SORT 36 (1) January-June 2012, 3-32

Analysing visual receptive fields through generalised additive models with interactions

María Xosé Rodríguez-Álvarez[1], Carmen Cadarso-Suárez[2] and Francisco González[3,4]

Abstract

Visual receptive fields (RFs) are small areas of the visual field where a stimulus induces a responses of a particular neuron from the visual system. RFs can be mapped using reverse cross-correlation technique, which produces raw matrices containing both spatial and temporal information about the RF. Though this technique is frequently used in electrophysiological experiments, it does not allow formal comparisons between RFs obtained under different experimental conditions. In this paper we propose the use of Generalised Additive Models (GAM) including complex interactions, to obtain smoothed spatio-temporal versions of RFs. Moreover, the proposed methodology also allow for the statistical comparisons of the RFs obtained across various experimental conditions. Data analysed here derive from studies of neurons' activity in the visual cortex of behaving monkeys. Our results suggest that the GAM-based technique proposed in this paper can be a flexible and powerful tool for assessing receptive field properties.

MSC: 65D07, 65D10, 62P10

Keywords: Visual receptive fields, reverse cross-correlation, visual cortex, smoothing, B-splines, P-splines, tensor product splines.

1. Introduction

One of the techniques used in neurophysiology is electrophysiology, which records the electrical activity produced by neurons. Electrophysiology allows to study the

[1] Clinical Epidemiology and Biostatistics Unit, Complexo Hospitalario Universitario de Santiago de Compostela (CHUS), Spain. maria.jose.rodriguez.alvarez2@sergas.es

[2] Unit of Biostatistics, Dept. of Statistics and OR. Universidade de Santiago de Compostela, Spain.

[3] Department of Phyisiology. School of Medicine. Universidade de Santiago de Compostela, Spain.

[4] Service of Ophthalmology, Complexo Hospitalario Universitario de Santiago de Compostela, Spain.

Received: December 2011

association between sensory stimuli and the neural response in any part of the brain, such as in the visual cortex. Neurons produce sudden changes in their membrane potential known as 'spikes', that can be recorded with microelectrodes. These spikes encode the information produced by the neurons. The analysis of the frequency of spike discharges gives us insights on how the neurons and the nervous system work.

Visual receptive fields (RF) are a basic feature of single cells in the visual system. They are small areas of the visual field that a particular visual neuron 'sees'. They have different sizes depending on the visual area we are considering. From the cell reponses (spikes) we can infer the spatio-temporal properties of the RFs. RFs can be mapped using different techniques. Initially, receptive field maps were manually mapped using simple dots, lines and edges. Despite its simplicity, this early mapping showed that the spatial structure of visual receptive fields as divided into 'on' and 'off' subregions (Barlow, 1953; Hartline and Ratliff, 1958; Hubel and Wiesel, 1962; Kuffler, 1953). Each of these regions responded to the onset of a bright ('on') or dark ('off') spot respectively and was the basis of the circular center-surround organization of visual receptive fields of retinal cells. In the primary visual cortex, Hubel and Wiesel (1962) classified cortical neurons into two groups, simple and complex. Simple cells had receptive fields divided with separate 'on' and 'off' regions, whereas complex cells did not have such division.

Visual cells from area V1 – which is a primary visual cortical area – show ocular dominance. This effect occurs because these cells receive stronger inputs from one eye than from the other. Some cells receive inputs with equal strength from both eyes. Ocular dominance makes that the RF of a given cell can be different depending on what eye is stimulated.

Further quantitative techniques of receptive field mapping have contributed significantly to refine our knowledge about receptive field properties and organization. Post-stimulus responses to flashing or moving bars or dots were used to map receptive fields in the visual system of both cats (Bishop et al., 1973; Henry and Bishop, 1972; Pei et al., 1994) and monkeys (Tsao et al., 2003).

Receptive field mapping techniques that use a reverse correlation analysis have been effective in providing detailed receptive field maps of neurons in early stages of the visual pathway (DeAngelis et al., 1993; Jones and Palmer, 1987; Krause et al., 1987; Reid et al., 1997). Binocular receptive fields have also been mapped using reverse correlation techniques and with stimuli with various binocular disparities (Gonzalez et al., 2001).

Briefly, reverse cross-correlation is a technique that can be used for studying how visual neurons process signals from different positions in their receptive field, and can provide both spatial and temporal information about their RF. A schematic illustration on how the reverse cross-correlation technique was used in our experiments is shown in Figure 1. The animal was viewing two monitors (Figure 1A) with a fixation target. Within a square area over the cell receptive field a bright or dark spot was flashed in different positions in a pseudorandom manner. Cell spikes were recorded while the stimulus was delivered (Figure 1B). When a spike was produced (t_0), the stimulus

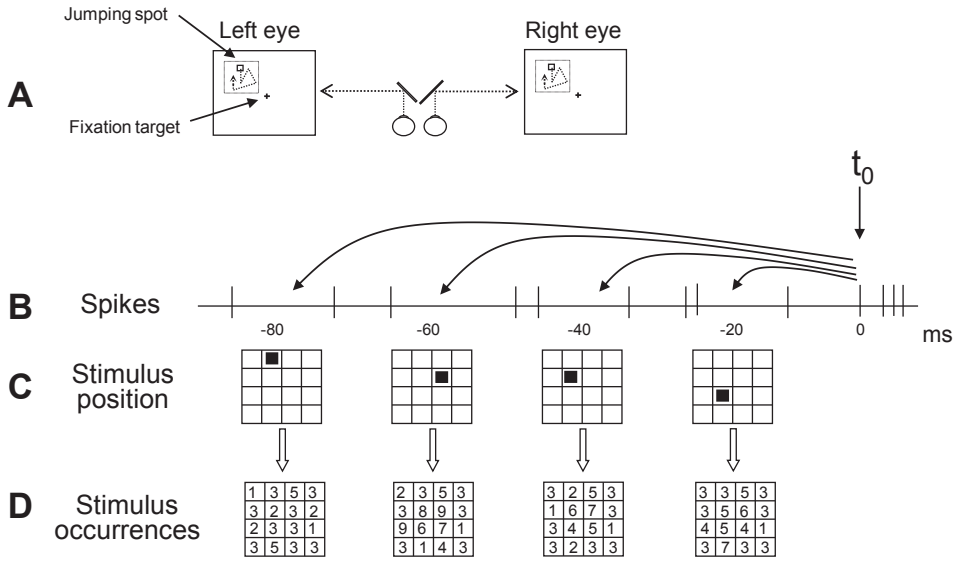


Figure 1: *Reverse cross-correlation technique. The animal was viewing two monitors (A) with a fixation target. Within a square area over the cell receptive field a bright or dark spot was flashed in different positions in a pseudorandom manner. Cell spikes were recorded while the stimulus was delivered (B). When a spike was produced (t_0), the stimulus position at several pre-spike times (-20, -40, …, -320 ms) was read (C) and the corresponding position was increased by one. The result was a numerical matrix containing the number of stimulus occurrences at each position (D).*

position at several pre-spike times $(-20, -40, \ldots, -320$ ms) was read (Figure 1C), and the corresponding position was increased by one. The result was a numerical matrix containing the number of stimulus occurrences at each position (Figure 1D). The graphical representation of this matrix is what we call receptive field map (RFmap).

Although this technique provides raw receptive field maps, it does not allow further aspects of receptive field analysis such as formal comparisons between left and right receptive field maps, 'on' and 'off' maps, or between monocular and binocular receptive field maps.

The main objective of this paper is to use flexible regression models including complex interactions for modelling the visual receptive fields over time, which in turn may vary across various experimental conditions. Specifically, we suggest the use of Poisson Generalised Additive Model (GAM, Hastie and Tibshirani, 1990) which expresses the cell response (i.e., number of spikes) as a smooth function of both space and time, including high-order interactions. Advantages of using this regression model-based approach include the following:

- construction of smoothed versions of RF maps, by including spatial effects in the model.;
- explanation of differences in the course of the cell activity in a unified way.

So far, several ways to smooth cell activity have been studied in the literature. An overview of the application of smoothing techniques in neuronal data can be found in Kass et al. (2003). Cadarso et al. (2006) and Roca-Padiñas et al. (2006, 2011), for example, employ a flexible modeling technique based on the logistic GAM with local linear kernel smoothers. Faes et al. (2008) apply a flexible method based on natural cubic splines to model synchrony in neuronal firing. Other recent techniques in this context include the Bayesian adaptive regression splines (Behseta and Kass, 2005; Behseta et al., 2005; DiMatteo et al., 2001). Flexible regression-based techniques come out favourably since they enjoy the flexibility of capturing the spatio-temporal evolution without the restriction of parametric modeling as well as the possibility to include covariate or factor information. We revisit this aspect in Section 3 where the models discussed happen to share similar properties.

Though GAMs were successfully applied in electrophysiological experiments, no attempt was made so far to use spatio-temporal GAM models when modelling neural data. To this aim, in this paper we follow the GAM approach suggested by Wood (2006a), as a flexible way to model the temporal evolution of visual RFs, across different conditions. The estimation algorithm used for fitting GAMs is based on penalised regression splines (Eilers and Marx, 1996), in combination with B-splines basis functions and the representation of the GAM as a mixed model. This representation is very appealing, since it allows to select the amount of smoothing automatically via restricted (or residual) maximum likelihood (REML). This statistical approach provides a computationally efficient way of estimating the model and making inference when dealing with neural data.

The paper is organized as follows. In Section 2, the electrophysiological experiment is discussed. GAMs including interactions are introduced in Section 3. In Section 4, we present the main results that were obtained for our study. Finally, we point out some conclusions in Section 5.

2. The electrophysiological experiment

2.1. Animal preparation

The experimental setup and physiological recording were reported in detail elsewhere (Gonzalez et al., 1993; Gonzalez and Krause, 1994; Gonzalez et al., 2001). Two monkeys (*Macaca mulatta*) were trained to perform with their head fixed a task that required a steady visual fixation on a small target (0.3×0.2 deg). The behavourial task consisted of a series of trials from 1 to 2 s separated by an intertrial interval of 1 s. Single-cell activity was recorded by means of metal microelectrodes (5 Mohm, AMSystems Inc., Washington, USA) inserted in the brain. For this, a stainless steel cylindrical chamber was attached to the skull covering area V1. To allow access to the visual cortex, small craniotomies of 5 mm diameter were made in the exposed skull within the cham-

ber. Microelectrode penetrations were made by using an electrohydraulic microdriver (Narishige, Japan) mounted on the chamber. Neural voltage signals were amplified and filtered using conventional electronic equipment. A time-amplitude window discriminator (Bak Electronics, Rockville, Maryland, USA) was used to convert the amplified, filtered neural signal to TTL pulses, which were collected by a conventional computer. At the end of the experiments the animal was deeply anaesthetized with nembutal, the brain perfused with 10% formalin and the region where the penetrations were made was blocked and sliced. Sections were stained with toluidine blue and the electrode tracks were reconstructed. Ocular movements were monitored by means of a video camera under infrared illumination. A frame grabber (Imagenation PXC200, Oregon, USA) attached to a conventional personal computer was used to detect the corneal reflex on the left eye and abort the trials when eye movements exceeded a fixation window of 1×1 deg. All surgical procedures were made under general anesthesia (ketamine i.m. 10 mg/kg, followed by sodium pentobarbital i.v. 27 mg/kg). Supplementary pentobarbital was given whenever necessary during the surgical procedure and analgesics and antibiotics (noramidopirine i.m. 150 mg/kg, penicillin i.m. 50000 IU/kg) were given at the end of the surgery. Cleaning and asepsis of the implant was made periodically. All animal procedures were performed in accordance with the guidelines of the Bioethic Committee of our institution.

2.2. Visual stimulation

The presentation of stimuli and the collection, storage and on-line display of analysed data were controlled by five conventional personal computers running software developed in our own laboratory (Gonzalez and Krause, 1994). The animal was placed in front of a two-mirror system allowing simultaneous and separate viewing of two monitors (Model CPD- 520GST, Sony) placed laterally 57.7 cm away from the monkey's eyes subtending 44.8×28 of visual field and set for a resolution of 320×200 pixels (1 pixel = 0.14) and 70 frames per second. Once the assessment of the basic properties of the RF of the cell under study was made, the stimulation procedure to obtain the RF maps (see below) was started.

2.3. Data acquisition

To obtain data to produce the RFmaps we first used a reverse cross-correlation technique (DeBoer and Kuyper, 1968; Jones and Palmer, 1987; Krause et al., 1987; Pérez et al., 2005). For this, we flashed a small bright square (jumping spot), 14.0 cd/m^2) on a grid with 16×16 spatial locations (2.2×2.2 deg). The spot size was adjusted to cover $1/16$th of the grid side and flashed for a duration of $1/70$s (one frame of the monitor) at different grid locations (Pérez et al., 2005). Figure 1 schematically shows this technique. The

background of the monitor screens had a constant luminance of 2 cd/m^2. To obtain 'on' responses, a bright (14.0 cd/m^2) jumping spot was used. To obtain 'off' responses, a dark (0.28 cd/m^2) jumping spot was used. For each successive presentation the location of the spot on the grid was chosen randomly. Each time a neural spike was detected we correlated it with the stimulus position at various prespike times and constructed series of matrices for different pre-spike times. To produce reliable maps, each sequence of stimulation typically required at least 20 presentations of the stimulus on each location of the stimulus grid. For each cell, a series of matrices with 256 grid positions each, covering a prespike time from 20 to 320 ms (at 20 ms interval), were obtained, under different experimental conditions. A local specific coupling between the stimulus and response at a particular prespike time and disparity will produce an emerging set of data with high values on this particular location in the matrix.

