

Multidimensional adaptive P-splines with application to neurons' activity studies

María Xosé Rodríguez-Álvarez^{1,2,3,4}  | María Durbán⁵ | Paul H.C. Eilers⁶ |
Dae-Jin Lee¹ | Francisco Gonzalez⁷

¹BCAM - Basque Center for Applied Mathematics, Bilbao, Spain

²IKERBASQUE, Basque Foundation for Science, Bilbao, Spain

³Current address: CINBIO, Universidade de Vigo, Department of Statistics and Operations Research, Vigo, Spain

⁴CITMAga, Galician Center for Mathematical Research and Technology, Santiago de Compostela, Spain

⁵Department of Statistics, Universidad Carlos III de Madrid, Leganés, Spain

⁶Erasmus University Medical Centre, Rotterdam, The Netherlands

⁷Department of Surgery and CiMUS, University of Santiago de Compostela, Santiago, Spain

Correspondence

María Xosé Rodríguez-Álvarez, CINBIO, Universidade de Vigo, Department of Statistics and Operations Research, Vigo, Spain.

Email: mxrodriguez@uvigo.es

Funding information

Instituto de Salud Carlos III, Spain, Grant/Award Number: RETICS RD16/0008/0003; Xunta de Galicia, Grant/Award Numbers: Accreditation 2019-2022, ED431G 2019/02; Spanish Ministry of Science, Innovation, and Universities, Grant/Award Number: BCAM Severo Ochoa accreditation SEV-2017-0718; Eusko Jaurlaritz, Grant/Award Number: BERC 2018-2021 program; Agencia Estatal de Investigación (AEI), Spain, Grant/Award Numbers: MTM2017-82379-R, PID2019-104901RB-I00; Ministerio de Ciencia e Innovación, Grant/Award Number: Ramon y Cajal Grant RYC2019-027534-I

Abstract

The receptive field (RF) of a visual neuron is the region of the space that elicits neuronal responses. It can be mapped using different techniques that allow inferring its spatial and temporal properties. Raw RF maps (RFmaps) are usually noisy, making it difficult to obtain and study important features of the RF. A possible solution is to smooth them using P-splines. Yet, raw RFmaps are characterized by sharp transitions in both space and time. Their analysis thus asks for spatiotemporal adaptive P-spline models, where smoothness can be locally adapted to the data. However, the literature lacks proposals for adaptive P-splines in more than two dimensions. Furthermore, the extra flexibility afforded by adaptive P-spline models is obtained at the cost of a high computational burden, especially in a multidimensional setting. To fill these gaps, this work presents a novel anisotropic locally adaptive P-spline model in two (e.g., space) and three (space and time) dimensions. Estimation is based on the recently proposed SOP (Separation of Overlapping Precision matrices) method, which provides the speed we look for. Besides the spatiotemporal analysis of the neuronal activity data that motivated this work, the practical performance of the proposal is evaluated through simulations, and comparisons with alternative methods are reported.

KEYWORDS

anisotropy, local adaptivity, penalized splines, smoothing, visual receptive fields

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Biometrics* published by Wiley Periodicals LLC on behalf of International Biometric Society.

1 | INTRODUCTION

A fundamental characteristic of neurons of the mammalian visual system, as well as all neurons involved in other sensory systems, is that they have what is known as a receptive field (RF). In the case of a visual neuron, the RF is the small area of the visual field that the neuron “sees” (i.e., it is the area that elicits neuronal responses). It is known that this RF has both spatial and temporal properties, and their study is fundamental to understanding how visual information is processed in the brain.

RF can be mapped using different techniques, such as reverse cross-correlation. These techniques allow studying how visual neurons process sensory stimuli from different positions in their visual field. In brief, since as a result of a sensory stimulus, a neuron can produce sudden changes in its membrane potential known as “spikes,” from the neuron responses (spikes), it is possible to infer properties of the RF (e.g., its spatiotemporal properties: where and when a sensory stimulus produces a response). Yet, neuronal responses usually produce noisy signals as they encode the information by means of spike trains of variable length and frequency, and therefore, the spatial and temporal properties of their RF might not be clearly defined. Postprocessing the noisy neural signals through smoothing techniques allows obtaining RF maps (RFmaps) with desirable properties and paves the way for the study of important characteristics of RFs, such as their size, center, or boundaries, as well as comparisons among neurons and/or experimental conditions.

Penalized splines (O’Sullivan, 1986; Eilers & Marx, 1996) has become one of the most popular smoothing methods in semiparametric regression for the estimation of curves and surfaces. Penalized splines combine a low-rank basis with a penalty term. This term controls the smoothness of the estimated function, and its influence is determined by a smoothing parameter. In the *classic* penalized spline approach, this parameter is global in the sense that it provides a constant amount of smoothing (i.e., *classic* penalized splines rely on assuming a smooth transition of the covariate effect across its whole domain). This might be a serious drawback in situations where the curves/surfaces are nonhomogeneous, showing rapid changes in some regions while being rather smooth in others (Ruppert & Carroll, 2000). Indeed, this is the case of the neuronal activity study that motivated the work described in this paper. Details regarding the experiment and the data are given in Section 2, but we advance that, in this setting, *classic* penalized spline models are not able to properly recover the sharp spatiotemporal transitions present in the data, which ask for spatially adaptive smoothness.

In the case of univariate smoothing, a wide range of solutions to adapt smoothness locally to the data have been

proposed. Ruppert and Carroll (2000) based their approach on locally varying smoothing parameters and a multivariate cross-validation (CV) approach for their selection. In Krivobokova et al. (2008), the varying smoothing (variance) parameters are modeled as a log-penalized spline, yielding a hierarchical mixed model; they use a Laplace approximation of the marginal likelihood for parameter estimation. More recently, Yue et al. (2014) introduced a class of adaptive smoothing spline models that is derived by solving certain stochastic differential equations with finite element methods. The task becomes more complicated when trying to achieve adaptivity in two dimensions. As far as we are aware, all attempts to two-dimensional adaptive smoothing have been proposed in the context of isotropic smoothing, i.e., the same amount of smoothing is used in both dimensions. For example, Lang and Brezger (2004) use locally adaptive smoothing parameters that are incorporated using a smoothness prior with spatially adaptive variances and Yue and Speckman (2010) improve this work by proposing a prior with a spatially adaptive variance component and taking a further Gaussian Markov random field prior for this variance function. A different approach is taken by Jang and Oh (2011) who replace the global smoothing parameter by a smooth step function that can be extended to higher dimensions. In Krivobokova et al. (2008), adaptivity in two dimensions is achieved by modeling the spatially variable smoothing parameters using (penalized) low-rank radial basis functions, and Reiss et al. (2017) use adaptive smoothing and a nonisotropic penalty, but the adaptivity is only in one direction, allowing only one smoothing parameter to vary along the other dimension. However, all these approaches are either computationally very demanding or unstable, or do not include general smoothing structures such as anisotropy.

In this work, we present a general framework for anisotropic multidimensional adaptive smoothing in the context of P-spline models (Eilers & Marx, 1996), where B-spline basis are combined with discrete difference penalties on the coefficients. Our proposal relies on the construction of locally adaptive penalties through assuming a different smoothing parameter for each coefficient difference in the penalty. Its use is not restricted to Gaussian responses (as it is in the case of most of the existing approaches) because it can be easily extended to responses within the exponential family of distributions. One of the possible drawbacks of adaptive smoothing is the computational cost of having to estimate, or select by CV methods or information criteria, multiple smoothing parameters. We solve this issue by using the connection between P-splines and generalized linear mixed models (Currie & Durban, 2002) and the recently developed SOP (Separation of Overlapping Precision Matrices) method

(Rodríguez-Álvarez et al., 2019). SOP allows estimating P-spline models with overlapping penalty matrices (or overlapping precision matrices) and uses restricted maximum likelihood (REML) to select the smoothing parameters (variance parameters). For a discussion on the benefits and disadvantages of RE/ML-based and CV-based smoothing parameter selection, we refer the reader to Krivobokova and Kauermann (2007) and Krivobokova (2013).

The rest of the paper is organized as follows. In Section 2, we describe the study that motivated this work. Section 3 presents our approach for the construction of locally adaptive multidimensional anisotropic penalties. The empirical performance of the approach is evaluated in a simulation study in Section 4. In Section 5, results for the neurons' activity study are shown, and we conclude with a discussion. Details on estimation and computation, extended simulations, and extra results for the neuronal activity data are available as [Supporting Information](#).

