Journal of the Royal Statistical Society Series C: Applied Statistics, 2023, **72**, 414–433
https://doi.org/10.1093/jrsssc/qlad015
Advance access publication 22 March 2023
Original Article

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS
Journal of the Statistics Society
Series C
Applied Statistics

Modelling intra-annual tree stem growth with a distributional regression approach for Gaussian process responses

Hannes Riebl[1], Nadja Klein[2] and Thomas Kneib[1]

[1]Chair of Statistics, Georg-August-Universität Göttingen, Göttingen, Germany
[2]Chair of Statistics and Data Science, Humboldt-Universität zu Berlin, Berlin, Germany

Address for correspondence: Hannes Riebl, Professur für Statistik, Georg-August-Universität Göttingen, Humboldtallee 3, 37073 Göttingen, Germany. Email: hriebl@uni-goettingen.de

Abstract

High-resolution circumference dendrometers measure the irreversible growth and the reversible shrinking and swelling due to the water content of a tree stem. We propose a novel statistical method to decompose these measurements into a permanent and a temporary component, while explaining differences between the trees and years by covariates. Our model embeds Gaussian processes with parametric mean and covariance functions as response structures in a distributional regression framework with structured additive predictors. We discuss different mean and covariance functions, connections with other model classes, Markov chain Monte Carlo inference, and the efficiency of our sampling scheme.

Keywords: generalised additive model for location, scale, and shape, growth curve model, Markov chain Monte Carlo simulation, Matérn covariance function, spatio-temporal regression, structured additive predictor

1 Introduction

Tree growth, and the growth of tree stems in particular, is a process that is of strong ecological and economic interest. Together with the height growth, the growth in the stem girth drives timber production, and at the same time, plays a key role in the global carbon cycle (Mencuccini et al., 2017). Unfortunately, it is difficult to measure the formation of new wood and bark cells in the cambium resulting in permanent stem growth, and while electronic dendrometers can record the variation of the stem circumference on small time scales of a few minutes (Klepper et al., 1971), these measurements also capture the reversible shrinking and swelling of the stem due to changes in its water content. Researchers have used additional measurement equipment such as sap flow sensors and controlled irrigation experiments to gain a better understanding of the permanent and temporary components of tree stem growth (Mencuccini et al., 2017), but these experiments are either expensive or not feasible under open field conditions.

We describe a novel statistical method for the analysis of high-resolution dendrometer measurements that does not require additional data about other tree-physiological processes. The method permits us to decompose the dendrometer measurements into a permanent and a temporary component through stochastic assumptions and explanatory variables. Our dataset contains 85 deciduous trees from Germany and the growing seasons 2012 and 2013. Figure 1 shows a subsample of the recorded growth curves between April 1 and September 30, each of which is assumed to be a realisation of a Gaussian process (GP). The GPs are conditionally independent from each other given a set of explanatory variables. We observe that the coloured ash grows primarily between mid-April and mid-July, while the coloured beech grows later and more during the

Received: March 4, 2022. **Revised:** December 5, 2022. **Accepted:** January 20, 2023

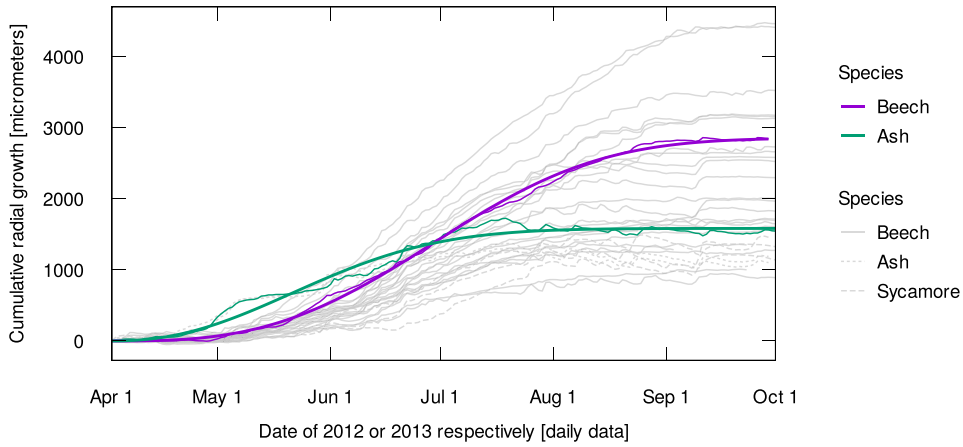

Figure 1. The cumulative radial growth of a subsample of the trees from our dataset over the course of one growing season. The coloured lines represent two exemplary trees, one beech and one ash, while the grey lines illustrate the diversity of the growth patterns in the dataset. The reversible shrinking and swelling is more pronounced for the coloured ash than the beech. The bold lines show the estimated irreversible growth.

vegetation period. These different patterns are captured in the estimated sigmoid mean functions of the GPs, which represent the irreversible growth of the tree stems. On the other hand, the temporary shrinking and swelling, which is more pronounced in the ash than the beech, is described by the within-season covariance functions. The tree species is one factor that we condition the GPs on. Other possible explanatory variables include the diameter at breast height (DBH) and the geographical location of the trees.

The way we use GPs is different from the standard approaches in machine learning or spatial statistics. Models from those fields typically assume a single latent GP. For instance, in many supervised learning problems in machine learning, GPs are used as priors over the hypothesis space of possible functions from the input to the output space (Rasmussen & Williams, 2006). GPs also play a key role in spatial statistics, where they are used to capture the spatial correlation of the data, and to avoid invalid, underestimated confidence intervals (Cressie, 1993). In contrast, we assume multiple *observed*, *conditionally independent* GPs as response structures in a regression model.

The fact that we link multiple properties of the mean and covariance functions of the GPs to explanatory variables puts our model in the domain of the so-called distributional regression models, also known as generalised additive models for location, scale, and shape (GAMLSS, Rigby & Stasinopoulos, 2005). Usually, this model class admits multiple structured additive predictors for different parameters of the conditional distribution of a response variable. Standard distributional regression models use univariate or low-dimensional multivariate response variables. Klein and Kneib (2016b) discuss distributional regression models with copula-based bivariate response distributions in a Bayesian setting, and Filippou et al. (2017) propose a trivariate probit model, which they estimate with a frequentist penalised likelihood method. A number of bivariate and trivariate response distributions is also available in the vector generalised additive model framework (Yee, 2015). Following this line of thought, we show that the distributional regression approach also works for more general, continuous response structures such as GPs.

The distributional regression literature offers different approaches to statistical inference. We build on the work of Klein et al. (2015), who propose a general Markov chain Monte Carlo (MCMC) algorithm for Bayesian inference in distributional regression models. To assess the posterior distribution of the model parameters, they use a Metropolis-within-Gibbs sampler with iterative weighted least squares (IWLS) proposals (Gamerman, 1997). The algorithm was implemented by Umlauf et al. (2018) in the R package `bamlss`. However, `bamlss` blocks the model parameters in a way that performs very poorly with GPs as response structures. We describe a more efficient way of blocking the model parameters in Section 3, and discuss the problem in more detail in the simulation study in Section 4.

All computations for this paper were performed using the R software environment for statistical computing (R Core Team, 2020). The relevant code is available under the MIT open source license as a supplement and on GitHub (https://github.com/hriebl/gp-responses).

The remainder of this paper is structured as follows: in the next section, we give the precise definition of our model, as well as some examples of mean and covariance functions. Structured additive predictors and the connection with a number of related statistical model classes (mixed models, functional data, etc.) are also discussed. In Section 3, we provide the specifics of the MCMC algorithm for the posterior estimation. The simulation study with three different scenarios is presented in Section 4, while Section 5 addresses the application to intra-annual tree stem growth in full detail. Finally, Section 6 concludes and discusses possible extensions and further applications of the model.

2 Model specification

2.1 Gaussian processes as response structures

We consider GPs $\{Y_i(\boldsymbol{t}); \boldsymbol{t} \in T\}$ as response structures in structured additive distributional regression, where the observation index i runs from 1 to N and the index set T is a metric space that can represent time, space, or space-time. The GPs are assumed to be conditionally independent given the covariate vectors $\boldsymbol{x}_i$,

$$\{Y_i(\boldsymbol{t}); \boldsymbol{t} \in T\} \mid \boldsymbol{x}_i \overset{\text{ind.}}{\sim} \mathcal{GP}(m_x(\boldsymbol{t}; \boldsymbol{x}_i), c_x(\boldsymbol{t}, \boldsymbol{t}'; \boldsymbol{x}_i)), \tag{1}$$

where $\boldsymbol{t}, \boldsymbol{t}' \in T$. As a specific feature of distributional regression, the mean function m_x and the covariance function c_x both depend on the covariates $\boldsymbol{x}_i$, which differ between the observations 1 to N but are constant within the index set T. An extension of the model to time or space-varying covariates is given below.

More precisely, the mean and the covariance function are linked to the covariates $\boldsymbol{x}_i$ via their respective parameter vectors $\boldsymbol{\theta}^m$ and $\boldsymbol{\theta}^c$,

$$m_x(\boldsymbol{t}; \boldsymbol{x}_i) = m(\boldsymbol{t}; \boldsymbol{\theta}^m(\boldsymbol{x}_i)) \quad \text{and} \quad c_x(\boldsymbol{t}, \boldsymbol{t}'; \boldsymbol{x}_i) = c(\boldsymbol{t}, \boldsymbol{t}'; \boldsymbol{\theta}^c(\boldsymbol{x}_i)).$$

For better readability, we use a subscript i as an abbreviation for the dependence of a variable on the covariates $\boldsymbol{x}_i$. Let $\boldsymbol{\theta}_i = [(\boldsymbol{\theta}_i^m)^\top, (\boldsymbol{\theta}_i^c)^\top]^\top = [\boldsymbol{\theta}^m(\boldsymbol{x}_i)^\top, \boldsymbol{\theta}^c(\boldsymbol{x}_i)^\top]^\top$ be the vector of all parameters of the GPs and K its dimension. In the terminology of distributional regression, $\boldsymbol{\theta}_i$ is the vector of the K distributional parameters. Each parameter θ_{ki} is linked to a structured additive predictor via a strictly monotonic link function (Section 2.3).