3. Statistical methodology

Regression analysis plays a fundamental role in statistics. The objective of this statistical methodology is to evaluate the influence of some explanatory variables (called covariates), $\boldsymbol{x} = (x_1, \ldots, x_p)$, on the mean of a measure of interest y. This relation is given by

$$E[y|x_1, \ldots, x_p] = E[y|\boldsymbol{x}] = m(x_1, \ldots, x_p), \tag{1}$$

where $m(\cdot)$ is a multivariate function, usually denoted as the mean regression function.

The most usual way to model the dependence between the response variable and the covariates is through the multiple linear regression model. In this model, the response y given covariates $\boldsymbol{x}$ is assumed to be normally distributed $\left(y|\boldsymbol{x} \sim N\left(m(\boldsymbol{x}), \sigma^2\right)\right)$, and the covariates are assumed to have a linear effect on the response. Specifically, the following model is assumed

$$E[y|\boldsymbol{x}] = m(x_1, \ldots, x_p) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p,$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^{\mathsf{T}}$ is a vector of unknown regression coefficients.

However, in some circumstances, the assumption of linearity in the effects of the continuous covariates is very restrictive and is not supported by the data at hand. In this setting non-parametric regression techniques are involved in modelling the dependence between y and $\boldsymbol{x}$, but without specifying, in advance, the function $m(\cdot)$ in (1) that relates the covariates and the response. However, if no restrictions are imposed to the funcion $m(\cdot)$, some problems arise. First, fully non-parametric data analysis can be afflicted by the so-called *curse of dimensionality* (Bellman, 1961): as the dimension of the number of covariates increases, it becomes exponentially more difficult to estimate the function $m(\cdot)$. Second, it is difficult to visualize a regression surface $m(\cdot)$ for more than two

covariates. To overcome these difficulties, additive regression models (AMs, Hastie and Tibshirani, 1990) are an alternative to unconstrained non-parametric regression with several covariates. The additive regression model is defined as

$$E[y|\boldsymbol{x}] = m(x_1,\ldots,x_p) = f_1(x_1) + \cdots + f_p(x_p), \tag{2}$$

where f_k $(k = 1,\ldots,p)$ are smooth and unknown functions.

In many applications, the response y can be discrete (e.g. count or binary data). These situations require a different model for the conditional expectation of y, since a direct connection between $E[y|\boldsymbol{x}]$ and the additive predictor $f_1(x_1) + \cdots + f_p(x_p)$ in (2) is no longer possible unless some constraints are imposed. In such cases, the generalised additive models (GAMs) extends the AM by allowing for different distributions of the response variable y. Specifically, GAMs assume that the distribution of the response variable y given covariates $\boldsymbol{x}$ belongs to the exponential family (McCullagh and Nelder, 1989). In these models, the relationship between $E[y|\boldsymbol{x}]$ and the covariates is specified through the following model

$$E[y|\boldsymbol{x}] = g\left(f_1(x_1) + \cdots + f_p(x_p)\right), \tag{3}$$

where $g(\cdot)$ is a monotonic known function (the inverse of link function).

A weakness of the AM and GAM given in equations (2) and (3) is that these models completely ignore the fact that the functional form of a covariate effect often varies according to the values taken by one or more of the remaining covariates. It is not unusual to find situations in which more complex models are needed. This is the case of our real data example, in which we could expect that the response of interest varies smoothly across the visual receptive field (defined in terms of x- and y- coordinates). In recent years, a number of papers have appeared which address the problem of estimating AMs and GAMs with interaction terms. Hastie and Tibshirani (1990) discussed various approaches using smoothing splines. Wahba (1990), among others, proposed the use of smoothing spline ANOVA methods. Roca-Padiñas et al. (2006, 2008) also proposed alternative methods based on kernel-type smoothers. In the context of penalised regression splines, Brezger and Lang (2006), Currie et al. (2006), Eilers and Marx (2003), Lee and Durbán (2011), Wood (2006b), among others, are several references related to multidimensional smoothing. This paper is focused on penalised regression splines, and more specifically on those approaches implemented in the `mgcv` package [Wood (2006a)] of the `R` (R Development Core Team, 2011) statistical software. We begin this section by presenting the idea of penalised spline smoothing in the univariate case and then we address the multivariate and multidimensional regression problem.

3.1. Penalised spline smoothing

For simplicity, we introduce the idea of penalised spline smoothing in the univariate Gaussian case. Specifically, let

$$y_i = f(x_i) + \varepsilon_i, \quad \varepsilon_i \sim N\left(0, \sigma^2\right), \quad i = 1, \ldots, n, \tag{4}$$

where $f(\cdot)$ is a smooth and unknown function of covariate x that needs to be estimated from the data points (x_i, y_i), $i = 1, \ldots, n$.

To estimate the function $f(\cdot)$, it is assumed that this function can be represented as a linear combination of d known basis functions B_j, i.e.

$$f(x) = \sum_{j=1}^{d} \theta_j B_j(x), \tag{5}$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)^{\mathsf{T}}$ is a vector of unknown regression coefficients. Under this representation model (4) is purely parametric, and therefore it can be easily estimated using ordinary least squares

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{B}^{\mathsf{T}}\mathbf{B}\right)^{-1}\mathbf{B}^{\mathsf{T}}\boldsymbol{y},$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)^{\mathsf{T}}$ and $\mathbf{B}$ is the design matrix

$$\mathbf{B} = \begin{pmatrix} B_1(x_1) & B_2(x_1) & \cdots & B_d(x_1) \\ \vdots & \vdots & \ddots & \vdots \\ B_1(x_n) & B_2(x_n) & \cdots & B_d(x_n) \end{pmatrix}.$$

There are several alternatives for the choice of the basis functions B_j $(j = 1, \ldots, d)$ in (5). The simplest basis functions are the polynomials, where $B_j(x) = x^{j-1}$. However, these basis functions have the disadvantage of not being very flexible, and, tend to produce an artificial behaviour at the boundaries. Some alternatives are the so-called *spline* basis, as truncated polynomial, B-splines (de Boor, 2001), or thin plate regression splines (Wood, 2003). For the sake of illustration, this paper is focused on B-splines. In this case, the function $f(\cdot)$ in (5) is specified with respect to a given set of knots

$$x_{min} = k_1 < k_2 < \cdots < k_k = x_{max}$$

placed at equidistant or noequidistant points over the domain of x. The knots divide the domain of x, such that each interval will be covered by $p+1$ B-splines of degree p, and the number of B-splines in (5) is $d = k + p - 1$. Figure 2(a) shows an example

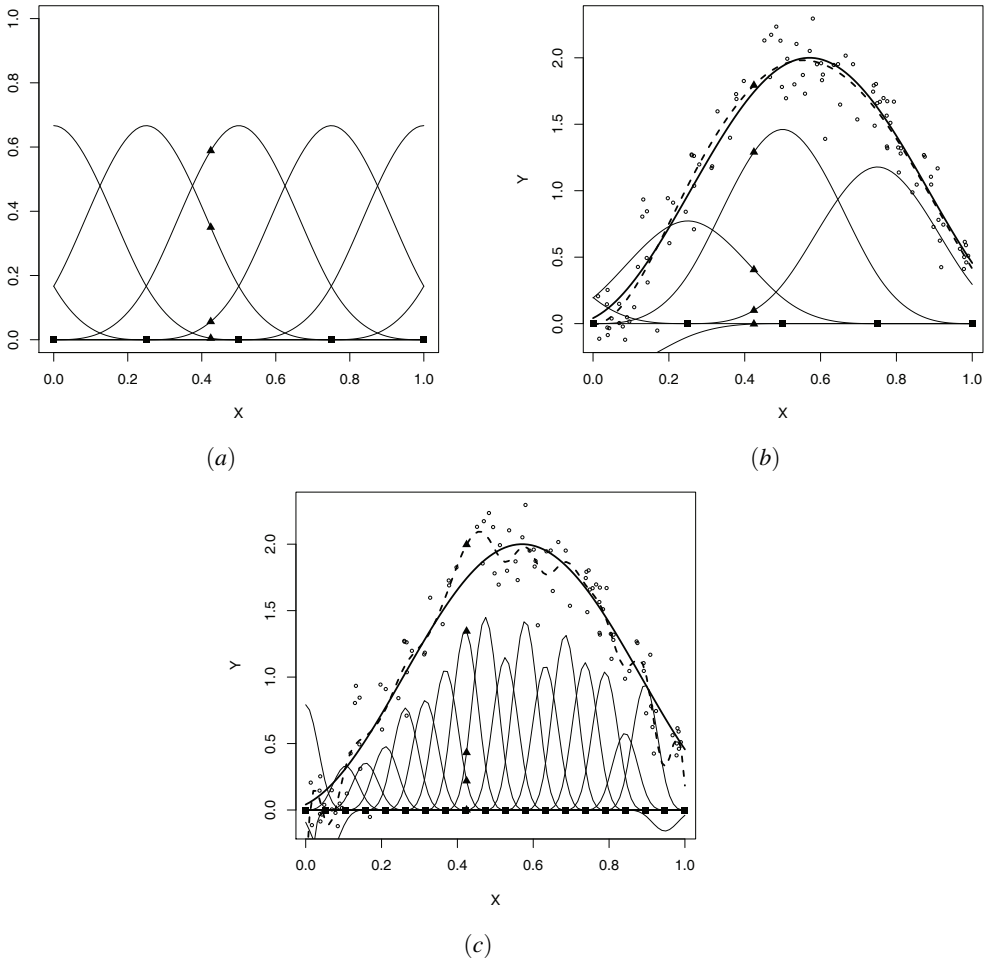


Figure 2: *(a) B-splines basis functions of degree $p = 3$ based on $k = 5$ knots (black squares). The solid triangles indicate the non-zero B-splines at $x = 0.42$. Shown in (b) and (c) are the "weighted" (accordingly to the coefficient estimate $\hat{\boldsymbol{\theta}}$) B-Spline basis function (solid lines), the true function $f(x) = 1 + \sin(5(x+1))$ (bold solid line), and the fitted curve (bold dotted lined) based on $k = 5$ and $k = 20$ knots respectively.*

of B-spline basis functions with $k = 5$ knots and degree $p = 3$. In order to estimate the function $f(\cdot)$ in (5) (or more precisely, the vector $\boldsymbol{\theta}$) we have therefore to specify the number and location of knots and the degree of the B-spline. As an illustration of the impact of the number of knots, we simulated $n = 100$ data from the function $f(x_i) = 1 + \sin(5(x_i+1))$, with $x_i \sim U[0,1]$ and independent $\varepsilon_i \sim N(0, 0.2)$. Figures 2(b) and 2(c) show the estimated curves (via ordinary least squares) with $k = 5$ and $k = 20$ equidistant knots respectively. As can be observed, as the number of knots increases, the estimated curve (in comparison with the true curve) becomes too wiggly, meaning that the data are overfitted. Thus, a wrong choice in the number (and also the location) of knots can lead to estimates, and therefore conclusions, that could be erroneous.

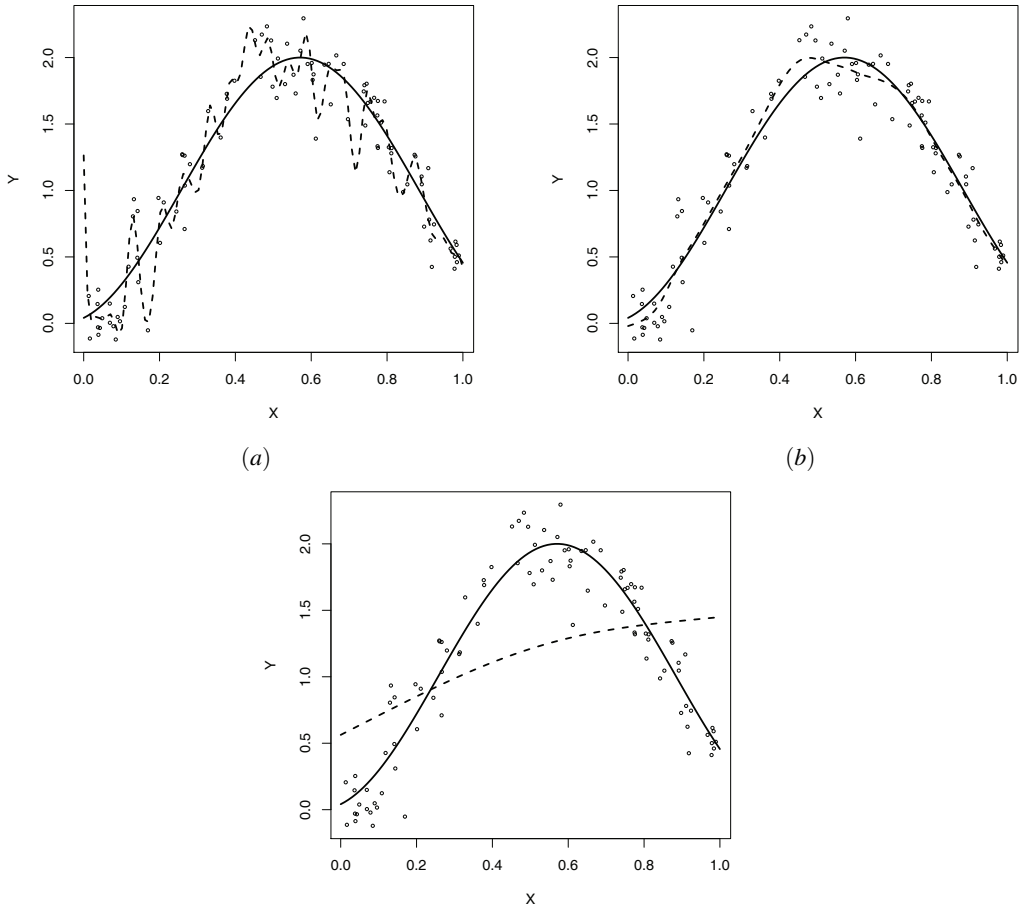


Figure 3: *True function* $f(x) = 1 + \sin(5(x+1))$ *(bold solid line), and the fitted curve (bold dotted lined) based on* $p = 3$ *degree,* $k = 40$ *knots; and* $q = 2$*-nd order derivative: (a)* $\lambda = 10^{-6}$*; (b)* $\lambda = 7$*; and (c)* $\lambda = 10^4$.