2 | MOTIVATING EXAMPLE: NEURONS' ACTIVITY STUDY

The work described in this paper was motivated by the research and electrophysiology studies conducted by Francisco Gonzalez, professor of Ophthalmology at the University of Santiago de Compostela (Galicia, Spain). The data we deal with here derive from experiments conducted to study neurons' activity in the visual cortex, in particular, the spatial and temporal properties of RFs. A detailed explanation of the electrophysiological experiment and the reverse cross-correlation technique discussed here can be found in Rodríguez-Álvarez et al. (2012) (and references therein). That paper was also the starting point of this work. Schematically, the experiment is as follows (Figure 1). The subject (a monkey) is viewing two monitors, one for each eye. In each monitor, there is a square area of dimension 16×16 (i.e., 256 spatial/grid locations). In a pseudorandom way, stimuli are delivered at these spatial locations (Figure 1A). The experiment is conducted under two different experimental conditions. Namely, the stimulus can correspond to the flash of a bright ("ON") or dark ("OFF") spot. While the stimuli are being delivered, the activity of the neuron is being recorded (Figure 1B). When a spike occurs, say at t_0 , the location of the stimulus at different prespike times ($-20, -40, \dots, -320$ ms; for this experiment, it is not expected that a stimulus would produce a response (spike) after 320 ms) is recovered (Figure 1C). Unfortunately, the way the experiment is performed does not allow to know which of these stimuli (and thus locations) is responsible for the spike (neuron's response), and all stimuli are considered as potentially responsible. As such, the number of spike occurrences

attributed to the location of the stimulus at the different prespike times is increased by one (Figure 1D, red bold numbers). However, since, during the experiment, stimuli are randomly delivered/presented many times, this allows determining where and when a stimulus produces a neuron's response. Summarizing, for each neuron (the complete experiment contains data for 17 different neurons), eye (left or right), and experimental condition ("ON" or "OFF"), the reverse cross-correlation technique provides a dataset consisting of a series of 16 matrices (one for each prespike time) of dimension 16×16 (256 grid positions that represent the square area). Each cell of each of the 16 matrices contains the number of spike occurrences attributed to this location at the corresponding prespike time. Besides, there is an extra matrix of dimension 16×16 with the number of stimulus presentations on each particular location of the square area. The graphical representation of each of the 16 matrices (normalized, i.e., each cell is divided by the number of stimulus presentations in this location) is called RF map (RFmap) and can be regarded as a representation of the firing rate of the neuron. Figure 2 depicts the evolution over time of the raw RFmap for a particular neuron (denoted by FAU3), eye (right), and experimental condition ("ON"). Despite the noisy data, it can be observed that for most prespike times, the firing rate is uniform (shows no structure), but there is a clear increase in the firing rate around -60 ms for a central area of the visual field. In other words, the results suggest that only stimuli in this central area produce a response of the neuron and that the response occurs around 60 ms after the stimulus is presented. This area corresponds to the RF of the neuron. In this paper, we aim to use P-spline models to provide smoothed (denoised) versions of RFmaps such as those shown in Figure 2. Yet, note that the raw RFmaps show that the transition from outside the RF into the RF is very sharp and that they are structured for a short period (between -40 and -80 ms). This suggests the need for multidimensional adaptive smoothing, which was also pointed out in the discussion of Rodríguez-Álvarez et al. (2012)'s paper.

3 | MULTIDIMENSIONAL ADAPTIVE PENALTY

Our proposal for the construction of locally adaptive penalties in more than one dimension builds upon the same principles as the adaptive penalty in one dimension (see, e.g., Rodríguez-Álvarez et al., 2019, and references therein). In consequence, to make the presentation of the new results more readable, we first briefly focus on the one-dimensional case, and then, we move onto the multidimensional setting. For clarity and brevity, we describe

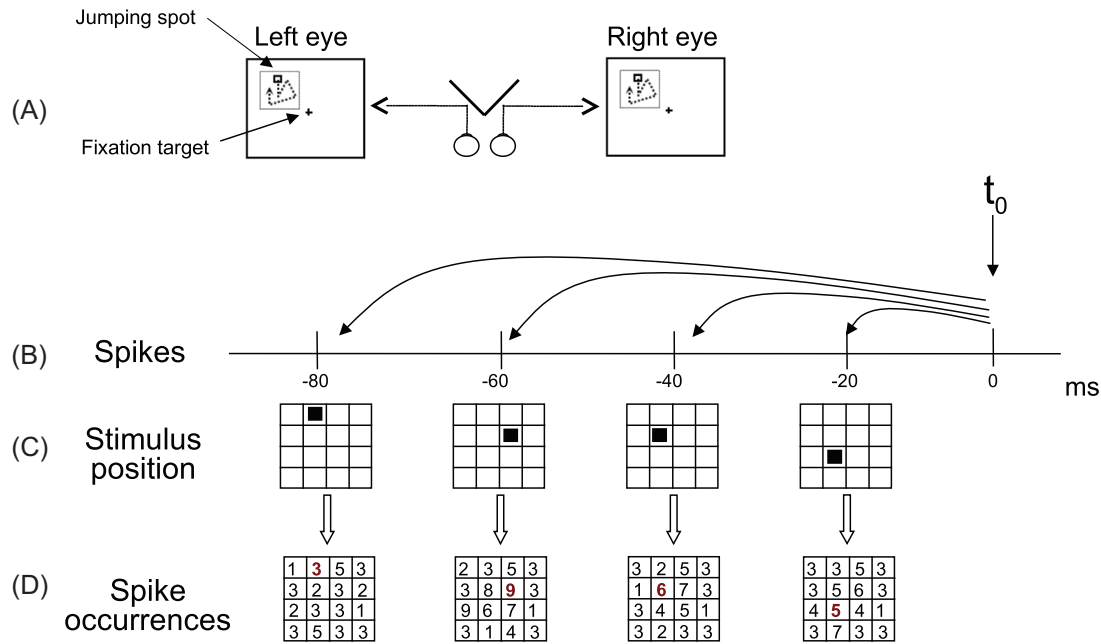


FIGURE 1 Reverse cross-correlation technique. The animal is viewing two monitors (A) with a fixation target. Within a square area over the neuron visual field a bright (“ON”) or dark (“OFF”) spot is flashed in different positions in a pseudorandom manner. Neuron spikes are recorded, whereas the stimulus is delivered (B). When a spike is produced (t_0), the stimulus position at several prespike times ($-20, -40, \dots, -320$ ms) is read (C) and the corresponding position is increased by one (D, red bold numbers). This figure appears in color in the electronic version of this article, and any mention of color refers to that version.

in full detail the rationale for the adaptive penalty in two dimensions and relegate the generalization to the three-dimensional case to Web Appendix C.

3.1 | Adaptive penalty in one dimension

Let $\mathbf{y} = (y_1, \dots, y_n)^\top$ be a vector of n observations, and consider the (simple) generalized model $g(\mu_i) = f(x_i)$ ($i = 1, \dots, n$), where $\mu_i = \mathbb{E}(y_i | x_i)$, $\text{Var}(y_i | x_i) = \phi\nu(\mu_i)$, $g(\cdot)$ is the link function, and $f(\cdot)$ is a smooth and unknown function. Here, $\nu(\cdot)$ is a specified variance function, and ϕ is the dispersion parameter that may be known or unknown. In P-splines (Eilers & Marx, 1996), the function $f(x)$ is modeled as a linear combination of B-spline basis functions, i.e., $f(x) = \sum_{j=1}^d \theta_j B_j(x)$, and smoothness is ensured by penalizing the differences of order q of coefficients associated with adjacent B-spline basis functions, i.e., the penalty takes the following form:

$$\lambda \sum_{k=q+1}^d (\Delta^q \theta_k)^2 = \lambda \boldsymbol{\theta}^\top \mathbf{D}_q^\top \mathbf{D}_q \boldsymbol{\theta}, \quad (1)$$

where Δ^q forms differences of order q , $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_d)^\top$, and \mathbf{D}_q is the matrix representation of Δ^q . Finally, λ is the smoothing parameter that controls the trade-off between fidelity to the data (when λ is small) and smoothness of

the function estimate (when λ is large). Note that Equation (1) penalizes all coefficient differences, $\Delta^q \theta_k$ ($k = q + 1, \dots, d$), by the same smoothing parameter λ (see Web Figure 1a). Implicit in Equation (1) is thus the assumption that the same amount of smoothing is needed across the whole domain of the covariate. The locally adaptive penalty relaxes this assumption by assuming a different smoothing parameter for each coefficient difference

$$\sum_{k=q+1}^d \lambda_{k-q} (\Delta^q \theta_k)^2 = \boldsymbol{\theta}^\top \mathbf{D}_q^\top \text{diag}(\boldsymbol{\lambda}) \mathbf{D}_q \boldsymbol{\theta}, \quad (2)$$

where $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_{d-q})^\top$ is a vector of smoothing parameters. That is, the adaptive penalty defined in (2) allows the amount of smoothing (driven by the smoothing parameters $\boldsymbol{\lambda}$) to vary locally depending on the covariate values. This is graphically illustrated in Web Figure 1(b). To reduce the complexity of the adaptive penalty in (2) (there are as many smoothing parameters as coefficient differences, i.e., $d - q$), the vector of smoothing parameters $\boldsymbol{\lambda}$ is further replaced by a smooth version of it $\boldsymbol{\xi} = (\xi_1, \dots, \xi_p)^\top$ (with $p < (d - q)$ so as to ensure that the number of smoothing parameters is reduced) using a B-spline basis expansion, i.e.,

