



# Density regression via Dirichlet process mixtures of normal structured additive regression models

María Xosé Rodríguez-Álvarez<sup>1,2</sup> · Vanda Inácio<sup>3</sup> · Nadja Klein<sup>4</sup>

Received: 23 May 2024 / Accepted: 12 January 2025  
© The Author(s) 2025

## Abstract

Within Bayesian nonparametrics, dependent Dirichlet process mixture models provide a flexible approach for conducting inference about the conditional density function. However, several formulations of this class make either restrictive modelling assumptions or involve intricate algorithms for posterior inference. We propose a flexible and computationally convenient approach for density regression based on a single-weights dependent Dirichlet process mixture of normal distributions model for univariate continuous responses. We assume an additive structure for the mean of each mixture component and incorporate the effects of continuous covariates through smooth functions. The key components of our modelling approach are penalised B-splines and their bivariate tensor product extension. Our method also seamlessly accommodates categorical covariates, linear effects of continuous covariates, varying coefficient terms, and random effects, which is why we refer our model as a Dirichlet process mixture of normal structured additive regression models. A notable feature of our method is the simplicity of posterior simulation using Gibbs sampling, as closed-form full conditional distributions for all model parameters are available. Results from a simulation study demonstrate that our approach successfully recovers the true conditional densities and other regression functionals in challenging scenarios. Applications to three real datasets further underpin the broad applicability of our method. An R package, *DDPstar*, implementing the proposed method is provided.

**Keywords** Bayesian nonparametrics · Bivariate smoothing · Conditional density · Gibbs sampling · Penalised B-splines

## 1 Introduction

In this article we are concerned about developing a flexible model for density regression that allows investigating how the distribution of a univariate continuous, real-valued, response variable  $y$  changes as a function of covariates  $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^p$ .

The literature on Bayesian flexible models for mean regression is vast and include, among others, (penalised) splines (Gustafson 2000; Lang and Brezger 2004), Gaussian processes (Williams and Rasmussen 2006, Chapter 2), neural

networks (Bishop 2006, Chapter 5), and additive regression trees (Chipman et al. 2010). These methods, although having the potential to approximate a wide range of regression functions, only allow for flexibility in the conditional mean of  $y$ , assuming that the scale and higher order moments are not affected by  $\mathbf{x}$ . Extensions to Bayesian location-scale regression models where both the mean and variance are covariate-dependent and flexibly modelled have been considered as well (e.g., Rodríguez and Martínez 2014; Pratola et al. 2020). When the goal is to extend beyond simply modelling specific functionals of the response distribution, such as the mean, variance, or even a quantile, through covariates, Bayesian distributional regression models (Klein et al. 2015) arise as a natural option. This class of models is able to provide a complete probabilistic characterisation of the response distribution by making a parametric assumption on the conditional density and by potentially relating each parameter of such a density to an additive predictor. However, choosing an appropriate parametric form for the conditional density may not be a trivial task in many applications. Furthermore, the underlying assumption that the same parametric distribution

✉ María Xosé Rodríguez-Álvarez  
mxrodriguez@uvigo.gal

<sup>1</sup> Departamento de Estatística e Investigación Operativa, Universidade de Vigo, 36310 Vigo, Spain

<sup>2</sup> CITMaga, Galician Centre for Mathematical Research and Technology, 15782 Santiago de Compostela, Spain

<sup>3</sup> School of Mathematics, University of Edinburgh, Edinburgh, Scotland, UK

<sup>4</sup> Scientific Computing Center, Karlsruhe Institute of Technology, Karlsruhe, Germany

family applies over the whole covariate space may be overly restrictive.

Within the Bayesian nonparametric literature, several methods relying on generalisations of mixture models for marginal density estimation have been proposed that provide flexible inference for conditional densities. This class of models does not require a specific parametric assumption about the conditional response distribution; it ‘only’ necessitates the specification of a family of distributions for the mixture components. A key advantage is that these methods can easily handle intricate distributional features, such as multimodality, skewness, and/or extreme variability, without the need to know of their existence in advance. Compared to regression models that focus on specific functionals, Bayesian nonparametric density regression yields coherent inference and allows to derive any quantity of interest, such as the conditional mean, variance, or quantiles.

In the context of marginal density estimation, Dirichlet process mixture models are a highly useful tool, as they combine the advantages of mixture modelling with the theoretical properties of nonparametric priors, such as full support and posterior consistency (see, e.g., Ghosal and Van der Vaart 2017, Chapter 7). These models are of the form

$$p(y | G) \equiv p(y) = \int k(y | \theta) dG(\theta), \quad G \sim \text{DP}(\alpha, G_0(\theta)), \quad (1)$$

where  $k(\cdot | \theta)$  is the density for the mixture kernel family of distributions with parameter  $\theta$  and the mixing distribution  $G$  follows a Dirichlet process (DP) prior (Ferguson 1973) with centring distribution  $E(G) = G_0(\theta)$  and precision parameter  $\alpha > 0$ . The stick-breaking representation of the DP (Sethuraman 1994), under which  $G$  can be written as an infinite weighted sum of point masses

$$G(\cdot) = \sum_{l=1}^{\infty} \omega_l \delta_{\theta_l}(\cdot), \quad \omega_l = \begin{cases} \eta_l, & \text{if } l = 1, \\ \eta_l \prod_{m < l} (1 - \eta_m), & \text{if } l > 1, \end{cases} \quad (2)$$

$$\eta_l \stackrel{\text{iid}}{\sim} \text{Beta}(1, \alpha), \quad \theta_l \stackrel{\text{iid}}{\sim} G_0,$$

with the collections  $\{\omega_l\}_{l \geq 1}$  and  $\{\theta_l\}_{l \geq 1}$  being independent of each other, further allows to write the density in (1) as a countable mixture of kernel densities

$$p(y) = \sum_{l=1}^{\infty} \omega_l k(y | \theta_l).$$

When covariate information is available, a model for the collection of conditional densities  $\{p(y | \mathbf{x}) : \mathbf{x} \in \mathcal{X} \subset \mathbb{R}^p\}$  can be obtained by allowing the mixing distribution  $G$  in (1) to depend on  $\mathbf{x}$ , i.e.,

$$p(y | G_{\mathbf{x}}) \equiv p(y | \mathbf{x}) = \int k(y | \theta) dG_{\mathbf{x}}(\theta).$$

One possibility for a prior on the collection of mixing distributions  $\{G_{\mathbf{x}} : \mathbf{x} \in \mathcal{X}\}$  is the dependent Dirichlet process (DDP) proposed by MacEachern (1999, 2000), which builds upon the stick-breaking representation of the DP in (2) and which in its full generality makes both the weights and the atoms dependent on covariates, that is,  $G_{\mathbf{x}}(\cdot) = \sum_{l=1}^{\infty} \omega_l(\mathbf{x}) \delta_{\theta_l(\mathbf{x})}(\cdot)$ . Here  $\{\eta_l(\mathbf{x})\}_{l \geq 1}$ , the inputs of the weights  $\omega_l(\mathbf{x})$  in  $G_{\mathbf{x}}$ , and  $\{\theta_l(\mathbf{x})\}_{l \geq 1}$ , the atoms of the mixture model, are collections of stochastic processes defined on  $\mathcal{X}$  and are independent of each other. The key idea behind the DDP construction is that at each  $\mathbf{x} \in \mathcal{X}$ ,  $G_{\mathbf{x}}$  is marginally a DP. An in-depth review of DDPs and related models is provided in Quintana et al. (2022) and Wade and Inacio (2025). In brief, most proposals in the literature fall into either: (i) models with covariate-dependent weights and atoms that are either constant across the values of  $\mathbf{x}$  or that rely on a linear formulation (see, among others, Griffin and Steel 2006; Dunson et al. 2007; Dunson and Park 2008; Chung and Dunson 2009; Antoniano-Villalobos et al. 2014; Rigon and Durante 2021, although we note that some of these correspond to a variation of MacEachern’s original DDP where, marginally,  $G_{\mathbf{x}}$  does not necessarily correspond to a DP), and (ii) models with covariate-dependent atoms but common (single) weights across the values of  $\mathbf{x}$ , i.e.,  $\omega_l(\mathbf{x}) = \omega_l$  (see, for instance, De Iorio et al. 2004; Gelfand et al. 2005; De la Cruz-Mesía et al. 2007; Rodriguez and Ter Horst 2008; De Iorio et al. 2009; Jara et al. 2010; Gutiérrez and Quintana 2011; Fronczyk and Kottas 2014; Xu et al. 2016, 2019, 2022).

Formulations that rely on covariate-dependent weights are very flexible but computations tend to be burdensome (see, e.g., Rigon and Durante 2021, Quintana et al. 2022, and Wade and Inacio 2025). An exception to the issue of cumbersome computations is the logit stick-breaking prior proposed by Rigon and Durante (2021), which is computationally tractable and posterior inference is available under several computational schemes. Additionally, the stick-breaking definition poses challenges in terms of the different choices that need to be made for functional shapes and hyperparameters when defining the weights’ inputs  $\eta_l(\mathbf{x})$ . These challenges are amplified by the lack of interpretation of the quantities involved (Antoniano-Villalobos et al. 2014, Wade and Inacio 2025).

In contrast, the single-weights DDP model is very popular, mainly due to the ‘ease’ of prior specification and its computational simplicity, as posterior simulation can be implemented using the same sampling algorithms available for DP mixtures. However, although such class of models, under its most general formulation, has desirable theoretical properties (Barrientos et al. 2012; Pati et al. 2013), it

may have limited flexibility in terms of the regression relationships it can capture (see more in Sect. 2). Nevertheless, quoting MacEachern (2000, p. 16): “*They (single-weights DDPs) thus provide a general framework that covers a vast territory*”.

In this article, we develop a single-weights DDP mixture of normal distributions model that overcomes the aforementioned lack of flexibility while retaining computational convenience. Specifically, we define the atoms of each normal component as  $\theta_l(\mathbf{x}) = (\mu_l(\mathbf{x}), \sigma_l^2)$  and assume an additive structure for  $\mu_l(\mathbf{x})$ , allowing the incorporation of different types of covariates and effects. Examples of effect types that can be accommodated include: (i) parametric effects of categorical covariates and linear or parametric effects of continuous covariates and their interactions, (ii) smooth nonlinear effects of univariate continuous covariates, (iii) bivariate smooth interactions, (iv) varying coefficient terms, and (v) random effects. For the specification of univariate and bivariate smooth functions, we use Bayesian penalised B-splines (P-splines, Eilers and Marx 1996; Lang and Brezger 2004). That is, smooth functions are represented by (the tensor-product of) B-spline basis functions, with a random walk prior placed on the basis coefficients to ensure smoothness. Additionally, for the other effect types, appropriate matching prior distributions are chosen for the respective vectors of regression coefficients. This predictor formulation corresponds to the one employed in the so-called structured additive regression models (STAR; Fahrmeir et al. 2022, Chapter 9). We therefore term our approach as DDP-star.

While previous works have explored integrating smooth effects into DP-based (or finite) mixture models using splines, this article expands on that research in three key ways. First, unlike prior contributions (e.g., Inácio de Carvalho et al. 2013, 2017; Inácio and Rodríguez-Álvarez 2022), we circumvent the critical and time-consuming task of selecting the optimal number (and location) of knots that define the spline basis by employing Bayesian P-splines (see more in Sect. 2). Second, in contrast to Chib and Greenberg (2010) and Wiesenfarth et al. (2014), who model covariate effects through a smooth additive predictor (based on splines) and the error distribution as a DP mixture—shifting the distribution of  $y$  by the smooth predictor—we greatly enhance flexibility, while still retaining interpretability, by additionally mixing over the regression coefficients, thus allowing for potentially different smooth covariate effects in the mean of each mixture component. Third, our approach permits complex and rich specifications for the components’ means, accommodating multiple covariates and effect types. This differs from previous research, such as that by Berrettini et al. (2023), which is constrained to using a single continuous covariate, thereby expanding the scope of possible models. Additionally, unlike previous approaches to STAR-based

distributional regression, our model does not require the specification of a parametric form for the conditional response distribution, allowing the distribution to adapt flexibly across the covariate space. All these advantages are achieved while ensuring that all parameters of the DDPstar model have conjugate full conditional distributions, enabling straightforward Gibbs sampling and eliminating the need for specialised techniques or tuning of Metropolis–Hastings steps. Furthermore, our method is implemented in the publicly available R package DDPstar, providing a user-friendly and versatile approach. The package provides tools for assessing model fit through posterior predictive checks and quantile residuals (Dunn and Smyth 1996). Additionally, it includes several information criteria—namely, the widely applicable information criterion (WAIC; Gelman et al. 2014), the log pseudomarginal likelihood (LPML; Geisser and Eddy 1979), and the deviance information criterion (DIC; Celeux et al. 2006)—to facilitate model and variable selection.

The rest of the paper is structured as follows. Section 2 starts with the basics of the single-weights DDP mixture of normals model before we detail our approach based on a structured additive predictor for the mean of each normal component and how to conduct posterior inference. The performance of our method is validated in Sect. 3 using simulated data under a variety of challenging scenarios. In Sect. 4 we apply our approach to data from toxicology, disease diagnosis, and agricultural studies. We conclude in Sect. 5 with a discussion. Details on posterior inference and extra results for both the simulation study and real data applications are available as Supplementary Material.

## 2 Dirichlet process mixture of normal structured additive regressions model

Our starting point is the single-weights DDP mixture of normal distributions model (De Iorio et al. 2009), which assumes that the conditional density function takes the following form

$$p(y | \mathbf{x}) = \int \phi(y | \mu, \sigma^2) dG_{\mathbf{x}}(\mu, \sigma^2), \text{ where} \\ G_{\mathbf{x}}(\cdot) = \sum_{l=1}^{\infty} \omega_l \delta_{(\mu_l(\mathbf{x}), \sigma_l^2)}(\cdot), \quad (3)$$

with the weights  $\{\omega_l\}_{l \geq 1}$  matching those from the standard stick-breaking representation in (2). For simplicity, in what follows we consider a single continuous covariate and model its effect on the mean of each mixture component linearly

$$\mu_l(x) = \beta_{l0} + \beta_{l1}x = \tilde{\mathbf{x}}^\top \boldsymbol{\beta}_l, \quad \tilde{\mathbf{x}} = (1, x)^\top, \\ \boldsymbol{\beta}_l = (\beta_{l0}, \beta_{l1})^\top, \quad l \geq 1,$$

which leads to a Dirichlet process mixture of normal linear (in the covariate’s effect) models

$$p(y | x) = \int \phi(y | \tilde{\mathbf{x}}^\top \boldsymbol{\beta}, \sigma^2) dG(\boldsymbol{\beta}, \sigma^2), \quad G \sim \text{DP}(\alpha, G_0(\boldsymbol{\beta}, \sigma^2))$$

$$= \sum_{l=1}^{\infty} \omega_l \phi(y | \tilde{\mathbf{x}}^\top \boldsymbol{\beta}_l, \sigma_l^2), \quad (\boldsymbol{\beta}_l, \sigma_l^2) \stackrel{\text{iid}}{\sim} G_0. \quad (4)$$

This model, by incorporating an infinite number of normal linear regression components, may seem very flexible. However, by assuming that the mixture weights are constant across  $x$ , the conditional density and its functionals are greatly restricted. For instance, the mean regression structure is linear (see, e.g., Dunson et al. 2007, p. 165). That is, this model is flexible in terms of non-Gaussian response, but not in terms of regression relationships. As a concrete example, Fig. 1, first row, shows the fit of model (4) to data generated from a homoscedastic simple normal regression with a nonlinear trend. As can be observed, both the regression function and several profiles of the conditional density are poorly recovered. As noted in Wade and Inacio (2025, p. 10), in single-weights DDPs, flexibility in the components-specific mean functions is key to achieve flexible, nonlinear mean regression structures. One possibility is to model the effect of continuous covariates within each mixture component through, for instance, B-spline basis functions (Fahrmeir et al. 2022, Chapter 8). In Fig. 1, second row, we show the results of fitting the model in (4) to the aforementioned simulated dataset but now considering a cubic B-spline basis expansion with ten interior knots. The model is now able to recover the true regression function and the profiles of the conditional density. However, it is widely acknowledged that the number (and location) of knots characterising the B-spline basis functions can significantly influence inferences. This is illustrated in the third and fourth rows of Fig. 1, which display the results using a cubic B-spline basis expansion with no interior knots and forty interior knots, respectively. It is evident that employing no interior knots lacks the necessary flexibility, whereas the use of 40 interior knots leads to overfitting, especially in the case of the regression function, and also to increased posterior uncertainty in the estimated density profiles. This example underscores the critical importance of selecting the ‘optimal’ number of knots.

