

# Boosting functional response models for location, scale and shape with an application to bacterial competition

Almond Stöcker<sup>1</sup>, Sarah Brockhaus<sup>2</sup>, Sophia Anna Schaffer<sup>3</sup>, Benedikt von Bronk<sup>3</sup>, Madeleine Opitz<sup>3</sup> and Sonja Greven<sup>1</sup>

<sup>1</sup>School of Business and Economics, HU Berlin, Germany.

<sup>2</sup>Department of Statistics, LMU Munich, Germany.

<sup>3</sup>Department of Physics, LMU Munich, Germany.

**Abstract:** We extend generalized additive models for location, scale and shape (GAMLSS) to regression with functional response. This allows us to simultaneously model point-wise mean curves, variances and other distributional parameters of the response in dependence of various scalar and functional covariate effects. In addition, the scope of distributions is extended beyond exponential families. The model is fitted via gradient boosting, which offers inherent model selection and is shown to be suitable for both complex model structures and highly auto-correlated response curves. This enables us to analyse bacterial growth in *Escherichia coli* in a complex interaction scenario, fruitfully extending usual growth models.

**Key words:** bacterial growth, distributional regression, functional data, functional regression, GAMLSS

Received May 2019; revised December 2019; accepted March 2020

## 1 Introduction

In functional data analysis (Ramsay and Silverman, 2005), functional response regression aims at estimating covariate effects on response curves (Morris, 2015; Greven and Scheipl, 2017). The response curves might, for instance, be given by annual temperature curves, growth curves or spectroscopy data. We propose a flexible approach to regression with functional response allowing for simultaneously modelling multiple distributional characteristics of response curves following potentially non-Gaussian distribution families. It therefore generalizes usual functional mean regression models.

The problem of non-Gaussian functional response appears in many applications and is, accordingly, addressed in several publications. Following early works on non-Gaussian functional data by Hall et al. (2008) and van der Linde (2009),

---

Address for correspondence: Almond Stöcker, School of Business and Economics, Chair of Statistics, Humboldt University of Berlin, Unter den Linden 6, D–100 99 Berlin, Germany.

E-mail: [almond.stoecker@hu-berlin.de](mailto:almond.stoecker@hu-berlin.de)



authors such as Goldsmith et al. (2015), Wang and Shi (2014) and Scheipl et al. (2016) proposed generalized linear mixed model (GLMM)-type regression models, which are suitable for, for example, positive, discrete or integer-valued response functions. The linear predictor for the mean function is typically composed of a covariate effect term and a latent random Gaussian error process accounting for auto-correlation, and combined with a link function. Li et al. (2014) jointly model continuous and binary-valued functional responses in a similar fashion, but without considering covariates. Other ideas to account for the auto-correlation of functional response curves include robust covariance estimation for valid inference after estimation under a working independence assumption (Gertheiss et al., 2015) and an overall regularization by early stopping a gradient boosting fitting procedure guided by curve-wise re-sampling methods as applied by Brockhaus et al. (2015, 2017) besides using curve-specific smooth errors. Moreover, this latter approach also offers quantile regression for functional data. The above approaches present important steps in generalizing functional regression models. In particular, our approach is a direct generalization of the framework of Brockhaus et al. (2015). However, the previous methods are restricted to one predictor such that none of them allows for simultaneously modelling also the response variance or other distributional parameters in a similar fashion as the mean. Staicu et al. (2012) propose a method for estimating mean, variance and other shape parameter functions nonparametrically and also for non-Gaussian point-wise distributions, while modelling auto-correlation via copulas. However, they do not allow for including covariate effects. Only the framework of Scheipl et al. (2016) now allows to perform simultaneous mean and variance regression in the Gaussian case (Greven and Scheipl, 2017) and is also the only one currently offering a comparable range of smooth/linear effects of scalar and functional covariates, which are implemented in the R package `refund` (Goldsmith et al., 2018). Thus, we will compare to their model in a simulation. However, they are so far restricted to the Gaussian special case and do not offer the flexibility to specify multiple predictors for other distributions and no more than two predictors, as we need, in particular, in the data scenario presented in this article. We overcome these limitations by introducing generalized additive models for location, scale and shape (GAMLSS) for functional responses. For the case of scalar response regression, GAMLSS were introduced by Rigby and Stasinopoulos (2005) extending usual generalized additive models (GAMs; Hastie and Tibshirani, 1990) to multiple distributional parameters. Each parameter of the assumed response distribution is modelled with a separate predictor depending on covariates, allowing, for example, for covariate effects on mean and variance. Hence, doubtful assumptions of homoscedasticity can be overcome. In addition to this extension, the range of applicable distributions of GAMs is also extended to non-exponential family distributions in the GAMLSS framework. Brockhaus et al. (2018a) discussed GAMLSS scalar-on-function regression based on the flexible gradient boosting regression framework introduced by Bühlmann and Hothorn (2007) combining a scalar GAMLSS framework developed by Mayr et al. (2012) and Thomas et al. (2018) and the flexible functional regression framework of Brockhaus et al. (2017). We further extend this to functional GAMLSS for

function-on-scalar and function-on-function regression, such that our framework now offers the full flexibility of scalar GAMLSS also for functional responses and covariates.

We apply this framework to analyse bacterial interaction: as bacterial resistances increase, producing effective antibiotics gets harder and harder. Understanding bacterial interactions might help finding alternatives. In particular, we analyse growth curves of two competing *Escherichia coli* bacteria strains (von Bronk et al., 2017)—a toxin producing ‘C-strain’ and a toxin sensitive ‘S-strain’—to obtain insights into the underlying growth affecting bacterial interaction. Our aim is to model the S-strain growth behaviour in dependence on the toxin emitting C-strain and under different experimental conditions, and allow this dependence to affect both mean and variability of growth as well as the extinction probability. This requires a functional response regression model for several parameters of the non-Gaussian response distribution with linear and smooth effects of functional and scalar covariates.

There are various approaches applied to modelling bacterial growth curves in the literature. Gompertz and Baranyi-Roberts models are two common parametric approaches to modelling growth curves (see, e.g., López et al., 2004; Perni et al., 2005). Weber et al. (2014) implement a model particularly for analysing bacterial interaction. The models are usually fitted using least squares methods, which corresponds to assuming a Gaussian distribution of bacterial propagation. This is problematic as response values are naturally positive and very small in the beginning, starting from single cell level. Thus, also assuming a constant variance over the whole time span seems not appropriate. Moreover, they do not offer the opportunity to include covariate effects for modelling, for example, the impact of external factors. Thus, these models are not applicable here. In addition, they are often highly non-linear, which may introduce problems in parameter estimation. Gasser et al. (1984) discuss this point and some further advantages of nonparametric growth curve regression over parametric models. They propose a kernel method for this purpose, which does, however, not include covariates and lacks the flexibility needed for our purposes. Still, non- or semi-parametric functional regression models present a natural choice from a statistical perspective, also because they can approximate the above parametric growth models very well when these are appropriate (Online Appendix E.2).

Besides providing the flexibility to meet all the challenges arising from the present analysis of bacterial interaction, we can show in extensive simulation studies that the presented approach is indeed well suited for complex scenarios with highly auto-correlated response curves despite working independence assumption: early stopping the gradient boosting algorithm based on curve-wise re-sampling techniques plays a key role in avoiding over-fitting and leads to highly improved estimation quality when comparing to the approach of Greven and Scheipl (2017).

The approach is implemented in the R (R Core Team, 2018) add-on package `FDboost` (Brockhaus and Ruegamer, 2018). Brockhaus et al. (2018b) provide a tutorial article to the package. Even though the discussion and illustration of GAMLSS focuses on the scalar-on-function and not the functional response case, we recommend it as a general software introduction.