$$\lambda_k = \sum_{l=1}^p \xi_l \psi_l(k). \quad (3)$$

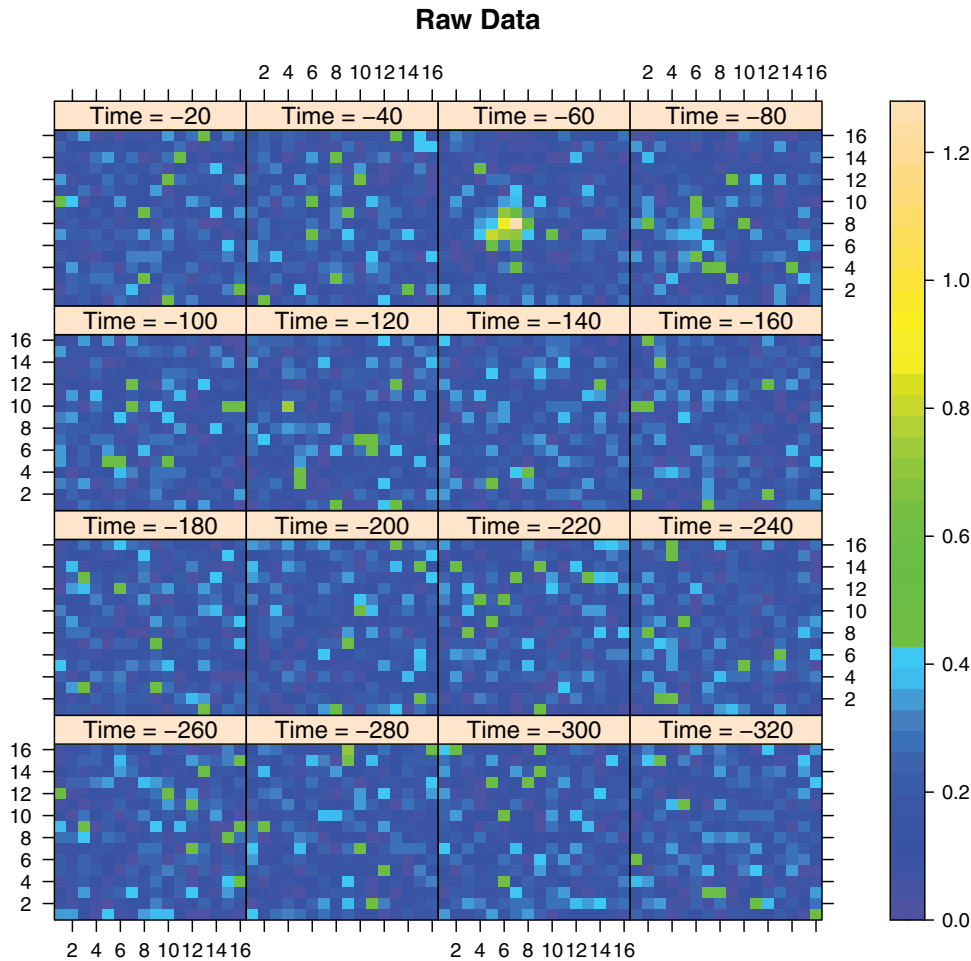


FIGURE 2 For the visual receptive field study: Level plot of the observed ON-RFmaps (firing rates) for the right eye of neuron FAU3

Plugging-in the right-hand side of previous equation into (2), the locally adaptive penalty is expressed as

$$\theta^T \left(\sum_{l=1}^p \xi_l D_q^T \text{diag}(\psi_l) D_q \right) \theta, \tag{4}$$

where $\psi_l = (\psi_l(1), \psi_l(2), \dots, \psi_l(d - q))^T$. Note that, in matrix form, Equation (3) is written as $\lambda = \Psi \xi$, where

$$\Psi = \begin{pmatrix} \psi_1(1) & \dots & \psi_p(1) \\ \vdots & \ddots & \vdots \\ \psi_1(d - q) & \dots & \psi_p(d - q) \end{pmatrix}, \tag{5}$$

is the B-spline design matrix of dimension $(d - q) \times p$; ψ_l in (4) is thus the column l of Ψ .

3.2 | Adaptive penalty in two dimensions

In the two-dimensional case, interest lies in the generalized model

$$g(\mu_i) = f(x_{i1}, x_{i2}), \quad i = 1, \dots, n, \tag{6}$$

where $\mathbf{x}_i = (x_{i1}, x_{i2})^T$ is a two-dimensional covariate vector, and $f(\cdot, \cdot)$ is a smooth and unknown bivariate surface, defined over covariates x_1 and x_2 . When it comes to extend the P-spline principles to two dimensions, we first model the bidimensional surface in terms of B-splines. This is accomplished by the tensor product of two marginal B-splines bases, i.e., $f(x_1, x_2) = \sum_{j=1}^{d_1} \sum_{k=1}^{d_2} \theta_{jk} B_{1j}(x_1) B_{2k}(x_2)$, where $B_{1j}(\cdot)$ and $B_{2k}(\cdot)$ are the marginal B-spline basis functions for, respectively, x_1 and x_2 . In matrix form, model (6) becomes

$$g(\boldsymbol{\mu}) = f(\mathbf{x}_1, \mathbf{x}_2) = (\mathbf{B}_2 \square \mathbf{B}_1) \boldsymbol{\theta} = \mathbf{B} \boldsymbol{\theta}, \tag{7}$$

where $\mathbf{B} = \mathbf{B}_2 \square \mathbf{B}_1 = (\mathbf{B}_2 \otimes \mathbf{1}_{d_1}^T) \odot (\mathbf{1}_{d_2}^T \otimes \mathbf{B}_1)$, and $\mathbf{B}_1 = [b_{1;ij}]$ with $b_{1;ij} = B_{1j}(x_{i1})$, $\mathbf{B}_2 = [b_{2;ik}]$ with $b_{2;ik} = B_{2k}(x_{i2})$, $\mathbf{1}_N$ is a column vector of ones of length N , \otimes denotes the Kronecker product, \odot the element-wise (Hadamard) product, and \square the “box” product (the face-splitting product or row-wise Kronecker product, Eilers et al., 2006; Slyusar, 1999). Finally, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$ and $\boldsymbol{\theta} = (\theta_{11}, \dots, \theta_{d_1 1}, \dots, \theta_{d_1 d_2})^T$. As for the one-dimensional case, in the two-dimensional setting, smoothness is