One important extension of Model (1) deals with the inclusion of time or space-varying covariates $\boldsymbol{z}_i : T \to S$, which change within the index set T depending on the coordinates $\boldsymbol{t}$ (as opposed to the previously discussed covariates $\boldsymbol{x}_i$). The covariates $\boldsymbol{z}_i$ may be understood as a mapping of the coordinates $\boldsymbol{t}$ into a different, more abstract metric space S, which no longer represents only space or time. The mean and the covariance function now depend on the coordinates $\boldsymbol{t}$ via this mapping,

$$\{Y_i(\boldsymbol{t})\} \mid \boldsymbol{z}_i, \boldsymbol{x}_i \overset{\text{ind.}}{\sim} \mathcal{GP}(m_x(\boldsymbol{z}_i(\boldsymbol{t}); \boldsymbol{x}_i), c_x(\boldsymbol{z}_i(\boldsymbol{t}), \boldsymbol{z}_i(\boldsymbol{t}'); \boldsymbol{x}_i)). \tag{2}$$

The simulation study in Section 4 includes a proof of concept for time-varying covariates.

In practice, each GP $\{Y_i(\boldsymbol{t})\}$ can only be observed at a finite number of points $\boldsymbol{t}_j \in T$, for $j = 1, \ldots, n_i$. The collection of random variables at these points has a multivariate normal distribution,

$$[Y_i(\boldsymbol{t}_1), \ldots, Y_i(\boldsymbol{t}_{n_i})]^\top \mid \boldsymbol{z}_i(\boldsymbol{t}_1), \ldots, \boldsymbol{z}_i(\boldsymbol{t}_{n_i}), \boldsymbol{x}_i \overset{\text{ind.}}{\sim} \mathcal{N}_{n_i}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), \tag{3}$$

where the elements of the mean vector $\boldsymbol{\mu}_i$ and the covariance matrix $\boldsymbol{\Sigma}_i$ are the evaluations of the mean function m and the covariance function c at the observed points,

$$\boldsymbol{\mu}_i = [\mu_{i,j}] = m(\boldsymbol{z}_i(\boldsymbol{t}_j); \boldsymbol{\theta}_i^m) \quad \text{and} \quad \boldsymbol{\Sigma}_i = [\sigma_{i,j,j'}] = c(\boldsymbol{z}_i(\boldsymbol{t}_j), \boldsymbol{z}_i(\boldsymbol{t}_{j'}); \boldsymbol{\theta}_i^c) \tag{4}$$

for $j, j' = 1, \ldots, n_i$. The number of observed values does not necessarily need to be the same for all GPs,

i.e., potentially $n_i \neq n_{i'}$ for some $i, i' \in \{1, \ldots, N\}$. Furthermore, our construction leaves the basic structure of the covariance function untouched, such that even after including dependence on covariates, it is ensured that the covariance function and therefore also the resulting covariance matrices $\boldsymbol{\Sigma}_i$ are positive definite.

2.2 Examples of mean and covariance functions

The most important condition for a GP to be valid is that the covariance function needs to be positive semi-definite. For Model (1), this means that $\sum_{t,t' \in U} a_t c(\boldsymbol{t}, \boldsymbol{t}') a_{t'}$ needs to be non-negative for all $U \subset T$ and the weights $a_t \in \mathbb{R}$ of each linear combination $\sum_{t \in U} a_t Y_i(\boldsymbol{t})$ (Adler, 1990, Section 1.1). When considering Model (2) with time or space-varying covariates, the positive semi-definiteness of the covariance function needs to hold on the index set S instead of T. The requirements for the mean function of a GP are less restrictive: Essentially any function $m : T \to \mathbb{R}$ or $m : S \to \mathbb{R}$ is a valid mean function of a GP.

The mean and the covariance function of a GP determine its continuity and differentiability properties. For example, a GP is mean-square continuous if and only if its mean and its covariance function are continuous. Mean-square continuity, however, does not imply sample continuity (Rasmussen & Williams, 2006, Section 4.1.1). The concept of sample continuity is discussed in a rigorous and abstract fashion in Adler (1990). For most applied modelling problems, a continuous mean function will be a reasonable assumption, but the same is not necessarily true for the covariance function: One reason for a discontinuous covariance function might be an idiosyncratic error term for each measurement. This idiosyncratic error term 'conceals' the GP of interest and is known as the 'nugget effect' in spatial statistics. It is usually modelled as an additive i.i.d. GP, rendering the resulting sum of two GPs discontinuous, even in the mean-square sense.

Note that we omit the observation index i in the following discussion of the mean and covariance functions for the sake of simplicity.

2.2.1 Mean functions

Linear Mean Function. The linear mean function is defined as the dot product of the covariates $z_i(\boldsymbol{t})$ and the parameters $\boldsymbol{\theta}^m$,

$$m^l(\boldsymbol{z}(\boldsymbol{t}); \boldsymbol{\theta}^m) = \boldsymbol{z}(\boldsymbol{t})^\top \boldsymbol{\theta}^m. \tag{5}$$

The linear mean function is mathematically convenient and provides considerable flexibility for statistical modelling. Polynomials or B-splines can be used, among others, to represent large classes of functions as linear combinations of basis functions. To do so, we choose $\boldsymbol{z}(\boldsymbol{t}) = [b_1(\boldsymbol{t}), \ldots, b_M(\boldsymbol{t})]^\top$, where $b_1, \ldots, b_M$ are the aforementioned basis functions and M is the number of basis functions or, equivalently, the number of distributional parameters of the mean function. This approach gives rise to *non-parametric* mean functions with very flexible shapes.

Weibull Growth Curve. The flexibility of a linear mean function with polynomials or B-splines comes at a cost: it requires a large number of parameters without an obvious interpretation. In many applications, however, interpretable distributional parameters are desirable. If prior knowledge about the shape of the response processes is available, a *parametric* mean function might be a more natural choice. For example, the intra-annual tree growth curves in Section 5 have a sigmoid shape. Any sigmoid function such as the logistic function or the hyperbolic tangent could serve as a mean function in this case, but we follow Metz et al. (2020) and use the Weibull growth curve for this purpose. The Weibull growth curve is a scaled version of the cumulative distribution function of the Weibull distribution,

$$m^w(\boldsymbol{z}(\boldsymbol{t}) = t; \boldsymbol{\theta}^m = [l, a, b]^\top) = l \times \left[1 - \exp\left(-\left(\frac{t}{b}\right)^a\right)\right], \tag{6}$$

where $t \geq 0$ is the point in time since the start of the growing season, and the parameters are the limit $l > 0$, the shape $a > 0$, and the scale $b > 0$. The scale parameter describes for how long a tree continues to grow during the summer, while the shape parameter represents the steepness

of the growth curve. As all parameters of the Weibull growth curve need to be positive, we use a log-link for these parameters.

2.2.2 *Covariance functions*

Throughout this paper, we use the Matérn covariance function and relate the standard deviation σ and the range ϕ to covariates. Without any time or space-varying covariates, it has the form

$$c^m(z(\boldsymbol{t}) = \boldsymbol{t}, z(\boldsymbol{t}') = \boldsymbol{t}'; \boldsymbol{\theta}^c = [\sigma, \phi]^\top) = \sigma^2 \times \rho\left(\frac{d(\boldsymbol{t}, \boldsymbol{t}')}{\phi}; \nu\right), \tag{7}$$

where ρ is the Matérn correlation function with the smoothness parameter ν, and $d(\boldsymbol{t}, \boldsymbol{t}')$ is a distance function. For $\nu = 1/2$, the Matérn correlation function simplifies to the exponential correlation function, and for $\nu \to \infty$, it converges to the squared exponential or Gaussian correlation function. GPs with the Matérn covariance function are $\lfloor \nu \rfloor$ times mean-square differentiable and even have differentiable sample paths (Rasmussen & Williams, 2006, Section 4.2.1; Paciorek, 2003, Section 2.5.4). Different values for ν can be used for different models, but we do not treat it as a distributional parameter in this paper. For the standard deviation and the range, we use a log-link, as these parameters need to be positive.

The Matérn covariance function (or any other covariance function) can be modified to include an additive i.i.d. measurement error, giving rise to the covariance function

$$c^*(\boldsymbol{t}, \boldsymbol{t}'; \boldsymbol{\theta}^c = [\sigma, \phi, \delta]^\top) = c^m(\boldsymbol{t}, \boldsymbol{t}'; \boldsymbol{\theta}^c = [\sigma, \phi]^\top) + \delta^2 \times \mathbb{I}(d(\boldsymbol{t}, \boldsymbol{t}') = 0),$$

where δ is the standard deviation of the idiosyncratic error, and $\mathbb{I}$ is the indicator function. In the distributional regression framework, δ can be interpreted as an additional distributional parameter of the covariance function. The estimation procedure discussed in Section 3 can be applied to δ in the same way as to any other distributional parameter.

It is important to note that the validity of a covariance function depends on the metric space it is defined on, i.e., on the distance function $d(\boldsymbol{t}, \boldsymbol{t}')$. While the Matérn covariance function is valid on the Euclidean space of any dimension, the situation on the sphere with the great circle distance is more complicated: Gneiting (2013) investigates the validity of different commonly used covariance functions on the one- to three-dimensional sphere and finds that, in this case, the Matérn covariance function is only valid for $0 < \nu \leq 1/2$.

There are no restrictions on the families of covariance functions that can be used in our model framework, and the number of covariance parameters can be greater than for the Matérn covariance function. Examples of alternative covariance functions include the power exponential, the rational quadratic, and the spherical covariance functions (Rasmussen & Williams, 2006, Section 4.2). Some index sets with special interpretations might require more elaborate covariance functions such as non-stationary or non-separable space-time covariance functions (Gneiting, 2002).