The remainder of the article is structured as follows: In Section 2, we formulate the general model and describe the fitting algorithm. In Section 3, we apply the proposed model to analysing *E. coli* bacteria growth. Section 4 provides the results of two simulation studies for Gaussian response curves as well as for the growth model. Section 5 concludes with a discussion. Further details concerning the model, application and simulation studies are provided as Online Supplement, as well as the code for the simulations and fitting of the model with the R-package `FDboost`.

## 2 Model formulation

Consider a data scenario with  $N$  observations of a functional response  $Y$  and respective covariates  $\mathbf{X}$ .  $Y$  is a stochastic process, such that its realized trajectories  $y_i : \mathcal{T} \rightarrow \mathbb{R}$ ,  $t \mapsto y_i(t)$  for  $i = 1, \dots, N$  represent the response curves over an index set  $\mathcal{T}$ . For notational simplicity, we assume that response curves are observed on a common grid  $\mathcal{T}_0 \subset \mathcal{T}$ , where  $\mathcal{T} = [0, t_{max}]$  is a real interval starting at zero and  $\mathcal{T}_0$  a finite discrete set of evaluation points. However, the curves could be measured on different grids as well. As this is the case in many applications, the variable  $t$  is referred to as time variable. Scalar response is contained as special case where  $\mathcal{T}$  is a one point set. Let  $\mathbf{x}_i = (x_{i,1}, \dots, x_{i,p})^\top$  denote the  $i$ -th observed covariate vector, that is, realization of  $\mathbf{X}$ , which can contain scalar and functional covariates. A functional covariate may have a different domain  $\mathcal{S}$  from the response and is denoted as  $x_{i,j} : \mathcal{S} \rightarrow \mathbb{R}$ ,  $s \mapsto x_{i,j}(s)$ . We suppress potential dependence of  $\mathcal{S}$  on  $j$  in our notation.

We assume that for all  $t \in \mathcal{T}$  the point-wise response distribution  $\mathcal{F}_{Y(t)|\mathbf{X}}$  is known up to the distribution parameters  $\vartheta(t) = (\vartheta^{(1)}(t), \dots, \vartheta^{(Q)}(t))^\top$ . For instance, for a Gaussian process, the parameters might represent the conditional mean and variance over time, that is,  $\vartheta^{(1)}(t) = \mathbb{E}(Y(t)|\mathbf{X} = \mathbf{x})$  and  $\vartheta^{(2)}(t) = \mathbb{V}\text{ar}(Y(t)|\mathbf{X} = \mathbf{x})$ , suppressing the dependence on  $\mathbf{x}$  in the notation. For each parameter an additive regression model is assumed. The model is specified by

$$g^{(q)}(\vartheta^{(q)}) = h^{(q)}(\mathbf{x}) = \sum_{j=1}^{J^{(q)}} h_j^{(q)}(\mathbf{x}) \quad , \quad q = 1, \dots, Q,$$

where  $g^{(q)}$  is a monotonic link function for the  $q$ th distribution parameter. This model structure corresponds to the GAMLSS introduced by Rigby and Stasinopoulos (2005). However, covariates and response may now be functions. Correspondingly,  $\vartheta^{(q)} := \vartheta^{(q)}(\cdot)$  and the predictor  $h^{(q)}(\mathbf{x}) := h^{(q)}(\mathbf{x}, \cdot)$  are now functions over the domain  $\mathcal{T}$  of the response, for  $q = 1, \dots, Q$ . For  $Q = 1$  parameter corresponding to the mean, the model reduces to the functional additive regression model of Scheipl et al. (2015, 2016) and Brockhaus et al. (2015).

Both covariate and time dependency are modelled within the additive predictor via the effect functions  $h_j^{(q)}(\mathbf{x}, t)$ . The predictor is typically composed of a functional

**Table 1** Overview of possible effect types (adapted from Brockhaus et al., 2015)

Covariate(s)	Type of Effect	$h_j^{(q)}$
(none)	Smooth intercept	$\beta_0(t)$
Scalar covariate $z$	Linear effect	$z \beta(t)$
	Smooth effect	$f(z, t)$
Two scalars $z_1, z_2$	Linear interaction	$z_1 z_2 \beta(t)$
	Functional varying coefficient	$z_1 f(z_2, t)$
	Smooth interaction	$f(z_1, z_2, t)$
Grouping variable $g$	Group-specific intercept	$\beta_g(t)$
Group. variable $g$ , scalar $z$	Group-specific linear effect	$z \beta_g(t)$
	Group-specific smooth effect	$f_g(z, t)$
Group. variables $g_1, g_2$ Functional covariate $x(s)$	Group-interaction	$\beta_{g_1, g_2}(t)$
	Functional linear effect	$\int x(s) \beta(s, t) ds$
Functional cov. $x(s)$ , scalar $z$	Linear interaction	$z \int x(s) \beta(s, t) ds$
	Smooth interaction	$\int x(s) \beta(z, s, t) ds$
Functional cov. $x(s)$ over $\mathcal{T}$	Concurrent effect	$x(t) \beta(t)$
	Historical effect	$\int_0^t x(s) \beta(s, t) ds$
	Effect with $t$ -specific integration limits	$\int_{l(t)}^{u(t)} x(s) \beta(s, t) ds$

intercept  $h_1^{(q)}(\mathbf{x}, t) = \beta_0(t)$  and linear or smooth covariate effects  $h_j^{(q)}(\mathbf{x}, t)$ , of which each depends on one or more covariates. The construction of the effects follows a modular principle, which allows for flexible specification of effect types and is outlined in the next subsection. Table 1 gives an overview of different effect types available and Section 2.2 will discuss inherent selection of effects within the fitting approach.

Apart from the Gaussian, a variety of other distributions can be specified for the response. In principle,  $\mathcal{F}_{Y(t)|X}$  can be any distribution for which both the likelihood and its derivatives are computable. The derivatives with respect to the parameters are required for the model estimation via gradient boosting. For functional response, usually only continuous distributions are under consideration. However, this does not necessarily have to be the case (see, e.g., Scheipl et al., 2016). As we built on the approach of Mayr et al. (2012) for scalar GAMLSS, all of the distributions implemented in their R package `gamboosLSS` (Hofner et al., 2017) are directly available for the present boosting approach. Moreover, they also provide an interface to use the comprehensive list of distributions available in the R package `gamlss.dist` (Stasinopoulos and Rigby, 2019) and custom distributions can be specified.

## 2.1 Construction of effect functions $h_j^{(q)}$

As both covariate and time dependency of the functional response are specified by the effect functions  $h_j^{(q)}$ , they play a key role in the framework. We briefly illustrate their modular structure and refer to Brockhaus et al. (2018b) and Greven and Scheipl (2017) for further examples and details, as their construction is not new to the functional GAMLSS. The novelty is, however, that we may now use them in multiple predictors for multiple parameter functions.

For each effect type,  $h_j^{(q)}$  is represented by a linear combination of specified basis functions, such that the predictor is linear in its coefficients. Multivariate basis functions are constructed as tensor products of univariate bases providing flexible modular means of specification (cf. Scheipl et al., 2015), giving the basis representation

$$h_j^{(q)}(\mathbf{x}, t) = \left( \mathbf{b}_{X_j}^{(q)}(\mathbf{x}, t) \otimes \mathbf{b}_{Y_j}^{(q)}(t) \right)^\top \boldsymbol{\theta}_j^{(q)} \quad , t \in \mathcal{T}. \quad (2.1)$$

A vector  $\mathbf{b}_{Y_j}^{(q)}$  of  $K_{Y_j}^{(q)}$  basis functions for the time variable is combined with a vector  $\mathbf{b}_{X_j}^{(q)}$  of  $K_{X_j}^{(q)}$  basis functions for the covariate effects. The basis  $\mathbf{b}_{X_j}^{(q)}(\mathbf{x}, t)$  might be time dependent, for example, for a functional historical effect. However, for many effect types, it only depends on covariates, such that we can write  $\mathbf{b}_{X_j}^{(q)}(\mathbf{x})$ . Applying the Kronecker product  $\otimes$ , a new basis is obtained. Its elements correspond to the pairwise products of elements in  $\mathbf{b}_{Y_j}^{(q)}$  and  $\mathbf{b}_{X_j}^{(q)}$ . For details, see Online Supplement A.1. The coefficient vector  $\boldsymbol{\theta}_j^{(q)} \in \mathbb{R}^{K_{Y_j}^{(q)} K_{X_j}^{(q)}}$  specifies the concrete form of the effect. Fitting the model corresponds to estimating  $\boldsymbol{\theta}_j^{(q)}$  for all effect functions.

