introduces the general term representation of Section 1.2.2.2 for models with conditionally independent functional responses with Gaussian errors, thereby extending the full flexibility of the structural model component of the scalar data setting to the functional data setting. FPC-based effects of (potentially sparsely or irregularly) observed functional covariates are introduced and empirically validated. A very general class of both spline-based or functional principal component-based functional random effects for this framework is defined and empirically validated on both simulated and real data sets, extending the applicability of our approach to correlated functional data. The analogy between our model formulation and well-known mixed model representations of scalar GAMMs is made explicit and the inferential properties of the resulting estimates are empirically validated on challenging synthetic data sets of different sizes and noise levels for a variety of different structures of the "true" additive predictor. We also provide an extensive case study of a geo-additive regression model for spatially correlated functional responses.

Scheipl, Gertheiss, and Greven [2016]

While the models described in Scheipl et al. [2015] remain restricted to conditionally Gaussian functional responses, this paper generalizes the *stochastic* component of the proposed model to the much broader model class with $\boldsymbol{\xi} = g \circ \mathbb{E}$, i.e., *generalized* functional responses for which only the latent expectation structure is assumed to be smooth and amenable to a suitable basis expansion over $\mathcal{T}$. Linear effects of functional covariates are generalized to so-called historical effects $h(x,t) = \int_{l(t)}^{u(t)} x(s)\beta(s,t)ds$, in which only a subset of the domain of $x(s)$ is assumed to be associated with the response at any given argument value t. Inferential properties and computational feasibility of the proposed approach are then validated on a variety of synthetic data sets from different response distributions as well as a case study on a data set of binomial longitudinal functional data.

Brockhaus, Scheipl, Hothorn, and Greven [2015]

In Brockhaus et al. [2015], a componentwise gradient boosting approach for inference for the model class described in Scheipl et al. [2015] and Scheipl et al. [2016] is developed. This very flexible and powerful framework
(1) extends the scope of the model class to scalar responses, (2) also allows for functional regression with robust loss functions such as the Huber loss and pointwise quantile regression for functional responses since boosting-based inference does not require proper stochastic models and (3) allows for automatic data-driven variable and model selection via early

stopping [Bühlmann and Hothorn, 2007] or stability selection [Meinshausen and Bühlmann, 2010], even in the $p > n$-setting

Greven and Scheipl [2017a]

This is an invited discussion paper and rejoinder showcasing the penalized likelihood and component-wise gradient boosting implementations of our approach. A running application example showcases the flexibility and performance of the approach presented in Scheipl et al. [2015] and Scheipl et al. [2016] as well as its boosting-based analogues [Brockhaus, Scheipl, Hothorn, and Greven, 2015] and its extension to GAMLSS-type models [Brockhaus et al., 2016] and is used to compare the respective strengths and weaknesses of the two approaches.

A series of papers describes and compares estimators of effects of functional covariates on scalar responses:

McLean, Hooker, Staicu, Scheipl, and Ruppert [2014]

McLean et al. [2014] describe, implement and evaluate semiparametric non-linear effects of functional covariates, i.e., effects of the form $h(x) = \int F(x(s), s)ds$, for scalar responses.

Fuchs, Scheipl, and Greven [2015]

Fuchs et al. [2015] describe, implement and evaluate semiparametric linear interaction effects of functional covariates, i.e., effects of the form $h(x) = \int x_1(s)x_2(r)\beta(r, s)drds$, for scalar responses.

Goldsmith and Scheipl [2014]

Goldsmith and Scheipl [2014] evaluates the predictive performance of a broad variety of available estimators of effects of functional covariates on scalar responses and describes and compares possible ensembling strategies for the resulting predictions.

The following two papers describe extensions of some of the ideas and concepts from our approach to semiparametric regression with functional covariates to the historically much older field of *distributed lag*-models [Almon, 1965, Schwartz, 2000, e.g.], which is concerned with modeling the cumulative effect $h(x) = \sum_{l=l_0}^{L} x_{T-l}\beta(l)$ of a time-series of covariate values $x_0, \ldots, x_T$ on a scalar response y_T in a regression setting.