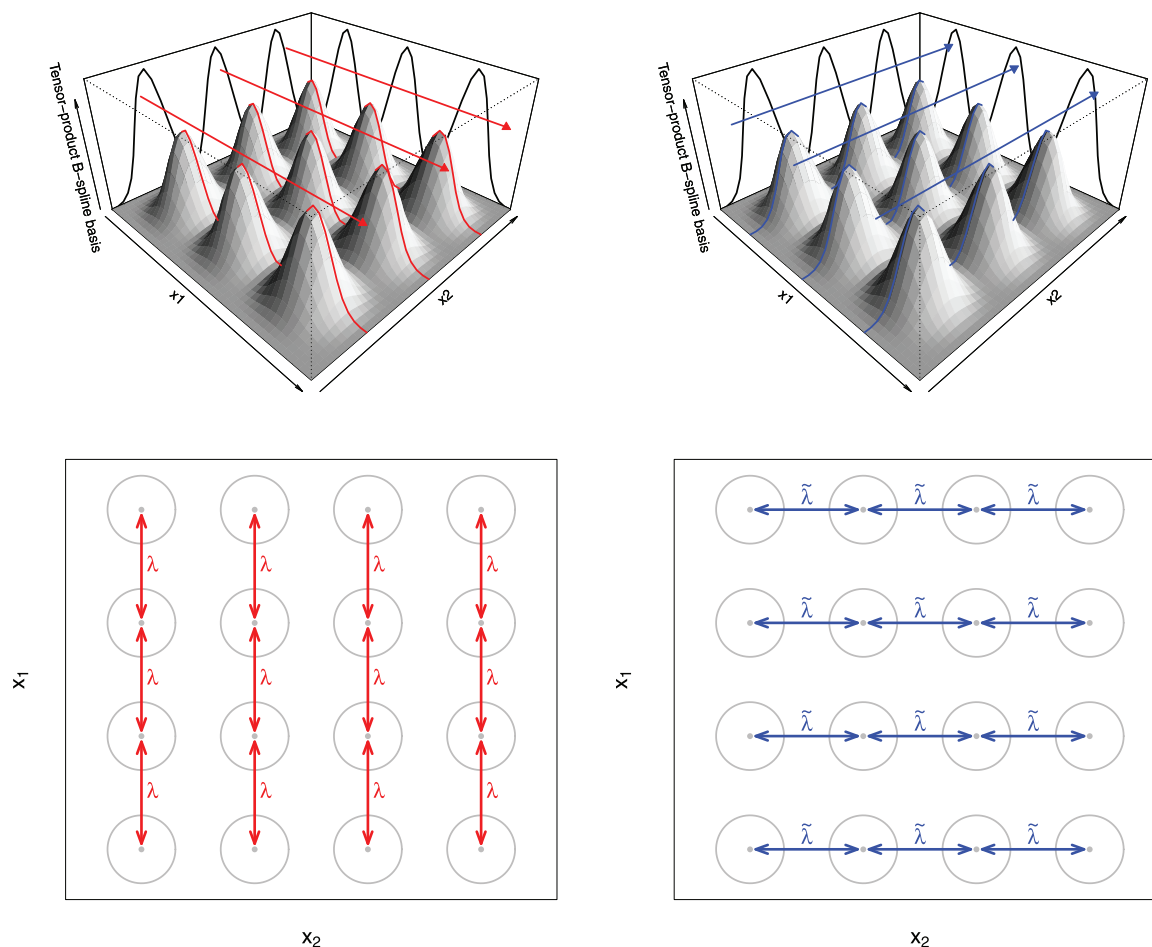


FIGURE 3 Illustration of the tensor product of marginal B-splines basis functions ($B_{1j}(x_1)B_{2k}(x_2)$) and the anisotropic penalty (based on coefficient differences along covariates x_1 and x_2) defined in (8). The top row shows the landscape of nine cubic B-spline tensor products—a portion of a full basis—and highlights why forming coefficient differences along x_1 and x_2 (i.e., on, respectively, the columns and rows of the matrix of coefficients) ensures smoothness along the corresponding covariate. The bottom row schematically illustrates the coefficient differences (arrows) and the smoothing parameters acting on them. In both cases, red is used for covariate x_1 and blue for covariate x_2 .

achieved by penalizing coefficient differences. In particular, the anisotropic penalty in two dimensions is defined as

$$\theta^\top (\lambda (\mathbf{I}_{d_2} \otimes \mathbf{D}_{q_1}^\top \mathbf{D}_{q_1}) + \tilde{\lambda} (\mathbf{D}_{q_2} \mathbf{D}_{q_2}^\top \otimes \mathbf{I}_{d_1})) \theta, \quad (8)$$

where \mathbf{D}_{q_m} ($m = 1, 2$) are difference matrices of possibly different order q_m , and λ and $\tilde{\lambda}$ are the smoothing parameters (Eilers & Marx, 2003). Before proceeding, it is worth seeing the vector θ as a $(d_1 \times d_2)$ matrix of coefficients, $\Theta = [\theta_{jk}]$; the rows and columns of Θ correspond to the regression coefficients in the x_1 and x_2 directions, respectively. Thus, $\theta^\top (\mathbf{I}_{d_2} \otimes \mathbf{D}_{q_1}^\top \mathbf{D}_{q_1}) \theta$ forms (the sum of squares of) differences of order q_1 on each column of the matrix of coefficients Θ ; it is thus responsible, in combination with the smoothing parameter λ , for the smoothness along covariate x_1 . Similarly, $\theta^\top (\mathbf{D}_{q_2} \mathbf{D}_{q_2}^\top \otimes \mathbf{I}_{d_1}) \theta$ forms (the sum of squares of) differences of order q_2 on each row of the matrix of coefficients, controlling, jointly with

$\tilde{\lambda}$, the smoothness along covariate x_2 . Figure 3 gives more insights about (the reasoning behind) the anisotropic penalty just presented, and helps presenting the adaptive penalty in two dimensions that will follow.

By considering two different smoothing parameters λ and $\tilde{\lambda}$ (i.e., anisotropy), the penalty in (8) permits a different amount of smoothing for x_1 and x_2 . However, in some circumstances, this flexibility may not be enough for the model to capture “local” behaviors in the data; all coefficient differences (along x_1 or along x_2) are penalized by the same smoothing parameter (λ or $\tilde{\lambda}$), and thus, the same amount of smoothing is assumed along each covariate. Following the same reasoning as in the one-dimensional case, we propose to overcome the (possible) lack of flexibility of penalty (8) by considering a different smoothing parameter for each coefficient difference, and we do that separately for the coefficient differences along x_1 and along x_2 . The idea is graphically

exemplified in Figure 4, which shows clearly that our approach gives rise to two matrices of smoothing parameters, $\mathbf{\Lambda} = [\lambda_{uv}]$ of dimension $(d_1 - q_1) \times d_2$, and $\tilde{\mathbf{\Lambda}} = [\tilde{\lambda}_{sp}]$ of dimension $d_1 \times (d_2 - q_2)$. Recall that d_m is the dimension of the marginal B-splines bases and q_m is the penalty order ($m = 1, 2$). In particular, the smoothing parameters in $\mathbf{\Lambda}$ act on the coefficient differences along x_1 , and then, control the amount of smoothing along that covariate, but permitting it to vary locally. The same applies to $\tilde{\mathbf{\Lambda}}$, which (adaptively) controls the amount of smoothing along x_2 . With this in mind, our two-dimensional anisotropic adaptive penalty takes the following form:

$$\begin{aligned} &\boldsymbol{\theta}^\top \left((\mathbf{I}_{d_2} \otimes \mathbf{D}_{q_1})^\top \text{diag}(\boldsymbol{\lambda})(\mathbf{I}_{d_2} \otimes \mathbf{D}_{q_1}) \right. \\ &\quad \left. + (\mathbf{D}_{q_2} \otimes \mathbf{I}_{d_1})^\top \text{diag}(\tilde{\boldsymbol{\lambda}})(\mathbf{D}_{q_2} \otimes \mathbf{I}_{d_1}) \right) \boldsymbol{\theta}, \end{aligned} \quad (9)$$

where $\boldsymbol{\lambda} = \text{vec}(\mathbf{\Lambda})$ and $\tilde{\boldsymbol{\lambda}} = \text{vec}(\tilde{\mathbf{\Lambda}})$, and $\text{vec}(\mathbf{M})$ denotes the vectorization of matrix \mathbf{M} . A possible drawback of the adaptive penalty defined in (9) is the number of smoothing parameters involved, which equals to $(d_1 - q_1) \times d_2 + d_1 \times (d_2 - q_2)$; i.e., the total number of coefficient differences along x_1 and x_2 . This may give rise, in addition to undersmoothing and unstable computations, to prohibitively long computing times. We propose to reduce the complexity of the multidimensional adaptive penalty in (9) through a reduction on the number of smoothing parameters. In a similar manner to the one-dimensional case presented in Section 3.1, this is done by considering a smoothed (and smaller) version of them. The underlying assumption is that smoothing parameters that are spatially proximate are more likely to be similar than those farther apart. Before proceeding, note that we now have two matrices of smoothing parameters $\mathbf{\Lambda}$ and $\tilde{\mathbf{\Lambda}}$ (see also Figure 4). It seems then reasonable to smooth them using the tensor product of marginal B-spline bases, and we do it separately for $\mathbf{\Lambda}$ and $\tilde{\mathbf{\Lambda}}$. Taking advantage of the “data” (smoothing parameters) being in an array structure, we write

$$\boldsymbol{\lambda} = (\boldsymbol{\Psi}_2 \otimes \boldsymbol{\Psi}_1) \boldsymbol{\xi}, \quad (10)$$

$$\tilde{\boldsymbol{\lambda}} = (\tilde{\boldsymbol{\Psi}}_2 \otimes \tilde{\boldsymbol{\Psi}}_1) \tilde{\boldsymbol{\xi}}, \quad (11)$$

where $\boldsymbol{\xi} = (\xi_1, \dots, \xi_{p_{11}p_{12}})^\top$ and $\tilde{\boldsymbol{\xi}} = (\tilde{\xi}_1, \dots, \tilde{\xi}_{p_{21}p_{22}})^\top$ are the new vectors of smoothing parameters, and $\boldsymbol{\Psi}_1^{(d_1 - q_1) \times p_{11}}$, $\boldsymbol{\Psi}_2^{d_2 \times p_{12}}$, $\tilde{\boldsymbol{\Psi}}_1^{d_1 \times p_{21}}$, and $\tilde{\boldsymbol{\Psi}}_2^{(d_2 - q_2) \times p_{22}}$ are B-spline design matrices (the superindices indicate their dimension). In particular, these matrices are constructed

as follows:

$$\begin{aligned} \boldsymbol{\Psi}_1 &= \begin{pmatrix} \psi_{11}(1) & \dots & \psi_{1p_{11}}(1) \\ \vdots & \ddots & \vdots \\ \psi_{11}(d_1 - q_1) & \dots & \psi_{1p_{11}}(d_1 - q_1) \end{pmatrix}, \\ \tilde{\boldsymbol{\Psi}}_1 &= \begin{pmatrix} \tilde{\psi}_{11}(1) & \dots & \tilde{\psi}_{1p_{21}}(1) \\ \vdots & \ddots & \vdots \\ \tilde{\psi}_{11}(d_1) & \dots & \tilde{\psi}_{1p_{21}}(d_1) \end{pmatrix}, \\ \boldsymbol{\Psi}_2 &= \begin{pmatrix} \psi_{21}(1) & \dots & \psi_{2p_{12}}(1) \\ \vdots & \ddots & \vdots \\ \psi_{21}(d_2) & \dots & \psi_{2p_{12}}(d_2) \end{pmatrix}, \\ \tilde{\boldsymbol{\Psi}}_2 &= \begin{pmatrix} \tilde{\psi}_{21}(1) & \dots & \tilde{\psi}_{2p_{22}}(1) \\ \vdots & \ddots & \vdots \\ \tilde{\psi}_{21}(d_2 - q_2) & \dots & \tilde{\psi}_{2p_{22}}(d_2 - q_2) \end{pmatrix}. \end{aligned} \quad (12)$$