Although several approaches have been proposed to select an optimal set of knots (see e.g. Fried and Silverman, 1989; Lee, 2002), all of them have the disadvantage of being computationally intensive. To overcome this problem O'Sullivan (1986), and later Eilers and Marx (1996), introduced the idea of penalised splines, where a smoothness penalty is added to the least squares criterion when estimating the regression coefficients $\boldsymbol{\theta}$ in (5). Although in Eilers and Marx (1996) a discrete penalty is proposed, for the sake of presentation we considered here a penalty based on the integrated derivatives of function f. Specifically, in the case of penalised splines a large amount of knots (e.g. $\min\{n/4, 40\}$) is chosen, and instead of fitting the model by minimising the sum of squares,

$$||\boldsymbol{y} - \mathbf{B}\boldsymbol{\theta}||^2,$$

it is fitted by minimising a *penalised* sum of squares

$$||\boldsymbol{y} - \mathbf{B}\boldsymbol{\theta}||^2 + \lambda \int_{x_{min}}^{x_{max}} [f^q(x)]^2 dx, \tag{6}$$

where the integral of the square of the q-th order derivative, f^q, penalises models that are too wiggly. The smoothing parameter λ controls the trade-off between the bias and the variance of the resulting estimates. The result of a using large smoothing parameter ($\lambda \to \infty$) is an oversmoothed curve, leading in the limit to the least squares $(q-1)$-order polynomial through the data. On the other hand, a small smoothing parameter ($\lambda \to 0$) tends reproduce the data. This can be observed in Figure 3, where the data presented in Figure 2 were re-analysed but incorporating the penalisation. Thus, the optimal amount of smoothing λ has to be chosen by compromising goodness of fit with complexity of the estimated function. This issue is the subject of the next subsection.

It sould be noted that since the function $f(\cdot)$ is linear in the coefficients θ_j (see (5)), the penalty can also be writen as (see, for instance, Marra and Radice, 2010)

$$\int_{x_{min}}^{x_{max}} [f^q(x)]^2 dx = \boldsymbol{\theta}^\top \mathbf{K} \boldsymbol{\theta}, \tag{7}$$

where $\boldsymbol{K}$ is a known $d \times d$ matrix, whose elements depend on the chosen spline basis. Accordingly, for a given λ, the minimiser of (6) is then

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{B}^\top \mathbf{B} + \lambda \mathbf{K}\right)^{-1} \mathbf{B}^\top \boldsymbol{y}.$$

In the case of B-splines, the integrated square of the q-th order derivative of function $f(\cdot)$ can be well approximated by differences of the sequence of regression coefficients $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$. Accordingly, the smoothness penalty presented in (7) is equivalent to impose a penalty on q-th order differences of adjacent B-Spline coefficients. For a more detailed review of this topic, see Eilers and Marx (1996).

3.1.1. Smoothing parameter selection

As pointed out before, a crucial point in penalised spline smoothing is the choice of the smoothing parameter λ in (6). If a large smoothing parameter is chosen, the resulting curve estimate is very smooth, but if a small smoothing parameter is chosen the resulting estimate becomes too wiggly. Therefore, it is important to have procedures for helping in the selection of the optimal smoothing parameter. In this section, we review several methods to choose the 'optimal'value of λ.

(Generalised) Cross Validation When using cross validation (CV) to select the the optimal smoothing parameter, the objective is to obtain the λ value which minimises the

cross validation expression

$$CV(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{f}^{-i}(x_i)\right)^2,$$

where $\hat{f}^{-i}$ denotes the fit obtained by leaving out the ith data point.

As can be observed in the expression above, CV implies to fit n different models (as many as the number of observations), which is computationally very intensive. Fortunately, a more efficient equivalent expression can be obtained Hastie and Tibshirani (1990):

$$CV(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{f}(x_i)}{1 - h_{ii}}\right)^2,$$

where h_{ii} are the diagonal elements of the *hat matrix*

$$\mathbf{H} = \mathbf{B}\left(\mathbf{B}^{\mathsf{T}}\mathbf{B} + \lambda\mathbf{K}\right)^{-1}\mathbf{B}^{\mathsf{T}},$$

that is the matrix such that $\hat{\boldsymbol{y}} = \hat{f}(\boldsymbol{x}) = \mathbf{H}\boldsymbol{y}$. It should be noted that the trace of the hat matrix $\mathbf{H}$ define the *effective degrees of freedom* (edf) (i.e. the 'effective' number of parameters) of a smoother (see Hastie and Tibshirani, 1990; Wood, 2006a).

However, the CV criterion suffers from several drawbacks (see Wahba, 1990; Wood, 2006a). For instance: (a) in the case of more than one smooth function, it becomes computationally expensive; and (b) it presents a lack of invariance. A modified version of this criterion, the generalised cross validation (GCV), was suggested by Craven and Wahba (1979), and presents some advantages over CV (see Craven and Wahba, 1979; Wahba, 1990). The GCV is defined as:

$$GCV(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{f}(x_i)}{1 - \sum_{j=1}^{n} h_{jj}/n}\right)^2.$$

An efficient implementation – in the multivariate additive case – of the GCV criterion (Wood, 2000) was the origin of the `mgcv` package, and for this reason the package deserves its name (multivariate generalised cross validation). Later, Wood (2004) proposed an alternative method to the Wood (2000) which overcame some of its limitations. Primarily the new method is particularly robust numerically, and can deal with rank deficiency in the model. This is, by now, the default method (for the additive case) to choose the smoothing parameters in the `mgcv` package.

Akaike Information Criterion A common approach for the λ selection is to optimise criteria such as the Akaike's information criterion (AIC). In this case, the optimal

smoothing parameter λ is that which minimises

$$AIC(\lambda) = \text{dev}(\boldsymbol{y}; \boldsymbol{\theta}; \lambda) + 2\text{trace}(\mathbf{H}),$$

where dev denotes the deviance of the model and $\mathbf{H}$ is the hat matrix.

In the `mgcv` package, the AIC, or more specifically a rescaled AIC – the unbiased risk estimator (UBRE) (Craven and Wahba, 1979) – is the default method when the scale parameter of the exponential family to which the response y pertains is known (for instance, in the binomial and Poisson cases) (see Wood, 2004, 2008).

Mixed Model Representation A different approach to choose the optimal smoothing parameter λ comes from the fact that a P-Spline regression model can be formulated as a linear mixed model (Brumback et al., 1999; Currie and Durbán, 2002). Although a detailed presentation of this approach is beyond the scope of this paper, we briefly describe here the main ideas behind it.

In this approach, the design matrix $\mathbf{B}$ and the vector of regression coefficients $\boldsymbol{\theta}$ are reformulated is such a way that

$$\boldsymbol{y} = \mathbf{B}\boldsymbol{\theta} + \boldsymbol{\varepsilon} = \mathbf{U}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{u} + \boldsymbol{\varepsilon}, \quad \text{with} \quad \boldsymbol{u} \sim N(0, \mathbf{G}) \quad \text{and} \quad \boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathbf{I}_n)$$

where $\mathbf{I}_n$ is the identity matrix, $\mathbf{U}$ and $\mathbf{Z}$ are the model matrices, and $\boldsymbol{\beta}$ and $\boldsymbol{u}$ are the fixed and random effects coefficients of the linear mixed model respectively. The random effects have covariance matrix $\mathbf{G}$, which depends on the variance σ_u^2, $\mathbf{G} = \sigma_u^2 \mathbf{I}_K$. Under this new configuration, it can be shown that λ is given by the ratio of the variance components, i.e., $\lambda = \frac{\sigma^2}{\sigma_u^2}$. Accordingly, the estimation of the P-Spline model can be obtained using standard procedures for the estimation of a linear mixed model, and the choice of the smoothing parameter becomes the estimation of the variance components (either via *maximum likelihood* (ML) or *restricted (or residual) maximum likelihood* (REML)).

Recently, Wood (2011) presented a computationally efficient way of estimating the smoothing parameters of P-splines models in which a Laplace approximation is used to obtain an approximate REML or ML. These methods are already implemented in the `gam()` function of the `mgcv` package.

Simulation results presented in Wood (2011) and Strasak et al. (2011) suggest that REML and ML methods offer some improvement in terms of the mean-square error and the stability of the estimator relative to GCV or AIC in most cases. Moreover, Reiss and Odgen (2009) also show that at finite sample sizes GCV or AIC are prone to undersmoothing and are more likely to develop multiple minima than REML. For all these reasons, the use of the REML method is recommended.

3.1.2. Additive models and identifiability constraints

In some circumstances, the objective of a real data analysis could be to jointly evaluate the effect of two (or more) covariates, x_1 and x_2, on the response of interest y. In this case, an appropiate model could be the addivive model

$$y_i = f_1(x_{i1}) + f_2(x_{i2}) + \varepsilon_i, \quad \varepsilon_i \sim N\left(0, \sigma^2\right), \quad i = 1, \ldots, n, \tag{8}$$

where $f_1(\cdot)$ and $f_2(\cdot)$ are smooth and unknown functions.

However, this model presents an identifiability problem due to the fact that it incorporates more than one covariate. We could subtract a constant c of any smooth function ($f_1(x_1) - c$), and add it to another one ($f_2(x_2) + c$), and the same regression model would be obtained. To avoid these free constants, it is necessary to impose some restrictions (that is, among all equivalent models, we have to choose one). In this case, the usual way to proceed to guarantee the identification of the model is to incorporate a constant α, and to "centre" each of the smooth functions in some way, for example by assuming:

$$\sum_{i=1}^{n} f_1(x_{1i}) = 0 \quad \text{and} \quad \sum_{i=1}^{n} f_2(x_{2i}) = 0,$$

yieding the model

$$y_i = \alpha + f_1(x_{i1}) + f_2(x_{i2}) + \varepsilon_i, \quad \varepsilon_i \sim N\left(0, \sigma^2\right), \quad i = 1, \ldots, n. \tag{9}$$

Once the identifiability problem has been solved, each of the smooth functions in (9) can be represented using regression splines, and the model can be estimated using penalised least squares (or the mixed model approach) as in the univariate case, with the smoothing parameters of each of the smooth functions being selected via any of the criteria previously presented.

3.1.3. Generalised additive models

The P-Spline methodology presented in the previous Sections can be easily extended to deal with a non-gaussian response y, given that the distribution of y conditional on the covariates $(x_1, \ldots, x_p)$ belongs to the exponential family. In this case, model (3) can be estimated on the basis of the penalised log-likelihood by means of penalised iterative re-weighted least squares (P-IRLS) (see Eilers and Marx, 1996; Wood, 2006a) or by its representation as a generalised linear mixed model (GLMM) (Breslow and Clayton, 1993).

3.2. Factor-by-curve interaction

In this section we are concerned with situations where the effect of a continuous covariate x on the response varies across groups defined by the levels of a categorical variable z (for example, a factor with M levels $\{1,\ldots,M\}$).

To face such situations, *varying coefficient* terms (Hastie and Tibshirani, 1993) can be used. Here, the effect of x is assumed to vary over the range of z, and the regression model then becomes

$$y_i = \alpha + \sum_{l=2}^{M} \alpha_l \mathbf{1}_{\{z_i=l\}} + \sum_{l=1}^{M} \mathbf{1}_{\{z_i=l\}} f_l(x_i) + \varepsilon_i, \tag{10}$$

where $\mathbf{1}_A$ denotes the indicator function of event A (used to construct the *dummy* variables).

As can be observed in equation (10), the model assumes a different effect of covariate x for each of the levels of the categorical covariate z. The inclusion of the "main" effect of z $\left(\sum_{l=2}^{M} \alpha_l \mathbf{1}_{\{z_i=l\}}\right)$ is needed due to identifiability contraints (given that each of the smooth functions $f_l(\cdot)$ is centred). Again, each function $f_l(\cdot)$ in (10) can be represented using regression splines, and the model can then be fitted using penalised least squares or the mixed model approach (or by P-IRLS and GLMM in the generalised case).