2.3 Structured additive predictors and effect priors

In the structured additive regression framework (Fahrmeir et al., 2004; Wood, 2017), each predictor η_{ki} can be expressed as a sum of L_k smooth terms f_{kl},

$$\eta_{ki} = \sum_{l=1}^{L_k} f_{kl}(\boldsymbol{x}_i; \boldsymbol{\beta}_{kl}),$$

where each function f_{kl} is expanded from a basis representation as $f_{kl} = \sum_{d=1}^{D_{kl}} B_{kld}(\boldsymbol{x}_i)\beta_{kld}$, and $\boldsymbol{\beta}_{kl}$ are the regression coefficients to be estimated. The predictor η_{ki} can attain any real value and needs to be mapped to the (possibly constrained) parameter space of the distributional parameter θ_{ki} with a strictly monotonic link function h_k, i.e., $h_k(\theta_{ki}) = \eta_{ki}$ or $\theta_{ki} = h_k^{-1}(\eta_{ki})$.

A smooth term usually depends on one or two elements of the covariate vector $\boldsymbol{x}_i$, but it can also be of a higher dimension, e.g., in the case of simple linear covariate effects. The interpretation of a smooth term depends on the choice of the basis functions and the prior of the regression coefficients. In many cases, a (proper or improper) normal prior is assumed for the regression coefficients,

$$p(\boldsymbol{\beta}_{kl} \mid \tau_{kl}) \propto \exp\left(-\frac{1}{2\tau_{kl}^2}\boldsymbol{\beta}_{kl}^\top \mathbf{P}_{kl}\boldsymbol{\beta}_{kl}\right),$$

where τ_{kl} is a hyperparameter that controls the smoothness of the covariate effect, and $\mathbf{P}_{kl}$ is a penalty matrix. The hierarchical prior of the parameters of smooth term l in predictor k is given by $p(\boldsymbol{\beta}_{kl}, \tau_{kl}) = p(\boldsymbol{\beta}_{kl} \mid \tau_{kl}) \times p(\tau_{kl})$, where $p(\tau_{kl})$ is often an inverse gamma distribution with fixed hyperparameters $a = b = 0.0001$ or some other small value. In more complex scenarios, we might also have covariate effects that depend on more than just one single, scalar hyperparameter (e.g., to achieve adaptive smoothness) or another hierarchical prior layer that connects the hyperparameters (e.g., to control the overall model complexity, Klein & Kneib, 2016a), but we will stick to the simple case in this article. Notationally, we will assume a vector of hyperparameters $\boldsymbol{\tau}_{kl}$ for better generality, including the scalar case as a special case.

The smooth terms in a structured additive predictor can represent a broad range of covariate effects (linear, random, non-linear, spatial, etc.). A simple non-linear effect of a single covariate can be constructed using a polynomial basis. Cubic or B-splines provide a numerically more stable and flexible alternative to standard polynomials. In Section 5, we use a kriging smooth to model a spatial effect, which we can also represent as a linear combination of basis function evaluations. See Fahrmeir et al. (2013) for more details on smooth terms and structured additive predictors.

2.4 Related model classes

Mixed models with within-group correlation structures

If the index set T reduces to a finite set, the GPs become finite collections of dependent observations. For grouped data like this, where the groups could represent longitudinal observations on one person, mixed models are the standard tool. Distributional regression models with GP responses are closely related to the marginal distribution implied by mixed models. For example, the random intercepts model corresponds to

$$[Y_{i,1}, \ldots, Y_{i,n_i}]^\top \mid \boldsymbol{x}_i \overset{\text{ind.}}{\sim} \mathcal{N}_{n_i}(\boldsymbol{\mu}_i, \xi^2\mathbf{E}_{n_i} + \sigma^2\mathbf{I}_{n_i}), \tag{8}$$

where $[Y_{i,1}, \ldots, Y_{i,n_i}]^\top$ is the vector of the n_i measurements on group i, $\mathbf{E}_i$ is a $(n_i \times n_i)$-dimensional matrix of ones, $\mathbf{I}_{n_i}$ is the n_i-dimensional identity matrix, and ξ^2 represents the random effect variance. In this case, there are two (co-)variance parameters ξ^2 and σ^2. Adding random slopes to the model would increase the number of covariance parameters.

Some mixed model implementations (such as the `nlme` package for R, Pinheiro et al., 2020) allow for within-group correlation structures of the residuals, which means that the identity matrix $\mathbf{I}_{n_i}$ in Equation (8) can be replaced with a more complex correlation matrix. This model extension is useful for temporally or spatially correlated data. As in our model, the correlation matrix is defined in terms of a parametric correlation function such as the Matérn correlation function, but to the best of our knowledge, its parameters (and the further covariance parameters of mixed models) are usually not linked to group-specific covariates. We are also not aware of any software package that supports this model structure out of the box. Moreover, the `nlme` package does not support structured additive predictors, but other mixed model software packages like `gamm4` (Wood & Scheipl, 2020) do.

Functional data

Our model also has a close conceptual relation with GP regression models for functional data (Shi & Choi, 2011). For example, consider the functional response model

$$Y_i(\boldsymbol{t}) \overset{\text{ind.}}{\sim} \mathcal{N}(\mu_i(\boldsymbol{t}) + \omega_i(\boldsymbol{x}_i(\boldsymbol{t})), \sigma^2),$$

where $\boldsymbol{x}_i(\boldsymbol{t})$ are functional covariates, while a GP prior is assumed for $\omega_i(\boldsymbol{x}_i(\boldsymbol{t}))$, which defines the covariance structure of $Y_i(\boldsymbol{t})$. As a major difference, Shi and Choi (2011) do not allow for the inclusion of group-specific covariate effects on the (hyper-)parameters of the covariance functions of $\omega_i(\boldsymbol{x}_i(\boldsymbol{t}))$ and $Y_i(\boldsymbol{t})$.

Another aspect where the work of Shi and Choi (2011) is similar to that of Greven and Scheipl (2017) or Scheipl et al. (2015) but different from ours is the way the mean functions $\mu_i(\boldsymbol{t})$ are modelled: while our motivation is to use mean functions whose parameters serve as distributional parameters and are linked to covariates, Greven and Scheipl (2017) focus on (non-parametric) linear representations of the mean functions via suitable basis expansions. Given this difference, our model can be considered more realistic and more stable, at least in situations where prior knowledge on the shape of the mean functions such as the sigmoid shape of the Weibull growth curves in Section 5 is available. On the other hand, parametric mean functions might be too restrictive for some applications. Finally, our index set T can represent different (potentially non-Euclidean) metric spaces, while the functional data literature is typically concerned with time-indexed data.

3 Posterior estimation

We stack all regression coefficients in a vector $\boldsymbol{\beta}$, all hyperparameters in a vector $\boldsymbol{\tau}$, and all covariates in a matrix $\mathbf{X}$ consisting of the N rows $\boldsymbol{x}_i$. The unnormalised log-posterior is then given by

$$\log \pi(\boldsymbol{\beta}, \boldsymbol{\tau} \mid \boldsymbol{y}_1, \ldots, \boldsymbol{y}_N, \mathbf{X}) \propto \sum_{i=1}^{N} \log p_{\mathcal{N}}(\boldsymbol{y}_i \mid \boldsymbol{x}_i, \boldsymbol{\beta}) + \sum_{k=1}^{K} \sum_{l=1}^{L_k} \log p_{kl}(\boldsymbol{\beta}_{kl}, \boldsymbol{\tau}_{kl}).$$

The first term on the right-hand side is the log-likelihood of the regression coefficients $\boldsymbol{\beta}$, where the observed values for the ith GP are denoted by $\boldsymbol{y}_i$ and the density of the multivariate normal distribution by $p_{\mathcal{N}}$. The second term on the right-hand side is the joint log-prior of all parameters $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$, where p_{kl} is the density of the prior distribution of the regression coefficients and hyperparameters of smooth term l in predictor k.

We perform fully Bayesian inference with an adjusted version of the generic MCMC sampler of Umlauf et al. (2018), which uses inverse gamma priors and Gibbs updates for each scalar element of the hyperparameters $\boldsymbol{\tau}$, and Metropolis–Hastings updates with locally adaptive IWLS proposals (Gamerman, 1997) for the regression coefficients $\boldsymbol{\beta}$. As the IWLS proposals involve the observed or expected Fisher information matrix, the regression coefficients are sampled in blocks for numerical stability and efficiency. Typically, one block consists of the regression coefficients of one smooth term, and the blocks are sampled in a nested loop over the distributional parameters first and the smooth terms second. As discussed in the next section, we sample the parameters of certain smooth terms in one joint block, which can reduce the autocorrelation of the MCMC chains substantially.

The required full conditionals for the Gibbs updates of the hyperparameters are independent of the specific response structure of a distributional regression model and are given e.g., in Umlauf et al. (2018).

3.1 Model-specific scores and Fisher information

For a general distributional regression model, the score and the Fisher information of the regression coefficients $\boldsymbol{\beta}_{kl}$ are given by

$$s(\boldsymbol{\beta}_{kl}) = \sum_{i=1}^{N} s(\theta_{ki}) \frac{\partial \theta_{ki}}{\partial \boldsymbol{\beta}_{kl}} \quad \text{and} \quad \mathcal{I}(\boldsymbol{\beta}_{kl}) = \sum_{i=1}^{N} \mathcal{I}(\theta_{ki}) \frac{\partial \theta_{ki}}{\partial \boldsymbol{\beta}_{kl}} \left(\frac{\partial \theta_{ki}}{\partial \boldsymbol{\beta}_{kl}} \right)^{\top},$$

where the distributional parameters θ_{ki} are functions of the regression coefficients $\boldsymbol{\beta}_{kl}$ composed of smooth terms, structured additive predictors, and inverse link functions. The derivatives $\partial\theta_{ki}/\partial\boldsymbol{\beta}_{kl}$ are usually easy to compute. For this reason, we only give $s(\theta_{ki})$ and $\mathcal{I}(\theta_{ki})$, the score and the Fisher information of the distributional parameters with respect to the response distribution, for the particular case of GP responses in the distributional regression framework.