Plugging-in the right-hand side of Equations (10) and (11) into (9), and after some algebraic operations, we obtain our proposal for the adaptive penalty in two dimensions

$$\begin{aligned} &\boldsymbol{\theta}^\top \left(\sum_{u=1}^{p_{11}p_{12}} \xi_u (\mathbf{I}_{d_2} \otimes \mathbf{D}_{q_1})^\top \text{diag}(\boldsymbol{\psi}_u)(\mathbf{I}_{d_2} \otimes \mathbf{D}_{q_1}) \right. \\ &\quad \left. + \sum_{s=1}^{p_{21}p_{22}} \tilde{\xi}_s (\mathbf{D}_{q_2} \otimes \mathbf{I}_{d_1})^\top \text{diag}(\tilde{\boldsymbol{\psi}}_s)(\mathbf{D}_{q_2} \otimes \mathbf{I}_{d_1}) \right) \boldsymbol{\theta}, \end{aligned} \quad (13)$$

where $\boldsymbol{\psi}_u$ and $\tilde{\boldsymbol{\psi}}_s$ denote, respectively, the columns u and s of $\boldsymbol{\Psi} = \boldsymbol{\Psi}_2 \otimes \boldsymbol{\Psi}_1$ (see (10)) and $\tilde{\boldsymbol{\Psi}} = \tilde{\boldsymbol{\Psi}}_2 \otimes \tilde{\boldsymbol{\Psi}}_1$ (see (11)).

3.3 | Simplifications and generalizations

The two-dimensional adaptive penalty presented in the previous section (expression (13)) is the most general one: A different smoothing parameter is assumed for each coefficient difference along both x_1 and x_2 . However, several simplifications may be made according to the data at hand. These are discussed and presented in detail in the [Supporting Information](#) (Web Appendix B), where we also cover the extension of the adaptive penalty to three dimensions (Web Appendix C).

3.4 | Estimation and inference

To estimate model (6) subject to the adaptive penalty defined in (13) (as well as the simplifications and the extension to the three-dimensional case discussed in the [Supporting Information](#)), we adopt the equivalence between P-splines and generalized linear mixed models for estimation

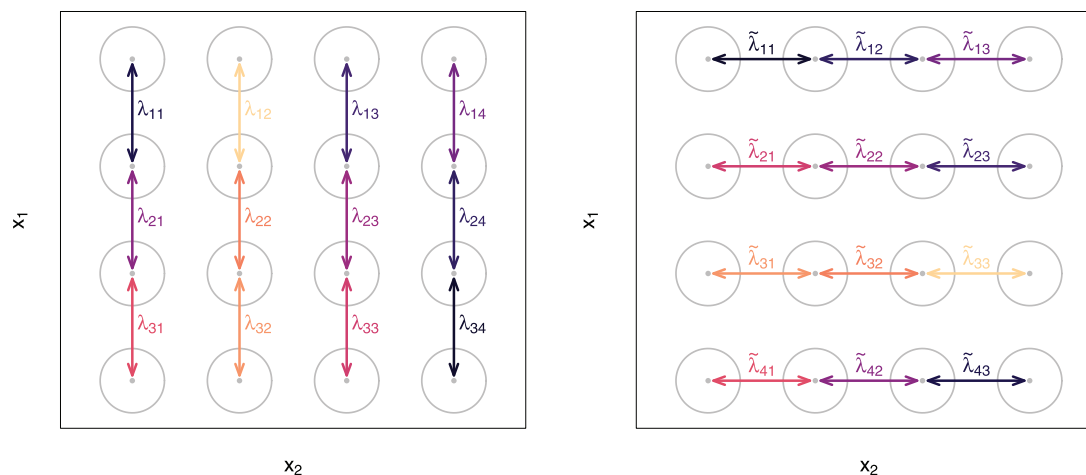


FIGURE 4 Illustration of the adaptive penalty in two dimensions. Separately for x_1 (left-hand side plot) and x_2 (right-hand side plot), each coefficient difference is penalized by a different smoothing parameter. The arrows (and colors) schematically represent different coefficient differences jointly with the smoothing parameters acting on them (λ_{uv} for the coefficient differences along x_1 and $\tilde{\lambda}_{sp}$ for the coefficient differences along x_2).

and inference (e.g., Currie & Durban, 2002; Wand, 2003). For brevity, the technical details and a discussion on computational aspects are given in Web Appendix D, but we note here that the model reparameterization we use gives rise to precision matrices for the random effects that are linear in the inverse of the variance parameters (i.e., the precision parameters). This feature allows using the SOP method (Rodríguez-Álvarez et al., 2019) for estimation. One of the main advantages of SOP is that it is computationally very efficient when it comes to estimating variance parameters; this is essential in our setting where the number of variance parameters (or equivalently smoothing parameters) to be estimated may be very large.

4 | SIMULATION STUDY

This section reports the results of a simulation study conducted to study the empirical performance of the multidimensional adaptive penalties described in Section 3 above. For conciseness, the study concentrates on the two-dimensional case, and only the full adaptive penalty specification is evaluated (expression (13)). We compare the performance of our proposal with that described in Krivobokova et al. (2008) and implemented in the R-package *AdaptFit* (Krivobokova, 2012). Also, a non-adaptive two-dimensional P-spline model (i.e., a model with a standard anisotropic penalty; see (8)) is explored. As for the adaptive case, estimation is based on the SOP method. Comparisons among the approaches are performed in terms of the mean square error (MSE), coverage probabilities and widths of the 95% nominal pointwise confidence intervals for the two-dimensional surfaces, and

computing time. All computations are performed in (64-bit) R 4.0.2 (R Core Team, 2022), and a 2.40GHz \times 4 Intel[®] Core™ i7 processor computer with 15.6GiB of RAM and Ubuntu 16.04 LTS operating system. Plots are generated using the R-package *ggplot2* (Wickham, 2009).

4.1 | Scenarios and setup

Three different scenarios are considered in this study. The first scenario is classical in the context of adaptive P-splines in two dimensions, and it has been considered, among others, by Lang and Brezger (2004), Crainiceanu et al. (2007), and Krivobokova et al. (2008). The second and third scenarios correspond to highly varying functions with sharp peaks, where adaptive smoothing seems a sensible choice. The third scenario has also been discussed in Yue and Speckman (2010). The scenarios considered follow.

Scenario I

$$g(\mu_i) = \eta_i = f(x_{i1}, x_{i2}) = x_{i1} \sin(4\pi x_{i2}), \quad (14)$$

with $x_{i1}, x_{i2} \stackrel{\text{iid}}{\sim} U[0, 1]$. The response data y_i are then generated as

$$y_i = \eta_i + \varepsilon_i \text{ where } \varepsilon_i \stackrel{\text{iid}}{\sim} N(0, s^2) \text{ with } s = \frac{1}{4}(\min(f) - \max(f)). \quad (15)$$

Scenario II

$$\eta_i = f(x_{i1}, x_{i2}) = \exp \left[-15 \left\{ \left(\frac{x_{i1}}{1.2} - 0.5 \right)^2 + \left(\frac{x_{i2} - 30}{50} \right)^2 \right\} \right]. \quad (16)$$

with $x_{i1} \stackrel{\text{iid}}{\sim} U[-5, 1.5]$ and $x_{i2} \stackrel{\text{iid}}{\sim} U[-50, 150]$. Here, y_i is generated under three different distributions

- $y_i = \eta_i + \varepsilon_i$, where $\varepsilon_i \stackrel{\text{iid}}{\sim} N(0, s^2)$ with $s \in \{0.1; 0.5\}$.
- $y_i \stackrel{\text{iid}}{\sim} \text{Poisson}(E_i \exp(\eta_i))$, with $E_i \stackrel{\text{iid}}{\sim} \text{DU}[5, 50]$ (DU denotes a discrete uniform distribution).
- $y_i \stackrel{\text{iid}}{\sim} \text{Bernoulli}(p_i)$, with $p_i = \exp(\tilde{\eta}_i) / \exp(1 + \tilde{\eta}_i)$, where $\tilde{\eta}_i = 7\eta_i - 3$.