3.3. Continuous bivariate interactions

In many applications, the additive structure of the model presented in (8) is not appropriate for the data at hand, and more complex models are needed. In this section we are concerned with the case where the response y is expressed on the basis of a bivariate surface defined by covariates x_1 and x_2:

$$y_i = f_{12}(x_{i1}, x_{i2}) + \varepsilon_i, \quad \varepsilon_i \sim N\left(0, \sigma^2\right), \quad i = 1,\ldots,n, \tag{11}$$

To date, several P-spline approaches to estimate model (11) have been suggested in the statistical literature. For instance, Wood (2003) proposed to model the surface $f_{12}(x_1, x_2)$ by means of two dimensional thin-plate regression splines. However, this approach presents the drawback that only one smoothing parameter λ is incorporated into the smoothness penalty (i.e., an isotropic penalty), meaning that the same smoothness is assumed for both covariates x_1 and x_2. Whereas this isotropy could be justified when modelling, for instance, a smooth function of latitude and longitude, this is not always the case when the covariates x_1 and x_2 are measured in different units. As an alternative, the tensor product of one-dimensional spline basis functions has been suggested (see Eilers and Marx, 2003; Wood, 2006b). In this case, univariate functions f_1 and f_2 are associated with x_1 and x_2 respectively, each of which is represented by

a regression spline:

$$f_1(x_1) = \sum_{j=1}^{d_1} \theta_j^1 B_j^1(x_1) \quad \text{and} \quad f_2(x_2) = \sum_{k=1}^{d_2} \theta_k^2 B_k^2(x_2).$$

Accordingly, the bivariate surface is then defined as:

$$f_{12}(x_1,x_2) = \sum_{j=1}^{d_1}\sum_{k=1}^{d_2} \theta_{jk} B_j^1(x_1) B_k^2(x_2) = \sum_{j=1}^{d_1}\sum_{k=1}^{d_2} \theta_{jk} B_{jk}(x_1,x_2),$$

where the two-dimensional basis functions are given by the tensor product (see also Figure 4)

$$B_{jk}(x_1,x_2) = B_j^1(x_1) B_k^2(x_2).$$

Thus, model (11) can be expressed in matrix notation as:

$$\boldsymbol{y} = \mathbf{B}_{12}\boldsymbol{\theta}_{12} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\theta}_{12} = (\theta_{11}, \ldots, \theta_{d_1 1}, \ldots, \theta_{d_1 d2})^\mathsf{T}$ and

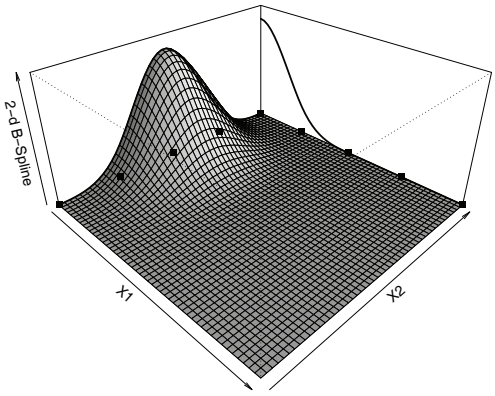
$$\mathbf{B}_{12} = \begin{pmatrix} B_{11}(x_{11},x_{12}) & \cdots & B_{d_1 1}(x_{11},x_{12}) & \cdots & B_{d_1 d_2}(x_{11},x_{12}) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ B_{11}(x_{n1},x_{n2}) & \cdots & B_{d_1 1}(x_{n1},x_{n2}) & \cdots & B_{d_1 d_2}(x_{n1},x_{n2}) \end{pmatrix}.$$

In this situation, how could be defined the penalty for the tensor product? First, it can be useful to view the vector of regression coefficients $\boldsymbol{\theta}_{12}$ as a two dimensional array:

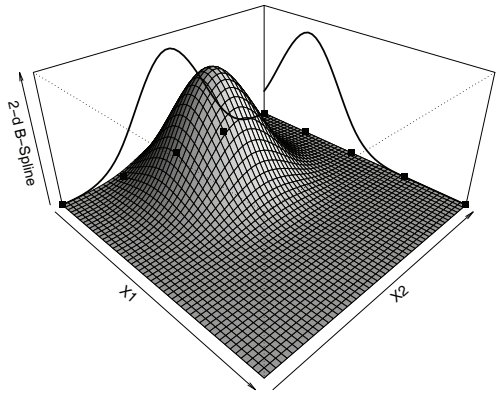
$$\boldsymbol{\Theta}_{12} = \begin{pmatrix} \theta_{11} & \cdots & \theta_{d_1 1} \\ \vdots & \ddots & \vdots \\ \theta_{1 d_2} & \cdots & \theta_{d_1 d_2} \end{pmatrix}.$$

Therefore, whereas the rows of $\boldsymbol{\Theta}_{12}$ correspond to the regression coefficients in the x_1 direction (see also Figure 4), the columns correspond to the x_2 direction. It then seems reasonable to consider the penalty for the bivariate surface separately for the rows and columns of $\boldsymbol{\Theta}_{12}$ (i.e., by considering separate penalties in the x_1 and x_2 directions). This leads to the penalty term in two dimensions (see Currie et al., 2006; Wood, 2006a, among others):

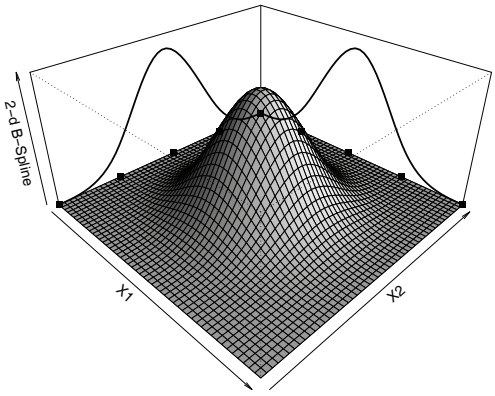
$$\lambda_1 \boldsymbol{\theta}_{12}^\mathsf{T} (\mathbf{I}_{d2} \otimes \mathbf{K}_1) \boldsymbol{\theta}_{12} + \lambda_2 \boldsymbol{\theta}_{12}^\mathsf{T} (\mathbf{K}_2 \otimes \mathbf{I}_{d1}) \boldsymbol{\theta}_{12},$$



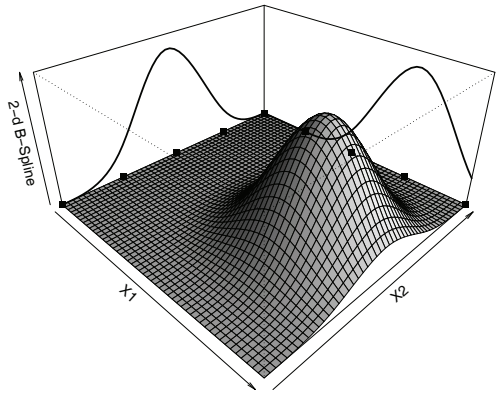
(a) $B_{24}(x_1,x_2)$



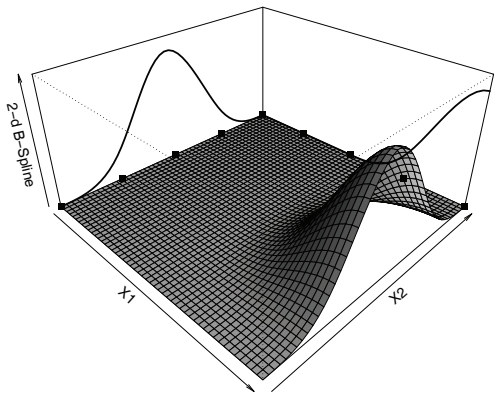
(b) $B_{34}(x_1,x_2)$



(c) $B_{44}(x_1,x_2)$



(d) $B_{54}(x_1,x_2)$



(e) $B_{64}(x_1,x_2)$

Figure 4: *Tensor product of univariate B-Spline basis functions of degree $p=3$ and $k=5$ knots. The figures show the tensor products obtained by varying the univariate B-Spline basis function along the x_1 direction $(B_j^1(x_1), j=2,\ldots,6)$ while maintaining fixed the B-Spline basis function in the x_2 direction $(B_4^2(x_2))$, i.e. $B_{j4}(x_1,x_2)=B_j^1(x_1)B_4^2(x_2), j=2,\ldots,6$.*

where $\otimes$ denotes the kronecker product, and $\mathbf{K}_1$ and $\mathbf{K}_2$ are univariate penalty matrices of dimension $d_1 \times d_1$ and $d_2 \times d_2$ respectively.

Parameters can now be estimated from the penalized least squares

$$||\boldsymbol{y} - \mathbf{B}_{12}\boldsymbol{\Theta}_{12}||^2 + \lambda_1 \boldsymbol{\theta}_{12}^{\mathsf{T}} (\mathbf{I}_{d2} \otimes \mathbf{K}_1)\, \boldsymbol{\theta}_{12} + \lambda_2 \boldsymbol{\theta}_{12}^{\mathsf{T}} (\mathbf{K}_2 \otimes \mathbf{I}_{d1})\, \boldsymbol{\theta}_{12},$$

or based on the mixed model representation.

3.4. Extension to the inclusion of higher order interactions

The methodology presented in Subsections 3.2 and 3.3 can also be extended to the inclusion of higher order interactions. For instance, in the analyses of visual receptive fields, we are interested on incorporating the time dimension to study the behaviour of the stimulus occurrences along pre-spike times. Thus, we are concerned on modelling a function of the form:

$$f_{12t}(x_1, x_2, t)\,.$$

Moreover, in order to evaluate the impact of different experimental conditions on cell response, we might need to incorporate this information into the regression model, and to allow for the previous three-dimensional function to vary across these experimental conditions ($l = 1, \ldots, M$):

$$\alpha + \sum_{l=2}^{M} \alpha_l \mathbf{1}_{\{z_i = l\}} + \sum_{l=1}^{M} \mathbf{1}_{\{z_i = l\}} f_{12t}^{l}(x_{i1}, x_{i2}, t_i)\,.$$

Based on the results presented in Subsection 3.3, $f_{12t}(x_1, x_2, t)$ (or, equivalently, $f_{12t}^{l}(x_{i1}, x_{i2}, t_i)$) can be represented by the tensor product of three univariate basis functions

$$f_{12t}(x_1, x_1, t) = \sum_{j=1}^{d_1} \sum_{k=1}^{d_2} \sum_{m=1}^{d_t} \theta_{jkm} B_j(x_1) B_k(x_2) B_m(t) = \sum_{j=1}^{d_1} \sum_{k=1}^{d_2} \sum_{m=1}^{d_t} \theta_{jkm} B_{jkm}(x_1, x_2, t)\,,$$

and the penalty term – organising the coefficients in some appropriate order – is now defined as

$$\lambda_1 \boldsymbol{\theta}_{12t}^{\mathsf{T}} (\mathbf{I}_{d_2} \otimes \mathbf{K_1} \otimes \mathbf{I}_{d_t})\, \boldsymbol{\theta}_{12t} + \lambda_2 \boldsymbol{\theta}_{12t}^{\mathsf{T}} (\mathbf{K_2} \otimes \mathbf{I}_{d_1} \otimes \mathbf{I}_{d_t})\, \boldsymbol{\theta}_{12t} + \lambda_t \boldsymbol{\theta}_{12t}^{\mathsf{T}} (\mathbf{I}_{c_2} \otimes \mathbf{I}_{c_1} \otimes \mathbf{K}_t)\, \boldsymbol{\theta}_{12t}.$$

4. Application to visual receptive fields

The dataset consists of a series of 16 matrices – with 256 grid positions each – of the form presented in Figure 1D, containing the number of stimulus occurrences at each position. Each matrix corresponds to the different pre-spike times considered (between -20 to -320 ms). The aim of the study was to obtain smooth RFmaps for a given pre-spike time, and to compare the obtained RFmaps between different experimental conditions. For illustration purposes, in the analyses presented in this paper we have considered the response to the onset of a bright spot ('ON') of two different cells (denoted by FAP0 and FBH4) of one monkey, and compared the RFmaps for the right and left eye.

4.1. Data analysis

We adopted a Poisson model with mean $n_{ijk}\lambda_{ijkt}$, where i indicates the row of the matrix, j the column ($i, j = 1, \ldots, 16$), k the eye (either left -0- or right -1-), and t the pre-spike time ($t = -20, \ldots, -320$). n_{ijk} denotes the number of stimulus presentations on each particular grid position. We considered different models for the intensity parameter (or firing rate) λ_{ijkt}, from the simplest model (and additive model) to the most complex (including the interaction between the row, column, time and eye). Specifically the following models were considered:

Model I

$$log\left(\lambda_{ijkt}\right) = \alpha_0 + \alpha_1 \mathbf{1}_{\{k=1\}} + f_{row}\left(i\right) + f_{col}\left(j\right) + f_{time}\left(t\right),$$

Model II

$$log\left(\lambda_{ijkt}\right) = \alpha_0 + \alpha_1 \mathbf{1}_{\{k=1\}} + f_{row,col}\left(i,j\right) + f_{time}\left(t\right),$$

Model III

$$log\left(\lambda_{ijkt}\right) = \alpha_0 + \alpha_1 \mathbf{1}_{\{k=1\}} + f_{row,col,time}\left(i,j,t\right),$$

Model IV

$$log\left(\lambda_{ijkt}\right) = \alpha_0 + \alpha_1 \mathbf{1}_{\{k=1\}} + \sum_{l=0}^{1} \mathbf{1}_{\{k=l\}} f^{l}_{row,col,time}\left(i,j,t\right),$$

The models were estimated using the package `mgcv`, version 1.7-9, in the `R` environment, version 2.14.0. In all cases, B-splines were used as spline basis functions, with degree $p = 3$ and penalty in the $q = 2$-nd order derivative, and tensor product smoothers

were used to model the multidimensional functions. For the univariate smooth functions, $k = 8$ knots placed at equidistant points were selected, whereas for the multidimensional case, $k = 3$ knots were chosen for each of the marginal basis functions. Different number of knots were also informally checked, but the results showed that the previous choices were enough. The smoothing parameters were automatically selected based on the REML criterion, and the models were compared in terms of the (conditional) AIC (cAIC).