A typical choice for  $\mathbf{b}_{Y_j}^{(q)}(t)$  is a spline basis. Then, in case of time-independent covariate basis functions  $\mathbf{b}_{X_j}^{(q)}$ ,  $h_j^{(q)}(\mathbf{x}_0, t)$  describes a spline curve for a fixed value  $\mathbf{x} = \mathbf{x}_0$  of the covariate. Usually, quadratic penalty terms are employed in order to control smoothness of the effect functions (see Section 2.2). A typical effect function  $h_j^{(q)}(\mathbf{x}, t)$  depends on a single covariate. For example, for a linear effect  $z\beta(t)$  of a scalar covariate  $z$ , this yields  $\mathbf{b}_{X_j}^{(q)}(\mathbf{x}, t) = \mathbf{b}_{X_j}^{(q)}(z) = z$ . In order to obtain a smooth covariate effect  $f(z, t)$ , a spline basis can be chosen for  $\mathbf{b}_{X_j}^{(q)}(z)$  just like for the time curve, yielding a tensor product spline basis in (2.1). For a functional covariate  $x : \mathcal{T} \mapsto \mathbb{R}$  a historical effect of the form  $\int_0^t x(s) \beta(s, t) ds$  can be constructed using a basis of time-dependent linear functionals  $\mathbf{b}_{j,k}^{(q)}(x, t) = \int_0^t x(s) \varphi_k(s) ds$ , where  $\varphi_k(s)$ ,  $k = 1, \dots, K_{X_j}^{(q)}$ , is a spline basis and the integral is numerically approximated over the observation grid of  $x$  in

$\mathcal{T}$ . Using also a spline basis for  $\mathbf{b}_{Y_j}^{(q)}(t)$ , this corresponds to specifying a tensor product spline basis for  $\beta(s, t)$ .

## 2.2 Model fit

Component-wise gradient boosting is a gradient descend method for model fitting, where the model is iteratively updated. In each iteration, the algorithm aims at minimizing a loss function following the direction of its steepest descent. Instead of updating the full additive predictor at once, the individual effect functions  $h_j^{(q)}$  are separately fit to the negative gradient in a component-wise approach. These individual effect models are called *base-learners*, as they present simple base models that jointly form the model predictor. In each iteration, only the effect function with the best fit is updated with a step length  $\nu$  in the direction of its fit. The component-wise and stepwise procedure yields automated model selection and allows for fitting models with more parameters than observations.

Let  $f(y(t) | \boldsymbol{\vartheta}(t)) = f_b(y(t) | \mathbf{h}(\mathbf{x}, t))$  with  $\mathbf{h} = (h^{(1)}, \dots, h^{(Q)})^\top$  denote the conditional probability density function (PDF) of the response at  $t \in \mathcal{T}$  for a given parameter setting. We define the point-wise loss function to be the negative log-likelihood

$$\varrho(y(t), \mathbf{h}(\mathbf{x}, t)) = -\log f_b(y(t) | \mathbf{h}(\mathbf{x}, t)) .$$

The functional GAMLSS loss function is then obtained as

$$\ell(y, \mathbf{h}(\mathbf{x})) = \int_{\mathcal{T}} \varrho(y(t), \mathbf{h}(\mathbf{x}, t)) dt ,$$

the integral over the point-wise loss functions over  $\mathcal{T}$ . Therefore, we assume that  $f_b$  and  $\mathbf{h}$  are chosen such that the integral exists, which is no restriction in practice. The aim of gradient boosting is to find the predictor

$$\mathbf{h}_{optimal} = \underset{\mathbf{h}}{\operatorname{argmin}} \mathbb{E}[\ell(Y, \mathbf{h}(\mathbf{X}))] = \underset{\mathbf{h}}{\operatorname{argmin}} \int_{\mathcal{T}} \mathbb{E}[\varrho(Y(t), \mathbf{h}(\mathbf{X}, t))] dt \quad (2.2)$$

minimizing the expected loss.

Based on data  $(y_i, \mathbf{x}_i)_{i=1, \dots, N}$ , this is estimated by optimizing the empirical mean loss. Hence, the estimated predictor vector  $\hat{\mathbf{h}} = (\hat{h}^{(1)}, \dots, \hat{h}^{(Q)})^\top$  is given by

$$\hat{\mathbf{h}} \approx \underset{\mathbf{h}}{\operatorname{argmin}} \frac{1}{N} \sum_{i=1}^N \hat{\ell}(y_i, \mathbf{h}(\mathbf{x}_i)) , \quad (2.3)$$

where  $\hat{\ell}(y_i, \mathbf{h}(\mathbf{x}_i)) = \sum_{t \in \mathcal{T}_0} \varrho(y_i(t), \mathbf{h}(\mathbf{x}_i, t))$  is an approximation of the loss. However, to avoid over-fitting, the optimization is generally not run until convergence. Instead, a re-sampling strategy is employed to find an optimal stopping iteration.

The minimization in (2.2) can be seen to minimize the Kullback–Leibler divergence (KLD) of the model density  $f_b$  to the true underlying density. Hastie and Tibshirani (1990) formulate a similar regression aim for GAMs. However, in the functional case, we consider the point-wise KLD integrated over the domain  $\mathcal{T}$ .

The base-learners fitted in each boosting iteration correspond to the effects  $h_j^{(q)}(\mathbf{x}, t)$  with  $j = 1, \dots, J^{(q)}$  and  $q = 1, \dots, Q$ . For any given loss function, they represent single regression models, which are fitted to the gradient of the loss function via penalized least squares. The coefficients  $\theta_j^{(q)}$  of the respective  $h_j^{(q)}$ , as defined in equation (2.1), are subject to a quadratic penalty of the form  $(\theta_j^{(q)})^\top \mathbf{P}_j^{(q)} \theta_j^{(q)}$ , where  $\mathbf{P}_j^{(q)}$  is a penalty matrix. As described for bivariate smooth terms, for example, in Wood (2006) or Brockhaus et al. (2015), the penalty matrix is constructed as  $\mathbf{P}_j^{(q)} = \lambda_{X_j}^{(q)} \left( \mathbf{P}_{X_j}^{(q)} \otimes \mathbf{I}_{K_{Y_j}^{(q)}} \right) + \lambda_{Y_j}^{(q)} \left( \mathbf{I}_{K_{X_j}^{(q)}} \otimes \mathbf{P}_{Y_j}^{(q)} \right)$  with smoothing parameters  $\lambda_{Y_j}^{(q)}, \lambda_{X_j}^{(q)} \geq 0$  and penalty matrices  $\mathbf{P}_{Y_j}^{(q)} \in \mathbb{R}^{K_{Y_j}^{(q)} \times K_{Y_j}^{(q)}}$  and  $\mathbf{P}_{X_j}^{(q)} \in \mathbb{R}^{K_{X_j}^{(q)} \times K_{X_j}^{(q)}}$  for the time basis  $\mathbf{b}_{Y_j}^{(q)}(t)$  and covariate basis  $\mathbf{b}_{X_j}^{(q)}(\mathbf{x}, t)$ , respectively. For instance, a common choice for B-spline bases is a first, or second-order difference penalty matrix yielding P-Splines (compare Eilers and Marx, 2010). Base-learners for group effects might be regularized with a ridge penalty. If no penalization should be applied for either the response or the covariates, this can also be obtained by setting  $\lambda_{Y_j}^{(q)} = 0$  or  $\lambda_{X_j}^{(q)} = 0$ , respectively. Thomas et al. (2018) compare different gradient boosting methods for GAMLSS, which can all be analogously generalized to functional response. While the ‘cyclic’ method and a ‘non-cyclic’ method are available in the R package `gamboostLSS`, only the algorithm of the ‘non-cyclic’ method is described here in detail. Comparing it to the ‘cyclic’ method, it performed better in simulations (Online Appendix Table 3), is faster (Online Appendix Figure 11) and provides the advantage of unified model selection across parameters  $\vartheta^{(q)}$ ,  $q = 1, \dots, Q$ .