Scenario III

$$\eta_i = f(x_{i1}, x_{i2}) = 2 \exp \left[-\frac{1}{0.4} \left\{ (x_{i1} - 2)^2 + (x_{i2} - 2)^2 \right\} \right] + \exp \left[-\frac{1}{3} (x_{i1}^2 + x_{i2}^2) \right], \quad (17)$$

with $x_{i1}, x_{i2} \stackrel{\text{iid}}{\sim} U[-5, 5]$. As for scenario II, y_i is generated according to

- $y_i = \eta_i + \varepsilon_i$, where $\varepsilon_i \stackrel{\text{iid}}{\sim} N(0, s^2)$ with $s \in \{0.1; 0.5\}$.
- $y_i \stackrel{\text{iid}}{\sim} \text{Poisson}(E_i \exp(\eta_i))$, with $E_i \stackrel{\text{iid}}{\sim} \text{DU}[5, 50]$.
- $y_i \stackrel{\text{iid}}{\sim} \text{Bernoulli}(p_i)$, with $p_i = \exp(\tilde{\eta}_i) / \exp(1 + \tilde{\eta}_i)$, where $\tilde{\eta}_i = 3.5\eta_i - 3$.

The true two-dimensional functions used in each scenario are shown in Web Figure 3. For all scenarios, a total of $R = 250$ replicates are performed and we consider $n \in \{300, 500, 1000, 2000\}$. For the P-spline models with and without adaptive smoothing, we use second-order differences ($q_m = 2$) and marginal cubic B-splines bases of dimension $d_m = 12$ (Scenario I) and $d_m = 20$ (Scenarios II and III) to represent the two-dimensional functions ($m = 1, 2$); larger bases dimensions are considered for Scenarios II and III due to the complexity of the simulated functions. For the adaptive approach, we consider the full adaptive penalty and choose $p_{mw} = 5$ ($m, w = 1, 2$; see (10) and (11)). This gives rise to a total of 50 (2×5^2) smoothing parameters (or variance components). For the proposal by Krivobokova et al. (2008) (hereafter denoted as AdaptFit), we use 144 (12×12) knots for the two-dimensional function in Scenario I and 200 (20×20) knots for Scenarios II and III. Also, and for the three scenarios, we consider 25 (5×5) knots for the variance parameters. Here, low-rank radial basis functions are used, and the knots are selected based on the *clara* algorithm by Kaufman and Rousseeuw (1990).

4.2 | Results

For brevity, the majority of graphical and numerical results are provided in Web Appendix E, and we focus here on

the main findings. We note that, for Scenarios II and III, AdaptFit presents severe convergence problems (e.g., for Scenario II—all response distributions—and Scenario III—Poisson and Bernoulli—it does not converge for any run). To avoid misleading conclusions on the performance of AdaptFit, results for these two scenarios are not shown. Figure 5 shows boxplots of $\log(\text{MSE})$ for all the scenarios, response distributions, sample sizes, and noise levels considered in the study. MSE is computed at the observed covariate values, where, for Gaussian data, the true linear predictor (η_i) is chosen as the target, and in the case of Poisson and binary data, it is computed on the response scale. We start with the results for Scenario I. To our surprise, in our study, the best approach is the P-spline model with the standard anisotropic penalty, followed closely by the adaptive approach proposed in this paper. AdaptFit is the one performing the worst. This result somehow contradicts that reported in Lang and Brezger (2004), where the adaptive approach (different to the one proposed here) performs better than the nonadaptive counterpart. We highlight that the results we obtain for AdaptFit are in concordance to those presented in Krivobokova et al. (2008) (median of $\log(\text{MSE})$ of -3.53 and -3.79 , respectively, for $n = 300$), which, in turn, outperform results in both Lang and Brezger (2004) and Crainiceanu et al. (2007). If we focus on the results for Scenarios II and III, we see that the full two-dimensional adaptive approach proposed in this paper is the one that performs the best, for (almost) all response distributions, sample sizes, and noise levels. Differences are particularly remarkable (in favor of the adaptive approach) for moderate to large sample sizes, and, in the Gaussian case, for low noise levels. Results regarding coverage probabilities (averaged over all the surface) are shown in Web Table 1. In general, average coverage probabilities are close to the nominal level except for small sample sizes and large noise levels in the Gaussian case (Scenarios II and III), and for the Bernoulli distribution in Scenario III. In these cases, average coverage probabilities are around 90%. Results for the adaptive and nonadaptive P-spline models are very similar, but at the cost of wider confidence intervals of the nonadaptive approach (see Web Figure 4). Despite results on coverage probabilities being similar on average, when looking at the pointwise coverage probability maps/surfaces (see Web Figures 5–11), we observe that, for Scenarios II and III, the nonadaptive approach presents very low coverage probabilities (in some cases lower than 10%) around the areas where the sharp peaks occur, which is likely where the interest lies (the nonadaptive approach has trouble capturing the sharp peaks, see Web Figures 13–21, and this translates into poor coverage probabilities). Finally, results regarding computing times (in seconds and logarithm scale) are presented in Web Figure 12. As could have been expected, in all cases, fitting the nonadaptive P-spline model is faster than

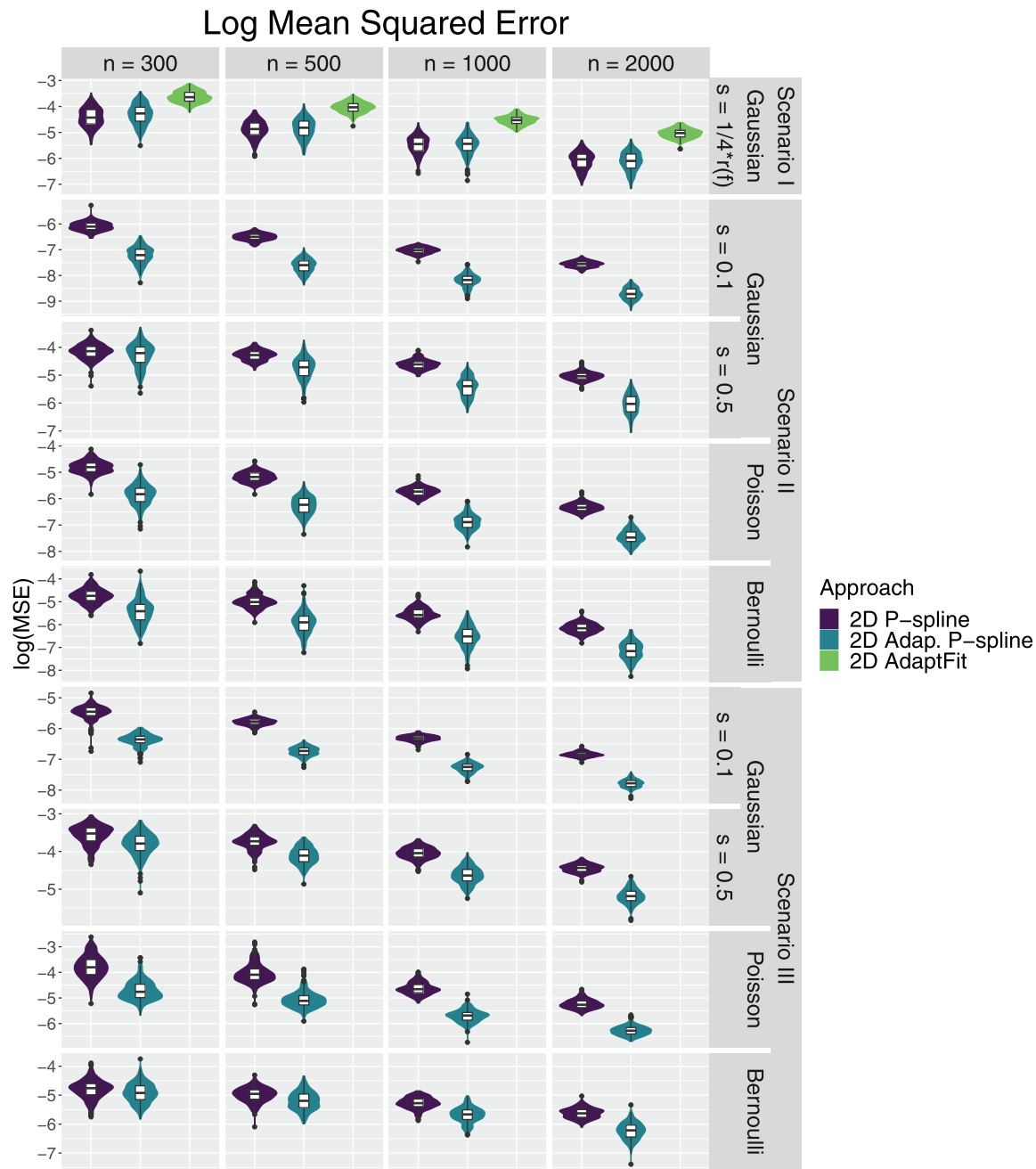


FIGURE 5 For the simulation study: Violin plots of $\log(\text{MSE})$ across $R = 250$ replicates for Scenarios I–III, different response distributions (Gaussian, Poisson, and Bernoulli), levels of noise (s), and sample sizes (n). “2D P-spline” stands for the two-dimensional P-spline model with the standard anisotropic penalty, “2D Adap. P-spline” for the model considering the full adaptive penalty in two dimensions proposed in this paper, and “2D AdaptFit” for the proposal by Krivobokova et al. (2008).

using the adaptive alternatives, but our approach outperforms AdaptFit. To give some numbers, for one of the worst-case situations (Scenario III, $n = 2000$ and Binomial distribution), the standard anisotropic two-dimensional P-spline model requires, in median, around 6.7 s, whereas the proposed adaptive approach requires around 32.6 s. In any case, computing times are more than reasonable. However, we are aware that these values cannot be considered as benchmarks: It is expected that computing times

will increase significantly for other (and more “generous”) bases specifications (see also Web Appendix D for a discussion on computational aspects).