Obermeier, Scheipl, Heumann, Wassermann, and Küchenhoff [2015]

Obermeier et al. [2015] introduce, implement and evaluate a flexible scheme for such distributed lag effects that allows to estimate both the shape of the coefficient function $\beta(l)$ and the length of the relevant lag period L from the data via an innovative double penalization scheme.

Gasparrini, Scheipl, Armstrong, and Kenward [2017]

Gasparrini et al. [2017] extends the spline-based approach of nonlinear distributed lag effects $h(x) = \sum_{l=l_0}^{L} f(x_{T-l}, l)$ presented in Gasparrini et al. [2010] to allow for data-driven penalization of the effect function $f(x, l)$ and evaluates inferential properties of the resulting estimators for a variety of penalization schemes.

1.4 Summary

In a nutshell, the framework summarized above and detailed in the submitted publications

- accommodates both scalars or functions as predictors and responses, even if functional responses are observed on sparse or irregular grids,
- accommodates multiple scalars and/or functions as predictors,
- accommodates linear or nonlinear association of such responses with scalar and functional predictors,
- accommodates dependence structures between either functional or scalar responses,
- allows for a large diversity of effect shapes (linear, smooth non-linear, interactions, etc.) of the predictors,
- allows for a broad range of response distributions even beyond the classical exponential families and more general loss functions (quantile regression, Huber loss, etc.) and has recently been extended to models for location, scale and shape with multiple additive predictors associated with different moments or parameters of the conditional response distribution,
- re-uses and adapts established methodology for additive models, with estimation and inference based on either the penalized likelihood framework that has been so successful in the context of additive and mixed models or model-based componentwise gradient boosting,
- is implemented in fully documented and open-source R packages for both inferential paradigms: package `refund` [Huang et al., 2016] provides the implementation for penalized likelihood-based inference using `mgcv` [Wood, 2016b] as its computational backend, while `FDboost` [Brockhaus and Rügamer, 2016] does the same for model-based componentwise gradient boosting-based inference, using `mboost` [Hothorn et al., 2016] and `gamboostLSS` [Hofner et al., 2016] as its computational back-ends. Fully Bayesian inference is available via `mgcv` and `RJAGS` [Plummer, 2016] as well, c.f. Wood [2016a].
- has cross-fertilized the somewhat related field of distributed lag models and improved the state of the art there by introducing and implementing recent advances for semi-parametric inference in this area.

2 Versicherung an Eides Statt

Hiermit versichere ich an Eides Statt durch meine Unterschrift, dass ich bei der Anfertigung der vorliegenden Habilitationsleistung keine weiteren als die hier angegebenen Hilfsmittel benutzt habe, und dass kein wissenschaftliches Fehlverhalten im Sinne der Richtlinien der Ludwigs-Maximilians-Universität München zur Selbstkontrolle in der Wissenschaft in der Fassung vom 16. Mai 2002 (geändert durch Beschlüsse des Senats vom 22.6.2006, 11.2.2010, 30.9.2014) vorliegt.

Ich versichere desweiteren dass ich an keiner anderen Hochschule außer der Ludwigs-Maximilians-Universität München zur Habilitation angemeldet bin oder jemals an einer anderen Hochschule außer der Ludwigs-Maximilians-Universität München zur Habilitation angemeldet gewesen war.

Fabian Scheipl

References

Shirley Almon. The distributed lag between capital appropriations and expenditures. *Econometrica: Journal of the Econometric Society*, pages 178–196, 1965.

Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015. doi: 10.18637/jss.v067.i01.

Christiane Belitz, Andreas Brezger, Nadja Klein, Thomas Kneib, Stefan Lang, and Nikolaus Umlauf. *BayesX - Software for Bayesian inference in structured additive regression models.*, 2017. URL http://http://www.bayesx.org.

Sarah Brockhaus and David Rügamer. *FDboost: Boosting Functional Regression Models*, 2016. URL http://CRAN.R-project.org/package=FDboost. R package version 0.2-0.