Using the definitions of the mean vector $\boldsymbol{\mu}_i$ and the covariance matrix $\boldsymbol{\Sigma}_i$ from Equation (4), the unnormalised log-likelihood contribution of the ith GP is

$$\log p_{\mathcal{N}}(\boldsymbol{y}_i \mid \boldsymbol{x}_i, \boldsymbol{\beta}) \propto -\frac{1}{2}(\log|\boldsymbol{\Sigma}_i| + (\boldsymbol{y}_i - \boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_i)).$$

For better readability, we omit the observation index i in the following formulas. Let θ_k be a distributional parameter of the *mean function*, then the score and the Fisher information of θ_k are

$$s(\theta_k) = \left(\frac{\partial\boldsymbol{\mu}}{\partial\theta_k}\right)^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}) \quad \text{and} \quad \mathcal{I}(\theta_k) = \left(\frac{\partial\boldsymbol{\mu}}{\partial\theta_k}\right)^\top \boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\mu}}{\partial\theta_k}.$$

Now, let θ_k be a distributional parameter of the *covariance function*, then the score of θ_k is

$$s(\theta_k) = -\frac{1}{2}\left[\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\right) - (\boldsymbol{y} - \boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{\mu})\right]$$

and the corresponding Fisher information is

$$\mathcal{I}(\theta_k) = \frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\theta_k}\right).$$

The derivatives $\partial\boldsymbol{\mu}/\partial\theta_k$ and $\partial\boldsymbol{\Sigma}/\partial\theta_k$ depend on the specific mean and covariance function of the GPs. Typically, one of the distributional parameters of the covariance function will be the standard deviation σ (or the variance σ^2, depending on the parameterisation). In this case, the covariance matrix is given by $\boldsymbol{\Sigma} = \sigma^2\mathbf{R}$, where $\mathbf{R}$ is the correlation matrix, and the score and the Fisher information of σ simplify to

$$s(\sigma) = -\frac{1}{\sigma}\left(n - (\boldsymbol{y} - \boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{\mu})\right) \quad \text{and} \quad \mathcal{I}(\sigma) = \frac{2n}{\sigma^2}.$$

3.2 Sampling the covariance parameters in one block

To sample the regression coefficients of the smooth terms l and $\tilde{l}$ in the predictors k and $\tilde{k}$ in one block, their joint Fisher information is required,

$$\mathcal{I}\left(\begin{bmatrix}\boldsymbol{\beta}_{kl}\\ \boldsymbol{\beta}_{\tilde{k}\tilde{l}}\end{bmatrix}\right) = \begin{bmatrix}\mathcal{I}(\boldsymbol{\beta}_{kl}) & \operatorname{Cov}(s(\boldsymbol{\beta}_{kl}), s(\boldsymbol{\beta}_{\tilde{k}\tilde{l}}))\\ \operatorname{Cov}(s(\boldsymbol{\beta}_{\tilde{k}\tilde{l}}), s(\boldsymbol{\beta}_{kl})) & \mathcal{I}(\boldsymbol{\beta}_{\tilde{k}\tilde{l}})\end{bmatrix},$$

where

$$\operatorname{Cov}(s(\boldsymbol{\beta}_{kl}), s(\boldsymbol{\beta}_{\tilde{k}\tilde{l}})) = \sum_{i=1}^{N}\operatorname{Cov}(s(\theta_{ki}), s(\theta_{\tilde{k}i}))\frac{\partial\theta_{ki}}{\partial\boldsymbol{\beta}_{kl}}\left(\frac{\partial\theta_{\tilde{k}i}}{\partial\boldsymbol{\beta}_{\tilde{k}\tilde{l}}}\right)^\top.$$

Specifically, we want to improve the sampling performance for the covariance parameters, so we need the covariance of the score of the standard deviation and the range, which is given by

$$\operatorname{Cov}(s(\sigma), s(\phi)) = \frac{1}{\sigma}\operatorname{tr}\left[\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\phi}\right].$$

4 Simulation study

We designed a simulation study with three scenarios: the first scenario shows that the sampling scheme from Section 3 can greatly improve the performance of the 'standard' IWLS sampler with separate blocks for each distributional parameter and smooth term. Scenario II resembles the real-world application to intra-annual tree stem growth in Section 5, extending it with an artificial time-varying covariate as a proof of concept. In the third scenario, we use GPs on a sphere, which can be understood as shapes of tree crowns. While this simulation is not immediately linked to the application of analysing tree stem radial growth, it underlines that the index set of the GPs does not need to be one-dimensional or Euclidean. To communicate a clear message with each scenario, we refrained from adding unnecessary complexity: All scenarios use normal priors with mean zero and standard deviation 1,000 for the regression coefficients, 100 replications of the data-generating process, and MCMC chains of length 1,000 after a burn-in of 200 iterations.

4.1 Scenario I: joint sampling

In this scenario, we use a constant mean function and the Matérn covariance function c^m from Equation (7). The smoothness parameter ν of the Matérn covariance function is fixed to 1.5, and the predictors and inverse link functions are defined to be

$$\mu_i = x_{i1}, \quad \sigma_i = \exp(x_{i2} + x_{i4}), \quad \text{and} \quad \phi_i = \exp(\beta_0 + x_{i3} + x_{i4}),$$

where $x_{i1}, x_{i2}, x_{i3}, x_{i4} \overset{\text{ind.}}{\sim} \mathcal{U}(0, 1)$ are the covariates. The observation index i runs from 1 to N, where $N = 30$ is the number of GPs, and $n_i = n = 30$ is the number of observed values per GP. The unit interval serves as the index set of the GPs.

For β_0, the intercept for the range ϕ_i, we use two different values: -3, such that $\phi_i \in [0.049, 0.368]$, and 0, such that $\phi_i \in [1, 7.39]$. We call $\beta_0 = -3$ the 'small-range scenario' and $\beta_0 = 0$ the 'large-range scenario.' Figure 2 shows 30 simulated GPs from one exemplary replication of the simulation scenario. The realisations of the GPs in the large-range scenario seem almost linear, while the realisations in the small-range scenario are much more wiggly.

We ran 100 replications of this setup, both with a small and a large range. In a next step, we sampled the (correctly specified) model for each replication and range, one time with separate blocks for the regression coefficients for the standard deviation and the range, and another time with one joint block for these coefficients. The bias of the posterior mean estimates is negligible for both samplers, but in terms of the autocorrelation of the chains, there are substantial differences between them. While the autocorrelation is similar for the regression coefficients for x_{i2} and x_{i3}, the joint sampler beats the one with separate blocks by far for the intercept and the regression coefficients for x_{i4} (the covariate with an effect on both the standard deviation and the range; see Figure 3 for a comparison of the trace plots). With a small range, the performance gap is less extreme but still very apparent.

We conclude from Scenario I that using a joint sampler for the covariance parameters is much more efficient. In contrast, with the sampling scheme with separate parameter blocks for the covariance parameters, the resulting MCMC chains have a high autocorrelation and must be inspected carefully. For Scenarios II and III, similar performance differences between the samplers can be observed, which we demonstrate in the supplementary material to this article.

4.2 Scenario II: time-varying covariates

The mean function of the GPs in this scenario is constructed as the sum of the Weibull growth curve from Equation (6) and the linear mean function from Equation (5), yielding

$$m(\boldsymbol{z}_i(\boldsymbol{t}) = [t, u_i(t)]^\top; \boldsymbol{\theta}_i^m = [l_i, a_i, b_i, w_i]^\top) = m^w(t; [l_i, a_i, b_i]^\top) + m^l(u_i(t); w_i),$$

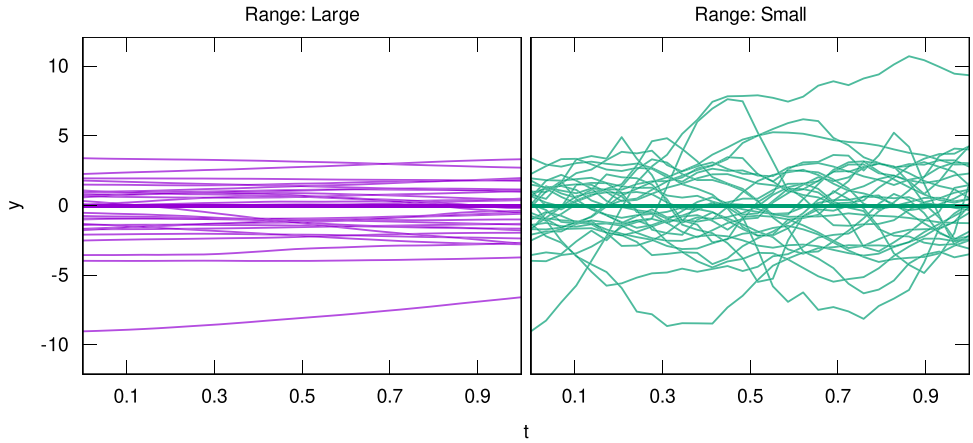

Figure 2. The GPs from one exemplary replication of Simulation Scenario I. The range takes values between 1 and 7.39 for the GPs on the left-hand side and between 0.049 and 0.368 for those on the right-hand side.

where $u_i(t)$ is a univariate time-varying covariate, and the parameter w_i determines the effect size of $u_i(t)$. The covariance function of the GPs is

$$c^s(z_i(\boldsymbol{t}) = t, z_i(\boldsymbol{t}') = t'; \boldsymbol{\theta}_i^c = [\sigma_i, \phi_i]^\top) = q(t) \times q(t') \times c^m(t, t'; [\sigma_i, \phi_i]^\top), \tag{9}$$

where c^m is the Matérn covariance function from Equation (7) with the smoothness parameter $\nu = 1.5$. The auxiliary function $q(t)$ is defined as

$$q(t) = \begin{cases} 0.1 + \frac{0.9}{30} \times t & \text{if } 0 \leq t < 30, \\ 1 & \text{otherwise,} \end{cases} \tag{10}$$

and scales the standard deviation of the GPs over time, such that it increases linearly on the interval [0, 30] and remains constant afterwards. The motivation for this step is that the growth curves in the application in Section 5 are defined to start at zero on April 1 of each year (i.e., at the beginning of each growing season) and have little variability in the first couple of weeks after that.