4.2. Results

Tables 1 and 2 show the results for the selection of models of cells FBH4 and FAP0 respectively. As can be observed, the cAIC indicates that for cell FBH4 the best model is that including the spatio-temporal interaction and the categorical covariate eye (Model III). However, in the case of cell FAP0 the best model is the most complex model, i.e., that including the spatio-temporal interaction and the interaction with eye (Model IV). Table 3 presents a detailed description of these fitted models. In both cases, the firing rate depends on the eye, with left eye presenting more activity (p-value $<$ 0.001 in both cases). This indicates that these cells show dominance of the left eye.

Figure 5(a) shows the change over time of the standard deviation of the estimated firing rates of an ON-RFmap of a cortical visual cell (FBH4). The abscissa represents pre-spike time in ms. As it can be seen there are important changes. The RFmap begins

Table 1: *Results for the selection of models of cell FBH4.*

Model	Formula in `mgcv`	cAIC	edf	Deviance explained (%)
I	eye + s(row, bs = 'ps') + s(col, bs = 'ps') + s(time, bs = 'ps')	31981.67	13.13	3.12
II	eye + te(row, col, bs = 'ps') + s(time, bs = 'ps')	32002.94	20.00	3.24
III	eye + te(row, col, time, bs = 'ps')	31502.67	98.04	7.15
IV	eye + te(row, col, time, by =eye, bs = 'ps')	31518.41	164.76	7.88

Deviance explained (%): Percentage of the null deviance (null.dev - deviance of the model with just one constant term) explained by the fitted model, i.e. (null.dev - dev)/null.dev, where dev is the deviance of the fitted model.

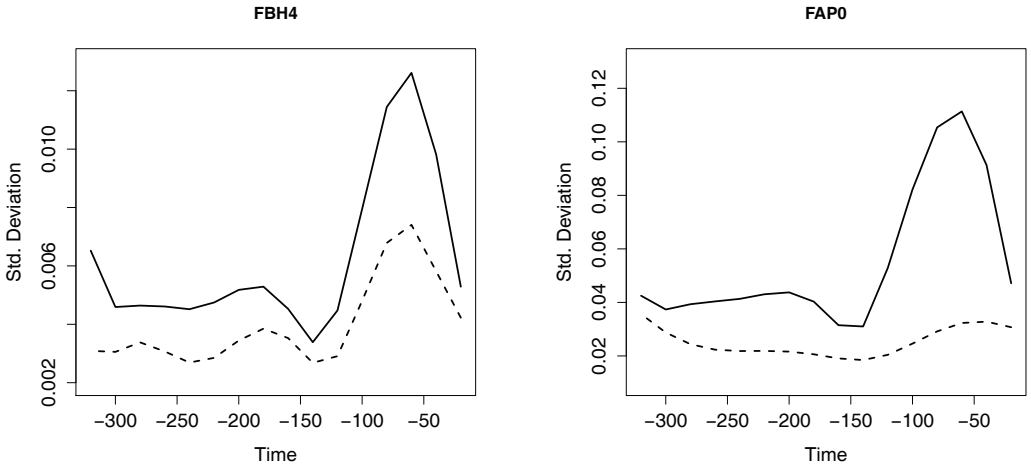
Table 2: *Results for the selection of models of cell FAP0.*

Model	Formula in `mgcv`	cAIC	edf	Deviance explained (%)
I	eye + s(row, bs = 'ps') + s(col, bs = 'ps') + s(time, bs = 'ps')	43248.77	12.04	16.1
II	eye + te(row, col, bs = 'ps') + s(time, bs = 'ps')	43207.59	21.61	16.6
III	eye + te(row, col, time, bs = 'ps')	42854.40	91.17	20.5
IV	eye + te(row, col, time, by =eye, bs = 'ps')	42831.73	138.94	21.5

s: smooth function
te: tensor product

Table 3: *Results of the fitted Models III and IV for cells FBH4 and FAP0 respectively.*

Term	Coefficient	edf	p-value
	Cell FBH4 - Model III		
Intercept	-3.448	—	< 0.001
eye_{Right}	-0.317	1	< 0.001
te(row, col, time)	—	96.04	< 0.001
	Cell FAP0 - Model IV		
Intercept	-0.485	—	< 0.001
eye_{Right}	-0.318	1	< 0.001
$\text{te(row, col, time)}_{Right}$	—	87.55	< 0.001
$\text{te(row, col, time)}_{Left}$	—	49.38	< 0.001



(*a*) (*b*)

Figure 5: *Change over time of the standard deviation of the estimated firing rates of the ON-RFmap of cells FBH4 and FAP0 for left (solid line) and right (dotted line) eye.*

to be structured about 140 ms pre-spike in both eyes and peaks approximately at 60 ms pre-spike, also in both eyes. Therefore the timing of the RFmap is similar for both eyes but the strength is different, indicating ocular dominance is constant over time. Figures 6 and 7 show a time series of ON-RFmaps obtained from a cortical visual cell (FBH4). As it can be seen, at about 200 ms pre-spike there is a central area with values below mean, and about 100 ms pre-spike a clear central area of values higher than the mean appears. This indicates that this area is the RF of the cell and that the optimal stimulus is a dot jumping from outside of the RF into the RF area.

Figure 5(b) is similar to Figure 5(a), however in this case the two cells studied show a different time-course of their RFmaps. The left eye stimulation produces a strong

RFmap with an onset at about 160 ms pre-spike and peaks at about 60 ms pre-spike. There is a clear dominance of the left eye, as it can be seen in Figures 8 and 9. Figure 8 shows the RFmaps of the non dominant eye (right eye). There is a very weak central area with almost no visible structure. Figure 9 shows the left ON-RFmap of the same cell shown in Figure 8. A clear central structure appears at 120 ms pre-spike and lasts until 20 ms prespike. The dominance of left eye induces stronger cell responses wich in turn produce stronger RFmap structures. As the cell of Figure 6, in this case there is a central area in the RFmaps from 220 to 180 ms prespike with low values, indicating that the cell response is better elicited when the spot jumps from outside the RF into the RF of the cell.

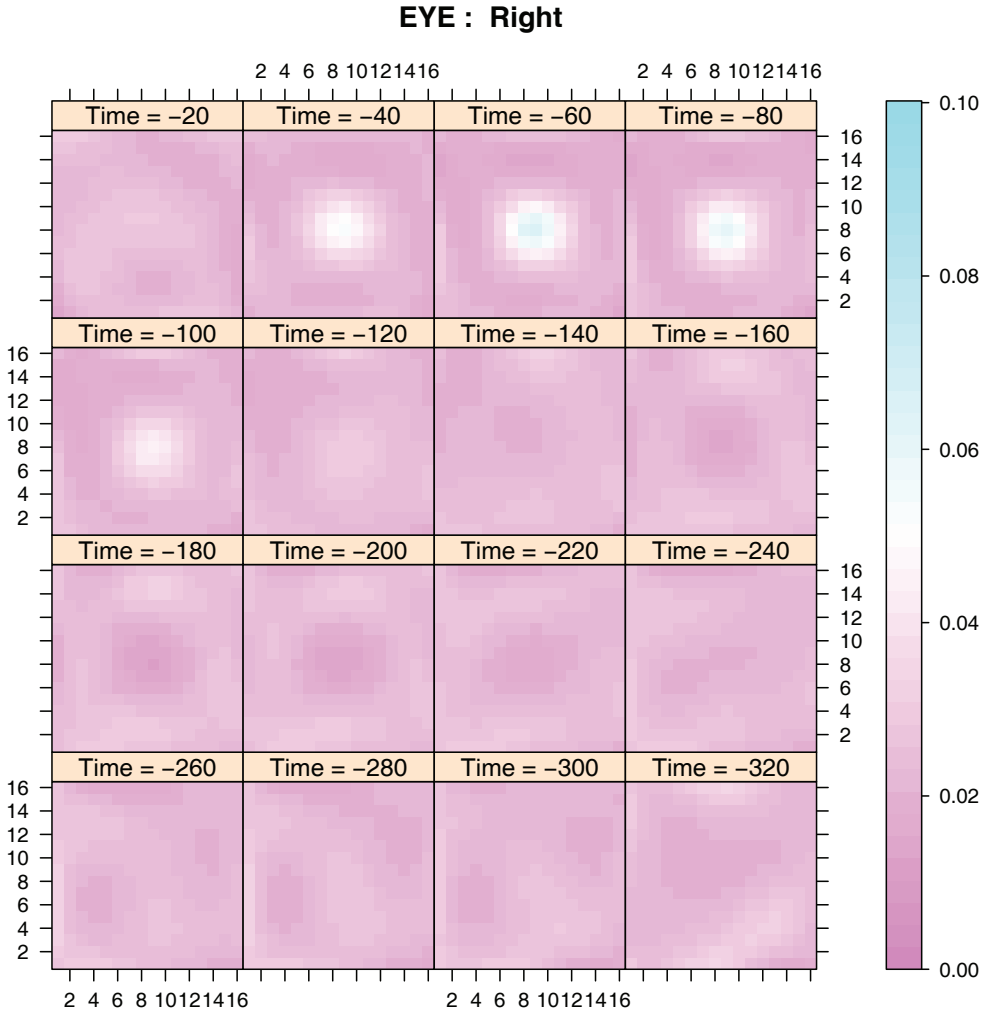


Figure 6: *Level plot of the estimated firing rates of the ON-RFmap based on Model III for the right eye of cell FBH4.*

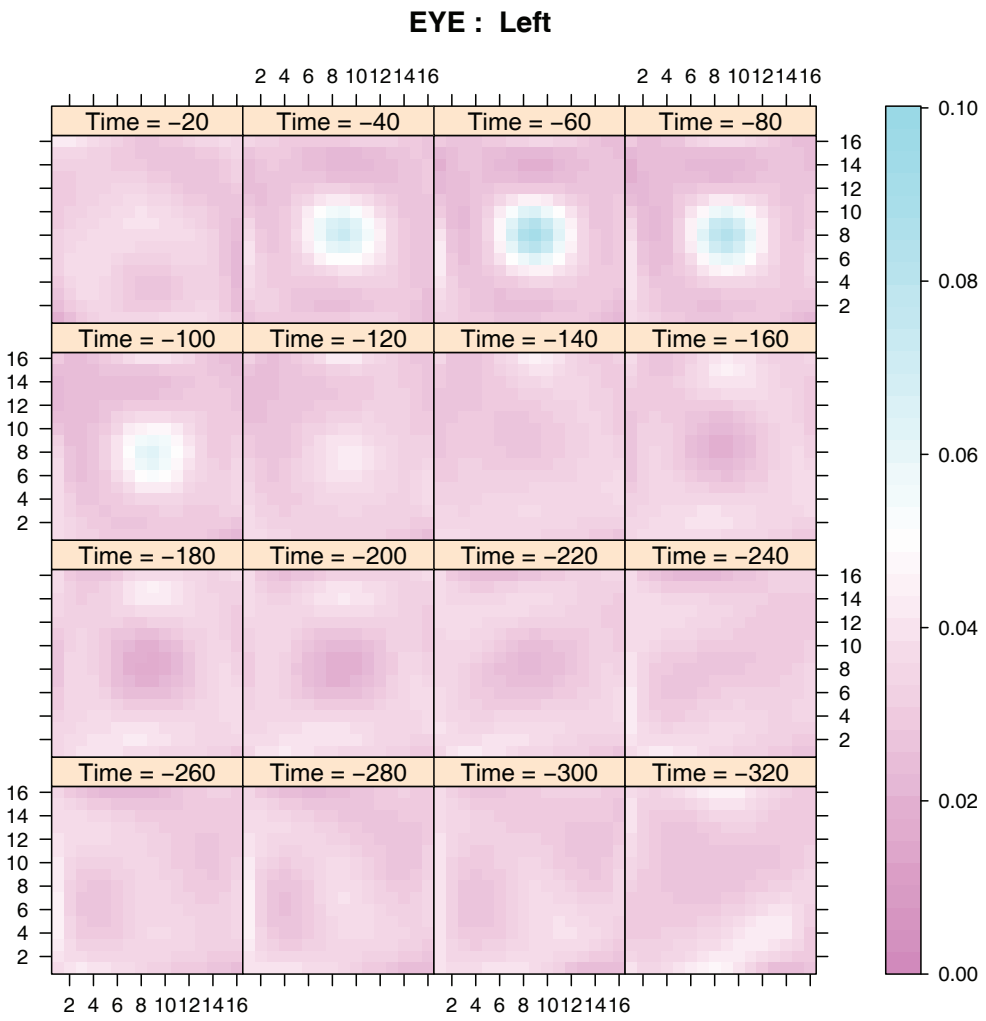


Figure 7: *Level plot of the estimated firing rates of the ON-RFmap based on Model III for the left eye of cell FBH4.*