One possible approach is to use a model selection criterion. This strategy, successfully applied by Inácio de Carvalho et al. (2013, 2017) and Inácio and Rodríguez-Álvarez (2022), is most effective when dealing with a single continuous covariate. However, when faced with multiple continuous covariates, it becomes, in principle, necessary to fit the model in (4) for every conceivable combination of the number of interior knots, a task that may prove impractical. Another possibility would be to place a prior distribution on the number of knots, e.g., extending the approach of DiMatteo et al. (2001), but this would require reversible jump Markov chain Monte Carlo techniques, which tend to be challenging to implement efficiently in practice.

In this work we favour the use of (Bayesian) penalised B-splines to model the effect of continuous covariates. This choice effectively circumvents the sensitivity of unpenalised approaches to the number of knots (see last row of Fig. 1), while preserving computational efficiency. The next sections are devoted to a detailed presentation of our proposal which is inspired by STAR models.

## 2.1 The DDPstar model

In its most general formulation, we write the mean of each mixture component in (3) as follows

$$\mu_l(\mathbf{x}) = \beta_{l0} + \sum_{a=1}^A f_{la}(\mathbf{v}_a), \quad (5)$$

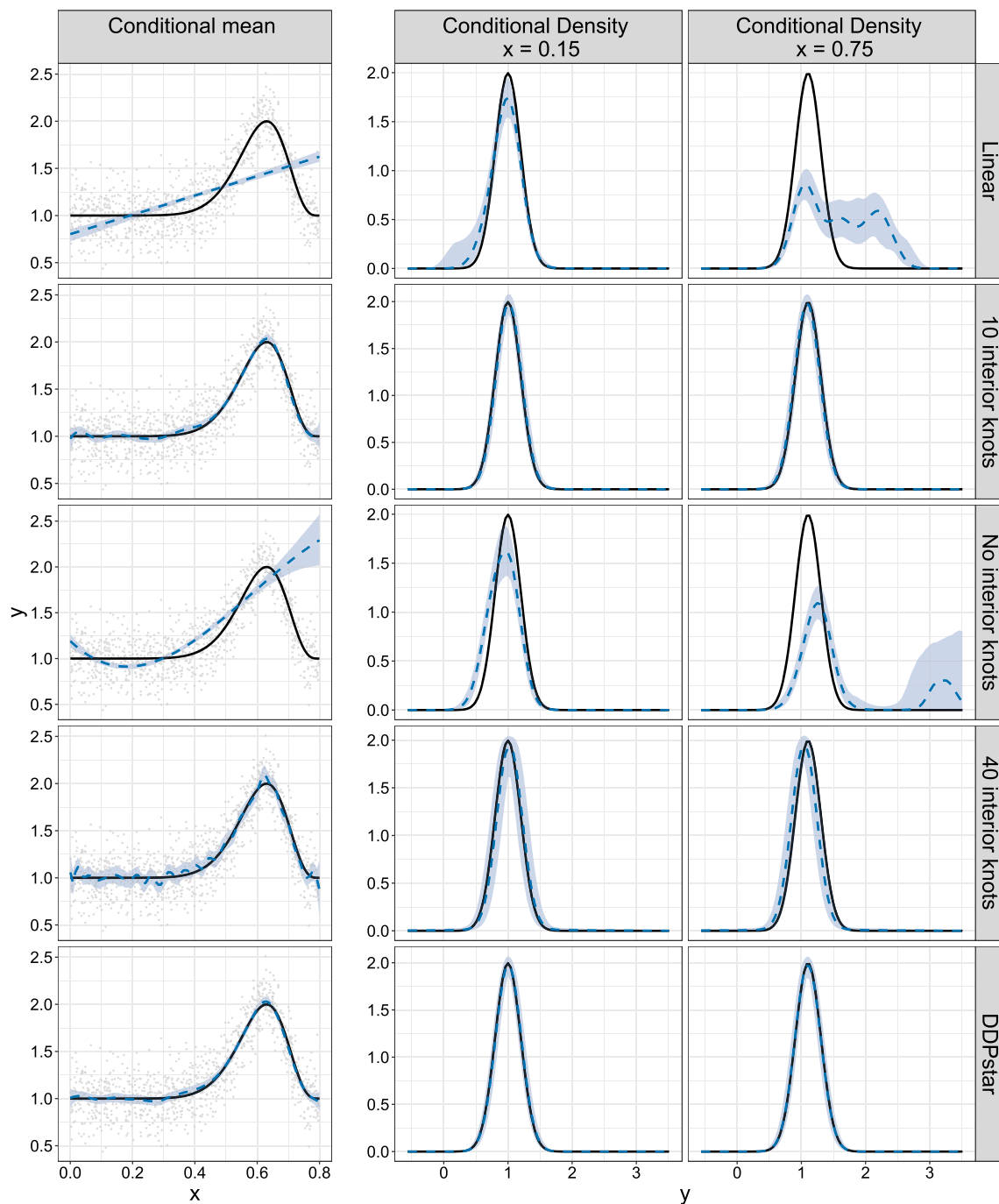
where  $\mathbf{v}_a$  denotes subsets of the  $p$  covariates in  $\mathbf{x}$  and  $f_{la}(\mathbf{v}_a)$  defines a generic representation of different types of functional effects depending on the covariate subset  $\mathbf{v}_a$ . As already noted, examples of effect types that can be accommodated into our framework include (i) parametric effects of categorical covariates and linear or parametric effects of continuous covariates and their interactions, (ii) smooth nonlinear effects of continuous covariates, (iii) bivariate smooth interactions or spatial effects, (iv) varying coefficient terms, and (v) random effects (see, e.g. Fahrmeir et al. 2022, Chapter 9). For example, for a linear effect of a continuous covariate, we have  $f_{la}(\mathbf{v}_a) = f_l^{\text{linear}}(v) = \beta_l v$ , where  $v$  is a univariate continuous covariate within the vector  $\mathbf{x}$ . In the case of nonlinear effects,  $f_{la}(\mathbf{v}_a) = f_l^{\text{smooth}}(v)$ , with  $f_l^{\text{smooth}}$  being a smooth univariate function. Of particular interest here are cases (ii) and (iii), which we thoroughly explain in the next sections, whereas cases (iv) and (v) are detailed in Supplementary Material A. For the sake of notational simplicity, we omit the specific function index ( $a$ ) in the subsequent discussion.

### 2.1.1 Smooth nonlinear effects

We start with the case where  $\mathbf{v}$  in (5) is constituted by a single continuous covariate, say  $v$ . To model smooth nonlinear effects, we consider the Bayesian analogue to P-splines (Eilers and Marx 1996) as introduced by Lang and Brezger (2004). In particular, the smooth function is approximated by a linear combination of  $J$  (cubic) B-spline basis functions on equidistant knots, i.e.,

$$f_l^{\text{smooth}}(v) = \sum_{j=1}^J \xi_{lj} B_j(v) = \mathbf{b}(v)^\top \boldsymbol{\xi}_l, \quad (6)$$

where  $\mathbf{b}(v) = (B_1(v), \dots, B_J(v))^\top$  is the vector of B-spline basis functions evaluated at  $v$  and  $\boldsymbol{\xi}_l = (\xi_{l1}, \dots, \xi_{lJ})^\top$  is the



**Fig. 1** Data were simulated from  $p(y | x) = \phi(y | \sin(2\pi x^3)^3 + 1, 0.2^2)$ , where  $x \sim U(0, 0.8)$ . Row 1: Linear DDP (LDDP). Row 2: B-splines LDDP with 10 internal knots at quantiles. Row 3: B-splines LDDP with no internal knots. Row 4: B-splines LDDP with 40 internal

knots at quantiles. Row 5: proposed DDPstar model with 40 equidistant internal knots. Sample size: 1000. In all cases, the infinite mixture was truncated to a finite number of components, with  $L = 20$  (details are given in Sect. 2.2)

vector of corresponding basis coefficients for the  $l$ th mixture component. P-splines rely on using a moderate to large number of basis functions, usually between 20 and 40, in combination with a penalty that enforces a smooth function estimation. Within a frequentist context, Eilers and Marx

(1996) proposed to penalise the squared  $q$ th order differences of adjacent basis coefficients. In the Bayesian framework, the  $q$ th order difference penalty is replaced by its stochastic analogue, i.e., a  $q$ th order random walk is used as a prior for the basis coefficients (Lang and Brezger 2004). The second-



order random walk –the most popular in the literature and our choice here– is defined by

$$\xi_{lj} = 2\xi_{l,j-1} - \xi_{l,j-2} + \epsilon_{lj}, \quad \epsilon_{lj} \sim N(0, \tau_l^2), \quad l \geq 1, \\ j = 3, \dots, J. \quad (7)$$

Usually,  $\xi_{l1}$  and  $\xi_{l2}$  are assigned noninformative priors, such that,  $p(\xi_{l1}) \propto \text{const}$  and  $p(\xi_{l2}) \propto \text{const}$ . The random walk prior distribution variance,  $\tau_l^2$ , controls the amount of smoothing, with small values corresponding to heavy smoothness and large values allowing considerable variation in the estimated smooth function. Note that we allow for a different amount of smoothing for each mixture component.

The second-order random walk prior in (7) induces the following joint Gaussian prior distribution for the vectors of basis coefficients  $\xi_l$

$$p(\xi_l | \tau_l^2) \propto |\mathbf{K}(\tau_l^2)|_+^{\frac{1}{2}} \exp\left(-\frac{1}{2}\xi_l^\top \mathbf{K}(\tau_l^2)\xi_l\right) \\ \times \left(\frac{1}{\tau_l^2}\right)^{\frac{\text{rank}(\mathbf{P})}{2}} \exp\left(-\frac{1}{2}\xi_l^\top \mathbf{K}(\tau_l^2)\xi_l\right), \quad (8)$$

where the precision matrix  $\mathbf{K}(\tau_l^2) = \frac{1}{\tau_l^2}\mathbf{P} \in \mathbb{R}^{J \times J}$ , with  $\mathbf{P} = \mathbf{D}^\top \mathbf{D}$  and  $\mathbf{D}$  being a second-order difference matrix, and  $|\mathbf{A}|_+$  denotes the pseudo-determinant of the matrix  $\mathbf{A}$  (i.e., the product of the non-zero eigenvalues). Because the penalty matrix  $\mathbf{P}$  is rank deficient,  $\text{rank}(\mathbf{P}) = J - 2$ , the prior in (8) is partially improper. This implies that there is a part of  $f_l^{\text{smooth}}$  that is not penalised by the prior precision matrix.

To better understand the unpenalised part of  $f_l^{\text{smooth}}$  as well as to make the mean function specification per component in (5) identifiable,  $f_l^{\text{smooth}}$  is decomposed into two parts: a penalised and an unpenalised part. There are different ways to obtain such a decomposition; we follow Currie et al. (2006) and use the eigendecomposition of the penalty matrix  $\mathbf{P}$ . Let  $\mathbf{P} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$  be the eigendecomposition of  $\mathbf{P}$ , where  $\mathbf{U}$  is the matrix of eigenvectors and  $\mathbf{\Lambda}$  is the diagonal matrix of eigenvalues (with eigenvalues arranged in order of increasing magnitude down the diagonal). Further denote by  $\mathbf{U}_+$  ( $\mathbf{\Lambda}_+$ ) and  $\mathbf{U}_0$  ( $\mathbf{\Lambda}_0$ ) the sub-matrices corresponding to the non-zero and zero eigenvalues, respectively, such that  $\mathbf{U} = [\mathbf{U}_0, \mathbf{U}_+]$  and  $\mathbf{\Lambda} = \text{blockdiag}(\mathbf{\Lambda}_0, \mathbf{\Lambda}_+)$ . As noted before, for second-order random walk priors there are two zero eigenvalues ( $\text{rank}(\mathbf{P}) = J - 2$ ). As such,  $\mathbf{\Lambda}_0$  is a  $2 \times 2$  matrix of zeroes, while  $\mathbf{\Lambda}_+$  is a full-rank,  $(J - 2) \times (J - 2)$ , diagonal matrix. It is easy to show that (6) can be reparameterised as

$$f_l^{\text{smooth}}(v) = \mathbf{b}(v)^\top \xi_l = \mathbf{b}(v)^\top \mathbf{U}\mathbf{U}^\top \xi_l = \mathbf{x}(v)^\top \beta_l + \mathbf{z}(v)^\top \gamma_l,$$

where

$$\mathbf{x}(v)^\top = \mathbf{b}(v)^\top \mathbf{U}_0, \quad \mathbf{z}(v)^\top = \mathbf{b}(v)^\top \mathbf{U}_+, \\ \beta_l = \mathbf{U}_0^\top \xi_l, \quad \gamma_l = \mathbf{U}_+^\top \xi_l.$$

We note that  $\beta_l^\top$  and  $\gamma_l^\top$  are vectors of length 2 and  $J - 2$ , respectively. Furthermore,  $\xi_l = [\mathbf{U}_0, \mathbf{U}_+] (\beta_l^\top, \gamma_l^\top)^\top$ , and thus the joint prior distribution in (8) can be rewritten, in terms of the new vector of coefficients,  $(\beta_l^\top, \gamma_l^\top)^\top$ , as

$$p\left((\beta_l^\top, \gamma_l^\top)^\top \mid \tau_l^2\right) \propto \left(\frac{1}{\tau_l^2}\right)^{\frac{\text{rank}(\tilde{\mathbf{P}})}{2}} \\ \times \exp\left(-\frac{1}{2\tau_l^2} (\beta_l^\top, \gamma_l^\top) \mathbf{\Lambda} (\beta_l^\top, \gamma_l^\top)^\top\right), \quad (9)$$

In other words, the joint prior distribution in (9) implies that  $\beta_l$  correspond to the unpenalised coefficients (with  $p(\beta_{l1}) \propto \text{const}$  and  $p(\beta_{l2}) \propto \text{const}$ ), while  $\gamma_l$  is the vector of penalised coefficients with proper Gaussian prior distribution given by

$$p(\gamma_l | \tau_l^2) \propto \left(\frac{1}{\tau_l^2}\right)^{\frac{J-2}{2}} \exp\left(-\frac{1}{2\tau_l^2} \gamma_l^\top \mathbf{\Lambda}_+ \gamma_l\right). \quad (10)$$

Note that although based on the eigendecomposition we have that  $\mathbf{x}(v)^\top = \mathbf{b}_l(v)^\top \mathbf{U}_0$ , this is equivalent to consider  $\mathbf{x}(v)^\top = (1, v)$  (see, e.g., Lee 2010). As such, in P-splines with second-order random walk priors, the space of functions that are not penalised corresponds to the polynomials of degree 1. This implies that when  $\tau_l^2 \rightarrow 0$ , the estimated function approaches a linear effect in that component. Moreover, this reparametrisation makes clear that the B-spline basis expansion of  $f_l^{\text{smooth}}$  in (6) includes an intercept (constant term). Given that there is already an intercept in the model for the mean of each component (see Eq. (5)), when constructing univariate smooth functions using P-splines, the intercept is removed to avoid identifiability issues (i.e., we consider  $\mathbf{x}(v)^\top = v$ ). Note that this approach also permits the incorporation of univariate smooth effects for multiple continuous covariates in (5), effectively circumventing identifiability issues.