---

*Algorithm: gradient boosting for functional GAMLSS*

---

1. To set up the model specify

- (a) a **functional loss function**  $\ell$  with point-wise loss  $\varrho$  corresponding to the assumed response distribution with  $Q$  distribution parameters
- (b) the **base-learners** by choosing the desired bases for the effects  $h_j^{(q)}(\mathbf{x}, t) = (\mathbf{b}_{X_j}^{(q)}(\mathbf{x}, t) \otimes \mathbf{b}_{Y_j}^{(q)}(t))^\top \theta_j^{(q)}$ , penalty matrices  $\mathbf{P}_j^{(q)}$  for all  $j = 1, \dots, J^{(q)}$  and  $q = 1, \dots, Q$  and their respective smoothing parameters.
- (c) **gradient boosting hyper-parameters**: the step-lengths  $v^{(q)} \in ]0, 1]$  for  $q = 1, \dots, Q$  and the maximum number of iterations  $m_{stop}$ .

Initialize the coefficients  $\theta_j^{(q)|0}$  for the initial predictor  $\mathbf{h}^{[0]}(\mathbf{x}_i, t)$ , for example, to 0, and set  $m = 0$ .

2. For  $m = 0, \dots, m_{stop} - 1$  iterate:

- i) Find best update for each distribution parameter.  
For  $q = 1, \dots, Q$  do:



(a) Evaluate negative partial gradients for  $i = 1, \dots, N$  at the current predictor  $\mathbf{h}^{[m]}$

$$u_i^{(q)}(t) := -\frac{\partial \mathcal{O}}{\partial b^{(q)}}(y_i(t), \mathbf{h}) \Big|_{\mathbf{h}=\mathbf{h}^{[m]}(\mathbf{x}_i, t)}$$

(b) Fit base-learners to the gradients, that is, for  $j = 1, \dots, J^{(q)}$  find  $\tilde{\theta}_j^{(q)}$  with

$$\tilde{\theta}_j^{(q)} := \operatorname{argmin}_{\theta_j^{(q)}} \left\{ \sum_{i=1}^N \sum_{t \in \mathcal{T}_0} \left( u_i^{(q)}(t) - (\mathbf{b}_{X_i}^{(q)}(\mathbf{x}_i, t) \otimes \mathbf{b}_{Y_i}^{(q)}(t))^\top \theta_j^{(q)} \right)^2 + (\theta_j^{(q)})^\top \mathbf{P}_j^{(q)} \theta_j^{(q)} \right\}$$

• Determine the best-fitting base-learner with index  $\bar{j}$  following the least squares criterion

$$\bar{j} := \operatorname{argmin}_j \sum_{i=1}^N \sum_{t \in \mathcal{T}_0} \left( u_i^{(q)}(t) - (\mathbf{b}_{X_i}^{(q)}(\mathbf{x}_i, t) \otimes \mathbf{b}_{Y_i}^{(q)}(t))^\top \tilde{\theta}_j^{(q)} \right)^2$$

• Determine updated predictor candidate, that is, determine  $\mathfrak{h}$  where only the coefficients of the best-fitting base-learner are updated, such that the coefficients are given by

$$\mathfrak{\theta}_k^{(p)} = \begin{cases} \theta_k^{(p)[m]} + \nu^{(p)} \tilde{\theta}_k^{(p)} & \text{for } p = q, k = \bar{j}, \\ \theta_k^{(p)[m]} & \text{else} \end{cases}$$

end for.

ii) Select best update across the distributional parameters and update the linear predictor accordingly

$$\mathbf{h}^{[m+1]} = \operatorname{argmin}_{\mathfrak{h}} \sum_{i=1}^N \hat{\ell}(y_i, \mathfrak{h}(\mathbf{x}_i))$$

end for.

The smoothing parameters for the penalty matrices  $\mathbf{P}_{jY}^{(q)}$  can be chosen indirectly specifying the base-learner degrees of freedom, as described by Hofner et al. (2011). They are typically specified such that equal degrees of freedom for all base-learners are attained to ensure a fair base-learner selection. Note that these degrees of freedom only specify the flexibility of each base-learner for one iteration, while the final effective degrees of freedom can be higher due to repeated selection of the same base-learner.  $\nu = 0.1$  is a popular choice for the step-length (Bühlmann and Hothorn, 2007). It should be chosen small enough to prevent overshooting. Yet, too small values greatly increase computation time. The optimal stopping iteration  $m_{stop}$ , with respect to equation (2.2), is the main tuning parameter. It can be estimated using, for example, curve-wise cross-validation or bootstrapping. As determining  $\mathbf{h}^{[m_{stop}]}$  involves computation of all earlier predictors, this can be done very efficiently (and in parallel over cross-validation folds). Early stopping induces regularization of effect functions and provides automated model selection: effect functions  $h_j^{(q)}$  which were never selected drop out of the model. As each base-learner is fitted separately, models with more covariates than observations can be fit and computational effort scales linearly in the number of covariate effects. By appropriately decomposing terms

into, for example, a linear and a non-linear base-learner, we cannot only select covariates, but also distinguish linear effects from smooth effects depending on the same covariate (compare Kneib et al., 2009) and covariate interactions from additive marginal effects (see Online Appendix A.2).

### 3 Analysis of bacterial interaction in *E. coli*

The coexistence of various bacterial species is a key factor in environmental systems. Equilibria in this biodiversity stand or fall with the species' interaction. Certain bacteria strains produce toxins and use them to assert themselves in bacterial competition. von Bronk et al. (2017) establish an experimental set-up with two cohabiting *Escherichia coli* bacteria strains: a 'C-strain' producing the toxin ColicinE2 and a colicin sensitive 'S-strain' pipetted together on an agar surface. Single bacteria of the C-strain population sacrifice themselves in order to liberate colicin. The emitted colicin diffuses through the agar and kills numerous S-strain bacteria on contact. On the other hand, the S-strain might outgrow the C-strain and starts in a favoured position of an initial ratio S:C of about 100:1. The arising population dynamics are influenced by external stress induced with the antibiotic agent Mitomycin C (MitC). MitC slightly damages the DNA of the bacteria. While it has little effect on the S-strain, it triggers colicin production in the C-strain as an SOS-response. A higher dose of MitC increases the fraction of colicin producing C-bacteria and, thus, colicin emission (von Bronk et al., 2017).