5 | NEURONS’ ACTIVITY STUDY: RESULTS

We now present the results for the study discussed in Section 2. Recall that the objective is to produce smoothed

(de-noised) versions of RFmaps (see Figure 2). To that aim, we adopt a Poisson model that expresses the neuronal response as a smooth function of both space (row and column position) and time

$$\log(\mathbb{E}[y_{rct} | r, c, t]) = \log(n_{rc}\lambda_{rct}) = \log(n_{rc}) + f(r, c, t), \quad (18)$$

where y_{rct} denotes the number of spike occurrences attributed to stimulus presentations at row r and column c of the square area ($r, c = 1, \dots, 16$) for the t th prespike time ($t = -20, \dots, -320$), n_{rc} is the total number of stimulus presentations during the experiment at row r and column c , and λ_{rct} is the intensity parameter (or firing rate). For fitting model (18), we consider the full three-dimensional adaptive penalty (see Web Appendix C). Also, we fit a model with the standard anisotropic penalty (i.e., non-adaptive smoothing). In both cases, we take advantage of the array structure of the data and generalized linear array methods (GLAMs, Currie et al., 2006) are used. Besides, we use second-order differences ($q_m = 2$) and marginal cubic B-splines bases of dimension $d_m = 11$ to represent the spatiotemporal surface. For the adaptive approach, we choose $p_{mw} = 6$ ($m, w = 1, 2, 3$), yielding a total of 648 (3×6^3) smoothing parameters (or variance components). These values are chosen to provide enough flexibility to the models. Figure 6 depicts (for the same neuron, eye, and experimental condition shown in Figure 2) the raw as well as the estimated (denoised) time-series of RFmaps (firing rates) using both approaches. For conciseness, results are not shown for the whole sequence of prespike times, but for five frames, from 20 to 100 ms prespike times (the complete results can be found in Web Appendix F). As noted before, neuronal responses usually produce noisy signals, making it difficult to study the spatial and temporal properties of their RF (see top row of Figure 6). Smoothing with the proposed adaptive approach yields clearly defined structures and time development of the RF (see middle row of Figure 6), whereas smoothing without the adaptive penalty yields a poorly structured RF (see bottom row of Figure 6). More precisely, the results provided by the adaptive approach show that the RFmap begins to be structured about 80 ms prespike and peaks approximately at 60 ms when a clear central area of high values (firing rates) appears that lasts until 40 ms. Outside the time range from -80 to -40 ms, the RFmaps show no structure and a uniform firing rate pattern (see Web Figure 34). In other words, and in concordance with the raw RFmaps, the adaptive approach clearly shows that the spike emission (neuronal response) is coupled with the stimulus from -80 to -40 ms (i.e., the time between sensory stimulus and neuronal response spans from 40 to 80 ms) and that the response is restricted to a small area (this area being the visual RF of the neuron) centered and stabilized over

time in the space covered by the stimulus presentation. By contrast, the smoothed RFmaps provided by the nonadaptive approach show almost no visible structure, neither spatial nor temporal (the approach is not able to recover neither the peak nor the temporal pattern; see also Web Figure 35 and Web Figure 37 where some cross-sections are shown jointly with 95% pointwise confidence intervals). As such, these results do not allow to detect the neuron's RF. Extra results can be found in the Web Appendix F, where we also include (approximate) standard errors for the smoothed RFmaps.

Regarding computing times, in the absence of adaptive smoothing, the SOP method takes about 3 min, which increases to 39 min with the adaptive penalty. In terms of degrees of freedom or effective dimensions (EDs), and despite the too smooth results of the nonadaptive approach, they are lower for the adaptive than for the nonadaptive counterpart (36.4 and 58.4, out of 1331, respectively). Using SOP, the total ED is obtained by adding the EDs associated with each smoothing/variance parameter in the model (plus the dimension of the fixed part). Details can be found in Rodríguez-Álvarez et al. (2019). For (multidimensional) adaptive P-spline models, most of the EDs (or equivalently the variance parameters) are almost zero, meaning a local linear fit (i.e., they “allow” the model to better adapt to the data regardless adaptivity is needed or not). According to the (conditional) Akaike information criterion (cAIC) proposed by Donohue et al. (2011), the best model is the one using the adaptive penalty (15,523.8 vs. 15,650.2 for the standard anisotropic penalty).

6 | DISCUSSION

This paper presents a novel approach for anisotropic multidimensional adaptive penalties in the context of P-spline models. The construction and rationale of the new adaptive penalties are described in full detail for the two-dimensional case, but the extension to three dimensions is also covered. Besides, we discuss several simplifications that can easily be implemented if desired. Model estimation is based on the connection between P-spline models and generalized linear mixed models, and the practical implementation is based on the recently proposed SOP method. However, nothing, in principle, precludes the use of other estimating methods/algorithms, provided that they can deal with precision/penalty matrices that are linear in the precision/smoothing parameters.

Both the simulation study and applications shown in the paper and Supporting Information highlight the gain value of using, when needed, the proposed multidimensional adaptive P-spline models. Also, the simulation study suggests that when adaptive smoothing is not necessary,

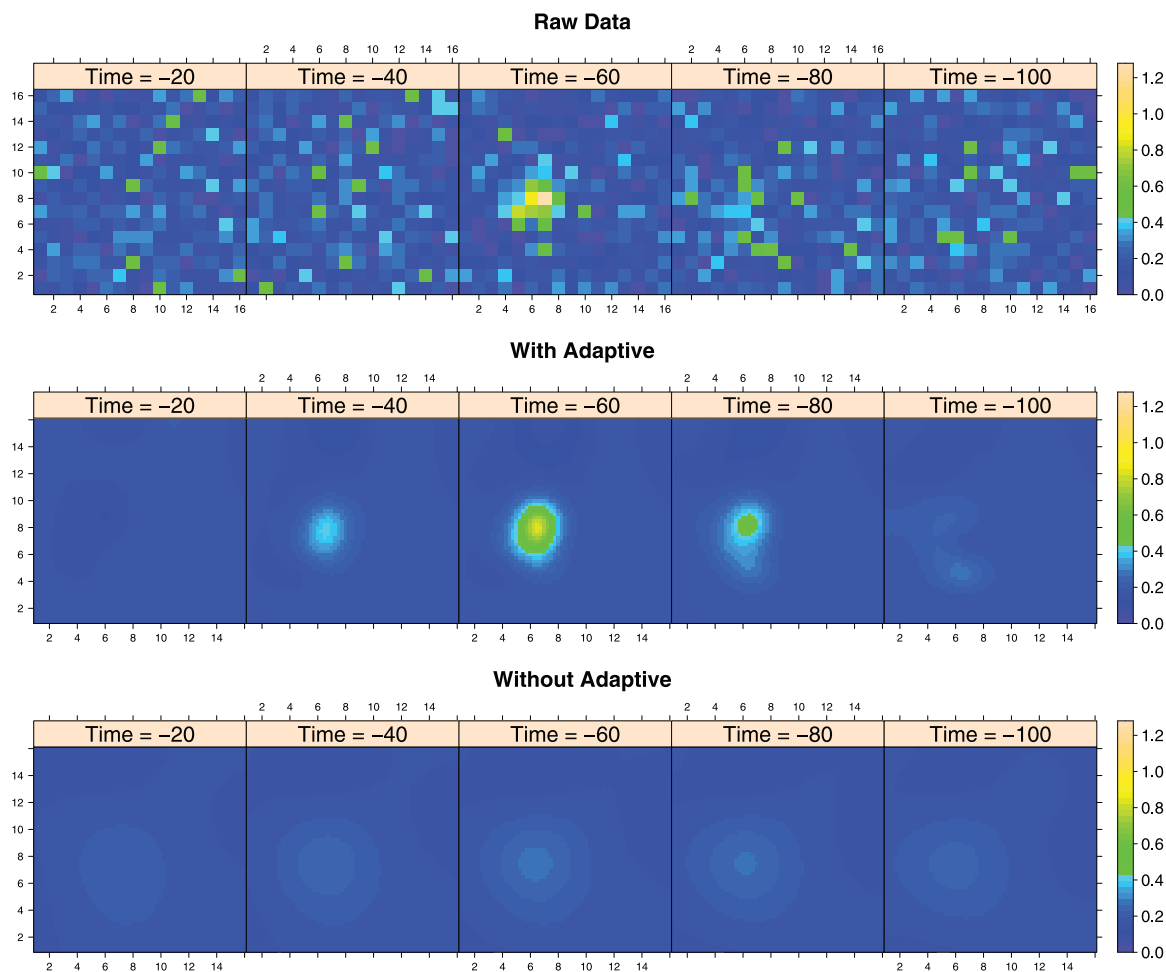


FIGURE 6 For the visual receptive field study: Level plot of the observed and smoothed ON-RFmaps (firing rates) for the right eye of neuron FAU3. First row: observed. Second row: estimates with locally adaptive smoothing. Third row: estimates without locally adaptive smoothing.