## 2.1.2 Bivariate smooth surfaces

We now move onto the case where  $\mathbf{v}$  in (5) is constituted by two continuous covariates, say  $v_1$  and  $v_2$ , and we are interested in modelling a smooth bivariate surface jointly defined over  $v_1$  and  $v_2$ . In the case of a spatial effect,  $v_1$  and  $v_2$  typically represent coordinate information about the spatial location.

When extending the principles of P-splines to the bivariate case, one first approximates the smooth bivariate surface using the tensor-product of two marginal B-splines bases, i.e.,

$$\begin{aligned} f_l^{\text{bivariate}}(v_1, v_2) &= \sum_{j_1=1}^{J_1} \sum_{j_2=1}^{J_2} \xi_{lj_1j_2} B_{1j_1}(v_1) B_{2j_2}(v_2) \\ &= (\mathbf{b}_1(v_1)^\top \otimes \mathbf{b}_2(v_2)^\top) \boldsymbol{\xi}_l, \end{aligned}$$

where  $\mathbf{b}_1(v_1)$  and  $\mathbf{b}_2(v_2)$  are the vectors containing the B-spline basis functions evaluations,  $\otimes$  denotes the Kronecker product, and  $\boldsymbol{\xi}_l = (\xi_{l11}, \dots, \xi_{l1J_2}, \dots, \xi_{lJ_11}, \dots, \xi_{lJ_1J_2})^\top$  is the vector of coefficients. Smoothness is achieved by penalising (the sum of squares of) second-order coefficient differences along  $v_1$  and  $v_2$  (for details, see, Eilers and Marx 2003), which translates into the following partially improper Gaussian prior distribution

$$\begin{aligned} p(\boldsymbol{\xi}_l \mid \tilde{\tau}_{l1}^2, \tilde{\tau}_{l2}^2) &\propto |\mathbf{K}(\tilde{\tau}_{l1}^2, \tilde{\tau}_{l2}^2)|_+^{1/2} \\ &\times \exp\left(-\frac{1}{2} \boldsymbol{\xi}_l^\top \mathbf{K}(\tilde{\tau}_{l1}^2, \tilde{\tau}_{l2}^2) \boldsymbol{\xi}_l\right), \end{aligned} \quad (11)$$

where

$$\mathbf{K}(\tilde{\tau}_{l1}^2, \tilde{\tau}_{l2}^2) = \frac{1}{\tilde{\tau}_{l1}^2} (\mathbf{P}_1 \otimes \mathbf{I}_{J_2}) + \frac{1}{\tilde{\tau}_{l2}^2} (\mathbf{I}_{J_1} \otimes \mathbf{P}_2),$$

with  $\mathbf{P}_1 = \mathbf{D}_1^\top \mathbf{D}_1$  and  $\mathbf{P}_2 = \mathbf{D}_2^\top \mathbf{D}_2$ . We note that by using two prior variances,  $\tilde{\tau}_{l1}^2$  and  $\tilde{\tau}_{l2}^2$ , the prior distribution in (11) permits a different amount of smoothing for  $v_1$  and  $v_2$ .

As for the univariate case, the penalty matrix  $\mathbf{P} = (\mathbf{P}_1 \otimes \mathbf{I}_{J_2}) + (\mathbf{I}_{J_1} \otimes \mathbf{P}_2)$  is rank deficient, with  $\text{rank}(\mathbf{P}) = J_1 J_2 - 4$ . We proceed similarly and decompose the tensor-product P-spline smooth bivariate surface into a penalised and an unpenalised part using the eigendecomposition of the marginal penalties  $\mathbf{P}_1 = \mathbf{U}_1 \boldsymbol{\Upsilon}_1 \mathbf{U}_1^\top$  and  $\mathbf{P}_2 = \mathbf{U}_2 \boldsymbol{\Upsilon}_2 \mathbf{U}_2^\top$ . It can be shown that (for details, see Lee 2010)

$$\begin{aligned} f_l^{\text{bivariate}}(v_1, v_2) &= (\mathbf{b}_1(v_1)^\top \otimes \mathbf{b}_2(v_2)^\top) \boldsymbol{\xi}_l \\ &= (\mathbf{x}_1(v_1)^\top \otimes \mathbf{x}_2(v_2)^\top) \boldsymbol{\beta}_l \\ &\quad + (\mathbf{u}_1(v_1)^\top \otimes \mathbf{x}_2(v_2)^\top, \mathbf{x}_1(v_1)^\top \otimes \mathbf{u}_2(v_2)^\top, \\ &\quad \mathbf{u}_1(v_1)^\top \otimes \mathbf{u}_2(v_2)^\top) \boldsymbol{\gamma}_l \\ &\equiv \mathbf{x}(v_1, v_2)^\top \boldsymbol{\beta}_l + \mathbf{u}(v_1, v_2)^\top \boldsymbol{\gamma}_l, \end{aligned}$$

where the symbol  $\equiv$  is used to indicate that the design vectors in the second and third rows have the same elements but in a

different order,

$$\begin{aligned} \mathbf{x}_1(v_1)^\top &= (1, v_1), \quad \mathbf{u}_1(v_1)^\top = \mathbf{b}_1(v_1)^\top \mathbf{U}_{1+}, \\ \mathbf{x}_2(v_2)^\top &= (1, v_2), \quad \mathbf{u}_2(v_2)^\top = \mathbf{b}_2(v_2)^\top \mathbf{U}_{2+}, \end{aligned} \quad (12)$$

and

$$\begin{aligned} \mathbf{x}(v_1, v_2)^\top &= (1, v_1, v_2, v_1 v_2), \\ \mathbf{u}(v_1, v_2)^\top &= (\mathbf{u}_1(v_1)^\top, \mathbf{u}_2(v_2)^\top, \\ &\quad v_2 \mathbf{u}_1(v_1)^\top, v_1 \mathbf{u}_2(v_2)^\top, \\ &\quad \mathbf{u}_1(v_1)^\top \otimes \mathbf{u}_2(v_2)^\top). \end{aligned} \quad (13)$$

In this case,  $\boldsymbol{\beta}_l^\top$  and  $\boldsymbol{\gamma}_l^\top$  are vectors of length 4 and  $J_1 J_2 - 4$ , respectively, with  $\boldsymbol{\beta}_l$  corresponding to the unpenalised coefficients (with  $p(\beta_{lk}) \propto \text{const}, k = 1, \dots, 4$ ), and  $\boldsymbol{\gamma}_l$  to the penalised coefficients, with proper Gaussian prior distribution given by

$$p(\boldsymbol{\gamma}_l \mid \tilde{\tau}_{l1}^2, \tilde{\tau}_{l2}^2) \propto |\tilde{\mathbf{K}}(\tilde{\tau}_{l1}^2, \tilde{\tau}_{l2}^2)|^{1/2} \exp\left(-\frac{1}{2} \boldsymbol{\gamma}_l^\top \tilde{\mathbf{K}}(\tilde{\tau}_{l1}^2, \tilde{\tau}_{l2}^2) \boldsymbol{\gamma}_l\right), \quad (14)$$

where the precision matrix is

$$\begin{aligned} \tilde{\mathbf{K}}(\tilde{\tau}_{l1}^2, \tilde{\tau}_{l2}^2) &= \text{blockdiag}\left(\frac{1}{\tilde{\tau}_{l1}^2} \boldsymbol{\Upsilon}_{1+}, \right. \\ &\quad \frac{1}{\tilde{\tau}_{l2}^2} \boldsymbol{\Upsilon}_{2+}, \frac{1}{\tilde{\tau}_{l1}^2} \boldsymbol{\Upsilon}_{1+}, \frac{1}{\tilde{\tau}_{l2}^2} \boldsymbol{\Upsilon}_{2+}, \\ &\quad \frac{1}{\tilde{\tau}_{l1}^2} \boldsymbol{\Upsilon}_{1+} \otimes \mathbf{I}_{J_2-2} \\ &\quad \left. + \frac{1}{\tilde{\tau}_{l2}^2} \mathbf{I}_{J_1-2} \otimes \boldsymbol{\Upsilon}_{2+}\right). \end{aligned} \quad (15)$$

A technical note is in order here. For notational convenience and simplicity, in (12) we have considered  $\mathbf{x}_d(v_d)^\top = (1, v_d)$ ,  $d = 1, 2$ . However, to ensure that the precision matrix associated with  $\boldsymbol{\gamma}_l$  matches the one presented in (15), certain adjustments are required. One possibility is to substitute  $v_d$  with a centred version based on the covariate observations. Another, more comprehensive method is discussed in Wood et al. (2013, p. 346), and this is the one we have used in our implementation.

Note that  $\mathbf{u}(v_1, v_2)^\top$  in (13), related to the penalised part of the tensor-product P-spline smooth bivariate surface, is composed by 5 building blocks (subvectors), and so  $\boldsymbol{\gamma}_l^\top$  can be seen as the concatenation of five subvectors of penalised coefficients, i.e.,  $\boldsymbol{\gamma}_l^\top = (\boldsymbol{\gamma}_{l1}^\top, \boldsymbol{\gamma}_{l2}^\top, \boldsymbol{\gamma}_{l3}^\top, \boldsymbol{\gamma}_{l4}^\top, \boldsymbol{\gamma}_{l5}^\top)$ . In fact, each block in the precision matrix (15) corresponds to one of these coefficients subvectors. Moreover, the block structure of  $\mathbf{u}(v_1, v_2)^\top$  also leads to an interesting ANOVA-type

decomposition of the penalised part of the bivariate surface in five different smooth terms

$$\begin{aligned}
 f_l^{\text{bivariate}}(v_1, v_2) = & \beta_{l1} + \beta_{l2}v_1 + \beta_{l3}v_2 + \beta_{l4}v_1v_2 \\
 & + \underbrace{f_{l1}(v_1)}_{\mathbf{u}_1(v_1)^\top \boldsymbol{\gamma}_{l1}} + \underbrace{f_{l2}(v_2)}_{\mathbf{u}_2(v_2)^\top \boldsymbol{\gamma}_{l2}} \\
 & + \underbrace{v_2 h_{l1}(v_1)}_{v_2 \mathbf{u}_1(v_1)^\top \boldsymbol{\gamma}_{l3}} + \underbrace{v_1 h_{l2}(v_2)}_{v_1 \mathbf{u}_2(v_2)^\top \boldsymbol{\gamma}_{l4}} \\
 & + \underbrace{f_{l(1,2)}(v_1, v_2)}_{(\mathbf{u}_1(v_1)^\top \otimes \mathbf{u}_2(v_2)^\top) \boldsymbol{\gamma}_{l5}}. \quad (16)
 \end{aligned}$$

In other words, there are two main *pure* smooth effects along  $v_1$  and  $v_2$ ,  $f_{l1}(v_1)$  and  $f_{l2}(v_2)$ , two varying-coefficient terms,  $v_2 h_{l1}(v_1)$  and  $v_1 h_{l2}(v_2)$ , and a *pure* nonlinear-by-nonlinear interaction term,  $f_{l(1,2)}(v_1, v_2)$ . For further details and insights we refer the reader to Lee et al. (2013) and Rodríguez-Álvarez et al. (2018).

Examining the precision matrix (15) reveals that, despite the five smooth terms, only two prior variances control their smoothness,  $\tilde{\tau}_{l1}^2$  and  $\tilde{\tau}_{l2}^2$ , and the same prior variances apply to both main effects and interaction terms. Moreover, the last block in (15) depends on both prior variances. As discussed in Kneib et al. (2019) and Bach and Klein (2022), this has important implications for posterior simulation. The full conditional posterior distribution of  $\tilde{\tau}_{l1}^2$  and  $\tilde{\tau}_{l2}^2$  involves the determinant  $|\tilde{\mathbf{K}}(\tilde{\tau}_{l1}^2, \tilde{\tau}_{l2}^2)|$ , and this precludes the use of Gibbs sampling for posterior simulation. To circumvent this problem and enable posterior simulation through Gibbs sampling, we extend here the so-called PS-ANOVA model proposed by Lee et al. (2013) to the Bayesian setting (BPS-ANOVA). The idea underlying BPS-ANOVA is to use a different prior variance for each smooth term in (16), i.e., each block in the prior precision matrix for  $\boldsymbol{\gamma}_l$ , see Eq. (15), will have its own prior variance. That is, under the BPS-ANOVA approach, the penalised part of the tensor-product P-spline smooth bivariate surface is represented as the sum of five sets of mutually independent coefficients  $\boldsymbol{\gamma}_{lk}$  ( $k = 1, \dots, 5$ ), each with a proper Gaussian prior distribution and a different prior variance, say  $\tau_{lk}^2$  ( $k = 1, \dots, 5$ ). To make things concrete, we first need to introduce further notation. Let

$$\begin{aligned}
 \mathbf{z}_1(\mathbf{v}) &= \mathbf{u}_1(v_1), \quad \mathbf{z}_2(\mathbf{v}) = \mathbf{u}_2(v_2), \quad \mathbf{z}_3(\mathbf{v}) = v_2 \mathbf{u}_1(v_1), \\
 \mathbf{z}_4(\mathbf{v}) &= v_1 \mathbf{u}_2(v_2), \quad \mathbf{z}_5(\mathbf{v}) = \mathbf{u}_1(v_1) \otimes \mathbf{u}_2(v_2) \\
 &\quad \text{(see Eqs. (13) and (16)),}
 \end{aligned}$$

and

$$\begin{aligned}
 \Lambda_{1+} &= \Lambda_{3+} = \Upsilon_{1+}, \quad \Lambda_{2+} = \Lambda_{4+} = \Upsilon_{2+}, \\
 \Lambda_{5+} &= \Upsilon_{1+} \otimes \mathbf{I}_{J_2-2} + \mathbf{I}_{J_1-2} \otimes \Upsilon_{2+} \quad \text{(see Eq. (15)).}
 \end{aligned}$$