Sarah Brockhaus, Fabian Scheipl, Torsten Hothorn, and Sonja Greven. The functional linear array model. *Statistical Modelling*, 15(3):279–300, 2015.

Sarah Brockhaus, Andreas Fuest, Andreas Mayr, and Sonja Greven. Signal regression models for location, scale and shape with an application to stock returns. *arXiv preprint arXiv:1605.04281*, 2016.

Peter Bühlmann and Torsten Hothorn. Boosting algorithms: Regularization, prediction and model fitting. *Statistical Science*, 22(4):477–505, 2007.

Antonio Cuevas. A partial overview of the theory of statistics with functional data. *Journal of Statistical Planning and Inference*, 147:1–23, 2014.

Iain D Currie, Maria Durban, and Paul HC Eilers. Generalized linear array models with applications to multidimensional smoothing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(2):259–280, 2006.

Ian L Dryden and Kanti V Mardia. *Statistical Shape Analysis: With Applications in R*. John Wiley & Sons, 2016.

Ludwig Fahrmeir, Thomas Kneib, and Stefan Lang. Penalized structured additive regression for space-time data: a Bayesian perspective. *Statistica Sinica*, pages 731–761, 2004.

Ludwig Fahrmeir, Thomas Kneib, Stefan Lang, and Brian Marx. *Regression: Models, Methods and Applications*. Springer Science & Business Media, 2013.

Manuel Febrero-Bande and Manuel Oviedo de la Fuente. Statistical computing in functional data analysis: The R package fda.usc. *Journal of Statistical Software*, 51(4):1–28, 2012. URL http://www.jstatsoft.org/v51/i04/.

Frédéric Ferraty and Philippe Vieu. *Nonparametric Functional Data Analysis: Theory and Practice*. Springer New York, 2006.

Frédéric Ferraty, André Mas, and Philippe Vieu. Nonparametric regression on functional data: Inference and practical aspects. *Australian & New Zealand Journal of Statistics*, 49 (3):267–286, 2007.

Jerome H Friedman. Multivariate adaptive regression splines. *The Annals of Statistics*, pages 1–67, 1991.

Karen Fuchs, Fabian Scheipl, and Sonja Greven. Penalized scalar-on-functions regression with interaction term. *Computational Statistics & Data Analysis*, 81:38–51, 2015.

Francis Galton. *Natural Inheritance*. Macmillan, 1894.

Antonio Gasparrini, Ben Armstrong, and Mike G Kenward. Distributed lag non-linear models. *Statistics in Medicine*, 29(21):2224–2234, 2010.

Antonio Gasparrini, Fabian Scheipl, Ben Armstrong, and Michael G Kenward. A penalized framework for distributed lag non-linear models. *Biometrics*, 2017. doi: 10.1111/biom.12645. in press.

Jeff Goldsmith and Fabian Scheipl. Estimator selection and combination in scalar-on-function regression. *Computational Statistics & Data Analysis*, 70:362–372, 2014.

Sonja Greven and Fabian Scheipl. A general framework for functional regression modelling. *Statistical Modelling*, 17(1-2):1–35, 2017a.

Sonja Greven and Fabian Scheipl. Rejoinder. *Statistical Modelling*, 17(1-2):100–115, 2017b.

Chong Gu and Grace Wahba. Semiparametric analysis of variance with tensor product thin plate splines. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 353–368, 1993.

Wolfgang Härdle. *Applied Nonparametric Regression*. Cambridge University Press, 1990.

Trevor Hastie and Robert Tibshirani. Varying-coefficient models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 55(4):757–796, 1993.

Trevor J Hastie and Robert J Tibshirani. *Generalized Additive Models*. CRC press, 1990.

Trevor J. Hastie, Robert John Tibshirani, and Jerome H Friedman. *The Elements of Statistical Learning: Data mining, Inference, and Prediction*. Springer New York, 2011.

R. Herrick. *WFMM*. The University of Texas M.D. Anderson Cancer Center, version 3.0 edition, 2015. URL https://biostatistics.mdanderson.org/SoftwareDownload.