At a total of  $N = 334$  observation sites, bacteria under consideration are exposed to one of four different MitC concentrations. Bacterial growth curves  $S_i(t)$  of the S-strain and  $C_i(s)$  of the C-strain,  $i = 1, \dots, N$ , are observed over 48 hours. Their values correspond to the propagation areas of the bacterial strains, which are obtained from the automated image segmentation procedure implemented by von Bronk et al. (2017). S- and C-strain areas can be distinguished as the bacteria are marked with red and green fluorescence, respectively. The resulting area growth curves are measured on a fixed time grid with  $G = 105$  measurements per curve. The experiments are conducted in batches of about 40 bacterial spots and with two batches for each MitC concentration. In order to keep track of bacterial growth, the zoom level of the microscope was adjusted after  $12^{1/4}h$ ,  $18^{1/2}h$  and  $33^{1/2}h$ . As the performance of the automatic bacterial area segmentation may depend on the zoom level, it has to be incorporated into the analysis.

#### 3.1 Model for S-strain growth

In order to obtain insights into bacterial interaction dynamics, we model the  $i$ -th propagation area curve of the S-strain  $S_i(t)$  in dependence on the C-strain growth and other covariates. While usually  $S_i(t) > 0$ , it might equal zero, if the S-strain is completely extinct or masked by the fluorescence of the C-strain. Therefore, we assume a conditional zero adjusted gamma (ZAGA) distribution

for  $S_i(t)$ , which is a mixed continuous and discrete distribution with its PDF given by  $f_{ZAGA}(s_i|\mu_i, \sigma_i/\mu_i, p_i) = p_i \delta_{s_i} + f_{GA}(s_i|\mu_i, \sigma_i/\mu_i)(1 - \delta_{s_i})$  with  $\delta_{s_i} = 1$  if  $s_i = 0$  and 0 otherwise and  $f_{GA}$  the density of a gamma distribution parametrized by its mean  $\mu_i(t)$  and the coefficient of variation  $\sigma_i(t)/\mu_i(t)$  with  $\sigma_i(t)$  the standard deviation (Stasinopoulos and Rigby, 2019). This corresponds to some extent to the zero adjustment in a zero-inflated Poisson model. However, unlike the Poisson distribution, the gamma distribution is continuous and does not have a point mass at zero by itself. For  $S_i(t) > 0$ , it offers the flexibility to model both a location and a scale parameter conditional on the survival of the S-strain at time  $t$ , while in addition, we model the probability of extinction of the S-strain,  $p_i(t) = P(S_i(t) = 0)$  over time.

Each component of the resulting parameter vector  $\vartheta_i(t) = \left( \vartheta_i^{(\mu)}(t), \vartheta_i^{(\sigma/\mu)}(t), \vartheta_i^{(p)}(t) \right)^\top = \left( \mu_i(t), \frac{\sigma_i(t)}{\mu_i(t)}, p_i(t) \right)^\top$  is modelled as

$$g^{(q)} \left( \vartheta_i^{(q)}(t) \right) = \beta_0^{(q)}(t) + \beta_{MitC_i}^{(q)}(t) + \beta_{Batch_i}^{(q)}(t) + h_1^{(q)}(C_i, t) + h_2^{(q)}(C'_i, t)$$

for  $q \in \{\mu, \sigma/\mu, p\}$ , with link-functions  $g^{(\mu)} = g^{(\sigma/\mu)} = \log$  and  $g^{(p)} = \text{logit}$ , and with historical effects  $h_j^{(q)}(C_i, t) = \int_0^t C_i(s) \beta_j^{(q)}(s, t) ds$  (compare Brockhaus et al., 2017).

For each distribution parameter, the model includes a functional intercept  $\beta_0^{(q)}(t)$ . As there are only four MitC concentrations employed, they are considered as categorical grouping variable and represented by group-specific intercepts  $\beta_{MitC_i}^{(q)}(t)$  per MitC level centred around the functional intercept. As a functional random intercept, we include an additional group-specific intercept  $\beta_{Batch_i}^{(q)}(t)$  to compensate for batch effects, which are centred around  $\beta_{MitC_i}^{(q)}(t)$  in order to preserve identifiability of the functional intercept. The impact of the C-strain on S-strain growth is modelled using historical effects with coefficient functions  $\beta_j^{(q)}(s, t)$ . Historical effects are included both for the current C-strain propagation  $C_i(s)$  and for its derivative  $C'_i(s)$  reflecting the current C-strain growth. The covariate curves are centred around their empirical point-wise mean curve, such that  $\frac{1}{N} \sum_{i=1}^N C_i(s) = 0$  for each  $s$ , and scaled with the corresponding standard deviation, such that  $\text{sd}(C(s)) = 1$ . For  $C'_i(s)$  correspondingly. Doing so, the coefficient functions can be uniformly interpreted over the whole time span. By integrating, the historical effect includes information about the curves from time point  $t = 0$  to the current time point  $t$ . For  $p$ , we include an additional step-function base-learner to capture the different zoom levels applied during the experiment at fixed known time points, which can lead to different visibility of small S populations. Corresponding effects are not expected to be necessary for  $\mu$  and  $\sigma/\mu$  and are thus not included, as they might even lead to spurious boundary effects in this experimental set-up. Apart from the step function, all effect functions are modelled with cubic P-splines and second-order difference penalties,

such that for the functional intercepts we penalize deviations from exponential growth when employing a log-link for the mean and the scale parameter. For the MitC and batch effects, a ridge-type penalty over factor levels is utilized to achieve the same number of effective degrees of freedom for all base-learners. A common step-length of  $\nu = 0.1$  is used for  $\mu$ ,  $\sigma/\mu$  and  $p$ . We fit the model with both implemented GAMLSS boosting methods and decide for the ‘non-cyclic’ method, described in Section 2.2, which performed better in ten fold curve-wise bootstrapping and is computationally more efficient. With a maximum of 3 000 boosting iterations, the model fit took less than 16 min on a 64-bit Windows laptop followed by 156 min of bootstrapping without parallelization. The latter can be easily accelerated by running it on several cores in parallel.

## 3.2 Results

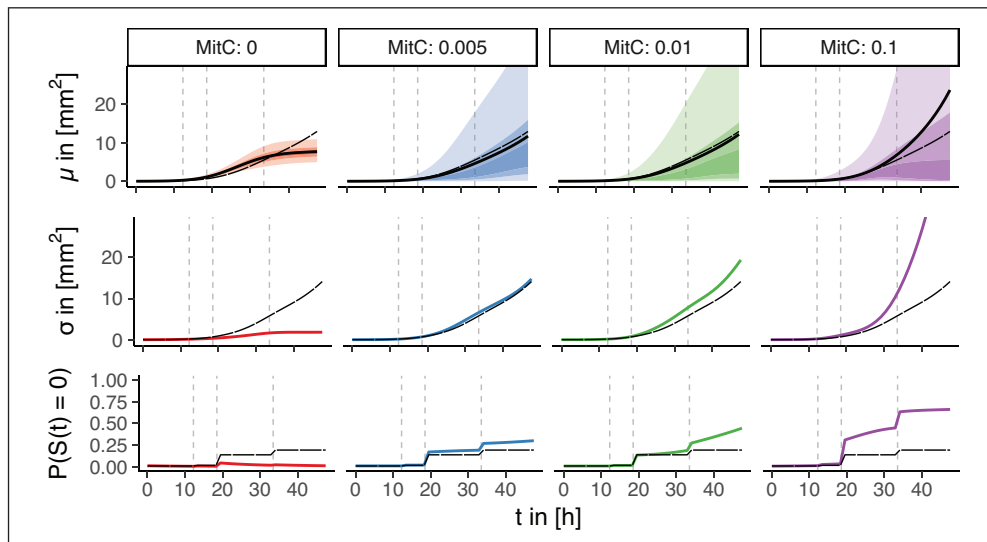
### 3.2.1 MitC effect and effect of experimental batches

An overview of the effects of the toxin MitC can be found in Figure 1. We observe that mean S-strain growth is slightly increasing for low MitC levels compared to no MitC, but is particularly higher for  $MitC_i = 0.1 \mu\text{g/ml}$ . This indicates that, if  $S_i(t) \geq 0$ , the S-strain even grows better under this condition.