Therefore, the tensor-product P-spline smooth bivariate surface in Eq. (16) can be expressed, excluding the intercept,

as

$$f_l^{\text{bivariate}}(v_1, v_2) = \beta_{l2}v_1 + \beta_{l3}v_2 + \beta_{l4}v_1v_2 + \sum_{k=1}^5 \mathbf{z}_k(\mathbf{v})^\top \boldsymbol{\gamma}_{lk},$$

and, under the BPS-ANOVA model, the prior Gaussian distributions for  $\boldsymbol{\gamma}_{lk}$ ,  $k = 1, \dots, 5$ , are given by

$$p(\boldsymbol{\gamma}_{lk} | \tau_{lk}^2) \propto \left( \frac{1}{\tau_{lk}^2} \right)^{\frac{\text{rank}(\Lambda_{k+})}{2}} \exp \left( -\frac{1}{2\tau_{lk}^2} \boldsymbol{\gamma}_{lk}^\top \Lambda_{k+} \boldsymbol{\gamma}_{lk} \right). \quad (17)$$

## 2.2 Prior specification and posterior inference

Using the elements introduced in Sect. 2.1, we can compactly write the mean of each normal component as

$$\mu_l(\mathbf{x}) = \underbrace{\mathbf{x}^\top \boldsymbol{\beta}_l}_{\text{Parametric}} + \underbrace{\sum_{r=1}^R \mathbf{z}_r(\mathbf{v}_r)^\top \boldsymbol{\gamma}_{lr}}_{\text{Smooth/Random}}. \quad (18)$$

Here  $\mathbf{x}^\top \boldsymbol{\beta}_l$  contains all parametric effects, including, by default, the intercept, parametric effects of categorical covariates and linear or parametric effects of continuous covariates and their interactions, as well as, the unpenalised terms associated with the P-splines formulation, whereas  $\sum_{r=1}^R \mathbf{z}_r(\mathbf{v}_r)^\top \boldsymbol{\gamma}_{lr}$  contains the P-splines' penalised terms and the random effects (for details on random effects, see Supplementary Material A). Our model for the conditional density can therefore be written as

$$\begin{aligned}
 p(y | \mathbf{x}) &= \int \phi(y | \mathbf{x}^\top \boldsymbol{\beta} + \sum_{r=1}^R \mathbf{z}_r(\mathbf{v}_r)^\top \boldsymbol{\gamma}_r, \sigma^2) \\
 &\quad dG(\boldsymbol{\beta}, \boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_R, \sigma^2), \quad \text{with} \quad (19) \\
 G(\cdot) &= \sum_{l=1}^{\infty} \omega_l \delta_{(\boldsymbol{\beta}_l, \boldsymbol{\gamma}_{l1}, \dots, \boldsymbol{\gamma}_{lR}, \sigma_l^2)}(\cdot),
 \end{aligned}$$

where the weights  $\{\omega_l\}_{l \geq 1}$  match those from the standard stick-breaking representation in (2). The model is completed with the prior specification and we specify a conditionally conjugate centring distribution

$$\begin{aligned}
 \boldsymbol{\beta}_l | \mathbf{m}_\beta, \quad \mathbf{H}_\beta &\stackrel{\text{iid}}{\sim} N(\mathbf{m}_\beta, \mathbf{H}_\beta), \\
 \boldsymbol{\gamma}_{lr} | \tau_{lr}^2 &\stackrel{\text{iid}}{\sim} N(\mathbf{0}, \tau_{lr}^2 \mathbf{K}_r^{-1}), \quad \sigma_l^2 \stackrel{\text{iid}}{\sim} \text{IG}(a_{\sigma^2}, b_{\sigma^2}),
 \end{aligned}$$

with conjugate hyperpriors

$$(\mathbf{m}_\beta, \mathbf{H}_\beta^{-1}) \sim N(\mathbf{m}_0, \mathbf{H}_0) \times W(v, (v\psi)^{-1}),$$



$$\tau_{lr}^2 \stackrel{\text{iid}}{\sim} \text{IG}(a_{\tau^2}, b_{\tau^2}),$$

where  $\text{IG}(a, b)$  denotes an inverse gamma distribution with shape parameter  $a$  and scale parameter  $b$  and  $W(\nu, (\nu\Psi)^{-1})$  denotes a Wishart distribution with  $\nu$  degrees of freedom and expectation  $\Psi^{-1}$ . The specific forms of  $\mathbf{K}_r$  in the prior distribution for  $\boldsymbol{\gamma}_{lr}$  can be seen in (10) and (17) (and in Eqs. (A2)–(A4) of Supplementary Material A for the case of varying coefficient terms and random effects). The precision parameter  $\alpha$  (see Eq. (2)) can either be fixed or assigned a prior distribution. In this work, we adopt the latter approach and, for conjugacy reasons, let  $\alpha \sim \text{Gamma}(a_\alpha, b_\alpha)$ .

For posterior inference, we use the blocked Gibbs sampler of Ishwaran and James (2001), which relies on truncating the stick-breaking representation to a finite number of components, say  $L$ . Hence the mixing distribution  $G$  in (19) is replaced by  $G^L(\cdot) = \sum_{l=1}^L \omega_l \delta_{(\beta_l, \boldsymbol{\gamma}_{l1}, \dots, \boldsymbol{\gamma}_{lR}, \sigma_l^2)}(\cdot)$ , with  $\eta_L = 1$  to ensure that the weights sum up to one. We shall emphasise that  $L$  is not the exact number of mixture components expected to be observed, but rather an upper bound on it, as some of the components may not be occupied. We favour the use of the blocked Gibbs sampler, over other posterior sampling algorithms, due to its ease of implementation and because it allows for straightforward full posterior inference for general regression functionals.

We further note an interplay between  $L$  and the precision parameter  $\alpha$ . The value of  $L$  can be chosen by considering distributional properties of the tail of the stick-breaking weights,  $U_L = \sum_{l=L+1}^\infty \omega_l$ . Ishwaran and Zarepour (2000) showed  $E(U_L | \alpha) = \{\alpha/(\alpha + 1)\}^L$ . This expression can be averaged over the prior for  $\alpha$  to estimate  $E(U_L)$ . Given a tolerance level for the approximation, the equation can be solved numerically to obtain the corresponding value of  $L$ . Furthermore, denoting by  $n$  the sample size and by  $L^* (\leq L \leq n)$  the number of occupied mixture components, we have that  $E(L^* | \alpha) \approx \alpha \log((\alpha + n)/\alpha)$  and  $\text{var}(L^* | \alpha) \approx \alpha \{\log((\alpha + n)/\alpha) - 1\}$  (see, e.g., Liu 1996). Although expressions exist for the case where  $\alpha$  is assigned a  $\text{Gamma}(a_\alpha, b_\alpha)$  prior distribution (Kottas et al. 2005), these can be averaged over the prior to obtain  $E(L^*)$  and  $\text{var}(L^*)$ . This information can be useful for determining an appropriate value of  $L$ . For instance, it may be reasonable to set  $L > E(L^*) + 2\sqrt{\text{var}(L^*)}$ . In the posterior samples, it is also possible to check the maximum index of the number of occupied components. If this value is close to or equal to  $L$  for most iterations, the analysis should be repeated with a larger  $L$ . Supplementary Material D provides further guidance on other key decisions for our proposed approach (e.g., the choice of  $a_\alpha$  and  $b_\alpha$ ) and includes a recommended workflow for applying our method.

Upon the introduction of latent variables that identify the mixture component to which each observation belongs to,

the model for the data can be written hierarchically. From this hierarchical representation, it is straightforward to derive the full conditional distributions of all model parameters, which are available in closed form, thus allowing for ready posterior simulation through Gibbs sampling. All details are available in Supplementary Material B. Each posterior sample for  $\{(\omega_l, \boldsymbol{\beta}_l, \boldsymbol{\gamma}_{l1}, \dots, \boldsymbol{\gamma}_{lR}, \sigma_l^2)\}_{l=1}^L$  provides a posterior realisation for  $G^L$  directly through its definition and therefore for any point of interest, say  $(y_0, \mathbf{x}_0)$ , we can obtain posterior realisations of the predictive conditional density,  $p(y_0 | \mathbf{x}_0)$ . Moreover, posterior realisations of the predictive conditional mean, variance, and quantiles can also be easily obtained (detailed expressions are given in Supplementary Material C).

### 3 Simulation study

This section reports the results of a simulation study conducted to evaluate the empirical performance of the proposed model, DDPstar, for conducting inference about different functionals of the conditional response distribution, namely the conditional densities, which are our key inferential objects, the conditional mean and variance, and the three conditional quartiles. We do not include comparisons against other methods as our goal is not to claim superior performance of DDPstar over competing covariate-dependent Dirichlet process mixture models. Instead, we reiterate that our aim is to provide a computationally convenient density regression model that performs well across various scenarios, whose performance can be readily assessed through model fitting checks, and for which posterior inference is straightforward to implement. Additional simulations are included in Supplementary Material E, where we (1) examine the impact of the number of B-spline basis functions  $J$  and the choice of hyperparameter  $b_{\tau^2}$ , and (2) evaluate the computational time required by our approach in a model incorporating up to ten univariate smooth functions. Simulations are performed in the R environment (R Core Team 2023), using the R package DDPstar that accompanies this paper. Plots are generated using the R package ggplot2 (Wickham 2016).

#### 3.1 Simulation scenarios and implementation details

We explore a wide range of simulation scenarios, including cases with a single continuous covariate as well as two continuous covariates. For the latter, both additive and interaction structures are examined. For each scenario, 100 datasets are generated with sample sizes  $n \in \{300, 500, 1000\}$ . The scenarios considered are as follows.

**Scenario I.** Heteroscedastic nonlinear normal regression

$$y_i | x_i \stackrel{\text{ind.}}{\sim} N\left(x_i - 4x_i^3, (1 + x_i)^2\right), \quad x_i \stackrel{\text{iid}}{\sim} U(0, 1).$$

**Scenario II.** Mixture of nonlinear non-normal regressions

$$y_i | x_i \stackrel{\text{ind.}}{\sim} 0.5 \text{SN}(\xi = x_i^2, \omega = 0.25, \alpha = 2) \\ + 0.5 t(\mu = \sin(\pi x_i), \sigma = 0.25, \nu = 5),$$

where  $x_i \stackrel{\text{iid}}{\sim} U(0, 1)$ ,  $\text{SN}(\xi, \omega, \alpha)$  denotes a skew normal distribution with location  $\xi$ , scale  $\omega$ , and skewness parameter  $\alpha$ , and  $t(\mu, \sigma, \nu)$  denotes a (scaled and shifted)  $t$  distribution with mean  $\mu$ , scale parameter  $\sigma$ , and degrees of freedom  $\nu$ .

**Scenario III.** Covariate-dependent mixture of normal homoscedastic regressions

$$y_i | x_i \stackrel{\text{ind.}}{\sim} \exp(-2x_i) N(x_i, 0.01) \\ + \{1 - \exp(-2x_i)\} N(x_i^4, 0.04), \quad x_i \stackrel{\text{iid}}{\sim} U(0, 1).$$

This scenario has been considered in previous studies, including Dunson et al. (2007) and Dunson and Park (2008), and it poses a challenge for our model due to our assumption of covariate-independent weights.

**Scenario IV.** Nonstationary mean function and covariate-dependent variance

$$y_i | x_i \stackrel{\text{ind.}}{\sim} N(m(x_i), \sigma^2(x_i)), \\ m(x_i) = \begin{cases} 0, & \text{if } -2 \leq x_i \leq 2, \\ 2x_i - 4, & \text{if } 2 < x_i \leq 5, \\ 6, & \text{if } 5 < x_i \leq 10, \end{cases} \\ \sigma^2(x_i) = \begin{cases} 0.2^2, & \text{if } -2 \leq x_i \leq 2, \\ 0.05^2, & \text{if } 2 < x_i \leq 5, \\ (x_i - 5)^2/15 + 0.01, & \text{if } 5 < x_i \leq 10, \end{cases}$$

and  $x_i \stackrel{\text{iid}}{\sim} U(-2, 10)$ . This scenario has been used by Wade and Inacio (2025) to illustrate drawbacks of single-weights DDPs.

**Scenario V.** Mixture of nonlinear normal regressions (two continuous covariates)

$$y_i | x_{1i}, x_{2i} \stackrel{\text{ind.}}{\sim} 0.5 N(g_{11}(x_{1i}) + g_{12}(x_{2i}), 0.25^2) \\ + 0.5 N(\mu = g_{21}(x_{1i}) + g_{22}(x_{2i}), 0.25^2).$$

**Scenario VI.** Mixture of nonlinear non-normal regressions (two continuous covariates)

$$y_i | x_{1i}, x_{2i} \stackrel{\text{ind.}}{\sim} 0.5 \text{SN}(\xi = g_{11}(x_{1i}) + g_{12}(x_{2i}), \omega = 0.25, \alpha = 2) \\ + 0.5 t(\mu = g_{21}(x_{1i}) + g_{22}(x_{2i}), \sigma = 0.25, \nu = 5).$$

For Scenarios V and VI, we consider  $x_{1i}, x_{2i} \stackrel{\text{iid}}{\sim} U(0, 1)$ ,  $g_{11}(x_1) = (x_1 - 0.5)^2$ ,  $g_{12}(x_2) = (x_2 - 0.5)^2$ ,  $g_{21}(x_1) = \exp(x_1) \sin(13(x_1 - 0.6)^2)$ , and  $g_{22}(x_2) = \exp(-x_2) \sin(7x_2)$ . Note that these two scenarios correspond to the case of two continuous covariates and an additive structure.

**Scenario VII.** Mixture of nonlinear normal regressions (interaction between two continuous covariates)

$$y_i | x_{1i}, x_{2i} \stackrel{\text{ind.}}{\sim} 0.5 N(g_1(x_{1i}, x_{2i}), 0.25^2) \\ + 0.5 N(g_2(x_{1i}, x_{2i}), 0.25^2).$$

**Scenario VIII.** Mixture of nonlinear non-normal regressions (interaction between two continuous covariates)

$$y_i | x_{1i}, x_{2i} \stackrel{\text{ind.}}{\sim} 0.5 \text{SN}(\xi = g_1(x_{1i}, x_{2i}), \omega = 0.25, \alpha = 2) \\ + 0.5 t(\mu = g_2(x_{1i}, x_{2i}), \sigma = 0.25, \nu = 5).$$

For Scenarios VII and VIII, we consider  $x_{1i}, x_{2i} \stackrel{\text{iid}}{\sim} U(0, 1)$ ,  $g_1(x_1, x_2) = \cos\left(2\pi\sqrt{(x_1 - 0.5)^2 + (x_2 - 0.5)^2}\right)$ , and  $g_2(x_1, x_2) = 1.9 \exp(x_1) \sin(13(x_1 - 0.6)^2) \exp(-x_2) \sin(7x_2)$ . Both scenarios involve two continuous covariates with and an interaction structure.