We define the predictors and inverse link functions $l_i = 1500 \approx \exp(7.313)$, $a_i = 3.5 \approx \exp(1.253)$, $b_i = 100 \approx \exp(4.605)$, $w_i = x_i$, $\sigma_i = 40 \approx \exp(3.689)$, and $\phi_i = 2 \approx \exp(0.693)$. The only explanatory variable x_i, where $i = 1, \ldots, N$, is independent and uniformly distributed on the interval [1, 2]. The number of GPs takes the values $N = 30$, 60, or 120, and the number of observed values per GP is $n_i = n = 60$ or 120. As the index set of the GPs, we use the interval [0, 182], representing the days during one growing season.

The focus of this scenario is on the time-varying covariate $u_i(t)$, which we simulate as i.i.d. GPs with mean zero and a squared exponential covariance function, scaled over time with the auxiliary function $q(t)$ from Equation (10). In the context of intra-annual tree stem growth, the time-varying covariate $u_i(t)$ could, for example, be a mean-centred moving average over the precipitation at the location of a tree i. How much the precipitation affects the growth dynamics of tree i depends on the explanatory variable x_i, which might represent the soil conditions around that tree. Figure 4 illustrates this simulation scenario.

For each possible N-n-combination, we ran 100 replications and found that the sampling scheme from Section 3 works very reliably. The posterior mean estimates do not show any systematic bias (Figure 5). They have very little variability around the true value for the parameters of the Weibull growth curve, while more variability is observed for the parameter of the time-varying covariate and the covariance parameters. As expected, the quality of the estimates improves with the sample size, where N has a stronger effect on the quality than n, because additional independent response GPs are more informative than additional dependent observations within each GP. In

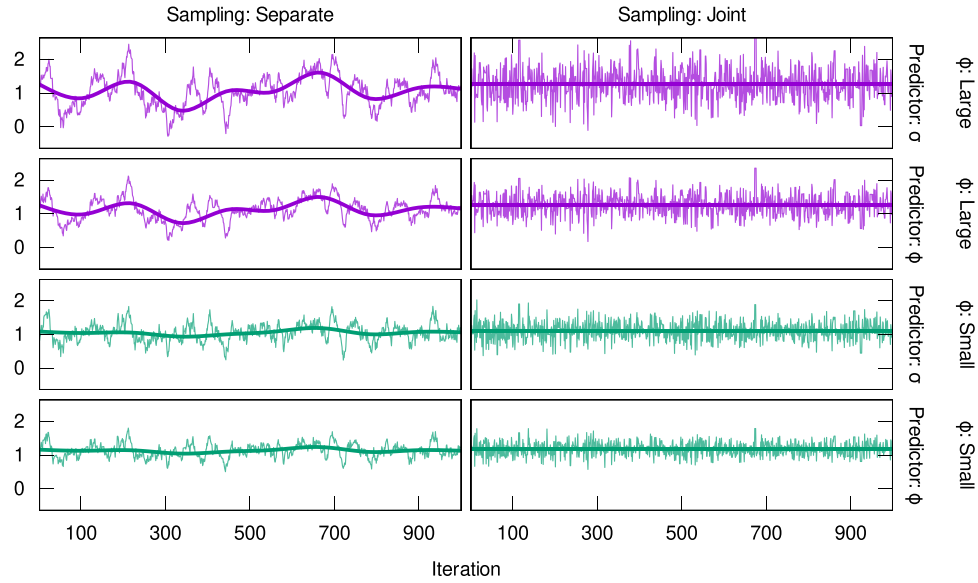

Figure 3. Trace plots for the regression coefficients for x_{i4} from one exemplary replication of Simulation Scenario I. The left-hand side shows the sampler with separate blocks for the regression coefficients for the covariance parameters, whereas the right-hand side shows the sampler with one joint block for these coefficients. No thinning was applied to the chains.

summary, this simulation scenario shows that our sampler has the expected properties for a correctly specified model, even if time-varying covariates are included.

Finally, we highlight that the boxplots in Figure 5 show the performance of the sampler under the assumption that the true smoothness parameter ν is known, which is usually not the case in practice. However, a misspecified smoothness parameter ($\nu = 0.5$ or 2.5 instead of 1.5 in this scenario) does not seem to have a strong adverse effects on the inferences drawn from such a model. In the supplementary material to this article, we show that a misspecified smoothness parameter is mostly compensated for by the estimated range parameter ϕ, while the other parameters remain essentially unaffected. Rather, we found that a reasonable choice for ν can increase the model stability and improve the MCMC mixing, as we will investigate in Section 5.

4.3 Scenario III: processes on a sphere

In this scenario, we show how our model can accommodate spatial or spatio-temporal processes as response structures. Generally speaking, the GPs can be defined on a one- or higher-dimensional Euclidean space, or even a non-Euclidean metric space when employing appropriate distances. The processes in this specific scenario are defined on a sphere, resembling shapes of tree crowns, and we use the great circle distance for quantifying distances. Figure 6 illustrates the design: the object on the left is an 'average' tree crown, from which we simulated a more realistic shape as a realisation of a GP with an exponential covariance function, shown on the right. In an application, the tree species or the light availability could be used as covariates to explain the properties of the mean and the covariance function of the crown shapes. The mean properties are, among others, the average radius, and the vertical elongation, while the covariance properties are the size and the persistence of the deviations from the mean.

The mean function in this scenario is defined in terms of the linear mean function from Equation (5), which is applied to a transformed coordinate vector as follows:

$$\begin{aligned} &m(z_i(\boldsymbol{t}) = [t_1, t_2]^\top; \boldsymbol{\theta}_i^m = [r_i, h_i, v_i]^\top) \\ =&m^l([1,\ \cos(t_2)(\cos(t_1) + 1), t_2 + \pi/2]^\top; [r_i, h_i, v_i]^\top) \\ =&r_i + \cos(t_2)(\cos(t_1) + 1) \times h_i + (t_2 + \pi/2) \times v_i. \end{aligned}$$

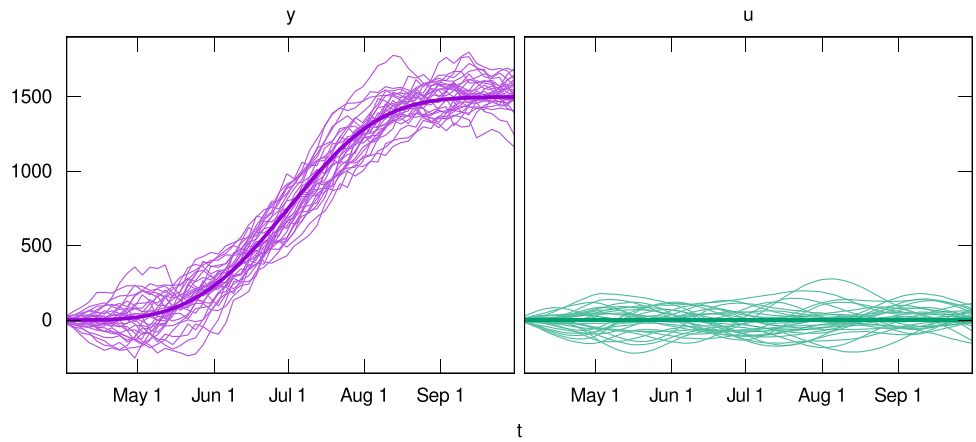

Figure 4. The response GPs $Y_i(t)$ (on the left-hand side) and the time-varying covariates $u_i(t)$ (on the right-hand side) from one exemplary replication of Simulation Scenario II with $N = 30$ and $n = 60$. The large-scale deviations of $Y_i(t)$ from the Weibull growth curve are mainly driven by mean shifts due to $u_i(t)$, while the small-scale variation comes from the covariance structure of $Y_i(t)$.

The coordinates $t_1 \in [-\pi, \pi]$ and $t_2 \in [-\pi/2, \pi/2]$ are the longitude and the latitude on the sphere in radians. The parameters are r_i, the minimum radius of the tree crown, h_i, the horizontal elongation towards the south, and v_i, the vertical elongation. Furthermore, we use the Matérn covariance function c^w from Equation (7) with the smoothness parameter $\nu = 0.5$, i.e., an exponential covariance function.

The predictors and inverse link functions are defined as $r_i = \exp(x_{i1})$, $s_i = \exp(x_{i2})$, $h_i = \exp(1 + x_{i3})$, $\sigma_i = \exp(x_{i4})$, and $\phi_i = \exp(x_{i5})$, where the explanatory variables x_{i1}, x_{i2}, x_{i3}, x_{i4}, and x_{i5} are independent and uniformly distributed on the unit interval and the observation index is $i = 1, \ldots, N$. The number of GPs is set to $N = 30$, and the number of observed values per GP is $n_i = n = 379$. A regular longitude–latitude grid is used for the observations of the GPs.

With a maximum value of 0.014 for the covariate effect of x_{i1} on the radius r_i, the average bias is negligible for all posterior mean estimates. The average mean squared errors (MSEs) are also very small, especially for the regression coefficients for the vertical elongation v_i, the standard deviation σ_i, and the range ϕ_i. Among these regression coefficients, the maximum average MSE is 0.006 for the covariate effect of x_{i3} on v_i. The average MSEs for the regression coefficients for r_i and the horizontal elongation h_i are higher but still uncritical with values between 0.02 and 0.26 (results not shown graphically for this scenario). These numbers indicate that we are able to estimate the model parameters reliably with our sampling scheme, despite the fact that the GPs are defined on a non-Euclidean space.

5 Intra-annual tree stem growth

In this section, we apply our method to the intra-annual stem growth of 72 beeches, 6 ashes, and 7 sycamores from three different regions in Germany. For each tree, the growing seasons 2012 and 2013 were recorded in a high temporal resolution using electronic circumference dendrometers. The original purpose of the data was a study on the effect of the neighbourhood identity on the growth patterns of beech trees in pure and mixed stands (Metz et al., 2020), which was conducted in the *Biodiversity Exploratories* (Fischer et al., 2010). The dataset can be downloaded from the information system of the project (*BExIS*, Metz & Ammer, 2018; Ostrowski et al., 2016), thanks to the open data policy of the *Biodiversity Exploratories*. Our analysis is fully replicable with the code in the supplements.