For the standard deviation, we observe a gradual but distinct rise with the MitC level. Due to the log-link we may not only interpret effects on the shape parameter  $\sigma/\mu$  but also on  $\sigma$ : effect functions  $h_j^{(\sigma)}$  for  $\sigma$  are obtained as  $h_j^{(\sigma)} = h_j^{(\mu)} + h_j^{(\sigma/\mu)}$ . In this plot, we choose to depict  $\sigma$  instead of  $\sigma/\mu$ , as it is more straightforward to interpret on the response level. We observe that positive skewness increases with MitC concentration.

It is important to note, that control experiments indicate no considerable effect of MitC on S-strain growth (von Bronk et al., 2017). Thus, present covariate effects of MitC reflect effects of C-cells which cannot be explained by the observed C-strain growth curves. Showing distinct shifts at the zoom points,  $p_i(t) = P(S_i(t) = 0)$  seems to depend highly on the zoom level of the microscope. This suggests, that besides full extinction of the S-strain,  $S_i(t) = 0$  is also linked to limitations in area recognition. Additionally, the probability for  $S_i(t) = 0$  is higher for positive MitC concentrations. Overall, the conditional mean for positive  $S_i(t)$  but also the variability and probability for zero increase with the MitC concentration.

The smooth functional effects for each of the eight experimental batches are relatively small in size. For the conditional mean  $\mu$ , they cause an average deviation of about 3% of the intercept growth curve (geometric mean over observed time points and batches); for the scale parameter  $\sigma/\mu$ , the average deviation is about 9%; and for  $p$  about 6%. While point-wise 95% bootstrap confidence interval, type uncertainty bounds (Online Appendix Section E.4) show less accuracy for the batch effects (in particular, those on  $p(t)$ ), they indicate a high estimation precision for the MitC effects and functional intercepts. This corresponds to our findings in the simulation study in Section 4.



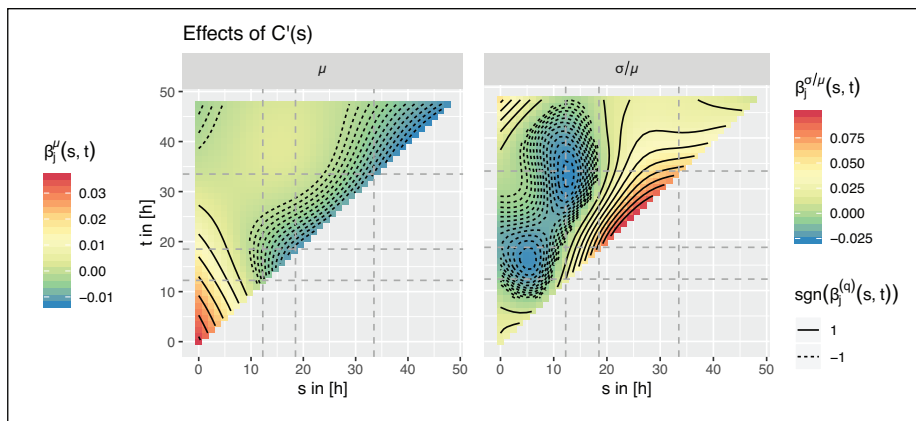
**Figure 1** Estimated point-wise mean (*top*) and standard deviation (*centre*) of the S-strain growth curves  $S_i(t)$  conditional on  $S_i(t) > 0$  and the extinction probabilities (*bottom*) of S-strain growth curves for each MitC concentration. Long-dashed curves correspond to the functional intercept, dashed vertical lines to the zoom level change-points. Thick solid lines indicate the estimates, transparent ribbons reflect the point-wise inner 25%, 50% and 90% probability mass intervals of the estimated gamma distributions conditional on  $S_i(t) > 0$  (*top*)

### 3.2.2 C-strain effect

The base-learner for the C-strain area propagation  $C_i(s)$  effect on  $\mu_i(t)$  is never selected throughout the boosting procedure and the effect on  $\sigma_i(t)/\mu_i(t)$  is small (Online Appendix Figure 16). Thus, we only discuss the effect of the area increment  $C'_i(s)$  here (Figure 2). Looking at the  $C'$ - $\mu$ -effect (effect of  $C'(s)$  on mean S area), we can distinguish two main impact phases.

In the earlier growth phase with  $s \leq 10 h$ , we observe a positive  $C'$ - $\mu$ -effect concerning almost the whole time curve of the S-strain. That means that C-strain growth above [below] the average indicates increased [decreased] S-strain propagation. Both colicin production and colicin secretion are costly to the population and slow down C-propagation. A low value of  $C'_i(s)$  indicates early colicin secretion. We conclude that this first phase delineates a time window, where colicin emission is able to severely harm the S-strain population.

In the second phase for  $s > 10 h$ , we observe a negative  $C'$ - $\mu$ -effect, which is maximal at short time lags and slowly fading. This likely reflects spatial competition of the S- and the C-strain (compare Online Appendix Figure 14). At this time, bacteria have grown together to coherent formations and strains obstruct expansion of each other. Even though the  $C'$ - $\mu$ -effect offers this clear interpretation, it is rather small compared to, for example, the MitC effect on  $\mu(t)$ . Moreover, while in simulation studies we observe a rather high estimation precision for most of the historical effects (Online



**Figure 2** *Left:* Coefficient function  $\beta^{(\mu)}(s, t)$  for the historical effects of  $C'_i(s)$  on the mean of S-strain growth curves. *Right:* the corresponding plot for the effect of  $C'_i(s)$  on the scale parameter  $\sigma_i(t)/\mu_i(t)$ . The y-axis represents the time line for the response curve, the x-axis represents the one for the C-strain growth curve. The change-points in zoom level are marked with dashed lines. For a fixed  $s = s_0$ ,  $\beta^{(\mu)}(s_0, t)$  and  $\beta^{(\sigma/\mu)}(s_0, t)$  describe the effect of the normalized covariate at time  $s_0$  on the S-strain growth curve over the whole remaining time interval.

Appendix Figure 12), 95% bootstrap confidence interval-type uncertainty bounds indicate distinctly less precision than for the MitC-effects (Online Appendix E.4). However, the historical  $C'$ - $\sigma/\mu$ -effect also corroborates the distinction into two phases of interaction: While in the first phase there is a negative effect of C-strain growth, the effect turns positive in the second phase. Thus, relative variability is increased for slow C-strain growth early in the experiment (colicin production) and for fast C-strain growth later in the experiment (areal competition).

For the probability  $p_i(t) = P(S_i(t) = 0)$  both the effects of  $C_i(s)$  and  $C'_i(s)$  were selected. Corresponding plots can be found in the Online Appendix Figure 16. However, as already indicated by the marked cuts between the different zoom levels (Figure 1), vanishing of the S-strain is particularly sensitive to the precision of the area recognition. Hence, we are careful with interpreting the effects further in terms of the bacterial dynamics.