For scenarios involving a single continuous covariate, Scenarios I–IV, the mean of each component is specified as  $\mu_l(x) = f_l(x)$ , and  $f_l$  is approximated using a cubic B-spline basis of dimension  $J = 23$ . For Scenarios V and VI we consider  $\mu_l(x_1, x_2) = f_{l1}(x_1) + f_{l2}(x_2)$ , whereas in Scenarios VII and VIII we consider  $\mu_l(x_1, x_2) = f_l(x_1, x_2)$  and represent the bivariate function using the tensor product of two marginal cubic B-splines bases; in both cases  $J_1 = J_2 = 23$ . In all cases, we use second-order random walks.

To facilitate prior specification, we standardise the responses and covariates and use the prior distributions outlined in Sect. 2.2. Following Lang and Brezger (2004), for the penalised or nonlinear part of our P-splines specification, we set  $a_{\tau^2} = 1$  and  $b_{\tau^2} = 0.005$ , which is a common default choice and has demonstrated good performance in the simulation study presented in Supplementary Material E. For  $\beta_l$ , the vector containing the coefficients associated to the unpenalised or linear terms, we use  $\mathbf{m}_0 = \mathbf{0}_Q$ ,  $\mathbf{H}_0 = 10I_Q$ ,  $\nu = Q + 2$ , and  $\Psi = I_Q$ , where  $Q$  is the length of vector  $\beta_l$  (including the intercept). Regarding the prior for  $\sigma_l^2$ , we choose  $a_{\sigma^2} = 2$  and  $b_{\sigma^2} = 0.5$ . Selecting  $a_{\sigma^2} = 2$  results in a prior distribution with infinite variance centred around a finite mean ( $b_{\sigma^2} = 0.5$ ), thereby favouring within-component variances smaller than one. Given that responses are standardised, and thus have a marginal variance of one, it is reasonable to expect the within-component variances to be smaller than the marginal variance. We further set  $a_\alpha = b_\alpha = 2$  and  $L = 20$ , resulting in a truncation error of  $E(U_{20}) \approx 0.0002$ . For the largest sample size ( $n = 1000$ ),

these choices yield an expected number of occupied components of 7 (with a standard deviation of 3), suggesting that  $L = 20$  is appropriate a priori (see Sect. 2.2). Lastly, for each scenario and simulated dataset, posterior inference is based on 6000 samples, following the discard of a 2000 iteration burn-in period of the Gibbs sampler.

### 3.2 Results

For brevity, most graphical results are provided in Supplementary Material E, while we focus here on the main findings. For Scenarios I to IV, Fig. 2 presents the average of the posterior medians of the predictive conditional density over the 100 simulated datasets, for  $x \in \{0.1, 0.2, 0.4, 0.6, 0.8, 0.9\}$ , along with the pointwise 2.5% and 97.5% quantiles of the ensemble of these posterior medians. Results for Scenarios V to VIII are shown in Fig. 3. For these four scenarios, which involve two continuous covariates, we consider  $(x_1, x_2) \in \{(0.1, 0.1), (0.4, 0.6), (0.6, 0.6), (0.6, 0.4), (0.8, 0.2), (0.9, 0.9)\}$ .

We begin with the results for Scenarios I to III, which all involve a single continuous covariate. Scenario IV, also involving a single continuous covariate, is discussed afterwards. We first note that, as depicted in Fig. 2, the conditional densities vary in shape depending on the covariate value, exhibiting features such as one or two modes and differing degrees of asymmetry. Similar patterns are observed for the other scenarios, as illustrated in Fig. 3. Figure 2 shows that our model effectively reconstructs the true profiles of the conditional densities, particularly in Scenarios I and II. In Scenario III, which features covariate-dependent weights, the reconstruction of the profiles of the conditional density is slightly less accurate than in Scenarios I and II, but remains highly satisfactory. As expected, across all three scenarios and for all covariates values, the estimates of the conditional density profiles become increasingly concentrated around the true conditional density profiles as the sample size grows. The results for the conditional mean, conditional variance, and the three conditional quartiles are presented in Web Figures 1, 6, and 10 of the Supplementary Material, respectively. These figures show that our proposed model successfully recovers the true shape of these functionals. However, some discrepancies are observed in the estimates of the conditional variance at the boundaries, particularly for smaller sample sizes. Notably, in our model, the variance of each component is independent of the covariates. Nevertheless, the variance of the mixture model as a whole depends on covariates indirectly, through their effect on the mean of each component (see Supplementary Material C). As such, accurately estimating the conditional variance poses a challenge to our approach. We now turn to Scenario IV, which involves a nonstationary mean function and a variance function that is nearly zero for all covariate values between  $-2$  and  $5$ , and

beyond this range, it has a quadratic form that increases quite rapidly. As shown in Fig. 2, some conditional density profiles, particularly the most peaked ones, are not well recovered. The conditional mean function, shown in Web Figure 1 of the Supplementary Material, is reasonably well recovered. However, the true functional shapes of the variance function and the conditional quartiles are not accurately reconstructed, as shown in Web Figures 6 and 10. In this scenario, flexibility in the mean function alone is insufficient to capture the covariate-dependent variance, which also affects the estimation of other functionals. It is important to note that model checking procedures can be used to evaluate the performance of our DDPstar model. This is illustrated in Web Figures 17 and 18 of the Supplementary Material, which show posterior predictive checks and quantile residuals, respectively, for a single generated dataset under Scenario IV. As these figures reveal, the DDPstar is not suitable for this particular scenario.

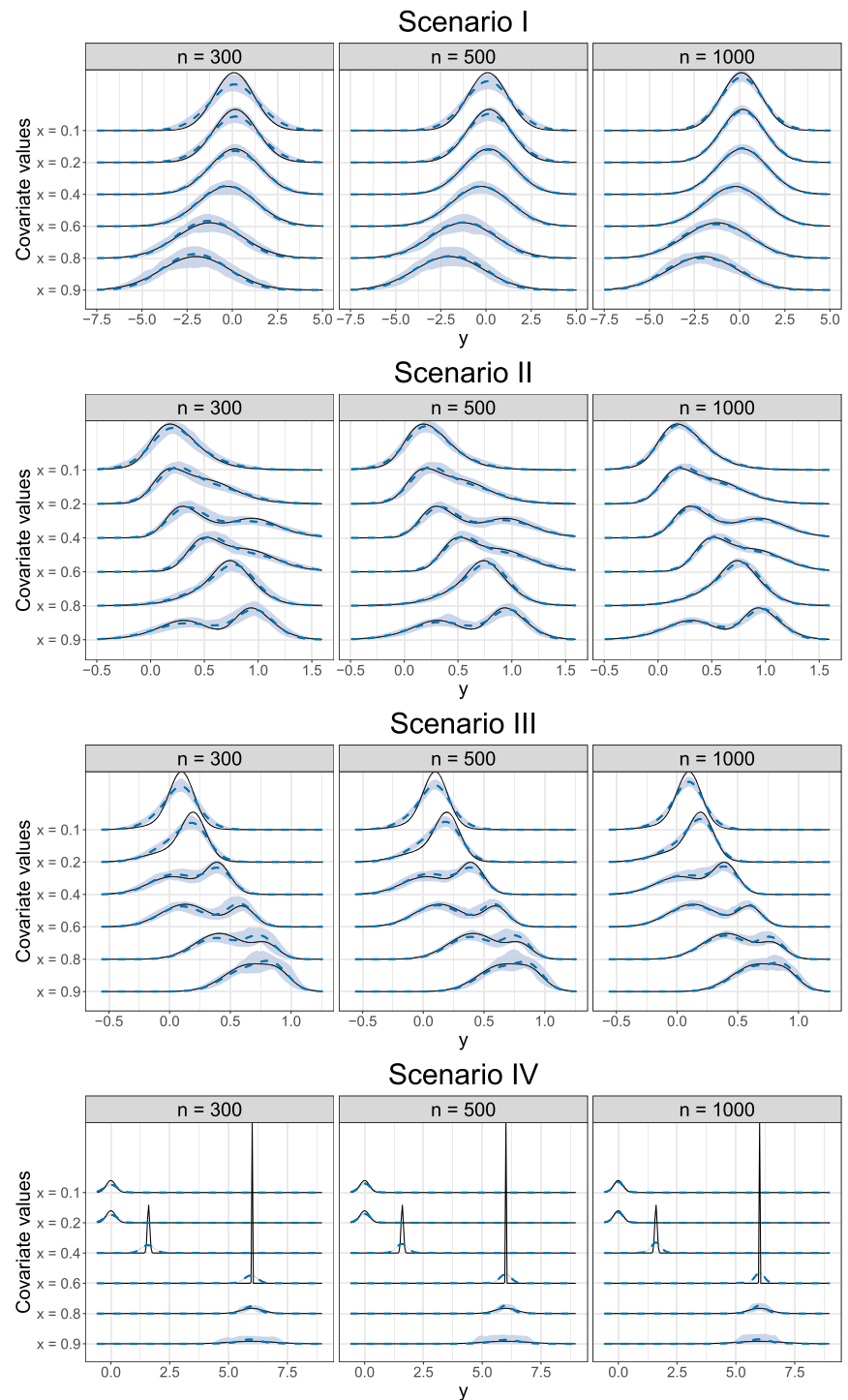
We now turn to scenarios involving two continuous covariates, Scenarios V–VIII. The results shown in Fig. 3 indicate that our model successfully recovers the true conditional densities. The performance is slightly better for Scenarios V and VII, which involve a mixture of two normal regressions and more closely align with our model specification. Results for the other functionals are presented in Web Figures 2, 3 and 4 (conditional mean), Web Figures 7, 8 and 9 (conditional variance), and Web Figures 11, 12, 13, 14, 15 and 16 (conditional quartiles) of the Supplementary Material. These figures demonstrate that our method consistently performs satisfactorily in reconstructing the true functionals for these scenarios. In Scenarios VII and VIII, which involve an interaction structure, the influence of sample size is more pronounced, requiring larger sample sizes to achieve satisfactory results. This is expected, as modelling interactions between smooth functions of continuous covariates generally demands substantial data.

## 4 Real data illustrations

This section demonstrates the broad applicability of the DDPstar model by applying it to three studies in different fields: toxicology, disease diagnosis, and agriculture. For each of these applications, comparative assessments are provided against alternative methods that have been applied to the same datasets. We employ, for the standardised data, the same hyperparameter values used in the Simulation Study (see Sect. 3.1). In all cases, the following workflow is followed:

1. The model is fitted.
2. Convergence checks are performed on label-invariant quantities (with the conditional density being the natural

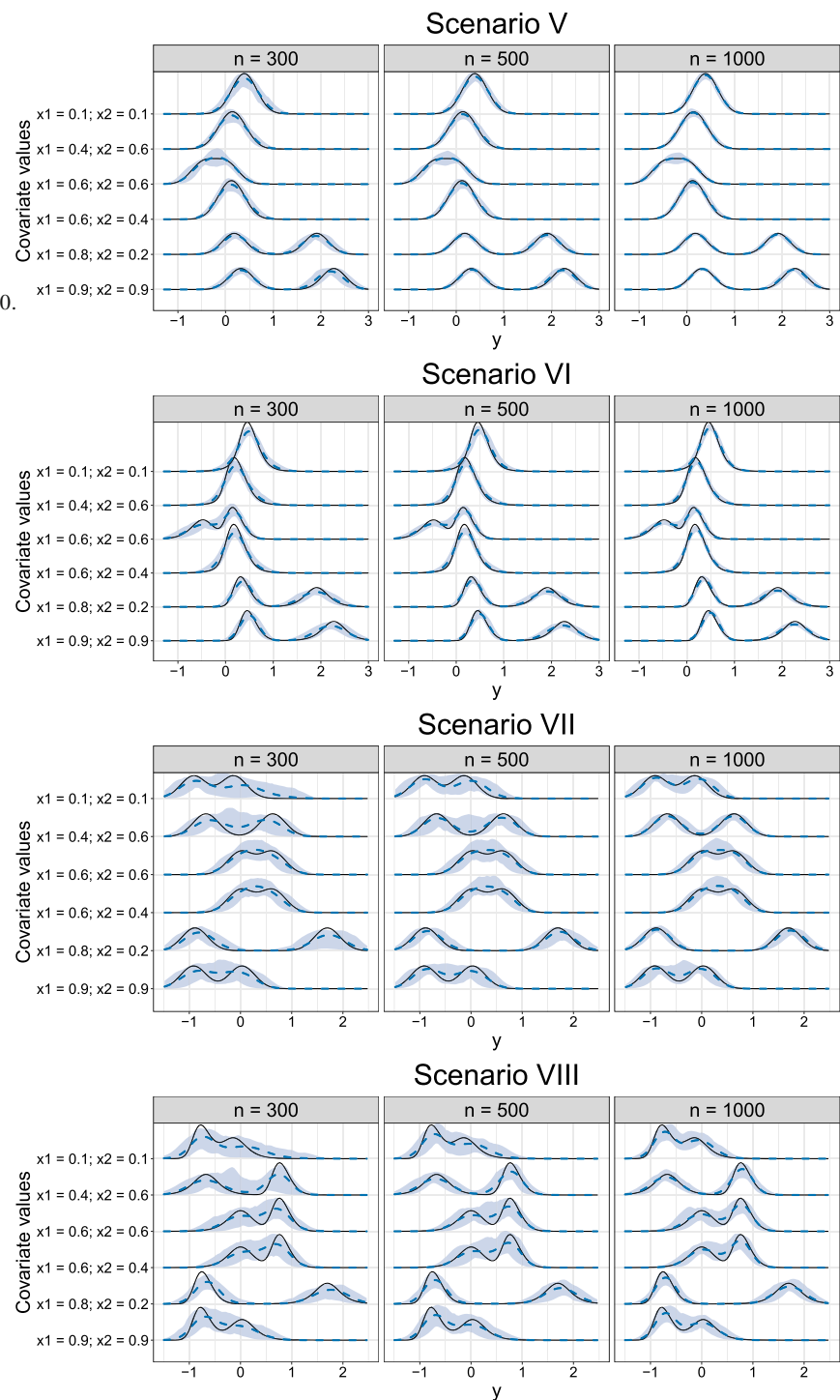
**Fig. 2** For the simulation study and scenarios involving one continuous covariate (Scenarios I, II, III and IV): True (solid black line) and average across the 100 simulated datasets (blue dashed lines) of the posterior median of the conditional density for  $x \in \{0.1, 0.2, 0.4, 0.6, 0.8, 0.9\}$  and different sample sizes ( $n$ ). The shaded areas are bands constructed using the pointwise 2.5% and 97.5% quantiles of the 100 posterior medians



- choice in most applications). These checks are conducted through the inspection of traceplots and Geweke statistics.
3. The effective sample size of label-invariant quantities is inspected.
  4. Model checking is conducted via quantile residuals and posterior predictive checks.