A first outlook on the dataset was given in Figure 1. We compared a beech to an ash and observed differences in the overall annual growth, the starting time of the growth process, and the shrinking and swelling. With our model, we formalise these observations and assess the effect of explanatory variables such as the DBH or the geographical location. We show that our model

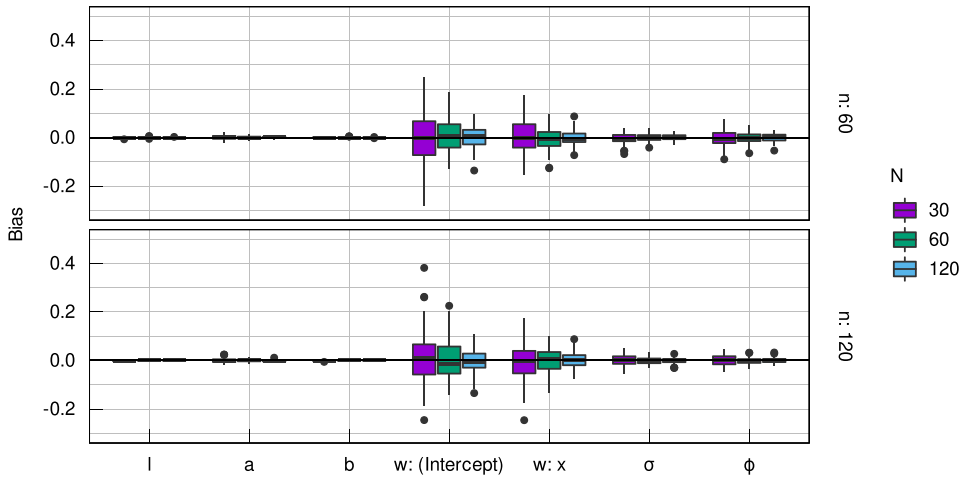

Figure 5. Boxplots of the bias of the posterior mean estimates in Simulation Scenario II. Each boxplot represents 100 replications for one combination of N, the number of response GPs, and n, the number of observed values per GP.

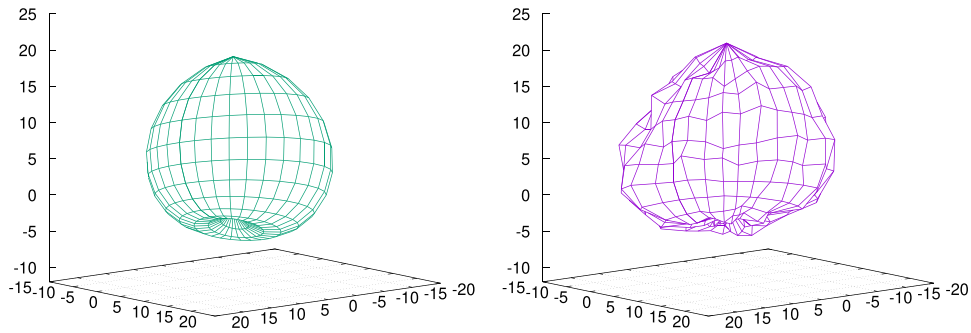

Figure 6. An exemplary mean function, shown on the left, and a corresponding realisation of a GP with an exponential covariance function, shown on the right, from one replication of Simulation Scenario III. The objects are designed to resemble shapes of tree crowns. The properties of the shapes can be related to explanatory variables such as the tree species or the light availability.

is instructive when applied to high-resolution dendrometer data. Our statistical approach is different from the one used by Metz et al. (2020), who estimate one Weibull curve per observed curve, each fitted individually by non-linear least squares, and then model the estimated parameters and other derived quantities. One downside of this two-step procedure is that the estimation uncertainty of the parameters is not systematically taken into account in the second step. We solve this problem with an explicit assumption about the probability distribution of the stochastic process of intra-annual tree stem growth and a one-step inference algorithm.

5.1 Model specification

Our analysis is based on the Weibull growth curve $m^{w}(t)$ from Equation (6) as a mean function and the scaled Matérn covariance function $c^{s}(t, t')$ from Equation (9). Consequently, we have the following five distributional parameters: the limit l, the shape a, and the scale b of the Weibull growth curve, and the standard deviation σ and the range ϕ of the covariance function.

The covariance function describes the shrinking and swelling of the tree stems and does not include any additional parameters to account for potential measurement errors from the dendrometers. The motivation for this approach is that electronic dendrometers are high-precision devices,

and the individual measurement errors are on a very small scale. Large errors occur only when the dendrometers are touched, e.g., by an animal or a researcher. These errors were already corrected in our dataset with a simple post-processing step. Finally and most importantly, the data were originally recorded in a very high temporal resolution of 30 min. As we are only working with daily data, a potential additive error per half-hourly measurement would average out over the 48 measurements on one day. From a theoretical perspective, however, including a measurement error in the model would be straightforward, see the discussion in Section 2.2.2.

The predictors and inverse link functions are defined as

$$
\begin{aligned}
l_i &= \exp(\beta_{l0} + (\text{Tree} * \text{Year})_i \times \boldsymbol{\beta}_{l1}),\\
a_i &= \exp(\beta_{a0} + \text{Species}_i \times \boldsymbol{\beta}_{a1} + \text{DBH}_i \times \beta_{a2} + (\text{Site} * \text{Year})_i \times \boldsymbol{\beta}_{a3}),\\
b_i &= \exp(\beta_{b0} + \text{Species}_i \times \boldsymbol{\beta}_{b1} + \text{DBH}_i \times \beta_{b2} + (\text{Site} * \text{Year})_i \times \boldsymbol{\beta}_{b3}),\\
\sigma_i &= \exp(\beta_{\sigma 0} + \text{Species}_i \times \boldsymbol{\beta}_{\sigma 1} + \text{DBH}_i \times \beta_{\sigma 2} + f_{\text{Year}_i}(x_i, y_i; \boldsymbol{\beta}_{\sigma 3})),\\
\phi_i &= \exp(\beta_{\phi 0} + \text{Species}_i \times \boldsymbol{\beta}_{\phi 1} + \text{DBH}_i \times \beta_{\phi 2} + (\text{Site} * \text{Year})_i \times \boldsymbol{\beta}_{\phi 3}),
\end{aligned}
$$

where $\beta_{\bullet,0}$ and $\beta_{\bullet,2}$ are scalar regression coefficients, while $\boldsymbol{\beta}_{\bullet,1}$ and $\boldsymbol{\beta}_{\bullet,3}$ are vectors of regression coefficients, and Species_i denotes the entries of the design matrix for the dummy variable for the species of the tree where the ith growth curve was recorded,

$$
\text{Species}_i = \begin{cases} [0, 0] & \text{if the } i\text{th growth curve is of a beech,} \\ [1, 0] & \text{if it is of a ash; and} \\ [0, 1] & \text{if it is of a sycamore.} \end{cases}
$$

Similarly, $(\text{Tree} * \text{Year})_i$ and $(\text{Site} * \text{Year})_i$ are the entries of the design matrix for the interaction of two dummy variables: in the first case, of the individual tree and the year, and in the second case, of the field site and the year. Finally, f_{Year_i} denotes a year-specific spatial kriging smooth.

The tree-year-interaction in the predictor for the limit parameter implies one degree of freedom for the limit of each growth curve. The other variables in the dataset do not explain the overall annual growth sufficiently, but the limit is identified well enough for each growth curve that these parameters can be estimated without problems. All other predictors include the species and the DBH as covariates. For the shape, the scale, and the range, the site-year-interaction captures the spatial and temporal differences between the field sites and the years. For the standard deviation, the smooth term f_{Year_i} serves this purpose and illustrates the flexibility of structured additive predictors when used with covariance parameters in the GP framework.

For the regression coefficients $\boldsymbol{\beta}$, we used uninformative $\mathcal{N}(0, 1000)$ priors, and an inverse gamma prior with fixed hyperparameters $a = b = 0.0001$ for the smoothing parameter τ^2 of the spatial kriging smooth f_{Year_i}.

5.2 Tree physiology

The dendrometer measurements of the stem radius of a tree are composed of an irreversible growth component and temporary shrinking and swelling dynamics, which can be further divided into water potential-driven and osmotic processes (Chan et al., 2016). Water potential-driven changes in the stem radius are caused by sap moving radially within the xylem or between the xylem and the phloem from areas with a higher to areas with a lower water potential. This process give rise to an approximately periodic fluctuation of the dendrometer measurements over the course of 24 hr, while osmotic changes occur more gradually, for example, if the tree draws water from the roots.

In our case, the mean curves can be interpreted as the irreversible growth and the fluctuations around them as temporary shrinking and swelling. These fluctuations are characterised by the covariance function of the GPs. As the employed Matérn covariance function is stationary, the GPs keep returning to their mean. How fast they return depends on the range parameter, which is estimated to be relatively small for most trees in the dataset. As we analyse the dendrometer measurements in a daily resolution, the water potential-driven changes are averaged out from the data for the most part, and the remaining fluctuations are primarily osmotic. In fact, the deviations of the

growth curves from the estimated mean curves that we observe in the data do typically last a few days or weeks, as expected for the osmotic processes in tree stems.

If a growth curve increases faster than the mean curve, we interpret this as the tree drawing more water from the roots than required for the formation of new cells and the irreversible growth at a given moment. Conversely, if a growth curve increases slower than the mean curve, more water is consumed by the irreversible growth than drawn from the roots. Finally, if a growth curve decreases, water is released from the stem. These processes are reflected in the stochastic part of the model, which is characterised by the covariance parameters: the standard deviation quantifies the magnitude of the osmotic changes in the stem radius, and the range parameter their persistence. For example, some tree species might store more water in the stem than others (which would imply a higher standard deviation), or they might store it for a longer time (which would imply a higher range parameter).