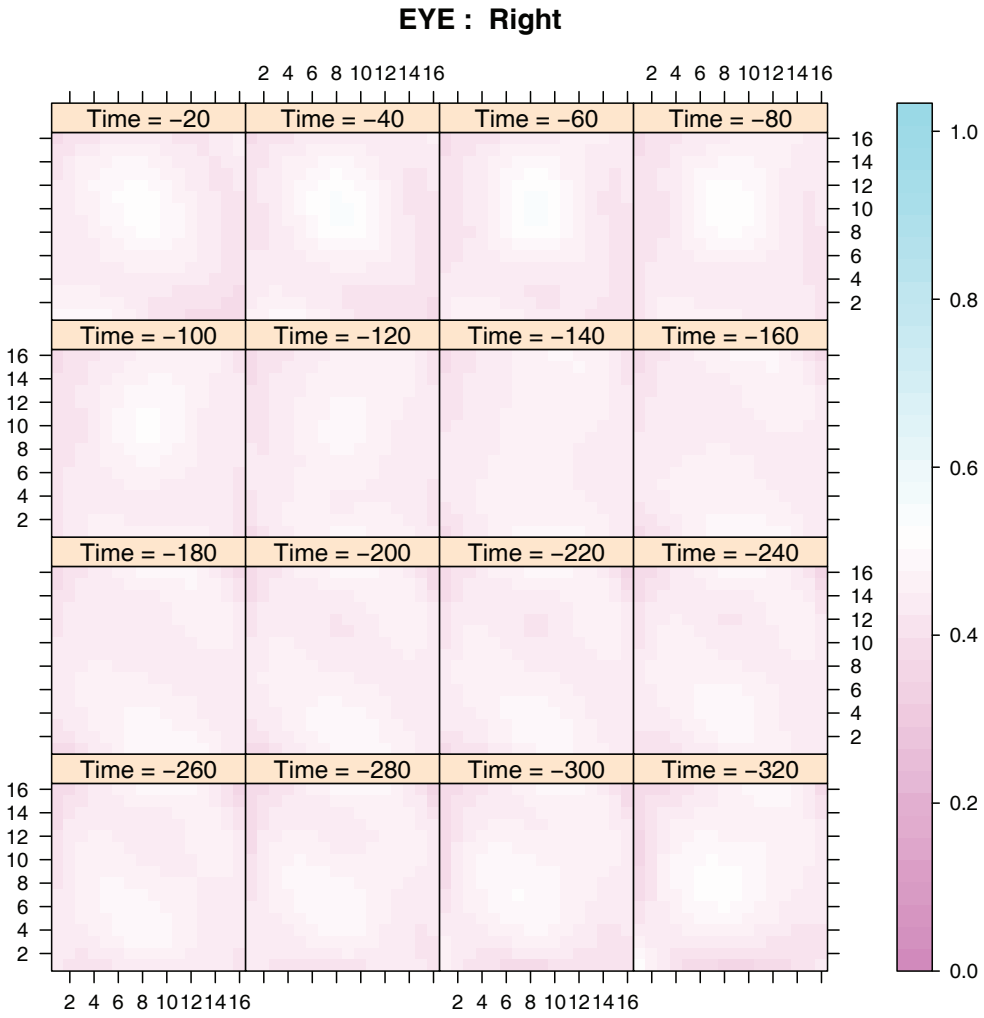


Figure 8: *Level plot of the estimated firing rates of the ON-RFmap based on Model IV for the right eye of cell FAP0.*

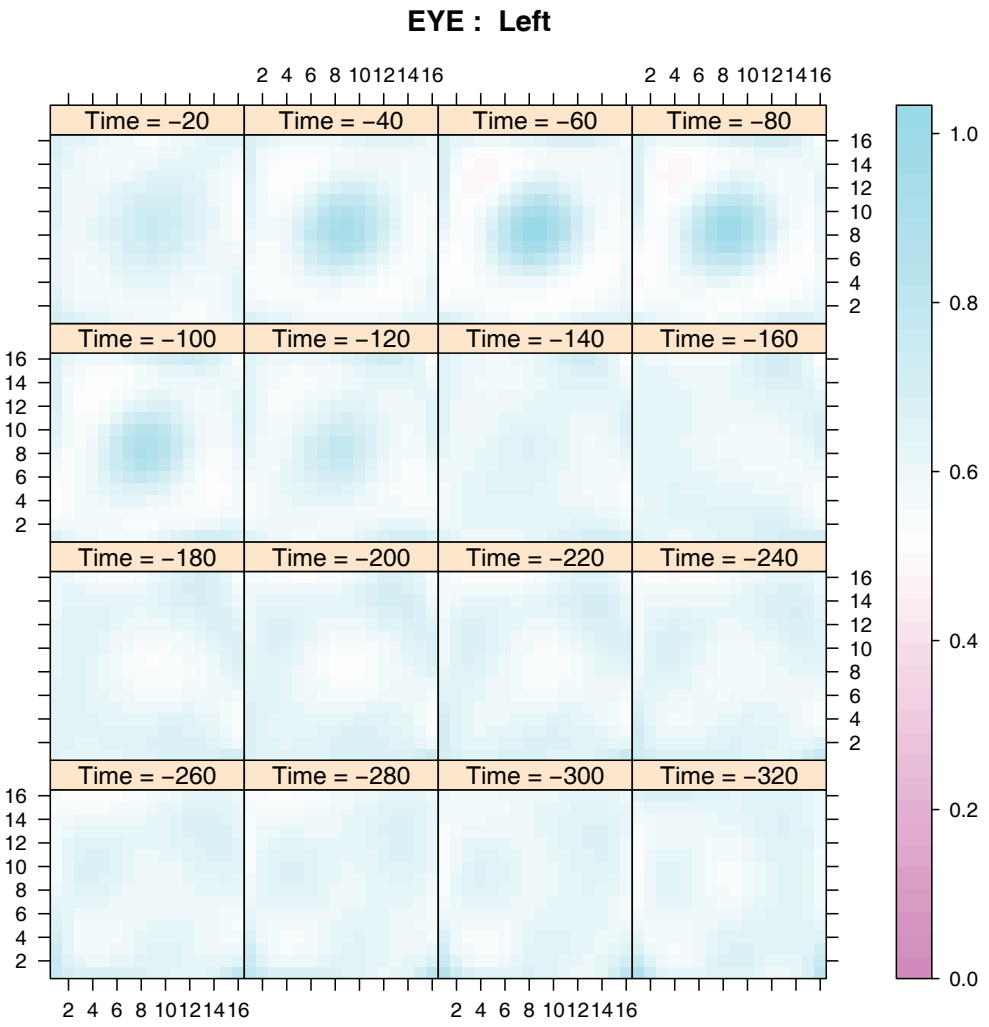


Figure 9: *Level plot of the estimated firing rates of the ON-RFmap based on Model IV for the left eye of cell FAP0.*

5. Conclusion

Receptive field mapping of cortical visual neurons is a critical procedure in neurophysiological studies of the visual system. In electrophysiological laboratories the reverse cross-correlation technique is frequently used to produce raw RFmaps – in the form of numerical matrices – which allow to give insights into the electrical activity of the visual neurons. However, the resulting RFmaps obtained from applying such technique, cannot be compared formally among different conditions.

In this paper, a Poisson GAM including interactions was proposed as a flexible statistical tool for (a) smoothing raw RFmaps of a single neuron, by including spatial effects in the model; and (b) estimating the temporal evolution of these maps, which in turn may vary across different experimental conditions.

The estimation algorithm used for fitting GAMs was based on P-splines and tensor product splines and the amount of smoothing was selected automatically via restricted (or residual) maximum likelihood (REML).

The proposed methodology was applied to study the activity of visual cells from area V1 (a primary visual cortical area). For the sake of illustration, in all the analyses perfomed in this paper, only visual neurons with response to the onset of a bright spot ('ON') were considered, and the temporal evolution of their RFmaps was compared for the right and left eye. However, it should be noted that our GAM methodology is very flexible and extensions to more complex interaction models are straightforward. This is important since physiologists are also interested in examining possible variations on RFmaps, when several levels of other covariates are considered. For example, comparisons between both monocular RFs and between monocular and binocular RFs are often required to assess the question on how the visual system handles binocular information. Therefore, GAM regression models designed to perform statistical comparisons between RFmaps would be very useful in electrophysiological experiments conducted in the visual system.

When calculating the time span of any RFmap, the onset and offset of the RFmap structures must be determined. In this paper, a relatively simple approach to solve this problem was proposed, and it consists of using the change of standard deviation of the matrix values over time. The authors are concisious, however, that a more refined analysis is needed, to extract accurate information from such RFmaps structures. Although it is beyond the scope of the current paper, it is an important topic for further research.

An `R` script implementing the nonparametric model estimation can be obtained by contacting the first author at maria.jose.rodriguez.alvarez2@sergas.es.

Acknowledgements

The authors would like to express their gratitude for the support received in the form of the MICINN (Spanish Ministry of Science and Innovation) grants MTM2008-01603 and DE2009-0030. Work of María Xosé Rodríguez-Álvarez was supported by grant CA09/0053 from the Instituto de Salud Carlos-III (Spanish Ministry of Science and Technology). Work of Francisco Gonzalez was supported by grant BFU-2010-14968 from the MCINN, Spain, FEDER, and 2010/PX142 from the Xunta de Galicia, Spain.

References

Barlow, H.B. (1953). Summation and inhibition in the frog's retina. *The Journal of Physiology*, 119, 69–88.

Behseta, S. and Kass, R. E. (2005). Testing equality of two functions using BARS. *Statistics in Medicine*, 24, 3523–3534.

Behseta, S., Wallstrom, G. L. and Kass, R. E. (2005). Hierachical models for assessing variability among functions. *Biometrika*, 92, 419–434.

Bellman, R. E. (1961). *Addaptive Control Processes*. Princenton University Press.

Bishop, P. O., Coombs, J. S. and Henry, G. H. (1973). Receptive fields of simple cells in the cat striate cortex. *The Journal of Physiology*, 231, 31–60.

de Boor, C. A. (2001). *A Practical Guide to Splines.* Revised Edition. Springer-Verlag, New York.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed model. *Journal of the American Statistical Association*, 88, 9–25. 116.

Brezger, A. and Lang, S. (2006). Generalized structured additive regression based on Bayesian P-splines. *Computational Statistics and Data Analysis*, 50, 967–991.

Brumback, B. A., Ruppert, D. and Wand, M. P. (1999). Variable selection and function estimation in additive nonparametric regression using a data-set prior: Comment. *Journal of the American Statistical Association*, 94, 794–797.

Cadarso-Suárez, C., Roca-Pardiñas, J., Molenberghs, G., Faes, C., Nácher, V., Ojeda, S. and Acuña, C. (2006). Flexible modelling of neuron firing rates across different experimental conditions. An application to neural activity in the prefrontal cortex during a discrimination task. *Applied Statistics*, 55, 431–447.

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions. *Numerische Mathematik*, 31, 377–403.

Currie, I. D. and Durbán, M. (2002). Flexible smoothing with P-splines: a unified approach. *Statistical Modelling*, 2, 333–349.

Currie, I. D., Durbán, M. and Eilers, P. H. C. (2006). Generalized linear array models with applications to multidimensional smoothing. *Journal of the Royal Statistical Society, Series B*, 68, 1–22.

DeAngelis, G. C., Ohzawa, I. and Freeman, R. D. (1993). Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. General characteristics and postnatal development. *Journal of Neurophysiology*, 69, 1091–1117.

DeBoer, E. and Kuyper, P. (1968). Triggered correlation. *IEEE Transactions on Biomedical Engineering*, 15, 169–179.

DiMatteo, I., Genovese, C. R. and Kass, R. E. (2001). Bayesian curve-fitting with free-knot splines. *Biometrika*, 88, 1055–1071.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11, 89–121.

Eilers, P. H. C. and Marx, B. D. (2003). Multivariate calibration with temperature interaction using two-dimensional penalized signal regression. *Chemometrics and Intelligent Laboratory Systems*, 66, 159–174.

Faes, C., Geys, H., Molenberghs, G., Aerts, M., Cadarso-Suárez, C., Acuña, C. and Cano, M. (2008). A Flexible Method to Measure Synchrony in Neuronal Firing. *Journal of the American Statistical Association*, 103, 149–161.

Fried, J. and Silverman, B. (1989). Flexible parsimonious smoothing and additive modelling. *Technometrics*, 31, 3–21.

Gonzalez, F., Krause, F., Perez, R., Alonso, J. M. and Acuña, C. (1993). Binocular matching in monkey visual cortex: single cells responses to correlated and uncorrelated dynamic random dot stereograms. *Neuroscience*, 52, 933–939.

Gonzalez, F. and Krause, F. (1994). Generation of Dynamic Random element stereograms in real time with a system based on a personal computer. *Medical and Biological Engineering and Computing*, 32, 373–376.

Gonzalez, F., Perez, R., Justo, M. S. and Bermudez, M. A. (2001). Receptive field organization of disparity-sensitive cells in macaque medial superior temporal cortex. *European Journal of Neuroscience*, 14, 167–173.

Hartline, H. K. and Ratliff, F. (1958). Spatial summation of inhibitory of inhibitory influences in the eye of Limulus, and the mutual interaction of receptor units. *The Journal of General Physiology*, 41, 1049–1066.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*. Monographs on Statistics and Applied Probability. Chapman & Hall.

Hastie, T. J. and Tibshirani, R. J. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society, Series B*, 55, 757–796.

Henry, G. H. and Bishop, P. O. (1972). Striate neurons: receptive field organization. *Investigative Ophthalmology*, 11, 357–368.

Hubel, D. H. and Wiesel, T. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160, 106–154.

Jones, J. P. and Palmer, L. A. (1987). The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58, 1187–1211.

Kass, R. E., Ventura, V. and Cai, C. (2003). Statistical smoothing of neuronal data. *Network: Computation in Neural Systems*, 14, 5–15.

Krause, F., Gonzalez, F., Nelson, J. I. and Eckhorn, R. (1987). A fast method for predicting coding properties of visual cortical simple cells. *Perception*, 16, 2–269 (A45).

Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology*, 16, 37–68.