Benjamin Hofner, Luigi Boccuto, and Markus Göker. Controlling false discoveries in high-dimensional situations: boosting with stability selection. *BMC bioinformatics*, 16(1):144, 2015.

Benjamin Hofner, Andreas Mayr, Nora Fenske, and Matthias Schmid. *gamboostLSS: Boosting Methods for GAMLSS Models*, 2016. URL http://CRAN.R-project.org/package=gamboostLSS. R package version 1.3-0.

Torsten Hothorn, Thomas Kneib, and Peter Bühlmann. Conditional transformation models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1):3–27, 2014.

Torsten Hothorn, Peter Buehlmann, Thomas Kneib, Matthias Schmid, and Benjamin Hofner. *mboost: Model-Based Boosting*, 2016. URL http://CRAN.R-project.org/package=mboost. R package version 2.6-0.

Lei Huang, Fabian Scheipl, Jeff Goldsmith, Jonathan Gellar, Jaroslaw Harezlak, Mathew W. McLean, Bruce Swihart, Luo Xiao, Ciprian Crainiceanu, and Philip Reiss. *refund: Regression with Functional Data*, 2016. URL https://CRAN.R-project.org/package=refund. R package version 0.1-15.

Andrada E Ivanescu, Ana-Maria Staicu, Fabian Scheipl, and Sonja Greven. Penalized function-on-function regression. *Computational Statistics*, 30(2):539–568, 2015.

EE Kammann and Matthew P Wand. Geoadditive models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52(1):1–18, 2003.

Roger Koenker. *Quantile Regression*. Cambridge University Press, 2005.

Piotr Kokoszka and Matthew Reimherr. Discussion of 'A general framework for functional regression modelling' by Greven and Scheipl. *Statistical Modelling*, 17(1-2):45–49, 2017.

Kung-Yee Liang and Scott L Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73:13–22, 1986.

Wenceslao González Manteiga and Philippe Vieu. Statistics for functional data. *Computational Statistics & Data Analysis*, 51(10):4788 – 4792, 2007.

J Steve Marron and Andrés M Alonso. Overview of object oriented data analysis. *Biometrical Journal*, 56(5):732–753, 2014.

Brian D Marx and Paul HC Eilers. Generalized linear regression on sampled signals and curves: a P-spline approach. *Technometrics*, 41(1):1–13, 1999.

Andreas Mayr, Nora Fenske, Benjamin Hofner, Thomas Kneib, and Matthias Schmid. Generalized additive models for location, scale and shape for high dimensional data - a flexible approach based on boosting. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 61(3):403–427, 2012.

Peter McCullagh. Regression models for ordinal data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 42(2):109–142, 1980.

Peter McCullagh and John A Nelder. *Generalized Linear Models*. Chapman and Hall, 1984.

Charles E McCulloch and John M Neuhaus. *Generalized linear mixed models*. Wiley Online Library, 2001.

Mathew W McLean, Giles Hooker, Ana-Maria Staicu, Fabian Scheipl, and David Ruppert. Functional generalized additive models. *Journal of Computational and Graphical Statistics*, 23(1):249–269, 2014.

Nicolai Meinshausen and Peter Bühlmann. Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(4):417–473, 2010.

Jeff S. Morris. Functional Regression. *Annual Review of Statistics and its Applications*, 2: 321–359, 2015.

Jeff S. Morris. Comparison and contrast of two general functional regression modelling frameworks. *Statistical Modelling*, 17(1-2):59–85, 2017.

Jeff S. Morris and Raymond J. Carroll. Wavelet-based functional mixed models. *Journal of the Royal Statistical Society, Series B*, 68(2):179–199, 2006.

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

Viola Obermeier, Fabian Scheipl, Christian Heumann, Joachim Wassermann, and Helmut Küchenhoff. Flexible distributed lags for modelling earthquake data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 64(2):395–412, 2015.

Martyn Plummer. *rjags: Bayesian Graphical Models using MCMC*, 2016. URL https://CRAN.R-project.org/package=rjags. R package version 4-6.

James 0 Ramsay and Bernhard W Silverman. *Functional Data Analysis*. Springer Science & Business Media, 2005.