For brevity, only some of these results are presented here. However, all necessary R code and datasets to reproduce the analyses are provided in the Supplementary Materials. Posterior inference for all three applications is based on 20000 iterations, with the first 5000 discarded as burn-in. We provide the computational times obtained using a computer equipped with an Intel® Core™ i9–14900×32 processor, 64

**Fig. 3** For the simulation study and scenarios involving two continuous covariates and an additive (Scenarios V and VI) or an interaction structure (Scenarios VII and VIII): True (solid black lines) and average across the 100 simulated datasets (blue dashed lines) of the posterior median of the conditional density for  $(x_1, x_2) \in \{(0.1, 0.1), (0.4, 0.6), (0.6, 0.6), (0.6, 0.4), (0.8, 0.2), (0.9, 0.9)\}$  and different sample sizes ( $n$ ). The shaded areas are bands constructed using the pointwise 2.5% and 97.5% quantiles of the 100 posterior medians



GiB of RAM, and running the Ubuntu 22.04 LTS operating system.

#### 4.1 Toxicology study

This first study is rooted in epidemiology and has become a benchmark for validating Bayesian nonparametric density regression models (see, e.g., Dunson and Park 2008; Canale

et al. 2018; Rigon and Durante 2021). Its primary objective is to explore the relationship between the concentration of Dichlorodiphenyldichloroethylene (DDE) in maternal serum during the third trimester of pregnancy and the risk of premature delivery (Longnecker et al. 2001). DDE is a metabolite of the pesticide DDT. Despite its potential adverse health effects, DDT is still used to combat malaria-transmitting mosquitoes in areas where malaria is endemic. The dataset



consists of 2312 women and the response variable is the gestational age at delivery (GAD), in weeks, and deliveries prior to 37 weeks of gestation are considered as preterm. Interest lies in investigating how the GAD distribution changes with increasing DDE levels, with a particular focus placed on the left tail of the distribution in order to assess the DDE effect on preterm deliveries.

We apply the DDPstar model to the data  $(y_i, x_i) = (\text{GAD}_i, \text{DDE}_i)$ , for  $i = 1, \dots, 2312$ , and consider the following model for the mean of each mixture component (for simplicity, the dependence on the component is omitted)

$$\mu(x_i) = \beta_0 + f^{\text{smooth}}(\text{DDE}_i),$$

where the smooth effect of DDE is approximated using a cubic B-spline basis of dimension  $J = 23$  (we exclude the intercept). Traceplots and Geweke statistics for the conditional density, mean and variance do not show any lack of convergence of the MCMC chains and effective sample sizes are satisfactory (for details, see the accompanying R code). Posterior predictive checks and quantile residuals are provided in Web Figures 29 and 30 in the Supplementary Material, respectively, and confirm an accurate fit to the data.

Figure 4 shows the scatterplot of the data along with the regression function (posterior median) and its pointwise 95% credible band. As can be observed, there is a slight decrease in the mean GAD as the DDE levels increase, although for DDE levels above 100 mg/L posterior uncertainty is quite substantial due to data scarcity. In the same figure, we also show the variance function and the three conditional quartiles. Figure 5 presents ‘conditional histograms’ for a range of DDE levels, with the estimated conditional densities superimposed. Similarly to Dunson and Park (2008) and Rigon and Durante (2021), we consider the 0.1, 0.6, 0.9, and 0.99 quantiles of the DDE. Following Rigon and Durante (2021), the conditional histograms are obtained by grouping the GAD values according to a binning of the DDE with cutoffs at the central values of subsequent quantiles. The estimated conditional densities closely follow the histogram, further indicating a good fit to the data. This figure suggests an increasing left tail of the GAD distribution, which is associated with preterm deliveries, as DDE levels increase. In addition, in Fig. 5 we also display the posterior inferences obtained using the approach of Rigon and Durante (2021), which allows for covariate-dependent weights. As is apparent, the posterior medians of the two methods are almost indistinguishable. For the 0.9 and 0.99 quantiles of the DDE, posterior uncertainty is notably higher with the DDPstar approach, especially for the 0.99 quantile. This is attributed to: (i) sparsity of the data in this region of the covariate space, and (ii) our model using a significantly larger number of parameters compared to that of Rigon and Durante (2021). To explore how to use the conditional density to infer conditional preterm

probabilities, and following what previous authors did, we compute the following four covariate-specific probabilities:  $\Pr(\text{GAD} \leq y^* \mid \text{DDE})$ , for  $y^* \in \{33, 35, 37, 40\}$ . Results for both our model and the one from Rigon and Durante (2021) are depicted in Web Figure 31 in the Supplementary Material. Once again, apart from some relatively minor differences at extreme DDE levels, the two approaches yield similar results. Finally, in terms of computing times, both approaches also perform similarly: for 20,000 Gibbs iterations, our approach takes approximately 85 s, compared to 80 s for the method proposed by Rigon and Durante (2021).

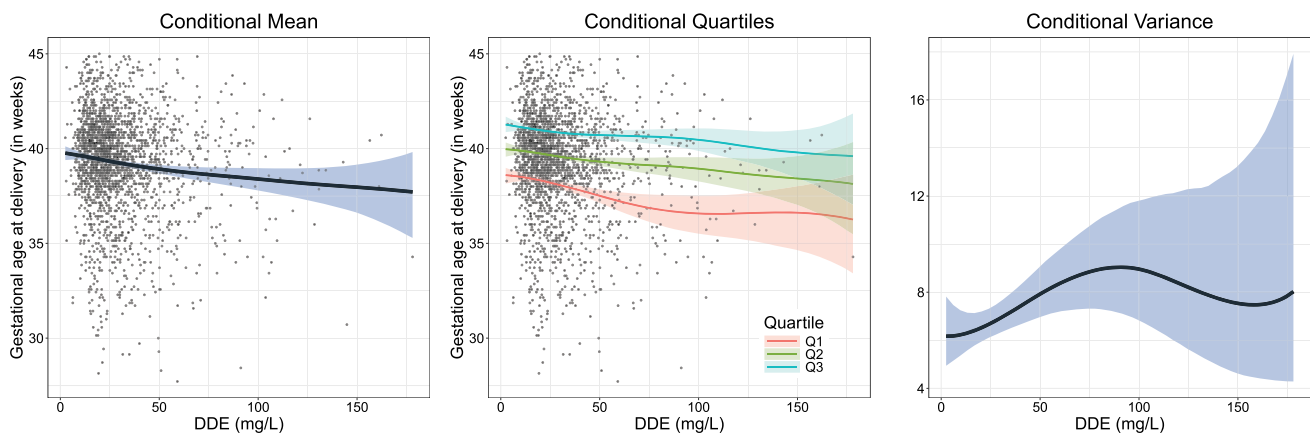
## 4.2 Disease diagnosis study

The receiver operating characteristic (ROC) curve stands as the preeminent tool in evaluating the ability of continuous-outcome diagnostic tests to distinguish between individuals with and without a well-defined condition (disease). It is recognised that in many instances, the test outcome and potentially its discriminatory capacity can be affected by covariates (e.g., age, gender, variations in testing conditions). In such cases, it is advisable to use the conditional or covariate-specific ROC curve, which is an ROC curve that is tailored to a specific covariate value, providing an assessment of the accuracy of the test within the particular subgroup defined by that value (see Inácio et al. 2021 and references therein). The covariate-specific ROC curve is defined as  $\{(t, \text{ROC}(t \mid \mathbf{x}))\}$ ,  $t \in [0, 1]$ , where

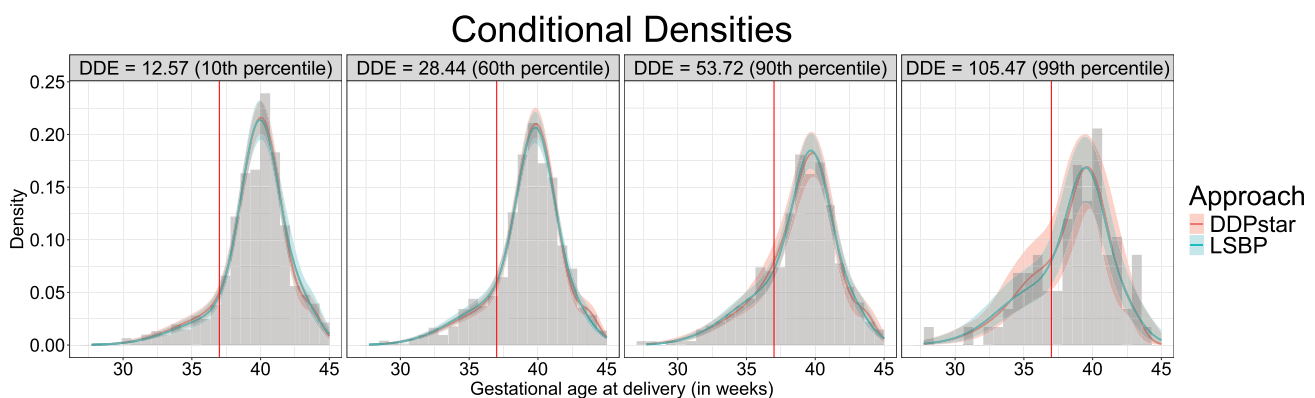
$$\text{ROC}(t \mid \mathbf{x}) = 1 - F_D\{F_D^{-1}(1 - t \mid \mathbf{x}) \mid \mathbf{x}\}, \quad (20)$$

and  $F_D(y \mid \mathbf{x}) = \Pr(Y \leq y \mid \mathbf{x}, D = 1)$  and  $F_{\bar{D}}(y \mid \mathbf{x}) = \Pr(Y \leq y \mid \mathbf{x}, D = 0)$ , with  $D$  being the binary variable indicating the presence ( $D = 1$ ) or absence ( $D = 0$ ) of disease. Related to the ROC curve, various summary indices serve as measures of a diagnostic test’s accuracy. The most commonly used one is the area under the ROC curve (AUC), which, in the conditional context, is defined as  $\text{AUC}(\mathbf{x}) = \int_0^1 \text{ROC}(t \mid \mathbf{x}) dt$ . An AUC of 1 indicates a test that can perfectly distinguish between diseased and nondiseased individuals, while an AUC of 0.5 signifies a test with no discriminatory ability.

Our goal is to assess the utility of the body mass index (BMI) as a diagnostic tool for distinguishing individuals with two or more cardiovascular disease (CVD) risk factors (classified as ‘diseased’) from those with none or just one CVD risk factor (classified as ‘nondiseased’). Since anthropometric measures, including BMI, vary with age and sex, we focus here on estimating age- and sex-specific ROC curves. The dataset comes from a cross-sectional study conducted by the Galician Endocrinology and Nutrition Foundation, consisting of 2840 individuals aged 18–85 years, with 691 classified as diseased (273 women and 418 men) and 2149 as



**Fig. 4** Toxicology study and DDPstar model: posterior median (solid lines) and 95% pointwise credible band (shaded areas) of the conditional mean (left plot), conditional quartiles (middle plot; Q1: first quartile, Q2: median, Q3: third quartile), and conditional variance (right plot)



**Fig. 5** Toxicology study: posterior median (solid lines) and 95% pointwise credible band (shaded areas) of the conditional density for GAD for selected percentiles of DDE. LSBP stands for the proposal of Rigon

and Durante (2021). The shaded grey areas correspond to ‘conditional histograms’ (details are given in the text)

nondiseased (1250 women and 899 men). A detailed description can be found in Tomé et al. (2009).

From Eq. (20), it is evident that the estimation of covariate-specific ROC curves can be approached through density regression, i.e., through the estimation of the conditional density function of the diagnostic test (in our case BMI) separately in the diseased and nondiseased population. Here we use the DDPstar model for that purpose. In particular, the following factor-by-smooth interaction models (see Supplementary Material A) are assumed for the mean of each mixture component in, respectively, the diseased and nondiseased populations (again here the dependence on the component is omitted)

$$\mu_D(\mathbf{x}_i) = \beta_{0D} + \beta_{1D}\mathbb{1}_{\{\text{sex}_i=\text{Women}\}} + f_{\text{sex}_i}^D(\text{age}_i),$$

$$i = 1, \dots, 691,$$

$$\mu_{\bar{D}}(\mathbf{x}_j) = \beta_{0\bar{D}} + \beta_{1\bar{D}}\mathbb{1}_{\{\text{sex}_j=\text{Women}\}} + f_{\text{sex}_j}^{\bar{D}}(\text{age}_j),$$

$$j = 1, \dots, 2149,$$

where  $\text{sex}_i$  (the same applies to  $\text{sex}_j$ ) denotes the sex of the  $i$ th diseased individual (either Men or Women), and  $\mathbb{1}_{\{\cdot\}}$  is the indicator function. Note that in our mean formulation, we allow the smooth function of age to be different in men and women and we approximate them using cubic B-spline bases of dimension  $J = 23$  (we exclude the intercept). The Gibbs sampler takes around 63 and 130 s in the diseased and nondiseased populations, respectively. The posterior predictive checks and quantile residuals shown in Web Figures 32 and 33 of the Supplementary Material, respectively, indicate a good fit of DDPstar to the data in both populations.

Figure 6 presents the estimated age- and sex-specific AUCs. For men, BMI accuracy remains relatively stable across age. In contrast, for women, age has a marked nonlinear influence: accuracy declines until around 70 years, then stabilises and remains steady until age 85. To provide further insights, Web Figure 34 in the Supplementary Material depicts the estimated conditional density functions of BMI for diseased and nondiseased individuals at different ages,

for both men and women. These plots show that the degree of separation between the BMI distributions in diseased and nondiseased individuals—which is what the ROC curve measures—varies with age and sex, explaining the changes in accuracy across these characteristics.

For comparison, we also obtain age- and sex-specific ROC curves using the proposal described in Inácio de Carvalho et al. (2013). The same models as for DDPstar are assumed for the mean of each component in the diseased and nondiseased populations, and the smooth functions of age are also approximated using cubic B-spline bases. However, in this approach, results depend heavily on the number of B-spline basis functions (knots are located at the quantiles of the age covariate) since no penalisation is considered. As such, we evaluate, separately for diseased and nondiseased individuals and for men and women, different numbers of B-spline basis functions ( $J \in \{3, 4, 5, 6, 7\}$ ), and select the ones that provide the lowest widely applicable information criterion. Note that this implies estimating a total of 20 ( $= 2 \times 2 \times 5$ ) models. For that purpose, the R package ROCnReg (Rodríguez-Álvarez and Inácio 2021) is used and we consider prior distributions that align with those from the DDPstar models. As can be seen from Fig. 6, which also shows the age and gender-specific AUCs obtained through this unpenalised approach, both DDPstar and ROCnReg approaches yield similar results except for individuals aged 70 or older. However, uncertainty becomes substantial beyond this age threshold due to limited available data.

In this application, the primary benefit of DDPstar is that it eliminates the need to examine or assess various B-spline basis dimensions. This leads to a substantial reduction in computational workload, as we only need to fit one model instead of the 20 considered in the case of ROCnReg.