Different approaches for the decomposition of dendrometer measurements have been proposed, among others, by Zweifel et al. (2005) and Chan et al. (2016). Zweifel et al. (2005), on the one hand, use a linear interpolation of the local maxima of the observed growth curves as an approximation of the irreversible growth and interpret the difference between the interpolation and the observed curves as the tree water deficit. They find that this measure of tree water deficit is explained well by soil water potential and vapour pressure deficit for pine, oak, and spruce under different environmental conditions in Switzerland. Chan et al. (2016), on the other hand, compute an estimate of the sum of the irreversible growth and the osmotic changes in the stem radius from dendrometer measurements of both the whole stem and the xylem radial thickness. The irreversible growth is then obtained as the difference of the minima of this estimate on two consecutive days.

Despite the apparent similarities between our model framework and the approaches of Zweifel et al. (2005) and Chan et al. (2016), the scope of the methods is quite different: While the other approaches are motivated from ecophysiological considerations, our model takes advantage of statistical assumptions about the parametric form of the mean and covariance function of the GPs in a regression setting. It can be used to decompose dendrometer measurements into a permanent and a temporary component for any given point in time, but this is not our primary goal, and the decomposition is likely to be less accurate than the ones from the other, more specialised methods. Instead, our model does focus on the relationship between structural patterns of both the irreversible growth and the temporary shrinking and swelling throughout the vegetation period and a set of explanatory variables. As mentioned before, such structural patterns could be for example the magnitude or persistence of the osmotic changes in the stem radius. In particular, we would like to point out the following advantages of our approach:

- The model has minimal data requirements: A single circumference dendrometer per tree is sufficient. Zweifel et al. (2005) and Chan et al. (2016) use one or two point dendrometers, respectively. The model is also agnostic about the temporal resolution of the data, while the method of Chan et al. (2016) requires multiple measurements per day.
- The model can be estimated with one integrated MCMC algorithm, which estimates the decomposition of the dendrometer measurements into irreversible growth and temporary shrinking and swelling, and the effects of the explanatory variables on the characteristics of the growth curves at the same time. The advantage is that the estimation uncertainty can be assessed in a sound statistical framework, and a two-step procedure can be avoided.
- The structural patterns of the growth curves that can be explained by covariates are not limited to the osmotic changes in the stem radius, but include characteristics of the overall intra-annual sigmoid growth curves as well, such as the total growth of a tree over an entire vegetation period or the steepness of a growth curve; for details, see Section 2.2.1.

The overall aim of our article is to present a model *framework* rather than one specific model for tree stem growth. The application in this section should illustrate how the different components of the framework can be interpreted in the context of tree stem growth. For the sake of simplicity, a relatively low temporal resolution and a limited number of covariates are used, but the model could easily be refined with a different mean or covariance function, more covariates, or a higher temporal resolution. In the view of this, the methods of Zweifel et al. (2005) and Chan et al. (2016) should not be considered as competitors of our model, but their studies could serve as a basis for

the development of more specific models in our framework that could provide further insights into the ecophysiological process of tree stem growth. In fact, recent studies by Zweifel et al. (2021) have confirmed vapour pressure deficit as an important driver of tree stem growth, so including it as a time-varying covariate in the mean or covariance function could be particularly instructive. In our application, the site-year-interactions in the predictors serve as proxies for the average weather conditions at a given site in a given year and ensure that the species and DBH effects do not suffer from an omitted variable bias, but the explicit inclusion of vapour pressure deficit or soil water potential would give rise to a more direct model.

Recent articles have addressed the question during which time of the day trees and other plants grow the most (Wiese et al., 2007; Zweifel et al., 2021). To investigate this problem in our model framework, the growth curves would need to be processed in a higher, e.g., hourly, temporal resolution, which would increase the computational cost of our model and require some adjustments of the covariance and possibly the mean function of the GPs: to account for the daily pattern of the water potential-driven changes in the tree stem radius, an additive periodic kernel could be included in the covariance function, for example, an exponential sine squared kernel (Rasmussen & Williams, 2006, Section 4.2). To address the question whether the irreversible growth primarily occurs during the day or the night, the mean function could be modified to represent a sequence of daily smooth step functions, where a parameter could be introduced to estimate the time of the day around which the steps are centred.

5.3 Estimation results

We used the sampling scheme from Section 3 to estimate our tree growth model with results presented in this section based on MCMC samples from the posterior distribution with a sample size of 10,000, excluding the 2,000 burn-in iterations. No thinning was applied to the chains. The effective sample size ranges from 703.169 to 8,793.660, as for some regression coefficients in the predictors for the mean parameters, the chains exhibit moderate to strong autocorrelation.

Table 1 summarises the posterior samples of the species effect of ash and the effect of DBH on the predictors. The reference category for the species effect is beech, so the negative effect on the scale implies that, on average, an ash stops growing earlier during the vegetation period than a beech. As different species allocate resources differently throughout the growing season, this is an expected result. The positive effect of DBH on the growth duration during the vegetation period might be due to the fact that trees with a greater DBH are more likely to be dominant in the stand. When the light availability decreases in the fall, the dominant trees still continue to grow, while the smaller trees cannot keep up their growth.

In terms of the covariance parameters, ash has a positive effect on the standard deviation. This is plausible because ash has a thicker bark than beech, which means it can store more water in the bark. The same argument applies for the effect of DBH on the standard deviation, as larger trees have a thicker bark. The effect of ash on the range parameter is estimated to be negative, which means that the osmosis-induced changes in the stem radius are less persistent for ash than for beech, possibly because the water is stored for shorter periods of time. Finally, the effect of DBH on the range is slightly positive, but the 95% credible interval does not exclude a zero effect.

The spatial kriging smooth in the predictor for the standard deviation is displayed in Figure 7. The six rectangles represent the three study regions Schwäbische Alb, Hainich-Dün, and Schorfheide-Chorin of the *Biodiversity Exploratories* and the years 2012 and 2013. The locations of the field sites in the study regions are marked with small crosses. A brighter colour indicates a higher standard deviation of the growth curves of the trees at a given location. The figure shows that the differences in the standard deviation are greater between than within the study regions: The trees on the Schwäbische Alb have the highest standard deviation, followed by those in the Hainich-Dün and the Schorfheide-Chorin. This pattern is stable over the years and very likely a result of differences in the precipitation, which are greater on a large than on a small scale. Note that the estimated effects are extrapolated considerably beyond areas supported by the observations, such that these parts of the rectangles in Figure 7 should be interpreted with care.

To check the robustness of the results with respect to the smoothness assumption of the correlation function, we also estimated the model with the smoothness parameters $\nu = 0.5$ and 2.5 instead of 1.5. We found that all three models produce similar results, both in terms of the model fit and the

Table 1. Summary statistics of the posterior samples of the species effect of ash (vs. beech) and the effect of DBH on the predictors

Ash	Coefficient	Mean	2.5%	Median	97.5%
Shape	$\beta_{a1,1}$	−0.289	−0.363	−0.289	−0.218
Scale	$\beta_{b1,1}$	−0.426	−0.452	−0.426	−0.400
Std. dev.	$\beta_{\sigma 1,1}$	0.602	0.529	0.602	0.678
Range	$\beta_{\phi 1,1}$	−0.101	−0.157	−0.101	−0.043
DBH	**Coefficient**	**Mean**	**2.5%**	**Median**	**97.5%**
Shape	β_{a2}	−0.027	−0.044	−0.027	−0.009
Scale	β_{b2}	0.050	0.045	0.050	0.055
Std. dev.	$\beta_{\sigma 2}$	0.085	0.064	0.085	0.106
Range	$\beta_{\phi 2}$	0.011	−0.004	0.011	0.027

estimated covariate effects. For $\nu = 0.5$, we had to use informative standard normal priors for the regression coefficients $\boldsymbol{\beta}$ instead of the default $\mathcal{N}(0, 1000)$ priors, and still some parameters appeared to be poorly identified, deteriorating the mixing of the MCMC chains. Comparing the results for $\nu = 1.5$ and 2.5, we obtained the better DIC for $\nu = 1.5$, making this model our preferred specification. The estimated covariate effects, particularly their signs and sizes, were comparable for all three models, such that the interpretation of the model is unaffected by the specific choice of the smoothness parameter. For more details, see the supplementary material to this article.

The discussion of the model results shows that our framework allows us to relate different properties of the tree growth curves to explanatory variables and complex covariate effects (such as the spatial kriging smooth in this example) in a very direct way. Using the methods of Zweifel et al. (2005) or Chan et al. (2016), similar results could be obtained by decomposing the dendrometer data, defining measures for the phenomena of interest, e.g., the persistence of the osmotic changes in the stem radius, based on the decomposition, and finally using these measures as response variables in different regression models. Both approaches have benefits and drawbacks depending on the goal of the analysis, but if the focus is on the effect of the explanatory variables, our model is arguably more comprehensive.

6 Discussion

In this paper, we embedded GPs as response structures into the framework of structured additive distributional regression as described by Klein et al. (2015) to study intra-annual tree stem growth and to decompose high-resolution dendrometer measurements into irreversible growth and temporary shrinking and swelling. It is of particular interest for the physiological understanding of stem growth that our model can explain certain properties of both components of the dendrometer measurements by covariates such as the tree species or the DBH. Based on a dataset of 85 individual trees from Germany, for which the variations in the stem radius were recorded during the growing seasons 2012 and 2013, we could identify different growth patterns for three deciduous tree species: for instance, ash grows more gradually and earlier during the vegetation period than beech, and its thick bark gives rise to a more pronounced temporary shrinking and swelling. Our model can quantify these differences between tree species and conditional on other explanatory variables with a sound and unified statistical approach.