Lang, S. and Brezger, A. (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics*, 13, 183–212.

Lee, T. C. M. (2002). On algorithms for ordinary least squares regression spline fitting: a comparative study. *Journal of Statistical Computation and Simulation*, 72, 647–663.

Lee, D.-J. and Durbán, M. (2011). P-spline ANOVA-type interaction models for spatio-temporal smoothing. *Statistical Modelling*, 11, 49–69.

Marra, G. and Radice, R. (2010). Penalised regression splines: theory and application to medical research. *Statistical Methods in Medical Research*, 19, 107–125.

Marx, B. D. and Eilers, P. C. H. (1998). Direct generalized additive modeling with penalized likelihood. *Computational Statistics and Data Analysis*, 28, 193–209.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Monographs on Statistics and Applied Probability. Second Edition. Chapman & Hall/CRC.

Nishimoto, S., Ishida, T. and Ohzawa, I. (2006). Receptive field properties of neurons in the early visual cortex revealed by local spectral reverse correlation. *The Journal of Neuroscience*, 26, 3269–3280.

O'Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems. *Statistical Science*, 1, 502–518.

Pei, X., Vidyasagar, T. R., Volgushev, M. and Creutzfeldt, O. D. (1994). Receptive field analysis and orientation selectivity of postsynamptic potentials of simple cells in cat visual cortex. *The Journal of Neuroscience*, 14, 7130–7140.

Pérez, R., Castro, A. F., Justo, M. S., Bermúdez, M. A. and Gonzalez, F. (2005). Retinal correspondence of monocular receptive fields in disparity-sensitive complex cells from area V1 in the awake monkey. *Investigative Ophthalmology & Visual Science*, 46, 1533–1539.

R Development Core Team. R: A language and environment for statistical computing (2011). R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Reid, R. C., Victor, J. D. and Shapley, R. M. (1997). The use of m-sequences in the analysis of visual neurons: linear receptive field properties. *Visual Neuroscience*, 14, 1015–1027.

Reiss, P. T. and Ogden, R. T. (2009). Smoothing parameter selection for a class of semiparametric linear models. *Journal of the Royal Statistical Society, Series B*, 71, 505–524.

Roca-Pardiñas, J., Cadarso-Suárez, C., Nácher, V. and Acuña, C. (2006). Bootstrap-based methods for testing factor-by-curve interactions in generalized additive models: assessing prefrontal cortex neural activity related to decision-making. *Statistics in Medicine*, 25, 2483–2501.

Roca-Pardiñas, J., Cadarso-Suárez, C., Tahoces, P. G. and Lado, M. J. (2008). Assessing continuous bivariate effects among different groups through nonparametric regression models: An application to breast cancer detection. *Computational Statistics and Data Analysis*, 52, 1958–1970.

Roca-Pardiñas, J., Cadarso-Suárez, C., Pardo-Vazquez, J. L., Leboran, V., Molenberghs, G., Faes, C. and Acuña, C. (2011). Assessing neural activity related to decision-making through flexible odds ratio curves and their derivatives. *Statistics in Medicine*, 14, 1695–1711.

Strasak, A. M., Umlauf, M., Pfeiffer R. M. and Lang, S. (2011). Comparing penalized splines and fractional polynomials for flexible modelling of the effects of continuous predictor variables. *Computational Statistics and Data Analysis*, 55, 1540–1551.

Tsao, D. Y., Conway, B. R. and Livingstone, M. S. (2003). Receptive fields of disparity-tuned simple cells in macaque V1. *Neuron*, 38, 103–114.

Wahba, G. (1990). *Spline Models for Observational Data*. Philadelphia: SIAM.

Wood, S. N. (2000). Modelling and smoothing parameter estimation with multiple quadratic penalties. *Journal of the Royal Statistical Society, Series B*, 62, 413–428.

Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society, Series B*, 65, 95–114.

Wood, S. N. (2004). Stable and efficient multiple smoothing parameter for generalized additive models. *Journal of the American Statistical Association*, 99, 673–686.

Wood, S. N. (2006a). *Generalized Additive Models. An Introduction with R*. Chapman & Hall/CRC.

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

Wood, S. N. (2008). Fast stable direct fitting and smoothness selection for generalized additive models. *Journal of the Royal Statistical Society, Series B*, 70, 495–518.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society, Series B*, 73, 3–36.

Discussion of "Analysing visual receptive fields through generalised additive models with interactions" by María Xosé Rodríguez-Álvarez, Carmen Cadarso-Suárez and Francisco Gonzalez

María L. Durbán
Universidad Carlos III de Madrid

It is a pleasure for me to comment on the paper by Rodriguez-Alvarez, Cadarso-Suarez and Gonzalez. I would like to thank the editors for their invitation. This paper proposes a Poisson generalized additive model (GAM) for smoothing visual receptive fields (RFs) over time, and uses multidimensional interactions to compare data under several experimental conditions. As the authors point out, the use of high-dimensional smoothing for the analysis of spatial and spatio-temporal data has been a subject of great interest in recent years (specially in the context of P-splines). Also, the possibility of fitting a factor by curve interaction has made these type of models even more flexible, and capable of capturing very complex structures present in the data (Durbán et al. (2005)). The authors go one step further and use multidimensional interactions, and use a unified approach for: i) spatio-temporal modelling, and ii) comparison across different experimental conditions. The models proposed and the results obtained in the paper unavoidably give rise to several related and interesting issues, which may help to improve and extend the scope of the models and their application.

One of the main problems encountered when multidimensional smoothing is used is the computational time that the fitting of the model might need. P-splines, and in general, low-rank smoothers, are the best approach to take. However, even in this case, great care needs to be taken when choosing the size of the bases used in the model. The authors have, obviously, come across this problem. I am concerned with their choice of size for the bases. First, the fact that the size of the bases is different in the univariate case ($k = 8$), and in the multivariate case ($k = 3$) result in models that are not nested (then, care must be taken when comparing them). To ensure that they are, bases for multidimensional smooth terms (models II, III and IV) should be constructed using marginal bases identical to those used for the main effects (model I). But, sometimes, using bases of the same size results in models that can be computationally very demanding, or even, impossible to fit. One possible solution is to reduce the number of parameters for the interaction term, by reducing the number of knots for the marginal bases for the tensor product, but preserving the nested nature of the models. This idea can be explained by analogy to classical ANOVA models, where, in general, the main effects are more significant than interactions. Lee and Durbán (2012) propose the use of nested B-spline bases for the interaction term: a B-spline basis such that the space spanned by the marginal basis in the interaction, is a subset of the space spanned by the basis in the unidimensional case and the hierarchical nature of the models is preserved. Another point related to the bases size is the fact that, using three knots for marginal bases might

not be enough to capture the structure of the firing rates. I agree that larger bases are difficult to handle by the package `mgcv` (specially when the data have an array structure). However, as a general rule (Ruppert, 2002), the minimum number of knots used to construct a basis is at least four. The results for generalized linear array models (GLAMs), in Eilers et al. (2006), can speed up the computation when the size of the bases increases.

A major contribution of the paper is the development of a model to compare the receptive fields properties by introducing interaction between factor and multidimensional smooth terms, and check, for example, if the firing rate depends on eye. Models III and IV test if the spatio-temporal effects are similar in both eyes or not, but do not separate the spatial and temporal component. It might be interesting to know if the change of firing rates over time is the same for both eyes, or to estimate a common spatial pattern in time and check if a space-time interaction is necessary. Maybe an ANOVA-type model (Lee and Durbán (2011)) could shed light on some interesting issues. Thus, the spatio-temporal term in model III and IV could be substituted by:

$$f_{\text{row, col, time}}(i,j,t) \Rightarrow f_{\text{time}} + f_{\text{row, col}}(i,j) + f_{\text{row, col, time}}(i,j,t),$$

and appropriate penalties and identifiability constraints can be easily imposed.

Let me finish by congratulating again the authors for an interesting and relevant paper for the statistical and medical scientific community.

References

Durbán, M., Harezlak, J., Wand, M. P. and Carroll, R. J. (2005). Simple fitting of subject-specific curves for longitudinal data. *Statistics in Medicine*, 24, 1153–1167.

Eilers, P. H. C., Currie, I. D. and Durbán, M. (2006). Fast and compact smoothing on large multidimensional grids. *Computational Statistics and Data Analysis*, 50, 61–76.

Lee, D-J. and Durbán, M. (2011). P-spline ANOVA-type interaction models for spatio-temporal smoothing. *Statistical Modelling*, 11, 49–69.

Lee, D-J. and Durbán, M. (2012). Efficient two-dimensional smoothing with P-spline ANOVA mixed models and nested bases (submitted).

Ruppert, D. (2002). Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics*, 11, 735–757.

Acknowledgements

The work of Maria Durbán is supported by the the Spanish Ministry of Science and Education (projects MTM 2008-02901 and MTM2011-28285-C02-02).

Thomas Kneib
Georg-August-Universität Göttingen

Thank you very much for the possibility to discuss this paper that introduces and discusses additive model including different types of interactions along the analysis of visual receptive. The authors do a remarkable job in reviewing the current state of the literature and in demonstrating the applicability of the methods in a complex setting that requires careful inclusion of interaction effects in different ways. Penalized spline smoothing forms the basis for the models considered and mixed model based inference for the smoothing parameters yields a data-driven amount of smoothness for all functions. My comments mainly focus on the use of mixed model methodology (and potential alternatives) and some specific modelling choices.

- As one advantage of the mixed model approach, the authors discuss the possibility of a formal comparison across models of different complexity. However, inference in mixed models has proved to be notoriously difficult due to deviations from regularity conditions underlying standard likelihood based inferential approaches. For example, likelihood ratio tests for the inclusion / exclusion of specific effects approached via testing smoothing variances for deviations from zero have been shown to lead to rather complex distributions for the test statistic that deviate considerably from the standard result that would be a χ^2 distribution with one degree of freedom (Crainiceanu and Ruppert 2004, Crainiceanu, Ruppert, Claeskens and Wand 2005, Scheipl, Greven and Küchenhoff 2008). In a similar vein, Greven and Kneib (2010) show that both marginal AIC and the conditional AIC (in their standard forms) are problematic when comparing models that differ by the inclusion / exclusion of specific effects obtained by setting the smoothing variance to zero. While this is not exactly the type of model comparison considered here, I wonder how generally applicable formal comparisons in the considered model class can be and how the authors suggest to deal with such problems.
- A related question concerns the performance of model choice and the p-values presented in Table 3. Given the difficulties discussed in the previous comment, more guidelines on model choice and details on how the p-values in Table 3 have been computed would be very helpful from the perspective of practitioners.
- When using varying coefficient terms for modelling interactions, the coding of the (categorical) interaction variable will strongly influence the results. For example, the model

$$y_i = \sum_{l=1}^{M} \mathbb{1}_{z_i=l} f_l(x_i) + \varepsilon_i$$

models separate curves per each of the M levels of the categorical covariate z while

$$y_i = f_1(x_i) + \sum_{l=2}^{M} \mathbb{1}_{z_i=l} g_l(x_i) + \varepsilon_i$$

models the effect in category 1 as $f_1(x)$ and deviations from this effect in remaining groups $l = 2,\ldots,M$ as $g_l(x)$. While both models would be equivalent when restricting $f_l(x)$ and $g_l(x)$ to be linear, this does no longer hold in the penalised spline smoothing context due to the estimation of the smoothing parameter that is not invariant under such transformations. In addition, when using effect coding instead of dummy coding for the categorical covariates, another set of results would be obtained. How would the authors deal with this difficulty and what would be their general recommendations for choosing a specific parameterisation?

- I wonder whether (spatial) smoothing is really desired in this application. I would presume that there may be sharp edges around the receptive fields while the given approach assumes a smooth transition from active to inactive cells. As a consequence, edge-preserving or adaptive procedures may be more appropriate at least for the effect representing the receptive field. In a Bayesian formulation, adaptiveness could for example be achieved by making the smoothing variance depending on the spatial location (see Lang and Brezger (2004) for such an approach).
- For the spatial effect, radial basis functions may also be a useful alternative to the bivariate P-splines proposed here since P-splines may induce some artificial structure in the estimates due to their inherent non-radiality. Although the amount of non-radiality is decreasing with the spline degree and may therefore be negligible for the cubic splines considered here, it would be good to have some general advice on the selection of either radial bases or bivariate penalized splines.

References

Crainiceanu, C. and Ruppert, D. (2004). Likelihood ratio tests for goodness-of-fit of a nonlinear regression model. *Journal of Multivariate Analysis*, 91, 35–52.

Crainiceanu, C., Ruppert, D., Claeskens, G. and Wand, M. P. (2005). Exact likelihood ratio tests for penalised splines. *Biometrika*, 92, 91–103.

Greven, S. and Kneib, T. (2010). On the behavior of marginal and conditional Akaike information criteria in linear mixed models. *Biometrika*, 97, 773–789.

Lang, S. and Brezger, A. (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics*, 13, 183–212.

Scheipl, F., Greven, S. and Küchenhoff, H. (2008). Size and power of tests for a zero random effect variance or polynomial regression in additive and linear mixed models. *Computational Statistics & Data Analysis*, 52, 3283–3299.

Rejoinder

First of all, we would like to thank the invited discussants for the time spent discussing our work and for all their valuable comments and sugggestions made on our paper.