## 4 Simulation studies

### 4.1 Simulation set-up

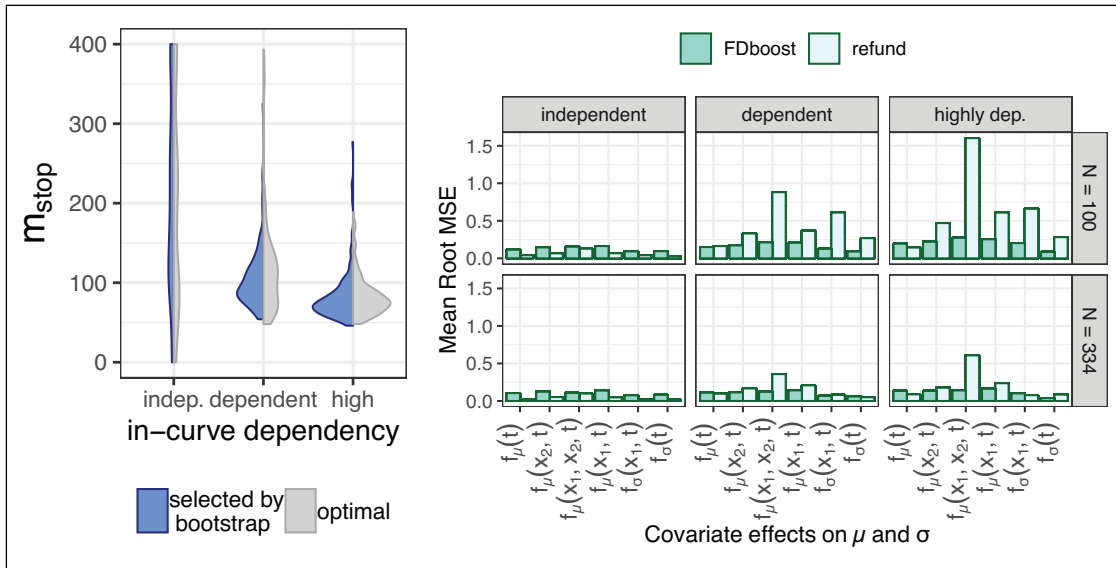
Model-based gradient boosting approaches to non-functional or one-parameter special cases of the present model are well tested with respect to their fitting performance and variable selection quality (e.g., see Brockhaus et al., 2018a; Brockhaus et al., 2015; Thomas et al., 2018; Mayr et al., 2012) showing a typically slow over-fitting behaviour. However, modelling functional response variables with GAMLSS presents an important additional challenge: high auto-correlation in

response functions may lead to severe over-fitting when estimating typically complex base-learners. While this is already the case for non-GAMLSS functional response models, it gets particularly acute for GAMLSS models with multiple predictors—if it is not properly controlled for by early stopping based on curve-wise re-sampling methods. We focus on this issue in an extensive simulation study investigating the fitting performance for different levels of in-curve dependency while also comparing different sample sizes, choices of hyper parameters, the non-cyclic and cyclic fitting method, and different (curve-wise) re-sampling methods. Moreover, we consider three different models in the simulation study: one model is directly based on the bacterial interaction scenario in Chapter 3 taking the model estimated on the original data as true underlying model; and two models with a Gaussian response distribution and categorical effects or more complex smooth (interaction) effects of metric covariates, respectively. There, we randomly generate different sets of true underlying effects in order to obtain as general results as possible. In the Gaussian case, where this is possible, we also compare to the penalized likelihood approach of (Greven and Scheipl, 2017) which is implemented in the R package `refund` (Goldsmith et al., 2018). For details concerning the simulation set-up, the data generation and a more thorough discussion of the results, please refer to the corresponding sections in the Online Supplement.

## 4.2 Simulation results

Considering the mean  $\overline{\text{KLD}}$  of the estimated to the true underlying model, we observe that for conditionally independent measurements within response curves, the optimal stopping iteration  $m_{\text{stop}}$  is typically far higher than for dependent or highly dependent measurements (Figure 3 (*left*)), that is, in the independent case a model can be fit distinctly longer without resulting in over-fitting. At the same time, we find that  $m_{\text{stop}}$  selected by curve-wise bootstrapping (performing slightly better than other curve-wise re-sampling methods) reflects these differences very well, which shows that it is desirably sensitive to in-curve dependency and prevents over-fitting. The resulting regularization improves the estimation accuracy strongly in particular for complex base-learners. The effect becomes especially visible when comparing it to the penalized likelihood (`refund`) approach (Figure 3 (*right*)), which currently lacks a corresponding regularization mechanism for GAMLSS models: When only modelling the response mean, there are typically curve-specific functional random intercepts included in order to account for in-curve dependency; however, they would interfere with modelling the marginal standard deviation in a separate predictor and are, thus, not included into the GAMLSS-type model. Measuring the fitting error in Root Mean Squared Error (RMSE), the `refund` approach shows a better performance in the independent case. However, it exceeds the RMSE of our `FDboost` approach by far in realistic scenarios with high in-curve dependency.

In the application motivated simulation scenario, we observe that most of the RMSEs for the estimated covariate effects are lower than 10% of the effect range even in the highly dependent setting (Online Appendix Figure 12). Exceptions are the functional intercept in the predictor for the extinction probability  $p(t)$  being



**Figure 3** Plots referring to a Gaussian model scenario including smooth covariate effects  $f_{\vartheta}(x_j, t)$  for  $\vartheta \in \{\mu, \sigma\}$ , the mean and standard deviation over time  $t \in [0, 1]$ , and for two metric covariates  $j \in \{1, 2\}$ , and a smooth interaction  $f_{\mu}(x_1, x_2, t)$  effect for  $\mu$  (200 model fits per combination of sample size  $N$  and in-curve dependency level). *Left:* Violin-plots reflecting the empirical density of the stopping iterations  $m_{stop}$  selected via 10-fold bootstrap (*left*) and for the KLD-optimal  $m_{stop}$  (*right*) for  $N = 334$  sampled curves. *Right:* Bar-plots indicating the mean RMSE of the different effects for our approach based on gradient boosting (FDboost, *dark*) and the approach based on penalized likelihood (refund, *light*). The highly dependent setting is the most realistic in many functional data scenarios and is—as far as the analogy can be drawn—the closest to the correlation structure in our application.

composed of a smooth functional intercept and a step function, and the  $\sigma/\mu$ -effect of  $C(t)$ , which has a comparably large relative RMSE, due to its small effect size, while having a quite small absolute RMSE. Although we do not focus on variable selection in this article, the  $C$ - $\mu$ -effect and the smooth (non-step) functional intercept for  $p(t)$ , which were not selected in the original model fit in Chapter 3, serve as nuisance effects in the application motivated simulation. While the sensitivity is quite high for most of the non-zero effects (mostly 100%, minimum 70%), the nuisance  $C$ - $\mu$ -effect is still selected in rather many simulation runs (44% independent, 49% dependent, 33% highly dependent scenario), see Online Supplement Figure 13. To improve on this, stability selection as applied, for example, by Brockhaus et al. (2017) and Thomas et al. (2018) might be used. However, the mean RMSE of the effect is still extremely low indicating that even if the effect is selected it is very small in size. Overall, we observe the effects to be estimated quite well despite in-curve dependency and the high complexity of the model in both the Gaussian and the application motivated simulation studies.



## 5 Discussion and outlook

The functional GAMLSS regression framework we present in this article allows for very flexible modelling of functional responses. We may simultaneously model multiple parameters of functional response distributions in dependence of time and covariates, specifying a separate additive predictor for each parameter function. In addition, point-wise distributions for the response curves beyond exponential family distributions can be specified. Doing so a vast variety of new data scenarios can be modelled. These new possibilities have shown to be crucial, when applying the framework to analyse growth curves in the present bacterial interaction scenario.

The results we obtain confirm and extend previous work: Focusing on the outcome after 48 hour and on the number of C-clusters at the edge of the S-colony after 12 hour, von Bronk et al. (2017) already identify a phase of 'stochastic toxin dynamics' followed by a phase of 'deterministic dynamics' similar to the two phases of bacterial interaction we find in the historical functional effects of the C-strain growth. The functional regression model not only provides new evidence for this distinction from a completely new perspective, but now also allows to quantitatively discuss the effect of the C-strain on the S-strain over the whole time range: We now observe C-growth to have a positive effect on S-growth in the early phase and a negative effect in the later phase. The separation of these two phases appears even more distinct in the effect on the relative standard deviation, which we would not be able to recognize without GAMLSS.