We want to point out that the design of the structured additive predictors, that is the selection of the explanatory variables and their effect type, requires special care for the proposed type of model. Theoretical considerations and subject-matter expertise must be taken into account to build models with meaningful interpretations for the research questions at hand. Concerning the analysis of intra-annual tree stem growth, additional, more detailed models including precipitation data and other time-varying explanatory variables could be a promising next step, especially since

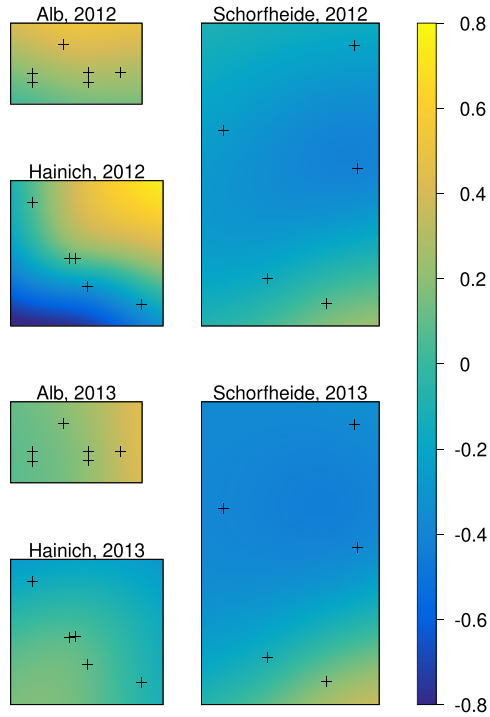

Figure 7. Posterior mean of the spatial kriging smooth f_{Year_i} in the predictor for the standard deviation. The field sites are marked with small crosses. Lighter colours indicate a higher standard deviation.

in the light of global climate change, a comprehensive understanding of the trees' reaction to drought stress is becoming more and more indispensable.

While the original use case for our model class is the analysis of intra-annual tree stem growth, the flexibility, and versatility of the model framework was discussed throughout the paper and demonstrated in particular in the simulation study. The general model class certainly deserves further investigation in the future. To explore its full potential in many other applications, it will be necessary to study the various members of the model class arising from specific choices of index sets, mean functions, and covariance functions, and to develop applications for research questions in different fields.

Other aspects that deserve further attention are both theoretical and software-related. We implemented an R package that can fit the models described in this paper but does not yet support arbitrary mean and covariance functions. The challenge will be to keep the performance cost of these generalisations as small as possible. To reach a greater audience, the package will also need a more complete documentation and a better user interface. In terms of open theoretical questions, the propriety of the posterior distribution and the ergodicity of the MCMC chains comes to mind. Finally, following up on the discussion in Sections 3 and 4, a more thorough investigation of the correlation structure of the model parameters could guide the development of more efficient MCMC sampling schemes.

Acknowledgments

We would like to thank Jérôme Metz, Peter Annighöfer, and Christian Ammer for providing us with the tree growth dataset, which was originally collected for the DFG Priority Program 1374: Biodiversity Exploratories, and two anonymous reviewers and the associate editor for their helpful comments.

Conflict of interest: None declared.

Funding statement

The work of H.R. on this article was funded by the German Research Foundation (DFG) through the Research Training Group 2300: Enrichment of European Beech Forests with Conifers (316045089/GRK2300). N.K. was funded by the German Research Foundation (DFG) through the Emmy Noether grant KL 3037/1-1.

Data availability

All data that were used in the article is freely available online and listed in the references.

Supplementary material

Supplementary material on the smoothness parameter of the Matérn correlation function and the different sampling schemes presented in the article are available at *Journal of the Royal Statistical Society: Series C* online. The code to replicate the results of the article is also available on GitHub (https://github.com/hriebl/gp-responses, https://github.com/hriebl/bamlssGP).

References

Adler R. J. (1990). *An introduction to continuity, extrema, and related topics for general gaussian processes*. Lecture Notes – Monograph Series, Vol. 12, Institute of Mathematical Statistics. http://www.jstor.org/stable/4355563.

Chan T., Hölttä T., Berninger F., Mäkinen H., Nöjd P., Mencuccini M., & Nikinmaa E. (2016). Separating water-potential induced swelling and shrinking from measured radial stem variations reveals a cambial growth and osmotic concentration signal. *Plant, Cell & Environment*, *39*(2), 233–244. https://doi.org/10.1111/pce.12541

Cressie N. A. C. (1993). *Statistics for spatial data*. Wiley.

Fahrmeir L., Kneib T., & Lang S. (2004). Penalized structured additive regression for space-time data: A Bayesian perspective. *Statistica Sinica*, *14*(3), 731–761.

Fahrmeir L., Kneib T., Lang S., & Marx S. (2013). *Regression: Models, methods and applications*. Springer.

Filippou P., Marra G., & Radice R. (2017). Penalized likelihood estimation of a trivariate additive probit model. *Biostatistics*, *18*(3), 569–585. https://doi.org/10.1093/biostatistics/kxx008

Fischer M., Bossdorf O., Gockel S., Hänsel F., Hemp A., Hessenmöller D., Korte G., Nieschulze J., Pfeiffer S., Prati D., Renner S., Schöning I., Schumacher U., Wells K., Buscot F., Kalko E. K. V., Linsenmair K. E., Schulze E.-D., & Weisser W. W. (2010). Implementing large-scale and long-term functional biodiversity research: The biodiversity exploratories. *Basic and Applied Ecology*, *11*(6), 473–485. https://doi.org/10.1016/j.baae.2010.07.009

Gamerman D. (1997). Sampling from the posterior distribution in generalized linear mixed models. *Statistics and Computing*, *7*(1), 57–68. https://doi.org/10.1023/A:1018509429360

Gneiting T. (2002). Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association*, *97*(458), 590–600. https://doi.org/10.1198/016214502760047113

Gneiting T. (2013). Strictly and non-strictly positive definite functions on spheres. *Bernoulli*, *19*(4), 1327–1349. https://doi.org/10.3150/12-BEJSP06

Greven S., & Scheipl F. (2017). A general framework for functional regression modelling. *Statistical Modelling*, *17*(1–2), 1–35. https://doi.org/10.1177/1471082X16681317

Klein N., & Kneib T. (2016a). Scale-dependent priors for variance parameters in structured additive distributional regression. *Bayesian Analysis*, *11*(4), 1071–1106. https://doi.org/10.1214/15-BA983

Klein N., & Kneib T. (2016b). Simultaneous inference in structured additive conditional copula regression models: A unifying Bayesian approach. *Statistics and Computing*, *26*(4), 841–860. https://doi.org/10.1007/s11222-015-9573-6

Klein N., Kneib T., Lang S., & Sohn A. (2015). Bayesian structured additive distributional regression with an application to regional income inequality in Germany. *The Annals of Applied Statistics*, *9*(2), 1024–1052. https://doi.org/10.1214/15-AOAS823

Klepper B., Browning V. D., & Taylor H. M. (1971). Stem diameter in relation to plant water status. *Plant Physiology*, *48*(6), 683–685. https://doi.org/10.1104/pp.48.6.683

Mencuccini M., Salmon Y., Mitchell P., Hölttä T., Choat B., Meir P., O'Grady A., Tissue D., Zweifel R., Sevanto S., & Pfautsch S. (2017). An empirical method that separates irreversible stem radial growth from bark water content changes in trees: Theory and case studies. *Plant, Cell & Environment*, *40*(2), 290–303. doi:10.1111/pce.12863

Metz J., & Ammer C. (2018). Dendrometer data of trees, neighbor, 1 MIP, 2012–2013. Biodiversity Exploratories Information System. Dataset. https://www.bexis.uni-jena.de/ddm/data/Showdata/17766?version=12.

Metz J., Annighöfer P., Westekemper K., Schall P., Schulze E.-D., & Ammer C. (2020). Less is more: Effects of competition reduction and facilitation on intra-annual (basal area) growth of mature European beech. *Trees*, *34*(1), 17–36. https://doi.org/10.1007/s00468-019-01894-7

Ostrowski A., Nieschulze J., Schulze E.-D., Fischer M., Ayasse M., Weisser W., & König-Ries B. (2016). Basic information and coordinates of field plots of the biodiversity exploratories project. Biodiversity Exploratories Information System. Dataset. https://www.bexis.uni-jena.de/ddm/data/Showdata/1000?version=7.

Paciorek C. J. (2003). *Nonstationary Gaussian processes for regression and spatial modelling* [Ph.D. thesis]. Carnegie Mellon University.

Pinheiro J., Bates D., DebRoy S., & Sarkar D., & R Core Team (2020). *nlme: Linear and nonlinear mixed effects models*. R package version 3.1-147. https://CRAN.R-project.org/package=nlme.

Rasmussen C. E., & Williams C. K. I. (2006). *Gaussian processes for machine learning*. MIT Press.

R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org.

Rigby R. A., & Stasinopoulos D. M. (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *54*(3), 507–554. doi:10.1111/j.1467-9876.2005.00510.x

Scheipl F., Staicu A.-M., & Greven S. (2015). Functional additive mixed models. *Journal of Computational and Graphical Statistics*, *24*(2), 477–501. https://doi.org/10.1080/10618600.2014.901914

Shi J. Q., & Choi T. (2011). *Gaussian process regression analysis for functional data*. CRC Press.

Umlauf N., Klein N., & Zeileis A. (2018). Bamlss: Bayesian additive models for location, scale, and shape (and beyond). *Journal of Computational and Graphical Statistics*, *27*(3), 612–627. https://doi.org/10.1080/10618600.2017.1407325

Wiese A., Christ M. M., Virnich O., Schurr U., & Walter A. (2007). Spatio-temporal leaf growth patterns of arabidopsis thaliana and evidence for sugar control of the diel leaf growth cycle. *New Phytologist*, *174*(4), 752–761. https://doi.org/10.1111/j.1469-8137.2007.02053.x

Wood S., & Scheipl F. (2020). *gamm4: Generalized additive mixed models using 'mgcv' and 'lme4'*. R package version 0.2-6. https://CRAN.R-project.org/package=gamm4.

Wood S. N. (2017). *Generalized additive models: An introduction with R* (2nd ed.), Chapman & Hall/CRC Texts in Statistical Science. CRC Press.

Yee T. W. (2015). *Vector generalized linear and additive models: With an implementation in R*. Springer.

Zweifel R., Sterck F., Braun S., Buchmann N., Eugster W., Gessler A., Häni M., Peters R. L., Walthert L., Wilhelm M., Ziemińska K., & Etzold S. (2021). Why trees grow at night. *New Phytologist*, *231*(6), 2174–2185. https://doi.org/10.1111/nph.17552

Zweifel R., Zimmermann L., & Newbery D. M. (2005). Modeling tree water deficit from microclimate: an approach to quantifying drought stress. *Tree Physiology*, *25*(2), 147–156. https://doi.org/10.1093/treephys/25.2.147