### 4.3 Agricultural study

Field trial experiments typically aim to assess the influence of specific genotypes on a targeted trait, which can include crop yield, resistance to pests, or quality measures. These trials often serve commercial purposes, identifying superior genotypes for future commercialisation. However, it is well-known that environmental conditions have a strong influence on trait or phenotypic expression. For instance, soil fertility can greatly affect crop yield, independent of genotype performance. Therefore, in analysing field trials, it is crucial to effectively separate between environmental and genetic influences on phenotypic expression.

As an illustrative example, we examine a Chilean wheat trial described in Lado et al. (2013). In this study, 384 diverse wheat genotypes were planted in an 800-plot field arranged in a 40-row by 20-column grid. Genotypes were randomly assigned to plots using an alpha-lattice design with 20 incomplete blocks, each block containing 20 genotypes.

Two replicates were planted for each genotype. The primary focus here is on grain yield as the trait of interest. The left-hand panel of Fig. 7 depicts the observed grain yield in each plot, revealing a pronounced spatial pattern. Because of the randomisation process and the inclusion of replicates, this pattern cannot be attributed to genetics, and it likely arises from environmental factors. Unfortunately, in many cases, there are no covariates available to relate to this effect, leading modelling strategies to depend on the position or spatial coordinates of the plots in the field only.

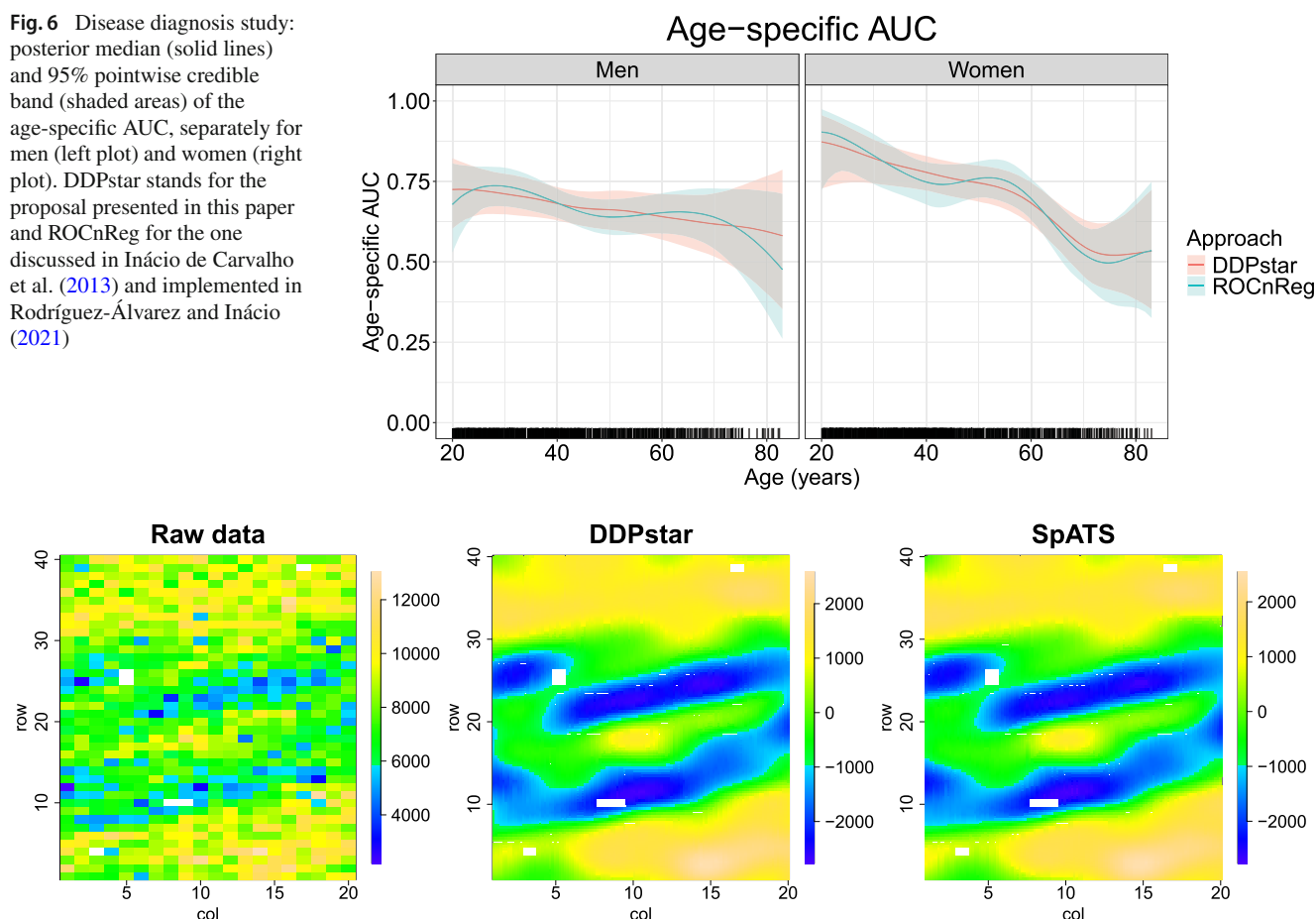
Recently, Rodríguez-Álvarez et al. (2018) proposed a new spatial model, called SpATS, for the analysis of field trials based on using tensor product P-splines to model spatial trends. Here, we adapt the original proposal to our DDPstar model, and specify the mean of each mixture component as a function of available covariates. These are the genotypes (`gen`), where the interest is focused, the spatial coordinates of the plots in the field ( $x_1$  and  $x_2$  for rows and columns, respectively) as well as the row (`row`) and column (`col`) position of the plots. The model for the mean of each mixture component follows (the dependence on the component's index is omitted)

$$\mu(x_i) = \beta_0 + \zeta_{g(i)}^{\text{gen}} + \zeta_{r(i)}^{\text{row}} + \zeta_{c(i)}^{\text{col}} + f(x_{1i}, x_{2i}), \\ i = 1, \dots, 800,$$

where  $f(x_1, x_2)$  is a smooth bivariate function, jointly defined over  $x_1$  and  $x_2$ , that models the spatial effect/trend. We use indices  $g$ ,  $r$ , and  $c$  to refer to genotype, row and column positions, respectively ( $g = 1, \dots, 384$ ;  $r = 1, \dots, 40$ ;  $c = 1, \dots, 20$ ). Accordingly,  $\zeta_g^{\text{gen}}$  represents the genotypic effect for the  $g$ th genotype (similarly for `row` and `col`; these effects are typically incorporated to account for the way machines work before and during sowing or planting). To denote the genotype, row and column positions of the  $i$ th plot, we use  $g(i)$ ,  $r(i)$ , and  $c(i)$ , respectively.

Following Rodríguez-Álvarez et al. (2018), the spatial effect  $f(x_1, x_2)$  is modelled by the tensor product of two marginal cubic B-splines bases (excluding the intercept). Since there are twice as many rows ( $x_1$  values) as columns ( $x_2$  values), we use bases of varying dimensions, with  $J_1 = 23$  for  $x_1$  and  $J_2 = 13$  for  $x_2$ . Furthermore, we consider  $\zeta_g^{\text{gen}} \stackrel{\text{iid}}{\sim} N(0, \sigma_{\text{gen}}^2)$ ,  $\zeta_r^{\text{row}} \stackrel{\text{iid}}{\sim} N(0, \sigma_{\text{row}}^2)$ ,  $\zeta_c^{\text{col}} \stackrel{\text{iid}}{\sim} N(0, \sigma_{\text{col}}^2)$  as prior distributions, i.e., these correspond to random effects, with  $\sigma_{\text{gen}}^2$ ,  $\sigma_{\text{row}}^2$  and  $\sigma_{\text{col}}^2 \sim \text{IG}(1, 0.005)$  (see Supplementary Material A). In contrast to the previous studies, here we opt for  $L = 5$  components to reduce computational complexity. The Gibbs sampler takes approximately 20 min, which is substantially longer than in the other applications. Posterior predictive checks and quantile residuals, shown in Web Figures 35 and 36 of the Supplementary Material, respectively, indicate an exceptionally accurate fit of DDPstar to the data.

**Fig. 6** Disease diagnosis study: posterior median (solid lines) and 95% pointwise credible band (shaded areas) of the age-specific AUC, separately for men (left plot) and women (right plot). DDPstar stands for the proposal presented in this paper and ROCnReg for the one discussed in Inácio de Carvalho et al. (2013) and implemented in Rodríguez-Álvarez and Inácio (2021)



**Fig. 7** Agricultural study: raw data (left), posterior median of the bivariate spatial trend using DDPstar (middle), and estimated spatial trend by means of SpATS (right)

The central panel of Fig. 7 displays the posterior median of the bivariate spatial surface. As can be seen, the spatial trend effectively captures the intricate spatial pattern observed in the raw data (left-hand panel of Fig. 7). For comparison purposes, we also fit the model using SpATS. In the context of this paper, a SpATS model corresponds to a model with only one component, estimated through an empirical Bayes approach (for details, see Rodríguez-Álvarez et al. 2018). Figure 7 also depicts the estimated spatial trend using SpATS, revealing minimal differences between both approaches. Figure 8 presents a comparison between the predicted genotypic effects obtained through our approach and SpATS. Overall, there is strong agreement in the genotypic predictions obtained by both methods, and this extends to the widths of the 95% credible and confidence intervals (see Web Figure 37 in the Supplementary Material).

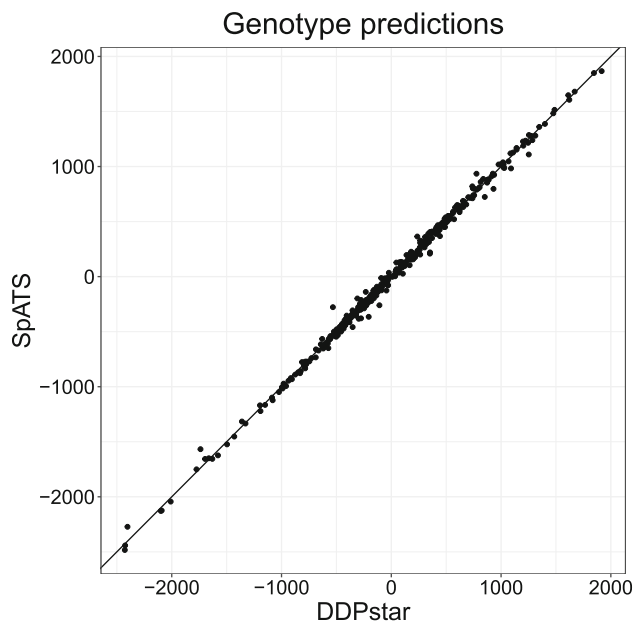
Overall, in this specific application, employing a more intricate model, such as the DDPstar approach, does not appear to offer any clear advantage compared to SpATS, which is also much faster, taking only about three seconds. In fact, the results show that only one component of the

mixture is predominantly occupied, receiving the majority of the weighting (see Web Figure 38 in the Supplementary Material). Nevertheless, these findings are significant as they demonstrate our approach's ability to handle complex analyses effectively.

## 5 Discussion

This paper introduced a novel density regression model, DDPstar, which harnesses the flexibility of (i) single-weights DDP mixture models, allowing for a broad class of response distributions and enabling the entire density to smoothly change with covariates; and (ii) structured additive regression models, allowing the accommodation of a variety of covariate effects. Through an extensive simulation study we demonstrated that DDPstar performs effectively even in scenarios where it is misspecified (e.g., in a scenario involving mixture models where the weights of each component depend on covariates). We also provided an example, involving a non-stationary mean function and covariate dependent variance,





**Fig. 8** Agricultural study: scatter plot of the predicted genotypic effects using SpATS (y-axis) against the posterior median of the genotypic effects using DDPstar (x-axis)

where our DDPstar model did not perform well. Although allowing the variance of each normal component to depend on covariates has been shown to be valuable (Villani et al. 2009), assuming a similar additive structure for the variance as for the mean would lead to a model with practically twice the number of parameters per normal component. We further illustrated the broad applicability of DDPstar using three real datasets, and its performance was on par with the state-of-the-art approach for each application considered. Our model is implemented in the publicly available R package `DDPstar` thereby providing a user-friendly and broadly applicable novel approach, and whose performance can be easily assessed through model checking procedures.

It is reasonable to inquire about the advantages of assuming an additive structure based on P-splines compared to a Gaussian process (MacEachern 2000; Xu et al. 2016, 2019, 2022). First, there is no closed form expression for the posterior distribution of the hyperparameters of the covariance function of the Gaussian process. These hyperparameters, along with the chosen forms for the covariance and mean functions, are crucial in determining the behaviour of the functionals of interest. Second, if we denote the total number of observations in each mixture component by  $n_l$ , Gaussian processes require the inversion of a  $n_l \times n_l$  matrix in each component, which is computationally expensive. In contrast, our model requires only the inversion of a block matrix, which does not depend on the sample size per component, but instead on the number of total B-splines basis functions (see Supplementary Material B). We also note that, although

our discussion has focused on the Bayesian nonparametric literature, there are closely related models in the machine learning literature known as ‘mixture of experts’ (see, e.g., Frühwirth-Schnatter et al. 2019, Chapter 12, and references therein).

There are a few limitations to mention. First, in our DDPstar model, allocation to a specific mixture component is independent of the covariates. Therefore, if the focus is on clustering, formulations that allow for covariate-dependent weights should be preferred. Second, except for the agricultural example, our MCMC scheme was quite fast. However, our approach does not scale well with increasing sample size or number of covariates. While we do not anticipate the DDPstar model being used in applications with a moderate to large number of covariates, variational inference (see, e.g., Blei and Jordan 2006, for DPMs) could be implemented for applications with large sample sizes. Third, although DDPstar provides a flexible formulation for density regression when the response variable is continuous and supported on the real line, it is clearly not suitable for count responses. However, we could easily adapt the proposal of Canale and Dunson (2011), based on DP mixtures of rounded Gaussian kernels, to our case. This extension would only involve an additional data augmentation step in our Gibbs sampler scheme. Alternatively, in the spirit of generalised linear and additive models, extending our modelling framework to incorporate discrete kernels (e.g., Poisson and binomial), along with developing computationally convenient posterior samplers for such kernels, represents a promising direction for future research.

A further interesting application of the DDPstar model is that of survival analysis. In this case, considering the logarithm of the event times, and handling right-censored times through a data augmentation step in the Gibbs sampler scheme, DDPstar can be interpreted as a mixture of lognormal accelerated failure time models. In contrast to accelerated failure time and Cox proportional hazards models, the survival curves from different covariate levels would be allowed to cross, a feature that is often appealing in practice. Furthermore, because DDPstar can deal with random effects, the potential inclusion of frailty terms in the aforementioned model would also be interesting to explore.

Another avenue for future research involves the choice of the prior distribution for the variance parameters within the DDPstar model framework. While this paper focused on the inverse gamma prior due to its conjugacy properties, we recognise the importance of exploring alternative prior distributions, such as those discussed in Klein and Kneib (2016).

Lastly, a generalisation of the DDPstar model could involve replacing the DP prior on the random mixing measure with a Pitman–Yor process (Pitman 1995), of which the DP is a particular case. Compared to the DP, the Pitman–Yor process, enables more robust inference regarding the num-



ber of components characterising the distribution of the data (Canale et al. 2022) and is particularly advantageous for modelling heavy-tailed distributions (Ramírez et al. 2024).