Regarding the fraction of S- and C-strain area after 48 hour von Bronk et al. (2017) categorized three different states of the bacterial interaction: for no MitC, there is either dominance of the S-strain or coexistence; for a moderate MitC concentration, there occurs a splitting into two extremes—either dominance of the S-strain or extinction; and for the highest MitC concentration, the toxin strategy of the C-strain fails and the S-strain either dominates or both strains go extinct. Now, referring to the complete growth curves, our results also reflect this categorization: If MitC is added, and conditional on a positive area, the mean S-strain growth increases, whereas also the probability for zero area and the variance increase. However, these differences would not be captured by non-GAMLSS regression models for the mean only, as the mean growth curves not conditioning on the response being positive are very similar for no and moderate MitC concentration (see Online Appendix Figure 17). Apart from that, the framework provides the flexibility to account for special challenges in the experimental set-up, such as dependencies between observations in the same experimental batch and differences between zoom levels of the microscope.

In simulation studies, we confirm that by fitting our models via component-wise gradient boosting we are capable of estimating even complex covariate effects on multiple distributional parameter functions. As it prevents over-fitting, early stopping of the boosting algorithm based on curve-wise re-sampling plays a key role: It enables us to face settings with highly auto-correlated response curves without explicitly modelling the correlation structure.

## Supplementary material

Supplementary material including an Online Appendix with further details and illustrations, as well as R code and data used for the analysis of bacterial interaction and simulations is available from <http://www.statmod.org/smij/archive.html>

## Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

## Funding

Financial support from the Deutsche Forschungsgemeinschaft (DFG) through Emmy Noether grant GR 3793/1-1 (AS, SB, SG) and through grant OP252/4-2 part of the DFG Priority Program SPP1617 is gratefully acknowledged. B.v.B was supported by a DFG Fellowship through the Graduate School of Quantitative Biosciences Munich (QBM). Additional financial support by the Center for Nanoscience (CeNS) and the Nano Systems Initiative - Munich (NIM) is gratefully acknowledged.

## References

- Brockhaus S, Fuest A, Mayr A and Greven S (2018a) Signal regression models for location, scale and shape with an application to stock returns. *Journal of the Royal Statistical Society: Series C*, **67**, 665–86.
- Brockhaus S, Melcher M, Leisch F and Greven S (2017) Boosting flexible functional regression models with a high number of functional historical effects. *Statistics and Computing*, **27**, 913–26.
- Brockhaus S and Ruegamer D (2018) *FDboost: Boosting functional regression models*. R package version 0.3–1.
- Brockhaus S, Rügamer D and Greven S (2018b) Boosting functional regression models with FDboost. *Journal of Statistical Software*. URL <https://arxiv.org/pdf/1705.10662.pdf> (last accessed 6 April 2020).
- Brockhaus S, Scheipl F and Greven S (2015) The functional linear array model. *Statistical Modelling*, **15**, 279–300.
- Bühlmann P and Hothorn T (2007) Boosting algorithms: Regularization, prediction and model fitting (with discussion). *Statistical Science*, **22**, 477–505.
- Eilers PHC and Marx BD (2010) *Splines, knots, and penalties*. *Wiley Interdisciplinary Reviews: Computational Statistics*, **2**, 637–53.
- Gasser T, Müller H-G, Kohler W, Molinari L and Prader A (1984) Nonparametric regression analysis of growth curves. *The Annals of Statistics*, **12**, 210–29.
- Gertheiss J, Maier V, Hessel EF and Staicu A-M (2015) Marginal functional regression models for analysing the feeding behaviour of pigs. *Journal of Agricultural, Biological, and Environmental Statistics*, **20**, 353–70.
- Goldsmith J, Scheipl F, Huang L, Wrobel J, Gellar J, Harezlak J, McLean MW, Swihart B, Xiao L, Crainiceanu C and Reiss PT (2018) *refund: Regression with Functional Data*. R

- package version 0.1–17.
- Goldsmith J, Zipunnikov V and Schrack J (2015) Generalized multilevel function-on-scalar regression and principal component analysis. *Biometrics*, **71**, 344–53.
- Greven S and Scheipl F (2017) A general framework for functional regression modelling (with discussion). *Statistical Modelling*, **17**, 1–35.
- Hall P, Mueller H-G and Yao F (2008) Modelling sparse generalized longitudinal observations with latent Gaussian processes. *Journal of the Royal Statistical Society: Series B*, **70**, 703–23.
- Hastie T and Tibshirani R (1990) *Generalized Additive Models*. London: Chapman & Hall.
- Hofner B, Hothorn T, Kneib T and Schmid M (2011) A framework for unbiased model selection based on boosting. *Journal of Computational and Graphical Statistics*, **20**, 956–71.
- Hofner B, Mayr A, Fenske N and Schmid M (2017) *gamboostLSS: Boosting methods for GAMLSS models*. R package version 2.0–0.
- Kneib T, Hothorn T and Tutz G (2009) Variable selection and model choice in geospatial regression models. *Biometrics*, **65**, 626–34.
- Li H, Staudenmayer J and Carroll RJ (2014) Hierarchical functional data with mixed continuous and binary measurements. *Biometrics*, **70**, 802–11.
- López S, Prieto M, Dijkstra J, Dhanoa M and France J (2004) Statistical evaluation of mathematical models for microbial growth. *International Journal of Food Microbiology*, **96**, 289–300.
- Mayr A, Fenske N, Hofner B, Kneib T and Schmid M (2012) 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**, 403–27.
- Morris JS (2015) Functional regression. *Annual Review of Statistics and Its Application*, **2**, 321–59.
- Perni S, Andrew PW and Shama G (2005) Estimating the maximum growth rate from microbial growth curves: Definition is everything. *Food Microbiology*, **22**, 491–95.
- R Core Team (2018) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. R version 3.5.1.
- Ramsay J and Silverman BW (2005) *Functional Data Analysis*. New York, NY: Springer Science & Business Media.
- Rigby RA and Stasinopoulos DM (2005) Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **54**, 507–54.
- Scheipl F, Gertheiss J and Greven S (2016) Generalized functional additive mixed models. *Electronic Journal of Statistics*, **10**, 1455–92.
- Scheipl F, Staicu A-M and Greven S (2015) Functional additive mixed models. *Journal of Computational and Graphical Statistics*, **24**, 477–501.
- Staicu A-M, Crainiceanu CM, Reich DS and Ruppert D (2012) Modeling functional data with spatially heterogeneous shape characteristics. *Biometrics*, **68**, 331–43.
- Stasinopoulos M and Rigby R (2019) *gamlss.dist: Distributions for generalized additive models for location scale and shape*. R package version 5.1–4.
- Thomas J, Mayr A, Bischl B, Schmid M, Smith A and Hofner B (2018) Gradient boosting for distributional regression: Faster tuning and improved variable selection via noncyclical updates. *Statistics and Computing*, **28**, 673–87.
- van der Linde A (2009) A Bayesian latent variable approach to functional principal components analysis with binary and count data. *AStA Advances in Statistical Analysis*, **93**, 307–33.
- von Bronk B, Schaffer SA, Götz A and Opitz M (2017) Effects of stochasticity and division of labor in toxin production on two-strain bacterial competition in *Escherichia coli*. *PLoS Biology*, **15**, e2001457.

- Wang B and Shi JQ (2014) Generalized Gaussian process regression model for non-Gaussian functional data. *Journal of the American Statistical Association*, **109**, 1123–33.
- Weber MF, Poxleitner G, Hebisch E, Frey E and Opitz M (2014) Chemical warfare and survival strategies in bacterial range expansions (online supplement). *Journal of The Royal Society Interface*, **11**, 20140172.
- Wood SN (2006) Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics*, **62**, 1025–36.