1. Comments from Prof. Thomas Kneib

Prof. Thomas Kneib primarily centres his comments on various problems arising in the context of the inference in Generalised Additive Models. Although Prof. Kneib mainly focuses his comments on the mixed model methodology, similar difficulties can be encountered as well using different approaches and smoothing parameters selectors. We agree with the discussant that this is an area of enormous interest and active research, not yet completely solved. We are conscious that our paper lacks an explicit discussion of this point, and the use of the word 'formal' probably overstates the main objective of the paper. We would like to thank Prof. Kneib for his valuable comments on this issue and for giving us the opportunity of shedding more light on this challenging point. However, we believe that this issue probably requires a new paper.

With respect to model comparisons and model choice (in the frequentist framework), in the context of GAMs (and GLMs), two different approaches are usually applied, depending on if the models are nested or not (see also the comments from Prof. María Durbán). If the models are not nested, the comparison between models can be based on some information criteria such as the Akaike information criterion (AIC) or the Bayesian information criterion (BIC). If the models are nested, the generalised likelihood ratio test or the F-ratio test can be applied. However, in GAMs the distribution of these test statistics under the null hypothesis is only approximate (due to the fact that the smoothing parameters are treated as if they are known), and the obtained p-values should be analised with caution (specially when they are close to the significance level) (Wood, 2006a). As pointed out by Prof. Simon Wood (in the help file of the functions `mgcv::summary.gam` and `mgcv::anova.gam`), various simulations studies suggested that the p-values obtained under ML and REML smoothness selection, have the best behaviour. This cautionary note can also be extended to the inference about each smooth term in a model, and the way in how p-values are computed. In this case, two different approaches can be used, based on the frequentist or the Bayesian covariance matrix of the coefficient estimates (Wood, 2006a). If the objective is testing for smooth terms in a GAM for equality to zero, the frequentist approach can be used (see Marra and Radice, 2010; Wood, 2006a). However, even in this case the performance of the

obtained p-values is not good, in terms of type I error. The Bayesian approach is then recommended in this case (as for the construction of the 'confidence' intervals; see the help of the `mgcv` package). Therefore, the p-values reported in this paper were based on the bayesian covariance matrix. Again, the (Bayesian) p-values obtained under ML and REML smoothness selection present the best behaviour. This is in accordance with the results of a small simulation study performed by us, in which type I errors (using REML) proved to be relatively close to nominal errors.

Under the representation of an additive model as a linear mixed model, some other alternatives (and difficulties) for model choice appear. In this framework, the random effects u parameterise the deviation of the smooth function from a given polynomial (depending on the degree p of the spline basis). Therefore, testing for a polynomial (or constant) function versus a smooth function is equivalent to test if the corresponding smoothing variance is zero ($\sigma_u^2 = 0$). As pointed out by Prof. Kneib, this field has received a considerably amount of research attention in the last years. For instance, in Scheipl et al. (2008) a comparative study of different restricted likelihood ratio tests (RLRT) and F-type tests for a zero variance (also in the context of penalised splines) is presented. The authors conclude that the RLRT statistic proposed by Crainiceanu and Ruppert (2004) – and extended by Greven et al. (2008) to test for more than one variance component – presents the best behaviour in terms of the computational time, with type I error rates and power almost equivalent to the bootstrap-based tests. Moreover, this test is implemented in the `R` package `RLRSim`. However, for the model comparisons performed in this paper, the use of the RLRT statistics present in this package is not possible unless we consider nested models and isotropic interactions (see Lee 2010), and, moreover, these tests may not be applicable to GAM. As regards the use of the AIC for comparison purposes, under the mixed model representation, two different alternative definitions can be used: the marginal and the conditional AIC. The conditional AIC (in its standard form) is the AIC supplied by the `mgcv` package, and its use is 'recommended' in the context of penalised splines. However, Greven and Kneib (2010) showed that the standard form of the conditional AIC has an undesirable performance, in the sense that it always chooses the inclusion of the random effect u, unless u is predicted to be exacly zero ($\hat{\sigma}_u^2 = 0$). Although the comparisons made in Greven and Kneib (2010) – where the authors compare the presence of a linear effect versus a smooth effect – are not the type of model comparison considered in our paper, it would be worthwhile evaluating the performance of the conditional AIC in this setting. Moreover, the extension of the corrected version of the conditional AIC Liang et al. 2008, Greven and Kneib 2010) to generalised linear mixed models and GAMs is is an open area of research.

The comment from Prof. Kneib related to the use of varying cofficient terms is connected to this previous issue. From an applied point of view, the decision about the coding of the categorical variable that should be used depends on both the problem at hand and the research question to be answered. On the one hand the use of the dummy coding allows to separately evaluate the effect of the continuous covariate in

each of the levels of the categorical covariate. However, this coding makes it difficult to 'visually' judge the presence of interaction from several curves that are estimated separately. On the other hand, when the interest is focused on finding out whether there actually is a difference among several groups, it may be advantageous to use the effect parameterisation ($y_i = f(x_i) + \sum_{l=1}^{M} \mathbf{1}_{\{z_i=l\}} g_l(x_i) + \varepsilon_i$, given that $\sum_{l=1}^{M} g_l(x_i) = 0$). In this case, the deviations from the main effect, $f(x_i)$, in each of the levels of the categorical covariate, $g_l(\cdot)$ ($l = 1, \ldots, M$), exactly tells us the amount of difference. Moreover, this approach also allows to evaluate the presence of interaction based on the inference about $g_l(\cdot)$, ($l = 1, \ldots, M$). If these functions are equal to zero, it implies that no interaction is present. The dummy coding is the way in which the `mgcv` package treats, by default, the interaction between continuous and categorical covariates, and this was the coding used in the paper. However, the use of the effect coding can also be implemented in the `mgcv` package, by first creating the effect coding variable. For illustration purposes, we have analysed the visual receptive field data (Model IV) using this approach:

```
R > fbh4$eff_eye <- as.numeric((fbh4$eye == 'Right') - (fbh4$eye == 'Left'))
R > fit.fbh4 <- gam(spikes ~ eye + te(row, col, time, by = eff_eye, bs = 'ps')
+   + offset(log(trial)), data = fbh4, family = poisson, method = 'REML')
```

and

```
R > fap0$eff_eye <- as.numeric((fap0$eye == 'Right') - (fap0$eye == 'Left'))
R > fit.fap0 <- gam(spikes ~ eye + te(row, col, time, by = eff_eye, bs = 'ps')
+   + offset(log(trial)), data = fap0, family = poisson, method = 'REML')
```

Table 1: *Results of the fitted models for cells FBH4 and FAP0.*

Term	Coefficient	edf	p-value
Cell FBH4 - Model IV			
Intercept	-3.466	—	< 0.001
eye_{Right}	-0.283	1	< 0.001
f(row, col, time)	—	96.00	< 0.001
g_{Right}(row, col, time)	—	20.13	0.107
Cell FAP0 - Model IV			
Intercept	-0.480	—	< 0.001
eye_{Right}	-0.330	1	< 0.001
f(row, col, time)	—	88.83	< 0.001
g_{Right}(row, col, time)	—	24.86	< 0.001

Table 1 presents a detailed description of these fitted models. It should be noted that in this case, due to the identifiability constraints imposed on the model, $g_{Right}(\cdot) = -g_{Left}(\cdot)$. As can be observed, whereas for cell FBH4 there is no evidence to suggest the presence of interaction with eye (p-value = 0.107), for cell FAP0 the results suggest that the RFmap is different in the left and right eye (p-value < 0.001). This is in

concordance with the results obtained using the dummy coding and the cAIC for comparison purposes.

As regards the use of radial basis for the spatial effect, we agree with Prof. Kneib that this approach could have been also used for the visual receptive field analysis in our paper. In contrast to tensor product smoothers, where the knots are placed separately in each direction (row and column in our example), radial bases (as thin plate splines, for example) consider bivariate knots located in the surface. Therefore, one of the main advantages of radial bases is that they allow to adapt the placement of knots to the data (i.e., placing more knots in areas with dense data), and therefore to take into account the possible correlation between the covariates (see Fahrmeir and Kneib 2011 for a detailed comparison between tensor product and radial basis). However, in contrast to tensor product smoothers, radial bases assume the same amount of smoothing in each direction. As pointed out in the paper, this isotropy could be justified when the covariates are measured in the same units (and, of course, the same amount of smoothing is expected in each direction). This is the case, for instance, of spatial effects, but not of the spatio-temporal interaction, due to the different scaling of time and space. For this reason, the use of radial bases is not recomended in this context. However, the `mgcv` package allows to combine in a trivariate smoother, the tensor product of a bivariate radial basis and a one dimensional smoother. We would like to point out that the visual receptive field data was also analised using this approach. Since the results provided by this model and the model using the tensor product of univariate smoothers were pretty similar, we decided to present in the paper the results of the trivariate tensor product.

Finally, as pointed out by Prof. Kneib the use of adaptive penalized splines could be more appropiate for this application, as can be observed from the figures. So far, the `mgcv` package allows the use of locally adaptive procedures but in the context of radial bases, and cannot be applied for tensor product smoothers. Although, as pointed out before, this `R` package allows to combine the tensor product of a bivariate radial basis and a one dimensional smoother, it is not possible to use adaptive smoothers for the bivariate radial basis. Moreover, as far as we know, no statistical software implements the use of locally adaptive procedures for trivariate smoothers. This is an interesting field for further research.

2. Comments from Prof. María Durbán

Prof. María Durbán comments on the importance of the selection of the basis sizes, due to its impact on: (a) the final estimates (if the basis dimension is not large enough), (b) the computational time; and (c) the inference procedures.

As pointed out by María Durbán, from a computational point of view, as the number of knots increases (specially in the multidimensional case), so it does the computational time. For instance, using an Intel(R) Core (TM) i5 CPU 2.40 GHz and 4.00 GB RAM computer, the estimation of model IV using $k = 3$ or $k = 4$ was around 100 sec. and 403

sec. respectively. For this reason, we chose $k = 3$ knots for each of the marginal basis functions in the multidimensional case. However, in order to evaluate the adequacy of the basis dimensions, we performed a sensitivity analysis by increasing the number of knots, and comparing the results obtained by the different fitted models. Since in this case the observed differences were almost negligible, we finally selected $k = 3$ knots. Moreover, we also evaluated the adequacy of the basis dimensions based on an analysis of the deviance residuals of the fitted model (see the help associated with the function `choose.k` in the `mgcv` package). For each smooth term in our models (I- IV), we fitted an equivalent, single, smooth to the residuals, using a larger number of knots to see if there were still a pattern in the residuals that could potentially be explained by increasing the basis dimension in the original models. Again, these analyses suggested that $k = 8$ and $k = 3$ were enough for the univariate and the multidimensional cases respectively. However, we are conscious that this could not be the always the case, and in some circumstances, the need of larger bases could make prohibitive the fit of the model using the package `mgcv`. As pointed out by María Durbán, when the data are in an array structure, as in the case of the visual receptive field data, the use of GLAMs can make the fit of the model feasible in a reasonable computing time. As far as we know, neither the `mgcv` package nor other `R` packages and statistical software implement this approach. This is, indeed, an interesting field for further work.

As Prof. Durbán indicates, the use of different sizes of the bases for the univariate and the multivariates case, results in models that are not nested. Accordingly, it is not possible to apply formal testing procedures, as the generalised likelihood ratio test or the F-ratio test, to make the comparison between alternative models. Although in this paper we have used the conditional AIC for comparison purposes, the use of formal formal testing procedures could be also an alternative approach, ensuring that the models are nested. Moreover, the ANOVA-type models proposed by Lee and Durbán (2011) is an interesting approach to study in future. We completely agree with Prof. Durbán that the use ANOVA-type models could help in a better understanding of the Visual receptive field behaviour. It should be noted, however, that in the visual receptive field data, the experiment design assigns each spike to the stimulus position in all pre-spike times. It implies that each spatial matrix contains exactly the same number of spike ocurrences. Therefore, the spatio-temporal interaction is probably needed in this context.

References

Crainiceanu, C. and Ruppert, D. (2004). Likelihood ratio tests for goodness-of-fit of a nonlinear regression model. *Journal of Multivariate Analysis*, 91, 35–52.

Fahrmeir, L. and Kneib, T. (2011). *Bayesian Smoothing and Regression for Longitudinal, Spatial and Event History Data*. Oxforf University Press.

Greven, S., Crainiceanu, C. M. and Küchenhoff, H. (2008). Restricted likelihood ratio testing for zero variance components in linear mixed models. *Journal of Computational and Graphical Statistics*, 17, 870–891.

Greven, S. and Kneib, T. (2010). On the Behavior of Marginal and Conditional Akaike Information Criteria in Linear Mixed Models. *Biometrika*, 97, 773–789.

Lee, D.-J. (2010). Smothing mixed model for spatial and spatio-temporal data. PhD thesis, Department of Statistics, Universidad Carlos III de Madrid, Spain.

Lee, D.-J. and Durbán, M. (2011). P-spline ANOVA-type interaction models for spatio-temporal smoothing. *Statistical Modelling*, 11, 49–69.

Liang, H., Wu, H. and Zou, G. (2008). A note on conditional AIC for linear mixed-effects models. *Biometrika*, 95, 773–778.

Marra, G. and Radice, R. (2010). Penalised regression splines: theory and application to medical research. *Statistical Methods in Medical Research*, 19, 107–125.

Scheipl, F., Greven, S. and Küchenhoff, H. (2008). Size and power of tests for a zero random effect variance or polynomial regression in additive and linear mixed models. *Computational Statistics & Data Analysis*, 52, 3283–3299.

Wood, S. N. (2006). *Generalized Additive Models. An Introduction with R*. Chapman & Hall/CRC.