The DDPstar package and the R-codes and datasets needed to reproduce the results of the simulation study and applications can be freely downloaded from <https://bitbucket.org/mxrodriguez/ddpstar>.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11222-025-10567-0>.

**Author Contributions** M.X. R-A: conceptualisation and methodology, software implementation, simulations; design and execution, data analysis, visualisation, original draft writing, review and editing. V.I.: conceptualisation and methodology, software implementation, simulations; design and execution, data analysis, visualisation, original draft writing, review and editing. N. K.: review and editing.

**Funding** Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. M.X. R-A was partially funded by the Spanish State Research Agency through the Ramón y Cajal grant RYC2019-027534-I. Funding for open access charge: Universidade de Vigo/CISUG.

**Data availability** Data is provided within the manuscript or supplementary information files.

**Code availability** The DDPstar package and the R-codes and datasets needed to reproduce the results of the simulation study and applications can be freely downloaded from <https://bitbucket.org/mxrodriguez/ddpstar>.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Antoniano-Villalobos, I., Wade, S., Walker, S.G.: A Bayesian non-parametric regression model with normalized weights: a study of hippocampal atrophy in Alzheimer's disease. *J. Am. Stat. Assoc.* **109**(506), 477–490 (2014)
- Bach, P., Klein, N.: Anisotropic multidimensional smoothing using Bayesian tensor product P-splines. *arXiv preprint (2022)* [arXiv:2211.16218](https://arxiv.org/abs/2211.16218)
- Barrientos, A.F., Jara, A., Quintana, F.A.: On the support of MacEachern's dependent Dirichlet processes and extensions. *Bayesian Anal.* **7**(2), 277–310 (2012)
- Berretini, M., Galimberti, G., Rancati, S.: Semiparametric finite mixture of regression models with Bayesian P-splines. *Adv. Data Anal. Classif.* **17**(3), 745–775 (2023)
- Bishop, C.M.: Pattern recognition and machine learning. Springer, New York (2006)
- Blei, D.M., Jordan, M.I.: Variational inference for Dirichlet process mixtures. *Bayesian Anal.* **1**(1), 121–143 (2006)
- Canale, A., Dunson, D.B.: Bayesian kernel mixtures for counts. *J. Am. Stat. Assoc.* **106**(496), 1528–1539 (2011)
- Canale, A., Durante, D., Dunson, D.B.: Convex mixture regression for quantitative risk assessment. *Biometrics* **74**(4), 1331–1340 (2018)
- Canale, A., Corradin, R., Nipoti, B.: Importance conditional sampling for Pitman-Yor mixtures. *Stat. Comput.* **32**(3), 40 (2022)
- Celeux, G., Forbes, F., Robert, C.P., Titterton, D.M.: Deviance information criteria for missing models. *Bayesian Anal.* **1**(4), 651–673 (2006)
- Chib, S., Greenberg, E.: Additive cubic spline regression with Dirichlet process mixture errors. *J. Econ.* **156**(2), 322–336 (2010)
- Chipman, H.A., George, E.I., McCulloch, R.E.: BART: bayesian additive regression trees. *Ann. Appl. Stat.* **4**(1), 266–298 (2010)
- Chung, Y., Dunson, D.B.: Nonparametric Bayes conditional distribution modeling with variable selection. *J. Am. Stat. Assoc.* **104**(488), 1646–1660 (2009)
- Currie, I.D., Durban, M., Eilers, P.H.C.: Generalized linear array models with applications to multidimensional smoothing. *J. Roy. Stat. Soc. B* **68**(2), 259–280 (2006)
- De Iorio, M., Müller, P., Rosner, G.L., MacEachern, S.N.: An ANOVA model for dependent random measures. *J. Am. Stat. Assoc.* **99**(465), 205–215 (2004)
- De Iorio, M., Johnson, W.O., Müller, P., Rosner, G.L.: Bayesian non-parametric nonproportional hazards survival modeling. *Biometrics* **65**(3), 762–771 (2009)
- De la Cruz-Mesía, R., Quintana, F.A., Müller, P.: Semiparametric Bayesian classification with longitudinal markers. *J. R. Stat. Soc. Ser. C* **56**(2), 119–137 (2007)
- DiMatteo, I., Genovese, C.R., Kass, R.E.: Bayesian curve-fitting with free-knot splines. *Biometrika* **88**(4), 1055–1071 (2001)
- Dunn, P.K., Smyth, G.K.: Randomized quantile residuals. *J. Comput. Graph. Stat.* **5**(3), 236–244 (1996)
- Dunson, D.B., Park, J.-H.: Kernel stick-breaking processes. *Biometrika* **95**(2), 307–323 (2008)
- Dunson, D.B., Pillai, N., Park, J.-H.: Bayesian density regression. *J. Roy. Stat. Soc. B* **69**(2), 163–183 (2007)
- Eilers, P.H., Marx, B.D.: Flexible smoothing with B-splines and penalties. *Stat. Sci.* **11**(2), 89–121 (1996)
- Eilers, P.H.C., Marx, B.D.: Multidimensional calibration with temperature interaction using two-dimensional penalized signal regression. *Chemom. Intell. Lab. Syst.* **66**, 159–174 (2003)
- Fahrmeir, L., Kneib, T., Lang, S., Marx, B. D.: Regression: Models, Methods, and Applications. Springer, second edition (2022)
- Ferguson, T.S.: A Bayesian analysis of some nonparametric problems. *Ann. Stat.* **1**(2), 209–230 (1973)
- Fronczyk, K., Kottas, A.: A Bayesian approach to the analysis of quantal bioassay studies using nonparametric mixture models. *Biometrics* **70**(1), 95–102 (2014)
- Fruhwirth-Schnatter, S., Celeux, G., Robert, C.P.: Handbook of mixture analysis. CRC Press, Boca Raton (2019)
- Geisser, S., Eddy, W.F.: A predictive approach to model selection. *J. Am. Stat. Assoc.* **74**(365), 153–160 (1979)
- Gelfand, A.E., Kottas, A., MacEachern, S.N.: Bayesian nonparametric spatial modeling with Dirichlet process mixing. *J. Am. Stat. Assoc.* **100**(471), 1021–1035 (2005)
- Gelman, A., Hwang, J., Vehtari, A.: Understanding predictive information criteria for Bayesian models. *Stat. Comput.* **24**(6), 997–1016 (2014)
- Ghosal, S., Van der Vaart, A.: Fundamentals of nonparametric Bayesian inference. Cambridge University Press, Cambridge (2017)
- Griffin, J.E., Steel, M.J.: Order-based dependent Dirichlet processes. *J. Am. Stat. Assoc.* **101**(473), 179–194 (2006)

- Gustafson, P.: Bayesian regression modeling with interactions and smooth effects. *J. Am. Stat. Assoc.* **95**(451), 795–806 (2000)
- Gutiérrez, L., Quintana, F. A.: Multivariate Bayesian semiparametric models for authentication of food and beverages. *The Annals of Applied Statistics*, pages 2385–2402 (2011)
- Inácio de Carvalho, V., Jara, A., Hanson, E., T., de Carvalho, M.: Bayesian nonparametric ROC regression modeling. *Bayesian Analysis* **8**(3), 623–646 (2013)
- Inácio de Carvalho, V., de Carvalho, M., Branscum, A.J.: Nonparametric Bayesian covariate-adjusted estimation of the Youden index. *Biometrics* **73**(4), 1279–1288 (2017)
- Inácio, V., Rodríguez-Álvarez, M.X.: The covariate-adjusted ROC curve: the concept and its importance, review of inferential methods, and a new Bayesian estimator. *Stat. Sci.* **37**(4), 541–561 (2022)
- Inácio, V., Rodríguez-Álvarez, M.X., Gayoso-Diz, P.: Statistical evaluation of medical tests. *Ann. Rev. Stat. Appl.* **8**(1), 41–67 (2021)
- Ishwaran, H., James, L.F.: Gibbs sampling methods for stick-breaking priors. *J. Am. Stat. Assoc.* **96**(453), 161–173 (2001)
- Ishwaran, H., Zarepour, M.: Markov chain Monte carlo in approximate Dirichlet and beta two-parameter process hierarchical models. *Biometrika* **87**(2), 371–390 (2000)
- Jara, A., Lesaffre, E., De Iorio, M., Quintana, F.: Bayesian semiparametric inference for multivariate doubly-interval-censored data. *Ann. Appl. Stat.* **4**(4), 2126–2149 (2010)
- Klein, N., Kneib, T.: Scale-dependent priors for variance parameters in structured additive distributional regression. *Bayesian Anal.* **11**(4), 1071–1106 (2016)
- Klein, N., Kneib, T., Lang, S., Sohn, A.: Bayesian structured additive distributional regression with an application to regional income inequality in germany. *Ann. Appl. Stat.* **9**(2), 1024–1052 (2015)
- Kneib, T., Klein, N., Lang, S., Umlauf, N.: Modular regression—a lego system for building structured additive distributional regression models with tensor product interactions. *TEST* **28**, 1–39 (2019)
- Kottas, A., Müller, P., Quintana, F.: Nonparametric Bayesian modeling for multivariate ordinal data. *J. Comput. Graph. Stat.* **14**(3), 610–625 (2005)
- Lado, B., Matus, I., Rodríguez, A., Inostroza, L., Poland, J., Belzile, F., del Pozo, A., Quincke, M., Castro, M., von Zitzewitz, J.: Increased genomic prediction accuracy in wheat breeding through spatial adjustment of field trial data. *G3 (Genes, Genomes, Genetics)*. **3**(12):2105–2114 (2013)
- Lang, S., Brezger, A.: Bayesian P-splines. *J. Comput. Graph. Stat.* **13**(1), 183–212 (2004)
- Lee, D.-J.: Smoothing mixed model for spatial and spatio-temporal data. PhD thesis, Department of Statistics, Universidad Carlos III de Madrid, Spain (2010)
- Lee, D.-J., Durbán, M., Eilers, P.: Efficient two-dimensional smoothing with P-spline ANOVA mixed models and nested bases. *Comput. Stat. Data Anal.* **61**, 22–37 (2013)
- Liu, J.S.: Nonparametric hierarchical Bayes via sequential imputations. *Ann. Stat.* **24**(3), 911–930 (1996)
- Longnecker, M.P., Klebanoff, M.A., Zhou, H., Brock, J.W.: Association between maternal serum concentration of the DDT metabolite DDE and preterm and small-for-gestational-age babies at birth. *The Lancet* **358**(9276), 110–114 (2001)
- MacEachern, S. N.: Dependent nonparametric processes. In *ASA Proceedings of the Section on Bayesian Statistical Science*, volume 1, pages 50–55. Alexandria, Virginia. Virginia: American Statistical Association; 1999 (1999)
- MacEachern, S. N.: Dependent Dirichlet processes. Technical Report, Department of Statistics, The Ohio State University (2000)
- Pati, D., Dunson, D.B., Tokdar, S.T.: Posterior consistency in conditional distribution estimation. *J. Multivar. Anal.* **116**, 456–472 (2013)
- Pitman, J.: Exchangeable and partially exchangeable random partitions. *Probab. Theory Relat. Fields* **102**(2), 145–158 (1995)
- Pratola, M.T., Chipman, H.A., George, E.I., McCulloch, R.E.: Heteroscedastic BART via multiplicative regression trees. *J. Comput. Graph. Stat.* **29**(2), 405–417 (2020)
- Quintana, F.A., Müller, P., Jara, A., MacEachern, S.N.: The dependent Dirichlet process and related models. *Stat. Sci.* **37**(1), 24–41 (2022)
- R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2023)
- Ramírez, V.P., de Carvalho, M., Gutiérrez, L.: Heavy-tailed NGG-mixture models. *Bayesian Anal.* **1**(1), 1–29 (2024)
- Rigon, T., Durante, D.: Tractable Bayesian density regression via logit stick-breaking priors. *J. Stat. Plann. Inference* **211**, 131–142 (2021)
- Rodríguez, A., Ter Horst, E.: Bayesian dynamic density estimation. *Bayesian Anal.* **3**(2), 339–365 (2008)
- Rodríguez, A., Martínez, J.C.: Bayesian semiparametric estimation of covariate-dependent ROC curves. *Biostatistics* **15**(2), 353–369 (2014)
- Rodríguez-Álvarez, M.X., Inácio, V.: ROCnReg: an R package for receiver operating characteristic curve inference with and without covariates. *R J.* **13**(1), 525–555 (2021)
- Rodríguez-Álvarez, M.X., Boer, M.P., van Eeuwijk, F.A., Eilers, P.H.: Correcting for spatial heterogeneity in plant breeding experiments with P-splines. *Spatial Stat.* **23**, 52–71 (2018)
- Sethuraman, J.: A constructive definition of Dirichlet priors. *Stat. Sin.* **4**(2), 639–650 (1994)
- Tomé, M.A., Botana, M.A., Cadarso-Suárez, C., Rego-Iraeta, A., Fernández-Mariño, A., Mato, J.A., Solache, I., Perez-Fernandez, R.: Prevalence of metabolic syndrome in galicia (nw Spain) on four alternative definitions and association with insulin resistance. *J. Endocrinol. Invest.* **32**(6), 505–511 (2009)
- Villani, M., Kohn, R., Giordani, P.: Regression density estimation using smooth adaptive gaussian mixtures. *J. Econo.* **153**(2), 155–173 (2009)
- Wade, S., Inacio, V.: Bayesian dependent mixture models: A predictive comparison and survey. *Stat. Sci.* **40**(1), 81–108 (2025)
- Wickham, H.: ggplot2: elegant graphics for data analysis. Springer-Verlag, New York (2016)
- Wiesenfarth, M., Hisgen, C.M., Kneib, T., Cadarso-Suarez, C.: Bayesian nonparametric instrumental variables regression based on penalized splines and Dirichlet process mixtures. *J. Bus. Econ. Stat.* **32**(3), 468–482 (2014)
- Williams, C. K., Rasmussen, C. E.: Gaussian Processes for Machine Learning. CMIT Press (2006)
- Wood, S.N., Scheipl, F., Faraway, J.J.: Straightforward intermediate rank tensor product smoothing in mixed models. *Stat. Comput.* **23**(3), 341–360 (2013)
- Xu, Y., Müller, P., Wahed, A.S., Thall, P.F.: Bayesian nonparametric estimation for dynamic treatment regimes with sequential transition times. *J. Am. Stat. Assoc.* **111**(515), 921–950 (2016)
- Xu, Y., Thall, P.F., Hua, W., Andersson, B.S.: Bayesian non-parametric survival regression for optimizing precision dosing of intravenous busulfan in allogeneic stem cell transplantation. *J. R. Stat. Soc.: Ser. C: Appl. Stat.* **68**(3), 809–828 (2019)
- Xu, Y., Scharfstein, D., Müller, P., Daniels, M.: A Bayesian nonparametric approach for evaluating the causal effect of treatment in randomized trials with semi-competing risks. *Biostatistics* **23**(1), 34–49 (2022)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.