the extra flexibility afforded by the adaptive penalties does not translate into an important loss in efficiency. The gain is remarkable for the study that motivated this work. In the absence of adaptivity, results are too smooth and do not appropriately recover the RF of the neuron; this may lead to erroneous conclusions. Although it has not been covered here (accurate), smoothed RFmaps will facilitate the study of important characteristics of RFs. For instance, smoothed RFmaps may help compare the results of bright (“ON”) with the results of dark (“OFF”) spot responses in such a way that overlapping in space and time can be determined. This, in turn, would allow inferring the functional properties of the cell. Also, comparisons between the results obtained from the monocular stimulation of each eye (right and left) could provide information about ocular dominance.

Of course, the flexibility of multidimensional adaptive P-spline models is at the cost of increasing the computing time. For applications of the adaptive approach in two dimensions, computing times are kept to a reasonable

level, even for rather large B-spline bases and number of smoothing parameters. However, the increase is substantial in the three-dimensional case. Even in the nonadaptive case, multidimensional P-splines can be very computationally demanding, especially if large basis dimensions are used. The SOP method allows alleviating the computational cost, which can otherwise be very large using alternative estimating methods. We note that we used the high-level R language to implement the SOP method for the adaptive approaches described in the paper. We expect that low-level languages (e.g., C or Fortran) could further improve computing times. This would also make computationally more feasible the estimation of more complex models, such as those described in Rodríguez-Álvarez et al. (2012), where, for the visual RF study, the authors jointly estimate and compare RFmaps under different experimental conditions.

A limitation of the proposal presented here is that, although relaxed, smoothness is still assumed, for both the functions to be estimated and the smoothing

parameters. Therefore, it is not suitable for functions with abrupt changes or discontinuities (see, e.g., Liu & Guo, 2010, for a one-dimensional adaptive approach in these settings). The study of appropriate multidimensional penalties (and efficient estimation strategies) for these situations represents an interesting topic for research. Finally, it would also be worth exploring the implementation of the approach presented in the paper within a Bayesian hierarchical framework.

ACKNOWLEDGMENTS

This research was funded by projects MTM2017-82379-R (AEI/FEDER, UE) and PID2019-104901RB-I00 (AEI), by the Ramon y Cajal Grant RYC2019-027534-I, by the Basque Government (BERC 2018-2021 program), by the Spanish Ministry of Science, Innovation, and Universities (BCAM Severo Ochoa accreditation SEV-2017-0718), by ISCIII (RETICS RD16/0008/0003), Xunta de Galicia (Accreditation ED431G 2019/02), and the European Regional Development Fund. We are grateful to Vanda Inácio for helpful comments and discussions. We are also grateful to the peer referees and associate editor for their constructive comments of the paper.

Funding for open access charge was provided by Universidade de Vigo/CISUG.

OPEN RESEARCH BADGES



This article has earned Open Data and Open Materials badges. Data and materials are available at <https://bitbucket.org/mxrodriguez/madaptsop>.

DATA AVAILABILITY STATEMENT

The data that support the findings in this paper are available in the [Supporting Information](#) of this article.

ORCID

Maria Xosé Rodríguez-Álvarez  <https://orcid.org/0000-0002-1329-9238>

REFERENCES

- Crainiceanu, C.M., Ruppert, D., Carroll, R.J., Joshi, A. & Goodner, B. (2007) Spatially adaptive Bayesian penalized splines with heteroscedastic errors. *Journal of Computational and Graphical Statistics*, 16, 265–288.
- Currie, I.D. & Durban, M. (2002) Flexible smoothing with P-splines: a unified approach. *Statistical Modelling*, 2, 333–349.
- Currie, I.D., Durban, M. & Eilers, P.H.C. (2006) Generalized linear array models with applications to multidimensional smoothing. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 68, 259–280.
- Donohue, M.C., Overholser, R., Xu, R. & Vaida, F. (2011) Conditional Akaike information under generalized linear and proportional hazards mixed models. *Biometrika*, 98, 685–700.
- Eilers, P.H.C., Currie, I.D. & Durban, M. (2006) Fast and compact smoothing on large multidimensional grids. *Computational Statistics and Data Analysis*, 50, 61–76.
- Eilers, P.H.C. & Marx, B.D. (1996) Flexible smoothing with B-splines and penalties. *Statistical Science*, 11, 89–121.
- Eilers, P.H.C. & Marx, B.D. (2003) Multidimensional calibration with temperature interaction using two-dimensional penalized signal regression. *Chemometrics and Intelligence Laboratory Systems*, 66, 159–174.
- Jang, D. & Oh, H.-S. (2011) Enhancement of spatially adaptive smoothing splines via parameterization of smoothing parameters. *Computational Statistics and Data Analysis*, 55, 1029–1040.
- Kaufman, L. & Rousseeuw, P.J. (1990) *Finding groups in data: an introduction to cluster analysis*. John Wiley, Hoboken, NJ.
- Krivobokova, T. (2012) *AdaptFit: adaptive semiparametric regression*. R package version 0.2-2.
- Krivobokova, T. (2013) Smoothing parameter selection in two frameworks for penalized splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75, 725–741.
- Krivobokova, T., Crainiceanu, C.M. & Kauermann, G. (2008) Fast adaptive penalized splines. *Journal of Computational and Graphical Statistics*, 17, 1–20.
- Krivobokova, T. & Kauermann, G. (2007) A note on penalized spline smoothing with correlated errors. *Journal of the American Statistical Association*, 102, 1328–1337.
- Lang, S. & Brezger, A. (2004) Bayesian P-splines. *Journal of Computational and Graphical Statistics*, 9, 183–212.
- Liu, Z. & Guo, W. (2010) Data driven adaptive spline smoothing. *Statistica Sinica*, 20, 1143–1163.
- O’Sullivan, F. (1986) A statistical perspective on ill-posed inverse problems (with discussion) *Statistical Science*, 1, 505–527.
- R Core Team (2022) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Reiss, P.T., Huang, L., Wu, P.-S., Chue, H. & Colombe, S. (2017) Pointwise influence matrices for functional-response regression. *Biometrics*, 73, 1092–1101.
- Rodríguez-Álvarez, M.X., Cadarso-Suárez, C. & Gonzalez, F. (2012) Analysing visual receptive fields through generalised additive models with interactions. *SORT-Statistics and Operations Research Transactions*, 36, 3–44.
- Rodríguez-Álvarez, M.X., Durban, M., Lee, D.-J. & Eilers, P. H.C. (2019) On the estimation of variance parameters in non-standard generalised linear mixed models: application to penalised smoothing. *Statistics and Computing*, 29, 483–500.
- Ruppert, D. & Carroll, R.J. (2000) Spatially-adaptive penalties for spline fitting. *Australian & New Zealand Journal of Statistics*, 42, 205–223.
- Slyusar, V.I. (1999) A family of face products of matrices and its properties. *Cybernetics and Systems Analysis*, 35, 379–384.
- Wand, M.P. (2003) Smoothing and mixed models. *Computational Statistics*, 18, 223–249.
- Wickham, H. (2009) *ggplot2: elegant graphics for data analysis*. Springer-Verlag, New York.

- Yue, Y.R., Simpson, D., Lindgren, F. & Rue, H. (2014) Bayesian adaptive smoothing splines using stochastic differential equations. *Bayesian Analysis*, 13, 397–424.
- Yue, Y.R. & Speckman, P.L. (2010) Nonstationary spatial Gaussian Markov random fields. *Journal of Computational and Graphical Statistics*, 19, 96–116.

SUPPORTING INFORMATION

Supporting Information S1

An R-package implementing the proposed two- and three-dimensional adaptive P-spline models as well as the R-codes and data needed to reproduce the simulations and applications shown in the paper are available at the

Biometrics website on Wiley Online Library, and they can also be downloaded from <https://bitbucket.org/mxrodriguez/madaptsop>.

How to cite this article: Rodríguez-Álvarez, M.X., Durbán, M., Eilers, P.H.C., Lee, D.-J. & Gonzalez, F. (2022) Multidimensional adaptive P-splines with application to neurons' activity studies. *Biometrics*, 1–14.
<https://doi.org/10.1111/biom.13755>