James O. Ramsay, Hadley Wickham, Spencer Graves, and Giles Hooker. *fda: Functional Data Analysis*, 2014. URL http://CRAN.R-project.org/package=fda. R package version 2.4-4.

J.O. Ramsay, Spencer Graves, and Giles Hooker. *Functional data analysis with R and MATLAB*. Springer, New York, 2009.

Philip T Reiss, Jeff Goldsmith, Han Lin Shang, and R Todd Ogden. Methods for scalar-on-function regression. *International Statistical Review*, 2016. in press.

Robert A Rigby and D Mikis Stasinopoulos. Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(3): 507–554, 2005.

David Ruppert, Matt P Wand, and Raymond J Carroll. *Semiparametric Regression*. Cambridge University Press Cambridge, 2003.

Fabian Scheipl and Sonja Greven. Identifiability in penalized function-on-function regression models. *Electronic Journal of Statistics*, 10(1):495 – 526, 2016.

Fabian Scheipl, Thomas Kneib, and Ludwig Fahrmeir. Penalized likelihood and bayesian function selection in regression models. *AStA Advances in Statistical Analysis*, 97(4): 349–385, 2013.

Fabian Scheipl, Ana-Maria Staicu, and Sonja Greven. Functional additive mixed models. *Journal of Computational and Graphical Statistics*, 24(2):477–501, 2015.

Fabian Scheipl, Jan Gertheiss, and Sonja Greven. Generalized functional additive mixed models. *Electronic Journal of Statistics*, 10(1):1455–1492, 2016.

Matthias Schmid and Torsten Hothorn. Boosting additive models using component-wise p-splines. *Computational Statistics & Data Analysis*, 53(2):298–311, 2008.

Joel Schwartz. The distributed lag between air pollution and daily deaths. *Epidemiology*, 11 (3):320–326, 2000.

Rajen D Shah and Richard J Samworth. Variable selection with error control: another look at stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(1):55–80, 2013.

Jian Qing Shi and Yafeng Cheng. *GPFDA: Apply Gaussian Process in Functional data analysis*, 2014. URL https://CRAN.R-project.org/package=GPFDA. R package version 2.2.

Jian Qing Shi and Taeryon Choi. *Gaussian Process Regression Analysis for Functional Data*. CRC Press Boca Raton, 2011.

Stan Development Team. *Stan Modeling Language Users Guide and Reference Manual, Version 2.12.0*, 2016. URL http://mc-stan.org.

Nikolaus Umlauf, Nadja Klein, Achim Zeileis, and Meike Koehler. *bamlss: Bayesian Additive Models for Location Scale and Shape (and Beyond).*, 2016. URL http://CRAN.R-project.org/package=bamlss. R package version 0.1-1.

Jane-Ling Wang, Jeng-Min Chiou, and Hans-Georg Mueller. Review of Functional Data Analysis. *Annual Review of Statistics and Its Application*, 3(1):257–295, 2016.

Simon N Wood. *Generalized Additive Models: An Introduction with R*. CRC Boca Raton, 2006a.

Simon N Wood. Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics*, 62(4):1025–1036, 2006b.

Simon N Wood. Just another Gibbs additive modeler: Interfacing JAGS and mgcv. *Journal of Statistical Software*, 75(7):1–15, 2016a.

Simon N Wood. *mgcv: Mixed GAM Computation Vehicle with GCV/AIC/REML Smoothness Estimation*, 2016b. URL http://CRAN.R-project.org/package=mgcv. R package version 1.8-12.

Simon N Wood, Fabian Scheipl, and Julian J Faraway. Straightforward intermediate rank tensor product smoothing in mixed models. *Statistics and Computing*, pages 1–20, 2013.

Simon N Wood, Natalya Pya, and Benjamin Säfken. Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association*, 111 (516):1548–1563, 2016.

Simon N Wood, Zheyuan Li, Gavin Shaddick, and Nicole H Augustin. Generalized additive models for gigadata: modelling the UK black smoke network daily data. *Journal of the American Statistical Association*, (accepted), 2017.